Amino acid stylization: a new approach for representing protein sequences and scoring protein alignments

Stefano M. Marino

Akdeniz University,

Department of Biotechnology,

Antalya, Turkey

correspondence

marinost@akdeniz.edu.tr

Synopsis

The amino acids (AA) stylization, introduced in this work, is a mean to obtain a qualitative and quantitative description of different AA; the goal is to allow, via a chemically meaningful representation, the analysis of protein similarity in multiple alignments, without any statistical and evolutionary information involved. The AA Stylization Matrix Method (aSMM) has been tested positively with different data sets, in line with most common statistically based approaches (e.g. BLOSUM62), while allowing the exploitation of some unique capabilities, e.g. to be rapidly configured for specific properties that can be (over-, or under-) weighted. Overall, applying *ad hoc* parameterization aSMM permits a tailored investigation to account for different types of substitutions in protein alignments.

Introduction

Protein sequence similarity searching programs like Blastp or Fasta use scoring matrices to account for differences between residues in two or more aligned proteins [1-4]. Amino-acid changes can range from biochemically conservative, e.g., leucine to valine or glutamic acid to aspartic acid, to dramatically different, e.g., tryptophan to alanine [1]. The earliest amino-acid scoring matrices were based on amino acid properties or genetic code differences; instead, in the last decades amino-acid scoring matrices based on empirical measurements of amino-acid replacement frequencies (from sets of homologous sequences) became more and more used. Particularly, the BLOSUM family of substitution matrices (in particular the BLOSUM62 matrix) is currently the *de facto* standard in protein database searches and sequence alignments (e.g. it is the default choice, in the widely used web version of blastp at ncbi) [5,6]. The primary motivation behind this work comes as a consequence: to date, the work on statistically derived matrices has reached a relatively stable status (and matrices like Blosum62 are expected to remain essential for the foreseeable future) [1,6]; meanwhile, protein alignment scoring based on amino acid properties has been comparatively neglected, and the derivation of an efficient scheme remains an important challenge. In the following text, an approach for amino acid similarity scoring relying solely on amino acid properties description, through an innovative scheme (called amino acid stylization), is presented and discussed.

Method

Amino acids, AA, share similarities and common features: while far from a continuous space (e.g. each amino acid being derived from another, by continuous additions or subtraction of one specific property), an opportune AA stylization scheme could permit a classification based on partial derivations (to represent, with approximation, AA diversity in an effective, albeit not strictly rigorous, manner). This was the goal of this work, and the stylization (with its underlying assumptions) represents the core of the innovation.

During the setup process, I developed and tested several alternative schemes, based on a manually curated, “human expertise” driven investigation. The pillars of the stylization, were:

- 1) side chains are represented as atoms in a grid; the representation is meant to be as fair as possible (but not strict), particularly in respect to size
- 2) functional groups and particularly non carbon atoms are considered of special importance
- 3) the stylization tries to capture the presence of other features, such as charge, aromaticity and carbon branching.

After preliminary analyses (i.e. on a range of alternative stylizations), a specific scheme was selected, as reported in Fig.1.

Figure 1. Amino acid stylization and features representation. First row and first column represent non carbon atoms and physicochemical features, respectively (see notation for Ala; O= oxygen, N=nitrogen, S=sulfur, Q+ = positive net charge, Q- = negative net charge). In the first column and first row (i.e. external layer, for features), a dark gray spot indicates positions with a full weight; Gly (special case) is scored with a default value(1, post assignment).

The example shows that for each amino acid, a 5x5 matrix composed of basically two regions is created: the 4x4 “inner core” region (inner section, from 2nd row and from 2nd column, representing the main non-hydrogen atoms of the side chains) and the “external” region, made

of the first row and the first column (Fig.1, first row includes the weights for non carbon atoms; first column the weights for some chemical/physical features: aromatic, branched, positive or negative charge). In terms of atomic representation, the main rationale of the stylization is: sidechain (non-hydrogen, nonH) atoms are represented in the “inner core” section (and treated equally, regarded of the atom type), while the presence of non-carbon atoms (e.g. oxygen, nitrogen, sulfur) is accounted in the first row; for example, delta position in Methionine is represented with an atom assignment in the inner section (e.g. as if it was a carbon) with a contemporary representation of sulfur in the external section (first row, positive weight for sulfur atom presence); in other words, all non-hydrogen sidechain atoms are treated equally in the inner section, while delegating the differences in functionalization to the external section.

This scheme was meant to have a visual component, and was meant to contain (for future developments) a specific color scheme; however, for the purpose of the present study, a grayscale stylization was developed, that could be directly implemented in matrices (one for each residue type).

As said before, the stylization includes the “inner core” of the matrix, and an “external” region for functional groups and features. Initially, all the non-empty positions in the matrix (squares in Fig. 1) are assigned a numerical value of 1 (i.e. nonH sidechain atoms=1, oxygen=sulfur=nitrogen=1, aromatic=branched=net_charge_positive=net_charge_negative=1). Then, a differential weighting is achieved by multiplying (via entrywise product, denoted with a $\circ$ symbol in the following text) an amino acid specific matrix, e.g. M_i , with a (5x5) weighting array, e.g. M_k . The weighting array can specify for 8 differential weights, for any of the categories described.

After preliminary tests (inclusive of some optimization runs, see also later on the text), the default parameters of the derivation scheme in Fig. 1 were set as follows:

$w(\text{inner_core_nonH_atoms})=1.0$; $w(\text{oxygen})=w(\text{nitrogen})=w(\text{sulfur})=1.5$; $w(\text{aromaticity})=w(\text{branched})=w(\text{positive charge})=w(\text{negative charge})=1.0$

An example of the “weighted” matrix for Serine, M_i , could be:

$$M_{i(i=Ser)} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1.5 & 1.5 & 1.5 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Generally, the score for each position in a two proteins sequence alignment is given by the absolute value of the difference between two matrices ($|Ma - Mb|$, where Ma is the weighted matrix for the amino acid x in sequence a , $Ma=M(x) \circ M_k$, and Mb is the same for sequence b), normalized by the sum of non zero positions (N) in both matrices.

In general, the score for the position i , $S(i)$, can be expressed as

$$S(i) = |[M(x) \circ M_k] - [M(y) \circ M_k]|/N = |Ma-Mb|/N$$

where, for a position i in the alignment, x is the aligned amino acid in sequence a , and y the corresponding amino acid in sequence b , M_k is the weighting matrix (previously defined, as well as the size normalization parameter, N).

Ultimately, the score of the alignment, is the sum of $S(i)$ scores iterated for each aligned position, and divided by the alignment length.

