

Practical:
Sequence searches, protein families,
function annotation

Protein families

- Members will share a structural core (homology modeling)
- Members may share aspects of function (function prediction)
- The whole set of members may reveal elements of protein and organism evolution (phylogenies)

Protein family databases (most but not all try to build families for domains)

Functions, organisms, structures

Marco Punta


Prokaryotes both domains and full-length equivalents



Signalling, extracellular and chromatin-associated proteins



Structural domains from SCOP

Gene3D
Structural domains from CATH

No limits, domains

Marco Punta
■ ■ ■ ■ ■



No limits, full-length proteins

Marco Punta
— ■ —



Integration

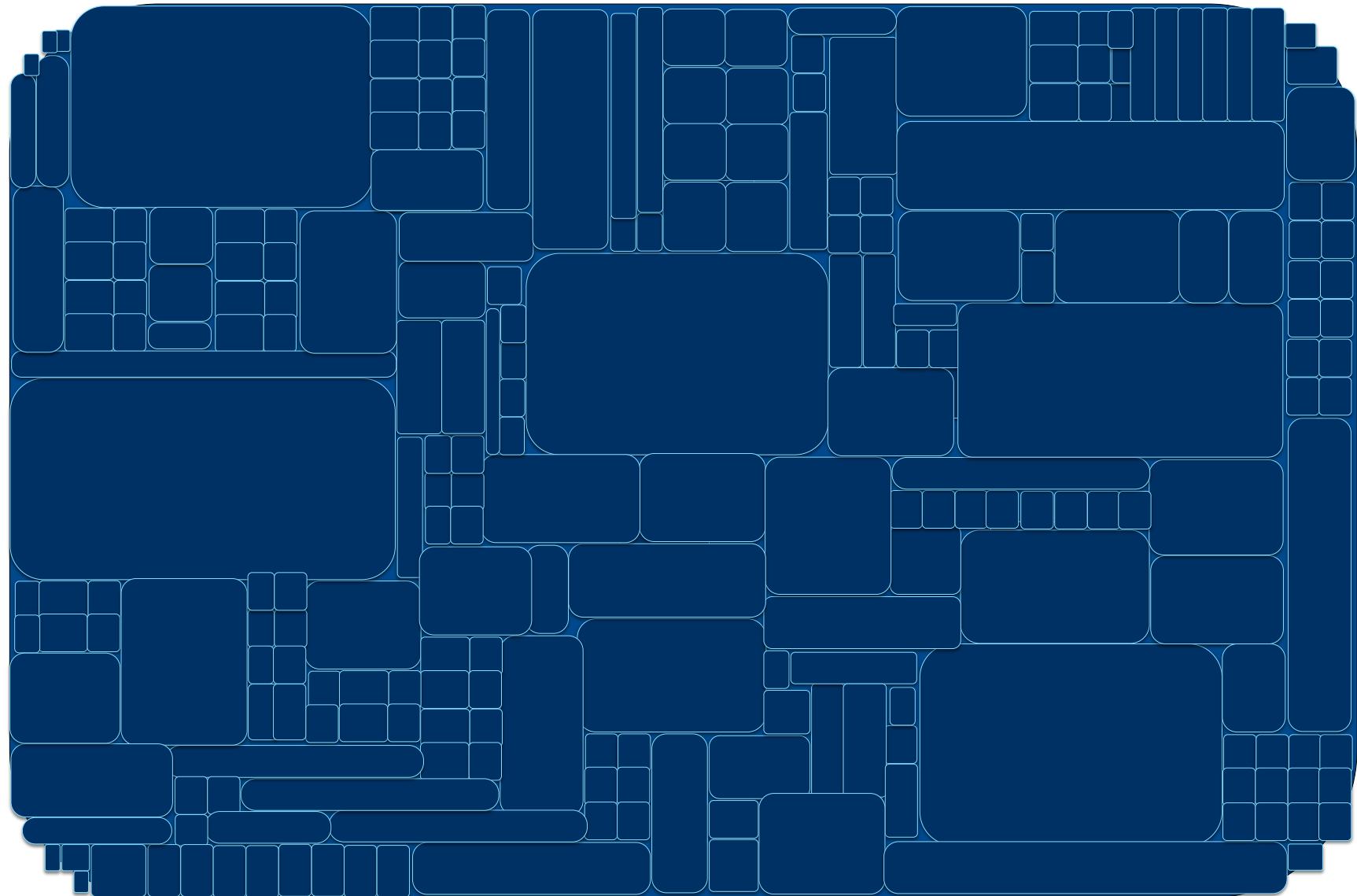
Marco Punta
■ ■ ■

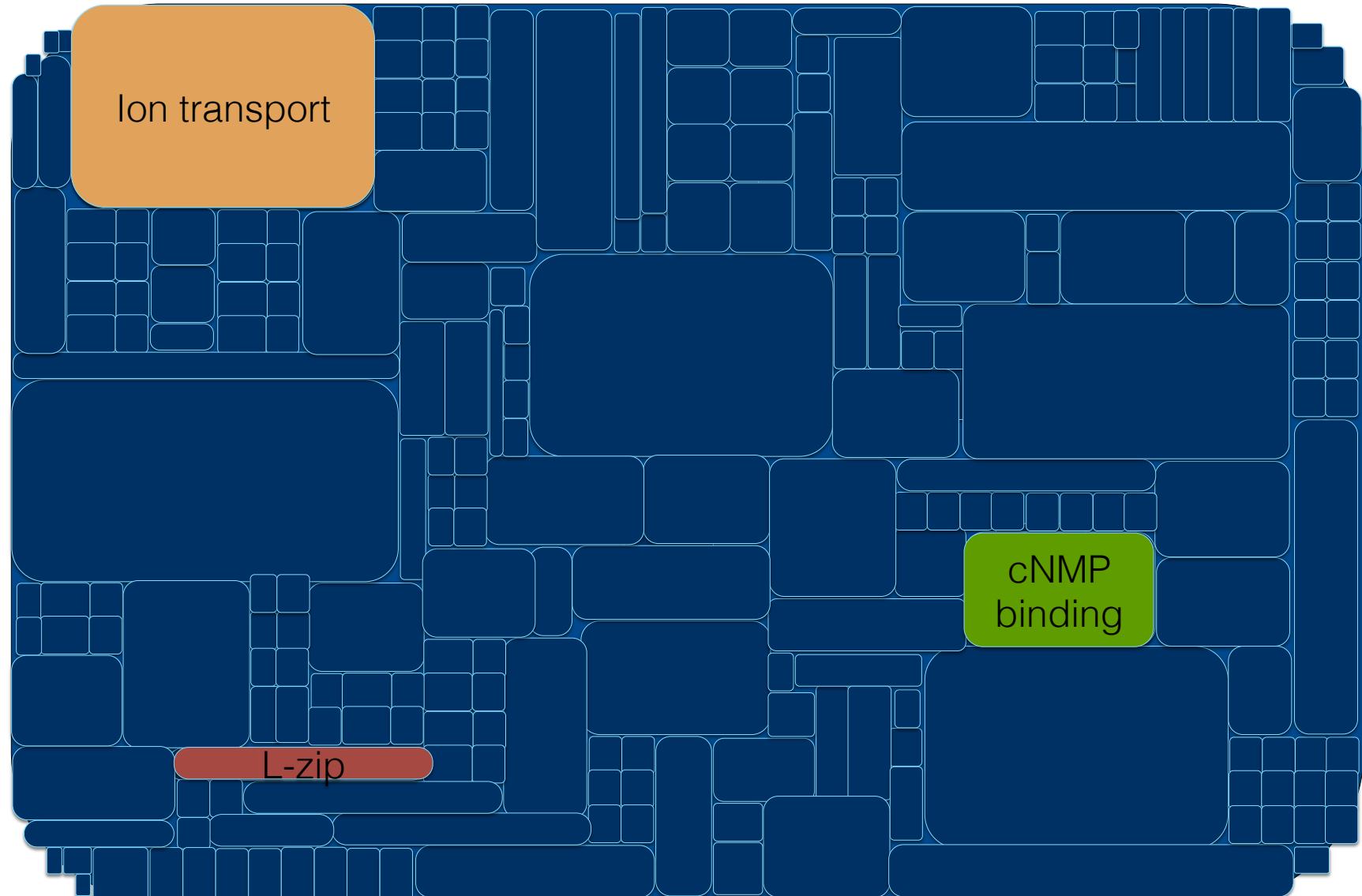
CDD



Marco Punta







How much of the sequence space is currently covered?

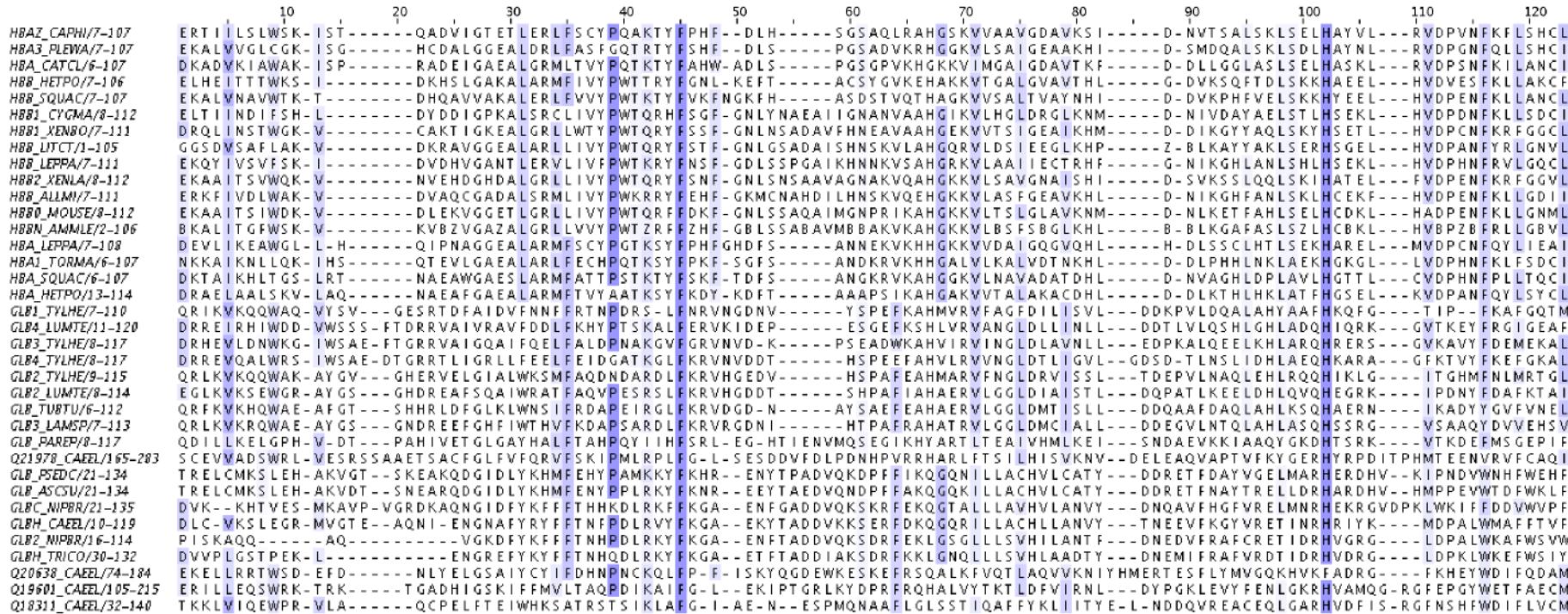
“The Uniprot reference proteomes set that we based Pfam 31.0 on contains 26.7 million sequences, which is an increase in size of 51% compared to when we made Pfam 30.0. Of the proteins in the Uniprot reference proteomes, 73% have a match to at least one Pfam entry, and 48% of all residues fall within a Pfam family.”

<https://xfam.wordpress.com/2017/03/08/pfam-31-0-is-released/>

Sensitivity

Family databases build multiple sequence alignments
between member sequences

MSAs



Functionally and/or structurally relevant?

Family power

Human: 1 MGLSDGEWQLVLNWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFHKLKSEDEMKA 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

Sequence-sequence alignments

Human: 1 MGLSDGEWQLVNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFHKLKSEDEMKA 60
 MGLSDGEWQLVNVWGKVEAD GHGQEVL LFK HPETL KFDKFH LKSE MK SE

Mouse: 1 MGLSDGEWQLVNVWGKVEADLAGHGQEVLIGLFKTHPETLDKFHNLKSEEDMKG 60



Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLSKH 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

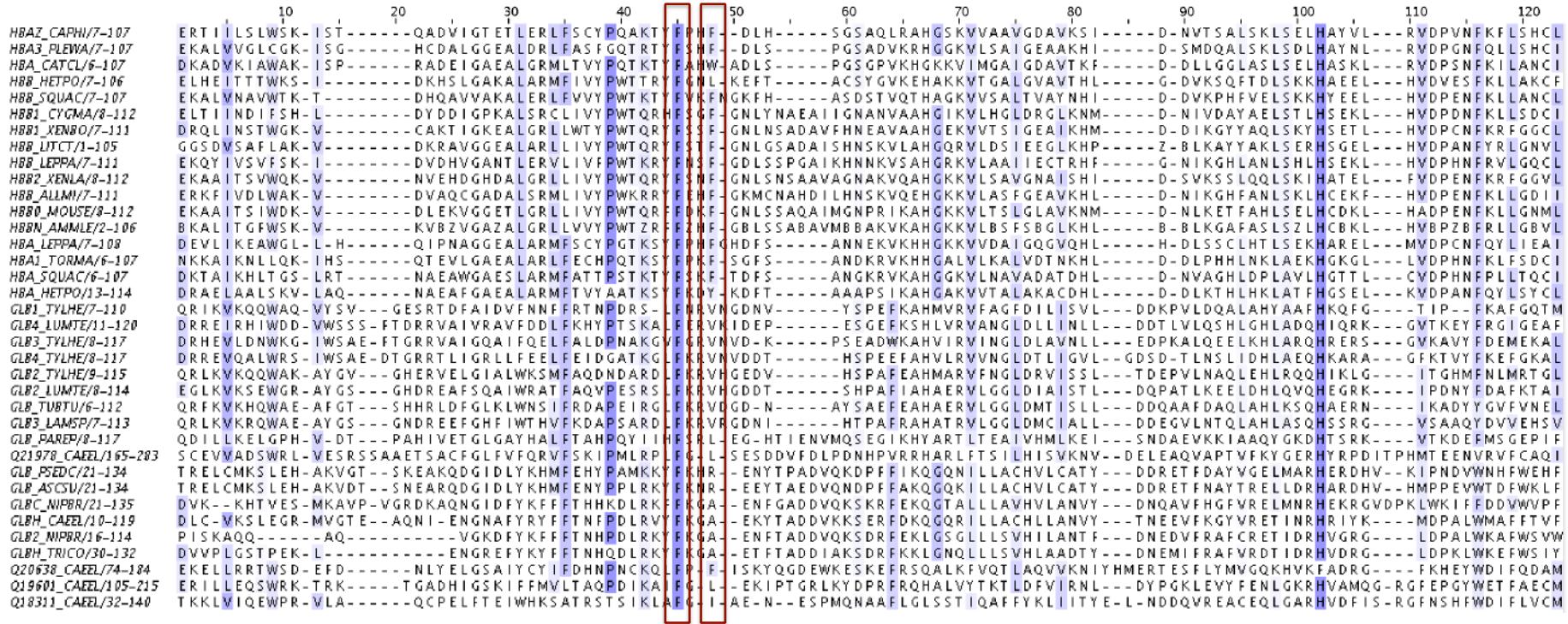
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4	-1	5	-2	0	6	-2	-3	-1	-3	-2	-1	-1	-2	-2	-1	-1	-3	-2	-1
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2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

e.g. F->F Same score irrespective of position along protein sequence

Profile-sequence alignments

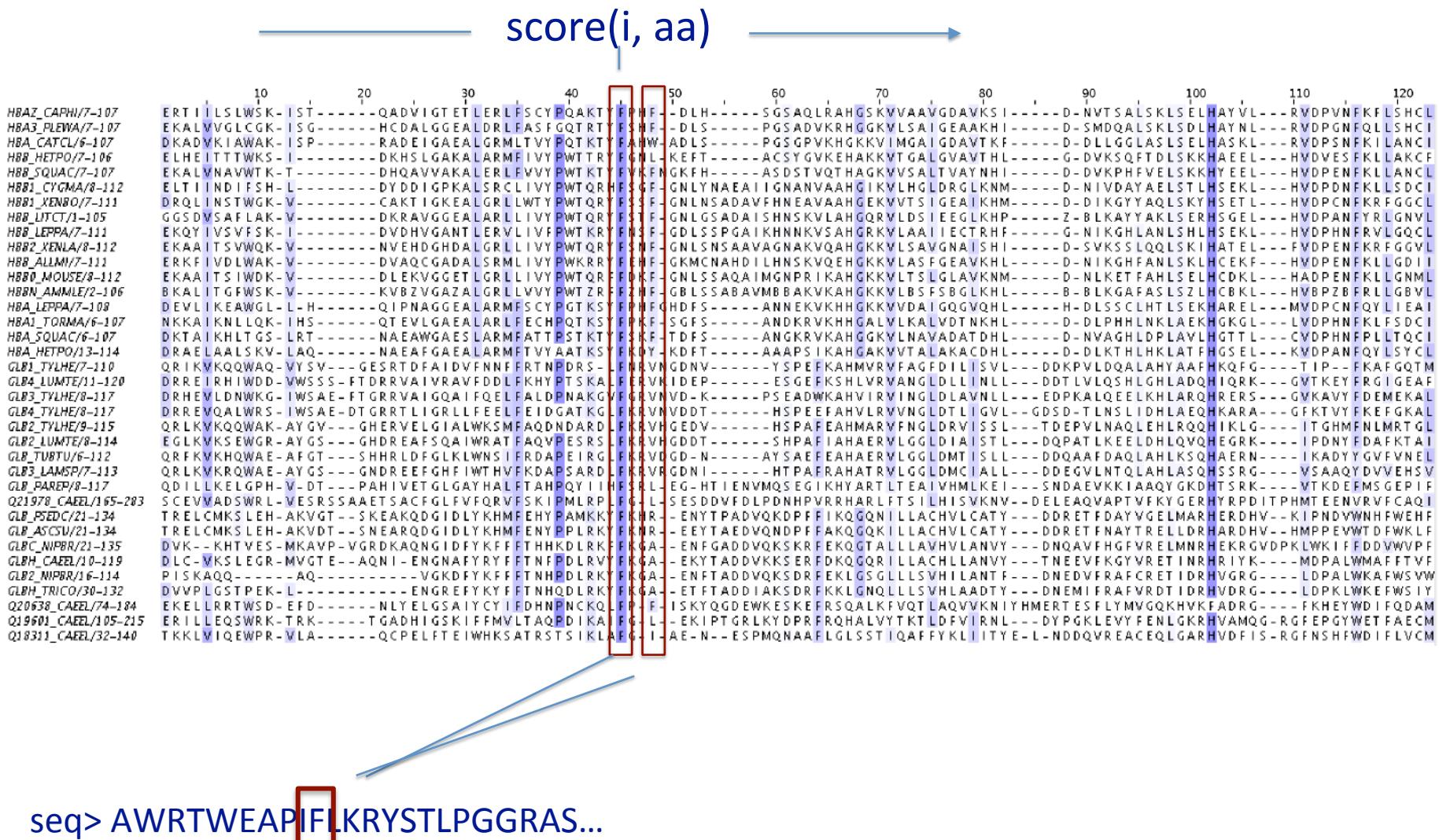
	10	20	30	40	50	60	70	80	90	100	110	120
HBAZ_CAPH/7-107	ERT LS LWSK-IST-----	QADVIGT ET LERL FSCYPQAKTY FPHF--DLH-----	S GSAQLRAHGSKVVAAGDAVKS I-----	D-NVT SALS KLS LSELHAY V-----	RVD PWNFKFLSHCI							
HBA3_PLEWA/7-107	EKA L VVG LCGK- ISG-----	HCD ALG GEAL DRL FAS FGQT RTY FSHF--DLS-----	PGS ADVK R HGG KVLS AIG EA AKH I-----	D-SMD QALS KLS LSELHAY NL-----	RVD PGNFQ LLSHC I							
HBA_CATCL/6-107	DKAD VKI AWA K- ISP-----	RADE I GAE AL GRML T Y PQT KTY FAHW-ADLS-----	PGS GPV K HGKKV I MGA IG DAVT K F-----	D-DLL GG L AS L SELHASK L-----	RVD PSN FK L LANCI							
HBB_HETPO/7-106	ELHE I VTTW KS- I-----	DKH S LGAK AL ARMF I VY PW TTRY FGNL-KEFT-----	ACSY G VKE HAKKV T GALGV AVT HL-----	G- DVQ S QFT DLS KKHA EEL-----	HVD VES FKL LAKC F							
HBB_SQUAC/7-107	EKA LVNA VWT K-T-----	DHQ A VVAKA LER L FV VY PW TTRY FV K FNG K F-----	ASD STV QTH AGK VV SALT VAY NH I-----	D-DVK PHF VELS KKHY EEL-----	HVD P ENF K L LANCI							
HBB1_CYGM A/8-112	ELT I INDIF SH-L-----	DY DDIG P KAL SR CL I VY PW TQR HF S GF-----	GNLY NAE AII GNAN VAAH GIK VLH GLR GLKNM-----	D-NIV DAY A ELS TLH SE K L-----	HVD P DNFK L L SDC I							
HBB1_XENBQ/7-111	DRC L IN STW GK-V-----	CAKT I GKE AL GR LL WTY PW TQRY FSS F-----	GNL NS A DADV FHNE VAA AHGE KV VT S I GEA IKHM-----	D-DIK GY TAOL SKY HS ET L-----	HVD PCN FK R FGG C L							
HBB_LTC T/1-105	GG SD V SFLAK L-V-----	DK RAV GGE ALAR LL I VY PW TQRY FST F-----	GNL GS A DAIS HNS KV LAH G R V L D S I EEG L KHP-----	Z-BLKAY YAK L S ERH SG E L-----	HVD PAN FY RL GNV L							
HBB_ LEPPA/7-111	EK QY IV S VFS K-I-----	DVD HVG A NT LER V L I V F PWT KRY FNS F-----	GDL SS PG A I KHN NK V SA HGR KV LA I I ECT RH F-----	G-NIK GH LAN L SHL HSE K L-----	HVD PHN FR V L GQ C L							
HBB2_XENLA/8-112	EKA AI TS V WQ K-V-----	NVE HDG H D AL GR LL I VY PW TQRY FS NF-----	GNL NS N SAA VAGNA KV QA H GKK V L S VAGN A SH I-----	D-SVK S S L QOL SKI HATE L-----	FVD P ENF K R FGG V L							
HBB_ALLMI/7-111	ER K F IVD L WAK-V-----	DVA QC GAD AL SR ML V-----	GNL NS N SAA VAGNA KV QA H GKK V L S VAGN A SH I-----	D-NIK GH FAN L SKL HCE K F-----	HVD P ENF K L L GDI I							
HBB0_MOUSE/8-112	EKA AI TS I WDK V-----	DLE KV GGET L GR LL I VY PW TQRF FD K F-----	GNL SS AQA I MG N P R I KA HG K KV LT S L G L A V K NM-----	D-NL KET FAH L S EL HCD K L-----	HAD P ENF K L L GNM L							
HBBN_AMMLE/2-10F	BKA L IT GFW SK-V-----	KVB ZV GAZ AL GR LL V VY PW T Z R F Z H F-----	GBL S SAB AV MB B A V K V KAH GKK V LB S FSB GL K H-----	B-BLK GAF A S L S Z L HCB K L-----	HVB PZ B F R L L G B V L							
HBA_ LEPPA/7-108	DEV L I CAE WGL-L-H-----	QIP NAG GE ALARM F S C PGT K SY F P F GHD F S-----	ANNE K V KHH GKK V V D A I GQ G V Q H L-----	H-DL SS CL HT L S EKHARE L-----	MVDP CNF QY L I EAI							
HBA1_TORMA/6-107	NKK A I KNL L QK- I HS-----	QTE VLG AE ALAR L F ECH P QTK TS F P K F-S G F S-----	AN DK R V K HH G GAL V L K AL D TN K H L-----	D-DLP H HLN K L A E K H G K G L-----	LVD PHN F K L F SDC I							
HBA_SQUAC/6-107	DKT A I KHL TGS -LRT-----	-N A E A W G A E S L A R M F AT T P S T K TY F S K F-T D F S-----	-ANG R V K V A H G G K V L N A V A D A T D H L-----	-D-NVAG HLD P L A V L H G T T L-----	CVD PHN F P L L T Q C I							
HBA_HETPO/13-114	DRA E A L A L S K V-LA Q-----	-N A E A F G A E AL A R M F T V Y A A T K S Y F K D Y -K D F T-----	-AA A P S I K A H G A K V V T A L A K C D H L-----	-D-DLK TH L H K L A T F H G S E L-----	KVDP AN F Q Y L S Y C L							
GLB1_TYLHE/7-110	QR L KV KQ QWA Q-VY SV-----	GES RT D FAID Y F 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 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 G Q T M							
GLB4_LUMTE/11-120	DR RE IR H I W D D-V W S S-----	FT DRR R V A I V R A F F D D L F K H Y P T S K A L F E R V K 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-----	GVT KEY F RG I G E A F							
GLB5_TYLHE/8-117	DR HE V LD N W K G -I W S A E-----	FT G R R V A I Q G A I F Q E L F A D L P N K A G V F G R V N V D-K-----	-P S E A D W K A H V I R V I N G L D L A V N L L-----	-EDPK A L Q E L K L H 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 A L W R S-----	DT 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 V N V D D T-----	-H S P E F A H V L R V N G L D T L I G V L-----	-GDS D-T L N S L I D H L A E Q H K A R A-----	GFK T V Y F K E F G K A L							
GLB2_TYLHE/9-115	Q R L KV KQ QWA K-A Y G V-----	G H E R V E L G I A L W K S M F A Q D N D A R D L F K 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 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 KV S E W R G-A Y G S-----	GHD R E A F S Q A I W R A T F A Q V P E S R L S F K 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 L K E E L D H L Q V Q H E G R K-----	I PD NY F D A F K T A I							
GLB_TUBTU/6-112	Q R F KV 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 K R V D G D-N-----	-A Y S A F F E 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 KV K R Q W A E-A Y G S-----	Q 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 R V R 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 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							
GLB_PAREP/8-117	Q D I L L K E L G P H-V D T-----	-PA 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 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								
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 G-L-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-----	-D E L E A Q V A P T V F K Y G E R H Y R P D I T P H M T E E N V R V F C A Q 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 I D 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 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_PSEDC/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 L Y K H M F E H Y P P L R K Y 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 N A Y T R E L L D R H A R D H V-----	-H M P P E V W T D F W K L F							
GLB_ACSVU/21-134	D V K -K H T V E S-M K A V P-----	-V O R D K A Q N G I D F Y K F 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 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 R H E K R G V D P K L W K I F F D D V W V P F								
GLB_NIPBR/21-135	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 K G A-----	-E K Y T A D D V K K S E R F D K Q Q G 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							
GLBH_CAEEL/10-119	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 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 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						
GLB2_NIPBR/16-114	D 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 K G A-----	-E T F T A D D I L A K S D R F K K L G N Q L 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 K L W K E F W S I Y							
GLB_TRICO/30-132	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-----	-I S K Y Q G D E W K E S K 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 K H V K F A D R G-----	-F K H E Y W D I F Q D A M								
Q20638_CAEEL/74-184	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 R 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								
Q39601_CAEEL/105-215	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 G -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								
Q18311_CAEEL/32-140												

AA probabilities within the family are position specific



$$F_i \rightarrow F_i \neq F_j \rightarrow F_j$$

Position-specific scores



Sequence-profile alignments

- Position specific substitution matrices
- profile-hidden Markov models

EDITORIAL

New computational approaches to understanding molecular protein function

Jacquelyn S. Fetrow^{1*}, Patricia C. Babbitt²

1 Office of the President, Albright College, Reading, Pennsylvania, United States of America, **2** Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, United States of America

* jfetrow@albright.edu

Defining function

Function is like beauty—its definition lies in the eye of the beholder or, in this case, the researcher. At the broadest level, we define organismal function—the function that the protein plays in the overall organism. This function can be observed by understanding the impact on the organism of deletion or mutation of the protein. Physiological function is the function the protein plays in pathways, such as metabolic or signaling pathways. Another level of function

Scenario

- We have identified (experimentally or/and computationally) some proteins that interact with our protein of interest
- We want to gather functional information about these proteins: are all their regions/domains annotated?
- What can we say about unannotated regions?
We are going to use sequence searches to have investigate them

Goals

- We will learn more on the advantages of profile-based searches vs single sequence searches
- We will learn how to run an iterative profile search (and associated caveats)
- We will learn more on how to edit a multiple sequence alignment
- We will learn how to build and annotate a new protein family

Ambitious!

Our protein

C-reactive protein (P02741):

- Displays several functions associated with host defense: it promotes agglutination, bacterial capsular swelling, phagocytosis and complement fixation through its calcium-dependent binding to phosphorylcholine. Can interact with DNA and histones and may scavenge nuclear material released from damaged circulating cells (UniProtKB annotation).
- Over-expressed in many diseases, including malignant tumors and cardiovascular disorders.
- Prognostic Marker in Patients with Hepatocellular Carcinoma
- ...

Identification of the C-Reactive Protein Interaction Network Using a Bioinformatics Approach Provides Insights into the Molecular Pathogenesis of Hepatocellular Carcinoma

Sha She Lingyu Jiang Zhenfang Zhang Min Yang Huaidong Hu
Peng Hu Yong Liao Yixuan Yang Hong Ren

Institute for Viral Hepatitis, Key Laboratory of Molecular Biology for Infectious Diseases (Ministry of Education), Department of Infectious Diseases, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, China

Key Words

C-reactive protein (CRP) • Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) • Gene Ontology (GO) • Kyoto Encyclopedia of Genes and Genomes (KEGG)

Abstract

Background/Aims: C reactive protein (CRP) levels are elevated in many diseases, including malignant tumors and cardiovascular disorders. In this study, the protein interaction network for CRP was evaluated to determine the importance of CRP and its interacting proteins in the molecular pathogenesis of hepatocellular carcinoma (HCC). **Methods:** Isobaric tags for relative and absolute quantitation (iTRAQ) and mass spectrometry were used to identify CRP interacting proteins in SMMC7721 cells. Moreover, Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to evaluate enriched genes and pathways for differentially expressed genes using DAVID and WebGestalt. Co-immunoprecipitation and western blot analyses were employed to assess interactions between CRP and KRT8, ANXA2, ENO2, and HSP90B1. **Results:** In total, 52 proteins that interact with CRP were identified. A GO analysis suggested that most of the interacting proteins were involved in CRP complexes and regulated metabolic processes. A KEGG pathway analysis suggested that most CRP-interacting proteins contribute to the TRAIL signaling pathway, Class I PI3K/Akt signaling pathway, plasma membrane estrogen receptor signaling, Nectin adhesion pathway, and S1P1 pathway. Immunoprecipitation and western blot analyses revealed interactions between CRP and KRT8, ANXA2, ENO2, and HSP90B1. **Conclusions:** iTRAQ based proteomic profiling revealed the network of CRP interacting proteins. This network may activate the PI3K/Akt signaling pathway, thereby contributing to the pathogenesis of HCC.

