

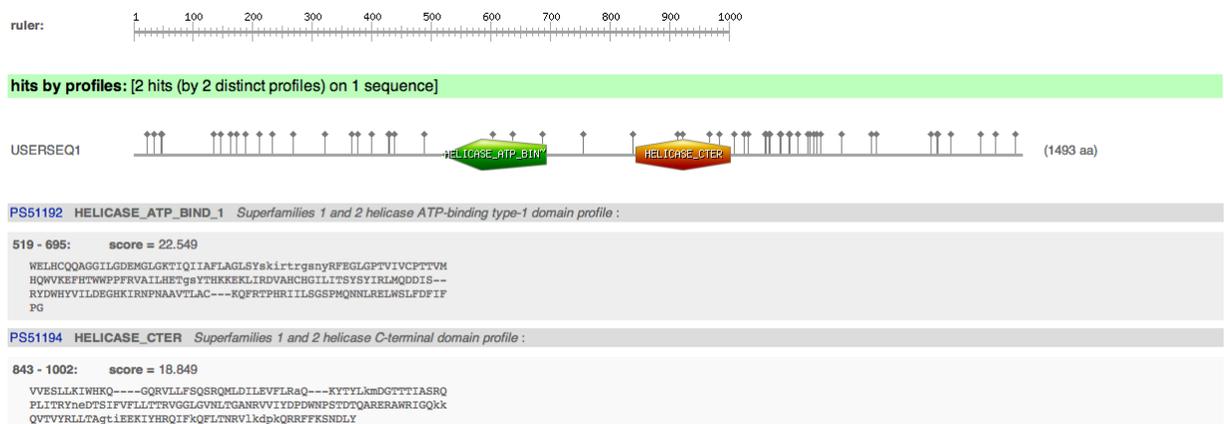
Alesandra Espinoza
 BioChem 118
 Doug Brutlag
 5 December 2012

ERCC-6 – Cockayne Syndrome CSB

The protein ERCC-6 is mainly involved in transcription-coupled nucleotide excision repair. This allows RNA polymerase II-blocking lesions to be quickly removed from the transcribed strand of active genes. It locally modifies DNA conformation by wrapping the DNA around itself, which modifies the interface between stalled RNA polymerase II and DNA. This protein recruits the CSA complex, nucleotide excision repair proteins, and EP300 to the sites of RNA polymerase II-blocking lesions. Defects and mutations in ERCC-6 cause Cockayne syndrome type B, or CSB. Cockayne syndrome is a disorder that is characterized by sensitivity to sunlight, abnormal and slow growth, impaired development of the nervous system, premature aging, and also possibly hearing loss and eye abnormalities. When the ERCC-6 is altered, DNA damage is not repaired, causing Cockayne syndrome type B.

1. MyHits / Prosite

The following are significant hits from Prosite. They show high sequence frequency with helicase ATP binding 1, CK2 phosphorylation site, PKC phosphorylation site, CAMP and TYR phosphorylation sites, amidation site, and ASN glycosylation site. Amidation site is a c terminus consensus sequence, required for c terminus amidation of peptides. Glycosylation refers to the enzymatic process that attaches glycans to proteins, lipids, or other organic molecules. This enzymatic process produces one of the fundamental biopolymers found in cells (along with DNA, RNA, and proteins). Some proteins do not fold correctly unless they are glycosylated first. With this information, ERCC-6 has a helicase ATP binding site. Also, according to the UniProt site, ERCC-6 does not contain cAMP induced TRY phosphorylation sites, amidated sites, or ASN glycosylation site. In addition, once it undergoes DNA damage, it is phosphorylated by ATM or ATR. UniProt also mentions that it does, however, undergo ubiquitination by the CSA complex. This leads to ERCC6 proteasomal degradation in a UV-dependent manner. It stabilized once it interacts with KIAA1530/UVSSA. This promotes recruitment of the enzyme USP7, which deubiquitinates ERCC6, preventing UV-induced degradation of ERCC6 by the proteasome.