Optimizations were run with all weights made to vary (independently) between 0.5 and 2.0 (with 0.5 intervals); the default scheme (all weights of 1.0, except oxygen, nitrogen, sulfur atoms, with a weight of 1.5,) was deliberated after rounds of optimization (of the weights distribution scheme, in parametric form) with reference datasets (made of externally generated multiple alignments, e.g. pfam and CDD); ultimately, the chosen default scheme was the most balanced in our tests, the one providing consistent performances in all the different scenarios tested. As a reference in the optimization, the program ClustalW was employed, with the scoring function, CW, derived as follows:

$$CW = (1.0 * \text{identities} + 0.5 * \text{similar} + 0.25 * \text{weakly_similar})/L$$

where similar are those positions marked by the symbol (" $:$ ") and weakly_similar are those positions marked by the symbol (" $.$ "), and L is the number of aligned positions.

After having defined its basic structure, the approach was further refined by introducing some "special" cases, for some peculiar amino acids: Gly, Pro, Cys. Indeed, with the stylization method presented, the description of the amino acids can be (with very little work) enriched with "custom" features; for example, in respect to the default scheme (see Fig.1 and related discussion), small variants were introduced (Fig. 2): "special cases" were defined, for (i) Cys (fractional weight, 0.5x, for negative charge, at the light of its "close to neutral pH" pKa; a significant proportion of Cys, particularly if exposed, can exist, more or less transiently, in the

thiolate form, [7]), (ii) Pro (fractional weights, 0.5x, to approximately represent its unique structural features, Fig. 2B); as mentioned in the legend of Fig.1, Gly was already treated as a special case, with an extra matrix value of 1 (so that, for example, Gly - Ala = $|0 - \text{Ala}| + 1$). To be noted, these modifications for the special cases, were introduced *a priori*, i.e. before testing their effect on the performances, mainly because these three adjustments were deemed as more opportune, fitting representations of these peculiar AA types. Two minor customization were also considered for Asn and Gln: to

Fig. 2. Customization for special cases.
Cys and Pro present some fractional (0.5x) weights, depicted in light grey.

account for the "mild" (e.g. if compared with their close relatives, Asp and Glu) functionalization of their (side chain) functional groups, their carboxamide N and O atoms were fractionally weighted (0.5x, e.g. for the oxygen spot in the matrix, Asn is scored half the value of Asp). While biochemically grounded, these modifications provided overall minor effects in our tests (on average, for the tests in the Results section, less than 1% on the final score; naturally, the difference would become more significant, in alignments with unusually high number of Asn

and Gln). For this study I opted not to use them (but it could be considered for future customizations or further refinement studies). Hereinafter, all values, numbers and stats reported in the Results section refers to the inclusion of the customized matrices for Cys and Pro in Fig. 2.

An important mention goes for the treatment of Amino acid identity: by design AA identities are all treated equally (score contribution = 0), in aSMM with default option; this is not the case for most common matrices (e.g. Blosom, PAM)), which factors in also the different abundance (e.g. different probability of identity by chance for different AA types); in aSMM it is possible to introduce a dedicated scoring to account for this; in a preliminary study, I introduced a (very mild) scoring scheme of: -0.1 for Cys-Cys and Trp-Trp identities, -0.05 for Met-Met, His-His, Tyr-Tyr (and 0.0 for all other cases): the analysis revealed that the correlation between aSMM (with separate identity scoring) and CW and Blosom62 had a minor improvement (on average, 1%) if tested with Example 1 (pfam families, Results section). However, it may be important to consider the possibility to employ differential weights for identities in future developments. Last, but not least: gaps are accounted for, by default, with a score of +1. This is a conservative choice; however, gap weighting is an important parameters, that could require adjustments (e.g. more impact, for alignments of similar proteins, with few expected gaps), depending on the user needs. In the following section, all results refers to aSMM with default options (described before) for gaps and identity scoring.

Results

To test the stylization approach introduced in the previous paragraphs I defined some alignments with (i) different protein families, and (ii) designed peptides with variable % id (from 38% to 87%, alignment length of 56 aa).

Overall, the scheme of Fig. 1 (aSMM, with the modifications of Fig.2) returned consistent results, in line with two widespread evaluation methods tested (Clustal W based similarity scoring, CW, and Blosom62 similarity scoring). Moreover, by playing with the parameters, aSMM can change the results in a predictable way, providing a clearer picture in relation to user specific needs, i.e. favoring the conservation of specific functional features, such as aromaticity over electrostatics, or functional groups over size, etc.

Example 1 - Comparative tests with different protein families. We used pre-generated alignments available in pfam (<http://pfam.xfam.org/>), for the following protein families: pfam16674, pfam15620, DUF1679, pfam00246 [8]. For the extended list of sequences employed, see Appendix. By analyzing the correlations between the scores of aSMM and ClustalW, for each alignment in the multiple alignment file, and then evaluating for each family the Pearson correlation (R^2 coefficient) we obtained the following results (in brackets the values for the correlation between aSMM and BLOSUM62): R^2 [pfam16674] = 0.9345 (0.9388), R^2 [pfam15620] = 0.9191 (0.9115), R^2 [DUF1679] = 0.9590 (0.9630), R^2 [pfam00246] = 0.9260 (0.9339). These results highlight the significant agreement of aSMM with the two other methods: this is particularly remarkable at the light of the fact that while both CW and Blosom62 are derived from statistical analysis of amino acid frequency and evolution in natural

proteins, aSMM is straightforwardly derived from a (chemically and structurally inspired) stylization, with no a priori knowledge of amino acid preferences in protein alignments.

Example 2 - A closer look: method-specific differences and features. Considering the following input set of sequences:

```
>p5/1-56
PCTWDDFRGHLGGGNRKLVLIGWDEHEDMRLEFFFGNPLDKWWNYSEFERLLELLA
>p6/1-56
PCTHDKRVGHLGGGGGGGGGGWNEHEDLPLEIFETNAQDKWWNYSDWSRAVELLA
>p9/1-56
CCCCCCCCCGGGGGGGGGGGWEEHEDMRLEIFEGNPLDKWWNYSEFERLLELLA
>p10/1-56
CCCCCCCCCGGGQHRLLIVGWEEHEDMRLEIFEGNPLDKWWNYSEFERLLELLA
```

These sequences were aligned (with clustalO, each pair aligned separately), scored with the three methods (CW, Blosum62, and aSMM), and sorted by similarity, as reported in Fig. 3 (from top to bottom, more similar to less similar; note that for aSMM lower scores stand for higher similarity, while the opposite is true for CW and Blosum62); the results show that, there is an overall consensus (indeed, all three methods individuate the best and the worst alignments); relative distance between different positions, however, was peculiar for each of the three methods, and reflect different takes on the evaluation of protein conservation (e.g. evolution driven for Blosum62, structural and chemical affinity for aSMM). Thus while one can confidently use different methods to detect the bigger trend, more subtle differences may require using different methods.

aSMM			CW			Blosum62		
p9	p10	0.1126	p9	p10	0.8750	p9	p10	5.2679
p5	p10	0.1714	p5	p10	0.7321	p5	p10	3.5536
p6	p9	0.2153	p6	p9	0.6786	p6	p9	3.3214
p5	p9	0.2645	p5	p9	0.6652	p5	p9	3.0000
p6	p5	0.2710	p6	p5	0.6429	p6	p5	2.9464
p10	p6	0.3279	p10	p6	0.5536	p10	p6	2.1964

Figure 3. Detailed view of the scoring and sorting for each pairwise alignment in Example 2, for the three methods tested. Color codes (consistent between different cases in the figure) are used to facilitate the reader; color codes for scores: from green (best) to red (worse).

