

Факультет биоинженерии и биоинформатики

Межфакультетский курс

Лабораторная работа N°2.

Выравнивание биологических последовательностей

Студент: Шутова Ю.М.

1. UnitPro

Ищем PEPSIN, упоминающийся в **protein names**. Результаты по поиску: **2 530**. С пометкой "Reviewed (Swiss-Prot)" нашлось **32** подходящих запроса (аннотированные белки).

Аннотированные белки — это белки, которые прошли процесс модификации, изменяющий их свойства или функции.

Белки с не соответствующей функцией: **major pepsin inhibitor 3** (защитный белок растений), **Pepsin inhibitor Dit33** (функция связана с модуляцией иммунного ответа хозяина и обеспечением выживания паразита).

2. Выравнивание двух ортологов

Выбрала два белка: **Pepsin A-4** (у человека) и **Pepsin A-4** (у макаки). Отмечаем в чекбокса, затем **Tools -> Align**. Нажимаем "Run align". Переходим по гиперссылке "Completed".

Выравнивание показало схожую последовательность, половина несоответствий штрафуются по низкой стоимости, так как являются схожими по структуре. Последовательности гомологичны.

CLUSTAL O(1.2.4) multiple sequence alignment

```
sp|P0DJD7|PEPA4_HUMAN      MKWLLLLGLVALSECIMYKVPLIRKKSLRRTLSEERGLLKDFLKHHNLPARKYFPQWEAP      60
sp|P27678|PEPA4_MACFU      MKWLLLLGLVALSECIIYKVPLVRKKSLRRNLSEHGLLKDFLKHHNLPASKYFPQAEAP      60
*****:*****:*****:***:***** ***** ***

sp|P0DJD7|PEPA4_HUMAN      TLVDEQPLENYLDMEYFGTIGIGTPAQDFTVVFDTGSSNLWVPSVYCSSLACTNHNRFNP      120
sp|P27678|PEPA4_MACFU      TLIDEQPLENYLDVEYFGTIGIGTPAQNFTVVFDTGSSNLWVPSVYCYSLACMDHNLFP      120
**:*:*****:*****:*****:***** ***** :** ***

sp|P0DJD7|PEPA4_HUMAN      EDSSTYQSTSETVSITYGTGSMTGILGYDTVQVGGISDTNQIFGLSETEPGSFLYYAPFD      180
sp|P27678|PEPA4_MACFU      QDSSTYRATSKTVSITYGTGSMTGILGYDTVKVGGISDTNQIFGLSETEPGFFLYFAPFD      180
:*****::**:*:*****:*****:*****:***** *****:****

sp|P0DJD7|PEPA4_HUMAN      GILGLAYPSISSSGATPVFDNIWNQGLVSQDLFSVYLSADDQSGSVVIFGGIDSSYYTGS      240
sp|P27678|PEPA4_MACFU      GILGLAYPSISSSGATPVFDNIWNQRLVSQDLFSVYLSADDQSGSVVIFGGIDSSYYTGS      240
***** *****

sp|P0DJD7|PEPA4_HUMAN      LNWVPVTVEGYWQITVDSITMNGEAIACAEGCQAIVDTGTSLLTGPTSPIANIQSDIGAS      300
sp|P27678|PEPA4_MACFU      LNWVPVSVEGYWQISVDSITMNGKTIACAKGCQAIVDTGTSLLTGPTSPIANIQSDIGAS      300
*****:*****:*****:.*:***:*****:*****
```

sp P0DJD7 PEPA4_HUMAN	ENSDGDMVVSCSAISSLPDIVFTINGVQYPVPPSAYILQSEGSCISGFQGMNLPTESGEL	360
sp P27678 PEPA4_MACFU	ENSDGEMVVSCSAISSLPDIVFTINGVQYPLPPSAYILQSQGSCSTSGFQGMVPTESGEL	360
	*****:*****:*****:*** *****:*****	
sp P0DJD7 PEPA4_HUMAN	WILGDVFIRQYFTVFD RANNQVGLAPVA	388
sp P27678 PEPA4_MACFU	WILGDVFIRQYFTVFD RANNQVGLAPVA	388

3. Вывравнивание нескольких паралогов

Задание 1. Пепсин (человек, макака, мышь)

