

A Comparative Analysis of Position Specific Scoring Matrixes for
the Identification of Related Proteins

Biochemistry 218
Computational Molecular Biology
Final Project

Matthew Taylor

June 6, 2002

Outline:

Introduction

- Question
- Background

Searches

- Description of Sequence identification
- Databases Used, Techniques, and Theory
- Map of Search Plan

Results

- Alignments for TGF-beta compare/contrast High vs. Total
 - eMATRIX
 - MEME/MAST
- Alignments for Auto-Inducer compare/contrast High vs. Total
 - eMATRIX
 - MEME/MAST
- Alignments for RRN3
 - eMATRIX
 - MEME/MAST

Conclusions:

- Overall
- Specific problems- outlier sequences affecting alignments
- Future areas
- Improvements

Figures

Appendix A: Sequence Lists

Appendix B: Clustal Alignments

Bibliography

Introduction:

There are many problems a researcher faces when trying to create a family of related proteins based on sequence analysis. The largest barrier is high sequence diversity compounded with the number of organisms that the data set is being chosen from. Often times, a researcher can group proteins based on conserved function between organisms that are far removed on an evolutionary ladder, but this requires experimental evidence to support the grouping and does not exhaust all the possible members of a family. For these reasons, searching the compendium of bioinformatics data available online opens new avenues of identification. Unfortunately, this identification is compounded by the same factors that made grouping difficult, and often the research only has a small number of highly divergent sequences to start with. Using search algorithms that utilize position specific probability modeling of amino acids for global sequences or for conserved motifs as a basis for searching can aid this search. These techniques may aid in the identification of sequences that may not be completely homologous, but are still related to the query proteins. For this project the questions that I am interested in answering are:

- How is the identification of sequence members affected by the sequence content using different position specific frequency scoring matrix?
- What program works best with large and/or small numbers of sequences?
- Can sequences that may represent new members of a protein family be identified with this method?

I propose to test the methods for their ability to identify sequence members from protein databases using three different sets of sequences groups that vary in their composition with respect to one another. I will test two algorithms, the Multiple Em for Motif Elicitation (MEME)^{1,2} program, and the eMATRIX maker program^{3,4}. The MEME software prepares a set of matrixes of conserved regions in the set of sequences, and can be queried against multiple databases using Multiple Alignment and Search Tool (MAST). The position scoring matrix created in eMATRIX will be used to query the Swiss-Prot database using eMATRIX search. The original plan had called for the use of BLOCKS maker^{5,6} to create an independent identification of conserved motifs, but this software did not provide the necessary output search function necessary for a direct comparison with the other programs. For these searches, three sequence groups have been compiled. Two were identified from the InterPro⁷ database of sequence patterns, and the third is an experimental set of sequences that have been grouped

based on functional conservation between organisms experimentally. These three groups differ in sequence number, conserved identity between members, and function of sequence members.

Methods

Sequence Query Compilation:

The two sequence groups derived from the InterPro database were created from two different patterns in that database, the Auto-inducer Binding domain⁸ and the TGF-beta receptor, type I/II extracellular region^{8,9}. The last group is formed from homologues of the RNA polymerase I transcription factor, RRN3¹⁰. These sequences were identified experimentally, and using different PSI-BLAST queries of each sequence identified.

The TGF-beta group comprised 187 protein members, of which 20 were selected for this analysis. The top and bottom 10 sequences of the family that were unique were selected. The top 10 were considered the “higher identity” group and the bottom 10 were considered the “lower identity” group. This was determined from pairwise alignment of 2 sequences from each group either within the group, or outside the group. There was a higher match within the group, rather than outside the group, so this sufficed for the determination. A more thorough test of the overall identity may have been useful, but for the further analysis it was not too important. The list was later modified to remove extremely divergent sequences of significantly larger size which prevented alignment of the total group to each other. These sequences were replaced by other unique sequences from the family. The list of sequences used for this group and all others is included in Appendix A at the end of this paper. The composition of this group is determined by its similarity to the TGF-beta external domain, and is spread across eukaryotic and prokaryotic sequences. This is a membrane bound protein that is used to transduce growth signals, and transport the molecule into the cell.

The Auto-inducer sequence group comprised 86 members, of which 30 were taken for analysis. The sequences were identified in a similar manner as described for the TGF-beta group selection. No testing of the “high” and “low” identity groups was performed. During the analysis of this group a single sequence was removed prior to alignment and use of the data set in the eMATRIX algorithm. Inclusion of this sequence led to a very poor alignment, and poor identification of sequences from the database, its exclusion increased the number and diversity of sequences, as well as creating a better alignment. The alignment that is included in Appendix B

for this group is excluding this sequence, which is identified as p55629 a hypothetical transcriptional regulator for a strain of *Rhizobium*. The overall composition of this group is exclusively prokaryotic in origin, and the group is selected for their similarity in binding site for the auto-inducer molecules of that organism. The auto-inducer is usually a small molecule excreted into the extracellular environment by the organism, where-upon it is taken up by similar species and then binds to this transcription factor to induce or repression expression of environmentally variable genes.

The third group of proteins was identified through a novel pathway separate from the other two groups. Using the sequences of human and yeast homologues of RRN3 in PSI-Blast a total group of 7 unique sequences of full-length proteins was identified. Individually, either query would only find a subset of the sequences that were closest to the evolutionary distance to itself, i.e. the human gene found mouse, *Drosophila*, and *Arabidopsis* homologues and not *C. elegans* or yeast homologues and vice versa. Sequence fragments, and small proteins were ignored from these searches, and could be excluded when low complexity or non-redundancy stringencies were applied. Experimentally, there is functional conservation between yeast and human homologues of this protein, and have the highest homology among all polymerase I transcription factors. It has been shown that deletion mutants of yeast RRN3, which are lethal, can be rescued by expression of human RRN3. As it has been described, the normal methods of finding similar proteins is not sufficient, so testing this protein with the methods in this paper may yield interesting results.

Search Techniques and Theory:

The two algorithms that will be used for this analysis is the eMATRIX and the MEME programs to create matrixes of amino acid composition. Both create a position specific matrix, but the differences are in the amount of sequence that is used, and how alignments are performed.

The eMATRIX software requires alignment of the group of sequences prior input into the eMATRIX maker program. This alignment can be performed on almost any multiple sequence alignment program. In this analysis the CLUSTALW^{11,12} program on the Decypher server was used for this purpose. The default alignment was used when possible, but some of the larger groups required suspension of end-gap penalties or removal of certain sequences to obtain alignments. The eMATRIX then uses the alignments to determine the amino acid expectations

for each residue in the alignment with baseline frequency based on pre-determined values that act as an assumption for the base occurrence of each amino acid at any residue. This makes the matrix more powerful than a Hidden Markov Model since the HMM determine residue frequencies from within the dataset. So with small sets where each amino acid does not have the chance to be represented, the HMM loses significant predictive value.

The MEME software package takes unaligned groups of related sequences and outputs groups of position scoring matrixes for each conserved motif within that group. It creates an optimal alignment of the sequences, identifies user defined minimal ungapped region which are homologous. For each motif identified, the program creates position specific amino acid frequencies. The output matrixes are then applied to the MAST search algorithm in the default nr database of sequences.

The output matches for both of these searches will be compared to the original group of sequences that were used to create them. Each search will be analyzed for how many original sequences can be identified, and how many new sequences can be found. These newly identified sequences are not really false positive, but need to be identified for the known function and relationship to the query group. For those groups that two alignments are produced, an internal comparison of performance will be made to determine the effects of sequence relationships to each other.

Results

TGF-Beta Searches:

eMATRIX. The searches using the eMATRIX algorithm were queried with matrixes built with the eMATRIX maker from CLUSTALW alignments. The two alignments produced from the “high” and “total” sequence group alignments are included in Appendix B. The total sequence alignment has large blocks of ungapped sequence at base pairs 134 to 201, 236 to 265, 270 to 298, 308 to 369, 424 to 447, 452 to 476, 492 to 526, 532 to 568, 593 to 632, and 637 to 655. Each of these blocks is separated by gaps in half of the sequences at the same position. The high sequence alignment does not have as many ungapped blocks, with only 5 such blocks at positions 160 to 194, 230 to 320, 340 to 413, 431 to 497, and 650 to 750. These blocks are fewer in number but larger in average size. The major difficulty in aligning the complete set of

sequences was the end gap penalties, which needed to be suspended to achieve a successful alignment.

The results of the eMATRIX searches was compiled together, and compared to the list of sequences applied. This information was tabulated into the graphs in figure 2, part a. It is interesting to note that the hits found in this search are redundant, with the same sequence being found multiple times. A total of 34 and 21 alignments were uncovered with the "high" and "total" alignments. Both searches were able to identify most of the query sequences that were derived from the Swiss-Prot, but more was found by the high- end search. The total search was unable to find most of the sequences input in the original alignment, but this may have been a limitation of the database that was searched. The low end sequences were primarily derived from the TrEMBL database, while the Swiss-Prot database was the only one searched by eMATRIX. This search was able to identify three new sequences that were similar to the queried group, these were orthologues of the PET1 and PET2 genes that was identified in rats and humans. Another sequence, YBGH from *E. coli* was also identified.

MEME. This search created its own alignment and then identified ungapped motifs were aligned, and scoring matrixes were developed for each motif. The program requires that the operator determine how many motifs would be searched for, and what would be the minimum and maximum size would be. For this analysis, the default parameters of 3 motifs with a minimum size of 8 and maximum size of 50 bases.

The MEME engine identified different motifs for the two sets of queries that were used. For the "high" sequence set, the three motifs were spread across the entire length of the protein. The first and third motif are located at the N-terminal region of the sequence set, while the second domain is located at the C-terminal end. This coincides with the fact that this group of sequences has an overall higher identity to each other, than to the second "low" group of sequences. Further, the MEME engine identified three motifs that were clustered around the N-terminal region of the "total" sequence set. This is expected, since the group was identified by its homology for the extracellular domain, located at the N-terminus of the proteins. The consensus sequences for both of these alignments can be found in figure 3. The actual file describing the motif discovery, the relative alignments, and the scoring matrixes was too large to be included in this paper.

The MEME model for the two groups of sequences was then input into the MAST search database. Initially, the “nr” database of sequences was queried, but a more comparative result of the search engines abilities was performed by querying the Swiss-Prot database only. It is interesting to note that the “nr” database yielded a larger list of sequences, many of which were not directly related to the query sequences. Secondly, this search only uncovered 5 of the ten input sequences for the “high” sequence list, and 8 of the ten sequence from the “total” list by the first 50 aligned sequences. These sequences were not limited to only the “high” sequence list, but spanned all the sequences from the “total” list, proving that the lack of identification from before is based on the limitations of the database.

The results from the queries of the Swiss-Prot database were analyzed as with the eMATRIX results, and compiled into the graph of figure 2b. Both searches yielded equivalent numbers of significant hits. The “high” query was able to align all ten of the original query sequences, while the “total” query could only uncover 9 of the original 20. The “total” query was able to uncover a larger list of new sequences, but both fell into the same set group as identified with the eMATRIX search.

Auto-Inducer Searches:

eMATRIX. The group of sequences identified for the Auto-inducer binding domain groups have a higher homology to one another, so alignments with these sequences is better described by analyzing where gaps are introduced. The N-terminal region is variable, and alignment does not begin until the first 6 to 10 residues for both alignments. The “total” group has four gaps greater than one residue at position 77, 100, 132, and 193. The “high” group only has two significant gaps at 110 and 160. These two gaps correspond to gaps 132 and 193 in the total alignment. Overall, there is a fewer number of gaps introduced in the “high” sequence set as compared to the “total” set and there is a greater number of conserved and similar residues that cluster around the C-terminal half of the protein. This is interesting. One puzzling fact to note is that earlier alignments were performed that included a sequence that was significantly different from the rest of the data. When used, this alignment yielded few hits on the eMATRIX software, and would not produce an alignment when used in the “total” sequence set of alignments without suspension of end-gap penalties. The overall alignment with this sequence was not that different, it just contained a large upstream sequence of 78 residues that affected its position in the grouping.

The results of the eMATRIX search yielded from these alignments was summarized as before, and plotted in the graph in figure 4a. The “total” sequence set yielded significantly lower numbers of aligned sequences, and found far fewer numbers of query sequences than the “high” group was able to identify. The “high” sequence set was able to identify far more new related sequences as well. Many of these were paralogous genes from related to the query set from the same organism. The list is as follows:

PHZR_PSEFL	TRAR_RHISN
EXPR_ERWCH	NARP_ECOLI
EXPR_ERWCA	CARR_ERWCA
SLR1_RALSO	DCTR_RHOCA
LUXS_VIBFI	

Similar genes were identified from the “total” query, but in a more limited number. 42 sequences were aligned in the “high” query, while only 11 were aligned from the “total” query. The total query was only able to identify half the original sequence set that was present in the database, rather than complete recovery with the “high” query. This is probably due to the lack of sequence similarity over the entire sequence within the total group, but the higher C-terminal homology within the “high” group.

MEME. The MEME motifs are virtually identical for both alignment methods. Motifs 1 and 3 are virtually identical to one another, except for the consensus sequence that is created from the alignments. The difference is at motif 2, which are in similar positions along the length of the aligned sequences but are formed from different ungapped regions. This second region is at the N-terminal domain of these transcription factors where the auto-inducer molecule binds the protein. The total group has a smaller than the high group motif 2, and is contained within the region that is used for motif 2 by the high group alignment. The consensus sequences of the motifs are included in figure 4.

The results from the MAST search were analyzed as before and compiled into the plot in figure 3b. Both programs worked equivalently at identifying the original sequences from the search results. The total query yielded more significant alignments of novel related sequences than the high query, yet both worked equally well for this. The sequences identified from this search are along the same family as identified from the eMATRIX search.

RRN3 Searches:

Both search engines failed to fully detect all the original sequences from this data set. This was due to the limitations of sequence presence in the query database, and could be circumvented by using the “nr” database in MAST. When analyzing only the Swiss-Prot database, either algorithm could only identify the *C. elegans*, *S.pombe*, and *S. cerevisiae* homologues of the RRN3 gene. When using the “nr” database, the majority of the hits were among human, rat, mouse, and drosophila sequences. The problem with this set of data, is that most of the information is from fragments, cDNA library screens, or genomic screening. There is a lack of cross-referencing in this information, so it is difficult to discern if the information is novel and represents a new member of the family, or if it is a redundant hit from the database. The alignment produced did not have any glaring homologous regions. There was more homology in the middle of the alignment than at the ends, and there was a greater incidence of similar and conserved sequences in ungapped blocks of this middle region. Better results from this alignment could have been obtained by modify the alignment parameters or using a more stringent alignment software, but this would not have significantly affected the final sequence search results.

Conclusions

The sum of this analysis was to determine how sequence composition affects search results from two different pattern development search methods. I used three different sequence sets of varying size and composition. The TGF-Beta group was a highly diverse set of sequences, that differed in ungapped blocks depending on the total composition of the alignment set. The Auto-Inducer group maintained a fairly conserved set of regions independent of the alignment group. The final RRN3 group contained a small number of sequences that didn't have any significant regions of ungapped alignment blocks, but did have the most conservation and similarity across the sequence sets. These were used in the eMATRIX and MEME engines for production of position specific sequence matrixes that were applied to search the Swiss-Prot database of proteins. The results were analyzed, and compared to each other to reach this critique.

From this analysis, the MEME engine for identification of conserved motifs to create a scoring matrix is the more powerful tool for this type of analysis. It is able to be modified the most by used defined parameters to create a more stringent or less stringent search matrix. The

output of this can be easily applied to a search algorithm that is able to search a variety of sequence databases to maximize the alignment results. Finally, it performed well under all variables tested, seemingly independent of sequence composition.

The eMATRIX engine has one major advantage over the MEME algorithm, which is ease of use, and rapid return on queries. While the MEME engine took almost 8 hours to return MEME matrix alignments, which then required re-transmission to search the database, the eMATRIX engine returned results in seconds, and was quickly applied to searching the Swiss-Prot database.

The major problem that was encountered with this analysis, was limitations placed on search results from database representation of the query proteins. The query was built from the InterPro database of pattern families, which includes Swiss-Prot and TrEMBL sequences. The limitation of searching a single database with eMATRIX limited the power of this analysis to produce the best output. The test sequence set of RRN3 was extremely hampered by this problem which prevented a thorough analysis of how the sequence composition or size may have affected the alignments.

To improve these engines would be to increase the speed of the MEME engine or to allow the user to search multiple sequence databases with eMATRIX. The eMATRIX algorithm is more sensitive to sequence diversity, so when searching for a conserved region, and its possible new family members, it is best to refine the query sequence to maximize the searching criteria.