For what it regards this document, a first major highlight is the ability of such a different and independent approach (aSMM) to provide reliable results (for the general trend, in an overall agreement with Blosum62, but also with CW).

However, a major reason to develop (and to use) the method introduced here is the ability to tailor the analysis, and to customize the weight of different components and features.

An example of this application is reported below.

Example 3 - Effects of weight schemes in aSMM scoring and alignment resolution.
Considering the following input set of sequences:

```
>p6
GGGWNEHEDLPLEIFETNAQDKWWNYSDWSRAVELLATASLIPHFLA
>p7
GAAWNDYDCAPLEIFEVQADERYWNYSKGTCTLA VGSIVVHYILG
>p8
GGSWNEHECLPLEGFETNAQDKVVCSCRWSRAEPLATAYLKHHHIS
```

The scores for each alignment, and for different scoring methods are reported in Table 1. In this case, aSMM is in agreement with Blosum62, while CW is not (however, the latter is in agreement with the scoring by amino acid identity, ID in Table 1); it is important to note, that if other common matrices are considered (for clarity, in Table 1 only a representative set of matrices tested were reported), the trend is a consensus with aSMM and Blosum62, in spite of the fact that amino acid identity is considerably higher for the alignment preferred by CW.

<i>protein 1</i>	<i>protein 2</i>	<i>aSMM</i>	<i>%ID</i>	<i>CW</i>	<i>Blosum62</i>	<i>Risler</i>	<i>structure</i>	<i>PAM30</i>
<i>p8</i>	<i>p6</i>	0.2309	0.6458	0.6771	2.8750	1.3833	2.5625	2.3333
<i>p7</i>	<i>p6</i>	0.1937	0.4792	0.6458	3.0417	1.4875	2.7917	2.4583
<i>p7</i>	<i>p8</i>	0.3408	0.3333	0.5000	1.7500	1.1083	1.4167	0.1458

Table 1. Results of different scoring matrices for Example 3.

This example was designed on purpose to create a “borderline” situation, to highlight that similarity scoring methods (e.g. the reference standard Blosum62, but also the aSMM approach) can overrule decisions made only on amino acid identity. Furthermore, the aSMM scores reported here refer to the default scheme of parameters; with a different weighting scheme, e.g. in this case favoring aromaticity and net charge, the relative scorings in Table 1 can become more separated: for example, with relative weights of 2.0 (for aromaticity and for both positive and negative charge) the scores become: p8-p6 = 0.2728, p7-p6 = 0.2073, p7-p8 = 0.3925 (which represents a +35% increase in the average difference between the best alignment, p7-p6, and the other two; the difference can increase much further, e.g. up to +95%, with higher relative weights for aromaticity and charge, e.g. weight = 5.0). Thus, the aSMM scoring allows for the possibility to enhance, specifically for a chosen set of features, the relative scoring (indeed, with the weighting scheme discussed above, aSMM allows for the best separation in the example, in respect to any other method in Table 1), and this is a characteristic feature of the approach presented here, which can be useful in several situations, where comparisons of differential scoring can highlight specific properties of the alignments. Naturally, the user should be aware of possible consequences of (heavily) biasing

the scoring system (i.e. with “strange” weighting schemes, alignments can deviate significantly from those obtained with the default scheme).

Discussion

The amino acid stylization method (aSMM) presented here is an original and orthogonal addition to current matrix based methods for protein alignment scoring. It has some unique features, which are well suitable for a range of present and future applications:

(i) evolutionary independent analysis of sequence similarity - The scheme is derived without *a priori* knowledge of amino acid distribution or substitution preferences (i.e. independent from biological or evolutionary observations), and thus it may prove particularly beneficial in studies where the evolutionary component is not critical, e.g. protein engineering and design of protein variants;

(ii) differential conservation of physicochemical properties in alignments - With its “2-shell” information layers (“inner core” and “external” regions) it was designed to make the customization (i.e. to favor one or more features over the others) of the similarity search powerful yet straightforward; due to its theoretical form, the customization is achievable by direct tuning of the weighting matrix (without any modification of the amino acid description); consequently, aSMM allows the sampling of a broad and diverse search space, to score alignments differentially, depending on (a chosen sets of) different properties; a practical application would be to score alignments with different parameterization schemes: for example, first scoring with the default scheme; and then with *(i)* a scheme favoring size, i.e. higher side chain atoms weight, and with *(ii)* a scheme favoring functional groups and electrostatics; afterwards, the comparison of the outputs could highlight differentially conserved properties, e.g. preferential preservation of electrostatic properties; with similar experiments, the user could be able to detect particular trends (conservation of some properties, over others) in protein alignments;

(iii) customization and extension to non standard amino acids - The description of the amino acids can be enriched with new features; indeed, in respect to the basic scheme (Fig 1), some small variants were introduced for Cys, Pro (Fig. 2) in our default scheme (see discussion in the Method section). Similarly, modifications can be introduced to account for non standard amino acids; a simple example could be the most common, naturally occurring, non standard amino acid, selenocysteine; this residue can easily be handled within the aSMM framework (e.g. by representing it as a sort of “super cysteine”, with selenium scoring in the first row, 4th position, with a “higher than sulfur” weight; and the low pKa - around 5, compared to a reference value of around 8 for Cys - reflected in the first column, last position, with a higher fractional weight, e.g. 0.75); in general, many modified amino acids with an overall conserved structure (i.e. similar to a natural type), such as Selenocysteine or Acetyl Phenylalanine, can be treated within the current aSMM framework with few, appropriate modifications;

(iv) a fundamental step toward Computer Vision based applications - aSMM was designed specifically to be visually informative, for future utilization with computer vision; indeed, it was developed with this future application specifically in mind (for example, by using specific color codes, data not shown), whereas the sequence and/or the alignments could be read and interpreted automatically (e.g. by an artificially intelligent, computer-vision based, sequence

analyzer); the advantages over the simplest analytical approach (i.e. employed here) could be, in time, a winning feature of the Amino Acid Stylization approach presented here.

Image recognition and amino acid stylization can also pave the way to an evolved scoring system, that is sensitive to the surroundings (rather than strictly position specific) and where similarities are scored beyond the isolated value for the *i*-th position, so that to provide a more “fluid” evaluation of protein similarities and of protein/sequence properties.