© 2018 The Author(s)
Published by S. Karger AG, Basel

S. She, L. Jiang and Z. Zhang contributed equally to this work.

Prof. Yixuan Yang
and Hong Ren

Department of Infectious Diseases, The Second Affiliated Hospital of Chongqing
Medical University, 76# Linjiang Road, Chongqing 400010 (China)
Tel. 86 23 63693215, Fax 86 23 63693215, E-Mail: yummy_8813@163.com

Sha She Lingyu Jiang Zhenfang Zhang Min Yang Huaidong Hu
Peng Hu Yong Liao Yixuan Yang Hong Ren

Institute for Viral Hepatitis, Key Laboratory of Molecular Biology for Infectious Diseases (Ministry of Education), Department of Infectious Diseases, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, China

Key Words

C-reactive protein (CRP) • Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) • Gene Ontology (GO) • Kyoto Encyclopedia of Genes and Genomes (KEGG)

Abstract

Background/Aims: C reactive protein (CRP) levels are elevated in many diseases, including malignant tumors and cardiovascular disorders. In this study, the protein interaction network for CRP was evaluated to determine the importance of CRP and its interacting proteins in the molecular pathogenesis of hepatocellular carcinoma (HCC). **Methods:** Isobaric tags for relative and absolute quantitation (iTRAQ) and mass spectrometry were used to identify CRP interacting proteins in SMMC7721 cells. Moreover, Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to evaluate enriched genes and pathways for differentially expressed genes using DAVID and WebGestalt. Co-immunoprecipitation and western blot analyses were employed to assess interactions between CRP and KRT8, ANXA2, ENO2, and HSP90B1. **Results:** In total, 52 proteins that interact with CRP were identified. A GO analysis suggested that most of the interacting proteins were involved in CRP complexes and regulated metabolic processes. A KEGG pathway analysis suggested that most CRP-interacting proteins contribute to the TRAIL signaling pathway, Class I PI3K/Akt signaling pathway, plasma membrane estrogen receptor signaling, Nectin adhesion pathway, and S1P1 pathway. Immunoprecipitation and western blot analyses revealed interactions between CRP and KRT8, ANXA2, ENO2, and HSP90B1. **Conclusions:** iTRAQ based proteomic profiling revealed the network of CRP interacting proteins. This network may activate the PI3K/Akt signaling pathway, thereby contributing to the pathogenesis of HCC.

Sha She Lingyu Jiang Zhenfang Zhang Min Yang Huaidong Hu
Peng Hu Yong Liao Yixuan Yang Hong Ren

Institute for Viral Hepatitis, Key Laboratory of Molecular Biology for Infectious Diseases (Ministry of Education), Department of Infectious Diseases, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, China

Key Words

C-reactive protein (CRP) • Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) • Gene Ontology (GO) • Kyoto Encyclopedia of Genes and Genomes (KEGG)

Abstract

Background/Aims: C reactive protein (CRP) levels are elevated in many diseases, including malignant tumors and cardiovascular disorders. In this study, the protein interaction network for CRP was evaluated to determine the importance of CRP and its interacting proteins in the molecular pathogenesis of hepatocellular carcinoma (HCC). **Methods:** Isobaric tags for relative and absolute quantitation (iTRAQ) and mass spectrometry were used to identify CRP interacting proteins in SMMC7721 cells. Moreover, Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to evaluate enriched genes and pathways for differentially expressed genes using DAVID and WebGestalt. Co-immunoprecipitation and western blot analyses were employed to assess interactions between CRP and KRT8, ANXA2, ENO2, and HSP90B1. **Results:** In total, 52 proteins that interact with CRP were identified. A GO analysis suggested that most of the interacting proteins were involved in CRP complexes and regulated metabolic processes. A KEGG pathway analysis suggested that most CRP-interacting proteins contribute to the TRAIL signaling pathway, Class I PI3K/Akt signaling pathway, plasma membrane estrogen receptor signaling, Nectin adhesion pathway, and S1P1 pathway. Immunoprecipitation and western blot analyses revealed interactions between CRP and KRT8, ANXA2, ENO2, and HSP90B1. **Conclusions:** iTRAQ based proteomic profiling revealed the network of CRP interacting proteins. This network may activate the PI3K/Akt signaling pathway, thereby contributing to the pathogenesis of HCC.

UniProtKB - P05787 (K2C8_HUMAN)

Display

Entry

Publications

Feature viewer

Feature table

None

 Function Names & Taxonomy Subcellular location Pathology & Biotech PTM / Processing Expression Interaction Structure Family & Domains Sequences (2+) Similar proteins Cross-references Entry information Miscellaneous

▲ Top

Protein | Keratin, type II cytoskeletal 8

Gene | KRT8

Organism | Homo sapiens (Human)

Status

 Reviewed - Annotation score:  - Experimental evidence at protein levelⁱ 1 Publication ▾Functionⁱ

Together with KRT19, helps to link the contractile apparatus to dystrophin at the costameres of striated muscle.

Miscellaneous

There are two types of cytoskeletal and microfibrillar keratin: I (acidic; 40-55 kDa) and II (neutral to basic; 56-70 kDa).

Sites

Feature key	Position(s)	Description
Stutter ⁱ	342	Stutter

GO - Molecular functionⁱ

- protein-containing complex binding 
- scaffold protein binding 
- structural molecule activity 

[View the complete GO annotation on QuickGO ...](#)

GO - Biological processⁱ

- cell differentiation involved in embryonic placenta development 
- cornification 
- extrinsic apoptotic signaling pathway 
- hepatocyte apoptotic process 
- keratinization 
- response to hydrostatic pressure 

We want to learn what is known
about the domain composition
of KRT8

Let's visit the Pfam website:
<https://pfam.xfam.org>

P05787



HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam
keyword search **Go**

Pfam 32.0 (September 2018, 17929 entries)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. [More...](#)

QUICK LINKS

[SEQUENCE SEARCH](#)

[VIEW A PFAM ENTRY](#)

[VIEW A CLAN](#)

[VIEW A SEQUENCE](#)

[VIEW A STRUCTURE](#)

[KEYWORD SEARCH](#)

JUMP TO

YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...

Analyze your protein sequence for Pfam matches

View Pfam annotation and alignments

See groups of related entries

Look at the domain organisation of a protein sequence

Find the domains on a PDB structure

Query Pfam by keywords

P05787

Go

[Example](#)

Enter any type of accession or ID to jump to the page for a Pfam entry or clan, UniProt sequence, PDB structure, etc.

Or view the [help](#) pages for more information

Recent Pfam [blog](#) posts

Hide this

[Pfam 31.0 is released](#) (posted 8 March 2017)

Pfam 31.0 contains a total of 16712 families and 604 clans. Since the last release, we have built 415 new families, killed 9 families and created 11 new clans. We have also been working on expanding our clan

<https://pfam.xfam.org>

Summary**Sequence****Structures****TreeFam****Jump to...** ⓘ

enter ID/acc

Go

SummaryThis is the summary of UniProt entry [K2C8_HUMAN](#) (P05787).**Description:** Keratin, type II cytoskeletal 8**Source organism:** [Homo sapiens \(Human\)](#) (NCBI taxonomy ID 9606)[View Pfam proteome data.](#)**Length:** 483 amino acids**Reference Proteome:** ✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)



[Download the data used to generate the domain graphic in JSON format.](#)

Source	Domain	Start	End
Pfam	Keratin_2_head	6	87
disorder	n/a	18	26
low_complexity	n/a	23	48
disorder	n/a	31	34
disorder	n/a	36	38
low_complexity	n/a	41	61
disorder	n/a	89	91
Pfam	Filament	90	401
coiled_coil	n/a	95	115
coiled_coil	n/a	144	192
coiled_coil	n/a	200	234
disorder	n/a	297	298
disorder	n/a	300	309
coiled_coil	n/a	300	380
disorder	n/a	311	312
low_complexity	n/a	393	404
disorder	n/a	402	411
disorder	n/a	413	415
low_complexity	n/a	415	431
low_complexity	n/a	426	459
disorder	n/a	477	479

Summary**Sequence****Structures****TreeFam****Jump to...**

enter ID/acc

Go

SummaryThis is the summary of UniProt entry [K2C8_HUMAN](#) (P05787).**Click**

Description:	Keratin, type II cytoskeletal 8
Source:	Homo sapiens (Human) (NCBI taxonomy ID 9606) with Pfam proteome data.
Reference:	3 amino acids

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)



[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
Pfam	Keratin_2_head	6	87
disorder	n/a	18	26
low_complexity	n/a	23	48
disorder	n/a	31	34
disorder	n/a	36	38
low_complexity	n/a	41	61
disorder	n/a	89	91
Pfam	Filament	90	401
coiled_coil	n/a	95	115
coiled_coil	n/a	144	192
coiled_coil	n/a	200	234
disorder	n/a	297	298
disorder	n/a	300	309
coiled_coil	n/a	300	380
disorder	n/a	311	312
low_complexity	n/a	393	404
disorder	n/a	402	411
disorder	n/a	413	415
low_complexity	n/a	415	431
low_complexity	n/a	426	459
disorder	n/a	477	479

Family: Filament (PF00038)

119 architectures

9043 sequences

3 interactions

327 species

43 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to... ⓘ

enter ID/acc

Go

Summary: Intermediate filament protein

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

[Wikipedia: Intermediate filament](#)

Pfam

InterPro

This is the Wikipedia entry entitled "[Intermediate filament](#)". [More...](#)

Intermediate filament Edit Wikipedia article

Intermediate filaments (IFs) are cytoskeletal components found in the cells of vertebrate animal species,^{[1][2]} and perhaps also in other animals, fungi, plants, and unicellular organisms.^[3] They are composed of a family of related proteins sharing common structural and sequence features. Initially designated 'intermediate' because their average diameter (10 nm) is between those of narrower microfilaments (actin) and wider myosin filaments found in muscle cells, the diameter of intermediate filaments is now commonly compared to actin microfilaments (7 nm) and microtubules (25 nm).^{[1][4]} Most types of intermediate filaments are cytoplasmic, but one type, the lamins, are nuclear. Unlike microtubules, IFs distribution in cells show no good correlation with the distribution of either mitochondria or endoplasmic reticulum.^[5]

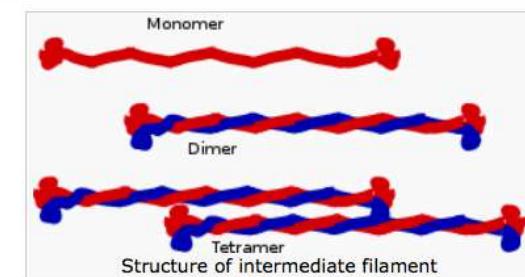
[Contents](#) [hide]

1 Structure

2 Biomechanical properties

3 Types

- 3.1 Types I and II – acidic and basic keratins
- 3.2 Type III
- 3.3 Type IV
- 3.4 Type V - nuclear lamins
- 3.5 Type VI
- 3.6 Unclassified



Intermediate filament tail domain



Click

Family: Filament (PF00038)

119 architectures 9043 sequences 3 interactions 327 species 43 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to... ⓘ

enter ID/acc

Go

Summary: Intermediate filament protein

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

[Wikipedia: Intermediate filament](#)

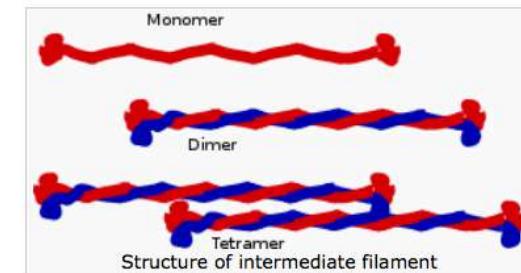
[Pfam](#)

[InterPro](#)

This is the Wikipedia entry entitled "[Intermediate filament](#)". [More...](#)

Intermediate filament [Edit Wikipedia article](#)

Intermediate filaments (IFs) are cytoskeletal components found in the cells of vertebrate animal species,^{[1][2]} and perhaps also in other animals, fungi, plants, and unicellular organisms.^[3] They are composed of a family of related proteins sharing common structural and sequence features. Initially designated 'intermediate' because their average diameter (10 nm) is between those of narrower microfilaments (actin) and wider myosin filaments found in muscle cells, the diameter of intermediate filaments is now commonly compared to actin microfilaments (7 nm) and microtubules (25 nm).^{[1][4]} Most types of intermediate filaments are cytoplasmic, but one type, the lamins, are nuclear. Unlike microtubules, IFs distribution in cells show no good correlation with the distribution of either mitochondria or endoplasmic reticulum.^[5]



[Contents](#) [hide]

1 Structure

2 Biomechanical properties

3 Types

- 3.1 Types I and II – acidic and basic keratins
- 3.2 Type III
- 3.3 Type IV
- 3.4 Type V - nuclear lamins
- 3.5 Type VI
- 3.6 Unclassified

Intermediate filament tail domain



<https://pfam.xfam.org>

Family: Filament (PF00038)

119 architectures

9043 sequences

3 interactions

327 species

43 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to... ⓘ

enter ID/acc

Go

Summary: Intermediate filament protein

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

[Wikipedia: Intermediate filament](#) [Pfam](#) [InterPro](#)

This tab holds annotation information from the [InterPro](#) database.

InterPro entry IPR039008 ⓘ

Intermediate filaments (IF) [PUBMED:2183847, PUBMED:28101862] are proteins which are primordial components of the cytoskeleton and the nuclear envelope. They generally form filamentous structures 8 to 14 nm wide. IF proteins are members of a very large multigene family of proteins which has been subdivided in six types:

- Type I: Acidic cytokeratins.
- Type II: Basic cytokeratins.
- Type III: Vimentin, desmin, glial fibrillary acidic protein (GFAP), peripherin, and plastin.
- Type IV: Neurofilaments L, H and M, alpha-internexin and nestin.
- Type V: Nuclear lamins A, B1, B2 and C.
- Type VI: 'Orphan' IF proteins, which are more distant in terms of their amino acid sequences.

All IF proteins are structurally similar in that they consist of: a central rod domain comprising some 300 to 350 residues which is arranged in coiled-coiled alpha-helices, with at least two short characteristic interruptions; a N-terminal non-helical domain (head) of variable length; and a C-terminal domain (tail) which is also non-helical, and which shows extreme length variation between different IF proteins.

While IF proteins are evolutionary and structurally related, they have limited sequence homologies except in several regions of the rod domain. The IF rod domain is approximately 310 residues long in all cytoplasmic IF proteins and close to 350 residues in the nuclear ones. The IF rod domain exhibits an interrupted alpha-helical conformation and reveals a pronounced seven-residue periodicity in the distribution of apolar residues. The heptad periodicity within the rod domain is interrupted in several places, which generates four consecutive alpha-helical segments: 1A and 1B, which together form the so-called coil 1, and 2A and 2B, which form coil 2. The four alpha-helical segments are interconnected by relatively short, variable linkers L1, L12 and L2 [PUBMED:12596228, PUBMED:22869704].

IF proteins have a very strong tendency to dimerize via the formation of an alpha-helical coiled coil (CC) by their rod domains [PUBMED:22869704].

Summary**Sequence****Structures****TreeFam****Jump to...** ⓘ

enter ID/acc

Go

SummaryThis is the summary of UniProt entry [K2C8_HUMAN](#) (P05787).**Description:** Keratin, type II cytoskeletal 8**Source organism:** [Homo sapiens \(Human\)](#) (NCBI taxonomy ID [9606](#))[View Pfam proteome data.](#)**Length:** 483 amino acids**Reference Proteome:** ✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)

[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
Pfam	Keratin_2_head	6	87
disorder	n/a	18	26
low_complexity	n/a	23	48
disorder	n/a	31	34
disorder	n/a	36	38
low_complexity	n/a	41	61
disorder	n/a	62	71
Pfam	Filament	90	401
coiled_coil	n/a	95	115
coiled_coil	n/a	144	192
coiled_coil	n/a	200	234
disorder	n/a	297	298
disorder	n/a	300	309
coiled_coil	n/a	300	380
disorder	n/a	311	312
low_complexity	n/a	393	404
disorder	n/a	402	411
disorder	n/a	413	415
low_complexity	n/a	415	431
low_complexity	n/a	426	459
disorder	n/a	477	479

Central Rod domain

<https://pfam.xfam.org>

Summary**Sequence****Structures****TreeFam****Jump to...**

enter ID/acc

Go

SummaryThis is the summary of UniProt entry [K2C8_HUMAN](#) (P05787).**Description:** Keratin, type II cytoskeletal 8**Source organism:** [Homo sapiens \(Human\)](#) (NCBI taxonomy ID 9606)[View Pfam proteome data.](#)**Length:** 483 amino acids**Reference Proteome:** ✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)

[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
Pfam	Keratin_2_head	6	87
disorder	n/a	18	26
low_complexity	n/a	23	48
disorder	n/a	31	34
disorder	n/a	36	38
low_complexity	n/a	41	61
disorder	n/a	89	91
Pfam	Filament	90	401
coiled_coil	n/a	95	115
coiled_coil	n/a	144	192
coiled_coil	n/a	200	234
disorder	n/a	297	298
disorder	n/a	300	309
coiled_coil	n/a	300	380
disorder	n/a	311	312
low_complexity	n/a	393	404
disorder	n/a	402	411
disorder	n/a	413	415
low_complexity	n/a	415	431
low_complexity	n/a	426	459
disorder	n/a	477	479

Head domain**Tail domain region**

Family: Keratin_2_head (PF16208)

 32 architectures 2224 sequences 0 Interactions 119 species 0 structures**Summary****Domain organisation**

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to... [enter ID/acc](#)[Go](#)

Summary: Keratin type II head

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

[No Wikipedia article](#) [Pfam](#) [InterPro](#)

This tab holds annotation information from the [InterPro](#) database.

InterPro entry [IPR032444](#)

All intermediate filament proteins feature a central alpha-helical rod domain and variable nonhelical domains located at the N-terminal (head) and C-terminal (tail). The central rod domain is the main driver of self-assembly into filaments, whereas the N- and C-terminal domains are involved in post-translational modifications and interactions with other proteins [[PUBMED:16710422](#)]. Type I and type II keratin form heteropolymeric intermediate filaments providing vital mechanical support in epithelia [[PUBMED:22705788](#)]. This entry represents the N-terminal domain (head) of type II keratins.

Summary**Sequence****Structures****TreeFam****Jump to...** ⓘ

enter ID/acc

Go

SummaryThis is the summary of UniProt entry [K2C8_HUMAN](#) (P05787).**Description:** Keratin, type II cytoskeletal 8**Source organism:** [Homo sapiens \(Human\)](#) (NCBI taxonomy ID 9606) View Pfam proteome data.**Length:** 483 amino acids**Reference Proteome:** ✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)



[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
Pfam	Keratin_2_head	6	87
disorder	n/a	18	26
low_complexity	n/a	23	48
disorder	n/a	31	34
disorder	n/a	36	38
low_complexity	n/a	41	61
disorder	n/a	89	91
Pfam	Filament	90	401
coiled_coil	n/a	95	115
coiled_coil	n/a	144	192
coiled_coil	n/a	200	234
disorder	n/a	297	298
disorder	n/a	300	309
coiled_coil	n/a	300	380
disorder	n/a	311	312
low_complexity	n/a	393	404
disorder	n/a	402	411
disorder	n/a	413	415
low_complexity	n/a	415	431
low_complexity	n/a	426	459
disorder	n/a	477	479

Tail domain region

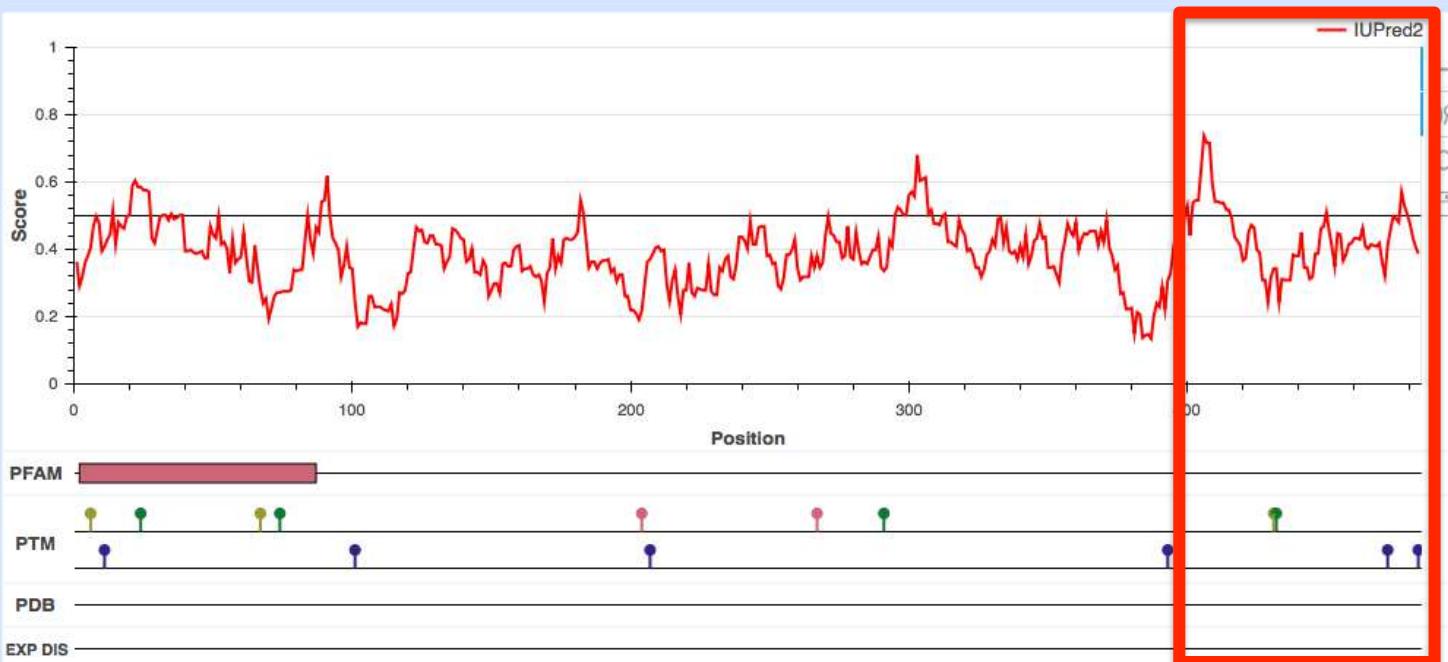
→ Predictions of disorder and low complexity in some parts of the tail region



Prediction of Intrinsically Unstructured Proteins

- IUPred2A
 - How to use
 - Examples
 - New features
 - Statistics
 - Related links
 - Downloads
- NEW**
- Just published in Science
Signaling: A review on
Degrons in Cancer

>sp|P05787|K2C8_HUMAN Keratin, type II cytoskeletal 8 OS=Homo sapiens GN=KRT8 PE=1 SV=7



[Download text](#)

[Download JSON](#)

Overview

Similar proteins (2628)

Structures

Filter view on

Entry type

- H Homologous superfamily
- F Family
- D Domains
- R Repeats
- S Site

Status

- Unintegrated

Per-residue features

- Residue annotation

Colour by

help

- domain relationship
- source database

Keratin, type II cytoskeletal 8 (P05787)

Export TSV

Export FASTA

Accession [P05787](#) (K2C8_HUMAN)

Species Homo sapiens (Human)

Length 483 amino acids (complete)

Source: UniProtKB

Protein family membership

↳ F Intermediate filament protein (IPR001664)

 ↳ F Keratin, type II (IPR003054)

Homologous superfamilies

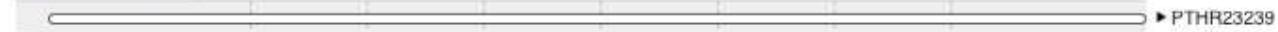
None predicted.

Domains and repeats



Detailed signature matches

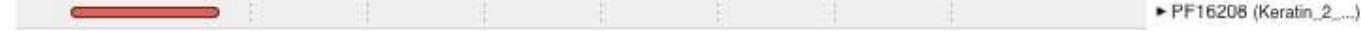
↳ IPR001664 Intermediate filament protein



↳ IPR003054 Keratin, type II



↳ IPR032444 Keratin type II head



↳ IPR039008 Intermediate filament, rod domain



Overview

Similar proteins (2628)

Structures

Filter view on

Entry type

- H Homologous superfamily
- F Family
- D Domains
- R Repeats
- S Site

Status

- Unintegrated

Per-residue features

- Residue annotation

Colour by

help

- domain relationship
- source database

Keratin, type II cytoskeletal 8 (P05787)

Export TSV

Export FASTA

Accession [P05787](#) (K2C8_HUMAN)

Species Homo sapiens (Human)

Length 483 amino acids (complete)

Source: UniProtKB

Protein family membership

↳ F Intermediate filament protein (IPR001664)

 ↳ F Keratin, type II (IPR003054)

Homologous superfamilies

None predicted.

Pfam domains

This image shows the arrangement of the describing that Pfam entry. The table below



Domains and repeats



Detailed signature matches

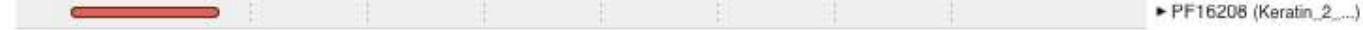
↳ IPR001664 Intermediate filament protein



↳ IPR003054 Keratin, type II



↳ IPR032444 Keratin type II head



↳ IPR039008 Intermediate filament, rod domain



Overview

Similar proteins (2628)

Structures

Filter view on

Entry type

-  Homologous superfamily
-  Family
-  Domains
-  Repeats
-  Site

Status

-  Unintegrated

Per-residue features

- Residue annotation

Colour by

help

- domain relationship
- source database

Keratin, type II cytoskeletal 8 (P05787)

 Export TSV

 Export FASTA

Accession  P05787 (K2C8_HUMAN)

Species Homo sapiens (Human)

Length 483 amino acids (complete)

Source: UniProtKB

Protein family membership

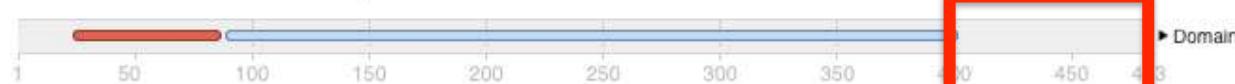
  Intermediate filament protein (IPR001664)

  Keratin, type II (IPR003054)

Homologous superfamilies

None predicted.

Domains and repeats



Detailed signature matches

 IPR001664 Intermediate filament protein

 IPR003054 Keratin, type II

 IPR032444 Keratin type II head

 IPR039008 Intermediate filament, rod domain

► Domain

► PTHR23239

► PR01276 (TYPE2KERATIN)

► PF16208 (Keratin_2,...)

► PS51842 (IF_ROD_2)

► SM01391 (Filament)

► PF00038 (Filament)

- We are going to look for homologs of KRT8's tail region and see if we can learn more about its function

>sp|P05787|K2C8_HUMAN Keratin, type II cytoskeletal 8 OS=Homo sapiens OX=9606 GN=KRT8 PE=1 SV=7
MSIRVTQSYKVSTSGPRAFSSRSYTSAGPGSRISSSSFSRVGSSNFRGGGGYGGASGM
GGITAVTVNQSLLSPLVLEVDPNIQAVRTQEKEQIKTLNNKFASFIDKVRFLEQQNKMLE
TKWSLLQQQKTARSNMDNMFESYINNLRRQLETLGQEKLKLEAELGNMQGLVEDFKNKYE
DEINKRTEMENEFVLIKKDVDEAYMNKVELESRLEGLTDEINFLRQLYEEEIRELQSQIS
DTSVVLSMDNSRSLDMDSIIAEVKAQYEDIANRSRAEAESMYQIKYEELQSLAGKHGDDL
RRTKTEISEMNRNISRLQAEIEGLKGQRASLEAAIADAERQGELAITKDANAKLSELEAAL
QRAKQDMARQLREYQELMVKLALDIEIATYRKLLGEESRLESGMQNMSIHTKTTSGYA
GGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFRTSSRAVVVKIETRDGKLVSESSDV
LPK

>sp|P05787|K2C8_HUMAN Keratin, type II cytoskeletal 8 OS=Homo sapiens OX=9606 GN=KRT8 PE=1 SV=7
MSIRVTQSYKVSTSGPRAFSSRSYTSAGPGSRISSSSFSRVGSSNFRGGGGYGGASGM
GGITAVTVNQSLLSPLVLEVDPNIQAVRTQEKEQIKTLNNKFASFIDKVRFLEQQNKMLE
TKWSLLQQQKTARSNMDNMFESYINNLRRQLETLGQEKLKLEAELGNMQGLVEDFKNKYE
DEINKRTEMENEFVLIKKDVDEAYMNKVELESRLEGLTDEINFLRQLYEEEIRELQSQIS
DTSVVLSMDNSRSLDMDSIIAEVKAQYEDIANRSRAEAESMYQIKYEELQSLAGKHGDDL
RRTKTEISEMNRNISRLQAEIEGLKGQRASLEAAIADAEGRGELAIKDANAKLSELEAAL
QRAKQDMARQLREYQELMVKLALDIEIATYRKLLGEES [REDACTED] IHTKTTSGYA
GGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFRTSSRAVVVKIETRDGKLVSESSDV
LPK

>sp|P05787|K2C8_HUMAN Keratin, type II cytoskeletal 8 OS=Homo sapiens OX=9606 GN=KRT8 PE=1 SV=7
MSIRVTQSKYKVSTSGPRAFSSRSYTSAGPSRISSSSFSGVSSNFRGGGGYGGASGM
GGITAVTVNQSLLSPLVLEVDPNIQAVRTQEKEQIKTLNNKFASFIDKVRFLEQQNKMLE
TKWSLLQQQKTARSNMDNMFESYINNLRRQLETLGQEKLKLEAELGNMQGLVEDFKNKYE
DEINKRTEMENEFVLIKKDVDEAYMNKVELESRLEGLTDEINFLRQLYEEEIRELQSQIS
DTSVVLSMDNSRSLDMDSIIAEVKAQYEDIANRSRAEAESMYQIKYEELQSLAGKHGDDL
RRTKTEISEMNRNISRLQAEIEGLKGQRASLEAAIADAEGRGELAIKDANAKLSELEAAL
QRAKQDMARQLREYQELMVKLALDIEIATYRKLLGEES [REDACTED] IHTKTTSGYA
GGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFSSRTSSRAVVVKIETRDGKLVSESSDV
LPK

KRT8-tail-long.fasta

Let's visit the HMMER website:

<https://www.ebi.ac.uk/Tools/hmmer/search>



HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

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 ⓘ

 ▼ Sequence Database ⓘFrequently used databases: [Reference Proteomes](#) [UniProtKB](#) [SwissProt](#) [PDB](#) [Ensembl](#)

Current database selection:

► Restrict by Taxonomy ⓘ**▼ Cut-Offs ⓘ** E-value Bit score<https://www.ebi.ac.uk/Tools/hmmer/search>

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

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 ⓘ

Click

Submit Reset

▼ Sequence Database ⓘ

Frequently used databases: Reference Proteomes UniProtKB SwissProt PDB Ensembl

Current database selection:

Reference Proteomes

► Restrict by Taxonomy ⓘ

▼ Cut-Offs ⓘ

E-value Bit score

<https://www.ebi.ac.uk/Tools/hmmer/search>



HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmmer hmmscan hmmsearch jackhmmer

Iterative search vs protein sequence database

Paste a Sequence or an Alignment | Upload a File | Accession Search

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

Submit Reset

▼ Sequence Database

Frequently used databases: [Reference Proteomes](#) [UniProtKB](#) [SwissProt](#) [PDB](#) [Ensembl](#)

Current database selection:

Reference Proteomes

► Restrict by Taxonomy

▼ Cut-Offs

<https://www.ebi.ac.uk/Tools/hmmer/search>

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmer hmmscan hmmsearch jackhmmer

Iterative search vs protein sequence database

Paste a Sequence or an Alignment | Upload a File | Accession Search

Paste in your sequence (example), HMM (example) or mmseqs sequence alignment (example) 

Click

Submit

Reset

▼ Sequence Database

Frequently used databases: Reference Proteomes UniProtKB SwissProt

KRT8-tail-long.fasta

Current database selection:

Reference Proteomes

► Restrict by Taxonomy

▼ Cut-Offs

<https://www.ebi.ac.uk/Tools/hmmer/search>

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmmer hmmscan hmmsearch jackhmmer

Iterative search vs protein sequence database

Paste a Sequence or an Alignment | Upload a File | Accession Search

Paste in your sequence (example), HMM (example) or multiple sequence alignment (example) ⓘ



or Copy and Paste Sequence

Submit Reset

▼ Sequence Database ⓘ

Frequently used databases: Reference Proteomes UniProtKB SwissProt PDB Ensembl

Current database selection:

Reference Proteomes

► Restrict by Taxonomy ⓘ

▼ Cut-Offs ⓘ

<https://www.ebi.ac.uk/Tools/hmmer/search>



HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmmer

hmmscan

hmmsearch

jackhmmer

Iterative search vs protein sequence database

Paste a Sequence or an Alignment | Upload a File | Accession Search

Paste in your sequence (example), HMM (example) or multiple sequence alignment (example) ⓘ

```
>KRT8_tail_long
IHTKTTSGYAGGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFSRTSSSRAVVKKIETRDGKLVSESSDVLPK
```

