

# Comparative Analysis of Methods for Multiple Sequence Alignment

Yong Yang

## Abstract

A total of five different multiple sequence alignment methods were analyzed in this study. The purpose of this study is to evaluate each method's ability to correctly align highly conserved motifs and secondary structure elements including  $\alpha$ -helix and  $\beta$  strand in the input sequences, especially, to avoid putting gaps in those regions. Forty-six protein sequences from the thymidylate kinase family were used to test the methods. Among the methods tested, the Hidden Markov Model (HMM) method generated the best alignment.

## Introduction

The simultaneous alignment of a number of nucleotide or amino acid sequences to produce optimal global or local alignment is of great utility in molecular biology. Multiple alignments are used to find motifs or conserved domains to characterize families of sequences. It can also be used to identify new members of existing protein families to assign function to new sequences. If the structure of some members of a family is known, then multiple sequence alignments can help predict the secondary and tertiary structure of other family members. Multiple sequence alignment can also be used for phylogenetic or evolutionary analysis.

Automatic multiple sequence alignment is an important topic in computational biology. The importance of multiple sequence alignment algorithms is clearly evidenced by the large number of programs available for this task. They can be divided into four categories: 1) Progressive global alignment of the sequences starting with alignment of the most similar pair of sequences and then adding more sequences to the alignment by aligning sequences to the consensus of the alignment already generated. ClustalW (41) and Pileup (6) belong to this category. 2) Iterative alignment methods that make an initial alignment of the input sequences and then try to iteratively refine the initial alignment to achieve better results. DIALIGN (25, 26) and SAGA (30) belong to this category. 3) optimal local alignments of common subsequences among input sequences. BLOCK-MAKER (9-11), MACAW (39), eMotif (13, 29), MOTIF and Profile-Maker belong to this category. 4) Use of statistical methods and probabilistic models of the sequences. Hidden Markov Models (14, 17, 19), MEME (2-4, 8), Gibbs sampler (23, 28) belong to this category.

Multiple sequence alignments are usually inferred from primary sequence alone. In multiple sequence alignment, residues among a set of sequences are aligned together in

columns. The residues in one column are supposed to be structurally or evolutionarily related. So ideally a column of aligned residues should occupy similar three-dimensional structural positions or all residues in that column diverge from a common ancestral residue. The gold standard for multiple structure alignments is the structural alignments. One difficulty in making sequence alignments reflect structural similarity is that gaps in the alignment should be confined to regions not in the core. Alignments that reflect structures in the core regions should have few if any gaps. The core region consists of secondary structure elements including  $\alpha$  helix and  $\beta$  strand. So in a good multiple alignment there should be few gaps in  $\alpha$  helix and  $\beta$  strand since gaps will usually disturb the structure. Also there should be few gaps in those very conserved motifs or domains for a protein family.

In this study, five different multiple sequence alignment methods were analyzed to evaluate their ability to correctly place gaps in  $\alpha$  helix,  $\beta$  strand and highly conserved motifs.

## Materials and Methods

### Test Data Set

Forty-six thymidylate kinases from prokaryotes and eukaryotes were chosen as data set to test multiple alignment methods. The Swiss-Prot ID for these forty-six proteins are as follows: KTHY\_AQUAE, KTHY\_THEMA, KTHY\_BACHD, KTHY\_BACSU, KTHY\_LACLA, KTHY\_HALN1, KTHY\_ECOLI, KTHY\_YERPE, KTHY\_HAEIN, KTHY\_PASMU, KTHY\_VIBCH, KTHY\_BUCAI, KTHY\_CAUCR, KTHY\_DEIRA, KTHY\_NEIMA, KTHY\_NEIMB, KTHY\_PSEAE, KTHY\_CHLMU, KTHY\_CHLTR, KTHY\_CHLPN, KTHY\_MYCGE, KTHY\_MYCPN, KTHY\_UREPA, KTHY\_RICPR, KTH1\_SULSO, KTHY\_ARCFU, KTHY\_THEAC, KTHY\_XYLFA, KTHY\_AERPE, KTHY\_METTH, KTHY\_METJA, KTHY\_HELPJ, KTHY\_HELPY, KTHY\_CAMJE, KTHY\_SYNY3, KTHY\_PYRAB, KTHY\_PYRHO, KTHY\_CAEEL, KTHY\_SCHPO, KTHY\_HUMAN, KTHY\_MOUSE, KTHY\_YEAST, KTHY\_VACCV, KTHY\_VARV, , KTH2\_SULSO, KTHY\_ASFB7.

The Prosite motif for this family is PS01331: [LIV]-[LIVMGSTC]-[DET]-[RH]-[FYHCS]-x(2)-S-[GSTNP]-x-[AVC]-[FY]-[STANQ].

The X-ray structure of *Escherichia coli* and yeast thymidylate kinase was solved. The PDB ID for these two structures are 4TMK and 3TMK respectively. The secondary structure of *Escherichia coli* thymidylate kinase (Fig. 1) and yeast thymidylate kinase (Fig. 2) are very similar to each other. As you can see from the superimposition of 3D structures of these two proteins (Fig. 3), the 3D structure of these two kinases are very similar, which suggests that thymidylate kinase in prokaryote and eukaryotes may take similar structure.

## Multiple alignment programs tested

Five different multiple alignment programs were tested in this study: ClustalW, PSI-BLAST, HMM, BLOCK-MAKER, Pileup. Please see Table 1 for details about these five programs.

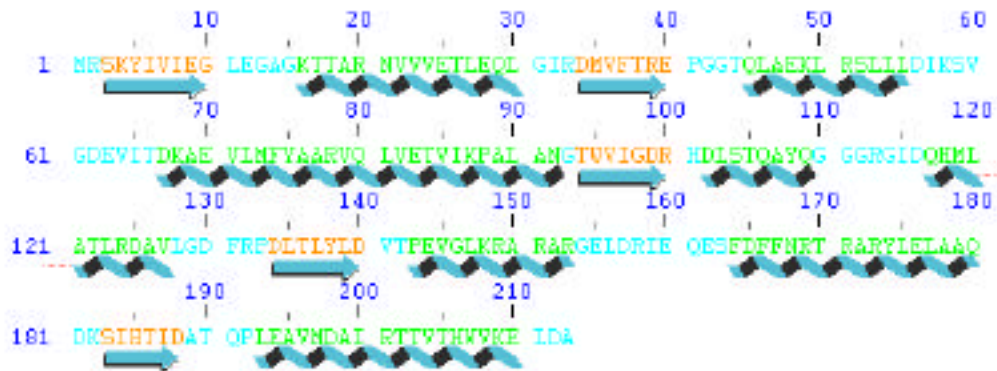


Fig. 1. Secondary structure of *Escherichia coli* thymidylate kinase (PDB ID: 4TMK).

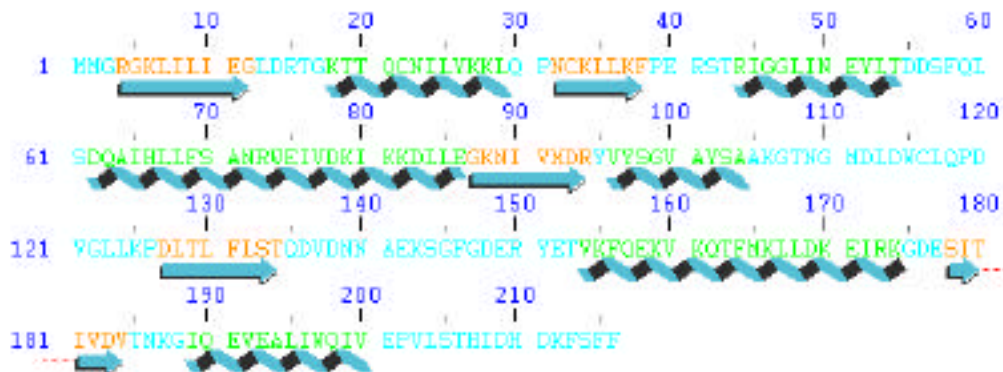


Fig. 2: Secondary structure of *Yeast* thymidylate kinase (PDB ID: 3TMK)



Fig. 3. 3D Superimposition of *Escherichia coli* thymidylate kinase and yeast thymidylate kinase.

Table 1. Multiple alignment programs used for comparative analysis in this study.

Program Name	Source	Reference
ClustalW	ClustalW 1.81 installed in Eli Lilly and Company Bioinfo server.	(41)
HMM	<a href="http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-tuneup.html">http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-tuneup.html</a>	(14, 17, 19)
Pileup	Pileup in GCG package installed in Eli Lilly and Company Bioinfo server	A part of GCG package of sequence analysis programs. Derived from (6)
BLOCK-MAKER	<a href="http://www.blocks.fhrc.org/blockmkr/make_blocks.html">http://www.blocks.fhrc.org/blockmkr/make_blocks.html</a>	(9-11)
PSI-BLAST	<a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>	(1)

## Results

Five different multiple alignment programs listed in Table 1 were analyzed to evaluate each program's ability to correctly place gaps in  $\alpha$  helices,  $\beta$  strands and those highly conserved motifs identified by eMOTIF-SEARCH and eMATRIX-SEARCH.

The structure of *Escherichia coli* thymidylate kinase were used as a gold standard for evaluating the output alignment from each of the five multiple alignment program. There are totally eight  $\alpha$  helices and four  $\beta$  strands in *Escherichia coli* thymidylate kinase.

Totally four statistically significant motifs were identified in *Escherichia coli* thymidylate kinase by both eMOTIF-SEARCH and eMATRIX-SEARCH. The four blocks are as follows:

IPB000062A Thymidylate kinase 5-YIVIEGLEGAGKTT-18

IPB000062B Thymidylate kinase 65-ITDKAEVLMFYAAR-78

IPB000062C Thymidylate kinase 86-IKPALANGTWVIGDRHDLSTQAYQ-109

IPB000062D Thymidylate kinase 133-PDLTLYLDVTPEVGLKR-149

### ClustalW

ClustalW was run using BLOSUM as scoring matrix, gap opening penalty of 10 and gap extension penalty of 0.1 for pair-wise alignment, gap opening penalty of 10 and gap extension penalty of 0.2 for multiple alignment, Gonnet series were used as protein weight matrix.

In the output alignment (Appendix 1) from ClustalW, gaps were put in two  $\alpha$  helical and two  $\beta$  strand regions of *Escherichia coli* thymidylate kinase. The size of the gaps ranges from one to nine. One gap was put in one block (IPB000062D).

One example of incorrect gap placement:

In *Escherichia coli* thymidylate kinase, 35-DMVFTRE-40 is a  $\beta$  strand region. But in the output alignment from ClustalW, a gap of size one was inserted into this region. In the alignment, this region is DMVFT-RE.

### HMM method (SAM-T99 Alignment Tuneup)

SAM-T99 was run without providing a seed alignment.

In the output alignment (Appendix 2) from SAM-T99 Alignment Tuneup program, one gap was put in one  $\alpha$  helical region of *Escherichia coli* thymidylate kinase. The gap size is three. So the HMM method generate better alignment than ClustalW in terms of gap placement in  $\alpha$  helices and  $\beta$  strands. One gap was put in one block (IPB000062B).

One example of incorrect gap placement:

In *Escherichia coli* thymidylate kinase, 67-DKAEVLMFYAARVQLVETVIKPALAN-92 is a  $\alpha$  helical region. But in the output alignment from HMM method, two gaps of size three and size one, respectively, were inserted into this region. In the alignment, this region is DKAEVLMFYAARVQLV...E.TVIKPALAN:

## Pileup

Pileup was run with gap opening penalty of 6 and gap extension penalty of 2 and other parameters as default.

In the output alignment (Appendix 3) from Pileup program, gaps were placed in three  $\alpha$  helical regions of *Escherichia coli* thymidylate kinase. The gap size ranges from one to five. Gaps were put in three out of four blocks identified by eMATRIX-SEARCH. So HMM method make a much better alignment than Pileup.

One example of incorrect gap placement:

In *Escherichia coli* thymidylate kinase, 67-DKAEVLMFYAARVQLVETVIKPALAN-92 is a  $\alpha$  helical region. But in the output alignment from Pileup, three gaps of size one, size three and size one, respectively, were inserted into this region. In the alignment, this region is DKAEV.L MFYAARVQLV...ETVIKPAL.AN

## Block-Maker

Block-Maker found totally four conserved blocks (Appendix 4). These four blocks is almost the same as the ones identified by eMOTIF-SEARCH and eMATRIX-SEARCH. The blocks found are all ungapped. So gap analysis could not be performed.

## PSI-Blast

*Escherichia coli* thymidylate kinase was used as a query to run PSI-Blast. Five iterations were run with inclusion E-value threshold of 0.05. In the output (Data not show due to big file size) for PSI-Blast There are few incorrect gaps in those alignments with hits very similar to *Escherichia coli* thymidylate kinase. But many gaps were incorrectly placed in those alignments with hits remotely related to *Escherichia coli* thymidylate kinase.

## Discussion

Among the five multiple alignment programs tested, the HMM method (SAM-T99 Alignment Tuneup) generated best alignment in terms of gap placement in secondary structure elements including  $\alpha$  helices and  $\beta$  strands and in those highly conserved motifs or domains of *Escherichia coli* thymidylate kinase. But the HMM method still incorrectly put few gaps in the  $\alpha$  helices and  $\beta$  strands and conserved motifs. A good multiple alignment programs should output alignment that has very few gaps in the core region consisting of  $\alpha$  helices and  $\beta$  strands. But most of the current multiple sequence alignment programs make alignments from primary sequences alone. Although the number of protein structure in PDB does not grow as fast as sequences, we should use the structure information to make more accurate alignment. To improve the quality of alignment, many additional strategies could be applied to the current multiple alignment methods.

First, for those families with some member's structure already solved, a multiple alignment program could make use of that structure information. The multiple alignment program could take the PDB file as input in addition to input primary sequences. It could automatically find secondary structure information in the PDB file and use that information to help decide gap penalties. In this case, position specific gap penalties should be used. A high gap opening or extension penalty should be imposed on those  $\alpha$  helices and  $\beta$  strands forming core regions.