The work presented is an introduction and a first description of a new approach for protein sequence similarity scoring; it describes a first step into AA stylization approaches, that could evolve in alternative schemes (i.e. also with significant differences from schemes in Fig. 1 and 2), while maintaining the general approach described here. A particular aim of this author is to develop it further as an innovative tool for synthetic biology and protein design, with (i) improved representations (that can accommodate some relevant “non standard” features) for non canonical amino acids and (ii) with employment of computer vision (i.e. with *ad hoc* color codes for different features) and artificial intelligence to enhance and extend the reach of the similarity search.

References

1. Pearson WR. Selecting the Right Similarity-Scoring Matrix. *Curr Protoc Bioinformatics*. 2013; 43:3.5.1-9.
2. Blast: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990 Oct 5;215(3):403-10.
4. Lipman DJ, Pearson WR. Rapid and sensitive protein similarity searches. *Science*. 1985 Mar 22;227(4693):1435-41.
5. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*. 1992 Nov 15;89(22):10915-9
6. Styczynski MP, Jensen KL, Rigoutsos I, Stephanopoulos G. BLOSUM62 miscalculations improve search performance. *Nat Biotechnol*. 2008 Mar;26(3):274-5.
7. Marino SM, Gladyshev VN. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *J Mol Biol*. 2010 Dec 17;404(5):902-16
8. Pfam: <http://pfam.xfam.org/>

Appendix

Below are the sequences of the pfam alignments used in *Example 1* (note: the employed sequences are a subset of the full pfam alignments, i.e. since the full alignments can be very large and with long stretches of gaps, which would be detrimental to this type of analysis); different families are separated by a # symbol, followed by pfam ID..

```
#pf15620
>P49452.2|CENPC_MOUSE/264-906
KAHPIEEAQLSVEENPATTC--TDELENDCRSPENKMQSETAKTPPAWERTTKQS-QRRVSKPKAAEE-LRKGQSSWENSNV--SNTGQ-DK
LQ-INSKRNMKDCEEVRNEPNPKKQKPAL-ENKKKTNST
>lcl|18_pfamImport/1-265
KAHPVGRSQPSEQKGVNKGCalSNEMENHCRSTKYELYENAEKSSGNRRRTVRQK-QRRKFKANVLEEQVDTEQSKDKNTNM--SHITQ-DK
LQ-RNSDRNMEECENTINVAISKQVLPV-GRKKNSTSV
```

```
>lcl|7_pfamImport/1-264
TAYPVEKSPSEQRRLGRSCalTDELENRDRSTKYETNYENAEKSSGKRTIKQK-QRRKFKDNVLEEQIDMEQSKNKNINM--SHITQ-GK
LQ-RNSDRNMEEYEEMINVDIS-KQMLPV-GNKKNSTNG
>lcl|3_pfamImport/1-263
TAYAVETSPPEQRRLGRSCasTNELENDYRSTKHEMYENAEKSSGKKTIQQQ-ERGKFKANILEEQVDMEQSKDNNINV--SHITQ-DK
LQ-RNSDRNMEECEKVINVGISKQMLPV-GSKKNSTNV
>lcl|6_pfamImport/1-258
KAHQVEKSPSEQRILCRSCtlTDEMENNYRSTYEYVYSKNARKSSGNKRTVNQN-QRRKFKANVVEEQVDMEHSKDKNINM--SHVAQ-DK
LQ-RNSGSNMKECKEMSNVHISKQMPV-GNEKSSTN-
>lcl|4_pfamImport/1-263
KAHPVEKSPPEQNIILGGSCasTDEMENNCKSAKYEMYSENGKSSGNKRTIKQK-QGTQFKANVVEEQVDMEHTKDKNVNI--SHIAQ-DK
LQ-RNSGQNMEEREI-GVQISKKEMPPM-GSKRGSTNI
>lcl|2_pfamImport/1-263
KAHPVEKSPSEQKILGGSCalSDDVENNYKSTKYEIHSENAKSSGKRTIKQN-QRTKSKANVPEEQVDVRQSKDTNINI--SHTVQ-DK
LQ-RNSDKNMEEGEEI-SVNISNNQLLSV-GSKKGSTNV
>lcl|8_pfamImport/1-249
KAHPLEKSPSQDRKLGRNCalaDEMEN--RSTKTEMYSKNAEKSSGNKRTTQK-QKKKLKVNE-----DMKQGKDKNINM--SPIAK-DK
SQ-RNSDRSMKACKEKRNHISKRVPPV-ANGKRSGan
>lcl|11_pfamImport/1-263
KAHPVEKPPPEFGKLGKSCasASEMENNCKSTKDALYSENAEKPSSENKGTIKQN-RRRKVKAKVVEEHLDMEQPKDKNINT--SFTVQ-GK
LQ-RNSDKNTEVCEEMRNDHISPKQIFFILGKKKSRTHV
>lcl|15_pfamImport/1-250
KAQPLEKSPSQGERLGTSC-----MENDSRSTKREIHSQD-KKSSGSKKAIEQK-QTRKLKVNVEEQVLMHMQSDDENLNT--SHIAQ-NK
FQ-RNSLKSMECEKMRNDHVSKKQMPV-ESKKSSS--
>lcl|17_pfamImport/1-254
GAHPVEKSPSQDENLCTNCaltDEMENSRCSTKYKMCSENAEKSPGNKRTVKQK-QRTKSKDSIDKDqLDVGQSKDGN--M--SHIAQ-KK
LQ-RNSDRNTECEEKRNRRISKQMPSM-GSKNST---
>lcl|12_pfamImport/1-260
K1-----SPEMGNASKSTRHTLYSEETEQFFKKKTVMQH-QSENCMDQNGVYkHKVGKHKQKRFNV--SQTNQ-DK
LQ-SASDRSLGQEEERAESCHSTKRKPSA-AASQQIQKK
>lcl|5_pfamImport/1-256
KARAVEKSPSKRKKLDs--LSNEMGSTSKSTRHKIYSEKSEQFLENKGAILRKcQKKDYKPKQVSVKhHGTEQTKKKRSNM--SQRNQ-DK
LH-RESDRILGQKERAESCHSAEQILSA-DGHLKKTs-
>lcl|9_pfamImport/1-257
KPHPVETSQPSDKTVLDTsYaltGETENNYRSTKYEMYSKNAEKPSRSKRTIKQK-QRRKFMAKPAEEqLDVGQSKDENIHT--SHITQ-DE
FQ-RNSDRNMEEEREEMGNDCVSKQMPV-GSKKSST--
>lcl|10_pfamImport/1-257
GSHPVEKSLPCDDKKQSMSCaltDEVDDNSRSTKCEHSHKNKGSSGNKTKVKQK-KKKKYKANVVEEQVLMHMQSKEENINAhkKHIAP--K
LQ-RSSSRNMKDCEEMRNDHIS-KMVPV-GSKQKKSTK
>lcl|1_pfamImport/1-251
KTGPVEEAQLSVEEKLGTTL--TNEVENDCRSTENKTHSENAKPSARKRTVKQK-QRRPSKPNVAQE-LSMGQNETENRNM--SKIGQ-DK
LQ-INSKRNMEDHEEVRNEPTPKKPPVPAI-GNKKEKDST
>lcl|19_pfamImport/1-266
TAYPVEKSPSKQKRLGKCalANELENHCGSTKYEMYNENAEKSSGKKTIQK-QRRKYKGNVLEEQVDVAQSKEKNTNM--SHITQgDN
TQgKNISRNRYEIEEADSEVYSIKQCEKL-GSKKNSTNV
>lcl|16_pfamImport/1-260
KAHPVEKSPSEQRILCRSCtlTDEMENNYRSTKYEISSENVKPSSENKRPIQN-QRRKFKANVVEEQVLDLEQSKDKNVNM--SHSAE-DK
LQ-KNLHRNGEDCEGTRNDNIFKTPV-GDKKINRG
>lcl|14_pfamImport/1-235
KTHPEEKSEPSCEQKRSINCTvIDKLENSHRSIK--MYSE-----KFKTNVVEEQVLDTGQSEEN--I--SDITQ-DK
IP-RNSERTIEDCEKMRNDHSATKGMPPV-ENMKTKCRT
>lcl|13_pfamImport/1-254
KAHPVEKSEPLNEKKPSTSCtlTDELESNCRSIK--MYSEKAEKPSGNKRTTKKK-QRRKLKTSVVEEQPDTGQTKDEN--M--SHIAQ-DK
LP-RNSDRITEGCEEMRNGHNAKKEMPPV-ENMKTKCST
#pf16674
>gi|356624665|pdb|3U12|A/1-141
QTGITKWKESFEIVEKENKVSIVVHYNTGG-IpR-I-FQLSHNIKNVVLRPSGAKQS-RLTLQDNSFLSIDKVPKDAEERLFLDAVHQ
```