Submit

Reset

▼ Sequence Database ⓘ

Frequently used databases:

[Reference Proteomes](#) [UniProtKB](#) [SwissProt](#) [PDB](#) [Ensembl](#)

Current database selection:

Reference Proteomes

► Restrict by Taxonomy ⓘ

▼ Cut-Offs ⓘ

<https://www.ebi.ac.uk/Tools/hmmer/search>

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

▼ Sequence Database

Frequently used databases:

[Reference Proteomes](#)[UniProtKB](#)[SwissProt](#)[PDB](#)[Ensembl](#)

Current database selection:

Reference Proteomes



► Restrict by Taxonomy

▼ Cut-Offs

 E-value Bit score

Significance E-values:

0.01

Hit

0.03

Report E-values:

1

Hit

1

EMBL-EBI Services Research Training About us  EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmmer

hmmscan

hmmsearch

jackhmmer

Iterative search vs protein sequence database

Paste a Sequence or an Alignment | Upload a File | Accession Search

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

```
>KRT8_tail_long
IHTKTTSGYAGGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFSRTSSRAVVKKIETRDGKLVSESSDVLPK
```

Submit

Click Submit

▼ Sequence Database

Frequently used databases: [Reference Proteomes](#) [UniProtKB](#) [SwissProt](#) [PDB](#) [Ensembl](#)

Current database selection:

Reference Proteomes

► Restrict by Taxonomy

▼ Cut-Offs

<https://www.ebi.ac.uk/Tools/hmmer/search>

JACKHMMER Results

[Search Again](#)

Score

Taxonomy

Domain

Download

Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)Sequence selection [?](#)

- above threshold [?](#)
- unselect all [?](#)

[Continue the search](#)[Start iteration 2](#)Distribution of Significant Hits [?](#)

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

« First « Previous Page 1 of 6 Next » Last »

Significant Query Matches (255) in *uniprotrefprot* (v.2018_08)[Customise](#)

Target	Description	Species	Cross-references	E-value
> K2C8_HUMAN	Keratin, type II cytoskeletal 8	Homo sapiens		1.5e-35 ✓
> K2C8_HUMAN	Isoform 2 of Keratin, type II cytoskeletal 8	Homo sapiens		1.6e-35 ✓
> A0A096NQU2_PAPAN	Uncharacterized protein	Papio anubis		3.1e-35 ✓
> G1S7N4_NOMLE	Uncharacterized protein	Nomascus leucogenys		4.3e-35 ✓
> A0A2K5L5Z8_CERAT	Uncharacterized protein	Cercopithecus atys		5.6e-35 ✓
> A0A2J8KDC3_PANTR	KRT8 isoform 11	Pan troglodytes		5.8e-35 ✓
> F6U9A3_MACMU	Keratin 8	Macaca mulatta		5.8e-35 ✓
> A0A2K6NLD5_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2K6NVV1_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2R9A6T6_PANPA	Keratin 8	Pan paniscus		5.8e-35 ✓
> A0A2K5N310_CERAT	Keratin 8	Cercopithecus atys		5.8e-35 ✓
> A0A2K6BEC7_MACNE	Uncharacterized protein	Macaca nemestrina		5.8e-35 ✓
> F7HME7_MACMU	Keratin 8	Macaca mulatta		6.2e-35 ✓
> A0A2K6BED4_MACNE	Uncharacterized protein	Macaca nemestrina		6.4e-35 ✓
> A0A2K5N2V1_CERAT	Keratin 8	Cercopithecus atys		6.4e-35 ✓
> A0A096MSQ9_PAPAN	Uncharacterized protein	Papio anubis		6.4e-35 ✓
> A0A2K6P3S2_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		1.2e-34 ✓

JACKHMMER Results

[Search Again](#)

Score Taxonomy Domain Download

Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)

Sequence selection [?](#)

above threshold [?](#)
 unselect all [?](#)

[Continue the search](#) **Start iteration 2**

Distribution of Significant Hits [?](#)



more significant

Legend: Bacteria, Eukaryota, Archaea, Viruses, Unclassified Sequences, Other Sequences

« First « Previous Page 1 of 6 Next » Last »

Significant Query 18.08 [Customise](#)

255 hits

Target	Species	Cross-references	E-value
> K2C8_HUMAN	Homo sapiens		1.5e-35 ✓
> K2C8_HUMAN	Homo sapiens		1.6e-35 ✓
> A0A096NQU2_PAPAN	Papio anubis		3.1e-35 ✓
> G1S7N4_NOMLE	Nomascus leucogenys		4.3e-35 ✓
> A0A2K5L5Z8_CERAT	Cercocebus atys		5.6e-35 ✓
> A0A2J8KDC3_PANTR	Pan troglodytes		5.8e-35 ✓
> F6U9A3_MACMU	Macaca mulatta		5.8e-35 ✓
> A0A2K6NLD5_RHIRO	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2K6NVV1_RHIRO	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2R9A6T6_PANPA	Pan paniscus		5.8e-35 ✓
> A0A2K5N310_CERAT	Cercocebus atys		5.8e-35 ✓
> A0A2K6BEC7_MACNE	Macaca nemestrina		5.8e-35 ✓
> F7HME7_MACMU	Macaca mulatta		6.2e-35 ✓
> A0A2K6BED4_MACNE	Macaca nemestrina		6.4e-35 ✓
> A0A2K5N2V1_CERAT	Cercocebus atys		6.4e-35 ✓
> A0A096MSQ9_PAPAN	Papio anubis		6.4e-35 ✓
> A0A2K6P3S2_RHIRO	Rhinopithecus roxellana		1.2e-34 ✓

Home | Search | Results | Software | Help | About | Contact

JACKHMMER Results

Score Taxonomy Domain Download

Iteration 1

Return to the Results Summary | Jump to threshold

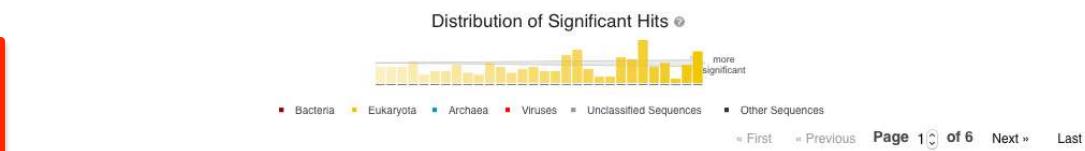
Sequence selection

above threshold ([?](#))

unselect all ([?](#))

Continue the search Start iteration 2

KRT8 is our top hit



Significant Query Matches (255) in *uniprotrefprot* (v.2018_08)

Customise

Target	Description	Species	Cross-references	E-value
> K2C8_HUMAN	Keratin, type II cytoskeletal 8	Homo sapiens		1.5e-35 ✓
> K2C8_HUMAN	Isoform 2 of Keratin, type II cytoskeletal 8	Homo sapiens		1.6e-35 ✓
> A0A096NQU2_PAPAN	Uncharacterized protein	Papio anubis		3.1e-35 ✓
> G1S7N4_NOMLE	Uncharacterized protein	Nomascus leucogenys		4.3e-35 ✓
> A0A2K5L5Z8_CERAT	Uncharacterized protein	Cercopithecus atys		5.6e-35 ✓
> A0A2J8KDC3_PANTR	KRT8 isoform 11	Pan troglodytes		5.8e-35 ✓
> F6U9A3_MACMU	Keratin 8	Macaca mulatta		5.8e-35 ✓
> A0A2K6NLD5_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2K6NVV1_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2R9A6T6_PANPA	Keratin 8	Pan paniscus		5.8e-35 ✓
> A0A2K5N310_CERAT	Keratin 8	Cercopithecus atys		5.8e-35 ✓
> A0A2K6BEC7_MACNE	Uncharacterized protein	Macaca nemestrina		5.8e-35 ✓
> F7HME7_MACMU	Keratin 8	Macaca mulatta		6.2e-35 ✓
> A0A2K6BED4_MACNE	Uncharacterized protein	Macaca nemestrina		6.4e-35 ✓
> A0A2K5N2V1_CERAT	Keratin 8	Cercopithecus atys		6.4e-35 ✓
> A0A096MSQ9_PAPAN	Uncharacterized protein	Papio anubis		6.4e-35 ✓
> A0A2K6P3S2_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		1.2e-34 ✓

JACKHMMER Results

[Search Again](#)**Click Taxonomy**

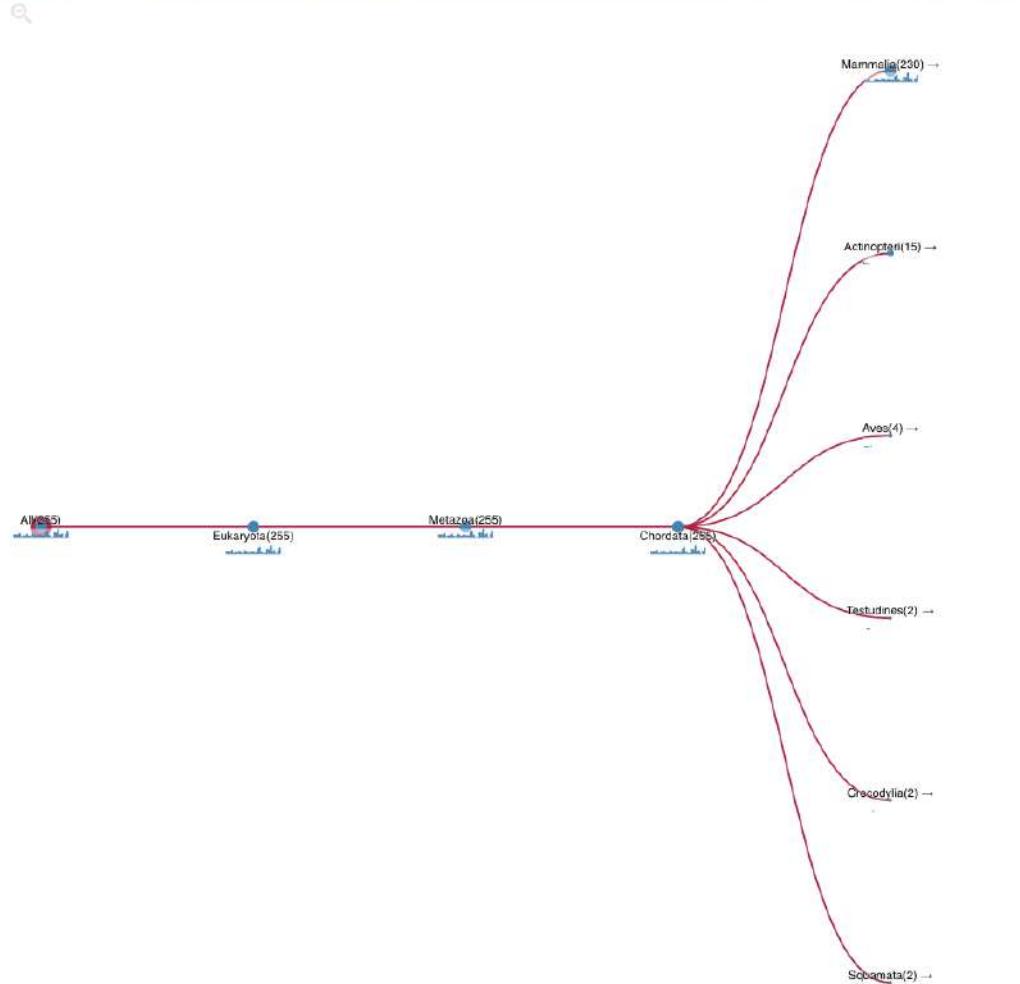
Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)Sequence selection [?](#)

- above threshold [?](#)
 unselect all [?](#)

[Continue the search](#) **Start iteration 2**Distribution of Significant Hits [?](#)[Bacteria](#) [Eukaryota](#) [Archaea](#) [Viruses](#) [Unclassified Sequences](#) [Other Sequences](#)[« First](#) [« Previous](#) **Page 1** of 6 [Next »](#) [Last »](#)Significant Query Matches (255) in *uniprotrefprot* (v.2018_08)[Customise](#)

Target	Description	Species	Cross-references	E-value
> K2C8_HUMAN	Keratin, type II cytoskeletal 8	Homo sapiens		1.5e-35 ✓
> K2C8_HUMAN	Isoform 2 of Keratin, type II cytoskeletal 8	Homo sapiens		1.6e-35 ✓
> A0A096NQU2_PAPAN	Uncharacterized protein	Papio anubis		3.1e-35 ✓
> G1S7N4_NOMLE	Uncharacterized protein	Nomascus leucogenys		4.3e-35 ✓
> A0A2K5L5Z8_CERAT	Uncharacterized protein	Cercopithecus atys		5.6e-35 ✓
> A0A2J8KDC3_PANTR	KRT8 isoform 11	Pan troglodytes		5.8e-35 ✓
> F6U9A3_MACMU	Keratin 8	Macaca mulatta		5.8e-35 ✓
> A0A2K6NLD5_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2K6NVV1_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		5.8e-35 ✓
> A0A2R9A6T6_PANPA	Keratin 8	Pan paniscus		5.8e-35 ✓
> A0A2K5N310_CERAT	Keratin 8	Cercopithecus atys		5.8e-35 ✓
> A0A2K6BEC7_MACNE	Uncharacterized protein	Macaca nemestrina		5.8e-35 ✓
> F7HME7_MACMU	Keratin 8	Macaca mulatta		6.2e-35 ✓
> A0A2K6BED4_MACNE	Uncharacterized protein	Macaca nemestrina		6.4e-35 ✓
> A0A2K5N2V1_CERAT	Keratin 8	Cercopithecus atys		6.4e-35 ✓
> A0A096MSQ9_PAPAN	Uncharacterized protein	Papio anubis		6.4e-35 ✓
> A0A2K6P3S2_RHIRO	Uncharacterized protein	Rhinopithecus roxellana		1.2e-34 ✓

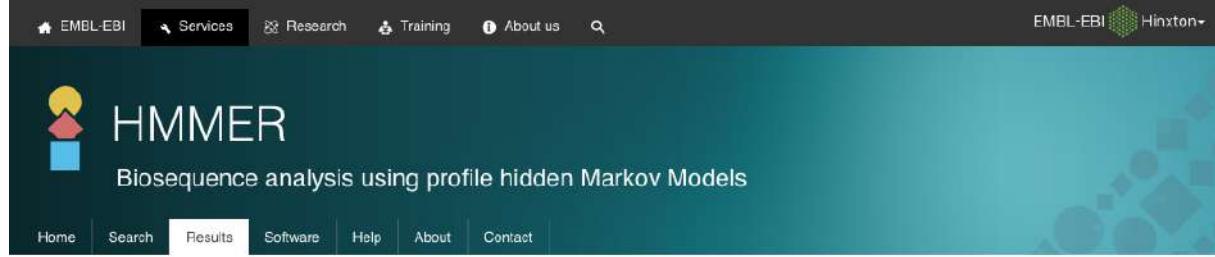


Lineage: → All

All (255)

**Species Distribution**

Species	Count	View
Aotus nancymai	14	Show
Gorilla gorilla gorilla	13	Show
Mandrillus leucophaeus	13	Show
Caimiri helveticus helveticus	44	Show



Home

Search

Results

Software

Help

About

Contact

JACKHMMER Results

Search Again

Score

Taxonomy

Domain

Click Domain

Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)

Taxonomic distribution of all search hits



A(55)

Eukaryota(255)

Metazoa(255)

Chordata(255)

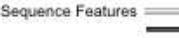
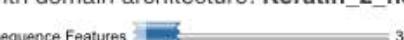
Mammalia(230) →

Actinopteri(15) →

Aves(4) →

Testudines(2) →

Domain Architectures

159 SEQUENCES	with domain architecture: Keratin_2_head, Filament , example: K2C8_HUMAN	View Scores
Show All	 483	
49 SEQUENCES	with domain architecture: Filament , example: A0A096NQU2_PAPAN	View Scores
Show All	 266	
22 SEQUENCES	with domain architecture: Keratin_2_head, Filament, Filament , example: F7GPG9_MACMU	View Scores
Show All	 395	
10 SEQUENCES	with domain architecture: Filament, Filament , example: A0A2K5M2T8_CERAT	View Scores
Show All	 258	
9 SEQUENCES	with domain architecture: Pkinase, Keratin_2_head, Filament, Filament , example: G3REJ4_GORGO	View Scores
Show All	 1649	
3 SEQUENCES	with no domain architecture, example: S9XNI4_CAMFR	View Scores
Show All	 74	
2 SEQUENCES	with domain architecture: Keratin_2_head, Filament, Filament, Filament , example: AOA2I3RRX0_PANTR	View Scores
Show All	 419	
1 SEQUENCE	with domain architecture: Keratin_2_head , example: H0XMA4_OTOGA	View Scores
Search Details	 347	

Domain Architectures

159
SEQUENCES

Show All

with domain architecture: Keratin_2_head, Filament, example: K8_HUM

Sequence Features Filament 483

Pfam domains

This image shows the arrangement of the describing that Pfam entry. The table below



View Scores

49
SEQUENCES

Show All

with domain architecture: Filament, example: A0A096NQU2_PAPAN

Sequence Features Filament 266

View Scores

22
SEQUENCES

Show All

with domain architecture: Keratin_2_head, Filament, Filament, example: F7GPG9_MACMU

Sequence Features Filament 395

View Scores

10
SEQUENCES

Show All

with domain architecture: Filament, Filament, example: A0A2K5M2T8_CERAT

Sequence Features Filament 258

View Scores

9
SEQUENCES

Show All

with domain architecture: Pkinase, Keratin_2_head, Filament, Filament, example: G3REJ4_GORGO

Sequence Features Pkinase Filament

1649

View Scores

3
SEQUENCES

Show All

with no domain architecture, example: S9XNI4_CAMFR

Sequence Features 74

View Scores

2
SEQUENCES

Show All

with domain architecture: Keratin_2_head, Filament, Filament, Filament, example: A0A2I3RRX0_PANTR

Sequence Features Filament 419

View Scores

1
SEQUENCE

Show All

with domain architecture: Keratin_2_head, example: H0XMA4_OTOGA

Sequence Features 347

View Scores

Search Details

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton+

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

JACKHMMER Results

Score **Click on Score**

Iteration 1

Return to the [Results Summary](#) | [Jump to threshold](#)

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

Domain Architectures

159 SEQUENCES	with domain architecture: Keratin_2_head, Filament, example: K2C8_HUMAN	View Scores
Show All	Sequence Features  483	
49 SEQUENCES	with domain architecture: Filament, example: A0A096NQU2_PAPAN	View Scores
Show All	Sequence Features  266	
22 SEQUENCES	with domain architecture: Keratin_2_head, Filament, Filament, example: F7GPG9_MACMU	View Scores
Show All	Sequence Features  395	
10 SEQUENCES	with domain architecture: Filament, Filament, example: A0A2K5M2T8_CERAT	View Scores
Show All	Sequence Features  258	
9 SEQUENCES	with domain architecture: PKinase, Keratin_2_head, Filament, Filament, example: G3REJ4_GORGO	View Scores
Show All	Sequence Features  1649	
3 SEQUENCES	with no domain architecture, example: S9XNI4_CAMFR	View Scores
Show All	Sequence Features  74	
2 SEQUENCES	with domain architecture: Keratin_2_head, Filament, Filament, Filament, example: A0A2I3RRX0_PANTR	View Scores
Show All	Sequence Features  419	
1 SEQUENCE	with domain architecture: Keratin_2_head, example: H0XMA4_OTOGA	View Scores
Show All	Sequence Features  347	

[Search Details](#)

JACKHMMER Results

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

Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)Sequence selection ?

- above threshold ?
 unselect all ?

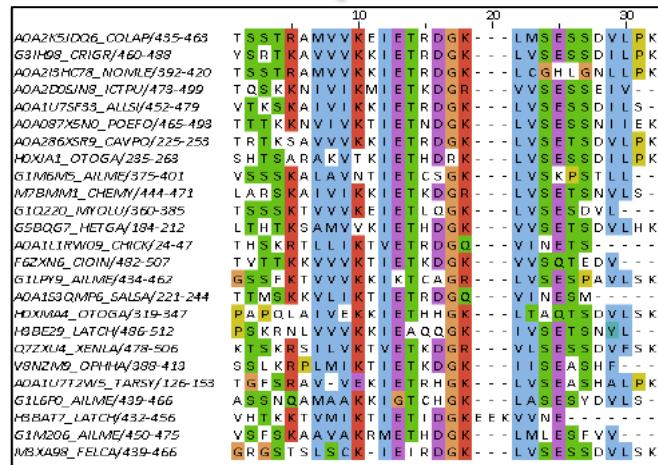
[Continue the search](#)[Start iteration 2](#)Distribution of Significant Hits ?■ Bacteria ■ Eukaryota ■ Archaea ■ Viruses ■ Unclassified Sequences ■ Other Sequences[« First](#) [« Previous](#) **Page 1** of 6 [Next »](#) [Last »](#)Significant Query Matches (255) in *uniprotrefprot* (v.2018_08)[Customise](#)

Target	Description	Species	Cross-references	E-value	
> K2C8_HUMAN	Keratin, type II cytoskeletal 8	Homo sapiens		1.5e-35	<input checked="" type="checkbox"/>
> K2C8_HUMAN	Isoform 2 of Keratin, type II cytoskeletal 8	Homo sapiens		1.6e-35	<input checked="" type="checkbox"/>
> A0A096NQU2_PAPAN	Uncharacterized protein	Papio anubis		3.1e-35	<input checked="" type="checkbox"/>
> G1S7N4_NOMLE	Uncharacterized protein	Nomascus leucogenys		4.3e-35	<input checked="" type="checkbox"/>

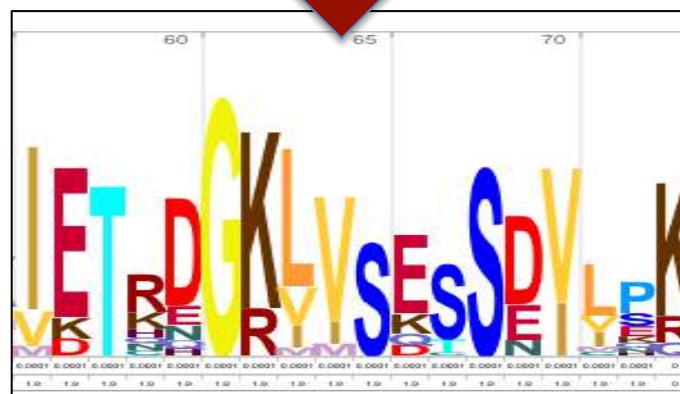
>KRT8_tail_long

IHTKTTSGYAGGLSSAYGGLTSPGLSYSLGSSFGSGA
GSSSFRTSSSRAVVVKIETRDGKLVSESSDVLPK

We run a sequence to
sequence search

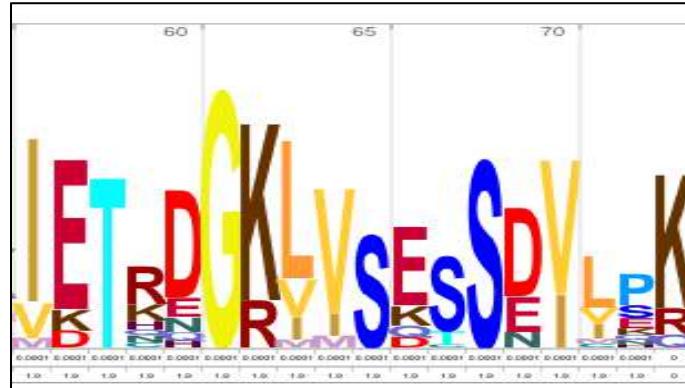


We find close homologs
of our protein



From the MSA we build
a family profile

From the MSA we build
a family profile

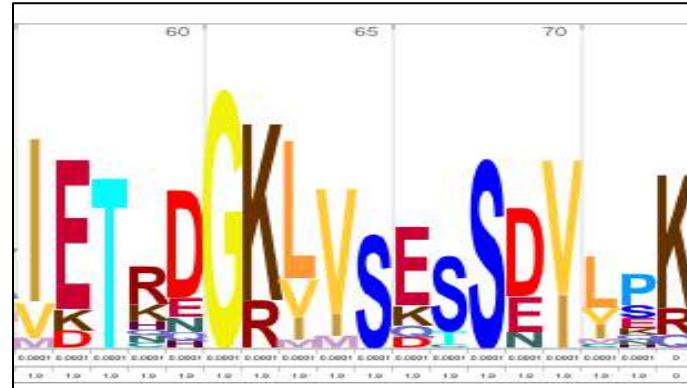


We run a profile
to sequence search

	10	20	30
A0A2K5JDQ6_COLAP/435-463	T S S T R A M V V K E I E T R D G K - - - L M S E S S D V L P K		
G3IM98_CRIGR/460-488	Y S R T K A V V V K K I E T R D G K - - - L V S E S S D I L P K		
F6YKZ7_HORSE/457-485	I G P P S A I V V V K K I E T R D G K - - - L V S E S S D I L P K		
A0A2BHC78_NOMLE/392-420	T S S T R A M V V K K I E T R D G K - - - L C Q H L G N L L P K		
A0A2D0SIN8_ICTPU/473-499	T Q S K K N I V I K M I E T K D G R - - - V V S E S S E I V - -		
A0A1U7SF33_ALSLV/452-479	V T K S K A I V I V K K I E T R D G K - - - L V S E S S D I L S -		
A0A087X5N0_POEFO/465-493	T T T K K N V I V K T I E T N D G K - - - L V S E S S S N I I E K		
A0A286XSR9_CAVPO/225-253	T R T K S A V V V K K I E T R D G K - - - L V S E F T S D V L P K		
H2L7BB_ORYLA/463-489	T Q S K K S V V V K T I E T K D G R - - - V V S E S S E I I - -		
HDXIA1_OTOGA/235-263	S H T S A R A K V T K I E T H D R K - - - L V S E S S D I L P K		
G1M6M5_ALMVE/375-401	V S S S K A L A V N T I E T C S G K - - - L V S K K P S T L L - -		
M7BMM11_CHEMY/444-471	L A R S K A I V I V K K I E T K D G R - - - L V S E T S N V L S -		
K7FDH1_PELS/460-487	V S K S K T V V V K K I E T R D G R - - - L V S E T S D I L S -		
G1Q220_MYOLU/360-385	T S S S K T V V V K E I E T L Q G K - - - L V S E S D V L - - -		
G5BQG7_HETGA/184-212	L T H S K T S A M V W K I E T H D G K - - - V V S E T S D V L H K		
W5ULZ9_ICTPU/446-470	I H S K K T V V I K T I E T R D G E - - - S K V V S E S - - -		
I3JR18_OREN/488-514	T Q T K K N V V I K M I E T K D G K - - - V V S E S S E V I - -		
H2RK76_TAKRU/432-455	M H S K K T V L I K T I E T R D G E - - - I I T R T S - - -		
A0A11IRW09_CHICK/24-47	T H S K R T L L I K T V E T R D G Q - - - V I N E S T S - - -		
F6ZXN6_CIONIN/482-507	T V T T K K V V V K T I E T K D G K - - - V V S Q T E D V - -		
A0A087XRH9_POEFO/456-478	T H G K K T V V V I K T V E T R D G E - - - V V K E S - - -		
A0A091EIM9_FUKDA/466-494	T P T K S A V V V K K I K T C D R K - - - L V S Q T S D V L P K		
G1LPY9_AILME/434-462	G S S F K T V V V K K I K T C A G R - - - L V S E S P A V L S K		
A0A1S3QMP6_SALSA/221-244	T T M S R K V L I K T I E T R D G Q - - - V I N E S M - - -		
HDXMA44_OTOGA/319-347	P A P Q L A I V E K K I E T H H G K - - - L T A Q T S D V L S K		
U81S84_ANAPL/402-424	A H P K R T I V V K T V E T R D G E - - - V I K E S - - -		
H5BE29_LATCH/486-512	P S K R N L V V V K K I E A Q Q G K - - - I V S E T S N Y L - -		
H2ZA20_CIOSA/191-213	T K T T K K V I V K T I E T R D G K - - - I V S S S - - -		
Q7ZXU4_XENLA/478-506	K T S K R S I L V K T V E T K D G R - - - V L S E S S D V F S K		
Q50207_DANRE/409-431	P A P G K K V V V I K T V E I R D G E - - - V V K E S - - -		
V8NZM9_OPHMA/388-413	S S L K R P L M I K T I E T K D G K - - - I I S E A S H F - - -		
A0A2BHMND8_PAPAN/415-438	L V F R K T L L I K T V E T R D G Q - - - V I N E S T S - - -		
A0A1U7T2WS_TARSY/126-153	T G F S R A V I V E K I E T R H G K - - - L V S E A S H A L P K		
G1L6P0_AILME/439-466	A S S N O C A M A A K K I G T C H G K - - - L A S E S Y D V L S -		
H8BA77_LATCH/432-456	V H I K K T V M I K T I E T I D G K E E K V V N E - - -		

What do we
expect to find?

From the MSA we build
a family profile



We run a profile
to sequence search

	10	20	30
A0A2K5JDQ6_COLAP/435-463	T S S T R A M V V K E I E T R D G K - - -	L M S E S S D V L P K	
G3IM98_CRIGR/460-488	Y S R T K A V V V K K I E T R D G K - - -	L V S E S S D I L P K	
F6YKZ7_HORSE/457-485	I G P P S A I V V V K K I E T R D G K - - -	L V S E S S D I L P K	
A0A2BHC78_NOMLE/392-420	T S S T R A M V V K K I E T R D G K - - -	L C Q H L G N L L P K	
A0A2D0SIN8_ICTPU/473-499	T Q S K K N I V I K M I E T K D G R - - -	V V S E S S E I V - -	
A0A1U7SF33_ALLSV/452-479	V T K S K A I V I V K K I E T R D G K - - -	L V S E S S D I L S -	
A0A087X5N0_POEFO/465-493	T T T K K N V V I V K T I E T N D G K - - -	L V S E S S S N I I E K	
A0A286XSR9_CAVPO/225-253	T R T K S A V V V K K I E T R D G K - - -	L V S E F T S D V L P K	
H2L7BB_ORYLA/463-489	T Q S K K S V V V I K T I E T K D G R - - -	V V S E S S E I I - -	
HDXA1I_OTOGA/235-263	S H T S A R A K V T K I E T H D R K - - -	L V S E S S D I L P K	
G1M6MS_ALMVE/375-401	V S S S K A L A V N T I E T C S G K - - -	L V S K K P S T L L - -	
M7BMM11_CHEMY/444-471	L A R S K A I V I V K K I E T K D G R - - -	L V S E T S N V L S -	
K7FDH1_PELS/460-487	V S K S K T V V V K K I E T R D G R - - -	L V S E T S D I L S -	
G1Q220_MYOLU/360-385	T S S S K T V V V V K E I E T L Q G K - - -	L V S E S D V L - -	
G5BQG7_HETGA/184-212	L T H T K S A M V W K I E T H D G K - - -	V V S E T S D V L H K	
W5ULZ9_ICTPU/446-470	I H S K K T V V I K T I E T R D G E - - -	S K V V S E S - - -	
I3JR18_ORENII/488-514	T Q T K K N V V I K M I E T K D G K - - -	V V S E S S E V I - -	
H2RK76_TAKRU/432-455	M H S K K T V V I K T I E T R D G E - - -	I I I R T S - - -	
A0A11IRW09_CHICK/24-47	T H S K R T L L I K T V E T R D G Q - - -	V I N E T S - - -	
F6ZXN6_CIONV/482-507	T V T K K V V V K T I E T K D G K - - -	V V S Q T E D V - -	
A0A087XRH9_POEFO/456-478	T H G K K T V V V I K T V E T R D G E - - -	V V K E S - - -	
A0A091EIM9_FUKDA/466-494	T P T K S A V V V K K I K T C D R K - - -	L V S Q T S D V L P K	
G1LPY9_AILME/434-462	G S S F K T V V V V K K I K T C A G R - - -	L V S E S P A V L S K	
A0A153QMP6_SALSA/221-244	T T M S R K V V I K T I E T R D G Q - - -	V I N E S M - - -	
HDXM44_OTOGA/319-347	P A P Q L A I V E K K I E T H H G K - - -	L T A Q T S D V L S K	
U8IS84_ANAPL/402-424	A H P K R T I V V V K T V E T R D G E - - -	V I K E S - - -	
H5BE29_LATCH/486-512	P S K R N L V V V V K K I E A Q Q G K - - -	I V S E T S N V L - -	
H2ZA20_CIOSA/191-213	T K T T K K V V I K T I E T R D G K - - -	I V S S S - - -	
Q7ZXU4_XENLA/478-506	K T S K R S I L V K T V E T K D G R - - -	V L S E S S D V F S K	
Q50207_DANRE/409-431	P A P G K K V V V I K T V E I R D G E - - -	V V K E S - - -	
V8NZM9_OPHMA/388-413	S S L K R P L M I K T I E T K D G K - - -	I I S E A S H F - - -	
A0A2BHMND8_PAPAN/415-438	L V F R K T L L I K T V E T R D G Q - - -	V I N E T S - - -	
A0A1U7T2WS_TARSY/126-153	T G F S R A V I V E K I E T R H G K - - -	L V S E A S H A L P K	
G1L6P0_AILME/439-466	A S S N O A M A A K K I G T C H G K - - -	L A S E S Y D V L S -	
H8BA77_LATCH/432-456	V H I K K T V M I K T I E T I D G K E E K V V N E - - -		

Previous
homologs +
more remote
homologs

JACKHMMER Results

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

Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)Sequence selection ?