Figure 1- Map of Search Plan

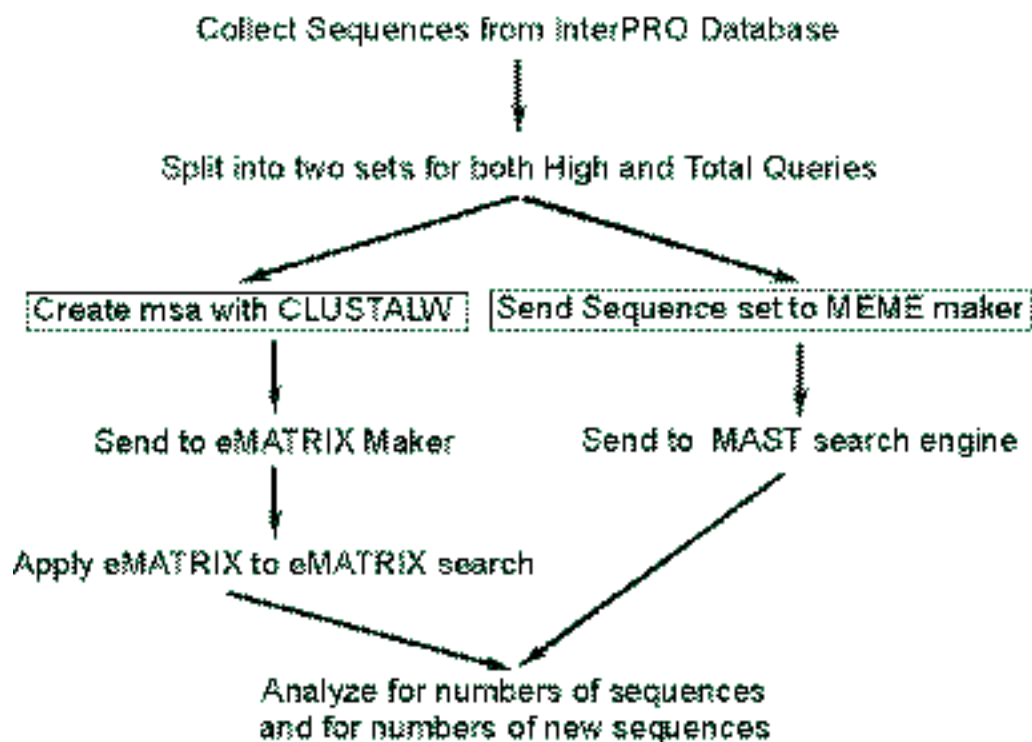


Figure 2: ROC plots based on query sequence set as true positive values for TGF-Beta Group Alignments using eMATRIX or the MEME/MAST methods

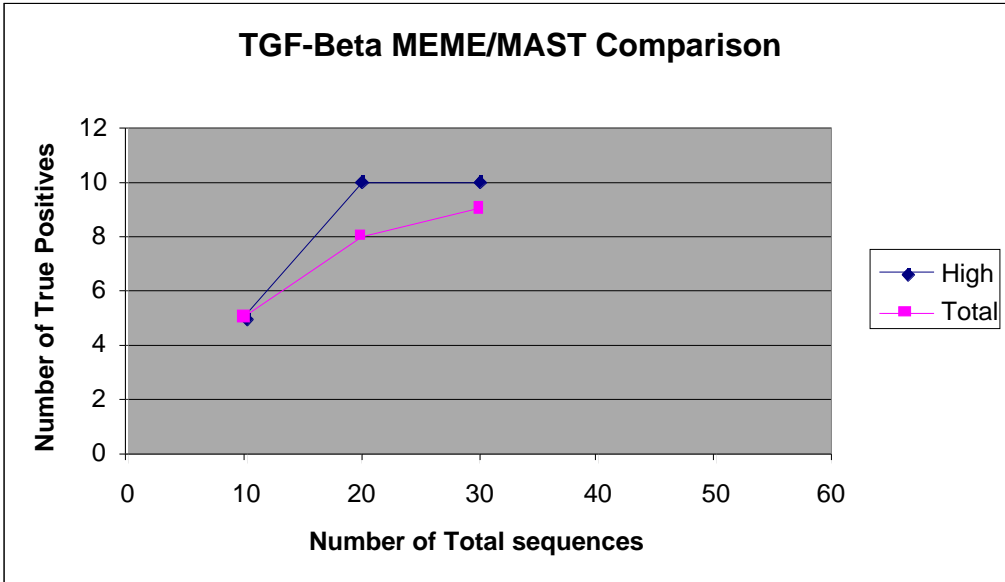
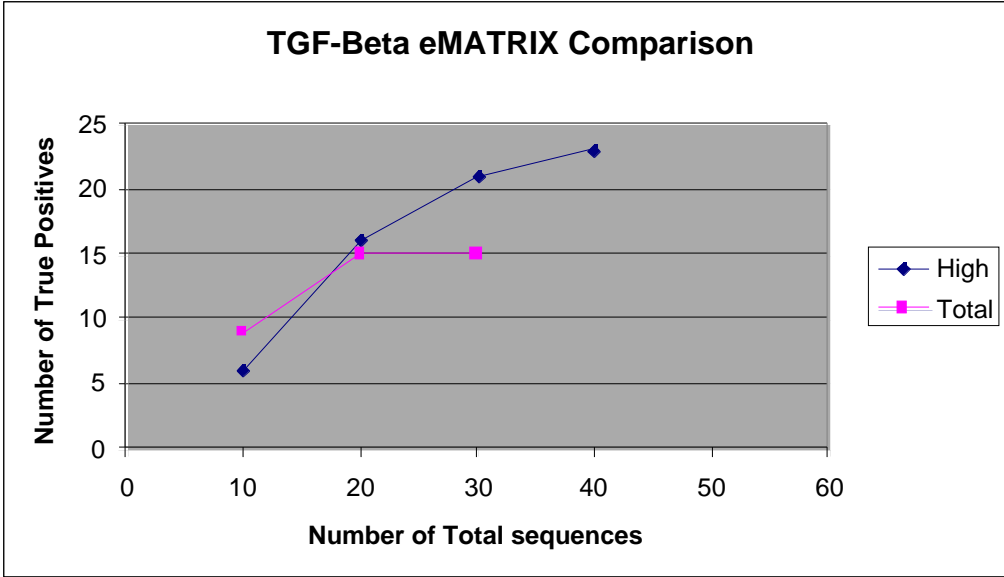


Figure 3- MEME motif identified consensus sequences

TGF-Beta Total Sequence List

MOTIF WIDTH BEST POSSIBLE MATCH

- 1 29 LIACGTGCIKPCVSAFGGDQYDEKDPRER
- 2 21 YVTPIIGGWIADHWWGRYKTI
- 3 28 NECCERFSYYGMРАНLVNYFTHQLHWDD

TGF-Beta High Sequence List

MOTIF WIDTH BEST POSSIBLE MATCH

- 1 29 QPRSCFMICINEFCERFSYYGMRAILQNY
- 2 42 QIPQYVLITAGEVFASITGLEFAYSQAPPSMKSVIQAMWLLT
- 3 24 YVTPIIGGWIADRWVGRYKTICCG

Figure 4- MEME motif identified consensus sequences

Auto-Inducer Total Sequence List

Auto-Inducer Total Sequence List

MOTIF WIDTH BEST POSSIBLE MATCH

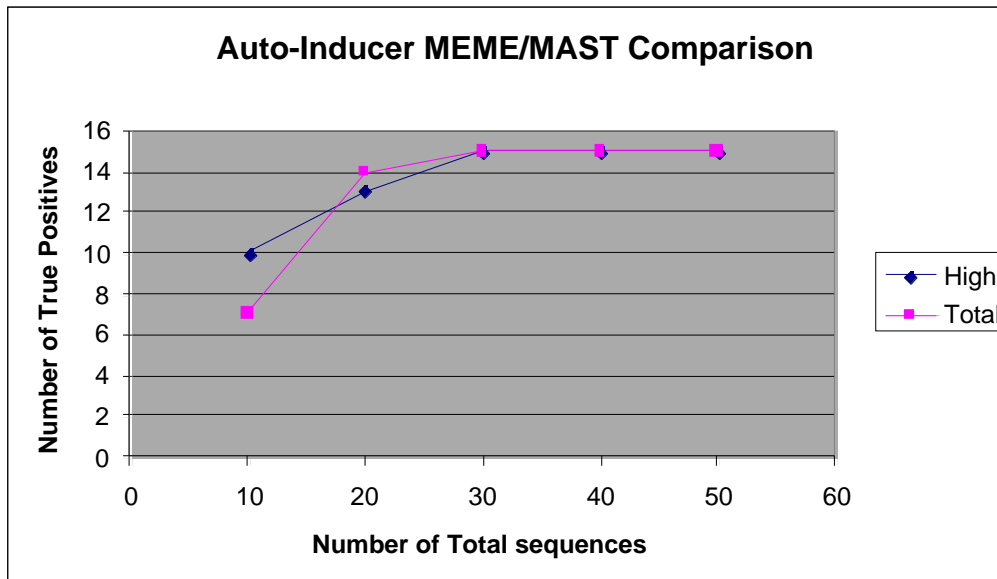
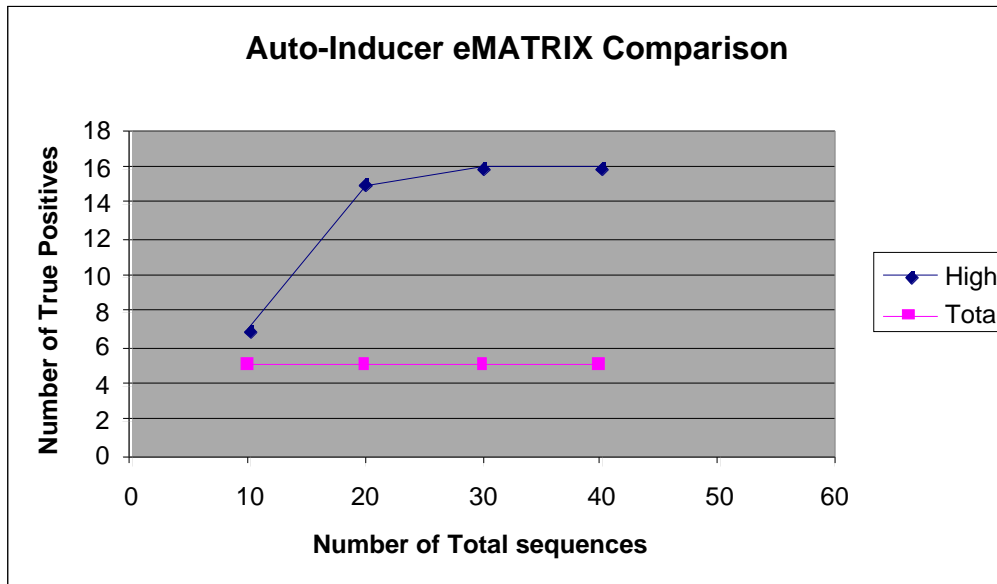
- 1 50 LSPRECECLCWTAMGKTYWEIACILGISERTVNFHMKNIQR
KLGVTNKKQ
- 2 26 NYPDDWVQHYQAQNYHHIDPVVKHGK
- 3 21 VWDEARDYGLCNGYTFPLHDH

Auto-Inducer High Sequence List

MOTIF WIDTH BEST POSSIBLE MATCH

- 1 50 LSPRECECLHWCAEGKTSWEIACILGISERTVNFHMKNVQK
KLGVTNKCQ
- 2 46 PMTPPHVFIVGNYPDEWVQHYQANNYQHIDPVVKHCKKCFHPFVWD
- 3 28 VWEEARDYGLCNGFTFPLHTANGELGML

Figure 3: ROC plots based on query sequence set as true positive values for the Auto-Inducer Group Alignments using eMATRIX or the MEME/MAST methods



Appendix A: Sequences

Part I- TGF-beta external binding domain Interpro Acc# IPR000109

“High Identity Sequences”

- >sp|O01840|OPT2_CAEEL Hypothetical oligopeptide transporter F56F4.5 - *Caenorhabditis elegans*.
- >sp|O07380|DTPT_LACHE DI-/tripeptide transporter - *Lactobacillus helveticus*.
- >sp|P32901|PTR2_YEAST Peptide transporter PTR2 (Peptide permease PTR2) - *Saccharomyces cerevisiae* (Baker's yeast).
- >sp|P36574|DTPT_LACLA Di-/tripeptide transporter - *Lactococcus lactis* (subsp. *lactis*) (*Streptococcus lactis*).
- >sp|P36836|PET1_RABIT Oligopeptide transporter, small intestine isoform (Peptide transporter 1) (Intestinal H+/peptide cotransporter) - *Oryctolagus cuniculus* (Rabbit).
- >sp|P36837|YHIP_ECOLI Hypothetical transporter yhiP - *Escherichia coli*.
- >sp|P39276|YJDL_ECOLI Hypothetical transporter yjDL - *Escherichia coli*.
- >sp|P46029|PET2_RABIT Oligopeptide transporter, kidney isoform (Peptide transporter 2) (Kidney H+/peptide cotransporter) - *Oryctolagus cuniculus* (Rabbit).
- >sp|P46030|PTR2_CANAL Peptide transporter PTR2 - *Candida albicans* (Yeast).
- >sp|P46031|PT2A_ARATH Peptide transporter PTR2-A - *Arabidopsis thaliana* (Mouse-ear cress).

“Low Identity Sequences”

- >tr|Q63422 Proton-dependent peptide transporter (Fragment) - *Rattus norvegicus* (Rat).
- >tr|Q8ZLD6 Putative POT family, peptide transport protein - *Salmonella typhimurium*.
- >tr|Q8XDS3 Putative peptide transporter - *Escherichia coli* O157:H7.
- >tr|Q43390 RCH2 protein - *Brassica napus* (Rape).
- >tr|Q9FRU2 Nitrate transporter NRT1-5 (Fragment) - *Glycine max* (Soybean).
- >tr|Q9SZY4 Nitrate transporter - *Arabidopsis thaliana* (Mouse-ear cress).
- >tr|Q9WU80 CAMP inducible 1 protein - *Mus musculus* (Mouse).
- >tr|Q9XIF3 F23H11.6 protein - *Arabidopsis thaliana* (Mouse-ear cress).
- >tr|Q9ZPS2 Putative peptide/amino acid transporter - *Arabidopsis thaliana* (Mouse-ear cress).
- >tr|Q9ZTX4 LeOPT1 - *Lycopersicon esculentum* (Tomato).

Part II- Auto-inducer family- Interpro Acc. # IPR005143

“Higher Identity Sequences”

- >sp|O30919|SLR2_RALSO Transcriptional activator protein solR - *Ralstonia solanacearum* (*Pseudomonas solanacearum*).
 - >sp|O54452|RAIR_RHIET Transcriptional activator protein raiR - *Rhizobium etli*.
 - >sp|O87971|YUKR_YERRU Transcriptional activator protein yukR (Fragment) - *Yersinia ruckeri*.
 - >sp|P07026|SDIA_ECOLI Regulatory protein sdiA - *Escherichia coli*.
 - >sp|P12746|LUXR_VIBFI Transcriptional activator protein luxR - *Vibrio fischeri*.
 - >sp|P25084|LASR_PSEAE Transcriptional activator protein lasR - *Pseudomonas aeruginosa*.
 - >sp|P33905|TRAR_AGRTU Transcriptional activator protein traR - *Agrobacterium tumefaciens*.
 - >sp|P54292|RHLR_PSEAE Regulatory protein rhlR (Elastase modulator) - *Pseudomonas aeruginosa*.
 - >sp|P54293|ESAR_ERWST Transcriptional activator protein esaR - *Erwinia stewartii*.
 - >sp|P54295|YENR_YEREN Transcriptional activator protein yenR - *Yersinia enterocolitica*.
 - >sp|P54303|PHZR_PSECL Transcriptional activator protein phzR - *Pseudomonas chlororaphis* (*Pseudomonas aureofaciens*).
 - *>sp|P55629|Y4QH_RHISN Hypothetical transcriptional regulator Y4QH - *Rhizobium* sp. (strain NGR234).
 - >sp|P74946|VANR_VIBAN Transcriptional activator protein vanR - *Vibrio anguillarum* (*Listonella anguillarum*).
 - >sp|Q03316|RHIR_RHILV Regulatory protein rhiR - *Rhizobium leguminosarum* (biovar *viciae*).
 - >sp|Q44059|AHYR_AERHY Transcriptional activator protein ahyR/asaR - *Aeromonas hydrophila*, and *Aeromonas salmonicida*.
- *- This sequence was removed for improved alignments and searches.

“Lower Identity Sequences”

- >tr|Q9I4J9 Probable transcriptional regulator - *Pseudomonas aeruginosa*.
- >tr|Q9KJ71 AHL-dependent transcriptional regulator CinR - *Rhizobium leguminosarum*.
- >tr|Q9KL60 Transcriptional regulator, LuxR family - *Vibrio cholerae*.
- >tr|Q9L3I8 SmaR protein - *Serratia* sp. ATCC 39006.
- >tr|Q9L6U5 Acylhomoserine lactone dependent transcriptional regulator - *Pseudomonas fluorescens*.
- >tr|Q9L6U7 Acylhomoserine lactone dependent transcriptional activator - *Pseudomonas fluorescens*.
- >tr|Q9R486 Conjugal transfer regulatory protein - *Agrobacterium tumefaciens*.
- >tr|Q9R6E3 TIORF116 PROTEIN - *Agrobacterium tumefaciens*.
- >tr|Q9R9G8 Transcriptional activator PhzR - *Pseudomonas chlororaphis* (*Pseudomonas aureofaciens*).
- >tr|Q9RH45 Transcriptional activator YpeR (Quorum-sensing transcriptional activator YpeR) - *Yersinia pestis*.
- >tr|Q9RMS5 PHZR (PROBABLE TRANSCRIPTIONAL REGULATOR) - *Pseudomonas aeruginosa*.
- >tr|Q9WWA5 Conjugal transfer regulatory protein - *Agrobacterium tumefaciens*.
- >tr|Q9WWB2 Conjugal transfer regulatory protein - *Agrobacterium tumefaciens*.
- >tr|Q9XDD1 LuxR homolog YpsR - *Yersinia pseudotuberculosis*.
- >tr|Q9ZIU0 Transcriptional activator - *Burkholderia cepacia* (*Pseudomonas cepacia*).

