

Практикум 9. Поиск предков кристаллинов ГОЛОВОНОГИХ МОЛЛЮСКОВ.

Белки кристаллины выполняют, главным образом, структурные функции. Они обеспечивают прозрачность хрусталика. У разных групп организмов белки кристаллины появлялись независимо и имеют, соответственно, разные предковые белки. Целью данного практикума является поиск предсказанных предковых белков S-кристаллина головоногих.

S-кристаллины – структурные компоненты хрусталика осьминогов и кальмаров. Имеют очень слабую глутатион S-трансферазную активность (собственно, GST и является предполагаемым предковым белком).

Последовательность S-кристаллина была найдена в UniProt – ID: SCRY1_OCTVU, организм: Осьминог обыкновенный.

Затем, для поиска предковых белков, я воспользовался алгоритмом Psi-blast в веб-интерфейсе NCBI blast. Все параметры выставлены по умолчанию, поиск проводился по базе swissprot.

В первой итерации (по-факту, дефолтный blastp), было найдено 95 последовательностей. Среди них обнаружены как другие S-кристаллины, так и (причем гораздо больше) GST (Рис.1)

RecName: Full=Glutathione S-transferase; AltName: Full=GST class-sigma [Nototodarus sloanii]

Sequence ID: [P46088.1](#) Length: 203 Number of Matches: 1

Range 1: 1 to 203 [GenPept](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps											
127 bits(320)	1e-35	Compositional matrix adjust.	72/216(33%)	112/216(51%)	15/216(6%)											
Query 1	MPSYTLHYFNHRGRARS	AVCCS	QLLV	SSTMT	ADRV	FRMG	QHEK	QDAM	SHDAN	VGIG	QQNP	60				
Sbjct 1	MP YTLHYF GRA +C L			DRV M		A + + + +						58				
Query 61	N - -SPEYAMARYLARE	FGFH	GRNN	MEMAR	VD	FISDC	FYD	I	LDDY	MRMY	QDGN	CRMM	FORS	118		
Sbjct 59	TKMSQSMCIARHLARE	FGLD	GKTS	SLEK	YRV	DEIT	ETLQ	D	I	FND	VVKI	-----	KF	107		
Query 119	RDMSSSSEKRMRFQET	CRRIL	PF	MERT	LEM	YNGG	SQY	F	MGD	QTM	ADMM	CY	CALE	NPLME	178	
Sbjct 108	APEAAKEAVQQNYEK	SCKRL	APF	LEGL	LV	SNGG	GG	F	V	GNS	MT	LAD	LHCY	VALE	VPLKH	167
Query 179	EPSMLSSYPKLMALRN	RV	MN	HSK	MSSYL	QRR	CRT	DF								214
Sbjct 168	TPELLKDCPKIVALR	KRVA	EC	PKIA	AAYL	KKR	PVR	DF								203

Рис.1. Выравнивание последовательности S-кристаллина с GST из кальмара (swissprot name: Glutathione S-transferase).

Запустим вторую итерацию с max number of sequences = 500.

На ней уже появилось 231 находка (136 новых).

Интересно, однако, что среди GST затесалось достаточно много ELF1-gamma (Рис.2.):