Second, for those families with no structure information available, secondary structure prediction algorithms including Chou-Fasman algorithm (5), neural net algorithm (12, 18, 27, 31-33, 40) and nearest-neighbor methods (7, 24, 36-38, 42), could be integrated into a multiple alignment program. Although the best secondary structure program could only achieve about 75% accuracy, it should still be helpful for making a more accurate alignment. Then a multiple alignment program could use the predicted secondary structure to help define position specific gap penalty used for scoring the alignment.

Third, a multiple alignment program could use threading technique (15, 16, 20-22, 34, 35) to get some structure information about the input sequences, though structure information obtained from threading may not be very accurate. Then the multiple alignment program could use that structure prediction obtained from threading to define position specific gap penalty.

## References

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17):3389-402.
2. **Bailey, T. L., and C. Elkan.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**:28-36.
3. **Bailey, T. L., and M. Gribskov.** 1996. The megaprior heuristic for discovering protein sequence patterns. *Proc Int Conf Intell Syst Mol Biol* **4**:15-24.
4. **Bailey, T. L., and M. Gribskov.** 1998. Methods and statistics for combining motif match scores. *J Comput Biol* **5**(2):211-21.
5. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* **47**:45-147.
6. **Feng, D. F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J Mol Evol* **25**(4):351-60.
7. **Frishman, D., and P. Argos.** 1996. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng* **9**(2):133-42.
8. **Grundy, W. N., T. L. Bailey, C. P. Elkan, and M. E. Baker.** 1997. Meta-MEME: motif-based hidden Markov models of protein families. *Comput Appl Biosci* **13**(4):397-406.
9. **Henikoff, S., and J. G. Henikoff.** 1991. Automated assembly of protein blocks for database searching. *Nucleic Acids Res* **19**(23):6565-72.
10. **Henikoff, S., J. G. Henikoff, W. J. Alford, and S. Pietrokovski.** 1995. Automated construction and graphical presentation of protein blocks from unaligned sequences. *Gene* **163**(2):GC17-26.
11. **Henikoff, S., J. G. Henikoff, and S. Pietrokovski.** 1999. Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. *Bioinformatics* **15**(6):471-9.
12. **Hirst, J. D., and M. J. Sternberg.** 1992. Prediction of structural and functional features of protein and nucleic acid sequences by artificial neural networks. *Biochemistry* **31**(32):7211-8.
13. **Huang, J. Y., and D. L. Brutlag.** 2001. The EMOTIF database. *Nucleic Acids Res* **29**(1):202-4.
14. **Hughey, R., and A. Krogh.** 1996. Hidden Markov models for sequence analysis: extension and analysis of the basic method. *Comput Appl Biosci* **12**(2):95-107.
15. **Jones, D. T.** 1999. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* **287**(4):797-815.
16. **Jones, D. T., R. T. Miller, and J. M. Thornton.** 1995. Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing. *Proteins* **23**(3):387-97.
17. **Karplus, K., C. Barrett, and R. Hughey.** 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics* **14**(10):846-56.



18. **Kneller, D. G., F. E. Cohen, and R. Langridge.** 1990. Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* **214**(1):171-82.
19. **Krogh, A., M. Brown, I. S. Mian, K. Sjolander, and D. Haussler.** 1994. Hidden Markov models in computational biology. Applications to protein modeling. *J Mol Biol* **235**(5):1501-31.
20. **Lathrop, R. H.** 1999. An anytime local-to-global optimization algorithm for protein threading in theta (m2n2) space. *J Comput Biol* **6**(3-4):405-18.
21. **Lathrop, R. H.** 1994. The protein threading problem with sequence amino acid interaction preferences is NP-complete. *Protein Eng* **7**(9):1059-68.
22. **Lathrop, R. H., and T. F. Smith.** 1996. Global optimum protein threading with gapped alignment and empirical pair score functions. *J Mol Biol* **255**(4):641-65.
23. **Lawrence, C. E., S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton.** 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **262**(5131):208-14.
24. **Levin, J. M.** 1997. Exploring the limits of nearest neighbour secondary structure prediction. *Protein Eng* **10**(7):771-6.
25. **Morgenstern, B.** 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**(3):211-8.
26. **Morgenstern, B., K. Frech, A. Dress, and T. Werner.** 1998. DIALIGN: finding local similarities by multiple sequence alignment. *Bioinformatics* **14**(3):290-4.
27. **Muggleton, S., R. D. King, and M. J. Sternberg.** 1992. Protein secondary structure prediction using logic-based machine learning. *Protein Eng* **5**(7):647-57.
28. **Neuwald, A. F., J. S. Liu, and C. E. Lawrence.** 1995. Gibbs motif sampling: detection of bacterial outer membrane protein repeats. *Protein Sci* **4**(8):1618-32.
29. **Nevill-Manning, C. G., K. S. Sethi, T. D. Wu, and D. L. Brutlag.** 1997. Enumerating and ranking discrete motifs. *Proc Int Conf Intell Syst Mol Biol* **5**:202-9.
30. **Notredame, C., and D. G. Higgins.** 1996. SAGA: sequence alignment by genetic algorithm. *Nucleic Acids Res* **24**(8):1515-24.
31. **Rost, B., and C. Sander.** 1993. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc Natl Acad Sci U S A* **90**(16):7558-62.
32. **Rost, B., and C. Sander.** 2000. Third generation prediction of secondary structures. *Methods Mol Biol* **143**:71-95.
33. **Rost, B., C. Sander, and R. Schneider.** 1994. PHD--an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci* **10**(1):53-60.
34. **Rost, B., R. Schneider, and C. Sander.** 1997. Protein fold recognition by prediction-based threading. *J Mol Biol* **270**(3):471-80.
35. **Russell, R. B., R. R. Copley, and G. J. Barton.** 1996. Protein fold recognition by mapping predicted secondary structures. *J Mol Biol* **259**(3):349-65.
36. **Salamov, A. A., and V. V. Solovyev.** 1995. Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. *J Mol Biol* **247**(1):11-5.
37. **Salamov, A. A., and V. V. Solovyev.** 1997. Protein secondary structure prediction using local alignments. *J Mol Biol* **268**(1):31-6.

38. **Salzberg, S., and S. Cost.** 1992. Predicting protein secondary structure with a nearest-neighbor algorithm. *J Mol Biol* **227**(2):371-4.
39. **Schuler, G. D., S. F. Altschul, and D. J. Lipman.** 1991. A workbench for multiple alignment construction and analysis. *Proteins* **9**(3):180-90.
40. **Stolorz, P., A. Lapedes, and Y. Xia.** 1992. Predicting protein secondary structure using neural net and statistical methods. *J Mol Biol* **225**(2):363-77.
41. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**(22):4673-80.
42. **Yi, T. M., and E. S. Lander.** 1993. Protein secondary structure prediction using nearest-neighbor methods. *J Mol Biol* **232**(4):1117-29.

# Appendix 1. Output Alignment from ClustalW

CLUSTAL W (1.81) multiple sequence alignment

```
KTHY_AQUAE      -----MLIAFEGIDGSGKTTQAKKLYEYLKQKGYF-----VSLYR--
KTHY_THEMA      -----MFITFEGIDGSGKSTQIQLLAQYLEKRGKK-----VILKR--
KTHY_BACHD      -----MTKGCFITVEGEGGAGKTSALDAIEEMLRENGLS-----VVRTR--
KTHY_BACSU      -----MSGLFITFEGPEGAGKTTVLQEIKNILTAEGLQ-----VMATR--
KTHY_LACLA      -----MNGILISLEGPDGAGKTTVLKEILPEIQMKMRE-----IVPTR--
KTHY_HALN1      -----MLVTLLEGLDGSKTTVWESLRASHDDG-----VTFTA--
KTHY_ECOLI      -----MRSKYIVIEGLEGAGKTTARNVVVETLEQLGIR-----DMVFT-R
KTHY_YERPE      -----MNSKFIVIEGLEGAGKTTTRDTRVAVLRAQGIN-----DIVFT-R
KTHY_HAEIN      -----MKGKFIVIEGLEGAGKSSAHQSVVRVLHELGIQ-----DVFVT-R
KTHY_PASMU      -----MTTGKFIIVLEGEGAGKTARDSIVRALHAHGIH-----DIVFT-R
KTHY_VIBCH      -----MNAKFIVIEGLEGAGKSTAIQVVVETLQQNGID-----HITRT-R
KTHY_BUCAI      -----MIKSKFIVIEGLEGAGKTNACICIKNLLKNSIK-----NVLLV-R
KTHY_CAUCR      -----MTQGFIFISFEGEGGAGKSTQIRRLADRLKAAGH-----DVIVT-R
KTHY_DEIRA      -----MSQGLFITLEGPEGAGKTTQLARLEARLRAAGH-----AVTVT-R
KTHY_NEIMA      -----MKPQFITLDGIDGAGKSTNLAVIKAWFERRGL-----PVLFT-R
KTHY_NEIMB      -----MKPQFITLDGIDGAGKSTNLAVIKAWFERRGL-----PVLFT-R
KTHY_PSEAE      -----MTGLFVTLLEGPEGAGKSTNRDYLAEERLRENGI-----IVQLT-R
KTHY_CHLMU      -----MFIVVEGEGGAGKTQFTQALSKRLEMEEGK-----EVVLT-R
KTHY_CHLTR      -----MFIVVEGEGGAGKTQFIQALSKRLEIEGR-----EIVTT-R
KTHY_CHLPN      -----MFIVIEGEGGSGKSSLAKALGDQLVAQDR-----KVLLT-R
KTHY_MYCGE      -----MNKGVFVIEGVDGAGKTALIEGFKKLYPTKFLN-----YQLTYTR
KTHY_MYCPN      -----MKQGVFVAIEGVDGAGKTVLLEAFKQRFPPQSFLG-----FKTLFSR
KTHY_UREPA      MILTKNSNEKKPLKGLFIVFEGIDGAGKTSILKQLEVLKPKLV-----NKIFLTR
KTHY_RICPR      -----MNKLTQGKFITFEGVDGIGKSTQSKMLYEYLSQ--K-----IPVILTR
KTH1_SULSO      -----MQKLIAIEGIDGSGKTTLANLLKEHLESKMK-----LNVIVTR
KTHY_ARCFU      -----MLIAVEGIDGAGKTTIAAYIAELLKEKG-----YKVKVLK
KTHY_THEAC      -----MFIAIEGIDGAGKTTLARGIGNMLLGE-----YRVYMTK
KTHY_XYLFA      -----MLVAIEGIDGAGKTTLARSLALKLRGVG-----LETVVS
KTHY_AERPE      -----MRIVALEGIDGSGVSTHSRLLHARLAGAG-----VKSCWLK
KTHY_METTH      -----MYICFIEGIDGSGKTTAALTAASWLRENG-----YMVHEVR
KTHY_METJA      -----MVDNMFIVFEGIDGSGKTTQSKLLAKKMD-----AFWTY
KTHY_HELPJ      -----MYVVEGVDGAGKSTQVGLLKDRFKN-----ALFTK
KTHY_HELPY      -----MYVVEGVDGAGKSTQVELLKDRFKN-----ALFTK
KTHY_CAMJE      -----MYVVEGIDCVGKSTQISLLKEIYKD-----AIFTL
KTHY_SYNY3      -----MAALFIVLEGIDGSGKTTQGDLLLAHFQRQG-----LAAVLS
KTHY_PYRAB      -----MRGYFVVELEGIDGSGKTTQAKLLAEWFEEQG-----WDVLLTK
KTHY_PYRHO      -----MRGYFIVLEGIDGSGKTTQAKLLAEWFEDKG-----YEVLLTK
KTHY_CAEEL      -----MATDQKRGLLIVFEGLDKSGKSTQAKRLVESINKKSTESGDASSPSAVLQA
KTHY_SCHPO      -----MSKQNRGLIVIEGLDRSGKSTQCQLLVDKL-----ISQHEKAELFK
KTHY_HUMAN      -----MAARRGALIVLEGVDKAGKSTQSRKLVEAL-----CAAGHRAELLR
KTHY_MOUSE      -----MASRRGALIVLEGVDKAGKTTQGLKLVLTAL-----CASGHRAELLR
KTHY_YEAST      -----MMGR-GKLILIEGLDRTGKTTQCNILYKKL-----QPN---CKLLK
KTHY_VACCV      -----MSR-GALIVFEGLDKSGKTTQCMNIMESI-----PAN--TIKYLN
KTHY_VARV      -----MSR-GALIVFEGLDKSGKTTQCMNIMESI-----PTN--TIKYLN
KTH2_SULSO      -----MRGLLIAFEGIDGSGKSSQAVLLKDWIEMRR-----DVYLTE
KTHY_ASFB7      -----MRGILITIEGINGVGKSTQAMRLKKALECMDYN-----AVCIR
: .:* : * :
```

```
KTHY_AQUAE      EPGG--TKVGEVLEIREILLT-----EELDERTELLLFEASRSKLIIEK---IIPDLKRD
KTHY_THEMA      EPGG--TETGEKIRKILLE-----EEVTPKAELFLFLASRN-LLVTE---IKQYLSEG
KTHY_BACHD      EPGG--IPIAEQIRSIILDVDH----TRMDPRTEALLYAAARRQHLVEK---VLPALQAG
KTHY_BACSU      EPGG--IDIAEQIREVILNENN----ILMDPKTEALLYAAARRQHLVEK---VKPALEQG
KTHY_LACLA      EPGG--VRVAEEIRQIILDVKN----TDIDSKTELMLFAAARRLHMQEK---MLPALQAG
KTHY_HALN1      EPTD--SQYQAVR---RSESA---ADADPIAELFLFTADHADHLSRV---VSPALDRG
KTHY_ECOLI      EPGG--TQLAEKLRSLVLDIKSVG-DEVITDKAEVLMFYAARVQLVETV---IKPALANG
KTHY_YERPE      EPGG--TPLAEKLRDLIK---QGID-GEVLTDKAEVLMFYAARVQLVENV---IKPALARG
KTHY_HAEIN      EPGG--TPLAEKLRHLIK---HET-EEPVTDKAELLMLYAARIQLVENV---IKPALMQG
```