Part III- RRN3 sequences

- >P36070 RRN3 – *Saccharomyces cerevisiae* (yeast)
- >tr|Q9NYV6 RRN3 - *Homo sapiens* (Human).
- >sp|P48322|YPJ1_CAEEL Hypothetical 75.1 kDa protein C36E8.1 in chromosome III - *Caenorhabditis elegans*.
- >sp|Q10110|YAQB_SCHPO Hypothetical protein C18G6.11c in chromosome I - *Schizosaccharomyces pombe* (Fission yeast).
- >(XM_156394) similar to CG3278 gene product *mus musculus* putative RRN3 sequence
- >tr|Q9V9M6 CG3278 protein - *Drosophila melanogaster* (Fruit fly).
- >NM_102796 *arabidopsis thaliana* putative RRN3 sequence

Appendix B: CLUSTAL Alignments

TGF-Beta External Domain Alignment- Total Sequences

```
sp|O01840|-----  
sp|O07380|-----  
sp|P32901|---MLNHPSQGSDDAQDEKQGDFPVIEEEKTQAVTLKDSYVSDDVANSTERYNLSPPED  
sp|P36574|-----  
sp|P36836|-----MGM-----  
sp|P36837|-----  
sp|P39276|-----  
sp|P46029|-----MNPFFQONESKETLFSVPVSTEETPPRLSSPAKKT-----  
sp|P46030|-----MVSSDFENEKQPDVVQVLTDEKNI SLDDKYDYEDPKNYSTNYVDDYNP  
sp|P46031|MSSIEEQITKSDSDFI I SEDQSYLSKEKKADGSATINQADEQSSTDELQKSMSTGVLVNG  
tr|Q63422|-----  
tr|Q8ZLD6|-----  
tr|Q8XDS3|-----  
tr|Q43390|-----MSLPETKT-----QTLDDAWDFQGR-----  
tr|Q9FRU2|-----  
tr|Q9SZY4|-----MESKGS-----WTVADAVDYKGR-----  
tr|Q9WU80|-----MSAPRAEEQPSRSGERQ-----  
tr|Q9XIF3|-----  
tr|Q9ZPS2|-----MASIDEERSLLEVEESLIQEEVKLYAEDGSIDIHGN-----  
tr|Q9ZTX4|-----MKYLF SKNGLLEDENS-----GLYTRDGSVDIKGN-----  
  
sp|O01840|---MEEKSLLQKLRSYPPAVF-----FMLGNEFCERFSFYGMKTILFIYL-----  
sp|O07380|-----MRAILLFYMYAVT-----  
sp|P32901|EDFEGPTEEEEMQTLRHVGGKIPMRCWLIAIVELSERFSYYGLSAPFQNYMEY-----  
sp|P36574|MQNLNKTEKTFFGQPRGLL-----TLFQTEFWERFSYYGMRAILVYYLYA-----  
sp|P36836|-----SKSLSCFGYPLSIF-----FIVVNEFCERFSYYGMRAILLILYF-----  
sp|P36837|---MNTTTPMGMLQQPRPFF-----MIFFVELWERFGYYGVQGVLAFFF-----  
sp|P39276|---MKTSPQPRAIY-----YIVAIQIWEYFSFYGMRAILLILYL-----  
sp|P46029|---PPKICGSNYPLSIA-----FIVVNEFCERFSYYGMKAVLTLYF-----  
sp|P46030|KGLRRPTPQESKSLRRVIGNIRYSTFMLCICEFAERASYSTTGILTNYIQRRIDPDSPH  
sp|P46031|DLYPSPTEEELATLPSVCGTIPWKAFI I I IVELCERFAYYGLTVPFQNYMQF-----  
tr|Q63422|-----XEFCEFRFSYYGMRAILLVLYF-----  
tr|Q8ZLD6|---MNTTAPTGLLQQPRPFF-----MIFFVELWERFGYYGVQGILAVFFF-----  
tr|Q8XDS3|---MKTSPQPRAIY-----YIVAIQIWEYFSFYGMRAILLILYL-----  
tr|Q43390|---PADRSKTGGWASAA-----MILCIEAVERLTTLGIGVNLVLYL-----  
tr|Q9FRU2|---ENDPKIDYRGWKAMP-----FIIGNETFEEKLGAIGTLANLLVYL-----  
tr|Q9SZY4|---PADKSKTGGWITAA-----LILGIEVVERLSTMGIAVNLVLYL-----  
tr|Q9WU80|---PLVARGPRGPRRWRRTAAA-----AVLLVQMLERAFFGVTSNLLVLYLN-----  
tr|Q9XIF3|-----  
tr|Q9ZPS2|---PPLKQTTGNWKACP-----FIFANECCERLAYYGIANKLITYF-----  
tr|Q9ZTX4|---PVLKSETGNWRACP-----FILGNECCERLAYYGIAANLVLYL-----  
  
sp|O01840|-----ITEHEFSPSKATFIYHLFTCIAYLTPLIGSIMADSVFGRFKVILYGS  
sp|O07380|-----KGGLGMSQTAAASIMSIYGLSVYLSTLVGGWLSDRVWGSRKTVFYGG  
sp|P32901|-----GPNDSPKGVLSLNSQGATGLSYFFQFWCVYTPVFGGYVADTFWGKYNTICCGT  
sp|P36574|-----LTTADNAGLGLPKAQAMAIVSIY GALVYLSTIVGGWVADRLLGASRTIFLGG  
sp|P36836|-----RNF IGWDDNLSTVIYHTFVALCYLTPILGALIADAWLGGFKTIVWLS  
sp|P36837|-----VKQLGFSQEQAFTVTFGAFALVYGLISIGGYVDHLLGKFKTIVLGA  
sp|P39276|-----THQLGFDDNHAI SLFSAYASLVYVTPILGGWLADRLLGNRTAVIAGA  
sp|P46029|-----LYFLHWNETSTSVYHAFSSSLCYFTPILGAAIADSWLGKFKTIIYLS  
sp|P46030|GWGAPPPGSPDASAGALGKGLQAASALTNLLTFLAYVFP LIGGYLGDSTIGRWKAIQWGV
```

sp | [P46031](#) -----GPKDATPGALNLAGETGADGLSNFFTFWCYVTPVGAALIADQFLGRYNTIVCSA
tr | [Q63422](#) -----RNFLGWDDDLSTAIYHTSVALRYLTPILGALIADSWLGKFKTIVSLS
tr | [Q8ZLD6](#) -----VKQLGFSQEQAFITFGAFAALVYGLISIGGYVGDHLLGTRKRTLVLGA
tr | [Q8XDS3](#) -----THQLGFDDNHAI SLFSAYASLVYVTPILGGWLADRL LGNRTAVIAGA
tr | [Q43390](#) -----TGTMHLGNATAANTVTNFLGTSFMLCLLGGFIADTFLGRYLTIAIFA
tr | [Q9FRU2](#) -----TTVFNLKNITATNIINIFNGSTNFATFIGAFLSDTYFGRYKTIGFCT
tr | [Q9SZY4](#) -----METMHLPSSTSANIVTDFMGTSFLLCLLGGFLADSF LGRFKTIGIFS
tr | [Q9WU80](#) -----SLNFNWDGQQASRATLLFLGASYLLAPVGGWLADVYLGRFLTISLSL
tr | [Q9XIF3](#) -----
tr | [Q9ZPS2](#) -----TNELHETNVSAARHVMTWQGTCTYITPLIGALIADAYWGRYWTIACFS
tr | [Q9ZTX4](#) -----TKKLHEGNVSAARNVTTWQGTCTYITPLIGAVLADAYWGRYWTIATFS

sp | [O01840](#) SIYVVGHVLLSLGAVPFLSYPI-----RSSLDFSGFLVI
sp | [O07380](#) VLIMLGHIVLALPAGVTVLYRS-----IALI
sp | [P32901](#) AIYIAGIFILFITSIPSVGNRDSAIGGF-----IAAII LIGIATGMIKANLSVLI
sp | [P36574](#) ILITLGHIALATPFGLSSLFVA-----LFLI
sp | [P36836](#) IVYTIQAVTSLSSVNELTDNNHDG-----TPDSLPHVHAVCMIGLLLI
sp | [P36837](#) LVLAI GYFMTGMSLLKPD LIFI-----ALGTI
sp | [P39276](#) LLMTLGHVV LGIDTNSTFSLYL-----ALAI I
sp | [P46029](#) LVNVLGHVIKLSAFPILGGKV-----VHTVLSLVGLCLI
sp | [P46030](#) FFGFVAHLFFIFASIPQAIENANAGLGLC-----VIAIITLSAGSGLMKNLLPLV
sp | [P46031](#) VIYFIGILILTCTAIPSVIDAGKSMGGF-----VVS LIIIGLGTGGIKSNVSPLM
tr | [Q63422](#) IVYTIQAVISVSSINDLTDHHDG-----SPNNLPHVALSMIGLGLI
tr | [Q8ZLD6](#) IVLAI GYFMTGMSLLNPD LIFI-----ALGTI
tr | [Q8XDS3](#) LLMTLGHVV LGIDTNSTFSLYL-----ALAI I
tr | [Q43390](#) AIQATGVSILTLSTIIPGLRPPRC DPTT-----SSHCVQANGIQ LTVLYLALYLT
tr | [Q9FRU2](#) FTSFLGLLVIQLTAVFKNLHPPHCGK-E-----MKTCKGPTAGQMAFLVSGFLL
tr | [Q9SZY4](#) TIQALGTGALAVATKLP E LRPPTC--H-----GEACIPATAFQMTILYVLSYLI
tr | [Q9WU80](#) LLYLAASGLLLTTITNDGRSRFCGEMPELPLEPAC P SSSCQGSWSSPYCATTLYLVLLLL
tr | [Q9XIF3](#) HVYSQGFILLSVQAHLPQLKPPKCNPLI-----DQTCEEAKGFKAMIFFMALYL V
tr | [Q9ZPS2](#) AIYFTGMVALTLSASVPGLKPAECI-----GSLCPPATMVQSTVLFSGLYLI
tr | [Q9ZTX4](#) TIYFIGMCTLTL SASVPAFKPPQCV-----GSVCPSASPAQYAIFFFGLYLI

sp | [O01840](#) AFATGCIKPCVSAFAADQFTEDQKD---LRSQFFSFFFYFAINGGSLFAIIITPILRGRVQ
sp | [O07380](#) VVGTGLLKPNVSDMVGGLYSVEDPR---RDAGFSIFVFGINLGSIIAPWLV P WAAQGGF
sp | [P32901](#) ADQLPKRKPSIKVLKSGERVIVDSN--ITLQNVFMFFYFMINVGSL--SLMATTELEYH--
sp | [P36574](#) ILGTGMLKPNISNMVGHLYSKDDSR---RDTGFNIFVVGINMGSLIAPLIVGT VGGQ--
sp | [P36836](#) ALGTGGIKPCVSAFGGDQFEEGQEK---QRNRFFSIFYLAINAGSLLSTIITPMVRVQ-Q
sp | [P36837](#) AVGNGLFKANPASLLSKCYPPKDPR---LDGAFTLFYMSINIGSLIALSLAPVIADR--
sp | [P39276](#) ICGYGLFKSNISCLLGELYDENDHR---RDGGFSLLYAAGNIGSIAAPIACGLAAQW--
sp | [P46029](#) ALGTGGIKPCVAAFGGDQFE EKHA E---ERTRYFSGFYLA INAGSLISTFITPMLRGDVQ
sp | [P46030](#) LDQYPEERDMVKVLP TGESIILDRE--KLSLRITNVFYLA INI GAF LQIATSYCERRV--
sp | [P46031](#) AEQLPKIPPYVVKTKKNGSKVIVDPV--VTTSRAYMIFYWTINVGSL--SVLATT SLEST--
tr | [Q63422](#) GLGTGGIKPCVSAFGGDQFEGEQEK---QGNPVFSIFYLAINAGSLLSTIITPILRVQ-Q
tr | [Q8ZLD6](#) AVGNGLFKANPASLLSKCYQPKDPR---LDGAFTLFYMSINIGSLLSLSLAPVIADK--
tr | [Q8XDS3](#) ICGYGLFKSNISCLLGELYDENDHR---RDGGFSLLYAAGNIGSIAAPIACGLAAQW--
tr | [Q43390](#) ALGTGGVKASVSGFGSDQFDDTEPKERSQMTYFFNRFFFCINVGSLMAVTVLVYIQDD--
tr | [Q9FRU2](#) LIGAAGVRPCNLAFGADQFNPNNTDSGKKGINSFFNWFYFTFTFAQMVS LTLIVYVQSN--
tr | [Q9SZY4](#) ALGTGGLKSSISGFGSDQFDDKDPKEKAHMAFFFNRF FFI SMGTL L AVTVLVYMQDE--
tr | [Q9WU80](#) ALAASSVRSTLTSFGADQVMDLGRD---ATTRFFNWFYWSINLGA ILSLLVVA FIEQN--
tr | [Q9XIF3](#) ALGSGCVKPNMIAHGADQFSQSHPKQSKRLSSYFNAA YFAF SMGELIALTLLVWVQTH--
tr | [Q9ZPS2](#) ALGTGGIKPCVSSFGADQFDKTDPSERVRKASFFNWFYFTINIGAFVSSTVLVWIQEN--
tr | [Q9ZTX4](#) ALGTGGIKPCVSSFGADQFDDTDPKERVKKGSFFNWFYFSINIGALISSSLIVWIQEN--

sp | [O01840](#) C--FGNAHCFPLAFGVPGVLM L LALILFLMGWSMYKKHPPSKENVGSKVVAVIYTS LRKM
sp | [O07380](#) VHIFGSQLNFHAGFSLAAVGMFFGLVQYVVLGGKKY LSTESLTPNDPIDKGDLLNVIKWVV
sp | [P32901](#) -----KGFWAAYLLPFCFFWIAVVTLIFGKKQYIQRPIGDKVI-AKSFKVC--WILT K

sp | [P36574](#) | -----VNYHLGFSLAAIGMIFALFAYWYGRRLRHFPEIGREPSNPMDSKARRNFLITLT
 sp | [P36836](#) | CGIHVKQACYPLAFGIPAILMAVSLIVFIIIGSGMYKKFKPQGNIL-SKVVKCICFAIKNR
 sp | [P36837](#) | -----FGYSVTYNLCGAGLIIGLLVYIACRGMVKDIGSEPDFR-PMSFSKLLYVLLGS
 sp | [P39276](#) | -----YGWHVGFALAGGGMFIGLLIFLSGHRHFQSTRSMDKKA-LTSVKFALPVWSWL
 sp | [P46029](#) | C---FGEDCYALAFGVPGLLMVIALVVFAMGSKMYKKPPPEGNIV-AQVVKCIWFAISNR
 sp | [P46030](#) | -----GFWLAFFVPMILYIIVPIFLFIVKPKLKI KPPQGVQVM-TNVVKILAVLFSGN
 sp | [P46031](#) | -----KGFVYAYLLPLCVFVIPLIILAVSKTAFTSTLLPPVPS-LFVLVKCSSLLKKT
 tr | [Q63422](#) | CGIHSQQACYPLAFGVPAAALMAVALIVFVLGSGMYKKFQPQGNIM-GKVAKCIGFAIKNR
 tr | [Q8ZLD6](#) | -----FGYAVTYNLCGAGLIVALLVYFACRGMVKNIGSEPDHK-PLRFRNLLLVLVLT
 tr | [Q8XDS3](#) | -----YGWHVGFALAGGGMFIGLLIFLSGHRHFQSTRSMDKKA-LTSVKFALPVWSWL
 tr | [Q43390](#) | -----VGRKWGYGICALAILVLSIFLAGTNRIRYRFFKLLIGSPM-TQVGTVLVAWRNR
 tr | [Q9FRU2](#) | -----VSWAIGLGI PAALMLISCVVYFMGSKIYVKVEPSGSP-AGIVQVFFVAVKRR
 tr | [Q9SZY4](#) | -----VGRSWAYGICTVSMIAIVIFLCGTKRYRYKKSQGSVP-VQIFQVIAAAFRKR
 tr | [Q9WU80](#) | -----ISFLWGYSIIIVGLVGLAFFIFL FATPVFITKPPTGSQV-SSMLKLA FQNCPC
 tr | [Q9XIF3](#) | -----SGMDIGFGVSAAAMTMGIISLVSGTMYFRNKRPRRSIF-TPIAHVIVAAILKR
 tr | [Q9ZPS2](#) | -----YGWELGFLIPTVFMGLATMSFFFGTPLYRFQKPRGSP-ITSVCQVLVAAYRKS
 tr | [Q9ZTX4](#) | -----AGWGLGFGIPAVFMGIIASFFFGTPLYRFQKPGGSPL-TRMCQVLVAVFHKW