CLUSTAL O(1.2.4) multiple sequence alignment

sp Q9D106 PEPA5_MOUSE	MKWLWVLGLValseclv-KIPLMKIKSMRENlRESQVLKDYLEKYPRsRAHVlLEQRRNP	59
sp P0DJD7 PEPA4_HUMAN	MKWLLLLGLValsecIMYKVPLIRKKSLRRtLSErGLLKDFLKKHNLNPARKYFPQWEAP	60
sp P27678 PEPA4_MACFU	MKWLLLLGLValsecIiYKVPLVRKKSLRRNLSEHGLLKDFLKKHNLNPASKYFPQAEAP	60
	**** :*****:: *:**:: **:*. * :***:***: . * : * . *	
sp Q9D106 PEPA5_MOUSE	-AVTYEPMRNYLDlVYIGIISIGTPPQEFrVVLDTGSSVLWVPSIYCSSPACAHHKAFNP	118
sp P0DJD7 PEPA4_HUMAN	TLVDEQPLENYLDMEYFGTIGIGTPAQDFtVVFDtGSSNLWVPSVYCSSLACTNHNRFNP	120
sp P27678 PEPA4_MACFU	TLIDEQPLENYLDVEYFGTIGIGTPAQNFtVVFDtGSSNLWVPSVYCYSLACMDHNLfNP	120
	: :*:*****: *: * *.***** *: * **:***** *****:*** * ** .*: ***	
sp Q9D106 PEPA5_MOUSE	LRsSTfLVsGRPVNVAYSGSGMSGfLAYDTVRIGDLTVVAQAFGLSLEEPGIFMEYAVFD	178
sp P0DJD7 PEPA4_HUMAN	EDSsTYQSTSETVSITYGTGSMTGILGYDTVQVGGISDTNQIFGLSETEPGSFLYYAPFD	180
sp P27678 PEPA4_MACFU	QDSsTYRATSKTVSITYGTGSMTGILGYDTVKVGGISDTNQIFGLSETEPGFFLYFAPFD	180
	: :.. *.:**:*.*:*.**:*.:: . * ***** *** *: :* **	
sp Q9D106 PEPA5_MOUSE	GILGLGYPNLGLQGITPVFDNLWLQGLIPQNLfAFYlSSKDEKGSMLMLGGVDPSYYHGE	238
sp P0DJD7 PEPA4_HUMAN	GILGLAYPSISSGATPVFDNIWNQGLVSQDLfSVYlSADDQSGSVVIFGGIDSSYYTGS	240
sp P27678 PEPA4_MACFU	GILGLAYPSISSGATPVFDNIWNQRLVSQDLfSVYlSADDQSGSVVIFGGIDSSYYTGS	240
	*****.**.:. . * *****:* * *: *:*.*****:*.::***:***:*** *	
sp Q9D106 PEPA5_MOUSE	LHWVPVSKPSYWQLAVDSISMNGEVIACDGGCQGIMDTGTSLLTGPRSSIVNIQNLIGAK	298
sp P0DJD7 PEPA4_HUMAN	LNWVPVTVEGYWQITVDSITMNGEAIACAEGCQAIVDTGTSLLTGPTSPIANIQSDIGAS	300
sp P27678 PEPA4_MACFU	LNWVPVSVEGYWQISVDSITMNGKTIACAKGCQAIVDTGTSLLTGPTSPIANIQSDIGAS	300
	*:****: .***:*****:***:.* ** *.*:***** * *.***. ***.	
sp Q9D106 PEPA5_MOUSE	ASGDGEYFLKCDTINTLPDIVFTIGSVTYVPASAYIRKDRSHNCRSNFEEGMDDPSDPE	358
sp P0DJD7 PEPA4_HUMAN	ENSDGDMVVSCSAISSLPDIVFTINGVQYPVPPSAYILQSE-GSCISGFQGMNLPtesGE	359
sp P27678 PEPA4_MACFU	ENSDGEMVVSCSAISSLPDIVFTINGVQYPLPPSAYILQSQ-GSCTSGFQGMVPTESGE	359
	..***: :..*.:*.:*****.* **:* ***** :.. . * *.*: .. *	
sp Q9D106 PEPA5_MOUSE	MWVLGDVFLRLYFTVFD RANNRIGLAPAA	387
sp P0DJD7 PEPA4_HUMAN	LWILGDVFIRQYFTVFD RANNQVGLAPVA	388
sp P27678 PEPA4_MACFU	LWILGDVFIRQYFTVFD RANNQVGLAPVA	388
	:*:*****:* *****:*****.*	