✓	RecName: Full=Elongation factor 1-gamma 1; Short=EF-1-gamma 1; AltName: Full=eEF-1B... <i>Oryza sativa</i> J...	56.5	56.5	94%	2e-08	16.51%	418	Q9ZRI7.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase theta-2; AltName: Full=GST class-theta-2 [Homo... <i>Homo sapiens</i>	55.8	55.8	82%	2e-08	16.20%	244	P0CG29.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase F12; Short=AtGSTF12; AltName: Full=GST class... <i>Arabidopsis th...</i>	54.6	54.6	66%	5e-08	15.58%	214	Q9FE46.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase F9; Short=AtGSTF9; AltName: Full=AtGSTF7; Alt... <i>Arabidopsis th...</i>	54.6	54.6	79%	5e-08	11.73%	215	O80852.1	✓	✓		
✓	RecName: Full=Protein Gsta [Rhizobium leguminosarum]		Rhizobium leg...	54.2	54.2	92%	7e-08	15.92%	203	Q52828.1	✓	✓
✓	RecName: Full=Glutathione S-transferase 3; AltName: Full=GST class-phi member 3; AltNa... <i>Zea mays</i>	54.2	54.2	78%	7e-08	13.53%	222	P04907.4	✓	✓		
✓	RecName: Full=Glutathione S-transferase theta-3 [Mus musculus]		<i>Mus musculus</i>	54.2	54.2	85%	8e-08	16.13%	241	Q99L20.1	✓	✓
✓	RecName: Full=Elongation factor 1-gamma 2; Short=EF-1-gamma 2; AltName: Full=eEF-1B... <i>Oryza sativa</i> J...	55.0	55.0	77%	8e-08	15.70%	418	Q6YW46.2	✓	✓		
✓	RecName: Full=Glutathione S-transferase theta-2; AltName: Full=GST 12-12; AltName: Full... <i>Rattus norvegi...</i>	54.2	54.2	76%	9e-08	13.86%	244	P30713.3	✓	✓		
✓	RecName: Full=Elongation factor 1-gamma 3; Short=EF-1-gamma 3; AltName: Full=eEF-1B... <i>Oryza sativa</i> J...	54.2	54.2	90%	2e-07	14.63%	416	Q52627.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase F14; Short=AtGSTF14; AltName: Full=GST class... <i>Arabidopsis th...</i>	53.5	53.5	86%	2e-07	11.11%	254	Q9C6C8.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase F13; Short=AtGSTF13; AltName: Full=GST class... <i>Arabidopsis th...</i>	53.1	53.1	80%	2e-07	13.81%	219	Q9LZ19.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase 1; AltName: Full=GST class-phi member 1; AltNa... <i>Zea mays</i>	51.9	51.9	95%	4e-07	16.22%	214	P12653.4	✓	✓		
✓	RecName: Full=Glutathione S-transferase F11; Short=AtGSTF11; AltName: Full=AtGSTF6;... <i>Arabidopsis th...</i>	51.5	51.5	86%	6e-07	13.64%	214	Q96324.1	✓	✓		
✓	RecName: Full=Probable glutathione S-transferase GSTF2; AltName: Full=GST-II [Oryza sa... <i>Oryza sativa</i> J...	51.5	51.5	97%	6e-07	15.62%	215	O82451.3	✓	✓		
✓	RecName: Full=Probable elongation factor 1-gamma 1; Short=EF-1-gamma 1; AltName: Full... <i>Arabidopsis th...</i>	52.3	52.3	80%	6e-07	20.11%	414	O04487.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase theta-1; AltName: Full=GST class-theta-1 [Bos ta... <i>Bos taurus</i>	51.1	51.1	76%	1e-06	15.57%	240	Q2NL00.3	✓	✓		
✓	RecName: Full=Glutathione S-transferase F8_chloroplastic; Short=AtGSTF8; AltName: Full... <i>Arabidopsis th...</i>	50.8	50.8	74%	2e-06	13.41%	263	Q96266.3	✓	✓		
✓	RecName: Full=Glutathione S-transferase D7 [Drosophila melanogaster]		<i>Drosophila me...</i>	50.0	50.0	99%	2e-06	13.00%	224	Q9VG93.1	✓	✓
✓	RecName: Full=Glutathione S-transferase psoE; AltName: Full=Pseurotin biosynthesis prote... <i>Aspergillus fu...</i>	49.6	49.6	79%	3e-06	15.43%	237	Q4WB03.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase-like protein OpS6; AltName: Full=Oosporein bios... <i>Beauveria bas...</i>	49.2	49.2	87%	4e-06	12.95%	218	J4UH08.1	✓	✓		
✓	RecName: Full=Dichloromethane dehalogenase; Short=DCM dehalogenase [Methylophilus J... <i>Methylophilus...</i>	49.6	49.6	89%	4e-06	14.29%	267	P43387.2	✓	✓		
✓	RecName: Full=Elongation factor 1-gamma; Short=EF-1-gamma; AltName: Full=eEF-1B ga... <i>Prunus avium</i>	50.0	50.0	80%	5e-06	16.48%	422	Q9FUM1.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase 1; AltName: Full=GST class-theta [Manduca sexta] <i>Manduca sexta</i>	48.4	48.4	81%	7e-06	17.22%	217	P46430.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase 1-1; AltName: Full=DDT-dehydrochlorinase; AltN... <i>Drosophila tej...</i>	48.1	48.1	57%	8e-06	15.75%	200	P30107.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase 1-1; AltName: Full=DDT-dehydrochlorinase; AltN... <i>Drosophila ma...</i>	48.1	48.1	57%	8e-06	15.75%	200	P67804.1	✓	✓		
✓	RecName: Full=Elongation factor 1-gamma; Short=EF-1-gamma; AltName: Full=eEF-1B ga... <i>Trypanosoma...</i>	48.8	48.8	65%	9e-06	17.24%	411	P34715.1	✓	✓		
✓	RecName: Full=Glutathione S-transferase theta-1; AltName: Full=GST class-theta-1 [Mus m... <i>Mus musculus</i>	48.1	48.1	68%	1e-05	15.44%	240	Q64471.4	✓	✓		

Рис.2. Часть новых находок из второй итерации.