>gi|545559401|ref|XP_005641887.1|/1-924
KKGMSSSREAYIETVEGKKKIRLVVYFTNGE-Y-L-T-FQLSNNIKNVVLSYGGKQN-HLHLTFQNNNFLFIERLSSKDAKELKMFLDKVH
Q

>gi|395848787|ref|XP_003797026.1|/1-1038
KSGMSKSKEVYIETVEGKKKIKLVLYFNNGQ-Y-K-I-FRLSNNIKNVVLSYGEKQN-HLHLTFQNNNFLFIDKLSSSDAEQLKFLDRVR
E

>gi|545886809|ref|XP_005673983.1|/1-904
KTGMSKSKEAFIQTVERKSKVRLVVYFRTGE-H-T-T-FRLSNNIKSVVLSYGGKQN-HVHLTFQDNSFLFIEKLSSRDAENLKAFLDRVH
E

>gi|558099407|ref|XP_006082382.1|/1-917
DTGMSKLKEAFIETVERKKKARLVIRFSTGE-Y-T-T-FQLGNNIKNIVFRSYGENQN-YLHLTFQNTSFLFIRKLSSRDAKNLTMFLDRVH
Q

>gi|664752888|ref|XP_008533903.1|/1-960
KTGITKSREVFIEETVEGKKKETSLLVVFNSGEFI-R-V-FQLSNNITNVVLRHCGERQS-CLNLTLKNNSSLFIDKLSSRDAEQLKMFLDIVH
Q

>gi|410989455|ref|XP_004000977.1|/1-920
NVGMSKPKEAYIGKVERKKNIRLLVYFSTGE-S-T-T-FQLTNNIKNVVLSYGGKQN-HLHLTFQNNNSSLFIEGLSSADAELKMFLDGF
Q

>gi|351699529|gb|EHB02448.1|/1-913
QTGITKWKEGSFEIIVEKENKVSLLVHYNTGG-IpR-T-FQLSHNIKNVVLRLPSGAKQS-RLMLTLQDNSFLSIDKVPSKDAEEMKFLDAVH
Q

>gi|297711051|ref|XP_002832167.1|/1-915
KTGMSKSKEAFIEAVEGEKKDRLLVLYFKSGK-Y-N-T-FQLSDNIQNVVLSYRGQN-YLHLTLQNNNVLFIEGLSSTDAEQLKIFLDRVH
Q

>gi|291408229|ref|XP_002720440.1|/1-907
NSGMSKSKEAFIETVPGRKKDKLVIFYNAEK-M-R-T-FQLTDNIKSVVLSRREDDQN-HLHLTFQNNNSSLFIERLSSSDAQELKLYLDRVH
Q

>gi|760997666|ref|NP_001292064.1|/1-988
QTGITKWKEGSFEIIVEKDNKVSLLVHYNVGG-IpK-T-FQLSHNIKSVTLRPNGRKLC-CLMLTLKDTSFLLTIDKVPLKDNEMRMVYLDVH
Q

>gi|511922818|ref|XP_004780578.1|/1-923
KMGMSRTKEAYIETVERKKRIKLVVCFSTGE-S-V-T-FQL-NNIKNVVFRSYGEKQN-HLRLIFQNNNSSLFIMERLYFRDAVQLKMFLDRVH
Q

>gi|532104335|ref|XP_005337761.1|/1-932
KTGMSKSTAAFIEPVEGKKKVKMLLYFNTGQ-V-K-T-FQLSDNIKSVAFRAYGQEN-HLHLTFQNNDFLVERLSSTHARELKTLLDRVH
Q

>gi|344298458|ref|XP_003420909.1|/1-911
KTGMTKSKEAFIEIIVEGKKKDKLVIHFSAEvI-K-A-FQLTNNIRNIVFRFFGGNQN-HFHVTFQNNNSSLFIDRLSYSDAEQLKMFLDK-H
Q

>gi|803310102|ref|XP_012013420.1|/1-904
KNGMSKSKEAVVEVSGKKKTKLVICFCNGE-S-R-T-FELHNNIKNVVRSYGEKVN-HLHLTFQNNNSSLFMEKISSKDVKKLKAFLDGVH
Q

>gi|958757013|ref|XP_014713938.1|/1-917
NIRMSKTKEAYIRTVEEKKKVILVVSFNTGE-Y-T-T-FQLV-NIKNVVFRSYGENEN-QLNLIFQNDLFLFIEKLSSLRDGKELKMFLSF
K

>gi|565305610|gb|ETE60666.1|/1-120
QTGTTKWKEGTFEIVDKDSKVILLVHYNAGG-IpK-I-FPLNHNINNVVVRQTKGKLY-CLMITLKDSSVLTLDKIPSKDANEMKLYLEAIQ
N