:

sp | [O01840](#) | VGGASRDKP---VTHWLDHA-----
 sp | [O07380](#) | IIIIAIVAILAAMAGVGLSVD-----NVITLLTILAIALP-----
 sp | [P32901](#) | NKFDFNAAKP--SVHPEKNY-----
 sp | [P36574](#) | IVVIVAIIGFFLLYQASPANFI-----NNFINVLSIIIGIVVPI-----I
 sp | [P36836](#) | FRHRSKQFPK--RAHWLDWA-----
 sp | [P36837](#) | VVMIFVCAW---LMHNVEVAN-----LV--LIVLSIVV-----
 sp | [P39276](#) | VVMLCLAPVFFTLLENDWSG-----YL--LAIIVCLIA-----
 sp | [P46029](#) | FKNRSEDIPK--RQHWDWA-----
 sp | [P46030](#) | FIKRLWNGTFWDHARPSHMEA-----RGTIYYNSKKK-----
 sp | [P46031](#) | NLI SKKLNHL--ALLLLERYV-----
 tr | [Q63422](#) | FRHRSKAFPK--REHWLDWA-----
 tr | [Q8ZLD6](#) | VVMIFLCAW---LMHNVKIAN-----LV--LIVLSIVV-----
 tr | [Q8XDS3](#) | VVMLCLAPVFFTLLENDWSG-----YL--LAIIVCLIA-----
 tr | [Q43390](#) | RLELPSDPSF--LYDLDVIAAEGSMKSKQKLPHTNQFRSLDKAAIKDQ----EMAMTQN
 tr | [Q9FRU2](#) | SIKLPAEHPMLSLFNYPVPPMSV----NS-KLPYTFQFRLLDKAAIVTPKD--KIKPDGS
 tr | [Q9SZY4](#) | KMELPQSIVY--LYEDNP-----EGIRIEHTDQFHLLDKAAI VAEGDFEQTLDGVA
 tr | [Q9WU80](#) | RKSSSRDSES---AHL LPDQR-----
 tr | [Q9XIF3](#) | KLASPSDPRM--LHGDHVVANDV---VPSSTLPHTPRFRFLDKACIKIQ-----DTNT
 tr | [Q9ZPS2](#) | NLKVPEDST-----DEGDA-----
 tr | [Q9ZTX4](#) | NLSVPDDSTL--LYETPKSSAI---EGSRKLLHTDELRLCLDKAAVSD----NELTTGD

sp | [O01840](#) | -APEHSQKMIDSTRGLLNVAVIFCPLI---FFWALFDQQGSTWVLQARRLDGRVGH----
 sp | [O07380](#) | IYFVMMFRSSKVTKIELGIHLLPVSLKNRLF FKKGYKRLKQIIQLELAIKRQSFII---
 sp | [P32901](#) | ---PWNDKFVDEIKRALAACKVFI FYP---IYWTQYGTMISSFITQASMMELHG-----
 sp | [P36574](#) | YFVMMFTSKKVESDERRKLTAYIPLFLSAIVFWAIEEQSSTIIAVWGESRSNLDPTWFGI
 sp | [P36836](#) | -KEYDERLIAQIKMVT RVLFLYIPLP---MFWALFDQQGSRWTLQATTMSGRIGI----
 sp | [P36837](#) | TIIFFRQAFKLDKGTGRNKMVFVFLM LEAVVFI ILYAQMP TSLNFFAINN VHEILG---
 sp | [P39276](#) | AQIIARMMIKFPEHRR-LWQIVLLMFVGT LFWVLAQQGGSTISL FIDRFVNRQAFN---
 sp | [P46029](#) | -AEKYPQLIMDVKTLTRVFLYIPLP---MFWALLDQQGSRWTLQATKMNGNLGF----
 sp | [P46030](#) | SAITWSDQWILDIKQTFDCKIFLYYI---IFNLADSGLGSVETSLIGAMKLDG-----
 sp | [P46031](#) | -KDQWDDLFI DELKRALRACKTFLFYP---IYWVCYQMTNNLISQAGQMOTGN-----
 tr | [Q63422](#) | -KEYDERLISQIKMVT KVMFLYIPLP---MSWALFDQQGSRWTLQATTMTGKIGT----
 tr | [Q8ZLD6](#) | TIIFFFREAFRLDKTGRNKMVFVFI LMI EAVLFI ILYAQMP TSLNFFAINN VHEILG---
 tr | [Q8XDS3](#) | AQIIARMMIKFPEHRR-LWQIVLLMFVGT LFWVLAQQGGSTISL FIDRFVNRQAFN---
 tr | [Q43390](#) | VYNKWTLSVTVDIEEVKQIVRMLPIWATCILFWTVHAQLT TLSVAQSETMDRHIGS----
 tr | [Q9FRU2](#) | AADPWNLCSIQQVEEAKCVVRVLP IWFAAIVYHLVIVQMHTLLV FQALQSDRRLGSSN--
 tr | [Q9SZY4](#) | IPNPWKLSSVTKVEEVKMMVRLLP I WATTIIFWTTYAQMITFSVEQASTMRRNIGS----
 tr | [Q9WU80](#) | -SNQPGPSPQEDMANFQVLVKILPVMVTLVPYWMVYFQMQSTYVLQGLHLHIPNIFRTNP
 tr | [Q9XIF3](#) | KESPWRLCTVTQVEQVKTLISLVP IFASTIVFNTILAQLQTF SVQQGSSMNTRL SNS---

tr | Q9ZPS2 NTNPWKLCVTVTQVEEVKILLRRLVPIWASGIIFSVLHSQIYTLFVQQGRCKMRTIGL-----
tr | Q9ZTX4 YSNAWRLCTVTQVEELKILIRMFPIWATGIVFSAVYAQMSTMFVEEQGMVMDTAVGS-----

sp | O01840 -----FSILPEQIHAINPVCVLILVPIFEGWVYPALRKIT-----RVTPLRKM
sp | O07380 -----LIALIIMASILIPNKVIIAKHLKLVLLVIFYWIGL-----NLIPFSTF
sp | P32901 -----IPNDFLQAFDSIALIIFIPIFEKVFVYPIRRYT-----PLKPITKI
sp | P36574 T-----FHIDPSWYQLLNPLFIVLLSPIFVRLWNKLGERQP-----STIVKF
sp | P36836 -----LEIQPDQMOTVNTILIIILVPIMDAVVYPLIAKCGL-----NFTSLKKM
sp | P36837 -----FSINPVSFQALNPFVWVVLASPILAGIYTHLGNKKGK-----DLSMPMKF
sp | P39276 -----IEVPTALFQSVNAIAVMLAGVVLAWLASPESRGN-----TLRVWLKF
sp | P46029 -----FVLQPDQMQLNPLLVLIIFIPILFDLVIYRLISKCGI-----NFTSLRKM
sp | P46030 -----VPNDLFNNFNPLTIIILIPILEYGLYPLLNKFKI-----DFKPIWRI
sp | P46031 -----VSNDLFQAFDSIALIIFIPICDNIYPLLRKYNI-----PFKPIILRI
tr | Q63422 -----IEIQPDQMOTVNAAILIVIMVPIVDAVVYPLIAKCGF-----NYTSLKKM
tr | Q8ZLD6 -----FAINPVSFQALNPFVWVVASPVLAAIYTRLGSKGK-----DLTMPMKF
tr | Q8XDS3 -----IEVPTALFQSVNAIAVMLAGVVLAWLASPESHGNS-----TLRVWLKF
tr | Q43390 -----FEIPPAAMAVFYIGGLLTTAVYDRLAIPCKKLFNYPH-GLRPLQRI
tr | Q9FRU2 -----FKIPGASFNVFLMLSMTLWLPIDYDRIVVPFSLSNTPKESGGITLLQSD
tr | Q9SZY4 -----FKIPAGSLTVFFVAAIILITLAVYDRAIMPFWKKWKGK-----GFSSLQRI
tr | Q9WU80 NISLLLRSDSSNYRIPEAWLLLANVAVILILIPVKDHLIDPLLLRCKL-----LPSSLQKM
tr | Q9XIF3 -----FHIPPASLQAIPIYIMLIFLVPLIYDSFLVVPFARKLTGHNS-GIPPLTRI
tr | Q9ZPS2 -----FEIPPATLGMFDTASVLIISVPIYDRIVPLVRRFTGLAK-GFTELQRM
tr | Q9ZTX4 -----FKIPAASLSTFDTISVIVWVPIYDKILVPIARRFTGIER-GFSELQRM

: . :

sp | O01840 AVGGLLTAFSFAIAGVLQKLVNETMEFPPSL-----GRIYLQRVGNESLISDFRYKSDG
sp | O07380 VLSFLFLDYIKHMFKKEGEQAKKTKEKSRIHH-----GIEIPLFLRQLIINIFTLIILEG
sp | P32901 FFGFMFGSFAMTWAAVLQSFVYKAGRPWYNEPL-----GHNTPNHVHVCWQIPAYVLIIFS
sp | P36574 GLGLMLTGISYLIIMTLPLGLLNGTSGRASAL-----WL-VLMFAVQMAG
sp | P36836 TIGMFLASMAFVAAAAILQVEIDKTLVPFPAK-----NEVQIKVLNVGSENMIISLPGQT
sp | P36837 TLGMFMCSLGLFTAAAAGMWFADAQGLTSP-----WFIVLV-----YLFQSLG
sp | P39276 AFGLLLMACGFMLLAF-DARHAAADGQASM-----GVMISG-----LALMGFA
sp | P46029 AVGMVLAACLAFAAAATVEIKINEMAPPQPGS-----QEILLQVLNLADDEVKLTVLGNN
sp | P46030 CFGFVVCSFSQIAGFVLQKQVYEQSPCGYYAT-----NCDSPAPITAWKASSLFIILAAAG
sp | P46031 TLGMFLATASMIYAAVLAQAIYQRGPCYANFT-----DTCVSNDISVWIQIPAYVLIASF
tr | Q63422 TVGMFLASMAFVAAIVQVEIDKTLVPFVPSG-----NQVQIKVLNIGNNDMAVYFPGKN
tr | Q8ZLD6 TLGMFLCALGFLTAAA-GMWFADAQGLTSP-----WFIVLV-----YLFQSLG
tr | Q8XDS3 AFGLLLMACGFMLLAF-DARHAAADGQASM-----GVMVSG-----LALMGFA
tr | Q43390 GLGLLLAAMGMAVAALVEIKRLRTAHAHGPT-----VKTLPLGFSLLIPQYLIVGIG
tr | Q9FRU2 GIGIFISALCMIVAGVVEEHRRLALTNPIGVQ---PRKGAISSMSGLWLIPQLSLAGLS
tr | Q9SZY4 AIGLVLSTAGMAAAALVEQKRLSVAKSSSQ-----KTLPIISVFLVLPQFFLVGAG
tr | Q9WU80 ALGMFFGFTSIIIVAGVLEKERLQYIAANQTVPQLIGKDLYYAAPLSIWWQIPQYLLIGVS
tr | Q9XIF3 GIGLFLSTFSMVSAAAMLEKKRRDSSVLDGR-----ILSIFWITPQFLIFGIS
tr | Q9ZPS2 GIGLFFSVLSLTFAAIVETVRLQLARDLDLV-----ESGDIVPLNIFWQIPQYFLMGTA
tr | Q9ZTX4 GIGLFLSMLCMSAAAIVEIRRLQLARDLGLV-----DEAVSVPLSIFWQIPQYFILGAA

. . .

sp | O01840 R-LIGDGMLPKGRTELDAGIYTFNTGLKNESQEI---DISTPNKGYVMVAVFRLKDAVEV-
sp | O07380 ETLFDENGVEVNIAEHPVQGYTELNINLLNKD-----SIDLWADWIQSVAKYLLNIMY--
sp | P32901 EIFASITGLEAYASKAPASMSKSFIMSFLLTN-----AFGSAIGCALSPVTVDPKFTW--
sp | P36574 ELLVSPVGLSVSTKLAPVAFQSQMMAMWFLAD-----STSQAINAQITPLFKAATEVH--
sp | P36836 VTLNQMSQTNEFMTFNEDTLTSINITSGSQ-----VTMITPSLEAGQRHTLLVW--
sp | P36837 ELFISALGLAMIAALVPQHLMGFILGMWFL-----TQAAAFLLGGYVATFTAVPD--
sp | P39276 ELFIDPVAIAQITRLK---MSGVLTGIYMLA-----TGAVANWLAGVVAQQTTESQ--
sp | P46029 NNSLLADSIKSFQKTPHYSKIHLNTKSQDFYFHLKYHNLSIYTEHSVEERNWYSLI IRED
sp | P46030 ECWAYTTAYELAYTRSPPALKSLVYALFLVMS-----AFSAALSLAITPALKDPNLHW--
sp | P46031 EIFASITGLEFAFTKAPPSMKSIITLFLFTN-----AFGAILSICISSTAVNPKLTW--
tr | Q63422 VTVAQMSQTDTFMTFDVDQLTINSINVSSPGSPG-----VTTVAHEFEFGHRHTLLVW--

tr | Q8ZLD6 ELLISALGLAMVAALVPQHLMGFILGMWFL-----TQAAAFLLGGYVATFTAVPE--
tr | Q8XDS3 ELFIDPVAIAQITRLK---MSGVLTGIYMLA-----TGAVANWLAGVVAQQTTESQ--
tr | Q43390 EALIYTGQLDFFLRECPKGMKTMSTGLLLSTL-----ALGFFFSSVLVTIVEKVTDKA--
tr | Q9FRU2 ESFTAVGQVEFYKQFPENMRSIAGSLFYCGM-----AGSSYLSTLLISIVHNTSEKS--
tr | Q9SZY4 EAFIYTGQLDFFITQSPKGMKTMSTGLFLTTL-----SLGFFVSSFLVSIKRVTS--
tr | Q9WU80 EIFASIPGLEFAYSEAPRSMQGAIMGIFFCLS-----GVGSLGSGLVALLSFPGGWM--
tr | Q9XIF3 EMFTAVGLIEFFYKQSAKGMESFLMALTYSY-----SFGFYFSSVLVSVVNKITSTS--
tr | Q9ZPS2 GVFFFVGRIEFFYEQSPDSMRSLCSAWALLTT-----TLGNYLSSLIITLVAYLS-----
tr | Q9ZTX4 EIFTFIGQLEFFYDQSPDAMRSLCSALSLLTT-----ALGNYLSSFILTVVTSITTRG--

sp | O01840 VKFDYKVEKTDNGATRNVFVVTAREADTLVYAINKK GKILSSCELKSGSYVDVIPGIISD
sp | O07380 -TADVIVIIIFYLVKMAALWWAWSYIPLSTVFGYKYSGKDESLQAALVLEVL-----
sp | P32901 -LFT-GLAVACFISGCLFWLFCFRKYNDTEEMNAMDYEEDEDFLNPISAPKANDIEILE
sp | P36574 -----FFAITGIIIGIIVGIIILLIVKKPILKLMGDVR-----
sp | P36836 APNNYRVVNDGLTQKSDKGENGIRFVNTYSQPINVTMSGKVYEHIASYNASEYQFFTSGV
sp | P36837 -NITDPLETLPVYTNVFGKIGLVTLGVAVVMLLMVPWLKRMIAATPESH-----
sp | P39276 -ISGMAIAAYQRFFSQMGEWTLACVAIIVVLAFAFTRFLFSTPTNMIQESND-----
sp | P46029 GKSIS SIMVKDMENETTYGMTAIRFINTLQENVNISLGTDISLNVGENYGVSA--YRTVQ
sp | P46030 -VFL-AIGLAGFLCAIVMLAQFWNLDKWMENETNERERLDREEEEEANRGIHDVDHP IEA
sp | P46031 -MYT-GIAVTAFIAGIMFWVCFHHYDAMEDEQNQLEFKRNDALTKKDVEKEVHDSYSMAD
tr | Q63422 GPNNYRVV-----
tr | Q8ZLD6 -NITDPLQTLPIYTGVSFKIGLVTLAVTVVMAIMVPWLNRMINTPGTEQ-----
tr | Q8XDS3 -ISGMAIAAYQRFFSQMGEWTLACVAIIVVLAFAFTRFLFSTPTNMVQESND-----
tr | Q43390 -HP--WIAD-DLNKGRLYN--FYWLVAVIVALNFLVFLVFSKWYVYKEKRLADLGIELED
tr | Q9FRU2 -ATGNWLPE-DLNKGRLDL--FYWMIAALEIMNLGYFLLCSKWYKYKEIGSSDLELNQVP
tr | Q9SZY4 -TDVGWLD-NINHGRLDY--FYWLLVILSGINLVVYIICALWFKPTKGDSD--VEKEN
tr | Q9WU80 ---YCPKDFGNINNCQMDR--YFFLLAGIEAVTAVLFLWIAGRYERTRQDPASQRSSSRV
tr | Q9XIF3 VDSKGLWLGENDLNKDRDL--FYWLLAVLSLLNFLSYLFWSRWNIKSSRRNNTNVVGDEN
tr | Q9ZPS2 -GKDCWIPSDNINNGHLDY--FFWLLVSLGSVNIPVVFVFSVKYTHMKV-----
tr | Q9ZTX4 -GKPGWIPN-NLNGGHLDY--FFWLLAALSFFNLVIYVFLCQMYKSKKAS-----