KTHY\_PASMU EP GG--TPLAEKLRQLIK---HET-EEPVTDKAELLMLYAARIQLVENV---IKPALAQG  
KTHY\_VIBCH EP GG--TL LAEKLRALVK--EEHP-GEELQDITELLLVYAARVQLVENV---IKPALARG  
KTHY\_BUCAI QPGS--TPIAEDIRRLIK--KKFN-DDNLIKETELLLMYAARIQLVEKK---IKPALKNG  
KTHY\_CAUCR EP GG--SPGAEAIRELLVNG----AADRWSFVTESELLMYAARRDHIERV---IRPGLARG  
KTHY\_DEIRA EP GG--TPLGTRVREVLLDP-----AVE-IEPLGEFLLYSASRAQLVREV---LRPALERG  
KTHY\_NEIMA EP GG--TPVGEALREILLNP-----ETKAGLRAETLMMFAARMQHIEDV---ILPALSDG  
KTHY\_NEIMB EP GG--TPVGEALREILLNP-----ETKAGLRAETLMMFAARMQHIEEV---ILPALSDG  
KTHY\_PSEAE EP GG--TPLAERIRELLLAP-----SDEPMAADTELLLMFAARAQHLAGV---IRPALARG  
KTHY\_CHLMU EP GG--SALGEQLRDLVLDV-----TQEISSYAELELLFLAARAQHIQEK---ILPALESG  
KTHY\_CHLTR EP GG--CSLGDSVRGLLLDP-----EQKISPYAELELLFLAARAQHIQEK---IIPALKSG  
KTHY\_CHLPN EP GG--CLIGERLRDLILEP-----PHLELSRCCLEFLFLGSRQHIQEV---IIPALRDG  
KTHY\_MYCGE EP GG--TL LAEKIRQLLLN-----ETMEPLTEAYLFAARTEHISKL---IKPAIEKE  
KTHY\_MYCPN EP GG--TPLAEKIRALLLH-----EAMEPLTEAYLFAASRTEHVRL---IQPALQQK  
KTHY\_UREPA EP GGKNNNAEMIREFFLKN-----LEVFDPLTLAYLYASSRAEHVKKT---INPHLEKD  
KTHY\_RICPR EVGG--TTVAEKMREILVN-----EELLPMSSELLQMAAARYDHMARK---IIPALKDG  
KTH1\_SULSO EPFS---EDIKLIKIEKIG-WNDPILL-----VLLFAADREIHV---NWLSKIK-DA  
KTHY\_ARCFU EP GD---SKFGKKIKSSEERLSPEEE-----LEFLKLDREIDARE--NILPALQ-SG  
KTHY\_THEAC EPTDG--MENYAGDGVLEFLKFTINR-----YAHQREIDRHIKN--GEIVICD-RY  
KTHY\_XYLFA EPTNG--PWGMLLRQSAATGRFSPEEE-----VDVLLRDRRQHVED--LIVPMIG-RG  
KTHY\_AERPE EPTEG--PVGRLIRGFLRSTEGVDS-----LMALLFAADRLWGLRL--GVVERCGGSP  
KTHY\_METTH EPTDS--NIGSLIRSMSSPDARTPDVQR---MLALLFAADRLT---L--RSKIEGDWAE  
KTHY\_METJA EPSNS--LVGKI IREILSKTEVDNK-----TLALLFAADRIEH-----TKLIKEELK  
KTHY\_HELPJ EP GGT--RMGESLRRIALNEN--ISELAR-----AFLFLSDRAEHIES---VIKPAKKEK  
KTHY\_HELPY EP GGT--RMGESLRRIALNEN--ISELAR-----AFLFLSDRAEHIES---VIKPAKKEK  
KTHY\_CAMJE EP GGT--ELGKHLREIILNKTHPINKRAE-----LLLFLADRAQHFE---ILK--TNQN  
KTHY\_SYNY3 EPTNG--PVGRLIRQALQGD LFTYNDARQFEAQMGYLF AADRHYHLYHPGDGVEAKLAQQ  
KTHY\_PYRAB EPTDT--EFGRLIRELVLKN--IIDGSRISYAEAEALLFAADRAEHVKK---VILPALEK  
KTHY\_PYRHO EPTDS--ELGKLIRRIILEES-VIDGSKISYAEAEALLFAADRAEHVKK---IILPALSEG  
KTHY\_CAEEF FPDRS--SSIGKLIDQYLRKEI-----DMDEHALHLLFSADRFSKNQ---MIRDNIAG  
KTHY\_SCHPO FPDRT-TAIGKKIDDYLKESV-----QLNDQVIHLLFSANRWETIQ---YIYEQINKG  
KTHY\_HUMAN FPERS-TEIGKLLSSYLQKKS-----DVEDHSVHLLFSANRWEQVP---LIKEKLSQG  
KTHY\_MOUSE FPERS-TEIGKLLNSYLEKKT-----ELEDHSVHLLFSANRWEQVP---LIKAKLNQG  
KTHY\_YEAST FPERS-TRIGGLINEYLTDDSF-----QLSDQAIHLLFSANRWEIVD---KIKKDLLEG  
KTHY\_VACCV FPQRS-TVTGKMIDDYLTRKK-----TYNDHIVNLLFCANRWEFAS---FIQEQLQEQ  
KTHY\_VARV FPQRS-TVTGKMIDDYLTRKK-----TYNDHIVNLLFCANRWEFAS---FIQEQLQEQ  
KTH2\_SULSO WNSSE--WIHDI IKEAKKKN-----MLTSITFSLIHATDFSDRYERY---ILPMLKSG  
KTHY\_ASFB7 FPNPD-TTTGGLILQVLNKM-----EMSSEQLHKLFTKHHSEFSAE---IAALLKLN

KTHY\_AQUAE KVVILDRFVLSTIAYQGYGKGLDVEFIKLNLEFATRG--VKPDITLLLDIPVDIALRRLK  
KTHY\_THEME YAVLLDRYTDSSVAYQGFGRNLGKEIVEELNDFATDG--LIPDLTFYIDVDVETALKRKG  
KTHY\_BACHD HVVLCDRFIDSSLAYQGYARGIGFEDILAIN--EFAIEGRYPDLTLLFRVDPDVGLSRIH  
KTHY\_BACSU FIVLCDRFIDSSPLAYQGYARGLGIDEVLSIN--EFAIGDMMPHVTVYFSDIDPEEGLKRIY  
KTHY\_LACLA KVVIVDRFIDSSVAYQGYGRDLGVVVDWLN--YFATDGLKPDLTLYFDVDTVALERIM  
KTHY\_HALN1 DVVISDRYSDRYAYQGATLADTVPRAMEYVRGIHQPWTRPPDVTLYFDVDPDTGAARSG  
KTHY\_ECOLI TWVIGDRHDLSTQAYQGGGRGIDQHMLATLRDAVLGD--FRPDLTLYLDVTPVGLKRAR  
KTHY\_YERPE SWVVGDRHDLSSQAYQGGGRGIDSQLMASLRDTVLGE--FRPDLTLYLDLPPAVGLRAR  
KTHY\_HAEIN KWVVGDRHDMSSQAYQGGGRQLDPHFMLTLKETVLGN--FEPDLTIYLDIDPSVGLRAR  
KTHY\_PASMU KWVIGDRHDMSSQAYQGGGRQLDQHLLHTLKQITLGE--FEPDLTLYLDIDPVLGLSRAK  
KTHY\_VIBCH EWVVGDRHDMSSQAYQGGGRQIAPSTMQSLKQTALGD--FKPDLTLYLDIDPKGLGLERAR  
KTHY\_BUCAI IWVISDRHDLSSLAYQGGGLGIPKKIIYQLQSLFLNN--FIPDLTIYLDVSP EIGLARAL  
KTHY\_CAUCR AVVLCDFADSTRAYQAGGDAPASLIAALEEHVLGG--TVPVLTLLIDLPAEVGLQRAE  
KTHY\_DEIRA ETVLCDRYADSSLAYQAGRGLSLPLLRQITAEVTGG--LTPGLTVLLDLPALGLQRAA  
KTHY\_NEIMA IHVVSDFRFTDATFAYQGGGRGMPSEDI EILEHWVQGG--LRPDLTLLLDVPLEVSMARIG  
KTHY\_NEIMB IHVVSDFRFTDATFAYQGGGRGMPSEDI EILEHWVQGG--LKPDLTLLLDVPLEVSMARIG  
KTHY\_PSEAE AVVLCDFRFTDATYAYQGGGRGLPEARIAALESFVQGD--LRPDLTLVFDLPVEIGLARAA  
KTHY\_CHLMU KTVICDRFHDSTIVYQGIAGGLGEAFVTDLCYRVVGDPEFLPDITFLLDLPEKEGLLRKT  
KTHY\_CHLTR KTVISDRFHDSTIVYQGIAGGLGESFVTNLCYHVVGDKPFLPDITFLLDIPAREGLLRKA  
KTHY\_CHLPN YIVICERFHDSTIVYQGI AEGLGADFVADLCSKVVGPTPFLPNFVLLLDIPADIGLQRKH  
KTHY\_MYCGE QLVISDRFVSSFAYQGLSKKIGIDTVKQINHHALRN--MMPNFTFILD CNFKEALQRMQ  
KTHY\_MYCPN QLVIIVDRFVSSYAYQGLIKKVGLDVVKLNADAVGD--SMPDFTFIVDCDFETALNRMA  
KTHY\_UREPA HIVISDRFVHSSYIYQGI VQNQSLDVIYQINQQAIGE--LEIDYVYFVNVNNAALNRMK  
KTHY\_RICPR YIVICDRFIDSTACYQGLELENGIDLVSLSLHKTLMPS--LMPDITFFIDVEPHTAIKRVN

KTH1\_SULSO DLIILDRYFSSIA YQ GALVDE-----QWIKMVNS-YFPK PDMVILLDLP IEVAISR I K  
 KTHY\_ARCFU YAVVM DRY YFSNIAYQSARGIDA-----RLIREMNEKIAPK PDLTILLDVEPEI ALERVR  
 KTHY\_THEAC IRSSYAYQFEGIAEFFGNSEKAW-----EWMDSVSEI I KIRPDVQIYVDVDEETAMERIS  
 KTHY\_XYLFA AVVILDRYF PSMVAYQGAAGLPV-----DALLEANAFAP-RPDVLLLLLDVPPAIGLQR I W  
 KTHY\_AERPE EVLVVD RYKYSSSLAYQGVG-SGL-----EWDVAVNR-KAPEAEILVYIDVPT E VALRRIT  
 KTHY\_METTH DVVVSDRCYYS SSMVYQGP-----E-----EWVCEINR-FAPRPDVVILLDIDVEVAMERCG  
 KTHY\_METJA RDVVCD RYLYSSIA YQSVAGVDE-----NF IKSINR-YALKPDI VFLLI VDIETALKR V K  
 KTHY\_HELPJ KLIISDRSLISGMAYS-QFSS-----LELNLLATQSVLPEKI I LLLINKENLKQRLS  
 KTHY\_HELPY KLIISDRSLISGMAYS-QFSS-----LELNLLATQSVLPAKI I LLLIDKEGLKQRLS  
 KTHY\_CAMJE KLIISDRS F ISGMAYAKDFENDL-----LFALNSFALENFFPQKI I FLKGDANLIQERLS  
 KTHY\_SYNY3 CHVITTRY YFSSLAYNCHTEAD-----WEFVQRLNQSFQPDWVIYLDLPVDLALQRLG  
 KTHY\_PYRAB KVVICDRYLYSSSLAYQWARGLS-----LEWLMQINSFAPRPDLA I LLDLPVKESIRRTK  
 KTHY\_PYRHO KVVICDRYFYSSSLAYQWARGLD-----LNWLIQVNSFAPRPDLA I LLDLPVKESLRRIK  
 KTHY\_CAEL IDVICDRYCYSGVAYS LAKGL-----PEQWVRSSDVGLPKPDAVLF F DVSP EV-AAQRG  
 KTHY\_SCHPO VTCILDRYAFSGIAFSAAGL-----DWEWCKSPDRGLPRPDLVI FLNVDPRI-AA TRG  
 KTHY\_HUMAN VTLVVD RYAFSGVAFTGAKEN-----FSLDWCKQPDVGLPKPDLVLF LQLQLAD-AAKRG  
 KTHY\_MOUSE VTLVLD RYAFSGVAFTGAKEN-----FSLDWCKQPDVGLPKPDLI FLQLQLLD-AAARG  
 KTHY\_YEAST KNIVMD RYVYSGVAYSAAKTNG-----MDLDWCLQPDVGLLKPDLTLFLSTQDV DNNAEKS  
 KTHY\_VACCV ITLIVDRYAFSGVAYAAKGA-----SMTLSKSYESGLPKPDLVIFLES GSKE---INR  
 KTHY\_VARV ITLIVDRYAFSGVAYATAKGA-----SMTLSKSYESGLPKPDLVIFLES GSKE---INR  
 KTH2\_SULSO FVVICDRYVYTAYARDVVRNVD-----FDWVKRLYSFAIKPNFTFYIRVTP EIALERIR  
 KTHY\_ASFB7 FIVIVDHYIWSGLAYA QADGITIETKNIFKPDYTF FLSSKKPLNEKPLTLQRLFETKEKQ