- above threshold ?
- unselect all ?

[Continue the search](#)**Start iteration 2****Click Start Iteration 2**Distribution of Significant Hits ?■ Bacteria ■ Eukaryota ■ Archaea ■ Viruses ■ Unclassified Sequences ■ Other Sequences[« First](#) [« Previous](#) **Page 1** of 6 [Next »](#) [Last »](#)Significant Query Matches (255) in *uniprotrefprot* (v.2018_08)[Customise](#)

Target	Description	Species	Cross-references	E-value	
> K2C8_HUMAN	Keratin, type II cytoskeletal 8	Homo sapiens		1.5e-35	<input checked="" type="checkbox"/>
> K2C8_HUMAN	Isoform 2 of Keratin, type II cytoskeletal 8	Homo sapiens		1.6e-35	<input checked="" type="checkbox"/>
> A0A096NQU2_PAPAN	Uncharacterized protein	Papio anubis		3.1e-35	<input checked="" type="checkbox"/>
> G1S7N4_NOMLE	Uncharacterized protein	Nomascus leucogenys		4.3e-35	<input checked="" type="checkbox"/>

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton+

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

JACKHMMER Results

Search Again

Jackhmmer Summary

Iteration	Results	Hits			
		New	Lost	Dropped	Total
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330

Start Next Iteration

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton+

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

JACKHMMER Results

Search Again

Jackhmmer Summary		Hits				
Iteration	Results	New	Lost	Dropped	Total	
1	Click	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255
2	 86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330	

Start Next Iteration



HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

JACKHMMER Results

[Search Again](#)

Score

Taxonomy

Domain

Download

Iteration 2

[Return to the Results Summary](#) | [Jump to first new match](#) | [Jump to threshold](#)

Sequence selection

- above threshold
- unselect all

[Rerun iteration 3](#)

Warning, this will replace iteration 3 and delete iterations that rely on the data produced by it.

[« previous iteration](#)[next iteration »](#)

Distribution of Significant Hits



- Bacteria
- Eukaryota
- Archaea
- Viruses
- Unclassified Sequences
- Other Sequences

[« First](#) [« Previous](#) **Page 1 of 11** [Next »](#) [Last »](#)

Significant Query Matches (330) in *uniprotrefprot* (v.2018_08)

[Customise](#)

	Target	Description	Species	Cross-references	E-value
>	G1S7N4_NOMLE	Uncharacterized protein	Nomascus leucogenys		1.4e-38
>	A0A096NQU2_PAPAN	Uncharacterized protein	Papio anubis		1.7e-38
>	A0A2K5L5Z8_CERAT	Uncharacterized protein	Cercopithecus atys		1.8e-38

	Target	Description	Species	Cross-references	E-value	
>	A0A2K6FM60_PROCO	Keratin 8	Propithecus coquerelii	XXX	1.1e-35	<input checked="" type="checkbox"/>
>	A0A2R8MTC4_CALJA	Uncharacterized protein	Callithrix jacchus	XXX	1.1e-35	<input checked="" type="checkbox"/>
>	G7PIE0_MACFA	Uncharacterized protein (Fragment)	Macaca fascicularis	XXX	1.5e-35	<input checked="" type="checkbox"/>
>	A0A096NMK7_PAPAN	Uncharacterized protein	Papio anubis	XXX	2.5e-35	<input checked="" type="checkbox"/>
>	A0A2I3TKS8_PANTR	Uncharacterized protein	Pan troglodytes	XXX	4.3e-35	<input checked="" type="checkbox"/>
>	A0A2K6TKB3_SAIBB	Uncharacterized protein	Saimiri boliviensis boliviensis	XXX	4.5e-35	<input checked="" type="checkbox"/>
>	A0A2I2ZRA7_GORGO	Uncharacterized protein	Gorilla gorilla gorilla	XXX	5.1e-35	<input checked="" type="checkbox"/>
>	H0X5F4_OTOGA	Uncharacterized protein	Otolemur garnettii	XXX	5.6e-35	<input checked="" type="checkbox"/>
>	G3REJ4_GORGO	Uncharacterized protein	Gorilla gorilla gorilla	XXX	6.9e-35	<input checked="" type="checkbox"/>
>	A0A2R9BVP0_PANPA	Uncharacterized protein	Pan paniscus	XXX	6.9e-35	<input checked="" type="checkbox"/>
>	F6WZZ8_CALJA	Uncharacterized protein	Callithrix jacchus	XXX	7.2e-35	<input checked="" type="checkbox"/>

(show all) alignments Your search took: 2.46 secs showing rows 1 - 50 of 535

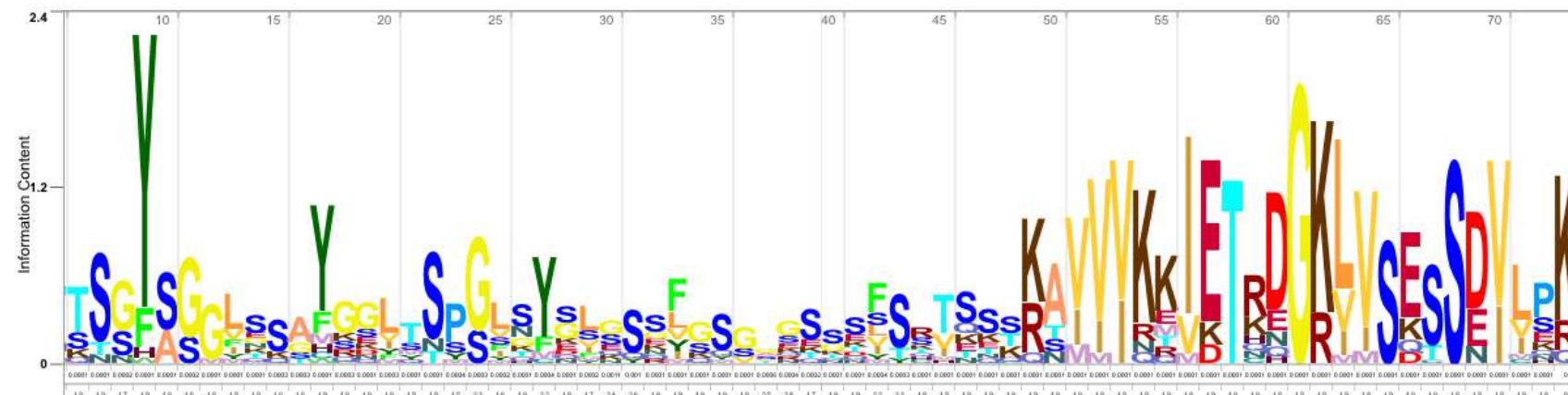
Rerun iteration 3

Warning, this will replace iteration 3 and delete iterations that rely on the data produced by it.

Search Details

« First « Previous Page 1 of 11 Next » Last »

Model Position



EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton+

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

JACKHMMER Results

Search Again

Jackhmmer Summary		Hits			
Iteration	Results	New	Lost	Dropped	Total
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330

Start Next Iteration

Click

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

JACKHMMER Results

Search Again

Jackhmmer Summary		Hits				
Iteration	Results	New	Lost	Dropped	Total	
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255	
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330	
3	86F38F56-DB84-11E8-9267-8E60E976C163.3	+236	-	-	566	

Start Next Iteration

This is cool!

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

JACKHMMER Results

Search Again

Jackhmmer Summary		Hits				
Iteration	Results	New	Lost	Dropped	Total	
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255	
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330	
3	86F38F56-DB84-11E8-9267-8E60E976C163.3	+236	-	-	566	

Start Next Iteration

Click up to 8th iteration

Let's take a 5 mins break



Jackhmmer Summary		Hits			
Iteration	Results	New	Lost	Dropped	Total
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330
3	86F38F56-DB84-11E8-9267-8E60E976C163.3	+236	-	-	566
4	86F38F56-DB84-11E8-9267-8E60E976C163.4	+189	-	1	754
5	86F38F56-DB84-11E8-9267-8E60E976C163.5	+46	-	1	799
6	86F38F56-DB84-11E8-9267-8E60E976C163.6	+70	-1	3	865
7	86F38F56-DB84-11E8-9267-8E60E976C163.7	+224	-1	2	1086
8	86F38F56-DB84-11E8-9267-8E60E976C163.8	+1533	-	1	2618
9	86F38F56-DB84-11E8-9267-8E60E976C163.9	+1752	-9	25	4336
10	86F38F56-DB84-11E8-9267-8E60E976C163.10	+1086	-23	98	5301
11	86F38F56-DB84-11E8-9267-8E60E976C163.11	+1689	-16	66	6908
12	86F38F56-DB84-11E8-9267-8E60E976C163.12	+12124	-80	336	18616

[Start Next Iteration](#)

Loads of hits!!!

Jackhmmer Summary		Hits			
Iteration	Results	New	Lost	Dropped	Total
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330
3	86F38F56-DB84-11E8-9267-8E60E976C163.3	+236	-	-	566
4	86F38F56-DB84-11E8-9267-8E60E976C163.4	+189	-	1	754
5	86F38F56-DB84-11E8-9267-8E60E976C163.5	+46	-	1	799
6	86F38F56-DB84-11E8-9267-8E60E976C163.6	+70	-1	3	865
7	86F38F56-DB84-11E8-9267-8E60E976C163.7	+224	-1	2	1086
8	86F38F56-DB84-11E8-9267-8E60E976C163.8	+1533	-	1	2618
9	86F38F56-DB84-11E8-9267-8E60E976C163.9	+1752	-9	25	4336
10	86F38F56-DB84-11E8-9267-8E60E976C163.10	+1086	-23	98	5301
11	86F38F56-DB84-11E8-9267-8E60E976C163.11	+1689	-16	66	6908
12	86F38F56-DB84-11E8-9267-8E60E976C163.12	+12124	-80	336	18616

[Start Next Iteration](#)

Let's have a look at the hits

Empty architecture, multiple hits

Domain Architectures

« First « Previous Page 1 of 45 Next » Last »

4604
SEQUENCES

with no domain architecture, example: U6LQH6_9EIME

[View Scores](#)

[Show All](#)

2254
SEQUENCES

with domain architecture: **Filament**, example: A0A2I4AJM3_9TELE

[View Scores](#)

[Show All](#)

1616
SEQUENCES

with domain architecture: **Keratin_2_head, Filament**, example: A0A1A6HI89_NEOLE

[View Scores](#)

[Show All](#)

698
SEQUENCES

with domain architecture: **Filament_head, Filament**, example: W5K3M6_ASTMX

[View Scores](#)

[Show All](#)

618
SEQUENCES

with domain architecture: **Filament, LTD**, example: A0A094ZM56_SCHHA

[View Scores](#)

[Show All](#)

173
SEQUENCES

with domain architecture: **Keratin_2_head, Keratin_2_head, Filament**, example: A0A2K6U6W7_SAIBB

[View Scores](#)

[Show All](#)

125
SEQUENCES

with domain architecture: **Filament, Filament**, example: A0A0F8BR19_LARCR

[View Scores](#)

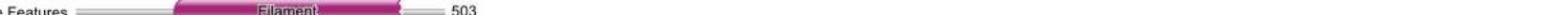
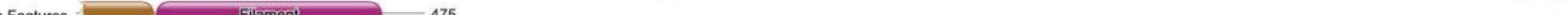
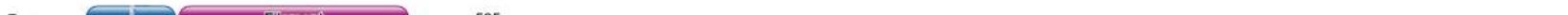
[Show All](#)

123
SEQUENCES

with domain architecture: **Ion_trans, BK_channel_a**, example: A0A2K6N6P6_RHIBE

[View Scores](#)

[Show All](#)

4604 SEQUENCES	with no domain architecture, example: U6LQH6_9EIME	View Scores
Show All		5407
2254 SEQUENCES	with domain architecture: Filament , example: A0A2I4AJM3_9TELE	View Scores
Show All		503
1616 SEQUENCES	with domain architecture: Keratin_2_head , Filament , example: A0A1A6HI89_NEOLE	View Scores
Show All		524
698 SEQUENCES	with domain architecture: Filament_head , Filament , example: W5K3M6_ASTMX	View Scores
Show All		475
618 SEQUENCES	with domain architecture: Filament , LTD , example: A0A094ZM56_SCHHA	View Scores
Show All		1854
173 SEQUENCES	with domain architecture: Keratin_2_head , Keratin_2_head , Filament , example: A0A2K6U6W7_SAIBB	View Scores
Show All		535
125 SEQUENCES	with domain architecture: Totally unrelated architecture, possible but would need further analysis	View Scores
Show All		
123 SEQUENCES	with domain architecture: Ion_trans , BK_channel_a , example: A0A2K6N6P6_RHIBE	View Scores
Show All		1236

4604
SEQUENCES

with no domain architecture, example: U6LQH6_9EIME

[View Scores](#)[Show All](#)**2254**
SEQUENCESwith domain architecture: **Filament**, example: A0A2I4AJM3_9TELE[View Scores](#)Sequence Features  503[Show All](#)**1616**
SEQUENCESwith domain architecture: **Keratin_2_head, Filament**, example: A1A6HI89_NEOLE[View Scores](#)Sequence Features  524[Show All](#)**698**
SEQUENCESwith domain architecture: **Filament_head, Filament**, example: W5K3M6_ASTMX[View Scores](#)Sequence Features  475[Show All](#)**618**
SEQUENCESwith domain architecture: **Filament, LTD**, example: A0A094ZM56_SCHHA[View Scores](#)Sequence Features  1854[Show All](#)**173**
SEQUENCESwith domain architecture: **Keratin_2_head, Keratin_2_head, Filament**, example: A0A2K6U6W7_SAIBB[View Scores](#)Sequence Features  535[Show All](#)**125**
SEQUENCESwith domain architecture: **Filament, Filament**, example: A0A0F8BR19_LARCR[View Scores](#)Sequence Features  406[Show All](#)**123**
SEQUENCESwith domain architecture: **Ion_trans, BK_channel_a**, example: A0A2K6N6P6_RHIBE[View Scores](#)Sequence Features  1236[Show All](#)**Hits overlap the core of unrelated domains**

What has happened?

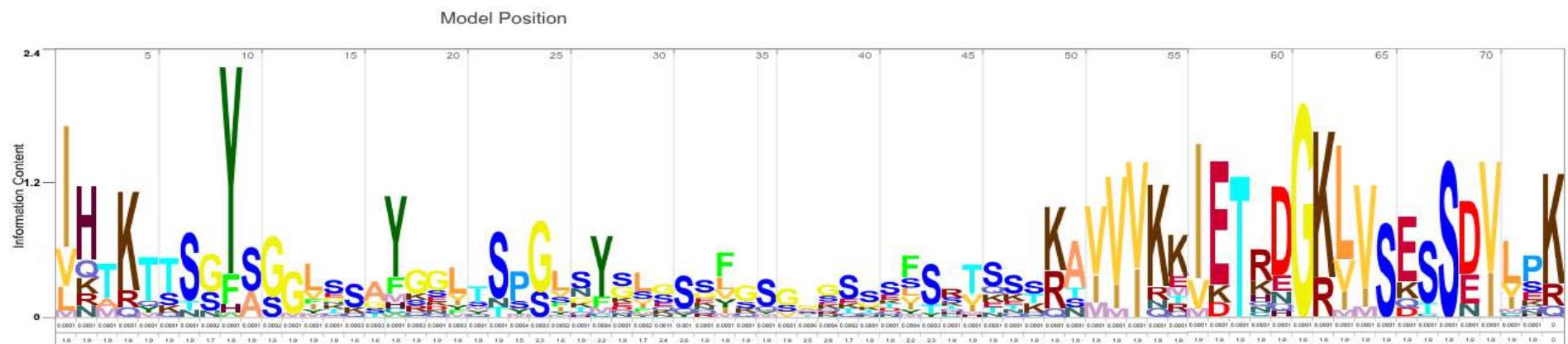
Iteration	Results	Hits			Total
		New	Lost ⓘ	Dropped ⓘ	
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330
3	86F38F56-DB84-11E8-9267-8E60E976C163.3	+236	-	-	566
4	86F38F56-DB84-11E8-9267-8E60E976C163.4	+189	-	1	754
5	86F38F56-DB84-11E8-9267-8E60E976C163.5	+46	-	1	799
6	86F38F56-DB84-11E8-9267-8E60E976C163.6	+70	-1	3	865
7	86F38F56-DB84-11E8-9267-8E60E976C163.7	+224	-1	2	1086
8	86F38F56-DB84-11E8-9267-8E60E976C163.8	+1533	-	1	2618
9	86F38F56-DB84-11E8-9267-8E60E976C163.9	+1752	-9	25	4336
10	86F38F56-DB84-11E8-9267-8E60E976C163.10	+1086	-23	98	5301
11	86F38F56-DB84-11E8-9267-8E60E976C163.11	+1689	-16	66	6908
12	86F38F56-DB84-11E8-9267-8E60E976C163.12	+12124	-80	336	18616

[Start Next Iteration](#)

We have been losing sequences! An indication that our profile is drifting away from our original sequence

Comparing profiles

After 1 Iteration

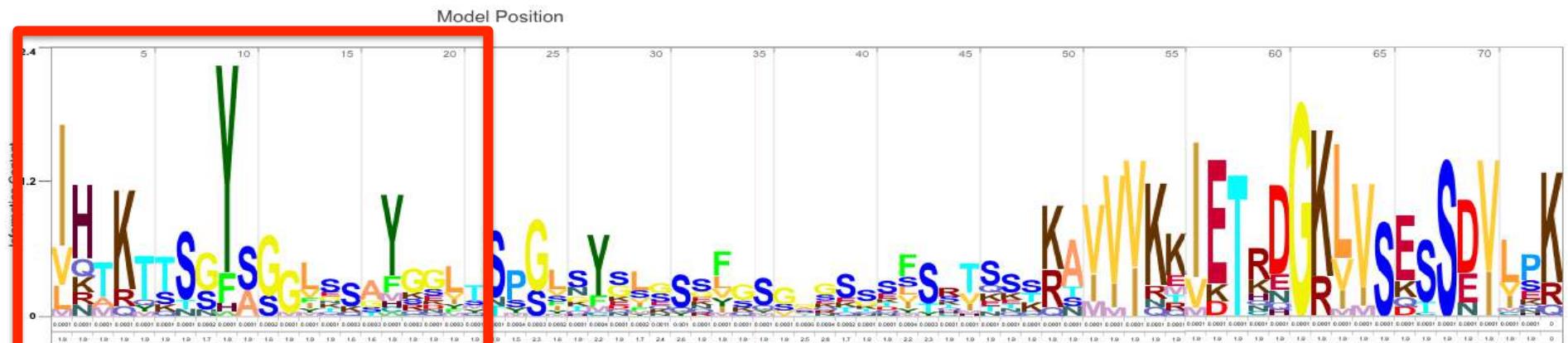


After 11 Iterations



Comparing profiles

After 1 Iteration

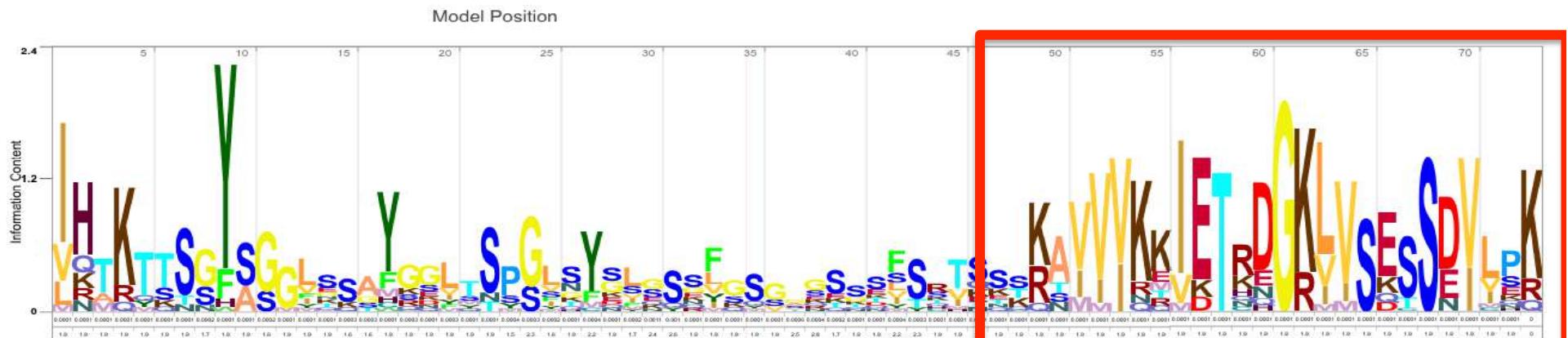


After 11 Iterations



Comparing profiles

After 1 Iteration

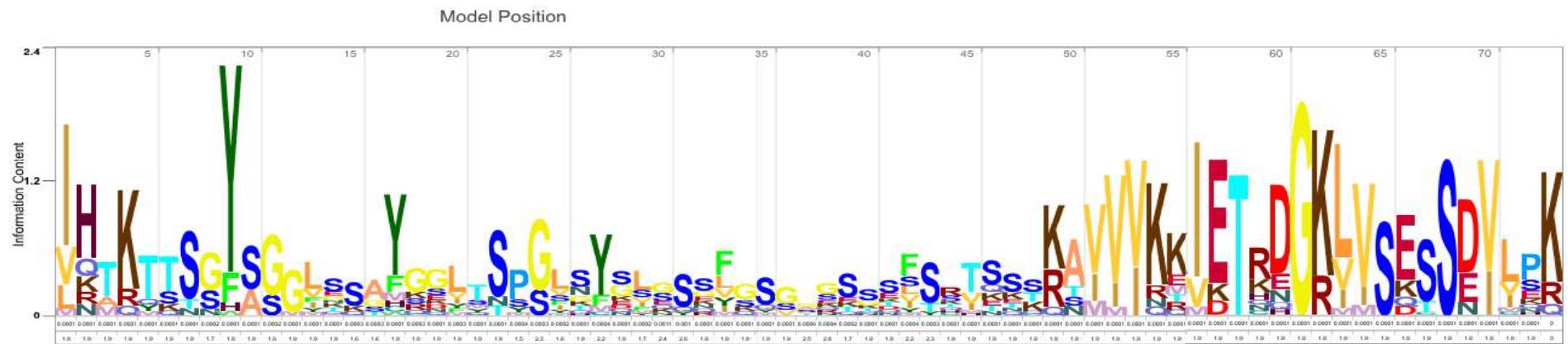


After 11 Iterations



Comparing hits profiles

After 1 Iteration



After 11 Iterations



Low complexity

Let's take a few steps back



HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

JACKHMMER Results

Search Again

Jackhmmer Summary

Iteration	Results	Hits			
		New	Lost	Dropped	Total
1	86F38F56-DB84-11E8-9267-8E60E976C163.1	+255	-	-	255
2	86F38F56-DB84-11E8-9267-8E60E976C163.2	+75	-	-	330
3	86F38F56-DB84-11E8-9267-8E60E976C163.3	+236	-	-	566
4	86F38F56-DB84-11E8-9267-8E60E976C163.4	+189	-	1	754
5	86F38F56-DB84-11E8-9267-8E60E976C163.5	+46	-	1	799
6	86F38F56-DB84-11E8-9267-8E60E976C163.6	+70	-1	3	865
7	86F38F56-DB84-11E8-9267-8E60E976C163.7	+224	-1	2	1086
8	86F38F56-DB84-11E8-9267-8E60E976C163.8	+1533	-	1	2618
9	86F38F56-DB84-11E8-9267-8E60E976C163.9	+1752	-9	25	4336
10	86F38F56-DB84-11E8-9267-8E60E976C163.10	+1086	-23	98	5301
11	86F38F56-DB84-11E8-9267-8E60E976C163.11	+1689	-16	66	6908
12	86F38F56-DB84-11E8-9267-8E60E976C163.12	+12124	-80	336	18616

Start Next Iteration

Iteration 8

[Return to the Results Summary](#) | [Jump to threshold](#)

[« previous iteration](#)

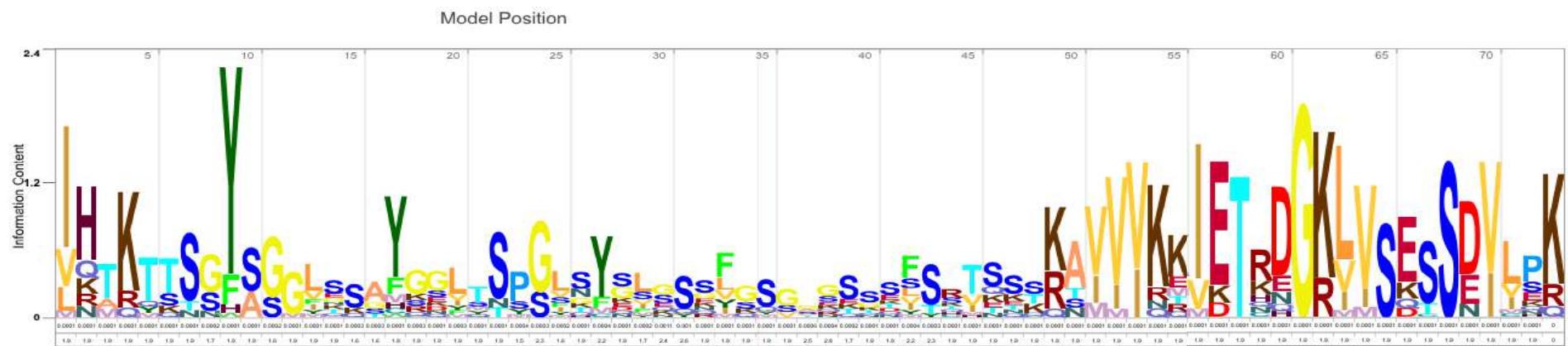
[next iteration »](#)

Domain Architectures 

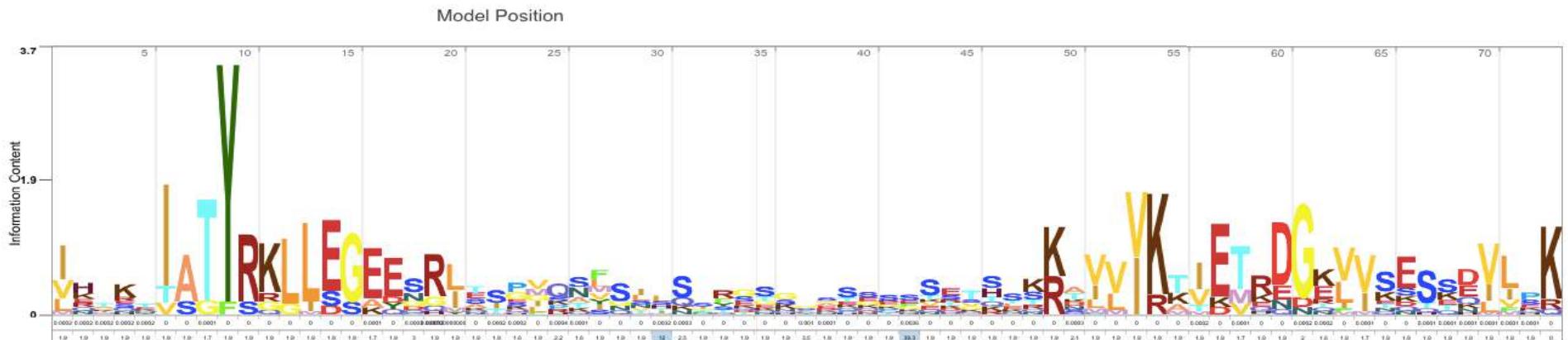
« First « Previous Page 1 of 2 Next » Last »

1101 SEQUENCES	with domain architecture: Keratin_2_head, Filament, example: K2C8_HUMAN 	View Scores
Show All		
621 SEQUENCES	with domain architecture: Filament, example: A0A096NQU2_PAPAN 	View Scores
Show All		
507 SEQUENCES	with domain architecture: Filament_head, Filament, example: H3DIE5_TETNG 	View Scores
Show All		
124 SEQUENCES	with domain architecture: Keratin_2_head, Keratin_2_head, Filament, example: M3Y5N2_MUSPF 	View Scores
Show All		
75 SEQUENCES	with domain architecture: Keratin_2_head, Filament, Filament, example: A0A2I3HEI2_NOMLE 	View Scores
Show All		
42 SEQUENCES	with domain architecture: Filament, Filament, example: A0A0F8BR19_LARCR 	View Scores
Show All		
31 SEQUENCES	with domain architecture: Filament, LTD, example: A0A087TU97_9ARAC 	View Scores
Show All		
19 SEQUENCES	with domain architecture: Filament_head, Filament, Filament, example: S9XPU7_CAMFR 	View Scores
Show All		
17 SEQUENCES	with no domain architecture, example: K7EPI4_HUMAN 	View Scores
Show All		
10 SEQUENCES	with domain architecture: Pkinase, Keratin_2_head, Filament, Filament, example: H2Q999_PANTR 	View Scores
Show All		

After 2 Iterations



After 8 Iterations





HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

**Click on Search Again
And select 'WITH THESE RESULTS'**



Search Again

Perform a new search:
 with new input
 with these results

Return to the [Results Summary](#) | 1 matches dropped below your threshold. | [Jump to first new match](#) | [Jump to threshold](#)

Sequence selection

- above threshold
- unselect all

Rerun iteration 9

Warning, this will replace iteration 9 and delete iterations that rely on the data produced by it.

< previous iteration

next iteration >

Distribution of Significant Hits



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

< First < Previous Page 1 of 87 Next > Last >

Target	Description	Species	Cross-references	E-value	
> A0A0F8BR19_LARCR	Glial fibrillary acidic protein	Larimichthys crocea		8.7e-43	<input checked="" type="checkbox"/>
> A0A218UCX9_9PASE	Keratin, type II cytoskeletal 75	Lonchura striata domestica		1.8e-35	<input checked="" type="checkbox"/>
> K2C8_HUMAN	Keratin, type II cytoskeletal 8	Homo sapiens		7.8e-34	<input checked="" type="checkbox"/>
> K2C8_HUMAN	Isoform 2 of Keratin, type II cytoskeletal 8	Homo sapiens		8.8e-34	<input checked="" type="checkbox"/>
> A0A2K5LZ8_CERAT	Uncharacterized protein	Cercopithecus atys		1.4e-33	<input checked="" type="checkbox"/>
< C1SEV2_GORILLA	Keratin 8	Gorilla gorilla gorilla		1.5e-32	<input checked="" type="checkbox"/>

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmer hmmsearch hmmscan jackhmmer

protein alignment/profile-HMM vs protein sequence database

Using Existing Search as Input

This search will be run with an HMM generated from the **2618** hits above threshold in your previous search **86F38F56-DB84-11E8-9267-8E60E976C163.8**.