Задание 2. Гемоглобин (9 позиций)

CLUSTAL O(1.2.4) multiple sequence alignment

sp P02042 HBD_HUMAN	MVHLTPEEKTAVNALWGKVN--VDAVGGEALGRLLVVYPWTQRFFESFGDLSSPDAVMGN	58
sp P68871 HBB_HUMAN	MVHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGN	58
sp P02100 HBE_HUMAN	MVHFTAEEKAAVTSLSWKMN--VEEAGGEALGRLLVVYPWTQRFFDSFGNLSSPSAILGN	58
sp P69891 HBG1_HUMAN	MGHFTTEEDKATITSLWGKVN--VEDAGGETLGRLLVVYPWTQRFFDSFGNLSSASAIMGN	58
sp P69892 HBG2_HUMAN	MGHFTTEEDKATITSLWGKVN--VEDAGGETLGRLLVVYPWTQRFFDSFGNLSSASAIMGN	58
sp Q6B0K9 HBM_HUMAN	--MLSAQERAQIAQVWDLIAGHEAQFGAELLRLFTVYPSTKVYFPHLS-----ACQDA	52
sp P02008 HBAZ_HUMAN	-MSLTKTERTIIIVSMWAKISTQADTIGTETLERLFLSHPQTKTYFPHFD-----LHPGS	53
sp P69905 HBA_HUMAN	-MVLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFD-----LSHGS	53
sp P09105 HBAT_HUMAN	-MALSAEDRALVRALWKKLGSNVGVYTTEALERTFLAFPATKTYFSHLD-----LSPGS	53
	:: ::: : * : * * * : .* *: :* :. .	
sp P02042 HBD_HUMAN	PKVKAHGKKVLGAFSDGLAHLNLTGTFSQLSELHCDKLHVDPENFRLLGNVLVCVLARN	118
sp P68871 HBB_HUMAN	PKVKAHGKKVLGAFSDGLAHLNLTGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHH	118
sp P02100 HBE_HUMAN	PKVKAHGKKVLTSLFGDAIKNMDNLKPAFAKLSELHCDKLHVDPENFKLLGNVMVIIILATH	118
sp P69891 HBG1_HUMAN	PKVKAHGKKVLTSLGDATKHLDDLKGTFAQLSELHCDKLHVDPENFKLLGNVLVTVLAIH	118
sp P69892 HBG2_HUMAN	PKVKAHGKKVLTSLGDAIKHLDDLKGTFAQLSELHCDKLHVDPENFKLLGNVLVTVLAIH	118
sp Q6B0K9 HBM_HUMAN	TQLLSHGQRMLAAVGAAGVQHVNDLRAALSPLADLHALVLRVDPANFPLLIQCFHVVLASH	112
sp P02008 HBAZ_HUMAN	AQLRAHGSKVVAAGDAVKSIDDIGGALSKLSELHAYILRVDPVNFKLLSHCLLVTLAAR	113
sp P69905 HBA_HUMAN	AQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAH	113
sp P09105 HBAT_HUMAN	SQVRAHGQKVADALSLAVERLDDLPHALSALSHLHACQLRVDPASFQLLGHCCLLVTLARH	113
	:: .*.*.: :. . :*:: ::: *:.**.* *:*** .* ** : : ** .	
sp P02042 HBD_HUMAN	FGKEFTPQMQAAYQKVAVGVANALAHKYH	147
sp P68871 HBB_HUMAN	FGKEFTPPVQAAYQKVAVGVANALAHKYH	147
sp P02100 HBE_HUMAN	FGKEFTPEVQAAWQKLVSVAIALAHKYH	147
sp P69891 HBG1_HUMAN	FGKEFTPEVQASWQKMVTAVASALSSRYH	147
sp P69892 HBG2_HUMAN	FGKEFTPEVQASWQKMVTGVASALSSRYH	147
sp Q6B0K9 HBM_HUMAN	LQDEFTVQMQAAWDKFLTGVAVVLTEKYR	141
sp P02008 HBAZ_HUMAN	FPADFTAEAHAAWDKFLSVSSVLTEKYR	142
sp P69905 HBA_HUMAN	LPAEFTPAVHASLDKFLASVSTVLTSKYR	142
sp P09105 HBAT_HUMAN	YPGDFSPALQASLDKFLSHVISALVSEYR	142