>gi|351698815|gb|EHB01734.1|/1-750
RTGMSEFPQGVFIEAVKGRKNFKLLIYLRTRGR-V-K-T-LKLNNNIKSIVLQSYGANLS-SLHVTFQNDGFLFVEKLSSTEAKQLKAFLDSVH
Q

>gi|296236438|ref|XP_002763324.1|/1-915
RTRMSKSKEALIGAMQEQNKRLVLYFKNGK-Y-V-T-FLLTDSIRSVVFRSYRGKN-HLHLSLKSNNKYLFEGLSSTDAEILKIFLDGFH
Q

>gi|344269718|ref|XP_003406695.1|/1-932
GTGMTNLKEALIQMVEGEKESRLVVIFNSGR-StR-V-FQLSNNITSMVLRHSREEQS-SLHLTFQKNCSLFIDKLSCGDAEKLKKFLDRVH
Q

```
>gi|395845350|ref|XP_003795403.1|/1-927
NTGMTKLKEAFIETVEEKKEIKLVILSKSGG-ItK-V-FRLSNNIRSVVLQHCEKRKS-HLCLTLQGN-ILFIGKLSFQDAEQMKMFLDVVR
Q
>gi|637247970|ref|XP_008108424.1|/1-968
QTGTTKWKEGTFEIVDKDSKVTLVHYNAGG-IpK-M-FQLNRNIKVVVQPNKGKLC-RLMITLKDSSFLMVDKIPLKDAEEMKLFLEGVQ
N
>gi|348553806|ref|XP_003462717.1|/1-939
RSGMSKTQNVFLEAVKGRKNFKLLIYFKTGK-V-K-T-FHLNNNIRSASFTSYGEDFT-HLHLVFRNNGFLIVERLASTDAEQLKVFLASVP
Q
>gi|556963968|ref|XP_005991835.1|/1-916
QTGTTKWKEGRFEIVEKDNANLMVHYNAGG-SpK-T-FQLNQNIKRVILQMSGAKQrRLLLMLKNSSSISIDKVPHKDAEELKQYLETLQ
Q
>gi|847165792|ref|XP_012826369.1|/1-953
KVGTSKWKEGCCEVIEKDNKYCLVVNYNAGG-M-P-TrFQLNQNIKTIVVKPNGGTQS-RFMLTLKDASSLTIDKVPLKDAENLKHFLSLED
K
>gi|344240562|gb|EGV96665.1|/1-535
HTRASQWKEAVIEIVERKQKVTMVFSFKLEK-R-KrV-FQLGNNVAGVVVSCCEMGLH-CLYLTTLKGDTSLLIDKLSSADIEQLKAFLDLAH
L
>gi|26332815|dbj|BAC30125.1|/1-198
HTRASQWKEAVIEIVERKQKVNLVVFSFKLEErR-R-V-FQLGDNVTGVVV-SGELGLY-HLDLTLRDDTSLLIDKLSSADVEHLKSFLDSst
P
>gi|562888237|ref|XP_006171489.1|/1-912
NGGMSKSEEACIEIEEKEKSKLSLSFSAGK-C-N-T-FQLNNNIESIVLSRYEEKRH-QLYLTFRNKEFLFIERLSFTDAKDKMFLDRIH
Q
>gi|973169364|ref|XP_015214688.1|/1-944
QLGTTKWKEGSFEIHEKDNKVNLFVRFNSGG-VpK-I-FQLNQNVKTVSLGP-KQNCQ-RIQVVLKDSSSLIIDKVPLLLAQKFRDYLEKLEK
Q
>gi|542255586|ref|XP_005463419.1|/1-959
DLGTTRWKEGTFEILEKDNKVNLCRFNCGGaS-K-T-FQLNQNVKNIS-----QNVN-RIVLTLKDNVSVITLDRMPQPLVQKTKEYLEKLEK
Q
>gi|688569768|ref|XP_009302661.1|/1-934
ELGTTKWREGVIEIQERDNKINLLVKFNSSGG-APR-V-FQLSHNVKSVSWFQ-THGPN-RMTLTLKDSCIVMMDKLNMLVAKKMKEYLETVK
L
>gi|992224821|ref|XP_015464014.1|/1-923
DVGTSRWKEGTFEIIYERDNKSNLVLFKFNSSGG-TpK-K-FQLSHNVKNVGYTQ-THSLT-RLALTLKDSSVVTFEKLSIMVAKKLDYLETIK
Q
>gi|410925226|ref|XP_003976082.1|/1-797
DLGTTKWREGTFEILEKDNKISLCRFNCGT-L-K-I-FQLNQNLKT-----FSQNL-SRIQLNLKDSSVIVLDKVPQMLVQRTYEYLKLEK
Q
>gi|1040126311|ref|XP_017194005.1|/1-850
RTGSSKLKEPITETVEREKDIQLIFMFQGGK-----PIDK-----RHYGERQN-GLYLTFWKKN-LCLGKLSRTHTERLQILLDIIH
Q
>gi|821001087|ref|XP_003277323.3|/1-912
RTGMTKLKEALIEIVQRKKETKLVVTFKSGQ-FiR-V-FQLSNNIRSVVLRHCKNRQS-HLHLTLKNNVFLFIDKLSYRDAKQLNMFLDIIQ
Q
>gi|342187096|sp|Q99MX1.2|UBP26_MOUSE/1-835
KAGMSKSRNALIETCVGKREVKLILYFSTGK-I-K-T-LQLHDNIKSVVLQTYGEDQN-YLHLTFKNNDFLFVEKLTTTDARRLRFLDKTS
Q
#pf00246
>p0gi|524933203|pdb|4BD9|A/39-122
GHSFENRPMYVLKFST-----GKGVR-RPAVWLNAGIHSREWISQATAIWTARKIVSDYQR-DP-----AITSILEKMDIFLLPVANPDGY
VYTQT
>p1gi|81716926|sp|Q829W4|Q829W4_STRAW/1-88
GKTLNGQDILALKLTK----GAKKTKDGAKPSALYMSNQHAREWITPEMTRRLMHYYLDNYSK-DK----RIKKIVDSTELWFLSANPDGY
DYTFQ
>p2gi|81671003|sp|P73902|P73902_SYNY3/38-124
```