KTHY\_AQUAE EKNR-----FENKEFLEKVRKGFL ELAKEE-ENV-----VVIDASG-  
 KTHY\_THEME ELNR-----FEKREFLERVREGYLVLAREHPERI-----VVL D GKR-  
 KTHY\_BACHD RDQSRE-----QNRLDQEALTFHQVKKEGYERIVETYPERV-----VEIDANQ-  
 KTHY\_BACSU ANGSRE-----KNRLDLEKLD FHTKVQEGYQELMKRFPERF-----HSVDAGQ-  
 KTHY\_LACLA KNRAD E-----VNRLDLERAEMHRKVREGYLEIVVKEPERF-----VKIDASQ-  
 KTHY\_HALN1 ATN-----KFETA AFLADVRANYEQLIDYTPERF-----VRIDATQ-  
 KTHY\_ECOLI ARGE-----LDRIEQESFDFNRTRARYLELAAQDK-SI-----HTIDATQ-  
 KTHY\_YERPE ARGE-----LDRIEQESLAF FTRARYLELAASDA-SI-----KTIDASQ-  
 KTHY\_HAEIN GRGE-----LDRIEQMDLDFH RTRARYLELVKDNP-KA-----VVINAEQ-  
 KTHY\_PASMU GRGA-----LDRIEQQNLDF FHRTRQRYQELVRHNP-KA-----VTIDASQ-  
 KTHY\_VIBCH GRGE-----LDRIEKMDISFFERARERYLELANSDD-SV-----VMIDAAQ-  
 KTHY\_BUCAI KRNP-----LDLIESRSLFFFKTRRCYLEKSKL DK-KT-----I IINANL-  
 KTHY\_CAUCR ARGG-----AARFESKGLAFHERLRAGYLEIARREPDR C-----VVIDADA-  
 KTHY\_DEIRA RRGQ-----PDRLEQADLTFHRRVRQGF LDLAHAEPQRF-----LVLDATR-  
 KTHY\_NEIMA QTRE-----KDRFEQE QADFFMRVRSVYLNRAAACPERY-----AVIDSNL-  
 KTHY\_NEIMB QTRE-----KDRFEQE QADFFMRVRGVYLDRAAACPERY-----AVIDSNR-  
 KTHY\_PSEAE ARGR-----LDRFEQEDRRFF EAVRQTYLQRAAQAPERY-----QVLDAGL-  
 KTHY\_CHLMU RQKN-----LDRFEQKPTSFHRAAREGFISLAERS PDRY-----KILDALL-  
 KTHY\_CHLTR RQKH-----LDKFEQKQPIFHRSVREGFLALAEKAPDRY-----KVLDALL-  
 KTHY\_CHLPN RQKV-----FDKFEK KPLSYHNR IREGFLSLASADPSRY-----LVLDARE-  
 KTHY\_MYCGE KRGNDNL-----LDEFIKGNDFDTVRSY---YLSLVDKKN C-----FLINGD--  
 KTHY\_MYCPN KRGQDNL-----LDNTVKKQADFN TMRQY---YHSLVDNKR V-----FLLDGQ--  
 KTHY\_UREPA NRFDNTN-----AFDSQNKQFYEKLLKQYPSVFKVYNQPKKI-----IFIDANK-  
 KTHY\_RICPR ARN-----MSNKFDIRSIDFYKKIYTCFKELSNR-----FPERIKT-  
 KTH1\_SULSO N--D-----KFNFE EKI KSLAKVR---EKYLKLAKEYN-----FYVVDA--  
 KTHY\_ARCFU KR-G-----KLSPF EKLDYLRKVR---KCFLENADET-----TVVVDA--  
 KTHY\_THEAC RRGL-----RNPHFENEQLRSVR---QIYKGFQWD-----LIVDG--  
 KTHY\_XYLFA ERGS-----TPNHFETTENLSRCR---DIFLAL E LPS-----KRVIDA--  
 KTHY\_AERPE AR-E-----RREVFETPEFLERVKSMYEEVLR LARARG-----VKVIRVEG  
 KTHY\_METTH G-----TDEFEDPLYLAGVR---ERYLELADKNG-----FYTVN---  
 KTHY\_METJA T-----KDIFEKKDFLKKVQ---DKYLELAE EYN-----FIVIDT--  
 KTHY\_HELPJ LKSL-----DKIENQIEKLLTIQ---QKLKTHAYALQ-----EKFG---  
 KTHY\_HELPY LKSL-----DKIENQIEKLLHIQ---QKLKTHAYALQ-----EKFG---  
 KTHY\_CAMJE QKEL-----DSIEKRGIEYFLSVQ---DKLEKVLHFLK-----EKIS---  
 KTHY\_SYNY3 DR-----QQLEDQAPRECYEQ R---EKLISVHRNYD-----RIFAHYQ-  
 KTHY\_PYRAB ARG-----NMSEFDKLELQ---RKVRMNYLKL A-----EMFK---  
 KTHY\_PYRHO LRG-----TLTEFDKIVELQ---RKVRHNYLKL A-----EMFP---  
 KTHY\_CAEL GFGE-----ERLETATIQQKVAAMP TLRDDAY-----WKTVNADGD  
 KTHY\_SCHPO QYGE-----ERYEKIEMQEKVLKNFQRLQKEFREEGLE---FITLDASS

KTHY\_HUMAN AFGH-----ERYENGAFQERALRCFHQLMKDT-TLN----WKMVDASKR  
 KTHY\_MOUSE EFGL-----ERYETGTFQKQVLLCFQQLMEEK-NLN----WKVVDASKR  
 KTHY\_YEAST GFGD-----ERYETVKFQEKVKQTFMKLLDKE-IRKGDSEITIVDVTK  
 KTHY\_VACCV NVGE-----EIYEDVTFQKQVLLQYKMKIEEG-DIH----WQIISSEFE  
 KTHY\_VARV NVGE-----EIYEDVAFQKQVLLQYKMKIEEGEDIH----WQIISSEFE  
 KTH2\_SULSO KAKRKIKPQEAGIDILGEIPLLEEGFLKYQSRIVEIYDKIAKEESNFI-----TIDGNR-  
 KTHY\_ASFB7 ETIFTN-----FTIIMNDVPKNRLCIIPATLNKEIHTMILTITIKVFDNNSC

KTHY\_AQUAE -----EEEEVFKE-----ILRALSGVLRV-----  
 KTHY\_THEMA -----SIEEIHRD-----VVREVKRRWKLVDV-----  
 KTHY\_BACHD -----SFDQVVAD-----AVRMIKQRLSL-----  
 KTHY\_BACSU -----SKDLVVQD-----VLKVIDEALKKIQL-----  
 KTHY\_LACLA -----PLEKVVAD-----TLSVLKKRFVSEF-----  
 KTHY\_HALN1 -----SPEAVIAD-----AEAALADALPDDAWA-----  
 KTHY\_ECOLI -----PLEAVMDA-----IRTTVTHWVKELDA-----  
 KTHY\_YERPE -----PIEQVSAS-----ISQALAQWLTNQEPV-----  
 KTHY\_HAEIN -----SIELVQAD-----IESAVKNWWSNEK-----  
 KTHY\_PASMU -----TMSKVAED-----VESAIETWLTTR-----  
 KTHY\_VIBCH -----SIEQVTAD-----IRRALQDWLSQVNRV-----  
 KTHY\_BUCAI -----NIKKVTON-----ITKKMLNWLNKQVI-----  
 KTHY\_CAUCR -----ELDAVTAA-----ISDVVQRLGL-----  
 KTHY\_DEIRA -----PEDELEAE-----IWAAVSERGH-----  
 KTHY\_NEIMA -----GLDEVNRNS-----IEKVLDRHFHGC-----  
 KTHY\_NEIMB -----NLDEVNRNS-----IEKVLDRHFHGC-----  
 KTHY\_PSEAE -----PLAEVQAG-----LDRLLPNLLERLNG-----  
 KTHY\_CHLMU -----PTEVSVDQ-----ALLQIRALI-----  
 KTHY\_CHLTR -----PTEASVDQ-----ALLQIRALI-----  
 KTHY\_CHLPN -----SLASLIDK-----VMLHTQLGLCT-----  
 KTHY\_MYCGE -----NKQEHLEK-----FIELLTRCLQOPHTY-----  
 KTHY\_MYCPN -----NQTGCLEQ-----FIEQLSQCLTQPTLS-----  
 KTHY\_UREPA -----NENEVLCE-----VKEQLLKIIFKEHKYI-----  
 KTHY\_RICPR -----IKASHLSP-----LEVHELIQKHL-----  
 KTH1\_SULSO -----SKDKNEVLEQ-----AIKIIQKNLF-----  
 KTHY\_ARCFU -----SKPLEEVKKE-----VRKVIESFNLKKNNSN-----  
 KTHY\_THEAC -----GRDKEAIISE-----TFEKI IARLRQEK-----  
 KTHY\_XYLFA -----TANAETVLSA-----ALALVMEVLRVRLGALGAVVLRRLAG  
 KTHY\_AERPE VRGGVERGIEDVQGE-----IAERVFEALGLARA-----  
 KTHY\_METTH ----AERGVNLIQRD-----IRRILAPHFGICSGGIM-----  
 KTHY\_METJA ---TKKSVEEVHNE-----IIGYLKNIPH-----  
 KTHY\_HELPJ -CEVLELDAQSAKN-----LHEKIATFIECVV-----  
 KTHY\_HELPY -CEVLELDAKESVKN-----LHEKIAAFIKCAV-----  
 KTHY\_CAMJE -VEILTLDAKESKEK-----LHQQIKEFLQ-----  
 KTHY\_SYNY3 -GQLCRLDASLPVEQ-----LHQAIITKVEEML-----  
 KTHY\_PYRAB --EMRIVNAMASVEE-----VHEDIVALVKHELLGL-----  
 KTHY\_PYRHO --EMRIVNALSSIED-----IHSDIVALVKHELLGL-----  
 KTHY\_CAEEEL ----LDSVEKNVFR-----IYENLDREKPFESLEKI-----  
 KTHY\_SCHPO ----LEDVHSQIVD-----LVSNNIHETLDVL-----  
 KTHY\_HUMAN ----LEAVHEELR-----VLSEDAIRTATEKPLG-ELWK-----  
 KTHY\_MOUSE TP---SETLHRGHWGSYGNKSASIANITIFWFCRKLVEGSHLYTISRS---  
 KTHY\_YEAST G----IQEVEALIW-----QIVPVLSTHIDHDKFSFF-----  
 KTHY\_VACCV ----EDVKKELIK-----NIVIEAIHTVTGPVQQLWM-----  
 KTHY\_VARV ----EDVKKELIK-----NIVIEAIHTVTGPVQQLWM-----  
 KTH2\_SULSO ----PLKDVQID-----IRKILGEYIDNSL-----  
 KTHY\_ASFB7 LNYIKMYDDKYLNVQDLN-LFDFDWQKCIEDNNDKKEYDDDDGFI I----

## Appendix 2. Output Alignment from HMM Method

```
; SAM: /projects/compbio/programs/hmm-server-2.0/bin/i686/prettyalign v3.2.1
(Noverber 14, 2000) compiled 11/17/00_00:59:04
; (c) 1992-2000 Regents of the University of California, Santa Cruz
;
;       Sequence Alignment and Modeling Software System
;       http://www.cse.ucsc.edu/research/compbio/sam.html
;
; ----- Citations (SAM, SAM-T99, HMMs) -----
; R. Hughey, A. Krogh, Hidden Markov models for sequence analysis:
; Extension and analysis of the basic method, CABIOS 12:95-107, 1996.
; K. Karplus, C. Barrett, R. Hughey, Hidden Markov models for detecting
; remote protein homologies, Bioinformatics 14(10):846-856, 1999.
; A. Krogh et al., Hidden Markov models in computational biology:
; Applications to protein modeling, JMB 235:1501-1531, Feb 1994.
; -----
; Sequences correspond to the following labels:
; 1 KTH1_SULSO
; 2 KTHY_AERPE
; 3 KTHY_AQUAE
; 4 KTHY_ARCFU
; 5 KTHY_ASFB7
; 6 KTHY_BACHD
; 7 KTHY_BACSU
; 8 KTHY_BUCAI
; 9 KTHY_CAEEL
; 10 KTHY_CAMJE
; 11 KTHY_CAUCR
; 12 KTHY_CHLMU
; 13 KTHY_CHLPN
; 14 KTHY_CHLTR
; 15 KTHY_DEIRA
; 16 KTHY_ECOLI
; 17 KTHY_HAEIN
; 18 KTHY_HELPJ
; 19 KTHY_HELPY
; 20 KTHY_HUMAN
; 21 KTHY_LACLA
; 22 KTHY_METJA
; 23 KTHY_METTH
; 24 KTHY_MOUSE
; 25 KTHY_MYCGE
; 26 KTHY_MYCPN
; 27 KTHY_PASMU
; 28 KTHY_PYRAB
; 29 KTHY_PYRHO
; 30 KTHY_RICPR
; 31 KTHY_SCHPO
; 32 KTHY_SYNY3
; 33 KTHY_THEAC
; 34 KTHY_THEMA
; 35 KTHY_VACCV
; 36 KTHY_VARV
; 37 KTHY_VIBCH
; 38 KTHY_YEAST
; 39 KTHY_YERPE
; 40 KTH2_SULSO
; 41 KTHY_HALN1
; 42 KTHY_NEIMA
; 43 KTHY_NEIMB
; 44 KTHY_PSEAE
```