2[Type text]

[Type text]

[Type text]

hits by profiles with a high probability of occurrence: [1 hit (by 1 profile) on 1 sequence]

Hits by **PS50313** **GLU_RICH** *Glutamic acid-rich region profile* :

USERSEQ1  (1493 aa)

362 - 392: score = 8.670

EaegdsegeeseyfpteeeeeedevegaE

hits by patterns with a high probability of occurrence or by user-defined patterns: [87 hits (by 7 distinct patterns) on 1 sequence]

USERSEQ1  (1493 aa)

PS00006 **CK2_PHOSPHO_SITE** *Casein kinase II phosphorylation site* :

24 - 27: SnnE

Predicted feature:

MOD_RES 24 Phosphoserine (By similarity) [condition: S]

35 - 38: SggD

Predicted feature:

MOD_RES 35 Phosphoserine (By similarity) [condition: S]

49 - 52: SvgD

Predicted feature:

MOD_RES 49 Phosphoserine (By similarity) [condition: S]

135 - 138: SvLD

Predicted feature:

MOD_RES 135 Phosphoserine (By similarity) [condition: S]

163 - 166: TsrD

Predicted feature:

MOD_RES 163 Phosphothreonine (By similarity) [condition: T]

212 - 215: SLeE

Predicted feature:

MOD_RES 212 Phosphoserine (By similarity) [condition: S]

234 - 237: TawE

Predicted feature:

MOD_RES 234 Phosphothreonine (By similarity) [condition: T]

269 - 272: SgEE

Predicted feature:

MOD_RES 269 Phosphoserine (By similarity) [condition: S]

322 - 325: SkkE

Predicted feature:

MOD_RES 322 Phosphoserine (By similarity) [condition: S]

367 - 370: SegE

367 - 370:	SegE			
Predicted feature:				
MOD_RES	367	Phosphoserine (By similarity)		[condition: S]
377 - 380:	TeeE			
Predicted feature:				
MOD_RES	377	Phosphothreonine (By similarity)		[condition: T]
400 - 403:	TdyE			
Predicted feature:				
MOD_RES	400	Phosphothreonine (By similarity)		[condition: T]
429 - 432:	SegE			
Predicted feature:				
MOD_RES	429	Phosphoserine (By similarity)		[condition: S]
430 - 433:	SgeE			
Predicted feature:				
MOD_RES	430	Phosphoserine (By similarity)		[condition: S]
438 - 441:	SvgE			
Predicted feature:				
MOD_RES	438	Phosphoserine (By similarity)		[condition: S]
489 - 492:	SdaE			
Predicted feature:				
MOD_RES	489	Phosphoserine (By similarity)		[condition: S]
636 - 639:	SryD			
Predicted feature:				
MOD_RES	636	Phosphoserine (By similarity)		[condition: S]
687 - 690:	SlfD			
Predicted feature:				
MOD_RES	687	Phosphoserine (By similarity)		[condition: S]
755 - 758:	SlpD			

PS00005 PKC_PHOSPHO_SITE *Protein kinase C phosphorylation site :*

46 - 48:	SER			
Predicted feature:				
MOD_RES	46	Phosphoserine (By similarity)		[condition: S]
146 - 148:	SLR			
Predicted feature:				
MOD_RES	146	Phosphoserine (By similarity)		[condition: S]
163 - 165:	TeR			
Predicted feature:				
MOD_RES	163	Phosphothreonine (By similarity)		[condition: T]
173 - 175:	SvK			
Predicted feature:				
MOD_RES	173	Phosphoserine (By similarity)		[condition: S]
189 - 191:	TaK			
Predicted feature:				
MOD_RES	189	Phosphothreonine (By similarity)		[condition: T]
322 - 324:	SxK			
Predicted feature:				
MOD_RES	322	Phosphoserine (By similarity)		[condition: S]
604 - 606:	ThK			
Predicted feature:				
MOD_RES	604	Phosphothreonine (By similarity)		[condition: T]
838 - 840:	SgK			
Predicted feature:				
MOD_RES	838	Phosphoserine (By similarity)		[condition: S]

4[Type text]

[Type text]

[Type text]