GQSYAGRDIWVAIATN----QATGDY-RNKPGYWIDANTHAGEVTGSAVALYCLHQLMTKYGE-DA----QITHLLDHYTVVYVLPRLAMDGA
 EKYLT
 >p3gi|81795132|sp|Q8PAS3|Q8PAS3_XANCP/65-151
 GTTPEGRPMKALVSTSGALDAASAAQRALPVVLVQGGIHAGEIDGKDAGFLALRELLDGTAG-----KEVLDKVVWIFVPVFNVDGH
 ERFG-
 >p4gi|81746867|sp|Q8ESH0|Q8ESH0_OCEIH/1-87
 GTSVEGRPIHLVRVGF--PEPPSDEDIANGNNMLVIGTQHGENAAPREMALELLRDLA--FTE-DE----ELVELLHDTTILFIPTANPDGR
 EADT-
 >p5gi|75542981|sp|Q897X5|Q897X5_CLOTE/1-85
 GKSVLGRELYYKLGN-----GPNKVFYNGAAHSHLEWITPPLMKFAENFIRSYAFGEEMRGYNIGDIWSRSTIYIVPMVNPBGY
 DLVLN
 >p6gi|74800085|sp|Q7PRP0|Q7PRP0_ANOGA/43-129
 GKSAQGRDLWVLVSA----S-PYEHMLGKPDVKYIGNIHGNEAVGRELLLLHIQYLVSSYGS-DP----YIKWLLDNTRIHLPSLNPDGY
 AASKE
 >p7gi|74803707|sp|Q7QC23|Q7QC23_ANOGA/1-87
 GKSVQGRDLWVMEVTE----Q-PGQHAPGKPEVKYIANMHGNEVVGRELLLLFATYLCENYNR-TQ----RITRLLNRTRLHLLFSMNPBGY
 ELADI
 >p8gi|13124695|sp|P42787.2|CBPD_DROME/1-87
 GKSVQGRDLWVLEIFA----T-PGSHVPGVPEFKYVANMHGNEVVGKELLLILTXYMLERYGN-DD----RITKLVNGTRMHFLYSMNPBGY
 EISIE
 >p9gi|62510319|sp|Q9JHW1.2|CBPD_RAT/1-87
 GKSVESRELYVMEISD----N-PGVHEPGEPEFKYIGNMHGNEVVGRELLLNIEYLCKNFGT-DP----EVTDLVRSTRIHLMPSMNPBGY
 EKSQE
 >p10gi|74961599|sp|P91359|P91359_CAEEL/1-87
 GKSVEGRELWVLIISD----K-PKEHKLMEPELKIVGNMHGNEVVGREAVLYLAEIILCLNYGK-NK----YLTDLVNNARIHLMPSMNPBGY
 EKGFP
 >p11gi|74966365|sp|Q22825|Q22825_CAEEL/1-87
 GQSVQGRELWVLVSR-----Y-PIEHRKLIPEFKYVANMHGNEVTGRVFLVSLAHTLLENYNS-NL----WIRQLVDSTRIHLMPSMNPBGY
 EHASE
 >p12gi|81884047|sp|Q640N1.1|AEBP1_MOUSE/1-88
 GKSSRGLKIYAMEISD----N-PGDHELGEPEFRYTAGIHGNEVLGRELLLLMQYLCQEYRDGNP----RVRNLVQDTRIHLVPSLNPDGY
 EVAAQ
 >p13gi|161783814|sp|P04836.2|CBPE_BOVIN/1-88
 GRSFEGRELLVLEISD----N-PGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCNEYQKGNE----TIVQLIHNTRIHIHIMPSLNPDGF
 EKAAS
 >p14gi|146386938|pdb|2NSM|A/33-120
 GRSVEGRHLYVLEFSD----H-PGIHEPLEPEVKYVGNMHGNEALGRELMLQLSEFLCEEFRNRNQ----RIVQLIQDTRIHLPSMNPBGY
 EVAAA
 >p15gi|74957758|sp|O17754|O17754_CAEEL/67-154
 GQSVGRPLVVIQFST----T-PGEHIPTKPEVKLIGNMHGNEPIGRELLLRFAETLCNGAINNDK----EIVQLLNSTSIHILPSMNPBGY
 ELALG
 >p16gi|57012713|sp|Q90240.1|CBPD_ANAPL/1-87
 GQSVQFRQIWSLEISN----K-PNHSEPEEPKIRFVAGIHGNAPVGTTELLALAEFLCMNYKK-NS----AVTKLIDRTRIVIVPSLNPDGR
 EIAQE
 >p17gi|74803707|sp|Q7QC23|Q7QC23_ANOGA/37-124
 GQSREGVPLSVLEIRP----NVNRPRPLMPMFKYVGNMHGDETVGRELLLYLAQYLLSNYGR-DP----EVSALVNETAIYLMPTMNPBGY
 ERSKE
 >p18gi|13124695|sp|P42787.2|CBPD_DROME/67-153
 GRSLEGRNLLALQISR----N-TRSRNLLTPPVKYIANMHGDETVGRQLLVYMAQYLLGNHER-IS----DLGQLVNSTDIYLVPTMNPBGY
 ALSQE
 >p19gi|74870680|sp|Q9VS64|Q9VS64_DROME/1-82
 GRSNEGRHIAALSISL-----NSRVRPRRVAYIQAATHGREWITTQTVLYLAYELLSNLRA-FT-----RVLQDVEIFLVPLVNPBGY
 EYTHT
 >p20gi|74870686|sp|Q9VS80|Q9VS80_DROME/65-146
 ARTYENRALKMAIITN-----GDGRPGKRVIFLDAALHSREWMTFAAALLTIHKLVEFAE-NS-----DLLTDYDWHIMPLANPDGY
 EYSRN
 >p21gi|74866546|sp|Q961J8|Q961J8_DROME/1-82