Submit **Reset**

▼ Sequence Database

Frequently used databases: Reference Proteomes UniProtKB SwissProt PDB **PDB**  **Click on PDB**

Current database selection: PDB

► Restrict by Taxonomy

▼ Cut-Offs

E-value Bit score

Significance E-values: Sequence Hit
0.01 0.03

Report E-values: Sequence Hit
1 1



HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

HMMSEARCH Results

Search Again

Score

Taxonomy

Domain

Download

Distribution of Significant Hits ⓘ



Significant Query Matches (11) in pdb (v.2018-08-24)

Customise

Target	Description	Species	Cross-references	E-value
> 1gk6_A PDB RCSB	VIMENTIN	Saccharomyces cerevisiae S288C		3.8e-07
> 1gk6_A PDB RCSB	VIMENTIN	Homo sapiens		3.8e-07
> 1gk4_A PDB RCSB	VIMENTIN	Homo sapiens		2.1e-06
> 2xv5_A PDB RCSB	LAMIN-A/C	Homo sapiens		8.5e-05
> 3tyy_A PDB RCSB	Lamin-B1	Homo sapiens		0.00025
> 1x8y_A PDB RCSB	Lamin A/C	Homo sapiens		0.0019
> 3tnu_B PDB RCSB	Keratin, type II cytoskeletal 5	Homo sapiens		0.0025
> 3v4q_A PDB RCSB	Prelamin-A/C	Homo sapiens		0.0029
> 3v4w_A PDB RCSB	Prelamin-A/C	Homo sapiens		0.0054
> 3v5b_A PDB RCSB	Prelamin-A/C	Homo sapiens		0.0054
> 3tnu_A PDB RCSB	Keratin, type I cytoskeletal 14	Homo sapiens		0.0097
> 4zry_B PDB RCSB	Keratin, type II cytoskeletal 1	Homo sapiens		0.07
> 4zry_A PDB RCSB	Keratin, type I cytoskeletal 10	Homo sapiens		0.18

(show all) alignments

Your search took: 0.01 secs

showing rows 1 - 13 of 13



HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

HMMSEARCH Results

Search Again

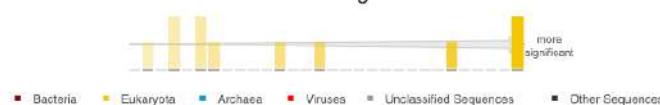
Score

Taxonomy

Domain

Download

Distribution of Significant Hits



Significant Query Matches (11) in pdb (v.2018-08-24)

Customise

Target				Description				Species			Cross-references		E-value	
Query	Target Envelope			Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value			
	start	end	start	end	start						Ind.	Cond.		
v	1gk6_A PDB RCSB	2	VIMENTIN		Saccharomyces cerevisiae S288C						3.8e-07			
	start	end	start	end	start									
	5	21	41	59	43	59	0.75	0.96	94.1 (16)	100.0 (17)	33.8	7.7e-07	2.6e-11	

.....
 Query 5 EIaTYRkLLEGeesRls 21
 EIaTYRkLLEGeesR+
 Target 43 DIATYRKILLEGESRIS 59
 PP 9*****97

>	1gk6_A PDB RCSB	2	VIMENTIN		Homo sapiens						3.8e-07			
>	1gk4_A PDB RCSB	2	VIMENTIN		Homo sapiens						2.1e-06			
>	2xv5_A PDB RCSB	2	LAMIN-A/C		Homo sapiens						8.5e-05			
>	3tyy_A PDB RCSB	2	Lamin-B1		Homo sapiens						0.00025			
>	1x8y_A PDB RCSB	2	Lamin A/C		Homo sapiens						0.0019			
>	3inu_B PDB RCSB	2	Keratin, type II cytoskeletal 5		Homo sapiens						0.0025			
>	3v4q_A PDB RCSB	2	Prelamin-A/C		Homo sapiens						0.0029			
>	3v4w_A PDB RCSB	2	Prelamin-A/C		Homo sapiens						0.0054			



Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

HMMSEARCH Results

[Search Again](#)

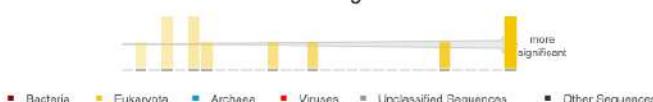
Soon

Taxonom

Domain

Download

Distribution of Significant Hits



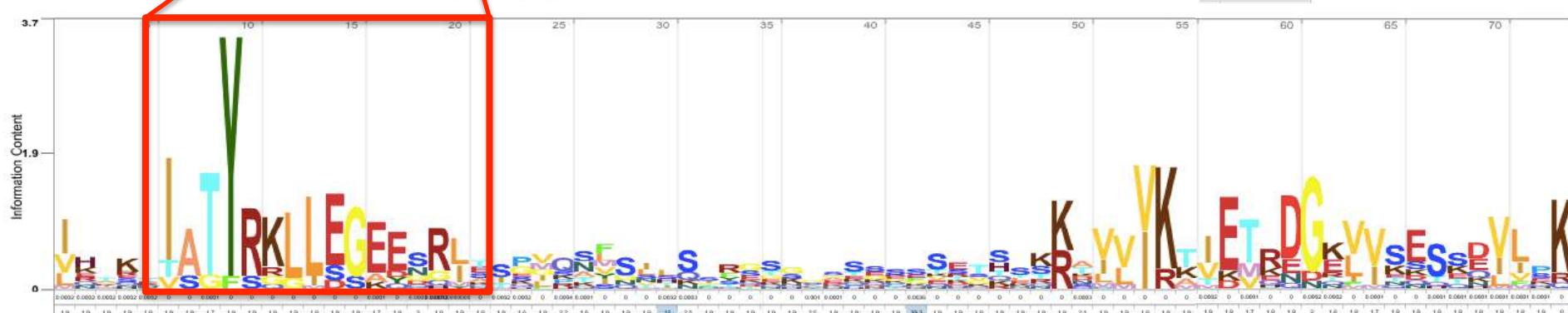
Significant Query Matches (11) in *pdb* (v.2018-08-24)

Customise

	Target		Description		Species				Cross-references		E-value		
v	1gk6_A PDBe RCSB		VIMENTIN				Saccharomyces cerevisiae S288C						3.8e-07
Query		Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value		
start	end	start	end	start	end						Ind.	Cond.	
5	21	41	59	43	59	0.75	0.96	94.1 (16)	100.0 (17)	33.8	7.7e-07	2.6e-11	

Query	5	EIA TYRkLLEGEesRIS	2
		EIA TYRkLLEGEesR+s	
Target	43	EIATYRKLLGEESRIS	5
PR		9*****9*****97	

Model Position



UniProt

UniProtKB

Advanced Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

Basket

UniProtKB - P08670 (VIME_HUMAN)

Display

[BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#)

[Feedback](#) [Help video](#) [Other tutorials and videos](#)

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (1+)

Similar proteins

Cross-references

Entry information

Miscellaneous

[Top](#)

Protein | **Vimentin**

Gene | **VIM**

Organism | *Homo sapiens (Human)*

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

Function

Vimentins are class-III intermediate filaments found in various non-epithelial cells, especially mesenchymal cells. Vimentin is attached to the nucleus, endoplasmic reticulum, and mitochondria, either laterally or terminally.  1 Publication 

Involved with LARP6 in the stabilization of type I collagen mRNAs for CO1A1 and CO1A2.  1 Publication 

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Site ⁱ	351	Stutter			1

GO - Molecular function

- double-stranded RNA binding  Source: MGI 
- identical protein binding  Source: IntAct 
- keratin filament binding  Source: AgBase 
- protein C-terminus binding  Source: UniProtKB 
- protein domain specific binding  Source: UniProtKB 
- scaffold protein binding  Source: BHF-UCL 
- structural constituent of cytoskeleton  Source: UniProtKB 
- structural constituent of eye lens  Source: Ensembl 

[View the complete GO annotation on QuickGO ...](#)

GO - Biological process

- astrocyte development  Source: Ensembl 
- Bergmann glial cell differentiation  Source: Ensembl 
- cellular response to interferon-gamma  Source: Ensembl 
- cellular response to lipopolysaccharide  Source: UniProtKB 
- cellular response to muramyl dipeptide  Source: UniProtKB 
- cytokine-mediated signaling pathway  Source: Reactome 

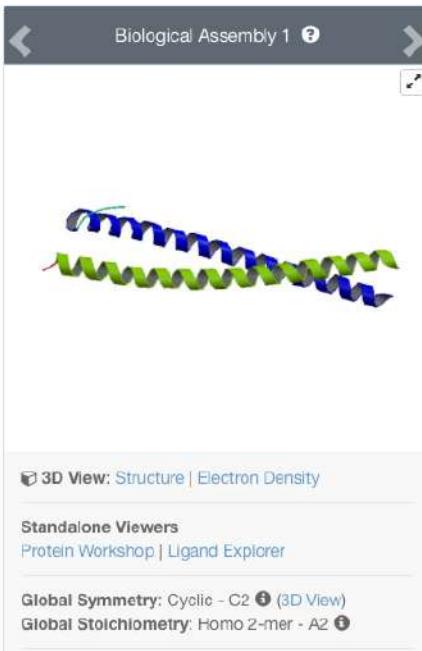
RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

RCSB PDB PROTEIN DATA BANK 145690 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands Go Advanced Search | Browse by Annotations

PDB-101 CPDB EMDDataBank NDB Worldwide Protein Data Bank Foundation

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



Macromolecule Content	
• Total Structure Weight:	13852.13
• Atom Count:	870
• Residue Count:	118
• Unique protein chains:	1

1GK6

Human vimentin coil 2B fragment linked to GCN4 leucine zipper (Z2B)

DOI: 10.2210/pdb1GK6/pdb Entry 1GK6 supersedes 1E7T

Classification: STRUCTURAL PROTEIN

Organism(s): *Saccharomyces cerevisiae* (strain ATCC 204508 / S288c), *Homo sapiens*

Expression System: *Escherichia coli* BL21(DE3)

Deposited: 2001-08-08 Released: 2002-03-15

Deposition Author(s): Strelkov, S.V., Herrmann, H., Geisler, N., Zimbelmann, R., Aebl, U., Burkhard, P.

Experimental Data Snapshot

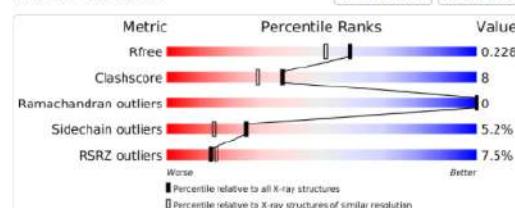
Method: X-RAY DIFFRACTION

Resolution: 1.9 Å

R-Value Free: 0.227

R-Value Work: 0.199

wwPDB Validation



This is version 1.2 of the entry. See complete history.

Literature

Download Primary Citation ▾

Conserved Segments 1A and 2B of the Intermediate Filament Dimer: Their Atomic Structures and Role in Filament Assembly.

Strelkov, S., Herrmann, H., Geisler, N., Wedig, T., Zimbelmann, R., Aebl, U., Burkhard, P.

(2002) *Embo J.* **21**: 1255

PubMed: 11889032 Search on PubMed | Search on PubMed Central

DOI: 10.1093/emboj/21.6.1255

Primary Citation of Related Structures:

1GK7, 1GK4

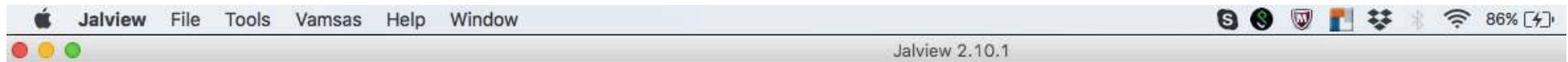
PubMed Abstract:

Intermediate filaments (IFs) are key components of the cytoskeleton in higher eukaryotic cells. The elementary IF 'building block' is an elongated coiled-coil dimer consisting of four consecutive alpha-helical segments. The segments 1A and 2B include ...+.

Let's open the alignment with
Jalview

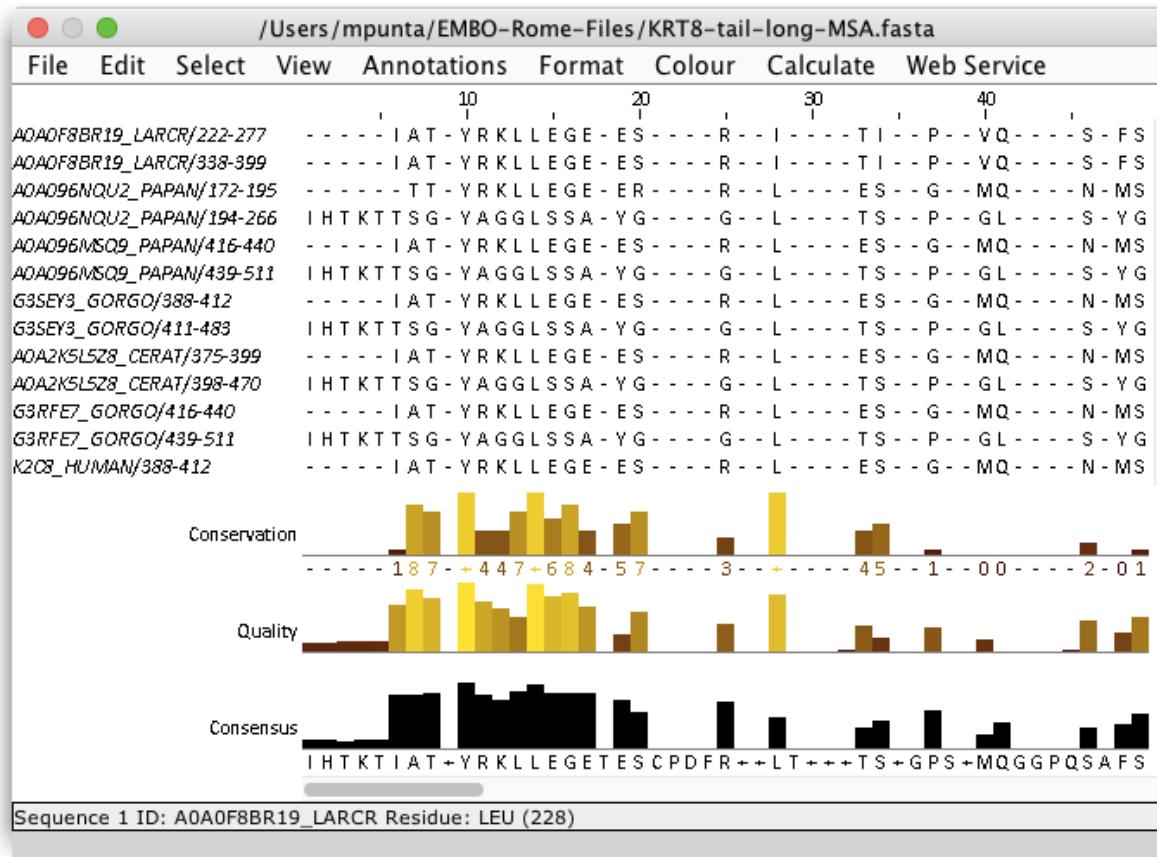
KRT8-tail-long-MSA.fasta

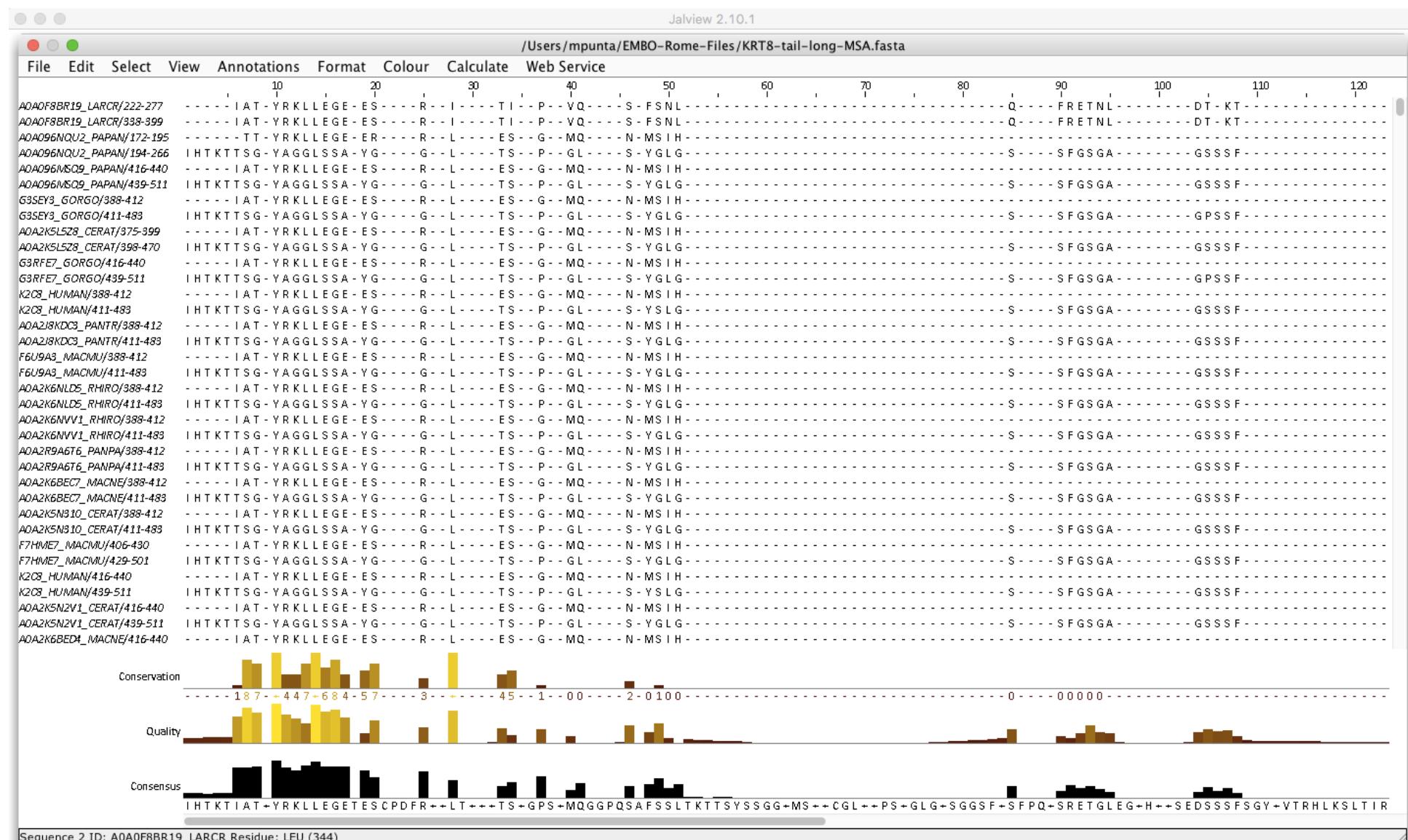
Start Jalview, close any window that might appear except for the main one, your screen should look like this

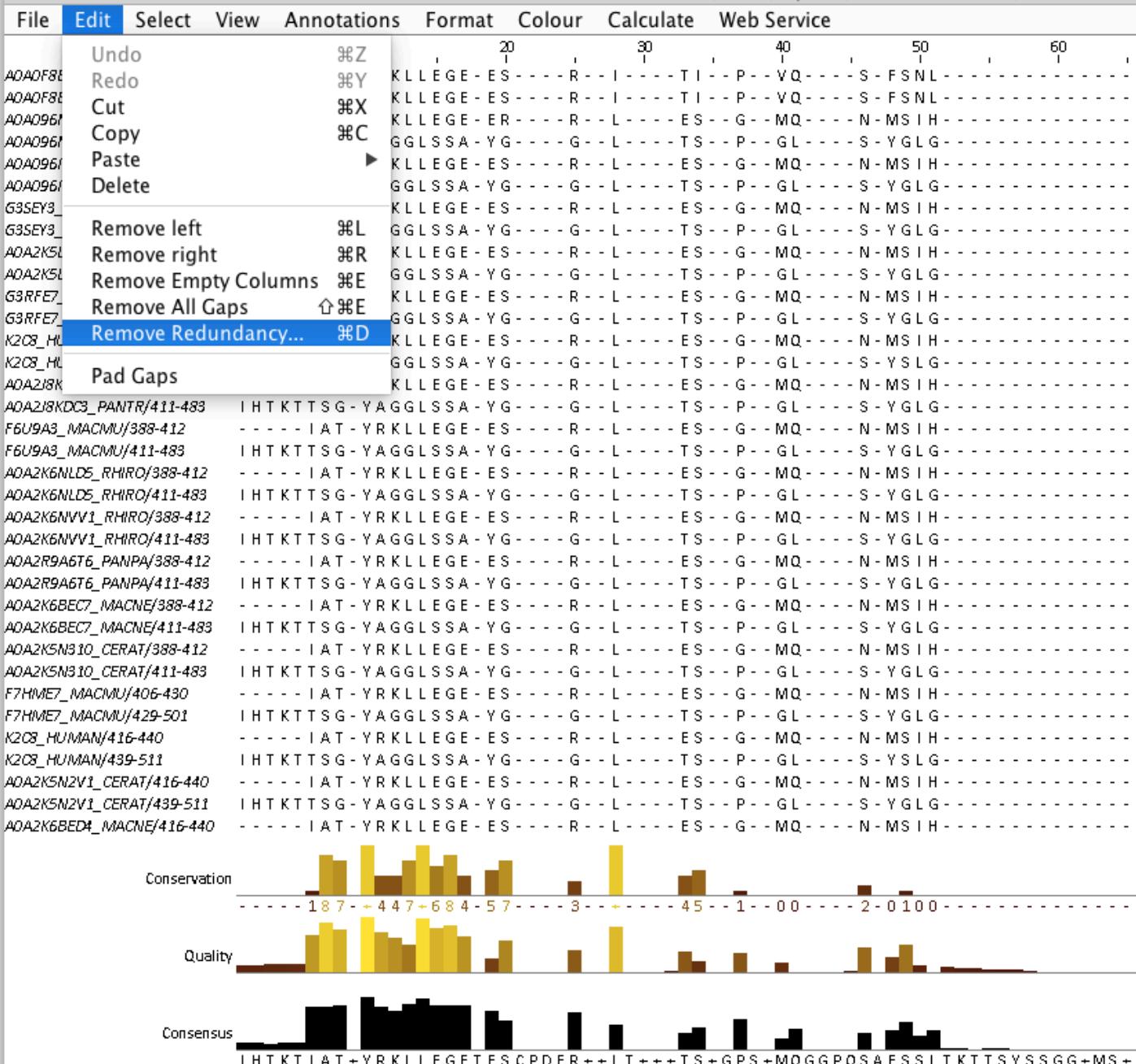


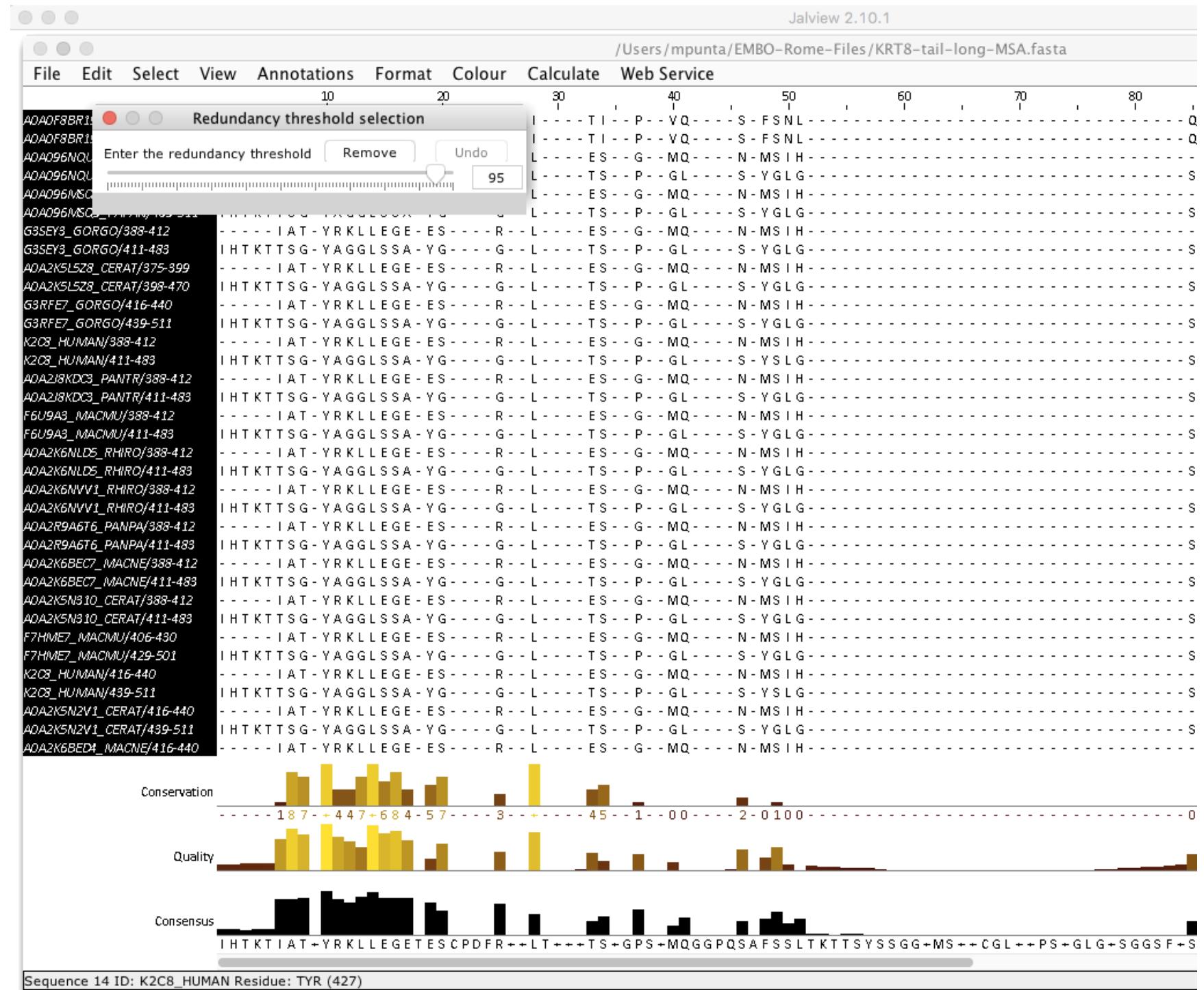


GO to the data file directory and select:
KRT8-tail-long-MSA.fasta





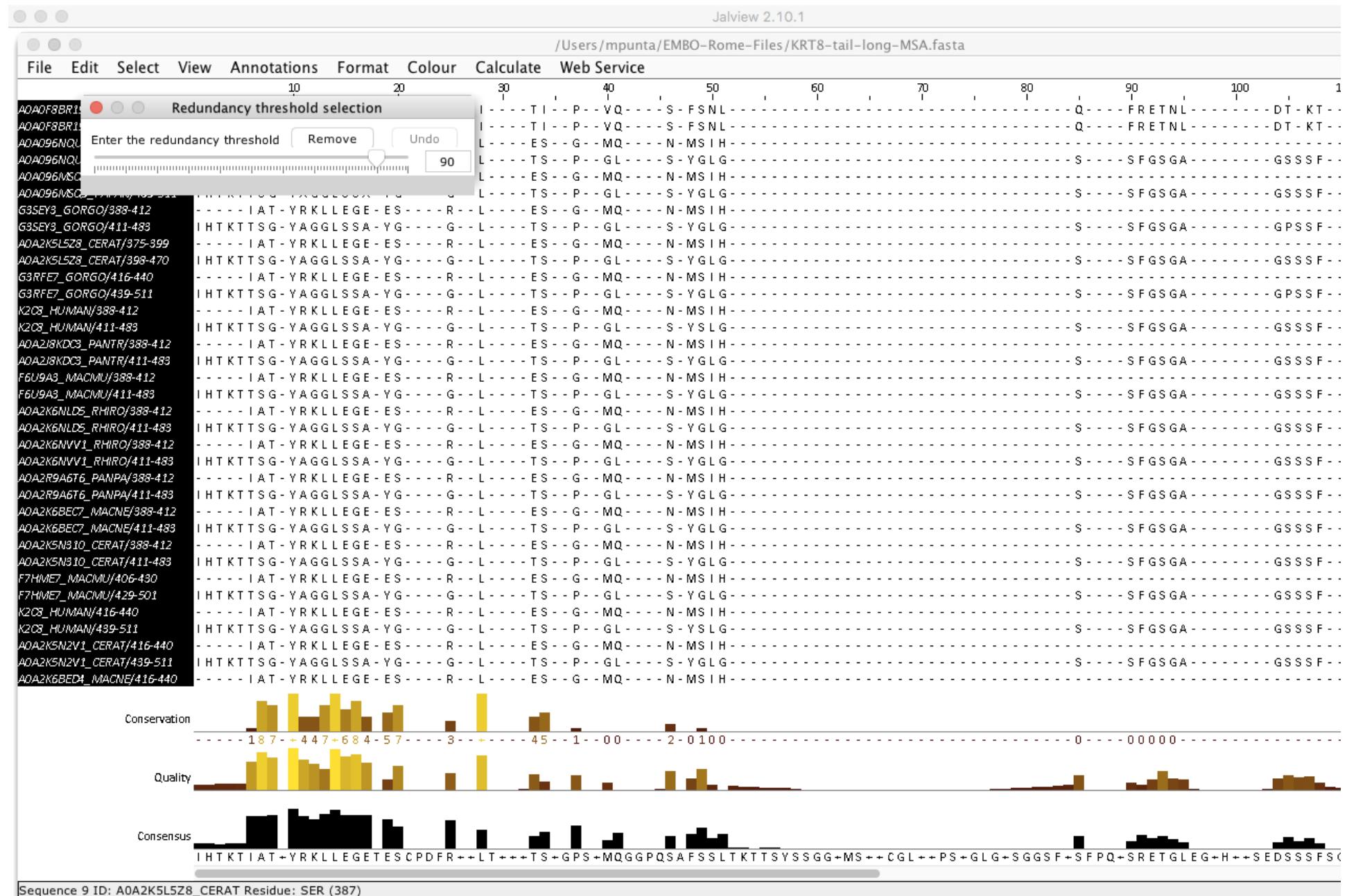




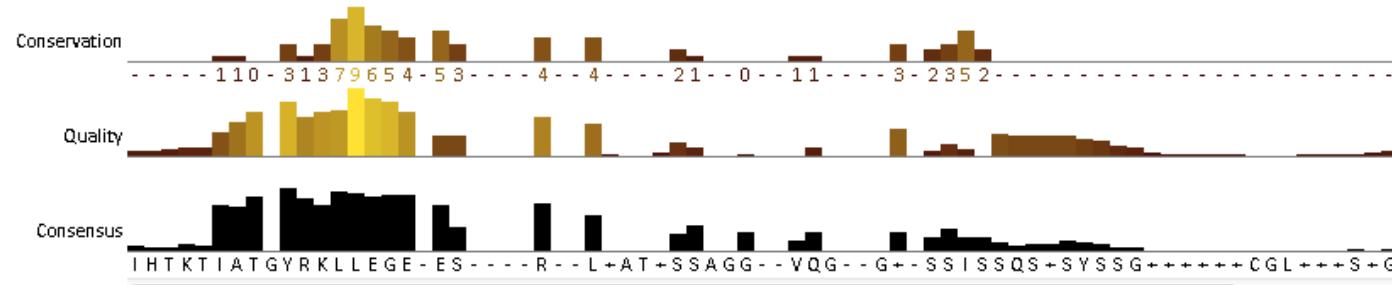
/Users/mpunta/EMBO-Rome-Files/KRT8-tail-long-MSA.fasta