PS00004 CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site :			
186 - 189:	KKIT		
1334 - 1337:	KRnS		
PS00009 AMIDATION Amidation site :			
410 - 413:	gGKR		
445 - 448:	gGRK		
1325 - 1328:	aGKK		
1331 - 1334:	fGKK		
PS00007 TYR_PHOSPHO_SITE Tyrosine kinase phosphorylation site :			
453 - 461:	RddgDedyY		
1172 - 1179:	Kqm.EnnfY		
PS00001 ASN_GLYCOSYLATION N-glycosylation site :			
922 - 925:	NLTG		
Predicted feature:			
CARBOHYD	922	N-linked (GlcNAc...) (Potential)	[condition: none]
1060 - 1063:	NISV		
Predicted feature:			
CARBOHYD	1060	N-linked (GlcNAc...) (Potential)	[condition: none]
1086 - 1089:	NRSD		
Predicted feature:			
CARBOHYD	1086	N-linked (GlcNAc...) (Potential)	[condition: none]
1100 - 1103:	NVTS		
Predicted feature:			
CARBOHYD	1100	N-linked (GlcNAc...) (Potential)	[condition: none]
1131 - 1134:	NSSG		
Predicted feature:			
CARBOHYD	1131	N-linked (GlcNAc...) (Potential)	[condition: none]
PS00001 ASN_GLYCOSYLATION N-glycosylation site :			
922 - 925:	NLTG		
Predicted feature:			
CARBOHYD	922	N-linked (GlcNAc...) (Potential)	[condition: none]
1060 - 1063:	NISV		
Predicted feature:			
CARBOHYD	1060	N-linked (GlcNAc...) (Potential)	[condition: none]
1086 - 1089:	NRSD		
Predicted feature:			
CARBOHYD	1086	N-linked (GlcNAc...) (Potential)	[condition: none]
1100 - 1103:	NVTS		
Predicted feature:			
CARBOHYD	1100	N-linked (GlcNAc...) (Potential)	[condition: none]
1131 - 1134:	NSSG		
Predicted feature:			
CARBOHYD	1131	N-linked (GlcNAc...) (Potential)	[condition: none]
1238 - 1241:	NKSE		
Predicted feature:			
CARBOHYD	1238	N-linked (GlcNAc...) (Potential)	[condition: none]
1338 - 1341:	NFSV		
Predicted feature:			
CARBOHYD	1338	N-linked (GlcNAc...) (Potential)	[condition: none]

Legend:



2. Interpro

These are all statistically significant hits. There were no insignificant hits to be found. The first one is biologically significant too because it reveals that they are SNF-2 related, which means that the protein is involved in many processes, including transcription regulation, DNA repair, DNA recombination, and chromatin unwinding. Proteins in this domain are distantly related to the DEAX box helicases, but not helicase activity has been demonstrated for these sort of proteins.

Helicase, C-terminal is also biologically significant because included in this group is the eukaryotic translation initiation factor 4A (eIF4A), a member of the DEA(D/H)-box RNA helicase family. Helicases have been classified in 5 superfamilies. Two superfamilies encompass a large number of DNA and RNA helicases from archaea, eubacteria, eukaryotes and viruses.

InterProScan Results

[Summary Table](#)
[Tool Output](#)
[Visual Output](#)
[Submission Details](#)
[Submit Another Job](#)

InterProScan Summary

IPR000330 SNF2-related

Method	Identifier	Description	Matches
PFAM	PF00176	SNF2_N	2.8E-81 [510-812] T

Parent No parent

Children No children

Found in No entries

Contains IPR014001

GO terms GO:0003677 DNA binding
GO:0005524 ATP binding

IPR001650 Helicase, C-terminal

Method	Identifier	Description	Matches
PFAM	PF00271	Helicase_C	1.1E-13 [874-952] T
SMART	SM00490	helicase superfamily c-terminal domain	3.3E-21 [869-952] T
PROFILE	PS51194	HELICASE_CTER	18.849 [843-1002] T

Parent No parent

Children No children

Found in IPR000185 IPR004576 IPR004589 IPR004609 IPR004807 IPR005259 IPR006293 IPR006474
IPR010222 IPR010225 IPR013444 IPR014018 IPR014412 IPR016438 IPR017170 IPR017575
IPR022307