; 45 KTHY\_UREPA  
; 46 KTHY\_XYLFA

```
          10          20          30          40          50
1  mq...---KLIAIEGIDGSGKTTLANLLKEHLESk.MKL....NVIVTREPf..SEDIKLIK
2  mr...----IVALEGIDGSGVSTHSRLLHARLAG..AGV....KSCLWKEPT..EGPVGRLIRG
3  ....---MLIAFEGIDGSGKTTQAKKLYEYLKQ..KGY....FVSLYREPG..GTKVGEVLRE
4  ....---MLIAVEGIDGAGKTTIAAYIAELLKE..KGY....KVKVLKEPG..DSKFGKIKS
5  ....MRGILITIEGINGVGKSTQAMRLKKALEC..MDY....NAVCIRFPNp.DTTTGGILIQ
6  m...TKGCFITVEGEGAGKTSALDAIEEMLRE..NGL....SVVRTREPg..GIPIAEQIRS
7  ....MSGLFITFEGEGAGKTTVLQEIKNILTA..EGL....QVMATREPg..GIDIAEQIRE
8  m...IKSKFIVIEGLEGAGKTNACICIKNLLKK..NSIk...NVLLVRQPG..STPIAEDIRR
9  matdqKRGLLIVFEGLDRSGKSTQAKRLVESINK..KSTes9spSAVLQAFpDr.SSSIGKLIDQ
10 ....---MYVVFEGIDCVGKSTQISLLKEIYK-...---....DAIFTLEPG..GTELGKHLRE
11 m...TQGFIFSFEGEGAGKSTQIRRLADRLKA..AGH....DVIVTREPg..GSPGAEAIRe
12 ....---MFIVVEGEGAGKTQFTQALSKRLME..EGK....EVVLTREPg..GSALGEQLRD
13 ....---MFIVIEGEGSGKSSLAKALGDQLVA..QDR....KVLRTREPg..GCLIGERLRD
14 ....---MFIVVEGEGAGKTQFIQALSKRLIE..EGR....EIVTTREPg..GCSLGDSVRG
15 ms...-QGLFITLEGPEGAGKTTQLARLEARLRA..AGH....AVTVTREPg..GTPLGTRVRE
16 ....MRSKYIVIEGLEGAGKTTARNVVETLEQ..LGIr...DMVFTREPg..GTQLAEKLRS
17 ....MKGKFIIVIEGLEGAGKSSAHQSVVRVLHE..LGIq...DVVFTREPg..GTPLAEKLRH
18 ....---MYVVEGVDGAGKSTQVGLLKDRFK-...---....NALFTKEPG..GTRMGESLRR
19 ....---MYVVEGVDGAGKSTQVELLKDRFK-...---....NALFTKEPG..GTRMGESLRR
20 maa.RRGALIVLEGVDRAGKSTQSRKLVEALCA..AGH....RAELLRFPeR.STEIGKLLSS
21 ....MNGLLISLEGPDGAGKTTVLKEILPEIQK..MKR....EIVPTREPg..GVRVAEIIRQ
22 m...VDNMFIVFEGIDGSGKTTQSKLLAKKM--...---....DAFWTYEPS..NSLVGKIIR
23 ....---MYICFEGIDGSGKTTAALTAWLRE..NGY....MVHEVREPT..DSNIGSLIRS
24 mas.RRGALIVLEGVDRAGKTTQGLKLVTAALCA..SGH....RAELLRFPeR.STEIGKLLNS
25 m...NKGVFVIEGVDGAGKTALIEGFKKLYPT..KFLNy...QLTYTREPg..GTLLAEKIRQ
26 m...KQGVFVAIEGVDGAGKTVLLEAFKQRFQsfLGF....KTLFSREPg..GTPLAEKIRA
27 m...TTGKFIIVIEGLEGAGKTTARDSIVRALHA..HGih...DIVFTREPg..GTPLAEKLRQ
28 ....MRGYFVIEGIDGSGKTTQAKLLAEWFEE..QGw...DVLLTKEPT..DTEFGRLIRE
29 ....MRGYFVIEGIDGSGKTTQAKLLAEWFED..KGY....EVLLTKEPT..DSELGKLIRR
30 mnk1.TQGKFITFEGVDGIGKSTQSKMLYEYLKS..QKI....PVILTREVG..GTTVAEKMRE
31 mskq.NRGRLIVIEGLDRSGKSTQCQLLVDKLIS..QHE....KAELFKFPDr.TTAIGKIDDD
32 ....MAALFIVLEGIDGSGKTTQGDLLLHAHFQR..QGL....AAVLSPEPT..NGPVGRLIRQ
33 ....---MFIAIEGIDGAGKTTLARGIGNMLLG..EGY....RVYMTKEPT..D-----
34 ....---MFITFEGIDGSGKSTQIQLLAQYLEK..RGK....KVILKREPg..GTETGEKIRK
35 ms...-RGALIVFEGLDKSGKTTQCMNIMESIPA..N--....TIKYLNFpQr.STVTGKMIDD
36 ms...-RGALIVFEGLDKSGKTTQCMNIMESIPT..N--....TIKYLNFpQr.STVTGKMIDD
37 ....MNAKFIIVIEGLEGAGKSTAIQVVETLQQ..NGid...HITRTREPg..GTLLAEKLRA
38 mmg...-RGKLLIEGLDRTGKTTQCINILYKQLP...---....NCKLLKFPeR.STRIGGLINE
39 ....MNSKFIIVIEGLEGAGKTTTRDVTVAVLRA..QGIN...DIVFTREPg..GTPLAEKLRD
40 ....MRGLLIAFEGIDGSGKSSQAVLLKDWIE-.MRR....DVYLTEWNS..SEWIHDIKE
41 ....---MLVTLEGLDGSKTTVWESLRASHDD...---....GVTFTAEPt..DSQYGOAVRR
42 ....MKPQFITLDGIDGAGKSTNLAVIKAWFER..RGL....PVLFTREPg..GTPVGEALRE
43 ....MKPQFITLDGIDGAGKSTNLAVIKAWFER..RGL....PVLFTREPg..GTPVGEALRE
44 ....MTGLFVITLEGPEGAGKSTNRDYLAEERLRE..RGI....EVQLTREPg..GTPLAERIRE
45 m13p1KKGLFIVFEGIDGAGKTSILKQLLEVLKE..PKLvn...KIFLTREPgGkNNAEMIRE
46 ....---MLVAIEGIDGAGKTTLARSLALKLRG..VGL....ETVVSKEPT..NGPWGMLLRQ
```

```
          60          70          80          90          100
1  IGWND--...-PILLVLLFAADREIHV...-NWLSKIKDADLIILDRIYFSSIAIYQ
2  FLRSTE-...-GVD...SDLMALLFAADRLWGL...RlGVVERCGGSPEVLVVDRIYKYSSESLAYQ
3  ILLTE-...-ELD...ERTELLLFEARSKLI...E.EKIIPDLKRDKVILDRFVLSTIAYQ
4  ----SE...ERLS...PEEELEFLKDRREIDA...R.ENILPALQSGYAVVMDRIYFNSIAYQ
5  VLNKMT-...-EMS...SEQLHKLFTKHHSEFS...-AEIAALLKLNFIIVVDHYIWSGLAYA
6  IILDVDH...TRMD...PRTEALLYAAARRQHL...V.EKVLPALEAGHVLCDFRFDSSSLAYQ
7  VILNENN...ILMD...PKTEALLYAAARRQHL...V.EKVKPALEQGFIVLCDFRFDSSSLAYQ
8  LIKKKFNd...DNLI...KETELLLMYAARIQLV...E.KKIKPALKNGIIVISDRHDLSSSLAYQ
9  YLRKEI-...-DMD...EHALHLLFSADRFsKN...-QMIRDNIAGIDVICDRYCYSGVAYS
```



10 ILLNKT-...HPIN...KRAELLLFLADRAQHF...E.EILK--TNQNKLIISDRSFISGMAYA  
11 LLVNGAA...DRWS...PVTESELLMYAARRDHI...E.RVIRPGLARGAVVLCDFRFDSTRAYQ  
12 LVLDVT-...QEIS...SYAELLLFLAARAQHI...Q.EKILPALESGKTVICDRFHDSTIVYQ  
13 LILEPPH...LELS...RCCELFLFLGSRAQHI...Q.EVIIPALRDGYIVICERFHDSTIVYQ  
14 LLLDPE-...QKIS...PYAELLLFLAARAQHI...Q.EKIIPALKSGKTVISDRFHDSTIVYQ  
15 VVLDDPA-...VEIE...PLGEFLLYSASRAQLV...R.EVLRPALERGETVLCDRYADSSLAYQ  
16 LVLDIKSVgdEVIT...DKAEVLMFYAARVQLV...E.TVIKPALANGTWVIGDRHDLSTQAYQ  
17 LIKHETE...EPVT...DKAELLLMYAARIQLV...E.NVIKPALMQGKVVVIGDRHDMSSQAYQ  
18 IALNE--...-NIS...ELARAFLLFLSDRAEHI...E.SVIKPALKEKKLIISDRSLISGMAYS  
19 IALNE--...-NIS...ELARAFLLFLSDRAEHT...E.SVIKPALKEKKLIISDRSLISGMAYS  
20 YLKKS-...-DVE...DHSVHLLFSANRWEQV...-PLIKEKLSQGVTLVVDRYAFSSVAFV  
21 IILDPKN...TDID...SKTELMLFAAARLHM...Q.EKMLPALQAGKVVIVDRFIDSSVAYQ  
22 ILSGKT-...-EVD...NKTLLALLFAADRIEHT...-KLIKEELKKR-DVVCDRYLYSSIAVYQ  
23 MLSSPDA...RTPDv...QRMLALLFAADRLLR...-SKIEGDWAED-VVSDRCYYSMVYQ  
24 YLEKKT-...-ELE...DHSVHLLFSANRWEQV...-PLIKAKLNQGVTLVLDRYAFSSVAFV  
25 LLLNE--...-TME...PLTEAYLFAAARTEHI...S.KLIKPAIEKEQLVISDRFVVFSSFAVYQ  
26 LLLHE--...-AME...PLTEAYLFAAARTEHV...R.QLIQPALQQQLVIVDRFVWSSYAYQ  
27 LIKHETE...EPVT...DKAELLLMYAARIQLV...E.NVIKPALAQGKVVIGDRHDMSSQAYQ  
28 LVLKNSIidgSRIS...YEAEALLFAADRAEHV...K.KVILPALEKGVVICDRYLYSSLAYQ  
29 IILEESVidgSKIS...YEAEALLFAADRAEHV...K.KIILPALSEGKVVICDRYFYSSLAYQ  
30 ILVNE--...-ELL...PMSSELLQAMAARYDHM...A.RKIIPALKDGIVICDRFIDSTACYQ  
31 YLKESV-...-QLN...DQVIHLLFSANRWETI...-QYIYEQINKGVTICILDRYAFSGIAFS  
32 ALQGDLF...TYNDarqfEAQMGYLFAADRHYHlyhpG.DGVEAKLAQQCHVITTRYFSSLAYN  
33 ---GMEN...YAGD...GVELFLKFTINRYAHQ...-REIDRHIKNGEIVICDRYIRSSYAYQ  
34 ILLEE--...-EVT...PKAELFLFLASRNLLV...-TEIKQYLSEGYAVLLDRYTDSSVAYQ  
35 YLTRKK-...-TYN...DHIVNLLFCANRWEFA...-SFIQEQLQGITLIVDRYAFSGVAYA  
36 YLTRKK-...-TYN...DHIVNLLFCANRWEFA...-SFIQEQLQGITLIVDRYAFSGVAYA  
37 LVKEEHPg...EELQ...DITELLLVYAARVQLV...E.NVIKPALARGEWVVGDRHDMSSQAYQ  
38 YLTDDS-...FQLS...DQAIHLLFSANRWEIV...-DKIKKDLLEGKNIVMDRYVYSSGVAYS  
39 LIKQGIDg...EVLt...DKAEVLMFYAARVQLV...E.NVIKPALARGSWVVGDRHDLSSQAYQ  
40 AKKK-...-MLT...SITFSLIHATDFSDRY...E.RYILPMLKSGFVICDRYVYATAYARD  
41 SESAA--...-DAD...PIAELFLFTADHADHL...S.RVVSPALDRGDVVISDRYSRDAYQ  
42 ILLNPE-...TKAG...LRAETLMMFAARMQHI...E.DVILPALSDGIHVVSDFRFTDATFAYQ  
43 ILLNPE-...TKAG...LRAETLMMFAARMQHI...E.EVILPALSDGIHVVSDFRFTDATFAYQ  
44 LLLAPSD...EPMA...ADTELLLMFAARAQHL...A.GVIRPALARGAVVLCDFRFTDATYAYQ  
45 FFLKKNL...EVFD...PLTLAYLYASSRAEHV...K.KTINPHLEKDHIVISDRFVHSSYIYQ  
46 SAATG--...-RFS...PEEEVDVLLRDRRQHV...E.DLIVPMIGRGAVVILDYFVPSMVAYQ

110 120 130 140 150  
| | | | |  
1 .....GALGVDEQWIKMVNSYF---.PKPDMVILLDLPIEVAISRKND....K---FNF  
2 .....G.--VSGLEWVDVAVNRKA---.PEAEILVYIDVPTVALRRITAR....E--REV  
3 .....G.YGKGLDVEFIKLNNEFATRG.VKPDITLLLDIPVDIALRRLKEK....N----R  
4 .....SARGIDARLIREMNEKI--A..PKPDLTILLDVEPEIALERVRKR....GKLS--P  
5 .....QADGITIETKNI-----..FKPDYTFLLSSKKPLNEKPLTLQ....R-----L  
6 .....G.YARGIGFEDILAINFAIEG.RYPDLTLLFRVDPDVGLSRIHRDqs...REQNRLD  
7 .....G.YARGLGIDEVLSINEFAIGD.MMPHVTVYFSDPEEGLKRIYANGs...REKNRLD  
8 .....G.GGLGIPKKIYYQLQSLFLNN.FIPDLTIYLDVSPEIGLARALKR....NPLDLIE  
9 .....LAKGLPEQWVRSSDVGL---.PKPDAVLFVDSPEVAAQR----....GGFGEER  
10 .....--KDFENDLLFALNSFALEN.FFPQKIIFLKGANLIQERLSQK....E-LDSIE  
11 .....G.AGGDAPASLIAALEEHVVG..TVPVLTLLLDLPAEVGLQRAEAR....GGAARFE  
12 .....G.IAGGLGEAFVTDLCYRVVGDpFLPDITFLLDLPEKEGLLRKTRQ....KNLDRFE  
13 .....G.IAEGLGADFDVADLCSKVVGtpFLPNFVLLLDIPADIGLQRKHRQ....KVFDKFE  
14 .....G.IAGGLGESFVTNLCHVVGdkpFLPDITFLLDIPAREGLLRKARQ....KHLDKFE  
15 .....G.AGRGLSLPLLRQITAEVTGG.LTPGLTVLLDLPALGLQRAARR....GQPDRLE  
16 .....G.GGRGIDQHMLATLRDAVLGD.FRPDLTLYLDVTPVGLKRARAR....GELDRIE  
17 .....G.GGRQLDPHFMLTLKETVVLGN.FEPDLTIYLDIDPSVGLARARGR....GELDRIE  
18 .....Q.FSS-----LELNLLATQS.VLPEKIILLINKENLQRLSLKsld.KIENQGI  
19 .....Q.FSS-----LELNLLATQS.VLPAKIILLIDKEGLKQRLSLKsld.KIENQGI  
20 .....G.AKENFSLDWCKQPDVGL---.PKPDLVLFVLLQLADAAKR----....GAFGHER  
21 .....G.YGRDLGVEVVDWLNLYFATDG.LKPDLTLYFDVDTDVALERIMKNra...DEVNRLD  
22 .....SVAGVDENFIKSNRYA---.LKPDIVFLLIVDIETALKRVKTK....D-----I