sp | O01840 PNVRLYWGPKNCSGVDPCNTVTTLNAQMGAVHVLHIHPSTTEGDFNLLVRPNSVSIWLSL
sp | O07380 -----
sp | P32901 PMESLRSTTKY-----
sp | P36574 -----
sp | P36836 KGFTVSSAGISEQCRDFESPYLEFGSAYTYLITSQATGCPQVTEFEDIPTNTMMAWQI
sp | P36837 -----
sp | P39276 -----
sp | P46029 RGEYPAVHCKTEDKDFSLNLGLLDFGASYL FVITNSTKQGLQAWKMEDI PANKVSI AWQL
sp | P46030 IVSIKS-----
sp | P46031 ESQYNLEKANC-----
tr | Q63422 -----
tr | Q8ZLD6 -----
tr | Q8XDS3 -----
tr | Q43390 EPDIPMGH-----
tr | Q9FRU2 KHI-----
tr | Q9SZY4 GKGFSVEDC-----
tr | Q9WU80 RG-----
tr | Q9XIF3 I-----

sp | O01840 PQYIIITLGEVLLSVTGLEFAYSQAAPNMKSVLTAMWLLTVFAGNLIDMMISGTRLIPHP
sp | O07380 -----
sp | P32901 -----
sp | P36574 -----
sp | P36836 PQYFLITSGEVVSITGLEFSYSQAPS NMKSVLQAGWLLTVAVGNIIVLIVAGAGQINKQ
sp | P36837 -----
sp | P39276 -----

sp | [P46029](#) PQYALVTAGEVMF SVTGLEFSYSQAPSSMKSVLQAAWLLTVAIGNIIVLVVAQFSGL-VQ
 sp | [O01840](#) ALEFFFYSTLMVIVMGV FILLAMQYTYVEDNDDEITITESEKKDVIALTEIESGTATSDK
 sp | [O07380](#) -----
 sp | [P32901](#) -----
 sp | [P36574](#) -----
 sp | [P36836](#) WAEYILFAALLLVV CVIFAIMARFYTYVNP AEIEAQFEEDEKKNPEKNDLYPSLAPVSQ
 sp | [P36837](#) -----
 sp | [P39276](#) -----
 sp | [P46029](#) WAEFVLF SCLLLV VCLIFS IMGYYYIPIKSEDIQGPEDKQIPHMQGNMINLETKKTKL--
 sp | [O01840](#) KE-
 sp | [O07380](#) ---
 sp | [P32901](#) ---
 sp | [P36574](#) ---
 sp | [P36836](#) TQM

TGF-Beta External Domain Alignment- High Sequences

sp | [O01840](#) -----MEEKSLLQKLR SYPPAVVFMLGNEFCERFSF
 sp | [O07380](#) -----
 sp | [P32901](#) -----MLNHPSQGSDD
 sp | [P36574](#) -----
 sp | [P36836](#) -----MGMSKSLSCFGYPLS IFFIVVNEFCERFSY
 sp | [P36837](#) -----
 sp | [P39276](#) -----
 sp | [P46029](#) MNPFQQNESKETL FSPVSTEETPPRLSSPAKKT PPKICGSNYPLS IAFIVVNEFCERFSY
 sp | [P46030](#) -----MVSSDFENEKQPDV
 sp | [P46031](#) -----MSSIEEQITKSDSD
 sp | [O01840](#) YGMKTILFIYLITEHEFSPSKATFIYHLFTCIAYLTPLIGSIMADSVFGRFKVILYGSSI
 sp | [O07380](#) -----
 sp | [P32901](#) AQDEKQGDFPVI EEEKTQAVTLKDSYVSD DVANSTERYNLSPSPEDEDFEGPTEEEMQTL
 sp | [P36574](#) -----MQNLNKTEK
 sp | [P36836](#) YGM RALLILYFRNF IGWDDNLSTVIYHTFVALCYLTPILGALIADAWLGKFKTIVWLSIV
 sp | [P36837](#) -----
 sp | [P39276](#) -----
 sp | [P46029](#) YGMKAVLTLYFLYFLHWNEDTSTSVYHAFSSL CYFTPILGAAIADSWLGKFKTIIYLSLV
 sp | [P46030](#) VQVLTDEKNI SLDDKYDYEDPKNYSTNYVDDYNPKGLRRPTPQESKSLRRVIGNIRYSTF
 sp | [P46031](#) FIIISEDQSYLSKEKKADGSATINQADEQSSTDELQKSMSTGVLVNGDLYPSPTEEELATL
 sp | [O01840](#) YVVGHVLLSLGAVPFLSYPI-----RSSLDFSGLFVIAFATGC IKPCVSAFAADQ
 sp | [O07380](#) -----MRAILLFYMYAVTKGGLGM
 sp | [P32901](#) RHVGGKIPMRCWLI AIVELS-----ERFSYYGLSAPFQNYMEYGPNDSPKGVLS
 sp | [P36574](#) TFFGQPRGLLTLFQTEFWER-----FSYYGMRAILVYYLYALTTADNAGLGL
 sp | [P36836](#) YTIGQAVTSLSSVNELT DNNDHGTGTPDSL PVHVAVCMIGLLLI ALGTGGIKPCVSAFGGDQ
 sp | [P36837](#) MNTTTPMGMLQQPRPFFMIF-----FVELWERFGYGVQGV LAVFFVKQLGF
 sp | [P39276](#) -----MKTPSQPRAIYYIV-----AIQIWEYFSFYGM RALLILYLTHQLGF
 sp | [P46029](#) NVLGHVIKLSA FPI LGGKV-----VHTVLSLVGLCLIALGTGGIKPCVA AFGGDQ
 sp | [P46030](#) MLCICEFAERASYSTTGIL-----TNYIQRRIDPDS PHGWGAPPPGSPDASAGALG
 sp | [P46031](#) PSVCGTIPWKAFIIIIIVELC-----ERFAYYGLTVPFQNYMQFGPKDATPGALN
 sp | [O01840](#) FTEDQKDLRSQFFSFFYFAINGGSLFAIIITPILRGRVQC---FGNAHCFPLAFGVPGVL
 sp | [O07380](#) SQTTAASIMSIYGS--LVYLST--LVGGWLSDRVWGS-----RKT VFYGGVL
 sp | [P32901](#) LNSQGATGLSYFFQ-FWCYVTP--VFGGYVADTFWGK-YNT--IC---CGTAIYIAGIFI

sp | [P36574](#) | PKAQAMAIVSIYGA--LVYLST--IVGGVADRLPGA-----SRTIFLGGIL
 sp | [P36836](#) | FEEGQEKQRNRFSSIFYLAINAGSLSTIITPMVRVQ-QCGIHVKQA-CYPLAFGIPAIL
 sp | [P36837](#) | SQEQAFVTFGAFAA--LVYGLI--SIGGYVGDHLLGT-----KRTIVLGAIV
 sp | [P39276](#) | DDNHAISLFSAYAS--LVYVTP--ILGGWLADRLGN-----RTAVIAGALL
 sp | [P46029](#) | FEEKHAEERTRYFSGFYLAINGSLISTFITPMLRGDVQC---FGED-CYALAFGVPGLL
 sp | [P46030](#) | KGLQAASALTNLLT-FLAYVFP--LIGGYLGDSTIGRWKA---IQ---WGVFFGFVAHLF
 sp | [P46031](#) | LGETGADGLSNFFT-FWCYVTP--VGAALADQFLGR-YNT--IV---CSAVIYFIGILI
 :
 sp | [O01840](#) | MLLALILFLMGWSMYKKHPPSKENVGSKVVAVIYTSLRKMGVGGASRDKP-VTHWLDHAAP
 sp | [O07380](#) | IMLGHIVLALPAGVTVLYRS-IALIVVGTGLLKNVSDMVGGLYSVEDPRRDAGFSIFVF
 sp | [P32901](#) | LFITSIPSVGNRDSAIGGFI-AAIILIGIA--TGMIKANLSVLIADQLPKRKPSI--KVL
 sp | [P36574](#) | ITLGHIALATPFGLSSLFVA-LFLIILGTGMLKPNISNMVGHLYSKDDSRRTGFINIFVV
 sp | [P36836](#) | MAVSLIVFIIIGSGMYKKFKP-QGNILSKVVKCICFAIKNRFRHRSKQFPKRAHWLDWAKE
 sp | [P36837](#) | LAIGYFMTGMSLLKPDILFIALGTIAVGNGLFKANPASLLSKCYPKDPRLDGAFTLFYM
 sp | [P39276](#) | MTLGHVVLGIDTNSTFSLYLALAIICGYGLFKSNISCLLGELYDENHRRDGGFSLLYA
 sp | [P46029](#) | MVIALVVFAMGSKMYKKPPP-EGNIVAQVVKCIWFAISNRFKNRSEDIKPRQHWLDWAAE
 sp | [P46030](#) | FIFASIPQAIENANAGLGLCVIAIITLSAG--SGLMKPNLLPLVLDQYPEERDMV--KVL
 sp | [P46031](#) | LTCTAIPSVIDAGKSMGGFV-VSLIIGLG--TGGIKSNVSPLMAEQLPKIPPYV--KTK
 :
 sp | [O01840](#) | EHSQKMIDSTRGLLNVAVIFCPLIFFWALFDQ-----GSTWVLQARRLDGRVGHFSI
 sp | [O07380](#) | GINLGSIIAPWLVPAWAAQGFVHIFGSQNLNFHAGFSLAAGMFFGLVQYVVLGGKYLSTE
 sp | [P32901](#) | KSGERVIVDSNITLQNVFMF--F--YFMINV-----GSLSLMATTELEYHKGFWAA
 sp | [P36574](#) | GINMGSLIAPLIVGTVGGQVNYHLGFSLAAI-----GMIFALFAYWYGRLRHFPEI
 sp | [P36836](#) | KYDERLIAQIKMVTBFLYIPLPMFWALFDQ-----GSRWTLQATTMSGRIGILEI
 sp | [P36837](#) | SINIGSLIALSLAPVIADRFYGVSVTYNLCGA-----GLIIALLVYIACRGMVKDI
 sp | [P39276](#) | AGNIGSIAAPIACGLAAQWYGVHVG-ALAG-----GGMFIGLLIFLSGHRHFQST
 sp | [P46029](#) | KYPKQLIMDVKTLTRVFLYIPLPMFWALLDQ-----GSRWTLQATKMNGNLGFFVL
 sp | [P46030](#) | PTGESIILDREKSLSRITNV--F--YLAINI-----GAFLQIATSYCERRVGFWLA
 sp | [P46031](#) | KNGSKVIVDPVVTTSRAYMI--F--YWTINV-----GSLSVLATTSLLESTKGFVYA
 :
 sp | [O01840](#) | LPEQIHAINPVCVLLIVPFIPEGWVYPALRKI-TRVTPLRKMVAVGGLLTAFSFA-----
 sp | [O07380](#) | SLTPNDPIDKGDLLNVIKWVVI I I I A I V A I L - A A M A G V G Q L S V D N V I T L L T I L A -----
 sp | [P32901](#) | YLLPFCFF-WIAVVTLIFGKKQYIQRPIGDKVIAKSFKVC--WILTKNKFDFN-----
 sp | [P36574](#) | GREPSNPMDSKARRNFLTITLIVVIVAIIGFFLLYQASPNFINNFINVLSIIGI-----
 sp | [P36836](#) | QPDQMOTVNTILIIILVPIMDAVVYPLIAKCGLNFTSLKKMTIGMFLASMAFV-----
 sp | [P36837](#) | GSEPDFRPMFSKLLYVLLGSVVMIFVCAWL-----MHNVEVANLVIVLSIV-----
 sp | [P39276](#) | RSMDDKALTSVKFALPVWSVLVVMLCLAPVF--FTLLENDWSGYLLAIVCL-----
 sp | [P46029](#) | QPDQMQLNPLLVLIIFIPFLDLVIYRLISKCGINFTSLRKMVAVGMVLAFLAFA-----
 sp | [P46030](#) | FFVPMILYIIVPIFLFIVKPKLKIKPPQGVMTNVVKILAVLFSGNFIKRLWNGTFWDHA
 sp | [P46031](#) | YLLPLCVF-VIPLIILAVSKTAFTSTLLPPVPSLFLVVKSSLLKTNLISKK-----
 :
 sp | [O01840](#) | -----IAGVLQKVNEMEFPPSLGRIYLQRVGNESLISDFRYKSDGRLIGDG-MLPKG
 sp | [O07380](#) | -----IALPIYFVMMFRSSKVTKIELGIHLLPVSLKNRLFKKGYKRLKQII--QLEL
 sp | [P32901](#) | -----AAKPSVHPEKNY---PWNDKFVDEIKRALAACKVFIYPIYWTQYGTM--ISSF
 sp | [P36574](#) | -----VVPIIYFVMMFTSKKVESDERRKLTAYIPLFLSAIVFWAIEEQSSTIIAVWGES
 sp | [P36836](#) | -----AAAILQVEIDKTLVPVFPKANEVQIKVLNVGSENMIISLPGQTVTLNQMSQTFNEF
 sp | [P36837](#) | -----VTIIFFRQAFKLDKTRNKMFVAVFLMLEAVVFYILYAQMPTSLNFFFAIN
 sp | [P39276](#) | -----IAAQI IARMMIKFPEHRRALWQIVLLMFVGTFLFWVLAQQGGSTISLFDIDRF
 sp | [P46029](#) | -----AAATVEIKINEMAPPQPGSQEILLQVLNLADDEVKLTIVLGNNNNSLLADSIKSF
 sp | [P46030](#) | RPSHMEARGTIYYNSKKSAITWSDQWILDIKQTFDCKIFLYYIIFNLADSGLSVETS
 sp | [P46031](#) | -----LNHLALLLLERYVKDQWDDLFIDELKRALRACKTFLFYPIYWVCYQGM--TNNL
 :
 sp | [O01840](#) | RTELDAGIYTFNTGLKN-----ESQEIDISTPNKGYVMAVFRKDAVEVVKFDY
 sp | [O07380](#) | AIKRQSFIIILIALIIMA-----S-ILIPNKVIAKHLK-----
 sp | [P32901](#) | IT--QASMMELHGIPND-----FLQAFDSIALIIFIFIPIFE-----K-
 sp | [P36574](#) | RS-NLDPTWFGITFHID-----PSWYQLLNPLFIVLLSPIFV-----
 sp | [P36836](#) | MTFNEDTLTSINITSGS-----QVTMITP-SLEAGQRHTLLVWAPNNY--RVVNDGL

sp | [P36837](#) | NV--HHEILGFSINPVS-----FQALNPFWVVLASPILA-----
sp | [P39276](#) | VN----RQAFNIEVPTA-----LFQSVNAIAVMLAGVVLA-----
sp | [P46029](#) | QKTPHYSKIHLNTKSQDFYFHLKYHNLSIYTEHSVEERNWYSLI IREDGKS SISSIMVKDM
sp | [P46030](#) | LI----GAMKLDGVPND-----LFNPNPLT I I I L I P I L E-----
sp | [P46031](#) | IS--QAGQMOTGNVSN-----LFQAFDSIAL I I F I P I C D-----N-