Contains No entries

GO terms GO:0003676 nucleic acid binding
GO:0004386 helicase activity
GO:0005524 ATP binding

6[Type text]

[Type text]

[Type text]

IPR014001 Helicase, superfamily 1/2, ATP-binding domain

Method	Identifier	Description	Matches
SMART	SM00487	DEAD-like helicases superfamily	1.8E-33 [503-703] T
PROFILE	PS51192	HELICASE_ATP_BIND_1	22.549 [519-695] T

Parent No parent

Children IPR006935 IPR011492 IPR011545

Found in IPR000330 IPR003450 IPR004473 IPR004483 IPR004589 IPR004807 IPR006293 IPR006310 IPR006474 IPR010222 IPR010225 IPR017548 IPR017575 IPR022307

Contains IPR000629 IPR002464 IPR010614 IPR014013

GO terms None

noIPR unintegrated

Method	Identifier	Description	Matches
GENE3D	G3DSA:3.40.50.300	no description	3.8E-20 [833-975] T
PANTHER	PTHR10799	SWI/SNF-RELATED MATRIX-ASSOCIATED ACTIN-DEPENDENT REGULATOR OF CHROMATIN SUBFAMILY-RELATED	0.0 [1-1056] T
PANTHER	PTHR10799:SF49	DNA EXCISION REPAIR PROTEIN ERCC-6 (COCKAYNE SYNDROME PROTEIN CSB)	0.0 [1-1056] T
SUPERFAMILY	SSF52540	P-loop containing nucleoside triphosphate hydrolases	1.4E-79 [468-745] T 3.9E-69 [747-1007] T

Parent No parent

Children No children

Found in No entries

Contains No entries

GO terms None

3. BLAST

Significant Hits:

The following two are significant hits because they discover function and sequence similarity. The first one is significant because it brings up the TATA binding protein. This binding protein is a general transcription factor that binds specifically to a DNA sequence called the TATA box. TBP, along with a variety of TBP-associated factors, make up the TFIID, a general transcription factor that in turn makes up part of the RNA polymerase II pre-initiation complex. It helps position RNA polymerase II over the transcription start site of the gene. It is involved in DNA melting (double strand separation) by bending the DNA by 80° (the AT-rich sequence to which it binds facilitates easy melting). The TBP is an unusual protein in that it binds the minor groove using a β sheet. When TBP binds to a TATA box within the DNA, it distorts the DNA by inserting amino acid side-chains between base pairs, partially unwinding the helix, and doubly kinking it. TBP binds with the negatively charged phosphates in the DNA backbone through positively charged lysine and arginine amino acid residues. The strain imposed on the DNA through this interaction initiates melting, or separation, of the strands. Separation of the two strands exposes the bases and allows RNA polymerase II to begin transcription of the gene.

The second hit is biologically significant because it is involved in the excision repair of DNA damage. It is a protein that recognizes and binds damaged DNA in an ATP-dependent manner during nucleotide excision repair.

Both of these are significant since they are extremely closely related to the known biological function of ERCC-6.

> [sp|O43065.4|MOT1_SCHPO](#) **G** RecName: Full=Probable helicase mot1; AltName: Full=Modifier of transcription 1; AltName: Full=TBP-associated factor mot1
Length=1953

[GENE ID: 2539633 mot1](#) | TATA-binding protein associated factor Mot1
[Schizosaccharomyces pombe 972h-] (10 or fewer PubMed links)