23 .....G.....PEEWVCEINRFA---.PRPDVVILLDIDVEVAMERCGGT.....D-----E  
 24 .....G..AKENFSLDWCKQPDVGL---.PKPDLILFLQLQLLDAAR-----GEGFLER  
 25 .....G..LSKKIGIDTVKQINHHALRN..MMPNFTFILD CNFKEALQRMQKRgn..DNLLDEF  
 26 .....G..LIKVKGLDVVKKLNADAVGD..SMPDFTFIVDCDFETALNRMAKRgq..DNLLDNT  
 27 .....G..GGRQLDQHLLHHTLKQTIILGE..FEPDLTLYLDIDPVLGLSRAKGR.....GALDRIE  
 28 .....-.WARGLSLEWLMQINSFA---.PRPDLAILLDLPVKESIRRTKAR....GNMSEFD  
 29 .....-.WARGLDLNLWLIQVNSFA---.PRPDLAILLDLPVKESLRRIKLR....GTLTEFD  
 30 .....G..LELENGIDLVS LHKTLMP S..LMPDITFFIDVEPHTAIKRVNAR....NMSNKFD  
 31 .....-.AAKGLDWEWCKSPDRGL---.PRPDLVIFLNVDPRIAATR-----GQYGEER  
 32 .....-.CHTEADWEFVQRLNQSF---.PQPDWVIYLDLPVDLALQRLGDRqqle.DQAPREC  
 33 fegiaE..FFGNSEKAWEMDSVSEI IK..IRPDVQIYVDVDEETAMERISRR....G-LRNP  
 34 .....G..FGRNLGKEIIVEELNDFATDG..LIPDLTFYIDVDVETALKRKGEL....N-----R  
 35 .....-.AAKGASMTLSKSYESGL---.PKPDLVIFLES GSKEINRN-----VGE EI  
 36 .....-.TAKGASMTLSKSYESGL---.PKPDLVIFLES GSKEINRN-----VGE EI  
 37 .....G..GGRQIAPSTMQSLKQTALGD..FKPDLTLYLDIDPKLGLERARGR....GELDRIE  
 38 .....AakGTNGMDLDWCLQPDVGL---.LKPDLTLFLSTQ---DVDNNAEK....SGFGDER  
 39 .....G..GGRGIDSQ LMASLRDTVLGE..FRPDLTLYLDLPPAVGLARARAR....GELDRIE  
 40 .....-.VVRNVDFDWKRLYSFA---.IKPNFTFYIRVTP EIALERIRKak11giDILGEIP  
 41 .....GatLADTVPRAMEYVRGIHQPWT..RPPDVTLYFDVDPDTGAARSGAT....N-----K  
 42 .....G..GGRGMPSEDI EILEHWVQGG..LRPDLTLLLDVPLEVSMARIGQT....REKDRFE  
 43 .....G..GGRGMPSEDI EILEHWVQGG..LKPDLTLLLDVPLEVSMARIGQT....REKDRFE  
 44 .....G..GGRGLPEARIAALESFVQGD..LRPDLTLVFDLPVEIGLARAAR....GRLDRFE  
 45 .....G..IVQNQSLDVIYQINQQAIGE..LEIDYVFFYFDVNVNNAALNRMKNrf...DNTNAFD  
 46 .....-.GAAGLPVDALLEANAF A---.PRPDVLLLLDVPPAIGLQRIWER....G-STPNH

	160	170	180	190	200
1	E E K . .	I K S L A K V R E K Y L K L A K E Y . . . . .	- - . N F Y V V D A S K . .	D K N E V L E Q A I K I I Q K - n l f . .	
2	F E T . .	P E F L E R V K S M Y E E V L R L A r a 8 v i r V .	E G V R G G V E R . .	G I E D V Q G E I A E R V F E A l g 6 r a .	
3	F E N . .	K E F L E K V R K G F L E L A K E - . . . . .	E E . N V V I D A S G . .	E E E E V F K E I L R A L S G V l r v . .	
4	F E K . .	L D Y L R K R K C F L E N A D E - . . . . .	- - . T T V V D A S K . .	P L E E V K E E V R K V I E S F l n 8 s n .	
5	F E T . .	K E K Q E T I F T N F T I I M N D V . . . . .	P K n R L C I I P A T L . .	N K E I I H T M I L T K T I K V f 5 l i i .	
6	Q E A . .	L T F H Q K V K E G Y E R I V E T Y . . . . .	P E . R V V E I D A N Q . .	S F D Q V V A D A V R M I K Q R l s l . .	
7	L E K . .	L D F H T K V Q E G Y Q E L M K R F . . . . .	P E . R F H S V D A G Q . .	S K D L V V Q D V L K V I D E A l k 6 q l .	
8	S R S . .	L F F F K T R R C Y L E K S K L - . . . . .	D K . K T I I I N A N L . .	N I K K V T Q N I T K K M L N W l n 6 v i .	
9	L E T . .	A I E I Q Q K V A A V M P T L R D - - . . . . .	D A . Y W K T V N A D G . .	D L D S V E K N V F R I Y E N L d 1 2 k i .	
10	K R G . .	I E Y F L S V Q D K L E K V L H F L k e k i . S V .	E I L T L D A K E . .	S K E K L H Q Q I K E F L Q - - . . . . .	
11	S K G . .	L A F H E R L R A G Y L E I A R R E . . . . .	P D . R C V V I D A D A . .	E L D A V T A A I S D V V V Q R l g l . .	
12	Q K P . .	T S F H R A A R E G F I S L A E R S . . . . .	P D . R Y K I L D A L L . .	P T E V S V D Q A L L Q I - - r a l i . .	
13	K K P . .	L S Y H N R I R E G F L S L A S A D . . . . .	P S . R Y L V L D A R E . .	S L A S L I D K V M L H T Q - - l g l c t .	
14	Q K P . .	Q I F H R S V R E G F L A L A E K A . . . . .	P D . R Y K V L D A L L . .	P T E A S V D Q A L L Q I - - r a l i . .	
15	Q A D . .	L T F H R R V R Q G F L D L A H A E . . . . .	P Q . R F L V L D A T R . .	P E D E L E A E I W A A V S E R g h . . . .	
16	Q E S . .	F D F F N R T R A R Y L E L A A Q - . . . . .	D K . S I H T I D A T Q . .	P L E A V M D A I R T T V T H W v k 6 d a .	
17	Q M D . .	L D F F H R T R A R Y L E L V K D - . . . . .	N P . K A V V I N A E Q . .	S I E L V Q A D I E S A V K N W w k 6 e k .	
18	E K L . .	L T I Q Q K L K T H A Y A L Q E K F . . . . .	G C . E V L E L D A Q K . .	S A K N L H E K I A T F I E C V v . . . . .	
19	E K L . .	L H I Q Q K L K T H A Y A L Q E K F . . . . .	G C . E V L E L D A K E . .	S V K N L H E K I A A F I K C A v . . . . .	
20	Y E N . .	G A F Q E R A L R C F H Q L M K D T . . . . .	T L . N W K M V D A S K . .	R L E A V H E E L R V L S E D A i l 4 w k .	
21	L E R . .	A E M H R K V R E G Y L E I V V K E . . . . .	P E . R F V K I D A S Q . .	P L E K V V A D T L S V L K K R f v s e f .	
22	F E K . .	K D F L K K V Q D K Y L E L A E E Y . . . . .	- - . N F I V I D T T k k .	S V E E V H N E I I G Y L K N I p h . . . .	
23	F E D . .	P L Y L A G V R E R Y L E L A D K - . . . . .	- N . G F Y T V N A E R . .	G V N L I Q R D I R R I L A P H f g 9 i m .	
24	Y E T . .	G T F Q K Q V L L C F Q Q L M E E K . . . . .	N L . N W K V V D A S K r t p S E T L H R G H W G S Y G N - k 2 8 r s .		
25	I K G . .	K N D F D T V R S Y Y L S L V D K K . . . . .	N C . F L I N G D N K Q e .	H L E K F I E L L T R C L Q Q - p t h y . .	
26	V K K . .	Q A D F N T M R Q Y Y H S L V D - - . . . . .	N K . R V F L L D G Q N . .	Q T G C L E Q F I E Q L S Q C L t q 6 l s .	
27	Q Q N . .	L D F F H R T R Q R Y Q E L V R H - . . . . .	N P . K A V T I D A S Q . .	T M S K V A E D V E S A I E T W l t t r . .	
28	- K L . .	L E L Q R K V R M N Y L K L A E M - . . . . .	F K . E M R I V N A M A . .	S V E E V H E D I V A L V K H - e l l g l .	
29	- K I . .	V E L Q R K V R H N Y L K L A E M F . . . . .	P - . E M R I V N A L S . .	S I E D I H S D I V A L V K H - e l l g l .	
30	I R S . .	I D F Y K I Y T C F K E L S N R F . . . . .	P E . R I K T I K A S H L .	S P L E V H E L I Q K H L - - - . . . . .	
31	Y E K . .	I E M Q E K V L K N F Q R L Q K E F r e e .	G L . E F I T L D A S S . .	S L E D V H S Q I V D L V S N V n i 9 v l .	
32	Y E Q . .	R E K L I S V H R N Y D R I F A H Y . . . . .	Q G . Q L C R L D A S L . .	P V E Q L H Q A I I T K V E E M l . . . . .	
33	F E N . .	E Q K L R S V R Q I Y K G F Q W D - . . . . .	- - . - - L I V D G G R . .	D K E A I I S E T F E K I I - - a r 8 k t .	
34	F E K . .	R E F L E R V R E G Y L V L A R E H . . . . .	P E . R I V V L D G K R . .	S I E E I H R D V V R E V K R R w k l d v .	
35	Y E D . .	V T F Q Q K V L Q E Y K K M I E E G . . . . .	D I . H W Q I I S S E F . .	E E D V K K E L I K N I V I E A i l 3 w m .	

36 YED..VAFQQKVLQEYKKMIEEGe....DI.HWQIISSEF..EEDVKKELIKNIVIEAil3wm.  
37 KMD..ISFFERARERYLELANS-.....DD.SVVMIDAAQ..SIEQVTADIRRALQDWls7rv.  
38 YET..VKFQEKVKQTFMKLLDKEirkg.DE.SITIVDVTNk.GIQEVEALIWQIVEPVl13ff.  
39 QES..LAFFERTRARYLELAAS-.....DA.SIKTIDASQ..PIEQVSASISQALAQWlt7pv.  
40 LEEgflKYQSRIVEIYDKIAKE-.....ES.NFITIDGnr..PLKDVQIDIRKILGEYidnsl.  
41 FET..AAFLADVRANYEQLIDYT....PE.RFVRIDATQ..SPEAVIADAEALADAlp7wa.  
42 QEQ..ADFFMRVRSVYLNRAAAC....PE.RYAVIDSNL..GLDEVRNSIEKVLDHRHfgc...  
43 QEQ..ADFFMRVRGVYLDRAAAC....PE.RYAVIDSNR..NLDEVRNSIEKVLDGHfgc...  
44 QED..RRFFEAVRQTYLQRAAQA....PE.RYQVLDAGL..PLAEVQAGLDRLLPNLle6ng.  
45 SQN..KQFYEKLLKQYPSVFKVYnq...PK.KIIFIDANK..NENEVLCEVKEQLLKIfk7yi.  
46 FET..TENLSRCRDIFLALPLP-.....--.SKRVIDATA..NAETVLSAALALVMEVl18ag.

### Appendix 3. Output alignment from Pileup

!!AA\_MULTIPLE\_ALIGNMENT 1.0

PileUp of: \*.pep

Symbol comparison table: GenRunData:blosum62.cmp CompCheck: 1102

GapWeight: 6

GapLengthWeight: 2

tmk.msf MSF: 285 Type: P June 6, 2002 15:02 Check: 7952 ..