::

sp | [O01840](#) | KVEKTDNGATR V F V V T A R E D A D T L V Y A I N K K G K I L S S C E L K S G S Y V D V I P G I I S D P N V R L
sp | [O07380](#) | -----LVL LVFYWIGL-----NLIPFSTFVLSFLFLDYIKHM-----
sp | [P32901](#) | -----FVYPFIRRYT-----PLKPITKIFFGFMFGSFAMTWAA-----
sp | [P36574](#) | -----RLWNKLG ERQP-----STIVKFGLGLMLTGISYLI MT-----
sp | [P36836](#) | TQKSDKGENGIRFVNTYSQP INV---TMSGKVYEHIA--SYNASEYQFFTSG-----VKG
sp | [P36837](#) | -----GIYTHLGNK GK D-----LSMPMKFTLGMFMCSLGLFLTAA-----
sp | [P39276](#) | -----WLASPE SRGNST-----LRVWLKFAFGLLLMACGFMLLA-----
sp | [P46029](#) | ENETTYGMTAIRFINTLQENVNI---SLGTDISLNVGENYGVSA YRTVQ-----RG
sp | [P46030](#) | -----YGLYPLL NKFKI---DFKPIWRICFGFVVC SFSQIAGF-----
sp | [P46031](#) | -----I I Y P L L R K Y N I---P F K P I L R I T L G F M F A T A S M I Y A A-----

sp | [O01840](#) | YWGPKNCSGVDPCNTVTLNAQMGAVHVLHIHPSTTEGD--FNLLVRP--NSVSILWSLPQ
sp | [O07380](#) | -----FKKEGEQAKKTKE---KSRIHHGIEIP-----LFLRQLIINIF
sp | [P32901](#) | -----VLQSFVYKAGPWY---NEPLGHNT-----P-NHVHVCWQIPA
sp | [P36574](#) | -----LPGLLNGTSG-----RASALWLVL M
sp | [P36836](#) | FTVSSAGISEQCRRDFESPYLEFGSAYTYLITSQATGCPQVTEFEDIPPNTMMAWQIPQ
sp | [P36837](#) | -----AAGMWFADAQ-----GLTSPWFIVLV
sp | [P39276](#) | -----FDARHAAADGQAS---MGVMISGL-----ALMGFAELFIDPV
sp | [P46029](#) | EYPAVHCKTEDKDFSLNLGLLDFGASYLFVITNSTKQGLQAWKMEDI PANKVSI AWQLPQ
sp | [P46030](#) | -----VLQKQVYEQSPCG---YYATNCD-----SPAPITAWKASSL
sp | [P46031](#) | -----VLQAKIYQRGPCY---ANFTDTCV-----S-NDISVWIQIPA

sp | [O01840](#) | Y I I I T L G E V L L S V T G L E F A Y S Q A A P N M K S V L T A M W L L T V F A G N L I D M M I S G T R L I P - H P A
sp | [O07380](#) | T L I I L E G E T L F D E N G V E V N I A E H P V Q Y T E L N I N L L N K D S I D L W A D W I Q S V A K Y L L N I M Y
sp | [P32901](#) | Y V L I S F S E I F A S I T G L E Y A Y S K A P A S M K S F I M S I F L L T N A F G S A I G C A L S P V T V D P K F T W
sp | [P36574](#) | F A V Q M A G E L L V S P V G L S V S T K L A P V A F Q S Q M M A M W F L A D S T S Q A I N A Q I T P L F K A A T E V H
sp | [P36836](#) | Y F L I T S G E V V F S I T G L E F S Y S Q A P S N M K S V L Q A G W L L T V A V G N I I V L I V A G A G Q I N - K Q W
sp | [P36837](#) | Y L F Q S L G E L F I S A L G L A M I A A L V P Q H L M G F I L G M W F L T Q A A A F L L G G Y V A T F T A V P D N I T
sp | [P39276](#) | A I A Q I T R L K M S G V L T G I Y M L A T G A V A N W L A G V V A Q Q T T E S Q I S G M A I A A Y Q R F F S Q M G E W
sp | [P46029](#) | Y A L V T A G E V M F S V T G L E F S Y S Q A P S S M K S V L Q A A W L L T V A I G N I I V L V V A Q F S G L -- V Q W
sp | [P46030](#) | F I L A A A G E C W A Y T T A Y E L A Y T R S P A L K S L V Y A L F L V M S A F S A A L S L A I T P A L K D P N L H W
sp | [P46031](#) | Y V L I A F S E I F A S I T G L E F A F T K A P S M K S I I T A L F L T N A F G A I L S I C I S S T A V N P K L T W

sp | [O01840](#) | L E F F F Y S T L M V I V M G V F I L L A M Q Y T Y V E D N D D E I T I T E - S E K K D V I A L T E I E S G T A T S D K
sp | [O07380](#) | T A D V I V I I I F Y L V K M A A L W W A W S Y I P L S T V F V G Y K Y S G K D E S L Q A A L E V L -----
sp | [P32901](#) | L F T G L A V A C F I S G C L - F W L C F R K Y N D T E E E M N A M D Y E E E D E F D L N P I S A P K A N D I E I L E P
sp | [P36574](#) | F F A I T G I I G I I V G I I L L I V K K P I L K L M G D V R -----
sp | [P36836](#) | A E Y I L F A A L L L V V C V I F A I M A R F Y T Y V N P A E I E A Q F E E - D E K K K N P E K N D L Y P S L A P V S Q
sp | [P36837](#) | D P L E T L P V Y T N V F G K I G L V T L G V A V V M L L M V P W L K R M I A T P E S H -----
sp | [P39276](#) | T L A C V A I I V V L A F A T R F L F S T P T N M I Q E S N D -----
sp | [P46029](#) | A E F V L F S C L L L V V C L I F S I M G Y Y Y I P I K S E D I Q G P ----- E D K Q I P H M Q G N M I N L - E T K K
sp | [P46030](#) | V F L A I G L A G F L C A I V M L A Q F W N L D K W M E N E T N E R E R L -- D R E E E E E A N R G I H D V D H P I E A
sp | [P46031](#) | M Y T G I A V T A F I A G I M - F W V C F H H Y D A M E D E Q N Q L E F K R N D A L T K K D V E K E V H D S Y S M A D E

sp | [O01840](#) | KE-----
sp | [O07380](#) | -----
sp | [P32901](#) | MESLRSTTKY
sp | [P36574](#) | -----
sp | [P36836](#) | TQM-----
sp | [P36837](#) | -----
sp | [P39276](#) | -----

sp | [P46029](#) TKL-----
 sp | [P46030](#) IVSIKS-----
 sp | [P46031](#) SQYNLEKANC

Auto-inducer Binding Domain Alignments – Total Sequences

sp | [O30919](#) -----MEPDFQ-----DAYHAFRTAEDEHQLFREI
 sp | [O54452](#) -----MSPSHAEQF--SFFLLSGPDLRIADIAGSGN
 sp | [O87971](#) -----I
 sp | [P07026](#) -----MQDKDFFSWRRTMLLRFORMETAEEVYHEI
 sp | [P12746](#) -----MKNINADDTY--RIINKIKACRSNNDINQCL
 sp | [P25084](#) -----MALV-----DGFLELERSGKLEWSAIL
 sp | [P33905](#) -----MQHWLD-----KLTDLAAIEGDECILKTGL
 sp | [P54292](#) -----MRNDGGFLLWWDGLRSEMQUIHDSQGVFAVL
 sp | [P54293](#) -----MFSFFLENQITITDTLQTYI
 sp | [P54295](#) -----MIIDYFDNESINEDIKNYI
 sp | [P54303](#) -----MELGQQ-LGWDAYFYFISFARTMDMQEFTAVA
 sp | [P74946](#) -----MYKIL-----RLIQENQQITSHDDLQENVL
 sp | [Q03316](#) -----MKEESSAVSNLVDFDLSESASAKSKDDVLLLF
 sp | [Q44059](#) -----MKQDQLL-----EYLEHFTSVTDGDRLAELI
 tr | [Q93UX7](#) MSRVAPAGRTVAPVTRGASRRGEWEARSCAVGEKSVGDGLRSLIDMTEAAHDERMIKSAL
 tr | [Q98A91](#) -----MQLVFE-----TLLEQLSLSVDEVDFHNAL
 tr | [Q9AM51](#) -----MELRWQ-----DAYQQFSAAEDEQQLFQRI
 tr | [Q9ANB1](#) -----MDDQSHGKMPLASSRRESRGPDLFSFIECAIQTRSIAALFDLL
 tr | [Q9I4J9](#) -----MTDYASLL--FDSVARLRKAATTEAVCREM
 tr | [Q9KJ71](#) -----MIENTYSE--KFESAFEQIKAAANVDAAI
 tr | [Q9KL60](#) -----MLKLSRNKIEQKEPMPAVDLITLISQLEESHFSFSTVQDIV
 tr | [Q9L3I8](#) -----MSNSFFNNTSINISIKNYL
 tr | [Q9L6U5](#) -----MERQQDFMHWWNDRMEMLKLQESAQV FALL
 tr | [Q9L6U7](#) -----MDSRLKY-----SDFIPADESKSNVCLKLEL
 tr | [Q9R9G8](#) -----MELGQQ-LGWDSYFYFISFARTMDMQEFTAVT
 tr | [Q9RH45](#) -----MIINFFDNESINEDIKNYI
 tr | [Q9RMS5](#) -----MHDERE-----GYLEILSRITTEEEFFSLV
 tr | [Q9XDD1](#) -----MIINFFDNESINEDIKNYI
 tr | [Q9ZIU0](#) -----MELRWQ-----DAYQQFSAAEDEQQLFQRI

sp | [O30919](#) AAIARQLGFDYCCYGARM----PLPVSKEPAVAIFDTPAGWWMQHYQASG---FLDIDPTV
 sp | [O54452](#) DAGRSRPHLCDIAYGSPC---DLAGATDSNPLMLTYPPPEWVKQYRDRD---YFSIDPVV
 sp | [O87971](#) DRKLERYSRPRYTYMV-----IDKKNPVDVFI VTSYPDEWADIYTSQN---YQHIDPIV
 sp | [P07026](#) ELQAQQLEDYDYSLCVRH----PVPFTRPKVAFYTNYPEAWVSYQAKN---FLAIDPVL
 sp | [P12746](#) SDMTKVMVHCEYYLLAIY----PHSMVKS DISILDNYPKKWRQYYDDAN---LIKYPDIV
 sp | [P25084](#) QKMASDLGFSKILFGLLP----KDSQDYENAFIVGNYPAAWREHYDRAG---YARVDP TV
 sp | [P33905](#) ADIADHFGFTGYAYLH-----IQHRHITAVTNYHRQWQSTYFDKK---FEALDPVV
 sp | [P54292](#) EKEVRRLGFDYYAYGVRH----TIPFTRPKTEVHGTYPKAWLERYQMQN---YGAVDPAI
 sp | [P54293](#) QRKLSPLGSPDYAYTV-----VSKKNPSNVLIISSYPDEWIRLYRANN---FQLTDPVI
 sp | [P54295](#) QRRIKTYGDLCSYLV-----MNKKTPLHPTIISNYPLDWVKYKKNKNS---YHLIDPVI
 sp | [P54303](#) LRALRELRFDFFRYGMCS---VTPFMRPRTYMYGNYPEDWVQRYQAAN---YAVIDPTV
 sp | [P74946](#) NGLNNLIGHEFFLFLGLSF----QPTLKTSETLVTDNYPNSWRQQYDESG---FMHIDPIV
 sp | [Q03316](#) GKISQYFGFSYFAISGIP----SPIERIDSYFVLGNWSVGFWDREYRENN---YVHADPIV
 sp | [Q44059](#) GRFTLGMGYDYRFALII---PMSMQRPKVVLFNQCPDSWVQAYTANH---MLACDPII
 tr | [Q93UX7](#) KTFAHACGFERFAYLQ-----TEGSEIRTFNSYPEEWQGVYLESH---YSRIDPVV
 tr | [Q98A91](#) ASAAGAFDIPAFAYLS-----LVSDRVTKPRLISNYHSGWTSHYLRHQ---YERIDPVI
 tr | [Q9AM51](#) AAYSKRLGFEYCCYGIRV----PLPVSKEPSVAIFDTPDGWMAHYQAHN---YIEIDSTV
 tr | [Q9ANB1](#) VNFASNEGFDKVAYGALSCSNERRLPEYLPPTINFPSDWQCRYAEQE---YQAIDPVV
 tr | [Q9I4J9](#) VEFAARLGFDRVIVCSLF----PRPDEDELIDELFFVHGDWAEGRSAQERDAYLLHCPVT

tr | Q9KJ71 RILQAEYNLDFVTTYHLAQ----TIASKIDSPFVRTTYPDAWVSRYL LNS---YVKVDP IV
tr | Q9KL60 RQRAHFYGYDKIVFFS-----AHSTLDGIIERIYWIEGDWFDDGENIDAATYIKYCPIT
tr | Q9L3I8 EKNLKVFNNIKYAYAI-----MNKKNPNDFAIISNR-MEWDFDYTKNN---LQFIDPVL
tr | Q9L6U5 EREVQYLGFDYAYGIRH----SIPFTRPKTEIYGSYPQPMERYQQQN---YAAIDPSI
tr | Q9L6U7 EKLLGDLQGASYAYFA-----APRNREVAPLIVSNYPARWLKAYKNAN---YHLIDPII
tr | Q9R9G8 LRVLRELRFDFYAYGMS----VTPFMRPRTCMYGNYPEDWVQRYQAAN---YAVIDPTV
tr | Q9RH45 QRRIKAYGNIRYSYLL-----MNKKVPLHPAIIISNYPLDWVKKYKKN---YHLIDPVI
tr | Q9RMS5 LEICGNYGFEFFSFGARA----PFPLTAPKYHFLSNYPGEWKSRYISED---YTSIDPVI
tr | Q9XDD1 QRRIKAYGNIRYSYLL-----MNKKVPLHPAIIISNYPLDWVKKYKKN---YHLIDPVI
tr | Q9ZIU0 AAYSKRLGFYCCYGIRV----PLPVS KPAVAIFDTPDGWMAHYQAQN---YIEIDSTV

*

sp | O30919 RAGASSSDLIVWPVSIR-----DDAARLWSDARDAGLN-IGVARSSWTAHGAFGLLTLA
sp | O54452 RLGRRGFLPVEWSASGWDS----GRAYGFKEAMAFGVGRQGVTL PVRGPQGERSLFTVTV
sp | O87971 LTAFKRISPFAWDENITILS--DLKSSKIFALS KKYNIV-NGFTFVLHDHMMNNLAML SLI
sp | P07026 NPENFSQGLHMWNDDLF-----SEAQPLWEAARAHGLR-RGVTQYLMLPNRALGFLSFS
sp | P12746 DYSNSNHSPINWNIFENNAV--NKKSPNVIKEAKTSGLI-TGFSFPIHTANNGFGMLSFA
sp | P25084 SHCTQSVLPFWEPSIYQT---RKQHEFEEASAAGLV-YGLTMPLHGARGELGALSLS
sp | P33905 KRARSRKHI FTWSGEHERPTL-SKDERAFYDHASDFGIR-SGITIPIKTANGFMSMFTMA
sp | P54292 LNGLRSSEM VVWSDSLF-----DQSRMLWNEARDWGLC-VGATLPIRAPNNLLSVLSVA
sp | P54293 LTAFKRTSPFAWDENITILMS--DLRFTKIFSLSKQYNIV-NGFTYVLDHMMNNLALLSVI
sp | P54295 LTAKDKVAPFAWDDNSVINK--KSTDSAVFKLAREYNIV-NGYTFVLHDNSNNMATLNIS
sp | P54303 KHSKVSSPILASNELF-----RGCPDLWSEANDSNLR-HGLAQPSFNTQGRVGLSLA
sp | P74946 KYSITNFLPIRWDDAKRVN----NDGRVIFEEARCNGLK-AGFSIPIHGLRGEFGMISFA
sp | Q03316 HLSKTCDHAFVWSEALRDQKL-DRQSRVMDEAREFKLI-DGFSVPLHTAAGFQSI VSFSG
sp | Q44059 QLARKQTLPIYWNRLDERARFLQEGSLDMGLAAEFGLR-NGISFPLHGAAGENGILSFI
tr | Q93UX7 WEAKTRMEIFPWTADWPSRG-KSELRQFRDQAINHGIR-SGVTIPVEGSFSGSTMMLTFA
tr | Q98A91 EWARCSECFQWGPGLGHAGI-STRQQQLFDEAAEFGIC-CGLTIPLVDRRGVAAAMTFA
tr | Q9AM51 RDGALSTNMIVWPDVDQ-----IEPCPLWQDARDFGLS-VGVAQSSWAARGAFGLLSIA
tr | Q9ANB1 RRTAMLPRPFLWDELTSRYEL-QPRELRLVREAKEAGLK-HGMSVPLFWSQGRSAGVVSFA
tr | Q9I4J9 RHILELDEPFFWKSSEDPE--ERMTRYIRVGRVLDLQGV-NGMQVPVFGRNLEGAVSFA
tr | Q9KJ71 KQGFERQLPFDWSEVEPT----PEAYAMLVDAQKHGIGNGYSIPVADKAQRALLSLN
tr | Q9KL60 RHIIETDRPFFWTKKPDVN----REQYRVVAKPKGSGI--HGLQIPIFGHLGLEGAVSLG
tr | Q9L3I8 ITASCCFTPFLWDENIMISS--GLKMPKIFNMAKNYDVI-NGYTFVLHDHNNHNLVLSII
tr | Q9L6U5 LNGLRSTEMVVWNDVLF-----GKGSVLWGEAQDWGLR-VGATVPIRTPDGALS VLSVA
tr | Q9L6U7 HHGLKSCAPFSWSDALQAAP--CERSRELFRSSQYRIC-TGATFTLHDAGGMFSSLSLC
tr | Q9R9G8 KHSKVSSAPILWSNELF-----RGCPDLWSEANDSNLC-HGLAQPSFNAQGRVGM LSLA
tr | Q9RH45 LTAKGKVAPFAWDDNSVINI--KSTDSAVFNLAREYNIV-NGYTFVLHDNNNNMATLNVS
tr | Q9RMS5 RHGLLEYTPLIWNGEDF-----QENRFFWEEALHHGIR-HGWSIPVRGKYGLISMLSLV
tr | Q9XDD1 LTAKGKVAPFAWDDNSVINI--KSTDSAVFNLAREYNIV-NGYTFVLHDNNNNMATLNVS
tr | Q9ZIU0 RDGALNTNMIVWPDVDR-----IDPCPLWQDARDFGLS-VGVAQSSWAARGAFGLLSIA