Score = 301 bits (770), Expect = 1e-81
Identities = 194/533 (36%), Positives = 280/533 (53%), Gaps = 47/533 (9%)

```

Query 495  EGFKVPGFLFKKLFKYQQTGVRWLWELHCQQAGGILGDEMGLGKTIQIIAFLAGLSYS-- 552
           E F +P + L KYQQ GV WL L+ + GIL D+MGLGKT+Q I +A Y+
Sbjct 1346  EAFSIPVPISADLRKYQQEGVNWLAFLNKYELHGILCDMGLGKTLQTICIVASDHYNRQ 1405

Query 553  KIRTRGSNRYFEGLGPTVIVCPTTVMHQVWKEFHTWPPFRVAILHETGSYTHKKEKLIR 612
           K+ + +F + P++IVCP+T+ W +E T+ P +V+ + K IR
Sbjct 1406  KLFEESGSPKFAHV-PSLIVCPSTLAGHWQQLSTYAPFLKVSAYVGPAAERAK----IR 1460

Query 613  DVAHCHGILITSYSYIRLMQDDISRYDWHYVILDEGHKIRNPNAAVTLACKQFRTPHRII 672
           +++TSY R D++ + DW+Y +LDEGH I+N A +T A K R+ HR+I
Sbjct 1461  SKMKKSDVVVTSYDICRNDVDELVKIDWNYCVLDEGHVKNARAKLTKAVKSLRSYHRLI 1520

Query 673  LSGSPMQNNLRELWLSLDFDFIFPGKLGTLFVFMQFVSPITMGGYSNASPVQVKTAYKCAC 732
           LSG+P+QNN+ ELWLSLDFD+ PG LGT F E+F PI + +S + +
Sbjct 1521  LSGTPIQNNVLELWLSLDFDFLMPGFLGTEKTFQERFVRPIAASRDAKSSSKERERGTLE 1580

Query 733  VLRDTINPYLLRRMKS DVKMSLSLPDKNEQVLFCLRTDEQHKVYQNFVDSKEVYRILNGE 792
           + + P++LRR+K DV LP K Q +C ++D Q K+ +FV + L +
Sbjct 1581  AIHKQVLPFMLRRLKEDV--LADLPPKIIQDYICDMSDLQRKLLNDFVSQLNINEELED 1638

Query 793  -----MQIFSGLIALRKICNHPDLFSGGPKNLKGLPDDELEEDQFGY--W 835
           IF L +RK+CNHP L + +L ++ G
Sbjct 1639  ETEKTQGTRKKKSQKAHIFQALQYMRKLCNHPALILTEKHPKRNAIVKQLAKENSGLHDL 1698

Query 836  KRSGKMIVVESLLK-----IWHKQG-----QRVLLFSQSRQMLDILEVFL-- 875
           K + K+ + LL+ + G RVL+F Q + MLD++E L
Sbjct 1699  KHAPKLTALGQLLRDCGLGNSSVNSNGIDSALTNVSEHRVLIQFCQLKDMMLDMVEKDLLQ 1758

Query 876  -RAQKYTYLKMMDGTTTTIASRQPLITRYNEDTSIFVLLTTRVGGGLGVNLTGANRVVIYDP 934
           TY+++DG+ RQ +T++N D SI V LLTT VGGLG+NLTGA+ V+ +
Sbjct 1759  ATPMDVTYMRLDGSVEPTKRQEAVTKFNNDPSIDVLLLTTHVGGGLGLNLTGADTVIFVEH 1818

Query 935  DWNPSTDTQARERAWRIGQKKQVTVYRLLTAGTIEEKIYHRQIFKQFLTNRVL 987
           DWNP D QA +RA RIGQKK V VYRL+T G +EEDI Q FK + + V+
Sbjct 1819  DWNPMRDLQAMDRAHRIGQKKVVVYRLLITRGCLEEKIMGLQRFKMNVAHVTVV 1871

```

8[Type text]

[Type text]

[Type text]

2.

> [sp|P31244.1|RAD16_YEAST](#) **C** RecName: Full=DNA repair protein RAD16; AltName: Full=ATP-dependent helicase RAD16
Length=790

[GENE ID: 852411 RAD16](#) | Rad16p [Saccharomyces cerevisiae S288c]
(Over 10 PubMed links)

Sort alignments for this subject sequence by:
E value [Score](#) [Percent identity](#)
[Query start position](#) [Subject start position](#)