File	Edit	Select	View	Annotations	Format	Colour	Calculate	Web Service
	Undo Remove Redundancy...	⌘Z			20 30 40 50 60 70 80			
AOA2K6F	Redo	⌘Y	SSA-YG-----G-L-----TS-P--GL---S-YGLS-----S					
G3UKR5	Cut	⌘X	GSA-YG-----G-L-----TS-P--GL---N-YGVss-----S					
AOA1S3C	Copy	⌘C	SSA-YG-----G-L-----TS-P--GF---N-YNMS-----S					
M8VUH2	Paste	▶	VSA-YG-----G-L-----MS-P--GL---S-YAMQ-----S					
G3IHM8	Delete		SSA-YG-----G-L-----TS-P--GF---S-YGVS-----S					
AOA1U7			NTA-YG-----G-L-----TS-P--GF---S-YGVS-----S					
I3ND60			NTS-YG-----G-L-----SS-P--GL---N-YGLS-----S					
G1SS18	Remove left	⌘L	SSG-FG-----S-L-----AS-P--GL---N-YGLSf-----qS					
L9KZY6	Remove right	⌘R	EGE-ES-----R-L-----ES-G-MQ---N-MSI Hpyed-----hqwLL					
AOA2K5J	Remove Empty Columns	⌘E	SLA-YG-----G-L-----TS-P--GL---S-YGLG-----S					
AOA2K5J	Remove All Gaps	⇧⌘E	SLA-YG-----G-L-----TS-P--GL---S-YGLG-----S					
F6SMLS			SSRnFR-----D-L-----TS-P--GL---N-YGLS-----S					
AOA2B3G	Remove Redundancy...	⌘D	SSA-YG-----D-L-----TS-P-SL---S-YGLG-----S					
F1R8W4			EGE-ES-----R-I-----SL-P-VQ---S-FSSL-----S					
AOA2K52	Pad Gaps		EGE-ES-----R-L-----ET-G-MQ---N-MSI Htkttsg-----sgggIss					
AOA2K5KH3_CERAT/410-479			IHMKTTS G-YASGLNL A-YG-----G-L-----TS-P-GS---S-SGSV-----A					
I9LD88_TUPCH/229-302			IHTKTTNG-YSGGL-T-YG-----G-L-----TS-P-SL---N-YGLSa-----fQS					
AOA2K5YC8N8_MANLE/388-456			IHTKTTSG-YANGLSSG-YK-----G-I-----TS-P-SL---S-LC-----					
AOA287A8E9_PIG/313-406			---IAT-YRKLL EGE-ES-----R-L-----ES-G-MQ---N-MSI Htkttsgy sggv-----tygtpgfnySIS					
G1P1Y9_MYOLU/414-514			---IAT-YRKLL EGE-EN-----R-L-----ES-G-MQ---N-MSI Htkttsy sgvs lpihcg lmspsyg n nyglgfqA					
AOA2K5FIY2_AOTNA/362-432			---TT-YRKLL EGE-ES-----R-L-----ES-G-MQ---N-MSI Htk-----tiT					
AOA2K6T4W3_SAIBB/380-461			---IAT-YRKLL ESE-ES-----R-L-----ES-G-MQ---N-MSI Htkttsgy-----adalySlgS					
AOA210LG9_COLLI/57-120			---IAT-YRKLL EGE-EN-----R-I-----SI-P-MH---qT-FASA-----L					
AOA2K6DCW7_MACNE/362-449			---IAT-YRKLL EGE-ES-----R-L-----ES-G-MQ---N-MSI Hmkttsgy agg-----lslasgg ltrL					
G3VJ06_SARHA/434-502			IHTKTTRGdYSSGFSSG-YG-----S-L-----TG-P-SL---N-YGLS-----S					
AOA00DRDZ6_CHLSB/396-476			---IAT-YRKLL ERE-ES-----R-L-----EF-G-MQ---N-MSI Ciyrktt-----rgskawgS					
AOA2K5K850_COLAP/401-470			-HAKTSSS-YAGGLSSA-FG-----G-L-----TS-P-RL---S-CSLG-----S					
AOA1L8FVK3_XENLA/390-452			---IAT-YRKLL EGE-ES-----R-I-----SL-P-VH---S-FSTM-----S					
AOA2R9B1G8_PANPA/383-460			---IAT-YRKLL EGE-ES-----R-L-----ES-V-IS---A-YGGLtspgi-----iyrligS					
I3JRY2_OREN/386-448			---IAT-YRKLL EGE-ES-----R-I-----ST-P-LP---N-FSSL-----N					
AOA1S3QH94_SALSA/54-138			---IAT-YRKLL EGE-ES-----R-I-----TV-S-GS---H-SAAHsaahsaahsa-----ahsaassystig					
AOA212UDL6_FELCA/357-436			---TT-YCKLL EGE-ES-----R-L-----ES-G-MQ---N-MSI Hktstg-----ylggivss					
AOA2B1MPF3_PAPAN/291-393			---IAT-YRKLL EGE-EN-----R-I-----TI-P-VQ---T-FSNL-----C					
F6WKL7_HORSE/417-492			IHTKTS-S-YSGGLSStFG-----G-L-----TS-P-GL---S-YGLS-----fQS					
AOA1S3RA50_SALSA/385-447			---IAT-YRKLL EGE-ES-----R-I-----TV-P-MQ---S-FSNL-----C					
	Conservation							
	Quality							
	Consensus							

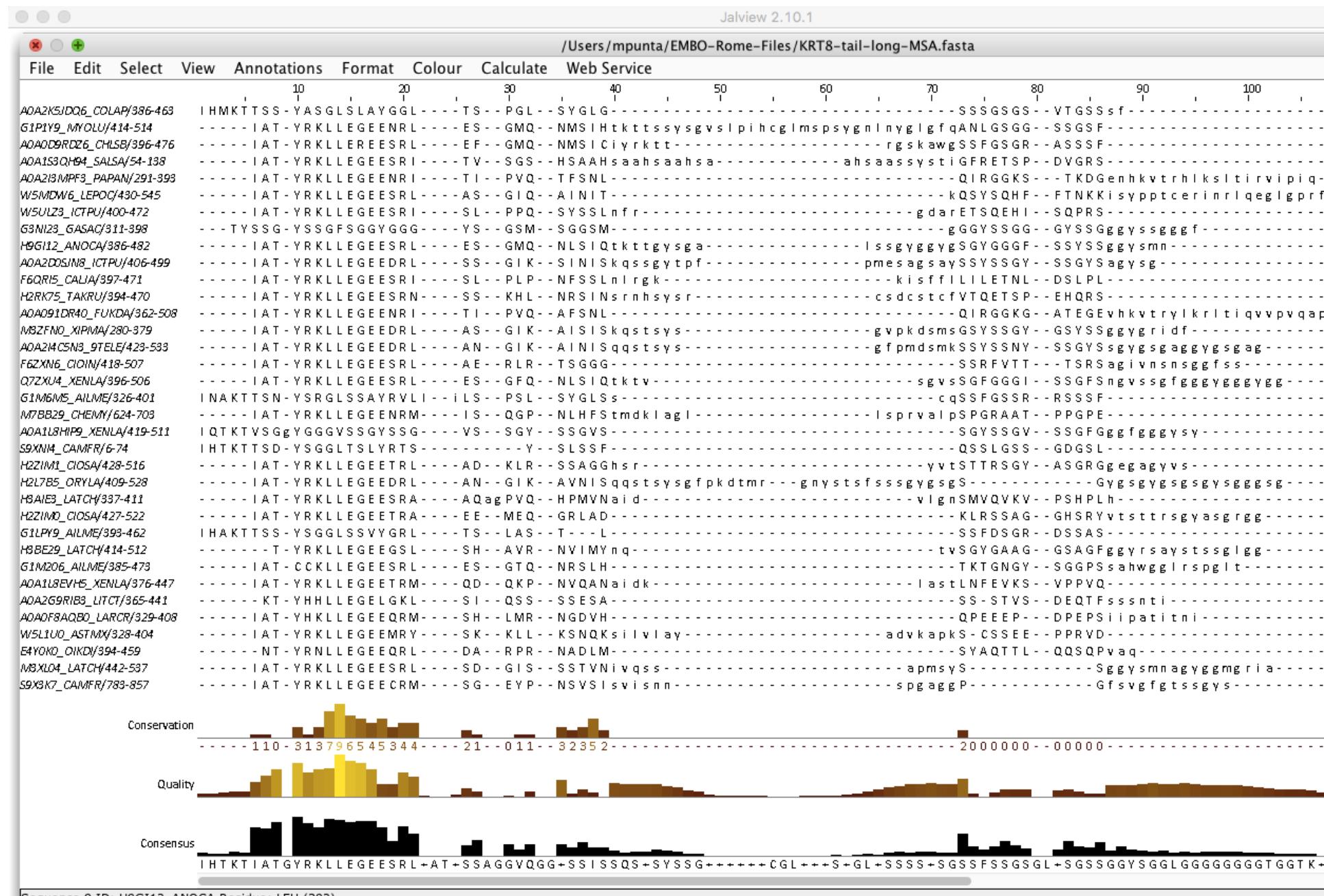
Sequence 1 ID: AOA2K6FM60_PROCO Residue: LEU (452)

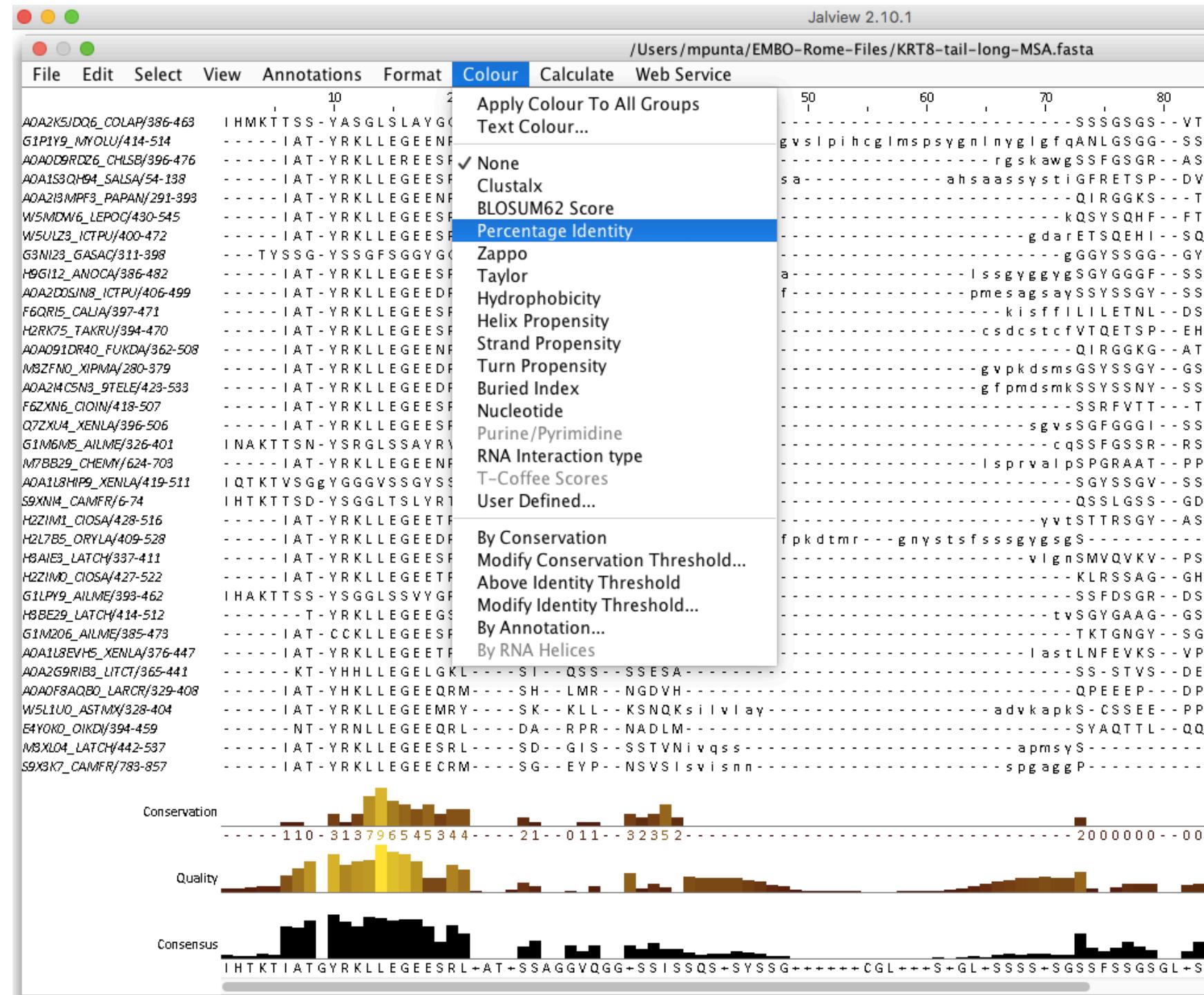


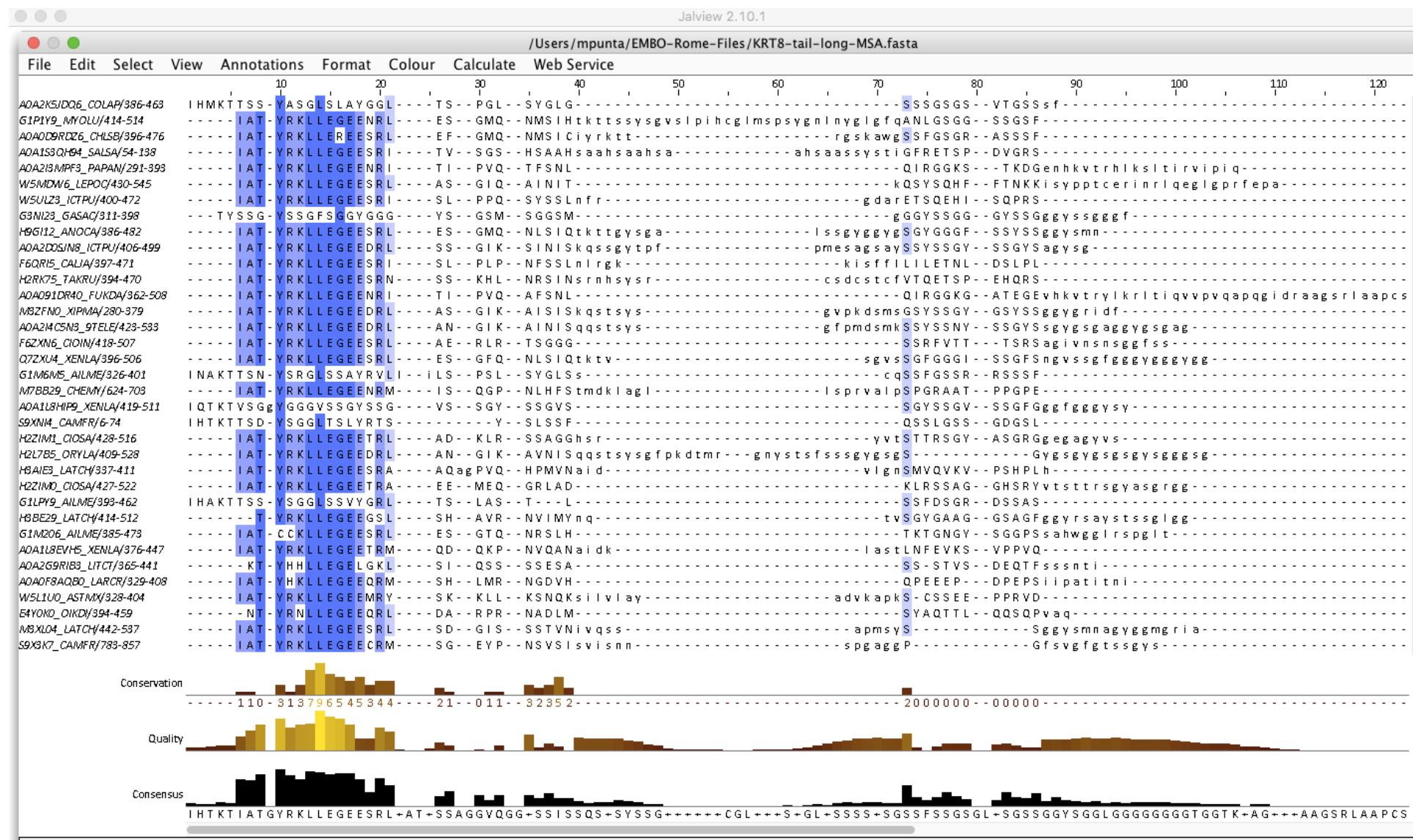
File	Edit	Select	View	Annotations	Format	Colour	Calculate	Web Service
	Undo Remove Redundancy...	⌘Z			20	30	40	50
	Redo	⌘Y			SLA-YG-----G--L----TS--P--GL---S-YGLG-----			60
	Cut	⌘X			EGE-EN----R--L----ES--G--MQ---N-MSIHkttssygvspihcgilmfspsyg			70
	Copy	⌘C			ERE-ES----R--L----EF-G--MQ---N-MSICiyrktt-----			
	Paste	▶			EGE-ES----R--I----TV-S--GS--H-SAAHsaahsaahsa-----ah			
	Delete				EGE-EN----R--I----TI--P--VQ---T-FSNL-----			
					EGE-ES----R--L----AS--G--IQ---A-INIT-----			
					EGE-ES----R--I----SL--P--PQ---S-YSSLnfr-----			
G3NI23_	Remove left	⌘L			SGG-YG----G--G----YS--G--SM---S-GGSM-----			
H9G112_	Remove right	⌘R			EGE-ES----R--L----ES--G--MQ---N-LSIQtkttgysga-----			
AOA2D05	Remove Empty Columns	⌘E			EGE-ED----R--L----SS--G--IK---S-INISKqssgytpf-----			
F6QR15_	Remove All Gaps	⇧⌘E			EGE-ES----R--I----SL--P--LP---N-FSSLnlrgk-----			
H2RK75_	Remove Redundancy...				EGE-ES----R--N----SS--K--HL---N-RSINsrnhsysr-----			
AOA0911_		⌘D			EGE-EN----R--I----TI--P--VQ---A-FSNL-----			
M8ZFNO_					EGE-ED----R--L----AS--G--IK---A-ISISKqstsys-----			
AOA2M4C_	Pad Gaps				EGE-ED----R--L----AN--G--IK---A-INISqqstsys-----			
F6ZXN6_CIOIN/418-507					-----IAT-YRKLLGE-E-----R--L----AE--R--LR---T-SGGG-----			
Q7ZXU4_XENLA/396-506					-----IAT-YRKLLGE-E-----R--L----ES--G--FQ---N-LSIQtktv-----			
G1M6M5_AILME/326-401					I NAKTTSN-YSRGLSSA-YR----V--LI--iLS--P--SL---S-YGLSS-----			
M7BB29_CHEMY/624-703					-----IAT-YRKLLGE-E-----R--M----IS--Q--GP---N-LHFStmdklagl-----			
AOA1L8HIP9_XENLA/419-511					I QT KTVSSG YGGGVSSG-YS----S--G----VS--S--GY---S-SGVS-----			
S9XN14_CAMFR/6-74					I HTKTTSD-YSGGLTSL-YR----T-S-----Y-----S-LSSF-----			
H22IM1_CIOSA/428-516					-----IAT-YRKLLGE-ET----R--L----AD--K--LR---S-SAGGhsr-----			
H2L7B5_ORYLA/409-528					-----IAT-YRKLLGE-ED----R--L----AN--G--IK---A-VNISqqstsysgfpkdtmr---gnysts			
H9AIE3_LATCH/337-411					-----IAT-YRKLLGE-E-----R--A----AQagP--VQ---H-PMVNaid-----			
H2ZIM0_CIOSA/427-522					-----IAT-YRKLLGE-ET----R--A----EE--M--EQ---G-RLAD-----			
G1LPY9_AILME/399-462					I HAKTTSS-YSGGLSSV-YG----R--L----TS--L--AS-----T-----L-----			
H9BE29_LATCH/414-512					-----T-YRKLLGE-EG----S--L----SH--A--VR---N-VIMYnq-----			
G1M206_AILME/385-473					-----IAT-CCKLLEGE-E-----R--L----ES--G--TQ---N-RSLH-----			
AOA1L8EVHS_XENLA/376-447					-----IAT-YRKLLGE-ET----R--M----QD--Q--KP---N-VQANaidk-----			
AOA2G9R1B3_LITCT/365-441					-----KT-YHHLLEGE-LG----K--L----SI--Q--SS----S-SESA-----			
AOA0F8AQBO_LARCR/329-408					-----IAT-YHKLLEGE-EQ----R--M----SH--L--MR---N-GDVH-----			
WSL1U0_ASTMX/328-404					-----IAT-YRKLLGE-EM----R--Y----SK--K--LL---K-SNQKsilvlay-----			
E4YOKO_OIKDI/394-459					-----NT-YRNLLEGE-EQ----R--L----DA--R--PR---N-ADLM-----			
M8XL04_LATCH/442-537					-----IAT-YRKLLGE-E-----R--L----SD--G--IS----S-STVNivqss-----			
S9X3K7_CAMFR/783-857					-----IAT-YRKLLGE-EC----R--M----SG--E--YP---N-SVSIsvisnn-----			

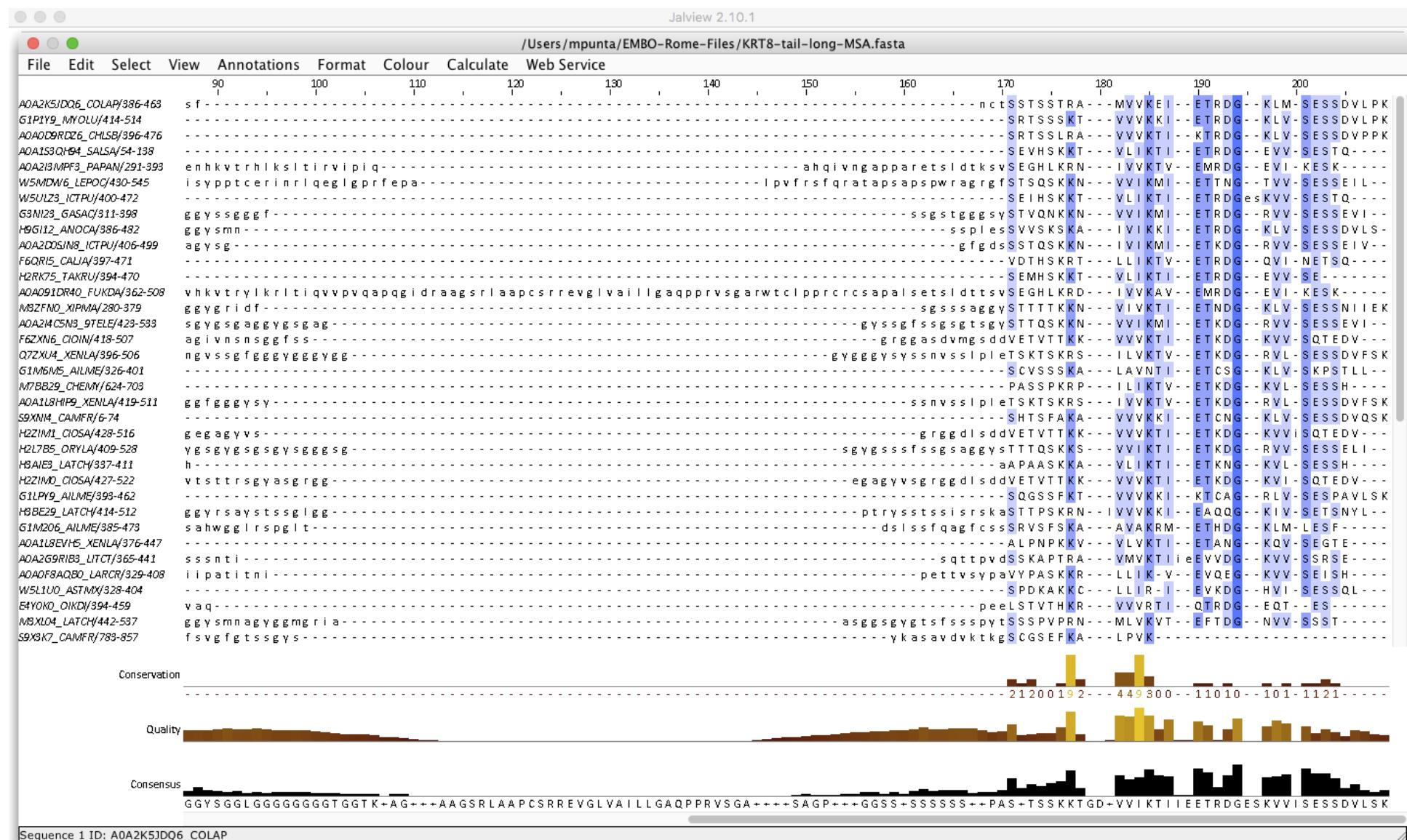


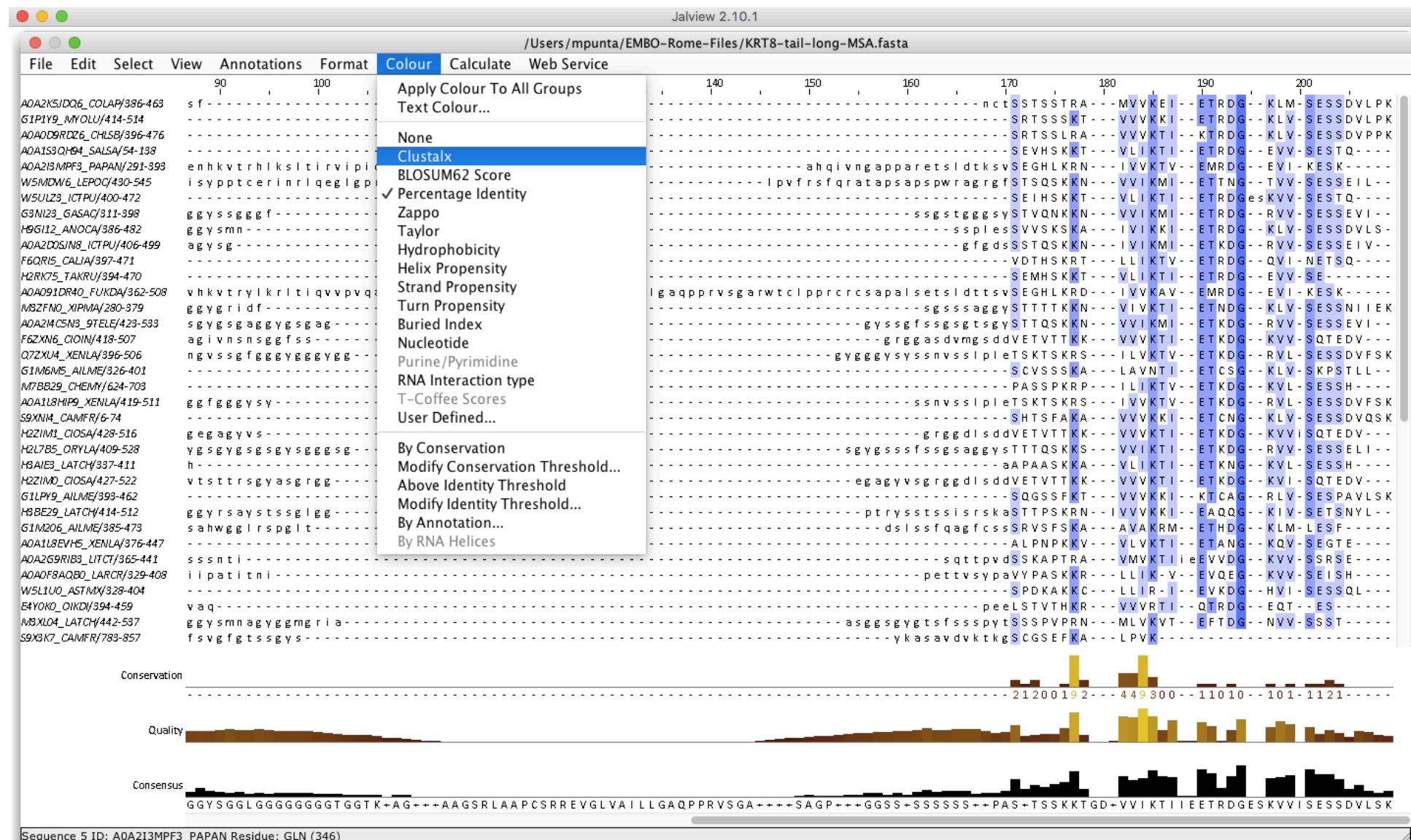
Sequence 9 ID: A0A2K5L5Z8_CERAT Residue: SER (387)