Name: kthy_vaccv	Len: 285	Check: 5453	Weight: 1.00
Name: kthy_varv	Len: 285	Check: 7296	Weight: 1.00
Name: kthy_human	Len: 285	Check: 7808	Weight: 1.00
Name: kthy_mouse	Len: 285	Check: 5725	Weight: 1.00
Name: kthy_schpo	Len: 285	Check: 9984	Weight: 1.00
Name: kthy_yeast	Len: 285	Check: 5416	Weight: 1.00
Name: kthy_caeel	Len: 285	Check: 1668	Weight: 1.00
Name: kthy_asfb7	Len: 285	Check: 7870	Weight: 1.00
Name: kthy_helpj	Len: 285	Check: 6785	Weight: 1.00
Name: kthy_helpy	Len: 285	Check: 5514	Weight: 1.00
Name: kthy_camje	Len: 285	Check: 8775	Weight: 1.00
Name: kthy_arcfu	Len: 285	Check: 1818	Weight: 1.00
Name: kthy_xylfa	Len: 285	Check: 7819	Weight: 1.00
Name: kthy_pyrab	Len: 285	Check: 3935	Weight: 1.00
Name: kthy_pyrho	Len: 285	Check: 3609	Weight: 1.00
Name: kthy_metja	Len: 285	Check: 9215	Weight: 1.00
Name: kth1_sulso	Len: 285	Check: 8486	Weight: 1.00
Name: kthy_metth	Len: 285	Check: 3120	Weight: 1.00
Name: kthy_aerpe	Len: 285	Check: 754	Weight: 1.00
Name: kthy_syny3	Len: 285	Check: 3659	Weight: 1.00
Name: kthy_mycge	Len: 285	Check: 4361	Weight: 1.00
Name: kthy_mycpn	Len: 285	Check: 4770	Weight: 1.00
Name: kthy_urepa	Len: 285	Check: 2551	Weight: 1.00
Name: kthy_haein	Len: 285	Check: 170	Weight: 1.00
Name: kthy_pasmu	Len: 285	Check: 923	Weight: 1.00
Name: kthy_ecoli	Len: 285	Check: 2166	Weight: 1.00
Name: kthy_yerpe	Len: 285	Check: 9296	Weight: 1.00
Name: kthy_vibch	Len: 285	Check: 505	Weight: 1.00
Name: kthy_buca1	Len: 285	Check: 1466	Weight: 1.00
Name: kthy_chlmu	Len: 285	Check: 8232	Weight: 1.00
Name: kthy_chltr	Len: 285	Check: 6572	Weight: 1.00
Name: kthy_chlpn	Len: 285	Check: 6898	Weight: 1.00
Name: kthy_neima	Len: 285	Check: 312	Weight: 1.00
Name: kthy_neimb	Len: 285	Check: 9059	Weight: 1.00
Name: kthy_pseae	Len: 285	Check: 7550	Weight: 1.00
Name: kthy_caucr	Len: 285	Check: 9502	Weight: 1.00
Name: kthy_deira	Len: 285	Check: 1457	Weight: 1.00
Name: kthy_bachd	Len: 285	Check: 4468	Weight: 1.00
Name: kthy_bacsu	Len: 285	Check: 9036	Weight: 1.00
Name: kthy_lacla	Len: 285	Check: 2734	Weight: 1.00
Name: kthy_aquae	Len: 285	Check: 9732	Weight: 1.00
Name: kthy_thema	Len: 285	Check: 9164	Weight: 1.00
Name: kthy_ricpr	Len: 285	Check: 7836	Weight: 1.00
Name: kthy_haln1	Len: 285	Check: 5319	Weight: 1.00
Name: kthy_theac	Len: 285	Check: 4218	Weight: 1.00
Name: kth2_sulso	Len: 285	Check: 4946	Weight: 1.00

//

kthy_vaccv	~~~~~	~MSRGALIV	FEGLDKSGKT	TQCMNIMESI	.....P
kthy_varv	~~~~~	~MSRGALIV	FEGLDKSGKT	TQCMNIMESI	.....P
kthy_human	~~~~~M	AAR.RGALIV	LEGVDRAGKS	TQSRKLVEAL	.....C
kthy_mouse	~~~~~M	ASR.RGALIV	LEGVDRAGKT	TQGLKLV TAL	.....C
kthy_schpo	~~~~~M	SKQNRRLIV	IEGLDRSGKS	TQCQLLV DKL	.....I
kthy_yeast	~~~~~	~MMGRGKLIL	IEGLDRTGKT	TQCNILYKKL	.....
kthy_caeel	~~~~~MA	TDQKRGLLIV	FEGLDRSGKS	TQAKRLVESI	NKKSTESGDA
kthy_asfb7	~~~~~	~MRGILIT	IEGINGVGKS	TQAMRLKKAL	ECMDY.....
kthy_helpj	~~~~~	~MYVV	LEGVDGAGKS	TQVGLLKDRF	K.....NAL
kthy_helpy	~~~~~	~MYVV	LEGVDGAGKS	TQVELLKDRF	K.....NAL
kthy_camje	~~~~~	~MYVV	FEGIDCVGKS	TQISLLKEIY	K.....DAI
kthy_arcfu	~~~~~	~MLIA	VEGIDGAGKT	TIAAYIAELL	KEK..GYKVK
kthy_xylfa	~~~~~	~MLVA	IEGIDGAGKT	TLARSLALKL	RGV..GLETV
kthy_pyrab	~~~~~	~MRGYFVV	LEGIDGSGKT	TQAKLLAEWF	EEQ..GWDVL
kthy_pyrho	~~~~~	~MRGYFIV	LEGIDGSGKT	TQAKLLAEWF	EDK..GYEVL
kthy_metja	~~~~~	~MVDNMFIV	FEGIDGSGKT	TQSKLLA...	.K..KMDAF
kth1_sulso	~~~~~	~MQKLI A	IEGIDGSGKT	TLANLLKEHL	ESK.MKLNVI
kthy_metth	~~~~~	~MYIC	FEGIDGSGKT	THAALTASWL	REN..GYMVH
kthy_aerpe	~~~~~	~MRIVA	LEGIDGSGVS	THSRL LHARL	AG..AGVKSC
kthy_syny3	~~~~~	~MAALFIV	LEGIDGSGKT	TQGDLLL AHF	Q.R.QGLAAV
kthy_mycge	~~~~~	~MNGKVFVV	IEGVDGAGKT	ALIEGFKKLY	PTKFLNYQLT
kthy_mycpn	~~~~~	~MKQGVFVA	IEGVDGAGKT	VILLEAFKQRF	PQSFLGFKTL
kthy_urepa	MILTKNSNEK	KPLKKGLFIV	FEGIDGAGKT	SILKQLLEVL	KEPKLVNKIF
kthy_haein	~~~~~	~MKGKFIV	IEGLEGAGKS	SAHQSVVRVL	HELGI.QDVV
kthy_pasmu	~~~~~	~MTGKFIV	LEGIEGAGKT	TARDSIVRAL	HAHGI.HDIV
kthy_ecoli	~~~~~	~MRSKYIV	IEGLEGAGKT	TARNVVVETL	EQLGI.RDMV
kthy_yerpe	~~~~~	~MNSKFIV	IEGLEGAGKT	TTRDTVVAVL	RAQGI.NDIV
kthy_vibch	~~~~~	~MNAKFIV	IEGLEGAGKS	TAIQVVVETL	QQNGI.DHIT
kthy_buca i	~~~~~	~MIKSKFIV	IEGLEGAGKT	NACICIKNLL	KKNSI.KNVL
kthy_chlmu	~~~~~	~MFIV	VEGEGAGKT	QFTQALSKRL	MEEG..KEVV
kthy_chltr	~~~~~	~MFIV	VEGEGAGKT	QFIQALSKRL	IEEG..REIV
kthy_chlpn	~~~~~	~MFIV	IEGEGSGKS	SLAKALGDQL	VAQD..RKVL
kthy_neima	~~~~~	~MKPQFIT	LDGIDGAGKS	TNLAVIKAWF	ERRG..LPVL
kthy_neimb	~~~~~	~MKPQFIT	LDGIDGAGKS	TNLAVIKAWF	ERRG..LPVL
kthy_pseae	~~~~~	~MTGLFVT	LEGPEGAGKS	TNRDYLAERL	RERG..IEVQ
kthy_caucr	~~~~~	~MTQGFIS	FEGEGAGKS	TQIRRLADRL	KAAG..HDVI
kthy_deira	~~~~~	~MSQGLFIT	LEGPEGAGKT	TQLARLEARL	RAAG..HAVT
kthy_bachd	~~~~~	~MTKGFIT	VEGEGAGKT	SALDAIEEML	RENG..LSVV
kthy_bacsu	~~~~~	~MSGLFIT	FEGPEGAGKT	TVLQEIKNIL	TAEG..LQVM
kthy_lacla	~~~~~	~MNGILIS	LEGPDGAGKT	TVLKEILPEI	QKMK..REIV
kthy_aquae	~~~~~	~MLIA	FEGIDGSGKT	TQAKKLYEYL	KQKG..YFVS
kthy_thema	~~~~~	~MFIT	FEGIDGSGKS	TQIQLLAQYL	EKRG..KKVI
kthy_ricpr	~~~~~M	NKLTQGFIT	FEGVDGIGKS	TQSKMLYEYL	KSQKI..PVI
kthy_haln1	~~~~~	~MLVT	LEGLDGS GKT	TVWESL...	.RASHDDGVT
kthy_theac	~~~~~	~MFIA	IEGIDGAGKT	TLARGIGNML	..LGEYRVY
kth2_sulso	~~~~~	~MRGLLIA	FEGIDGSGKS	SQAVLLKDWI	E...MRRDVY

51

100

kthy_vaccv	A..NTIKYLN	FPQRSTVTGK	MIDDYLTRKK	T.YNDHIVNL	LFCANRWEFA
kthy_varv	T..NTIKYLN	FPQRSTVTGK	MIDDYLTRKK	T.YNDHIVNL	LFCANRWEFA
kthy_human	AAGHRAELLR	FPERSTEIGK	LLSSYL.QKK	SDVEDHSVHL	LFSANRWEQV
kthy_mouse	ASGHRAELLR	FPERSTEIGK	LLNSYL.EKK	TELEDHSVHL	LFSANRWEQV
kthy_schpo	SQHEKAELFK	FPDRTTAIGK	KIDDYL.KES	VQLNDQVIHL	LFSANRWETI
kthy_yeast	..QPCKLLK	FPERSTRIGG	LINEYLTDDS	FQLSDQAIHL	LFSANRWEIV
kthy_caeel	SSSPSAVLQA	FPDRSSSIGK	LIDQYL.RKE	IDMDEHALHL	LFSADRFSKN
kthy_asfb7	...NAVCIR	FPNPDTTGG	LILQVL.NKM	TEMSSEQLHK	LFTKHHSEFS
kthy_helpj	FTKEP.GG..	TRMGESLRR I	ALNE...N.	.ISELARA.F	LFLSDRAEHI
kthy_helpy	FTKEP.GG..	TRMGESLRR I	ALNE...N.	.ISELARA.F	LFLSDRAEHT
kthy_camje	FTLEP.GG..	TELGKHLREI	LLNK...TH	PINKRAEL.L	LFLADRAQHF
kthy_arcfu	VLKEP...GD	SKFGKKIKS.	....SEE..	RLSPEEEL.E	LFLKDREIDA
kthy_xylfa	VSKEP...TN	GPWGMLLRQ.	....SAATG	RFSPEEEV.D	VLLRDRRQHV
kthy_pyrab	LTKEP...TD	TEFGRLIREL	VLKNSIIDGS	RISYEAEA.L	LFAADRAEHV
kthy_pyrho	LTKEP...TD	SELGKLIRRI	ILEESVIDGS	KISYEAEA.L	LFAADRAEHV

kthy\_metja WTYEP...SN SLVGKIIREI LSGKTEVDNK TL.....A.L LFAADRIEHT  
kthl\_sulso VTREP...FS EDIIKLIEKI GWNDPIL... . . . . .LV.L LFAADREIHV  
kthy\_metth EVREP...TD SNIGSLIRSM LSSPDA..RT PDVQRMLA.L LFAADRLTLR  
kthy\_aerpe LWKEP...TE GPVGRLLIRGF LRSTEGVDS. . . . .DLMA.L LFAADRLWGL  
kthy\_syny3 LSREP...TN GPVGRLLIRQA LQGDLLFTYND ARQFEAQMGY LFAADRHYHL  
kthy\_mycge YTREP.GG.. TLLAEKIRQL LLN.....E TMEPLTEA.Y LFAAARTEHI  
kthy\_mycpn FSREP.GG.. TPLAEKIRAL LLH.....E AMEPLTEA.Y LFAASRTEHV  
kthy\_urepa LTREP.GGKN NNAEMIREF FLK...NLE VFDPLTLA.Y LYASSRAEHV  
kthy\_haein FTREP.GG.. TPLAEKLRHL IKHET...EE PVTDKAEL.L MLYAARIQLV  
kthy\_pasmu FTREP.GG.. TPLAEKLRQL IKHET...EE PVTDKAEL.L MLYAARIQLV  
kthy\_ecoli FTREP.GG.. TQLAEKLRSL VLDIKSVGDE VITDKAEV.L MFYAARVQLV  
kthy\_yerpe FTREP.GG.. TPLAEKLRDL IK..QGIDGE VLTDKAEV.L MLYAARVQLV  
kthy\_vibch RTREP.GG.. TLLAEKLRAL VKEEH..PGE ELQDITEL.L LVYAARVQLV  
kthy\_bucal LVRQP.GS.. TPIAEDIRRL IK..KKFNDD NLIKETEL.L LMYAARIQLV  
kthy\_chlmu LTREP.GG.. SALGEQLRDL VLDVT...Q. EISSYAE.L LFLAARAQHI  
kthy\_chltr TTREP.GG.. CSLGDSVRGL LLDPE...Q. KISPYAE.L LFLAARAQHI  
kthy\_chlpm LTREP.GG.. CLIGERLRDL ILEPP...HL ELSRCCEL.F LFLGSRQHI  
kthy\_neima FTREP.GG.. TPVGEALREI LLNPET..KA GLRAET...L MMFAARMQHI  
kthy\_neimb FTREP.GG.. TPVGEALREI LLNPET..KA GLRAET...L MMFAARMQHI  
kthy\_pseae LTREP.GG.. TPLAERIREL LLAPSD..EP .MAADTEL.L LMFAARAQHL  
kthy\_caucr VTREP.GG.. SPGAEAIREL LVNGAA..D. RWSPTES.L LMYAARRDHI  
kthy\_deira VTREP.GG.. TPLGTRVREV VLDPAV..E. .IEPLGEF.L LYSASRAQLV  
kthy\_bachd RTREP.GG.. IPIAEQIRSI IILDVD...HT RMDPRTEA.L LYAAARRQHL  
kthy\_bacsu ATREP.GG.. IDIAEQIREV ILNEN...NI LMDPKTEA.L LYAAARRQHL  
kthy\_lacla PTREP.GG.. VRVAEEIRQI ILDPK...NT DIDSKTEL.M LFAAARRLHM  
kthy\_aquae LYREP.GG.. TKVGEVLREI LLTEE..... .LDERTEL.L LFEASRSKLI  
kthy\_thema LKREP.GG.. TETGEKIRKI LLEEE..... .VTPKAEL.F LFLASRNLLV  
kthy\_ricpr LTREV.GG.. TTVAEKMREI LVNEE..... .LLPMSEL.L QAMAARYDHM  
kthy\_haln1 FTAEP...TD SQYQAVRR. . . . .SESAA DADPIAEL.F LFTADHADHL  
kthy\_theac MTKEPTDGM NYAGDGV... . . . . .ELFL KFTINRYAHQ  
kth2\_sulso LT.....E WNSEWIHDI IKEAKKNML TSITFSLIHA TDFSDR....