Score = 89.4 bits (220), Expect = 1e-16
Identities = 102/412 (25%), Positives = 165/412 (40%), Gaps = 77/412 (19%)

```
Query 469 NKLRLQDKEKRLK-LEDDSEESDAEFDEGFKVPGFLFKKLFKYQQTGVRWLW-ELHCQQA 526
N LRL + L+ + D + + + K P + KL +Q G+ WL + A
Sbjct 145 NTLRLYEHHPELRNVFTDLKNAPPYVQPSKQPDGMTIKLLPFQLEGLHWLISQEESIYA 204

Query 527 GGILGDEMGLGKTIQIIAFLAGLSYSKIRTRGSNYRFEGLGPTVIVCPTTVMHQWVKEF- 585
GG+L DEMG+GKTIQ IA L +K P+++V PT + QW E
Sbjct 205 GGVLADEMGMGKTIQTIALLMN-DLTK-----SPLVVPPTVALMQWKNIEIE 250

Query 586 -HTWPPFRVAILHETGSIYTHKKEKLRDVAHCHGILITS-----YSYIRL-----MQD 633
HT ++ I H T K+ DV ++ S Y + R
Sbjct 251 QHTKGQ-LKIYIYHGASRTTDIKDLQGYDVVLTYYAVLESVFRKQNYGFRRNGLFKQPS 309

Query 634 DISRYDWHYVILDEGHKIRNPNAAVTLACKQFRTPHRIILSGSPMQNNLRELWLSLDFDIF 693
+ D++ VILDE H I++ + A +T R LSG+P+QN + E++SL F+
Sbjct 310 VLHNIDFYRVILDEAHNIKDRQSNTRAVNNLKTQKRWCLSGTPLQNRIGEMYSLIRFL- 368

Query 694 PGKLGTLVPFMEQFSVP-----ITMGYSNASPVQVKTAYKAC 732
+ F + F + M + + +K K
Sbjct 369 -----NINPFTKYFCTKDCASKDWKFTDRMHCDHCSHVIMQHTNFFNHFMLKNIQKFGV 423

Query 733 V-----LRDTINPYLLRRMKS DVKMSLSLDPKNEQVLFCLRTDEQHKVYQN-FVDS 782
++ + +LRR K + L LP + V +E+ +Y++ + DS
Sbjct 424 EPGGLESFNNIQTLKNIMLRRTKVERADDLGLPPRIVTVRRDFNEEEKDLYRSLYTDS 483

Query 783 KEVYR-----ILNGEMQIFSGLIALRKICNHPDLFSGGPKNLKGLPDEL 827
K Y +LN IF+ + +R++ +HPDL K L P D++
Sbjct 484 KRKYNFVEEGVVLNNYANIFTLITRMRQLADHPDLVL---KRLNNFPGDDI 532
```

Score = 82.0 bits (201), Expect = 2e-14
Identities = 49/146 (34%), Positives = 82/146 (56%), Gaps = 2/146 (1%)

```
Query 833 GYWKRSQGM-IVVESLLKIW-HKQQRVLLFSQSRQMLDILEVFLRAQKYTYLKMDGTTT 890
G W+ S K+ +VE L K+ +K+ + +FSQ MLD++E L+ + +K+ G+ +
Sbjct 613 GKWQSSTKIEALVEELYKLRNKRITIKSIVFSQFTSMLDLVEWRLKRAGFQTVKLGQSMS 672

Query 891 IASRQPLITRYNEDTSIFVLLTTRVGLGVNLTGANRVVIYDPDWNPSDTPQARERAWR 950
R I + + VFL++ + GG+ +NL A++V I DP WNPS + Q+ +R R
Sbjct 673 PTQRDETIKYFMNNIQCEVFLVSLKAGGVALNLCEASQVFILDPWNNPSVEWQSGDRVHR 732

Query 951 IGQKKQVTVYRLLTAGTIEEKIYHRQ 976
IGQ + V + R +IE +I Q
Sbjct 733 IGQYRPVKITRFCIEDSIEARIIEIQ 758
```

Insignificant Hits:

10[Type text]

[Type text]

[Type text]

The following are insignificant hits because they do not have a good enough expectation values, given that they have an E-value of 1 or higher, meaning that they could occur by pure chance.