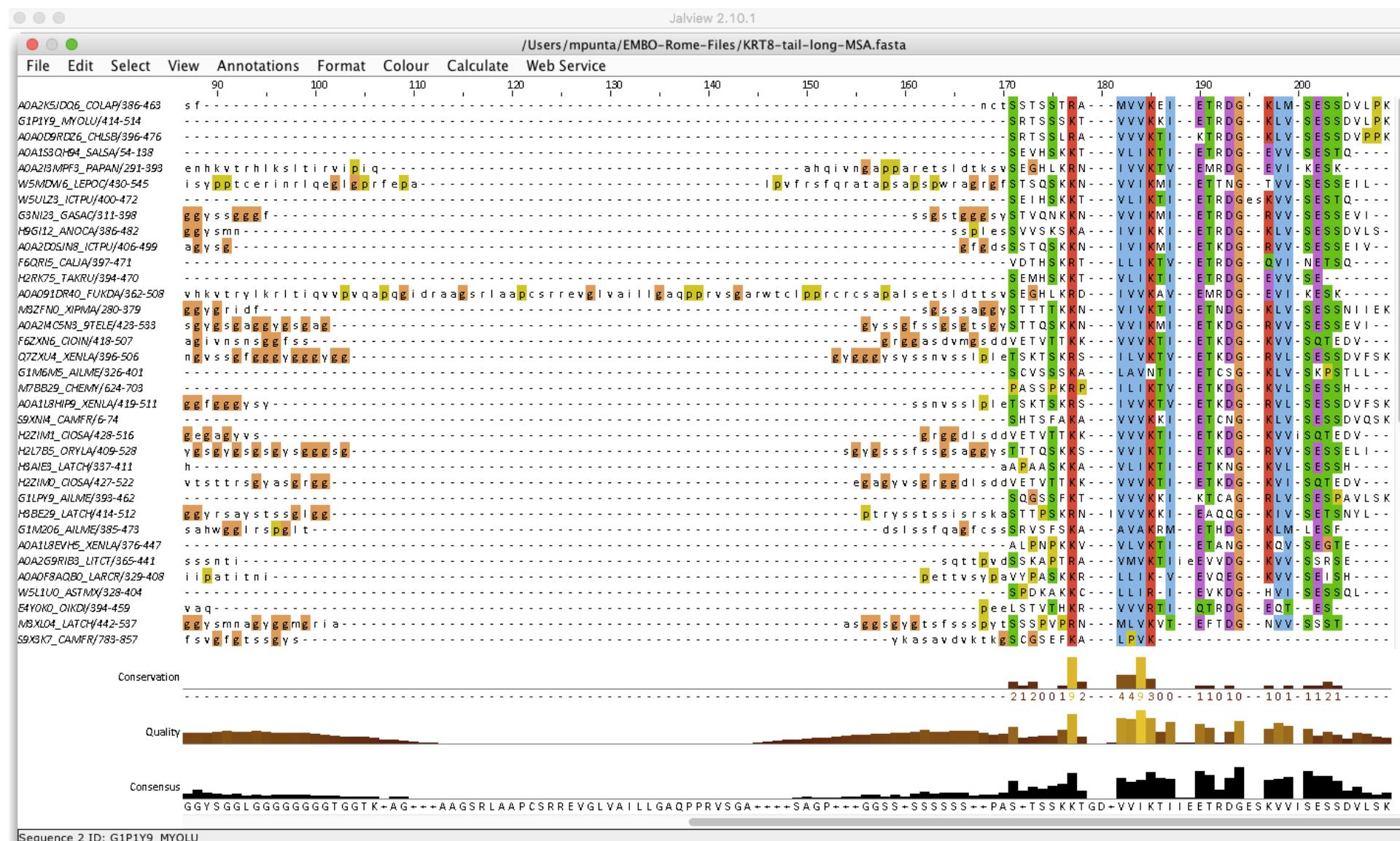


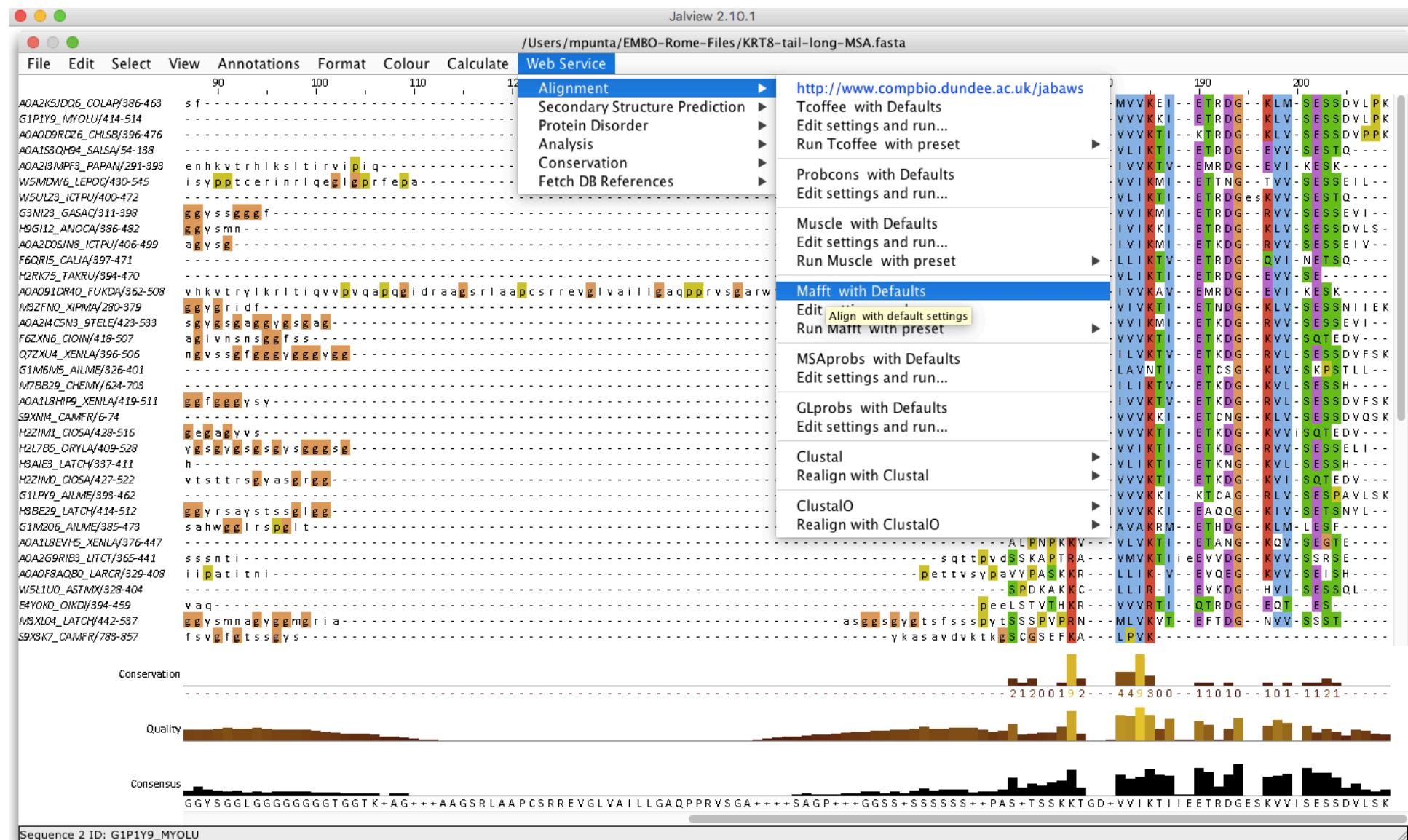




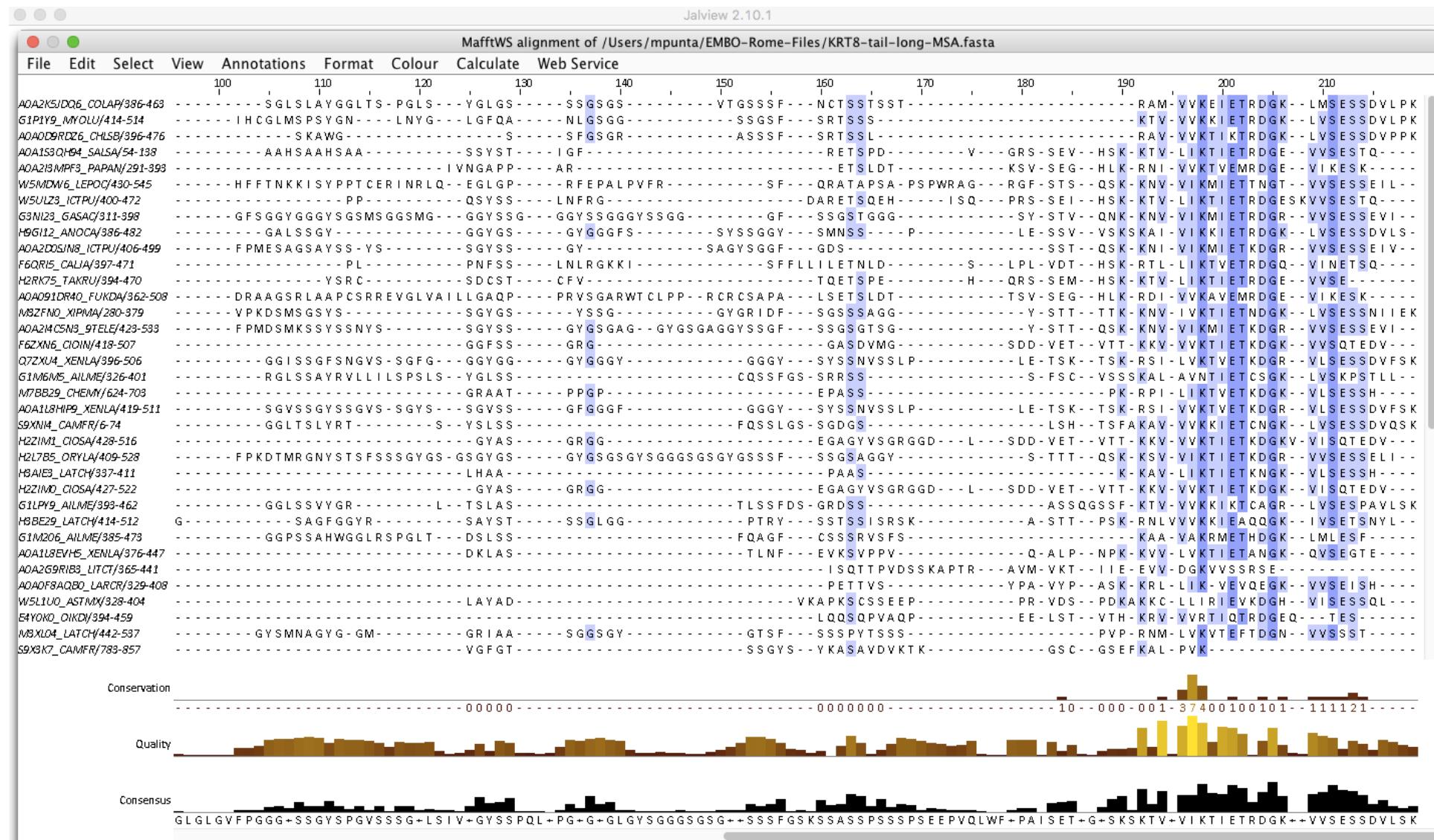


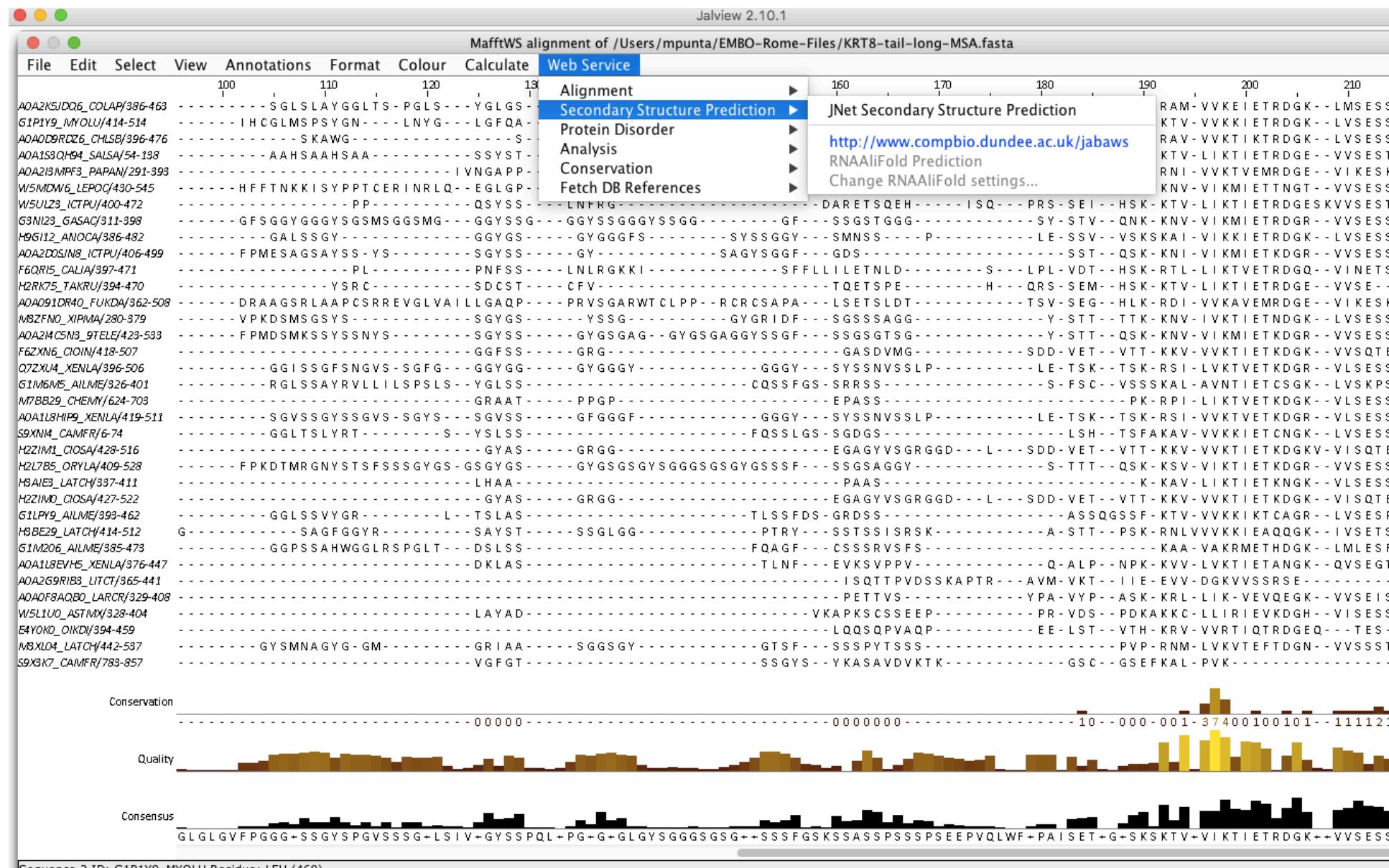


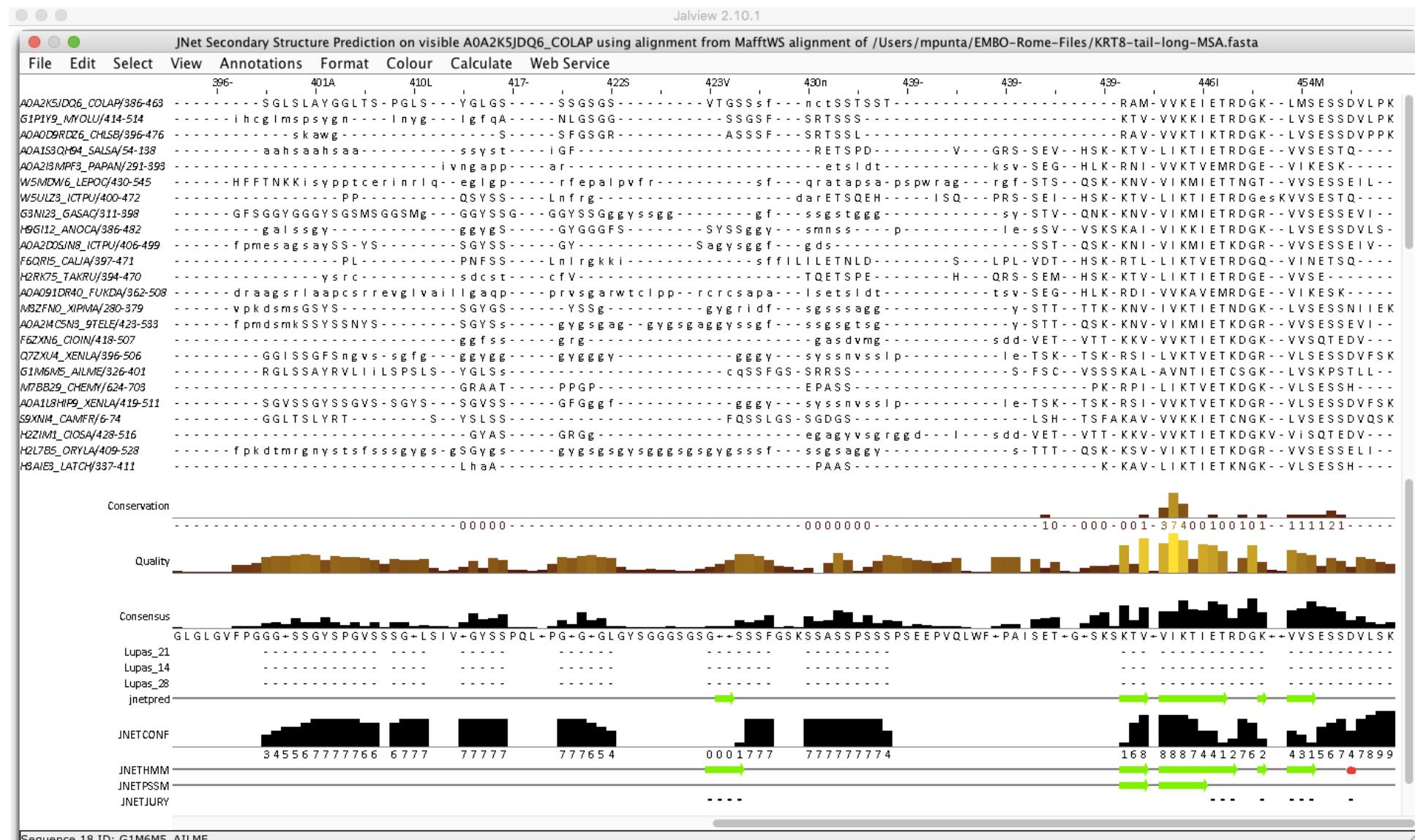


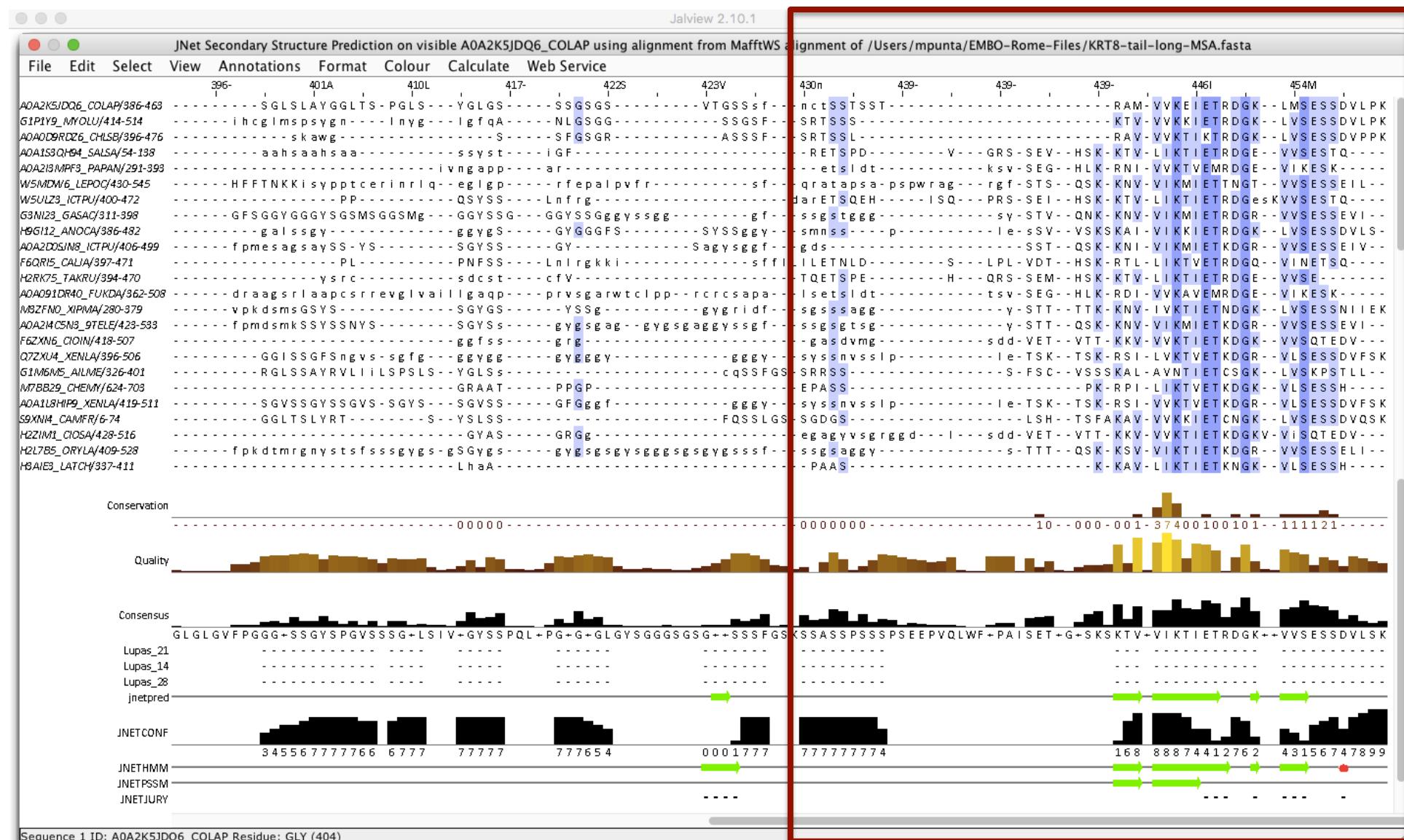












>sp|P05787|K2C8_HUMAN Keratin, type II cytoskeletal 8 OS=Homo sapiens OX=9606 GN=KRT8 PE=1 SV=7
MSIRVTQSYKVSTSGPRAFSSRSYTSAGPSRISSSSFSGVSSNFRGGLGGYGGASGM
GGITAVTVNQSLLSPLVLEVDPNIQAVRTQEKEQIKTLNNKFASFIDKVRFLQQNKMLE
TKWSLLQQQKTARSNMDNMFESYINNLRRQLETLGQEKLKLEAELGNMQGLVEDFKNKYE
DEINKRTEMENEFVLIKKDVDEAYMNKVELESRLEGLTDEINFLRQLYEEEIRELQSQIS
DTSVVLSMDNSRSLDMDSIIAEVKAQYEDIANRSRAEAESMYQIKYEELQSLAGKHGDDL
RRTKTEISEMNRNISRLQAEIEGLKGQRASLEAAIADAERQGELAIKDANAKLSELEAAL
QRAKQDMARQLREYQELMVNLALDIEIATYRKLLGEES

FSRTSSSSRAVVVKKIETRDGKLVSESSDV

LPK

KRT8-tail-short.fasta





HMMER

Biosequence analysis using profile hidden Markov Models

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

Next release within a week, think about downloading your results

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

iterative search vs protein sequence database

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

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

```
>KRT8_tail_short
FSRTSSSSRAVVVKKIETRDGKLVSESSDVLPK
```

▼ [Sequence Database](#) [?](#)

Let's take a 5 mins break





HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

JACKHMMER Results

[Search Again](#)

Jackhmmer Summary

Iteration	Results	Hits			
		New	Lost ⓘ	Dropped ⓘ	Total
1	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.1	+199	-	-	199
2	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.2	+72	-	-	271
3	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.3	+118	-1	1	387
4	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.4	+216	-	2	601
5	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.5	+116	-	1	716
6	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.6	+25	-	-	741
7	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.7	+14	-	1	754
8	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.8	+7	-	1	760
9	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.9	-	-	-	760

Your search has converged. No more iterations will be run.



HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

JACKHMMER Results

[Search Again](#)

Jackhmmer Summary

Iteration	Results	Hits			
		New	Lost	Dropped	Total
1	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.1	+199	-	-	199
2	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.2	+72	-	-	271
3	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.3	+118	-1	1	387
4	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.4	+216	-	2	601
5	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.5	+116	-	1	716
6	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.6	+25	-	-	741
7	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.7	+14	-	1	754
8	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.8	+7	-	1	760
9	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.9	-	-	-	760

Your search has converged. No more iterations will be run.



HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

JACKHMMER Results

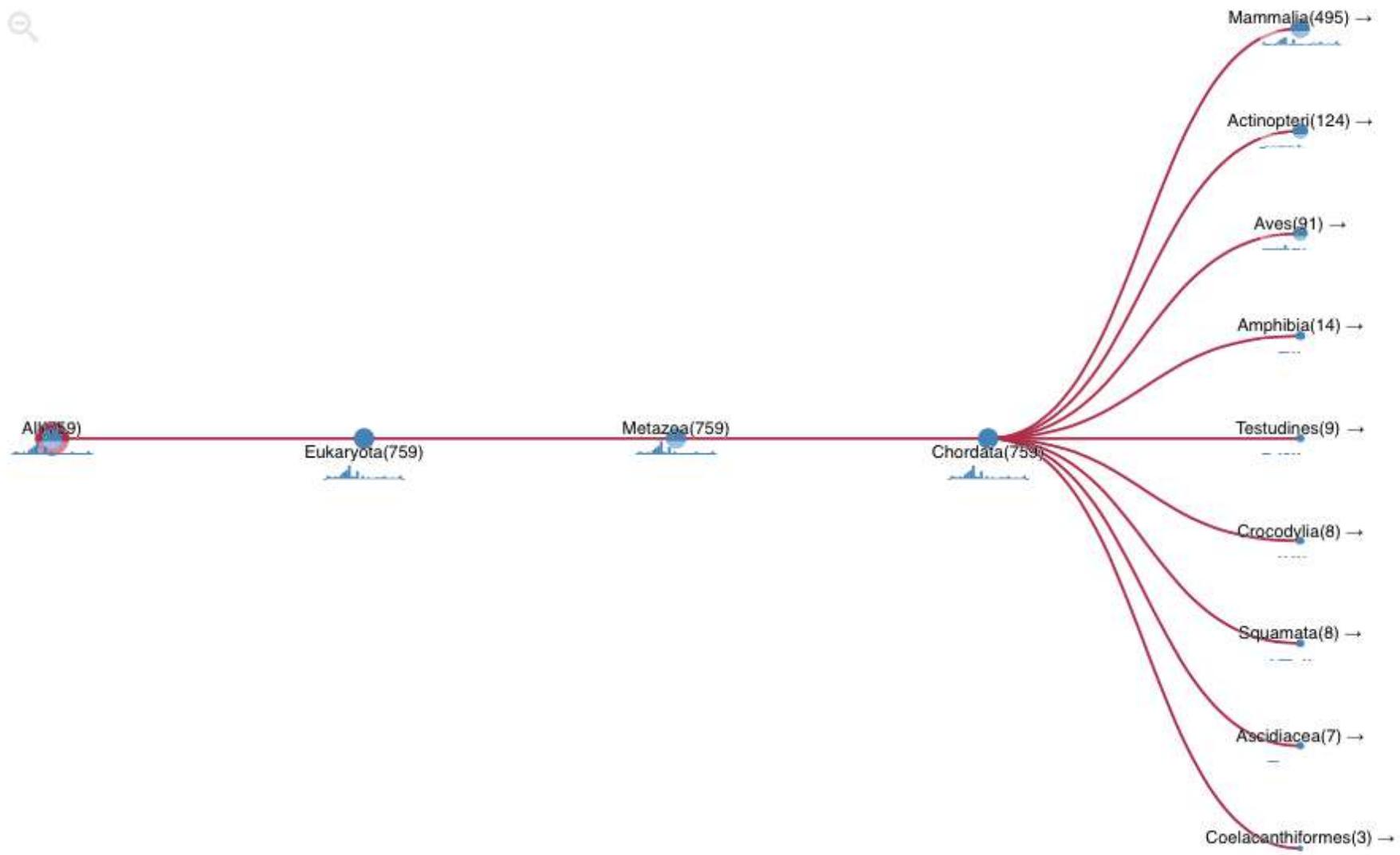
[Search Again](#)

Jackhmmer Summary		Hits			
Iteration	Results	New	Lost	Dropped	Total
1	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.1	+199	-	-	199
2	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.2	+72	-	-	271
3	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.3	+118	-1	1	387
4	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.4	+216	-	2	601
5	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.5	+116	-	1	716
6	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.6	+25	-	-	741
7	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.7	+14	-	1	754
8	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.8	+7	-	1	760
9	FE4CF6E2-DB9C-11E8-B3A9-A462E976C163.9	-	-	-	760

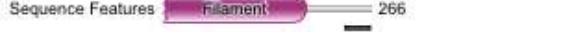
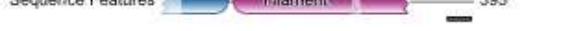
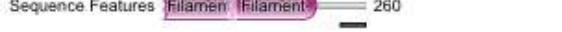
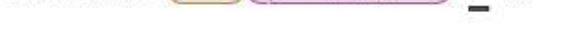
Your search has converged. No more iterations will be run.

Click

Taxonomic distribution of all search hits

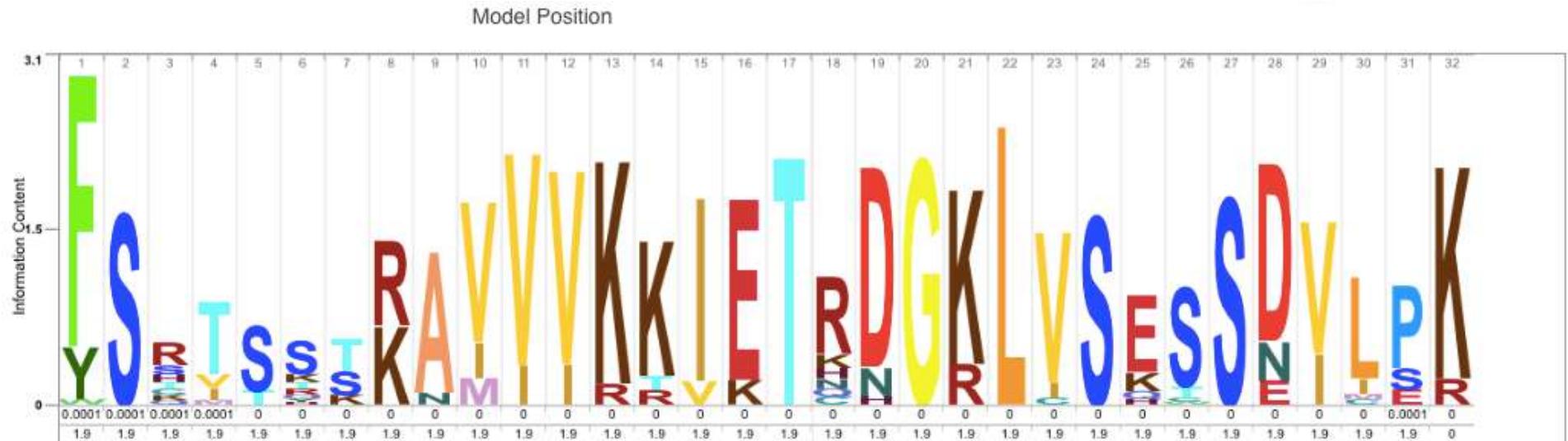


Domain Architectures

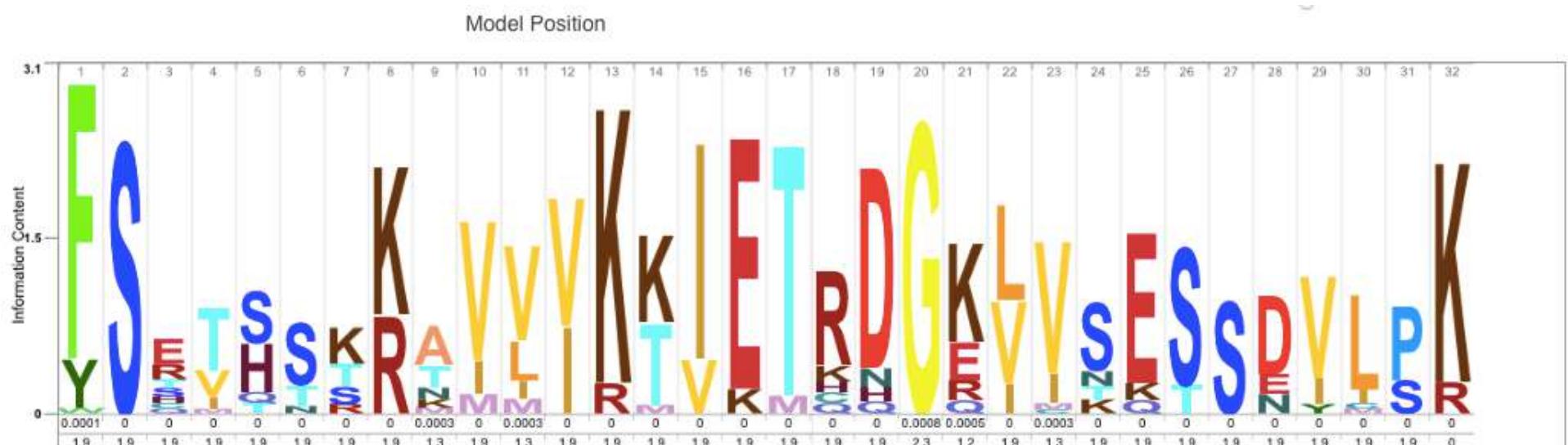
397 SEQUENCES	with domain architecture: Filament_head, Filament , example: A0A087YJM2_POEFO Sequence Features :  472	View Scores
Show All		
161 SEQUENCES	with domain architecture: Keratin_2_head, Filament , example: G3SEY3_GORGO Sequence Features :  483	View Scores
Show All		
127 SEQUENCES	with domain architecture: Filament , example: A0A096NMK7_PAPAN Sequence Features :  266	View Scores
Show All		
22 SEQUENCES	with domain architecture: Keratin_2_head, Filament, Filament , example: F7GPG9_MACMU Sequence Features :  395	View Scores
Show All		
14 SEQUENCES	with domain architecture: Filament, Filament , example: L9KZY6_TUPCH Sequence Features :  260	View Scores
Show All		
12 SEQUENCES	with no domain architecture, example: S9XNI4_CAMFR Sequence Features :  74	View Scores
Show All		
11 SEQUENCES	with domain architecture: Filament_head, Filament, Filament , example: A0A287AQY1_PIG Sequence Features :  424	View Scores
Show All		
6 SEQUENCES	with domain architecture: Pkinase, Keratin_2_head, Filament, Filament , example: G3REJ4_GORGO Sequence Features :  1649	View Scores
Show All		
2 SEQUENCES	with domain architecture: Keratin_2_head, Filament, Filament, Filament , example: A0A2I3RRX0_PANTR Sequence Features :  266	View Scores

Comparing hits profiles

After 1 Iteration



After 8 Iterations



score results | HMMER score results | HMMER Marco Punta - Outlook Web App score results | HMMER

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

estimated local date: 11/7/2018

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

Next release within a week, think about downloading your results

JACKHMMER Results

Score Taxonomy Domain Download

Iteration **9** Converged

Your jackhmmer search has reached convergence. No further iterations will be performed.

Return to the [Results Summary](#) | [Jump to threshold](#)

[« previous iteration](#)

Click search again and “with these results”

Perform a new search

- with new input
- with these results



Distribution of Significant Hits 



HMMER

Biosequence analysis using profile hidden Markov Models

Home

Search

Results

Software

Help

About

Contact

Next release within a week, think about downloading your results

phmmmer

hmmscan

hmmsearch

jackhmmer

iterative search vs protein sequence database

Paste a Sequence or an Alignment | Upload a File | Accession Search

Upload a file

KRT8-tail-short-MSA.fasta

Click

Submit

Reset

▼ Sequence Database



Frequently used databases:

Current database selection:



HMMER

Biosequence analysis using profile hidden Markov Models

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

Next release within a week, think about downloading your results

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

Iterative search vs protein sequence database

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

Upload a file

KRT8-tail-short-MSA.fasta

Click



▼ Sequence Database

Frequently used databases:

[Reference Proteomes](#)

[UniProtKB](#)

[SwissProt](#)

[PDB](#)

[Ensembl](#)

Current database selection:

PDB



HMMER

Biosequence analysis using profile hidden Markov Models

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

Next release within a week, think about downloading your results

JACKHMMER Results

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

Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)

No hits were found for your query.