GKSYENRDIKTITITN-----GDGKTGKNVFLDAGIHAREWIAHAGALYVIHQLVENFAA-NS-----ELLKDFDWVILPVVNPBGY
EYSHT
>p22gi|74948450|sp|Q9VS67|Q9VS67_DROME/1-82
GRSYEGRWLKTITITN-----GDARRKNVILVDGGFHAREWISPAATYILINQLVYNLED-NA-----DLLLLDFDWVILPVVNPBGY
EYQTL
>p23gi|74870682|sp|Q9VS66|Q9VS66_DROME/1-82
GWSYEQRVLKTITITN-----GDGKANKKVI FMDGGFHAREWISPAALYVIDQLVEQFEE-NA-----HLLKDYDWVILPLVNADGY
EHTQT
>p24gi|74872637|sp|Q9W475|Q9W475_DROME/1-80
GYSVEKRPLKILKISN-----GNARNPGIWI DGGMHAREWISPATVITYIANQLVEGWED-LP-----EHMRSINWYIHPVANPDGY
EYSHT
>p25gi|74872640|sp|Q9W478|Q9W478_DROME/1-80
GQTAEGRDLKVLRISE-----NPRENKKVWIDGGI HAREWISPATVTFILYQLMSDWEN-QP-----AHIRGLTWYIMPVMNPBGY
EYSRT
>p26gi|350854883|emb|CAZ36103.2|/1-77
GYTAEGRPMKGVMISS-----DSTRPIIWI DAGIHAREWIA PATALSINKLMRPSGQ-----TLLKHFFQFFIVPVVNPBGY
EYTRT
>p27gi|11514604|pdb|1QMU|A/37-123
GKSVLELRELYVMEISD----N-PGIHEAGEPEFKYIGNMHGNEVVGRELLLNIEYLCNFGT-DP----EVTDLVQSTRIHIMPSMNPBGY
EKSQE
>p28gi|884981810|ref|XP_012987561.1|/61-148
GESFEGRELLVLEMSD----N-PGTHEPGEPEFKYIGNMHGNEAVGRELIVYLAQYLCNQYQQGNE----SIIELIHTTRIHLMPSMNPBGY
EKAAS
>p29gi|767961614|ref|XP_011537587.1|/1-88
GKSHQGLKLYAVEISD----H-PGEHEVGEPEFHYIAGAHGNEVLGRELLLLLVQFVCQEYLARNA----RIVHLVEETRIHVLPSLNPBGY
EKAYE
>p30gi|930428091|ref|WP_054191037.1|/1-85
GKSVKDRKLMCKLTS-----SANTGKKPRVLYTSTMHGDETTGYVLLRLIDHLLSNYES-DP----RIKNILDKTEVWICPLTNPDGA
YRAGN
>p31gi|490064710|ref|WP_003966921.1|/1-86
GKTYQGRDVIAVKVSD-----NVATDEAEPEVLFTAQHAREHLTVEMALYLLRELQGGYGS-DS----RITQAVNGRELWIVPDMNPBGY
EYDIA
#DUF1679
>Q9U368_CAEEL/112-206
WAFENEAQQLHNREVNLKYITEKWNKNETMSPKIYFYKKFD---AENKTKGILGMEFVSDVTIRHLYCNAKPYELHPVLRLSLATLQAGSL
HLT---EDE
>Q20584_CAEEL/135-230
SDEFISFLPPCHNSENYFYKYIQL--PKLDILPDFYFGNQIElekDGNYSKGCIAIELVEDIKTSLPLENFSDDQMLQVLDALAKLQVQFI
NLS---EDK
>Q9XUP2_CAEEL/108-202
WAMFENEAQHLHNREVNLYVLAEKWNKPEELLNAKIFFSKKFD---SENKLGKFLGMEYVDDVTIRHLYCNLKPYPYELHPVLKAVQLQAESL
HLS---DEE
>Q9XUP7_CAEEL/109-203
MALFENESRLFHNREVNLKITKKWNKNDELLSPKIYFYKKFD---SENQTKGILGMEAVDDVTVRHLYYNVKPFELHSLVLSIARLQAESL
HLT---DQE
>Q9XUP8_CAEEL/110-204
WAMFEREAQNVHNREVNLRYITEKWNKNDALLSPKIYFYKKFD---CENKTQGVLGMEYVDDAVVRHLYCNAKPHLHPILQSLATLQAGSL
HLT---EDE
>O61882_CAEEL/110-199
HAHFEKSCQKGNHLEVEFCEAFGHLe---GLLLPKVFFSQKFE---EDNPNKGFVGMFVSGSVVRHCYENVTVDELQPIKALARLQALSL
S-----TES
>Q8IA57_CAEEL/110-206
YAYFESSCKMHNQEMNFYEVAGKFn-SKTL LIPKVYFYTKLD---EKNSNKGFIGMEYVEGSIVRHSYDTCTIEEIQPILRAIAKLQALSL
QNPaeiSKD
>Q19034_CAEEL/103-195
LKTFGNVTRECHNREVAAYKMLIKFn-NSDIPFTKVYHLKPFD---DANDLKGFMIMEFIPNVHSIPMYEAI PADDLISIVRG IATFAALGE
T-S---EGD


```
>Q9N5R6_CAEEL/103-196
LKFLGKFIREGHNREVEAYKLEKLn-HPDIPHTKAYYLKPFK---DKFDLKGFMILDFVPNVHPMPMYESIPADDLISLVRGVATFAGHGE
SLT---AEQ
>P91071_CAEEL/121-214
LKYLAkILRDAHNREIETyKLEKFn-HANIPYTKIYGLKPFY---DENDLKGYIILEYIPNIHTTSMSENIPADDLISTIRAVATFGALGA
CLP---ADQ
>Q19030_CAEEL/103-196
LKKFSKLTRKSHNIEVETyKVLTKFn-HPDIPYTKVYSLKPFN---GEDDLKGYLITDFIPNVHVIEAYKSIPADNIAATIRGIATFSALAE
HLE---REE
>Q9U3D0_CAEEL/103-195
MKSMTSLVKELHNREVDMyRIIMREk--PACPTVNVLSLEAFT---ELSPLKAYIISEYIPNLHHVGMNDcISIEEiWAVVDGIAAFsAMGE
SMS---EDE
>Q20974_CAEEL/104-197
LKGMGEVTRLLHNREVATyKILMREk-HPKIPFTKVYASKPFD---DENKLKAYLISEYYPNIHHIGMHESIPAE DLIPVIHAIAAFSAIGM
KLS---EEE
>XP_002636148.1|/137-230
AEHFLISILAMHNQEI NFYEILSSAn-AENILTPKVYFSQKFD---ETNRCKGFIGMEFLEGIGIRHSYENCTVQEFQPILKLIASIQALSF
SVP---NED
>XP_002647572.1|/111-204
NANFEISIQRiHNQEVDFYNVLKKGe-PENLLILKVYFTSKFD---ENNKTGKFIGMEYIEGSETRHSYENCTVKEIQPVLKAIaHIQALTv
STP---EEE
>XP_002636400.1|/103-196
LKGMGDMTRMLHNREVSTyKILMREk-HPRIPTKIYALKPFD---DENTLKAYIISEYIPDLHHIGMHESISAEDLLPIVHAIAAFSAFGM
KLS---EEE
>XP_002635271.1|/102-194
MKSMTSHVKELHNREVHMYKILGREq--PDCPTVNVLALEEFT---EYTPLKAYIISEYIPNLHDVAMSDNVPLEDIWAVVDGIAAFsAMGE
SMS---DEE
>XP_002637539.1|/103-196
LKKLSALTRECHNREVDVYKILMMFn-HPDIPYTKVYSLKSFD---DENDLKGYMVIEFISNIHQIEMFRSIPADDLFPLVRGIATFAALSE
SLS---PEE
>XP_002637540.1|/103-196
LKHLSKMTRECHNREVKAYQLLQKFn-HPNIPYTKVYSLKSFE---DENHLKGYMAIEFVSNiHSLLMFQPIKADDLIALVRGIATFAALAE
TLS---PEE
>XP_002635404.1|/103-196
LKELGKILREAHNREIGAYKLEKFn-HPDVPFTKVYALKPFT---SESDLKGYMIMEYIPDI FTTSMTPEFSADDLIPTIKGVSTFAALGH
KLS---DDE
```