*

sp | O30919 RHADPLTAAE--LGQLSIATHWLANLAHTLMSPFL-----VPQLVPESN-AVLTTRRE
sp | O54452 SNHPDAYWRQ-FRMDSMRDLQFLAHLHDRAMVLS-----GMRKVADLPRLSRRE
sp | O87971 MDNNADKGLNSRIESDKDR LQMNLIKIEHKMLMLEQNKLGVSNGKNTDTS GK-GILSPRE
sp | P07026 RCSAREIPIL--SDELQ LKMQLLVRESLMALMRLN-----DEIVMTPE-MNFSKRE
sp | P12746 -HSEKDNIDSLFLHACMNIPLIVPSLVDNYRKIN-----IANNKSNNDLTKRE
sp | P25084 VEAENRAEANRFMESVLP LWM LKDYALQSGAGLA-----FEHPVSKPVVLT SRE
sp | P33905 SDKPVIDLDR---EIDAVAAAATIGQIHARISFLR-----TTPTAEDA AWLDPKE
sp | P54292 RDQQNISSFE--REEIRLRLRCMIELLTQKLTDL E-----HPMLMSNP--VCLSHRE
sp | P54293 IKGNDQTALEQRLAAEQGTMQMLLIDFNEQMYRLAGTE-GERAPALNQSADK-TIFSSRE
sp | P54295 NGSDDSISFDERIEINKEKIQMLLIITHEKMLGLY-QSNSDKNENRNTQIER-DIFSPRE
sp | P54303 RKDNPI SLQE--FEALKVVTKAFAAAVHEKISELE-----SDVRVFNTD-VEFSGRE
sp | P74946 TSDTKSYDLN---QOSIHTSQLIVPLLAHNIGNIT-----RYHKDAKPRAVLTARE
sp | Q03316 AEKVELSTCD--RSALYLMAAYAHSL LRAQIGNDA-----SRKIQALPMITTR E
sp | Q44059 TAERASSDLL--LESSPILSWMSNYIFEAAIRIV-----RVSLREDDPQEALTDRE
tr | Q93UX7 SSAPAADVSK---LQDGQKAVRAVLAIH YRLKMLA-----ATTIVPPKRLLSPRE
tr | Q98A91 ADRLDPTFLR-VAEQYEQALEIMAMCFHIGVRRKL-----ARGLAVDG-VSLTPRE

tr | Q9AM51 RHADRLTPAE--INMLTLQTNWLANLSHSLMSRFM-----VPKLSPAAG-VTLTARE
tr | Q9ANB1 SPFDDADPQD---RMAHLT--TLASAFHNAVAQIT-----PPLDESCETDIPLTPRE
tr | Q9I4J9 GERLDLNAGC-----KLAAQAFCPILFFALKRLR-----GPVSAGDAQALSERE
tr | Q9KJ71 ARIPADEWTE---LVRRRCRNEWIEIAHLIHRKAVY-----ELHGENDPVPALSPRE
tr | Q9KL60 GKAISSPRA-----RCELSLLSTYAFFAARRLL-----ESSDPNRSALLSKRE
tr | Q9L3I8 MDKSCDDDIKIIIVDKKNDLQMLLLTTHEKLITLY--QEINDTHQFNKKNQK-EILSKRE
tr | Q9L6U5 RRNTAISAVE--REVIRLRLRCIIELAAGRLSELG-----HSNTQSDP--QELSARE
tr | Q9L6U7 -NSGPQAEFERRMADQQGHLQMALIRFHGRLLSLR-----AMDELFPPEPQTGPLSARE
tr | Q9R9G8 RKDNPI SLQE--FEALKLMTKAFAAAAIHEKISELE-----SDVRVFNTD-VEFSGRE
tr | Q9RH45 SGDDDSIFFDESI EVNKEKI QMLLIFIHDKMLGLYNKSHHENNTLNKKENKR-EIFSPRE
tr | Q9RMS5 RSSESIAATE--ILEKESFLWITSMLQATFGDLL-----APRIVPESN-VRLTARE
tr | Q9XDD1 SGDDDSIFFDESI EVNKEKI QMLLILFHDKMLGLYNKSHHENNTLNKKENKR-EIFSPRE
tr | Q9ZIU0 RHADRLTPAE--INMLTLQTNWLANLSHSLMSRFM-----VPKLSPAAG-VTLTARD

: ::

sp | O30919 REVL CWTGEGKTAYEIGQILRISERTVNFHVNNVLLKLAATNKVQAVVKA IATGLI----
sp | O54452 LQCLEMTANGLLAKQICARLSISVSAVQLYLASARRKLTVA TTSEQLLGP RRSN-----
sp | O87971 NEVLHWASMGKTYPEIALIAGITTRTVKHHMGNVVKLGVINARQAIRLGVELELIKPV L
sp | P07026 KEILRWTAEGKTSAEIAMILSISENTVNFHQKNMQK KINAPNKTQVACAAAATGLI----
sp | P12746 KECLAWACEGKSSWDISKILGCERTVTFHLTNAQMKLN TTNRCQSSISKAILTGAIDCPY
sp | P25084 KEVLQWCAIGKTSWEISVICNCSEANVNFHMGNI RRKFGVTSRRVAAIMAVNLGLITL--
sp | P33905 ATYLRWIAVGKTMEEIADVEGVKYNSVRVKLREAMKR FFDVRSKAHLTALAIRRKL I----
sp | P54292 REILQWTADGKSSGEIAIILSISESTVNFH HKNIQK KFDAPNKT LAAAYAAALGLI----
sp | P54293 NEVLYWASMGKTYAEIAAITGISVSTVKFHIKNV VVKLGVSNARQAIRLGVELDLIRPAA
sp | P54295 NEILYWASVGKTYAEISII LGIKRSTVKFHI GNVVVKLGVLNAKHAIRLGIELKLIKPI-
sp | P54303 CDVLRWTADGKTSEEIGVIMGVCTDTVNYHHRNIQRK IGASNRVQASRYAVAMGYI----
sp | P74946 VQCLAWAAEGKSAWEIATIINTSERTVKFHF SNACKKLGATNRYQAITKAILGGYINPYL
sp | Q03316 REI IHWCAAGKTAIEIATILGRSHRTIQNVILNIQRK LNVVNT PQMIAESFRLRIIR--
sp | Q44059 TECLFWASEGKTSGETAILTGITERTVNYHLNQVTRKTGSMNRYQAI AKGVSSGILLPNL
tr | Q93UX7 AMCLKWAAKGTAPETAILTGINPRTVQHYLDKAREKLEAATVPHLVAI AKDHGLL----
tr | Q98A91 YECLLEWTAKGKSAWEIGCILGIKERTAAFHLDN AKKKGVRTKNQAVTLLASSRSSIL--
tr | Q9AM51 REVL CWTAEGKTACEIGQILSISERTVNFHVNNI LEKLGATNKVQAVVKAISAGLIEAP-
tr | Q9ANB1 TECLYWVAEGKSAWVIGQLVNVTDN TVNFHMKNVIRKLGAA NRTNAVAKATRRGII----
tr | Q9I4J9 REVLQWIALGKQQA EVAAILMISERTVENHLRAARRRLGAA STAQAVARALRLGDIEV--
tr | Q9KJ71 IECLHWTALGKDYKDISVILGISEHTTRDYLKTARFKLGCATI SAAASRAVQLRIINP--
tr | Q9KL60 KEVLSLTALGRRQADIAIALGVSPRTIENHLRNARLKLGGATTAETIRIAIQRGDINSHS
tr | Q9L3I8 NEILYWASMGKSYQEI ALILGIKLT TVKYHVGNAVVKLGV TNAKHAIRLGVELKLIRPIL
tr | Q9L6U5 REILQWTADGKSSGEIALILGISVNTVNFHLKVIQK KFGAANKT LAAAYA AVLGLI----
tr | Q9L6U7 LGVLLKWVMMGKAYREIAVIC AISERTVKFHMSNISGKLQVCNAKQAVYEAQRQGIL----
tr | Q9R9G8 CDVLRWTADGKTSEEIGVIMGVCTDTVNYHHRNIQRK IGASNRVQAVSYAVAMGYI----
tr | Q9RH45 NEILYWASVGKTYSEIAIILGIKSTVKFHI GNIVRKLGVLNAKHAIRLGIELQLIKPI-
tr | Q9RMS5 TEMLKWTA VGKTYGEIGLILSIDQRTVKFHI VNAMRKLNSSNKAEATMKAYAIGLLN---
tr | Q9XDD1 NEILYWASVGKTYSEIAIILGIKSTVKFHI GNIVRKLGVLNAKHAIRLGIELQLIKPI-
tr | Q9ZIU0 REVL CWTAEGKTACEIGQILSISERTVNFHVNNI LEKLGATNKVQAVVKAISAGLIEAP-

: * :

sp | O30919 -----
sp | O54452 -----
sp | O87971 V-----
sp | P07026 -----
sp | P12746 FKN-----
sp | P25084 -----
sp | P33905 -----
sp | P54292 -----
sp | P54293 SAAR-----
sp | P54295 -----
sp | P54303 -----
sp | P74946 -----
sp | Q03316 -----

```

sp | Q44059      EQVVVTNFPKLMQ
tr | Q93UX7      -----
tr | Q98A91      -----
tr | Q9AM51      -----
tr | Q9ANB1      -----
tr | Q9I4J9      -----
tr | Q9KJ71      -----
tr | Q9KL60      I-----
tr | Q9L3I8      PDAE-----

```

Auto-inducer Binding Domain Alignments – High Sequences

```

sp | O30919      -----MEPDFQDAYHAFRTAEDEHQLFREIAAIAARQLGFDYCCYGARMPLPV-SKPAV
sp | O54452      ---MSPSHAEQFSFFLLSGPDLRIADIAGSGNDAGRSRPHLCDIAYGSPCDLAGATDSNP
sp | O87971      -----IDRKLERYDSPRYTYMVIDKKN---PVDV
sp | P07026      --MQDKDFFSWRRTMLLRFORMETAEEVYHEIELQAQQLEYDYSLCVRHPVPF-TRPKV
sp | P12746      ---MKNINADDTYRIINKIKACRSNNDINQCLSDMTKMVHCEYYLLAIIYPHSM-VKSDI
sp | P25084      -----MALVDGFLELERSSSGKLEWSAILQKMASDLGFSKILFGLLPKDSQ-DYENA
sp | P33905      -----MQHWLDKLTDLAAIEGDECILKTGLADIADHFQFTGYAYLHIQHRH-----I
sp | P54292      -MRNDGGFLLWWDGLRSEMQPIHDSQGVFAVLEKEVRRLLGFDYYAYGVRHTIPF-TRPKT
sp | P54293      -----MFSFFLENQTIT--DTLQTYIQRKLSPLGSPDYAYTVVSKKN---PSNV
sp | P54295      -----MIIDYFDNESIN--EDIKNIYIQRRIKTYGDLCSYSLVMNKKT---PLHP
sp | P54303      --MELGQQLGWDAYFYYSIFARTMDMQEFTAVALRALRELRFDFFRYGMCSVTPF-MRPRT
sp | P74946      -----MYKILRLIQENQQITSHDDLENVLNLNLIIGHEFFLFGLSFQPTL-KTSET
sp | Q03316      MKEESSAVSNLVDFDLSESASAKSKDDVLLLFGKISQYFGFSYFAISGIPSPIE-RIDSY
sp | Q44059      -----MKQDQLLEYLEHFHTSVTDGDRLAELIGRFTLGMGYDYRFFALIIPMSM-QRPKV

```

```

sp | O30919      AIFDTPAGWMQHYYQASGLFDIDPTVRAGASSSDLIVWPVSIIRDDAARLW-----SD-AR
sp | O54452      LLMLTYPPPEWVKQYRDRDYFSIDPVVRLGRRGFLPVEWSASGWDSGRAYG----FFKEAM
sp | O87971      FIVTSTYPDEWADIYTSQNYQHIDPVLVTAFAKRISPFWDENITILSDLKSS--KIFALSK
sp | P07026      AFYTYNYPEAWVSYQAKNFLAIDPVLNPNENFSQGHLMWNDDLFSQAQPLW-----EAAR
sp | P12746      SILDNYPKKWRQYDDANLIKYPIDVDYSNSNHSPINWNI FENNAVNKKSP--NVIKEAK
sp | P25084      FIVGNYPAAWREHYDRAGYARVDPTVSHCTQSVLPWFWEPSIYQTRKQHE----FFEEAS
sp | P33905      TAVTNYHRQWQSTYFDKKFEALDPVVKRARSRKHI FTWSGEHERPTLSKDER-AFYDHAS
sp | P54292      EVHGTYPKAWLERYQMNYGAVDPAILNGLRSSEM VVWSDSLFDQSRMLW-----NE-AR
sp | P54293      LIISSYPDEWIRLYRANNFQLTDPVILTAFKRTSPFAWDENITLMSDLRFT--KIFSLSK
sp | P54295      TIISNYPLDWVKKYKKNYSYHLIDPVILTAKDKVAPFAWDDNSVINKKSTDS--AVFKLAR
sp | P54303      YMYGNYPEDWVQRYQAANYAVIDPTVKHRSKVS SSPILASNELFRGCPDLW-----SEAN
sp | P74946      LVTDNYPNSWRQQYDESGFMHIDPVIKYSITNFLPIRWDDAKRVNNDGRV----IFEEAR
sp | Q03316      FVLGNWSVGFWDYRENNYVHADPVIHLSKTCDHAFVWSEALRDQKLDQRQSR-RVMDEAR
sp | Q44059      VLFNQCPDSWVQAYTANHMLACDPIIQLARKQTLPIYWNRLDERARFLQEGSLDVMGLAA

```

```

sp | O30919      DAGL-NIGVARSSWTAHGAFGLLTLARHADPLTAAELGQLS-----IATHWLANLA
sp | O54452      AFGVGRQGVTLVVRGPQGERSLFTVTSNHPDAYWRQFRMDS-----MRDLQFLAH
sp | O87971      KYNI-VNGFTFVLHDHMNNLAMLSLIMDNNADKGLNSRIESDKDRLQMNLIKIHEKMLML
sp | P07026      AHGL-RRGVTQYLMLPNRALGFLSFSRCSAREIPILSDELQ-----LKMQLLVRE
sp | P12746      TSGL-ITGFSFPIHTANNGFGMLSFAHSEKDNYIDSLFLHA-----CMNIPLIVPS
sp | P25084      AAGL-VYGLTMPLHGARGELGALSLSVEAENRAEANRFMES-----VLPTLWMLKDY
sp | P33905      DFGI-RSGITIPIKTANGFMSMFTMA-SDKPVIDLREIDA-----VAAAATIGQ
sp | P54292      DWGL-CVGATLPIRAPNNLLSVLSVARDQQNISSFEREEIR-----LRLRCMIEL
sp | P54293      QYNI-VNGFTYVLHDHMNNLALLSVI IKGNDQTALEQRLAAEQGTMQMLLIDFNEQMYRL
sp | P54295      EYNI-VNGYTFVLHDNSNNMATLNI SNGSDDSI SFDERIEINKEKIQMLLIITHEKMLGL
sp | P54303      DSNL-RHGLAQPSFNTQGRVGVLSLARKDNPI SLQEFELK-----VVTKAFAAA
sp | P74946      CNGL-KAGFSIPIHGLRGEFGMISFATSDTKSYDLNQQS-----IHTSQLIVPL
sp | Q03316      EFKL-IDGFSVPLHTAAGFQSI VSF GA EKVELSTCDRSALY-----LMAAYAHSL
sp | Q44059      EFGL-RNGISFPLHGAAGENGILSFITAERASSDLLLLLES-S-----PILSWMSNYI

```

```

      : * :
sp|O30919 HTLMSPFLVPQLVPESNA-VLTTREREVLCWTGEGKTAYEIGQILRISERTVNFHVNNVL
sp|O54452 HLHDRAMVLSGMRKVADLPRLSRRELQCLEMTANGLLAKQICARLSISVSAVQLYLASAR
sp|O87971 EQNKLGVSNGKNTDTSKGKILSPRENEVLHWASMGKTYPEIALIAGITTRTVKHHMGNVV
sp|P07026 SLMALMRLNDEIVMTPEM-NFSKREKEILRWTAEGKTSAEIAMILSISENTVNFHQNMQ
sp|P12746 LVDNYRKINIANNKSNN--LTKREKECLAWACEGKSSWDISKILGCERTVTFHLTNAQ
sp|P25084 ALQSGAGLAFEHPVSKPV-VLTSREKEVLQWCAIGKTSWEISVICNCSEANVNFHMGNIR
sp|P33905 IHARISFLRTPPTAEDAA-WLDPKEATYLRWIAVGTMEIADVEGVKYNVVRVKLREAM
sp|P54292 LTQKLTDLHPMLMSNPV-CLSHREREILQWTADGKSSGEIAIILSISESTVNFHKNIQ
sp|P54293 AGTE-GERAPALNQSAKTISSRENEVLWASMGKTYAEIAAITGISVSTVKFHIKNVV
sp|P54295 YQSN-SDKNENRNTQIERDIFSPRENEVLWASMGKTYAEISIIILGKRSTVKFHIGNVV
sp|P54303 VHEKISELESDVRFVNTDVEFSGRECDVLRWTADGKTSEEIGVIMGVCTDVTNYHHRNIQ
sp|P74946 LAHNIGNITRYHKDAKPRAVLTAREVQCLAWAAEGKSAWEIATIINTSERTVKFHFNSNAC
sp|Q03316 LRAQIGNDASRKIQALPM--ITTREREIHWCAAGKTAIEIATILGRSHRTIQNVILNIQ
sp|Q44059 FEAAIRIVRVSLREDDPQEALTDRETECLFWASEGKTSGEIACILGITERTVNYHLNQVT