В оригинальной статье (Tomarev, S.I., R.D. Zinovieva, and J. Piatigorsky, Crystallins of the octopus lens. Recruitment from detoxification enzymes. J Biol Chem, 1991. 266(35): p. 24226-31.) не было сказано ничего о ELF1 как о предполагаемом предке S-кристаллинов. Более того, ферментативная активность ELF1 никак не связана с глутатион S-трансферазой. Интересно.

Запустим третью итерацию.

Суммарно уже 397 находок, среди них самые разные белки (maiA, URE2...).

На 4-й итерации было 443 находки, на 5-й – 465, на 6-й – 478. Думаю, на этом итерации можно прекращать.

В принципе, такое большое число итераций было и не нужно – предковые белки нашлись буквально на первой.

На поздних итерациях GST практически не находились (Рис. 3)

<input checked="" type="checkbox"/>	RecName: Full=Aminoacyl tRNA synthase complex-interacting multifunctional protein 2; Alt...	<i>Cricetulus gris...</i>	50.4	50.4	36%	3e-06	16.67%	320	Q9WVM7.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Chloride intracellular channel protein 1; AltName: Full=Glutaredoxin-like oxi...	<i>Sus scrofa</i>	47.3	47.3	40%	4e-06	7.95%	110	Q29238.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Metaxin-2; AltName: Full=Mitochondrial outer membrane import complex pr...	<i>Sus scrofa</i>	49.2	49.2	71%	6e-06	12.30%	267	Q2L969.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Metaxin-2; AltName: Full=Mitochondrial outer membrane import complex pr...	<i>Mus musculus</i>	48.8	48.8	71%	8e-06	12.30%	263	Q88441.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Metaxin-2; AltName: Full=Mitochondrial outer membrane import complex pr...	<i>Homo sapiens</i>	48.4	48.4	76%	1e-05	12.69%	263	Q75431.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Probable methionine--tRNA ligase_cytoplasmic; AltName: Full=Methionyl-t...	<i>Schizosaccha...</i>	48.4	48.4	56%	2e-05	8.20%	782	Q9UUF2.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta 1; Short=EF-1-beta 1; AltName: Full=Elongation fa...	<i>Arabidopsis th...</i>	46.9	46.9	27%	3e-05	24.56%	228	Q84WM9.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=p30 [Xenopus la...	<i>Xenopus laevis</i>	46.5	46.5	32%	3e-05	15.71%	227	P30151.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Probable aminoacyl tRNA synthase complex-interacting multifunctional pro...	<i>Drosophila me...</i>	46.5	46.5	26%	5e-05	34.55%	334	Q9VUR3.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta [Gallus gallus]	<i>Gallus gallus</i>	45.7	45.7	32%	6e-05	15.71%	225	Q9YGO1.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Failed axon connections homolog [Rattus norvegicus]	<i>Rattus norvegi...</i>	46.1	46.1	81%	7e-05	9.36%	409	D3ZAT9.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Failed axon connections homolog [Mus musculus]	<i>Mus musculus</i>	46.1	46.1	81%	7e-05	9.36%	409	Q3UMF9.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Prostaglandin E synthase 2; AltName: Full=Membrane-associated prostagl...	<i>Homo sapiens</i>	46.1	46.1	91%	7e-05	10.00%	377	Q9H7Z7.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Prostaglandin E synthase 2; AltName: Full=Microsomal prostaglandin E sy...	<i>Macaca fascic...</i>	46.1	46.1	91%	7e-05	10.00%	377	Q9N0A4.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Failed axon connections homolog [Homo sapiens]	<i>Homo sapiens</i>	45.7	45.7	81%	1e-04	9.85%	409	Q5TGI0.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=Elongation factor...	<i>Oryza sativa J...</i>	45.0	45.0	31%	1e-04	14.93%	224	P29545.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Prostaglandin E synthase 2; AltName: Full=GATE-binding factor 1; Shorte...	<i>Mus musculus</i>	45.0	45.0	98%	2e-04	7.64%	384	Q8BWM0.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=Elongation factor...	<i>Triticum aestiv...</i>	44.2	44.2	26%	2e-04	19.64%	216	P29546.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta [Xenopus tropicalis]	<i>Xenopus tropi...</i>	44.2	44.2	32%	3e-04	15.71%	228	Q6DET9.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=eEF-1B alpha [H...	<i>Homo sapiens</i>	43.8	43.8	32%	3e-04	17.14%	225	P24534.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=eEF-1B alpha [R...	<i>Rattus norvegi...</i>	43.8	43.8	33%	3e-04	16.67%	225	A6IPG1.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=eEF-1B alpha [S...	<i>Sus scrofa</i>	43.8	43.8	32%	4e-04	17.14%	225	P29412.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=eEF-1B alpha [B...	<i>Bos taurus</i>	43.8	43.8	32%	4e-04	17.14%	225	Q5E983.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Chloride intracellular channel Clic [Drosophila melanogaster]	<i>Drosophila me...</i>	43.8	43.8	89%	4e-04	7.00%	260	Q9VY78.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=eEF-1B alpha [M...	<i>Mus musculus</i>	43.0	43.0	33%	5e-04	16.67%	225	Q70251.5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta 2; Short=EF-1-beta 2; AltName: Full=Elongation fa...	<i>Arabidopsis th...</i>	43.0	43.0	26%	5e-04	19.64%	224	Q9SCX3.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Elongation factor 1-beta; Short=EF-1-beta; AltName: Full=eEF-1B alpha [O...	<i>Oryctolagus c...</i>	42.7	42.7	32%	7e-04	17.14%	225	P34826.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Glutathione S-transferase C-terminal domain-containing protein [Xenopus l...	<i>Xenopus laevis</i>	42.7	42.7	34%	0.001	15.58%	613	Q6DDT5.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=S-crystallin 3; AltName: Full=Crystallin SIII [Nototodarus gouldi]	<i>Nototodarus g...</i>	38.4	38.4	18%	0.002	31.58%	38	P15534.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Glutathione S-transferase omega-like 1; AltName: Full=Glutathione-depend...	<i>Saccharomyc...</i>	42.3	42.3	66%	0.002	13.02%	356	P48239.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Prostaglandin E synthase 2; AltName: Full=Membrane-associated prostagl...	<i>Bos taurus</i>	41.9	41.9	98%	0.002	8.45%	372	Q66LN0.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Рис.3. Новые находки 6-й итерации.