>|sp|B0Z503.1|YCF2_OENBI|G RecName: Full=Protein ycf2 Length=2325

GENE ID: 5951957 ycf2 | ycf2 [Oenothera biennis] (10 or fewer PubMed links)
GENE ID: 5952025 ycf2 | ycf2 [Oenothera biennis] (10 or fewer PubMed links)

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Score = 40.8 bits (94), Expect = 0.11
Identities = 42/155 (27%), Positives = 68/155 (44%), Gaps = 7/155 (5%)
Query 356 ESDMRPEAEGDSEGESEYFPTEEEEE-EDDEVEGAADLSGDGTDYELKPLPKGGKRQ 414
E ++ + + EG E E TE+EE E ++EVEG E D +GT+ E++
Sbjct 1890 EEEVEGTEDEVEGTEEEVEGTEDEEGEGTEEEVEGTE-DEEGEGTEEEVEGTEDEEGEG 1948
Query 415 KKVVPQEIIDDDFFPSSGEEAASVGE GGGGGRKVG RYRDDGDEDYKQLRRWNKLRQL 474
+ V+ +++ G E E + G G + + D D R + N L +Q
Sbjct 1949 TEEVEGTEEEV---EGTEEEVEGTEDEEGEGTEKDSSQFDNDRVTL L L R F K P R N P L D I Q 2005
Query 475 DK-EKRLKLEDDSEESDAEFDEGFKVPGFLFKLFL 508
+ K E + EE D + DE P + + LF
Sbjct 2006 RLIIYQH QKYESELEEDDDD-DEDVFAPQKMLEDLF 2039

Score = 40.0 bits (92), Expect = 0.19
Identities = 28/92 (30%), Positives = 43/92 (47%), Gaps = 6/92 (7%)
Query 364 EGDSEGESEYFPTEEEEEEDDEVEGAADLSG-----DGTDYELKPLPKGGKRQKKV 417
E + EG E E TEEE E ++EVEG E ++ G +GT+ E++ +
Sbjct 1862 EEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEDEEEVEGTEEEVEGTEDEEGEGT 1921
Query 418 PVQEIIDDDFFPSSGEEAASVGE GGGGGRKV 449
V+ +D+ + EE E EG G +V
Sbjct 1922 EVEGTEDEEGEGTEEEVEGTEDEEGEGTEEV 1953

Score = 38.1 bits (87), Expect = 0.72
Identities = 26/88 (30%), Positives = 41/88 (47%), Gaps = 5/88 (6%)
Query 376 PTEEEEEEDDEVEGAADLSG-----DGTDYELKPLPKGGKRQKKVQEIIDDDFFPSS 430
PTEEE E ++EVEG E ++ G +GT+ E++ + V+ +D+ +
Sbjct 1860 PTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEDEEEVEGTEEEVEGTEDEEGEGT 1919
Query 431 GEEAASVGE GGGGGRKVG RYRDDGDE 458
EE E EG G +V D+ E
Sbjct 1920 EEEVEGTEDEEGEGTEEEVEGTEDEE 1947

2.

>|sp|C5DF79.1|RRP36_LACTC|G RecName: Full=rRNA biogenesis protein RRP36; AltName: Full=Ribosomal RNA-processing protein 36 Length=317

GENE ID: 8295512 KLTH0D12892q | KLTH0D12892p
[Lachancea thermotolerans CBS 6340]

Score = 37.0 bits (84), Expect = 1.0
Identities = 39/147 (27%), Positives = 60/147 (41%), Gaps = 27/147 (18%)
Query 358 DMRPEAEGDSEGESEYFPTE-----EEEEEEEDDEVEG-----AeadLSGD----- 398
+++P E D E E+ E T +E+E DDE+ AE L +
Sbjct 7 NLKPGYESDEELEDD EILKTL SKRYADEDESASDDELSSLSFDALRRAERQLEEE SRKEK 66
Query 399 -GTDYELKPLPKGGKRQKKVQEIIDDDFFPSS-----GEEAASVGE GGGGGRKVG R 451
G+ E KP K KR K+ + E+ D F S GE ++++S EG +V R
Sbjct 67 AGSQNESKPA LKNAKRSKEKEL-ELADSFKA KSYTEESFGENS DSSSENEGLFEEEEVVR 125
Query 452 YRDDGDEDYKQLRRWNKLR LQDK EK 478
+ D+ K R + K+K
Sbjct 126 GNKNSKADHGKNRK KSHAPSEQSSKKK 152