```

```

      : * :
sp|O30919 LKLAATNKVQAVVKAIATGLI-----
sp|O54452 RKLTVATTSEQLLGP RRSN-----
sp|O87971 KKLGVINARQAIRLGVELELIK-PVLV-----
sp|P07026 KKINAPNKTQVACYAAATGLI-----
sp|P12746 MKLNNTNRCQSI SKAILTGAIDCPYFKN-----
sp|P25084 RKFVGT SRRVAAIMAVNLGLITL-----
sp|P33905 KRFDVRSKAHLTALAIRRCLI-----
sp|P54292 KKFDPANKT LAAAYAAALGLI-----
sp|P54293 VKLGVSNARQAIRLGVELDLIR-PAASAAR-----
sp|P54295 RKLGVLNAKHAIRLGIELKLIK-PI-----
sp|P54303 RKIGASNRVQASRYAVAMGYI-----
sp|P74946 KKLGATNRYQAITKAILGGYIN-PYL-----
sp|Q03316 RKLNVVNT PQMIAESFRLRIIR-----
sp|Q44059 RKTGSMNRYQAIAGVSSGILL-PNLEQVVVTFNPKLMQ

```

RRN3 Alignments

```

P36070 -MMAFENTSKRPPQDFVAPIDQKK-RKVQFSDSTGLVTLQPEEIKDEVFSAAMYSRFVKS
tr|Q9NYV6 MAAPLLHT--RLPGDAAASSSAVKKLGASRTGISNMRALENDFFNSPPRKTVRFGGTVTE
sp|P48322 -----MKRSTANAPKLSPKHESESDPKVKVLEEEAKPTVNQAPTGREIVE
sp|Q10110 MPSIISSSTNPQYINKCVNNGTMASSSTNVPDRTVGSKSFASSVSKNDGRLMQMLRAFVVK
XM_156394 MAAPLLHT--RLSGDVTAASAT--LSASRTGLSDMLALESDFFNSPPKKTVRFGGTVTE
tr|Q9V9M6 -----MSFYTSKTDVTSILKTFSGVKRNQANAVAINKVRFSTPKEKGLAE
NM_102796 -----MGAEFPSPVPFNSNAMDNAEYTDTDLVFAVRK

```

```

P36070 ALDDLDKNDSTQIGIIANQVALPSKNPERINDKNLNILLDILSSNINRIESSRGTFLIQS
tr|Q9NYV6 VLLKYYKKGETNDFELLKNQLLDP-DIKDD---QIINWLLLEFRSSIMYLTKDQFEQLIS--
sp|P48322 NYLKGDVTAAVLYRKICNALETFEQWESE---APKIQLLDQFLNIADAMEARTETLVK--
sp|Q10110 ALDDKAEGNFAGYEDLRRQFAAKSDTKDAPSSLQLQNLLSALTCNVSRLLDSSNSSLVM--
XM_156394 VLLKYYKKGETNDLELLKNQLSDP-DIKDD---QIINWLLLEFRSSVMYLTKDQFEQLIN--
tr|Q9V9M6 SVRVALEERNFHLVKEFTYFLREAELSDD---EVV I I I KDAKRIVHNLTPFVTVVE--
NM_102796 ALASVQNGD TDDYSQLKTVMLCTEDADFD-AVAQLETVLKSLSVSAVAIDLVHKKDLL--

```

```

P36070 IINFEEKWELPPHTLSKYIYFIKILCSSIPKWWQDVSMILVSCFI-----
tr|Q9NYV6 IILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRVI I KEGDVDVSD
sp|P48322 RLLSLRWDKIPGSVIERFRNFLCELAIRHLCFTEEVYSAVVERLVLPQISVTEETGV-VTL
sp|Q10110 SVLDSVWVSRDESFVRCYTRFLGNLISAQSNYLPLVMTMLIQHML-----Y--
XM_156394 IILRLPWLNRSQRVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRVI I VKEGGIDVSD
tr|Q9V9M6 ALLSLNWKKRSEIIEAYTEFCIELMVTHNRYLPIGVSKLIVHWIPGDMDESD----VWH
NM_102796 EAMSLWYHSHRPSVMDALVDLIIISLAATSGKYLDPCLNMLVRNFSQPTFKHK-----VSQ

```

P36070 -----LPIKQTVCHHDMLKYFLRMIPSSMGFIDITYLAKFFPNKNDTR
tr | Q9NYV6 SDDEDD-----NLPANFDTCHR-ALQIIARYVVPSTPWFLMPILVEKFPFVRKSE
sp | [P48322](#) ILTEKV-----QNEHFEMAHH-I I SSVLRCFPLSARALLKCVKRVMPHFTRPS
sp | [Q10110](#) --RPDS-----LAIHYEHAHM-ALKYVLELV PRAHSFLYSSILEEFPYKDESL
XM_156394 SDDEDDSEYDLRTWVLHDLPAIFDTCHR-ALQII TRYVVPSTPWFLMPILVEKFPFVRKSE
tr | Q9V9M6 GCPSEH-----VRNELKPIHK-VLDRILTAVPMAFDI I IDA I SAKFPYFKKPY
NM_102796 TQLVKK-----MQEVHPRVHA-ALHKISYL I PLAPWNLVSILAQNMRKIDKKD

P36070 RKLNVYTSNLLKLRGYCS-ELGFQIWSLLIEKII SIDVELQNELDELDDDDDDLEEDV
tr | Q9NYV6 RTLECYVHNLLRISVYFP-TLRHEILEL I IEKLLKLDVNASRQ-----GIEDAEET--
sp | [P48322](#) VTVAGYMRNL I LMQKYIPASISKDVWEAVFERLAKDDTHNWKC-----EQNEEMSKS
sp | [Q10110](#) LAQMTYISNVLSICEYVP-SIKGPVLHAIIDKIIQIDVEIQVE-----VDDDDEEED
XM_156394 RTLECYVHNLLRISLYFP-TLRREILELVI EKLLKLDVSVSRQ-----DIEDAEK--
tr | Q9V9M6 HVTAGYLYNVLWLLLEYKP-IFEELVFQLVLQKLLLLLDVNAPKD-----
NM_102796 PSIVTYVDNLLRLE-----NSSIGEYVGSVILMMVMERMLD-----

P36070 LEDDDDLDDSDGDDDDENCGNSNEELRSGAADGSQSDSEDM-DIIEGMDGTEEYNVELTQ
tr | Q9NYV6 -----ATQTCGGT D STEGLFNMD EDEETEHETKAGPERLDQ-MVHPVAERLDILMSLVLS
sp | [P48322](#) PRLFALNDDILIEEVVEGNTNDS EDTVPEQLEQRKGEQMIQYLDVCTDVTIFIRSSVDS
sp | [Q10110](#) -----VVTDDDGT SNADSEVITASTLYERHTAISSEMTSSTILT PPSLTDTRQLMQQLDQ
XM_156394 -----AAQTCGGTDTTEGLFNMD EDEDTDPEKKADQEQPNQ-MAHPTAERLDVLLC LLLS
tr | Q9V9M6 -----EIELEMNDEEEKVEADTLFEIDDVCAENVPKSEKSV--SHPVGKTLDVCLFLLYK
NM_102796 -----LDLVSGCDDSNNGMFMDELEDAVESTMN EGFVPGALKQNTSGGNVSELDDK

P36070 GIKELSTKLD S I L T L V S T H V E E Q V T P E S L E S G E G V G V F N T L T T L F K T H V L P T Y Y T R S I Q Y
tr | Q9NYV6 YMKDVCYVDGKVDNGKTKD-----LYRDLINIFDKLLLP THASCHVQF
sp | [P48322](#) E I D E E N G N E R T K L N D K W L R N F K I T G D K V L P K --- E K L F D T F L E C L E S T M L N A T H V Q Y V S F
sp | [Q10110](#) L L Y T L F S Y L D S N L K S T S R D R ----- Y L V Y N S L I K S F V N T V L K T F R C R Y T Q F
XM_156394 Y I E D V C R V H G K I D N N K T K D ----- L Y R D L I S I F D K L L P T H A S C H V Q F
tr | Q9V9M6 F F D G K C R L N E N S S E E Q R R T A ----- N R L F K M L L Y L F D E V L L P S H N T H H V Q F
NM_102796 L M V L F F H H L E S C Q N S D R L D E ----- V F E I L F K S F E N Y I L N T Y K T K F S Q F

P36070 I M F H V S Q Q Q L E L M D S - F L V T L I D I S F A - V N E A A E K K I K -- S L Q Y L G S Y I A R A K K L S R T Q I
tr | Q9NYV6 F M F Y L C S F K L G F A E A - F L E H L W K K L Q D - P S N P A I I R Q A -- A G N Y I G S F L A R A K F I P L I T V
sp | [P48322](#) I W L Y F C S L S Q E Y E K K - M L E H L W Q V T I R M P R A P A D A R K S Q G A A S Y L A A F L A R A K Y V K K S T A
sp | [Q10110](#) L I F W A S Q L D P E F T D I - F L G V L T E V C L D - P S Q P Y T L R L S -- G A M Y I G S Y V A R A K A L E K N T I
XM_156394 F M F F L C S F K L G F A E A - F L E H L W K K L Q D - P N N P A I I R Q A -- A A N Y I G S F L A R A K F I P L I T V
tr | Q9V9M6 I L F Y A T S L R S V Y S E A - F L D S L W T K V Q N - P N V S A V I R H A -- A V S V D G L L H T D T V N D H M A L T
NM_102796 L M F Y A C S L D P E N C G V R F A S K L L D I Y L S - S N T C R L T R M S -- A V A Y L A S Y L S R G K F L P A S F V

P36070 I F V A S Y L T S W L N R Y V I E R E E E V D Q R G - G M E R F K H F Y A A F Q A L C Y I F C F R H N I F R D T D G N W
tr | Q9NYV6 K S C L D L L V N W L H I Y L N N Q D S G T K A F C - D V A L H G P F Y S A C Q A V F Y T V F V R H K Q L L S G N L K E
sp | [P48322](#) F T W L E E V Y I W L R H Y V D Q F G S G S S Q I L P G L Q R H G T F Y S V S Q A F F L V F A F R Y K E F V K N K D M L
sp | [Q10110](#) Q I I V N M M T R W V E A Y L - D Q C E N E L S D D - L L S K H S V F Y A I N Q S I F Y I F C F R W R E L C V S D E S E
XM_156394 K T C L D L L V N W L H M Y L T N Q D S G T K A F C - D V A L H G P F Y S A C Q A V F Y T V V F R H K Q L L S G N L K Q
tr | Q9V9M6 G P F A F L F A F E I F Q A K I N S C R S D S N ----- G V Y -----
NM_102796 A S M L K R L V D E C A E Y C G T C N D D V K P ----- E A H Q V F Y S G C Q A I L Y V L C F R M R S I V E I P R F Q

P36070 E C E ----- L D K F -- F Q R M V I S K F N P L K F C N E N V M L M F A R I A Q Q E S V A Y C F S I I E N N N N E
tr | Q9NYV6 G L Q ----- Y L Q S L - N F E R I V M S Q L N P L K I C L P S V V N F F A A I T N K Y Q L V F C Y T I I E R N N R Q
sp | [P48322](#) E T I ----- R R W - G V G R V V H S P L E P L K Y V S K P V A R C F S A I T R S L Q L V Y C N H I I P I E E V Q
sp | [Q10110](#) S M E P R P N E W I P G L E I L H R S V L S R L N P L R Y C S P N I V L Q F A K V A N H L N F M Y V Y S I I E Q N R K G
XM_156394 G L Q ----- Y L Q S L - N F E R I V L S Q L N P L K I C L P Q V V N F F A A I T N K Y Q L V F C Y T I M E R N S R Q
tr | Q9V9M6 -----
NM_102796 S Q F ----- R S L --- E S I L S H K L N P L L V C L P S V V S E F L K Q A K A G G L F I V S E S F I F D D L H

P36070 RLRGIIGKADSDKKENSAQANTTSSSWSLATRQQFIDLQSYFPYDPLFLKNYKILMKEYY
tr|Q9NYV6 MLPVIRSTAGGDSVQICT-----NPLDTFFFPDPCVLKRSKKFIDPIY
sp|P48322 RP-----FDDMFPPDCYHLKESKFMTPML
sp|Q10110 IF---R--EGFDT-----MDAYFPDPYRLTKSSIIIVQPFY
XM_156394 MLPVIRSTAGGDSVQCT-----NPLDTFFFPDPCVLKRSKKFIDPIY
tr|Q9V9M6 -----
NM_102796 ESELSRAFGGFER-----LDTFFFPDPCLLKMSSSYISPNF

P36070 IEWSEASGEYESD-----GSDD-----
tr|Q9NYV6 QVWEDMSAEELQEFKKPMKKDIVEDDDDFLKGVEP--QND--TVIGITPSSFDTHFRSP
sp|P48322 RKFSPLAEDMSTLTALCWNAATADKSEKSAEAVSSSEGLDFLDEDDAMMMGGSSGYRER
sp|Q10110 NEWQQIPGLDDDE-----EEEDTDYESSTV-----M-LGESPF-----
XM_156394 QIWEDGSAEELQEFKKSTKKEVVEDEDDDFLKGVEP--QSD--TVTGLTPSSFDTHFQSP
tr|Q9V9M6 -----
NM_102796 NFWSMVKTTYGEDGDEELCDEVI VNGDADSAEEPDDDVELD--SEMNTMSTTPKHSFMRE

P36070 -----
tr|Q9NYV6 SSSVGSPVLYMQ-PSPL-----
sp|P48322 TFSCGQSSLINYSATPGLQTFNV-
sp|Q10110 -----
XM_156394 SSSVGSPVLYIPGQSPLLTRIYD
tr|Q9V9M6 -----
NM_102796 TERLLKMPSRIRPSTSPPEFLI-

¹ Timothy L. Bailey and Michael Gribskov, "Methods and statistics for combining motif match scores", *Journal of Computational Biology*, Vol. 5, pp. 211-221, 1998.

² L. Bailey and Michael Gribskov, "Score distributions for simultaneous matching to multiple motifs", *Journal of Computational Biology*, Vol. 4, pp. 45-59, 1997.

³ Thomas D. Wu, Craig G. Nevill-Manning, and Douglas L. Brutlag, "Minimal-risk scoring matrices for sequence analysis," *Journal of Computational Biology* 6: 219-235, 1999.

⁴ Thomas D. Wu, Craig G. Nevill-Manning, and Douglas L. Brutlag, "Fast probabilistic analysis of sequence function using scoring matrices," *Bioinformatics* 16: 233-244, 2000.

⁵ S. Henikoff, J. G. Henikoff, W. J. Alford & S. Pietrokovski, "Automated construction and graphical presentation of protein blocks from unaligned sequences", *Gene-COMBIS*, *Gene* 163 (1995) GC 17-26.

⁶ J.G. Henikoff, E.A. Greene, S. Pietrokovski & S. Henikoff, "Increased coverage of protein families with the blocks database servers", *Nucl. Acids Res.* 28:228-230 (2000).

⁷ R.Apweiler, T.K.Attwood, A.Bairoch, A.Bateman, E.Birney, M.Biswas, P.Bucher, L.Cerutti, F.Corpet, M.D.R.Croning, R.Durbin, L.Falquet, W.Fleischmann, J.Gouzy, H.Hermjakob, N.Hulo, I.Jonassen, D.Kahn, A.Kanapin, Y.Karavidopoulou, R.Lopez, B.Marx, N.J.Mulder, T.M.Oinn, M.Pagni, F.Servant, C.J.A.Sigrist, E.M.Zdobnov.
 "The InterPro database, an integrated documentation resource for protein families, domains and functional sites." *Nucleic Acids Research* vol 29(1):37-40 (2001).

⁸ Skurray,R.A., Paulsen,I.T., (1994) "The POT family of transport proteins." *Trends Biochem. Sci.* 19: 404-404

⁹ Steiner,H.-Y., Naider,F., Becker,J.M., (1995)"The PTR family: a new group of peptide transporters." *Mol. Microbiol.* 16: 825-834

¹⁰ 11) Moorefield B, Greene EA, Reeder RH.(2000) "RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human." *Proc Natl Acad Sci U S A.*, 25;97(9):4724-9.

¹¹ Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, 266, 383-402.

¹² Thompson, J. D., Higgins, D.G. and Gibson, T.J. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice." *Nucleic Acids Research*, 22:4673-4680.