Тем не менее, на 6-й итерации нашлась одна GST из дрожжей (Рис.4.):

RecName: Full=Glutathione S-transferase omega-like 1; AltName: Full=Glutathione-dependent dehydroascorbate reductase [Saccharomyces cerevisiae S288C]

Sequence ID: [P48239.1](#) Length: 356 Number of Matches: 1

Range 1: 140 to 306 [GenPept](#) [Graphics](#)

▾ Next Match ▲ Previous Match

Related Information

[Gene](#) - associated gene details
[AlphaFold Structure](#) - 3D structure displays

Score	Expect	Method	Identities	Positives	Gaps
42.3 bits(98)	0.002	Composition-based stats.	22/169(13%)	50/169(29%)	29/169(17%)
Query 52		NVIGIGQQ-----NPNSPEYAMARYLAREFGFHGRNNMEMARVDFISDCFYDILDDYMRMY			106
		V + N+ + R L + E +D + D +D ++			
Sbjct 140		TVPVLWDSKTRKIVNNESGDIIRILNSGVFDEFIQSEETNVIDLVPDLIDEIDKNIKWV			199
Query 107		QDGNCRMMFQSRDMSSSSEKRMRFQETCRRILPFME-----RTLEMYNGGSQ-----			154
		+++ + + +T L ME + LE G++			
Sbjct 200		HPKINLGVYKVGLAENG--KIYETEVKTLFENLQKMECVLKENYKRLLEEQFSGNKQKILA			257
Query 155		--YFMGDQMTMADMMCYCALENPLMEEPS-----MLSSYPKLMALR 193			
		+ +G ++T AD+ Y ++ + +P L			
Sbjct 258		KYFVLGQRLTEADIRLYPSIIRFDVVYVQHFKCNLKTIRDGFPYLHLWL 306			

Рис.4. Выравнивание S-кристаллина с GST из дрожжей (swissprot name: Glutathione S-transferase omega-like 1).

Какие выводы можно сделать? Во-первых, psi-blast не был необходим для поиска предков S-кристаллина – предковые белки нашлись и на первой итерации.

Во-вторых, S-кристаллины, видимо, достаточно консервативные белки – их гены, произойдя от генов, кодирующих GST (видимо, вследствие паралогизаций (см. Рис.1.)), не претерпевали значимых изменений до сегодняшних дней, поэтому и предковые белки можно найти стандартным blastp.