[Search Details](#)[Model Position](#)

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

Next release within a week, think about downloading your results

JACKHMMER Results

Search Again

Score

Taxonomy

Domain

Download

Perform a new search
• [with new input](#)
• [with these results](#)

Iteration 1

[Return to the Results Summary](#) | [Jump to threshold](#)

No hits were found for your query.

Search Details

Model Position



HMMER

Biosequence analysis using profile hidden Markov Models

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

Next release within a week, think about downloading your results

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

iterative search vs protein sequence database

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

Upload a file

KRT8-tail-short-MSA.fasta

▼ Sequence Database 



Frequently used databases:

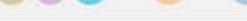
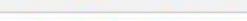
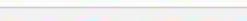
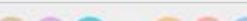
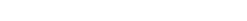
[Reference Proteomes](#)[UniProtKB](#)[SwissProt](#)[PDB](#)[Ensembl](#)

Current database selection:

SwissProt ▾

Distribution of Significant Hits



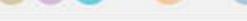
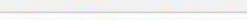
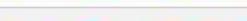
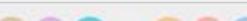
	Target	Description	Species	Cross-references	E-value	
>	K2C8_HUMAN 	Keratin, type II cytoskeletal 8	Homo sapiens 		1.4e-15	<input checked="" type="checkbox"/>
>	K2C8_RAT 	Keratin, type II cytoskeletal 8	Rattus norvegicus 		3.5e-13	<input checked="" type="checkbox"/>
>	K2C8_BOVIN 	Keratin, type II cytoskeletal 8	Bos taurus 		1.1e-12	<input checked="" type="checkbox"/>
>	K2C8_MOUSE 	Keratin, type II cytoskeletal 8	Mus musculus 		2.8e-12	<input checked="" type="checkbox"/>
>	VIM1_XENLA 	Vimentin-1/2	Xenopus laevis 		3.6e-12	<input checked="" type="checkbox"/>
>	DESM_XENLA 	Desmin	Xenopus laevis 		6.3e-11	<input checked="" type="checkbox"/>
>	DESM_MESAU 	Desmin	Mesocricetus auratus 		8.0e-11	<input checked="" type="checkbox"/>
>	DESM_HUMAN 	Desmin	Homo sapiens 		8.1e-11	<input checked="" type="checkbox"/>
>	DESM_RAT 	Desmin	Rattus norvegicus 		8.2e-11	<input checked="" type="checkbox"/>
>	DESM_BOVIN 	Desmin	Bos taurus 		8.2e-11	<input checked="" type="checkbox"/>
>	DESM_CHICK 	Desmin	Gallus gallus 		8.2e-11	<input checked="" type="checkbox"/>
	DESM_MOUSE 	Desmin	Mus musculus 		8.2e-11	<input checked="" type="checkbox"/>

>	DESM_CANLF	Desmin	Canis lupus familiaris		8.3e-11	<input checked="" type="checkbox"/>
>	VIM4_XENLA	Vimentin-4	Xenopus laevis		9.0e-10	<input checked="" type="checkbox"/>
>	VIME_PIG	Vimentin	Sus scrofa		1.7e-09	<input checked="" type="checkbox"/>
>	PERI_BOVIN	Peripherin	Bos taurus		1.7e-09	<input checked="" type="checkbox"/>
>	VIME_BOVIN	Vimentin	Bos taurus		1.7e-09	<input checked="" type="checkbox"/>
>	VIME_MOUSE	Vimentin	Mus musculus		1.7e-09	<input checked="" type="checkbox"/>
>	VIME_RAT	Vimentin	Rattus norvegicus		1.7e-09	<input checked="" type="checkbox"/>
>	VIME_PANTR	Vimentin	Pan troglodytes		1.8e-09	<input checked="" type="checkbox"/>
>	VIME_HUMAN	Vimentin	Homo sapiens		1.8e-09	<input checked="" type="checkbox"/>
>	VIME_MESAU	Vimentin	Mesocricetus auratus		1.8e-09	<input checked="" type="checkbox"/>
>	VIME_CHLAE	Vimentin	Chlorocebus aethiops		1.8e-09	<input checked="" type="checkbox"/>
>	VIME_MACFA	Vimentin	Macaca fascicularis		1.8e-09	<input checked="" type="checkbox"/>
>	VIME_CRIGR	Vimentin (Fragment)	Cricetulus griseus		1.8e-09	<input checked="" type="checkbox"/>
>	VIME_CHICK	Vimentin	Gallus gallus		1.8e-09	<input checked="" type="checkbox"/>

Distribution of Significant Hits



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

	Target	Description	Species	Cross-references	E-value	
>	K2C8_HUMAN 	Keratin, type II cytoskeletal 8	Homo sapiens 		1.4e-15	<input checked="" type="checkbox"/>
>	K2C8_RAT 	Keratin, type II cytoskeletal 8	Rattus norvegicus 		3.5e-13	<input checked="" type="checkbox"/>
>	K2C8_BOVIN 	Keratin, type II cytoskeletal 8	Bos taurus 		1.1e-12	<input checked="" type="checkbox"/>
>	K2C8_MOUSE 	Keratin, type II cytoskeletal 8	Mus musculus 		2.8e-12	<input checked="" type="checkbox"/>
>	VIM1_XENLA 	Vimentin-1/2	Xenopus laevis 		3.6e-12	<input checked="" type="checkbox"/>
>	DESM_XENLA 	Desmin	Xenopus laevis 		6.3e-11	<input checked="" type="checkbox"/>
>	DESM_MESAU 	Desmin	Mesocricetus auratus 		8.0e-11	<input checked="" type="checkbox"/>
>	DESM_HUMAN 	Desmin	Homo sapiens 		8.1e-11	<input checked="" type="checkbox"/>
>	DESM_RAT 	Desmin	Rattus norvegicus 		8.2e-11	<input checked="" type="checkbox"/>
>	DESM_BOVIN 	Desmin	Bos taurus 		8.2e-11	<input checked="" type="checkbox"/>
>	DESM_CHICK 	Desmin	Gallus gallus 		8.2e-11	<input checked="" type="checkbox"/>
	DESM_MOUSE 	Desmin	Mus musculus 		8.2e-11	

	Target	Description				Species			Cross-references			E-value	
>	K2C8_HUMAN	Keratin, type II cytoskeletal 8				Homo sapiens							1.4e-15
>	K2C8_RAT	Keratin, type II cytoskeletal 8				Rattus norvegicus							3.5e-13
>	K2C8_BOVIN	Keratin, type II cytoskeletal 8				Bos taurus							1.1e-12
>	K2C8_MOUSE	Keratin, type II cytoskeletal 8				Mus musculus							2.8e-12
>	VIM1_XENLA	Vimentin-1/2				Xenopus laevis							3.6e-12
>	DESM_XENLA	Desmin				Xenopus laevis							6.3e-11
>	DESM_MESAU	Desmin				Mesocricetus auratus							8.0e-11
v	DESM_HUMAN	Desmin				Homo sapiens							8.1e-11

Let's visit the UniProt website:

<https://www.uniprot.org>

<https://www.uniprot.org>

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB
UniProt Knowledgebase
Swiss-Prot (558,590)
Manually annotated and reviewed.
TrEMBL (126,780,198)
Automatically annotated and not reviewed.

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

Supporting data

Literature citations

Cross-ref. databases

Taxonomy

Diseases

Subcellular locations

Keywords

News

Forthcoming changes
Planned changes for UniProt

UniProt release 2018_09
Tubulin code: a long sought-after player identified

UniProt release 2018_08
Human brain development: slow and steady wins the race | 'Enzyme regulation' becomes 'Activity regulation' | New advanced search interfac...

News archive

Getting started

UniProt data

Protein spotlight

YouTube

UniProt

UniProtKB ▾ Advanced ▾ Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

UniProtKB - P17661 (DESM_HUMAN)

Basket ▾

Display

BLAST Align Format Add to basket History

Feedback Help video Other tutorials and videos

Entry

Protein | **Desmin**

Gene | **DES**

Organism | *Homo sapiens (Human)*

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

Function

None

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

Muscle-specific type III intermediate filament essential for proper muscular structure and function. Plays a crucial role in maintaining the structure of sarcomeres, inter-connecting the Z-disks and forming the myofibrils, linking them not only to the sarcolemmal cytoskeleton, but also to the nucleus and mitochondria, thus providing strength for the muscle fiber during activity (PubMed:[25358400](#)). In adult striated muscle they form a fibrous network connecting myofibrils to each other and to the plasma membrane from the periphery of the Z-line structures (PubMed:[24200904](#), PubMed:[25394388](#), PubMed:[26724190](#)). May act as a sarcomeric microtubule-anchoring protein: specifically associates with detyrosinated tubulin-alpha chains, leading to buckled microtubules and mechanical resistance to contraction. Contributes to the transcriptional regulation of the NKX2-5 gene in cardiac progenitor cells during a short period of cardiomyogenesis and in cardiac side population stem cells in the adult. Plays a role in maintaining an optimal conformation of nebulette (NEB) on heart muscle sarcomeres to bind and recruit cardiac alpha-actin (By similarity). 1 Publication ▾ By similarity ▾ 3 Publications ▾

GO - Molecular function

Let's take a 5 mins break



	Target	Description				Species			Cross-references				E-value		
>	K2C8_HUMAN	Keratin, type II cytoskeletal 8				Homo sapiens								1.4e-15	<input checked="" type="checkbox"/>
>	K2C8_RAT	Keratin, type II cytoskeletal 8				Rattus norvegicus								3.5e-13	<input checked="" type="checkbox"/>
>	K2C8_BOVIN	Keratin, type II cytoskeletal 8				Bos taurus								1.1e-12	<input checked="" type="checkbox"/>
>	K2C8_MOUSE	Keratin, type II cytoskeletal 8				Mus musculus								2.8e-12	<input checked="" type="checkbox"/>
>	VIM1_XENLA	Vimentin-1/2				Xenopus laevis								3.6e-12	<input checked="" type="checkbox"/>
>	DESM_XENLA	Desmin				Xenopus laevis								6.3e-11	<input checked="" type="checkbox"/>
>	DESM_MESAU	Desmin				Mesocricetus auratus								8.0e-11	<input checked="" type="checkbox"/>
v	DESM_HUMAN	Desmin				Homo sapiens								8.1e-11	<input checked="" type="checkbox"/>

```

.....*.....*.....
Query      1 SetsskKaVVVKkIETRDGK1VsES 25
          Se+++kK+V++K+IETRDG++VsE
Target     438 SEVHTKKTVMKIETRDGEVVSEA 462
PP          9*****5

```

Display

Family & Domainsⁱ

Get all the reaction from England's Quilter Intern...

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcell. location

Pathol./Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar proteins

Cross-references

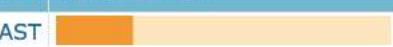
Entry information

Miscellaneous

Domains and Repeats

Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain ⁱ	108 – 416	IF rod PROSITE-ProRule annotation	Add BLAST		309

Region

Feature key	Position(s)	Description	Actions	Graphical view	Length
Region ⁱ	1 – 108	Head	Add BLAST		108
Region ⁱ	109 – 141	Coil 1A	Add BLAST		33
Region ⁱ	142 – 151	Linker 1			10
Region ⁱ	152 – 252	Coil 1B	Add BLAST		101
Region ⁱ	253 – 268	Linker 12	Add BLAST		16
Region ⁱ	268 – 415	Interaction with NEB 1 Publication	Add BLAST		148
Region ⁱ	269 – 287	Coil 2A	Add BLAST		19
Region ⁱ	288 – 295	Linker 2			8
Region ⁱ	296 – 412	Coil 2B	Add BLAST		117
Region ⁱ	413 – 470	Tail	Add BLAST		58
Region ⁱ	438 – 453	Interaction with CRYAB 1 Publication	Add BLAST		16

Sequence similaritiesⁱ

Belongs to the intermediate filament family. PROSITE-ProRule annotation

Keywords - Domainⁱ

Coiled coil

Phylogenomic databases

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcell. location

Pathol./Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar proteins

Cross-references

Entry information

Miscellaneous

Family & Domainsⁱ

Domains and Repeats

Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain ⁱ	108 – 416	IF rod PROSITE-ProRule annotation	Add BLAST		309

Region

Feature key	Position(s)	Description	Actions	Graphical view	Length
Region ⁱ	1 – 108	Head	Add BLAST		108
Region ⁱ	109 – 141	Coil 1A	Add BLAST		33
Region ⁱ	142 – 151	Linker 1			10
Region ⁱ	152 – 252	Coil 1B	Add BLAST		101
Region ⁱ	253 – 268	Linker 12	Add BLAST		16
Region ⁱ	268 – 415	Interaction with NEB 1 Publication	Add BLAST		148
Region ⁱ	269 – 287	Coil 2A	Add BLAST		19
Region ⁱ	288 – 295	Linker 2			8
Region ⁱ	296 – 412	Coil 2B	Add BLAST		117
Region ⁱ	413 – 470	Tail	Add BLAST		58
Region ⁱ			Add BLAST		16

Manual assertion based on experiment inⁱ

"alphaB-crystallin is a sensor for assembly intermediates and for the subunit topology of desmin intermediate filaments."

Sharma S., Conover G.M., Elliott J.L., Der Perng M., Herrmann H., Quinlan R.A.
Cell Stress Chaperones 22:613-626(2017) [PubMed] [Europe PMC] [Abstract]

Cited for: CHARACTERIZATION OF VARIANTS MFM1 MET-451 AND TDP-43A INTERACTION WITH CRYAB

Phylogenomic databases

αB-crystallin is a sensor for assembly intermediates and for the subunit topology of desmin intermediate filaments.

Sharma S¹, Conover GM², Elliott JL³, Der Perng M⁴, Herrmann H^{1,5}, Quinlan RA⁶.

Author information

Abstract

Mutations in the small heat shock protein chaperone CRYAB (αB-crystallin/HSPB5) and the intermediate filament protein desmin, phenocopy each other causing cardiomyopathies. Whilst the binding sites for desmin on CRYAB have been determined, desmin epitopes responsible for CRYAB binding and also the parameters that determine CRYAB binding to desmin filaments are unknown. Using a combination of co-sedimentation centrifugation, viscometric assays and electron microscopy of negatively stained filaments to analyse the *in vitro* assembly of desmin filaments, we show that the binding of CRYAB to desmin is subject to its assembly status, to the subunit organization within filaments formed and to the integrity of the C-terminal tail domain of desmin. Our *in vitro* studies using a rapid assembly protocol, C-terminally truncated desmin and two disease-causing mutants (I451M and R454W) suggest that CRYAB is a sensor for the surface topology of the desmin filament. Our data also suggest that CRYAB performs an assembly chaperone role because the assembling filaments have different CRYAB-binding properties during the maturation process. We suggest that the capability of CRYAB to distinguish between filaments with different surface topologies due either to mutation (R454W) or assembly protocol is important to understanding the pathomechanism(s) of desmin-CRYAB myopathies.

Alignments

[Download](#) [Graphics](#)

sp|P17661|DESM_HUMAN Desmin OS=Homo sapiens OX=9606 GN=DES PE=1 SV=3
Sequence ID: Query_170373 Length: 470 Number of Matches: 1

“Our results indicate that CRYAB binding most likely occurs between the amino acid residues 442–453 of desmin.”

Sbjct 72 SR+G++ YG + A VNQ L+ RT EK +++ L
SRLGTTR----TPSSYGAGELLDFSLADAVNQEFLT-----TRTNEKVELQEL 115

ii. Sensor Function during *in vitro* assembly

A. Coassembly of CRYAB with desmin – modulation of binding by mutants



RRQLETLGQEK 158
RRQ+E L ++ 159
RRQVEVLTNQR 173
VELESRLEGLT 218
++LE R+E L 219
IDLERRIESLN 233
EDIANRSRAEA 278
E IA ++ +EA 279
ETIAAKNISEA 292
RASLEAAIADA 338
SL + + 339
Sbjct 293 EEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDLRLQMREL 352

Query 339 EQRGELAIKDANAKLSELEAAALQRAKQDMARQLREYQELMNVKLALDIEIATYRKLLGE 398
E R ++ LE ++ K +MAR LREYQ+L+NVK+ALD+EIATYRKLLGE

Sbjct 353 EDRFASEASGYQDNIAIRLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLGE 412

Query 399 ESRLESGMQNMSIHTKTTSGYAGGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFRTSSS 458
ESR+ N+ I T Y L +F + S +

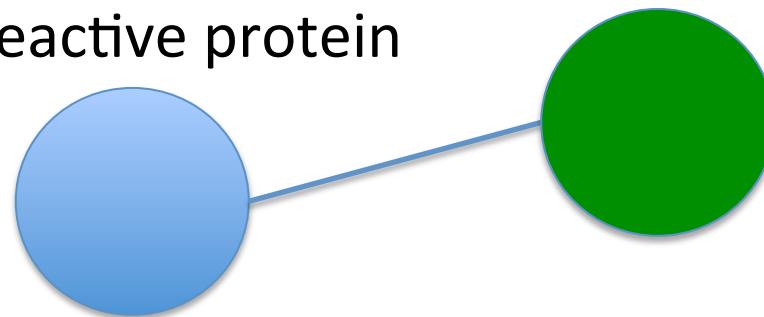
Sbjct 413 ESRI-----NLPIQT-----YSAL-----NFRETSPEQRGSEVHTK 443

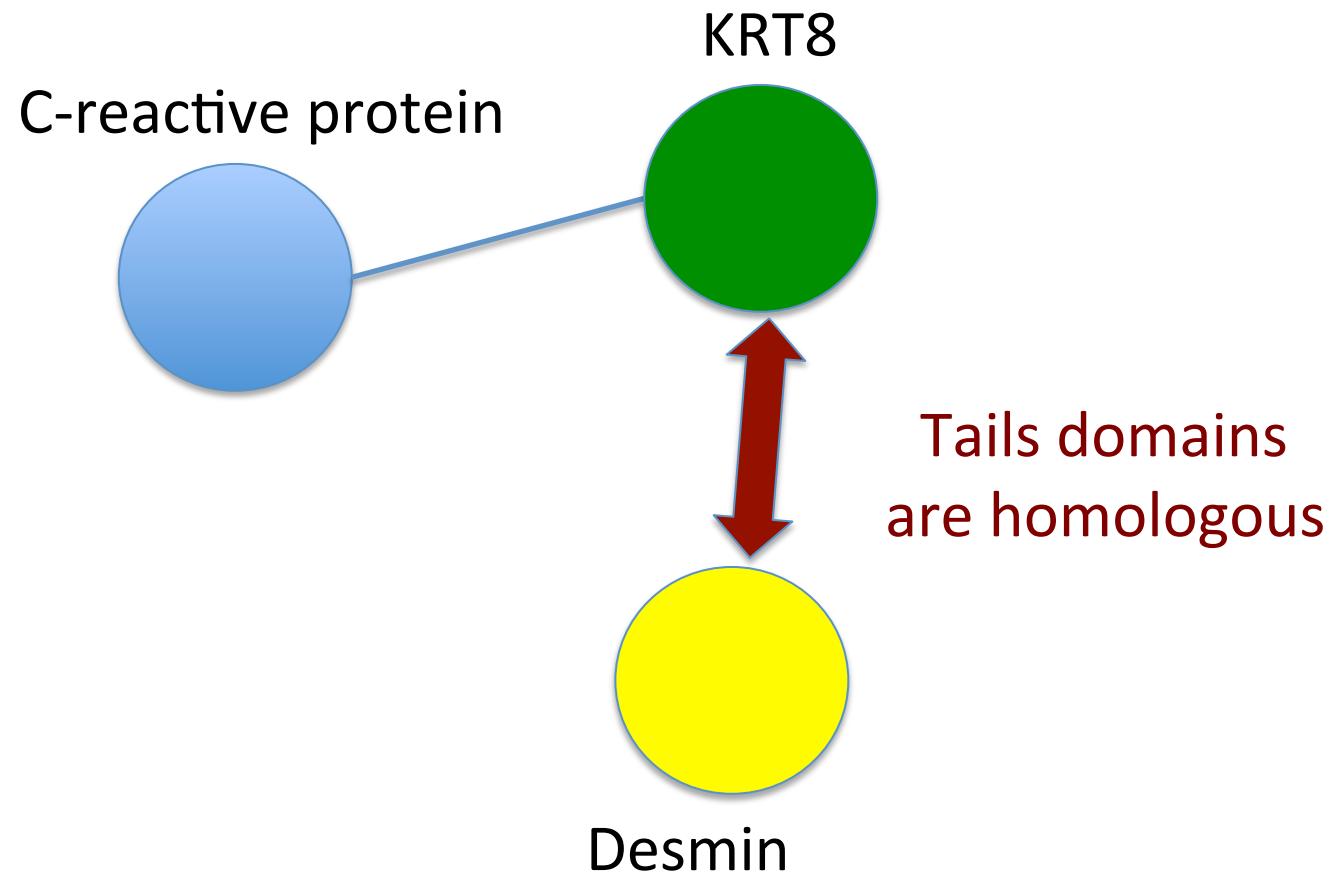
Query 459 RAVVVKKIETRDGKLVSESS 478

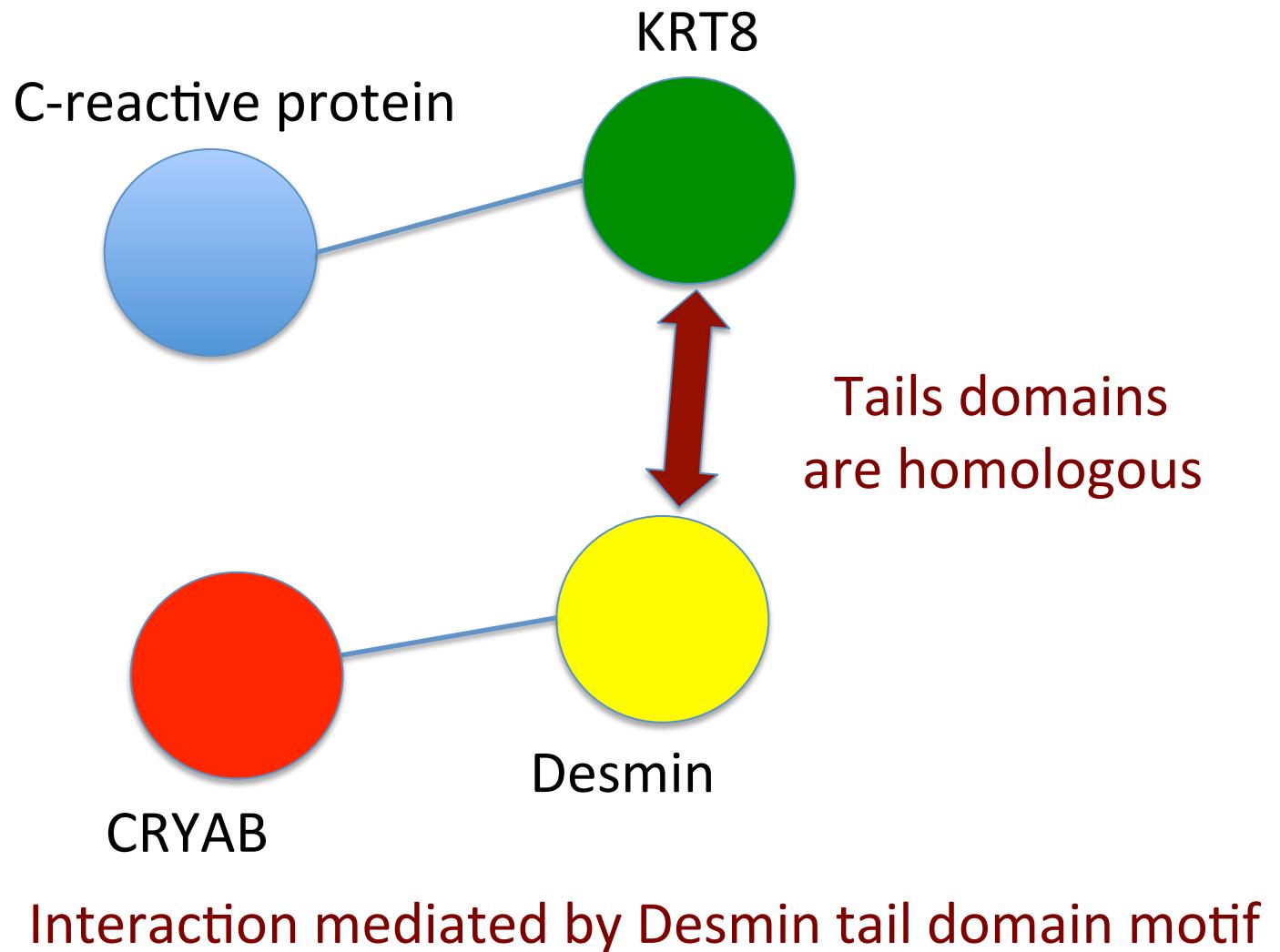
+ V++K IETRDG++VSE++

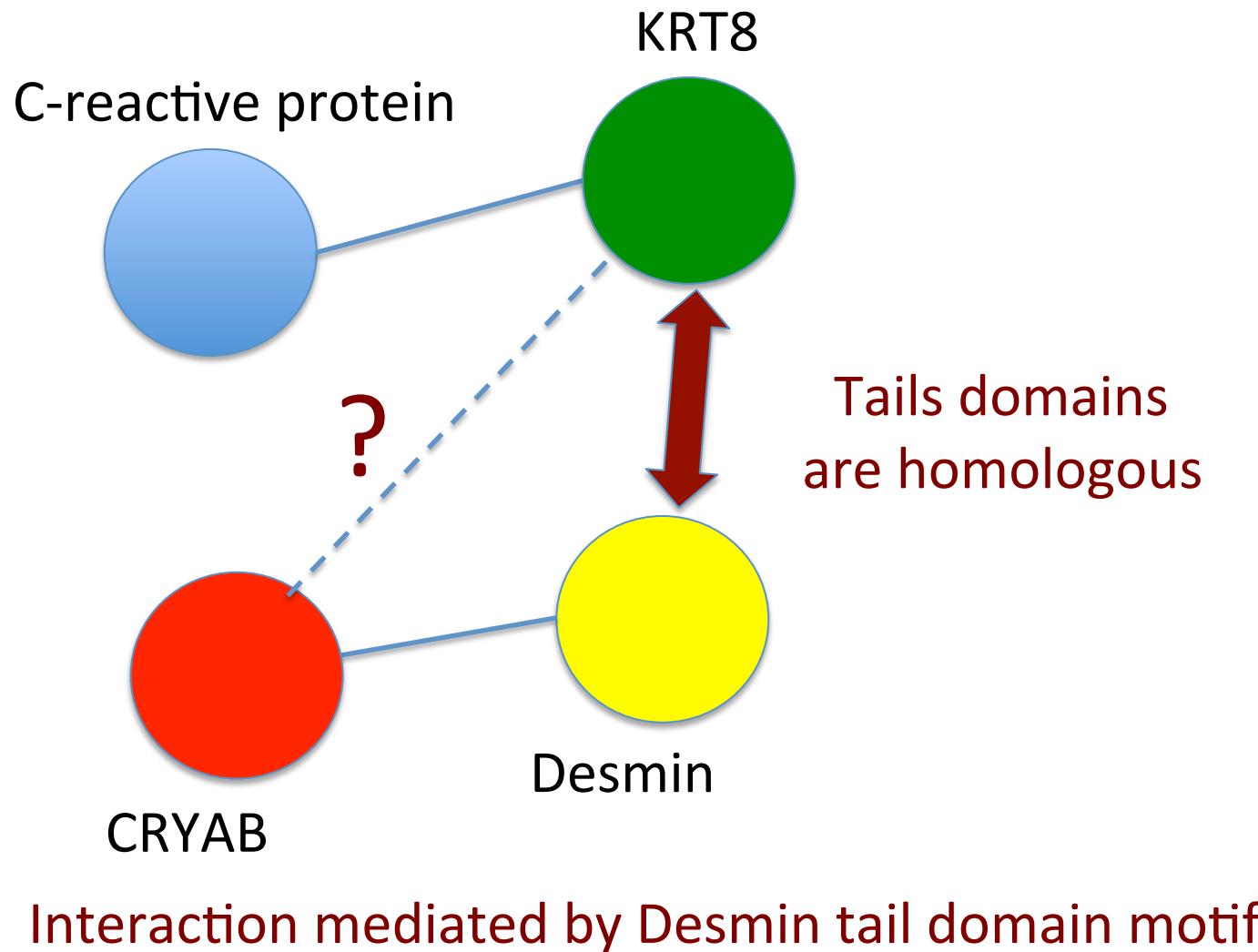
Sbjct 444 KTVMIKTIETRDGEVVSEAT 463

C-reactive protein KRT8









αB-crystallin is a sensor for assembly intermediates and for the subunit topology of desmin intermediate filaments.

Sharma S¹, Conover GM², Elliott JL³, Der Perng M⁴, Herrmann H^{1,5}, Quinlan RA⁶.

Author information

Abstract

Mutations in the small heat shock protein chaperone CRYAB (αB-crystallin/HSPB5) and the intermediate filament protein desmin, phenocopy each other causing cardiomyopathies. Whilst the bind CRYAB unknown electron filament subunit desmin. two dise surface assemb properti disting

responsible for filaments are says and of desmin status, to the tail domain of desmin and for the forms an B-binding B to ion (R454W)

For now it's just speculation: we need to understand more about the proteins and the interaction motif to see if it's a reasonable hypothesis that merits experimental validation

or assembly protocol is important to understanding the pathomechanism(s) of desmin-CRYAB myopathies.

Systemic augmentation of α B-crystallin provides therapeutic benefit twelve hours post-stroke onset via immune modulation

Ahmet Arac^{a,b,c,1}, Sara E. Brownell^{b,d,e,1}, Jonathan B. Rothbard^f, Charlene Chen^{b,c,e}, Rose M. Ko^f, Marta P. Pereira^{a,b,c}, Gregory W. Albers^{b,c,e}, Lawrence Steinman^{b,e,2,3}, and Gary K. Steinberg^{a,b,c,2,3}

^aDepartment of Neurosurgery, ^bStanford Institute for Neuro-Innovation and Translational Neurosciences, ^cStanford Stroke Center, ^dDepartments of Biology, ^eNeurology and Neurological Sciences, and ^fMedicine, Stanford University School of Medicine, Stanford, CA 94305

Edited by Marc Feldmann, Imperial College London, London, United Kingdom, and approved June 27, 2011 (received for review May 10, 2011)

Tissue plasminogen activator is the only treatment option for stroke victims; however, it has to be administered within 4.5 h after symptom onset, making its use very limited. This report describes a unique target for effective treatment of stroke, even 12 h after onset, by the administration of α B-crystallin (Cryab), an endogenous immunomodulatory neuroprotectant. In *Cryab*^{-/-} mice, there was increased lesion size and diminished neurologic function after stroke compared with wild-type mice. Increased plasma Cryab was detected after experimental stroke in mice and after stroke in human patients. Administration of Cryab even 12 h after experimental stroke reduced both stroke volume and inflammatory cytokines associated with stroke pathology. Cryab is

(TTC) staining, which stains the viable tissue in red due to the activity of the mitochondrial dehydrogenases (Fig. 1A). This difference remained at 7 d after stroke as assessed by silver stain (used to detect neuronal and axonal degeneration; Fig. 1B and *SI Appendix*, Fig. S1), indicating that the deficiency of Cryab affected both the early and delayed phases of ischemic damage. Functional outcome was assessed by a 28-point neurobehavioral scoring test (14, 15). The *Cryab*^{-/-} mice had significantly worse scores at both 2-d and 7-d time points compared with wild-type controls (Fig. 1C). No differences were seen in cerebral blood flow measured by laser Doppler flow meter immediately after the occlusion and at 15 and 30 min of reperfusion between the groups (*SI Appendix*, Fig. S2).

small heat shock protein after stroke. There are several potential biomarkers in plasma (e.g., C-reactive protein, S100B, and matrix metalloproteinase-9) that can be used for diagnosis or outcome prediction of stroke (30). Given that Cryab is secreted from human retinal pigment epithelial cells (31), whether it can be used as a biomarker after stroke requires more extensive studies. Variations with age of the patient must be considered in such studies.

THANK YOU!!