101

150

kthy\_vaccv SFIQEQL... .EQGITLIV DRYAFSGVAY AAAK...GAS M.TLSKSYES  
kthy\_varv SFIQEQL... .EQGITLIV DRYAFSGVAY ATAK...GAS M.TLSKSYES  
kthy\_human PLIKEKL... .SQGVTLV DRYAFSGVAF TGAK..ENFS L.DWCKQPDV  
kthy\_mouse PLIKAKL... .NQGVTLV DRYAFSGVAF TGAK..ENFS L.DWCKQPDV  
kthy\_schpo QYIYEQL... .NKGVTICIL DRYAFSGIAF SAAK...GLD W.EWCKSPDR  
kthy\_yeast DKIKKDL... .LEGKNIVM DRYVYSGVAY SAAKGTNGMD L.DWCLQPDV  
kthy\_caeel QMIRDNI... .AKGIDVIC DRYCYSGVAY SLAK...GLP E.QWVRSSDV  
kthy\_asfb7 AEIAALL... .KLNFIIVIV DHYIWSGLAY AQA...DGIT I.E....TK  
kthy\_helpj ...ESVIKPA LKEK.KLIIS DRSLISGMAY .....SQ.FS .S...LELNL  
kthy\_helpy ...ESVIKPA LKEK.KLIIS DRSLISGMAY .....SQ.FS .S...LELNL  
kthy\_camje ...EEILK.. TNQN.KLIIS DRSFISGMAY .....AKDFE .NDLLFALNS  
kthy\_arcfu ...RENILPA L.QSGYAVVM DRYYFSNIAY Q..S.ARGID .ARLIREMNE  
kthy\_xylfa ...EDLIVPM I.GRGAVVIL DRYFPSMVAY Q..G.AAGLP .VDALLEAN.  
kthy\_pyrab ...KKVILPA L.EKGKVVIC DRYLYSSLAY Q..W.ARGLS .LEWLMIQIN.  
kthy\_pyrho ...KKIILPA L.SEGKVVIC DRYFYSSLAY Q..W.ARGLD .LNWLIQVN.  
kthy\_metja ...KLIKEE L.KK.RDVVC DRYLYSSIAY Q..S.VAGVD .ENFIKSIN.  
kthl\_sulso ...NWL..SK I.KDADLIIL DRYYFSSIAY Q..G.ALGVD .EQWIKMVN.  
kthy\_metth ...SKIEGDW A.ED..VVVS DRCYYSMVY Q..G.....P .EEWVCEIN.  
kthy\_aerpe ...RLGVVER CGGSPEVLV DRYKYSSLAY Q..G.V.GSG .LEWVDVAVNR  
kthy\_syny3 YHPGDGVEAK LAQQCH.VIT TRYFYSSLAY N..C.HTEAD .WEFVQRLNQ  
kthy\_mycge ...SKLIKPA I.EKEQLVIS DRFVFSFAY Q..GLSKKIG .IDTVKQINH  
kthy\_mycpn ...RQLIQPA L.QKQLVIV DRFVWSSYAY Q..GLIKKVG .LDVVKKLNA  
kthy\_urepa ...KKTINPH L.EKDHIIVIS DRFVHSSYIY Q..GIVQNS .LDVIYQINQ  
kthy\_haein ...ENVIKPA L.MQGKVVVG DRHDMSSQAY Q..GGGRQLD .PHFMLTLKE  
kthy\_pasmu ...ENVIKPA L.AQGKVVIG DRHDMSSQAY Q..GGGRQLD .QHLLHTLKQ  
kthy\_ecoli ...ETVIKPA L.ANGTWVIG DRHDLSTQAY Q..GGGRGID .QHMLATLRD  
kthy\_yerpe ...ENVIKPA L.ARGSWVVG DRHDLSSQAY Q..GGGRGID .SQLMASLRD  
kthy\_vibch ...ENVIKPA L.ARGWVVG DRHDMSSQAY Q..GGGRQIA .PSTMQSLKQ  
kthy\_bucal ...EKKIKPA L.KNGIWWIS DRHDLSSLAY Q..GGGLGIP .KKIYYQLQS  
kthy\_chlmu ...QEKILPA L.ESGKTVIC DRFHDSTIVY Q..GIAGGLG .EAFVTDLCY

kthy_chltr	...	QEKIIPA	L.KSGKTVIS	DRFHDSTIVY	Q..	GIAGGLG	.ESFVTNLCY
kthy_chlpn	...	QEVIIIPA	L.RDGYIVIC	ERFHDSTIVY	Q..	GIAEGLG	.ADFVADLCS
kthy_neima	...	EDVILPA	L.SDGIHVVS	DRFTDATFAY	Q..	GGGRGMP	.SEDIEILEH
kthy_neimb	...	EEVILPA	L.SDGIHVVS	DRFTDATFAY	Q..	GGGRGMP	.SEDIEILEH
kthy_pseae	...	AGVIRPA	L.ARGAVVLC	DRFTDATYAY	Q..	GGGRGLP	.EARIAALES
kthy_caucr	...	ERVIRPG	L.ARGAVVLC	DRFADSTRAY	Q..	GAGGDAP	.ASLIAALEE
kthy_deira	...	REVLRPA	L.ERGETVLC	DRYADSSLAY	Q..	GAGRGLS	.LPLLRQITA
kthy_bachd	...	VEKVLPA	L.EAGHVVLC	DRFIDSSLAY	Q..	GYARGIG	.FEDILAIN
kthy_bacsu	...	VEKVKPA	L.EQGFIIVLC	DRFIDSPLAY	Q..	GYARGLG	.IDEVLSINE
kthy_lacla	...	QEKMLPA	L.QAGKVIV	DRFIDSSVAY	Q..	GYGRDLG	.VEVVDWLN
kthy_aquae	...	EKIIPD	L.KRDKVIL	DRFVLSTIAY	Q..	GYGKGLD	.VEFIKNLNE
kthy_thema	...	TE.IKQY	L.SEGYAVLL	DRYTDSSVAY	Q..	GFRNLG	.KEIVEELND
kthy_ricpr	...	ARKIIPA	L.KDGYIVIC	DRFIDSTACY	Q..	GLELENG	.IDLVYSLHK
kthy_haln1	...	SRVVSPA	L.DRGDVVIS	DRYSDSRAY	Q..	GATLADT	.VPRAMEYVR
kthy_theac	REIDRHIK..	...	NGEIVIC	DRYIRSSAY	QFEGIAEFFG	NSEKAWEMD	
kth2_sulso	..	YERYILPM	L.KSGFVVIC	DRYVYTAYAR	DVV...	RNVD	F.DWVKRLYS

151

200

kthy_vaccv	GL.....	PKP	DLVIFLESGS	KEINR...NV	GEEIYEDVTF	QOKVLQEQYK	
kthy_varv	GL.....	PKP	DLVIFLESGS	KEINR...NV	GEEIYEDVAF	QOKVLQEQYK	
kthy_human	GL.....	PKP	DLVLFQLQL	AD.AAKRGAF	GHERYENGAF	QERALRCFHQ	
kthy_mouse	GL.....	PKP	DLILFLQLQL	LD.AAARGE	GLERYETGTF	QKQVLLCFQ	
kthy_schpo	GL.....	PRP	DLVIFLNVDP	RI.AATRGOY	GEERYEKIEM	QEKVLKNFQR	
kthy_yeast	GL.....	LKP	DLTLFLSTQD	VDNNAEKSGF	GDERYETVKF	QEKVKQTFMK	
kthy_caeel	GL.....	PKP	DAVLFFDVSP	.EVAAQRGGF	GEERLETATI	QOKVAAVMPT	
kthy_asfb7	NI.....	FKP	DYTFFLSSK	KPLNEKPLTL	.QRLFETKEK	QETIFTNF.T	
kthy_helpj	LATQSVL..	P	EKIILLINK	.ENLKQRLS	LKSLDKIEN	.QGIEKLLTI	
kthy_helpy	LATQSVL..	P	AKIILLIDK	.EGLKQRLS	LKSLDKIEN	.QGIEKLLHI	
kthy_camje	FALENFF..	P	QKIIFLKGDA	.NLIQERLS	QKELDSIEK	.RGIEYFLSV	
kthy_arcfu	....	KIAPKP	DLTILLDVEP	.EIALERVR	KRGKL.S..P	FEKLDYLRKV	
kthy_xylfa	....	AFAPRP	DVLLLLDVPP	.AIGLQRIW	ERGSTPN..H	FETTENLRS	
kthy_pyrab	....	SFAPRP	DLAILLDLPV	.KESIRRTK	ARGNMSE..F	DKLLELQRKV	
kthy_pyrho	....	SFAPRP	DLAILLDLPV	.KESLRRIK	LRGTLTE..F	DKIVELQRKV	
kthy_metja	....	RYALKP	DIVFLIVDI	.ETALKRVK	TKDI.....F	EK.KDFLKKV	
kth1_sulso	....	SYFPPK	DMVILLDLPI	.EVAISRK	..NDKFN..F	EKIKSLAKV	
kthy_metth	....	RFAPRP	DVVILLDIDV	.EVAMERC..	..GGTDE..F	EDPL.YLAGV	
kthy_aerpe	....	K.APEA	EILVYIDVPT	.EVALRRIT	ARERREV..F	.ETPEFLERV	
kthy_syny3	....	SF.PQP	DWVIYLDLPV	.DLALQRLG	DRQQLEDQAP	RECYEQREKL	
kthy_mycge	HALRN..	MMP	NFTFILD CNF	.KEALQRMQ	K.R.GNDNLL	DEFIKGKNDF	
kthy_mycpn	DAVGD..	SMP	DFTFIVDCDF	.ETALNRMA	K.R.GQDNLL	DNTVKKQADF	
kthy_urepa	QAIGE..	LEI	DYVFYFDVNV	.NNALNRMK	N.RFDNTNAF	DS..QNKQFY	
kthy_haein	TVLGN..	FEP	DLTIYLDIDP	.SVGLARAR	G..RGELDRI	EQ..MDLDF	
kthy_pasmu	TILGE..	FEP	DLTYLDIDP	.VLGLSRAR	G..RGALDRI	EQ..QNLDF	
kthy_ecoli	AVLGD..	FRP	DLTYLDVTP	.EVGLKRAR	A..RGELDRI	EQ..ESDF	
kthy_yerpe	TVLGE..	FRP	DLTYLDLPP	.AVGLARAR	A..RGELDRI	EQ..ESLAF	
kthy_vibch	TALGD..	FKP	DLTYLDIDP	.KLGLERAR	G..RGELDRI	EK..MDISFF	
kthy_bucal	LFLNN..	FIP	DLTYLDVSP	.EIGLARAL	K..RNPLDI	ES..RSLFF	
kthy_chlmu	RVVGDEPFLP		DITFLLDLPE	.KEGLLR...	KTRQKNLDRF	EQ..KPTSFH	
kthy_chltr	HVVGDKPFLP		DITFLLDIPA	.REGLLR...	KARQKHLDF	EQ..KPQIFH	
kthy_chlpn	KVVGPTPFLP		NFVLLLDIPA	.DIGLQR...	KHRQKVDFK	EK..KPLSYH	
kthy_neima	WVQGG..	LRP	DLTLLLDVPL	.EVSMARI..	.GQTREKDRF	EQ..EQADFF	
kthy_neimb	WVQGG..	LKP	DLTLLLDVPL	.EVSMARI..	.GQTREKDRF	EQ..EQADFF	
kthy_pseae	FVQGD..	LRP	DLTLVFDLPV	.EIGLARA..	.AARGRLDRF	EQ..EDRRFF	
kthy_caucr	HVLGG..	TVP	VLTLLLDLPA	.EVGLQRA..	.EARGGAARF	ES..KGLAFH	
kthy_deira	EVTGG..	LTP	GLTVLLDLDP	.ALGLQRA..	.ARRGQPDRL	EQ..ADLTFH	
kthy_bachd	FAIEG..	RYP	DLTLLFRVDP	.DVGLSRIH	RDQSREQNRL	DQ..EALTFH	
kthy_bacsu	FAIGD..	MMP	HVTVYFSIDP	.EGLKRIY	ANGSREKNRL	DL..EKLDFH	
kthy_lacla	FATDG..	LKP	DLTYFDVDT	.DVALERIM	KNRADEVNRL	DL..ERAEMH	
kthy_aquae	FATRG..	VKP	DITLLLDIPV	.DIALRRL..	...KEKNRF	...ENKEFL	
kthy_thema	FATDG..	LIP	DLTFYIDVDV	.ETALKRK..	...GELNRF	...EKREFL	
kthy_ricpr	TLM..	PSLMP	DITFFIDVPE	.HTAIKRV..	.NARNMSNKF	DI..RSIDFY	
kthy_haln1	GIHQPWTRPP		DVTLYFDVDP	.DTGAAR...	...SGATNKF	ETAA...FL	
kthy_theac	SVSEIIKIRP		DVQIYVDVDE	.ETAMERIS	RRRLRN.PHF	ENE.QKLRV	

kth2\_sulso FAI....KP NFTFYIRVTP .EIALERIR. KAKRKIKPQ. EAGIDILGEI

201

250

kthy\_vaccv MIEEG...D IHWQII...S SEFEEDVKKE LIKNIVIEAI HT.VTGPVQG  
kthy\_varv MIEEG...ED IHWQII...S SEFEEDVKKE LIKNIVIEAI HT.VTGPVQG  
kthy\_human LM.K...DTT LNWKVVDASK RL..EAVHE. ELRVLSEDAI RTATEKPLGE  
kthy\_mouse LM.E...EKN LNWKVVDASK RTPSETLHRG HWGSYGNKSA SIANT...I  
kthy\_schpo LQ.KEFREEG LEFITLDASS SL..EDVHSQ IVDLVSNVNI HETLDVLT~~~  
kthy\_yeast LLDKEIRKGD ESITIVDVTN K.GIQEVEAL IWQIV.EPVL STHIDHDKFS  
kthy\_caeel L.....RDD AYWKTVDNADG DL..DSVEKN VFRIYENLDR EKPFESLEKI  
kthy\_asfb7 IIMNDVPKNR LCIIIPATLNK EIIHTMILTK TIKVFDNNSC LNYIKMYDDK  
kthy\_helpj ..QQKLIKTHA .YALQ..... EKFGCEVLEL DA.QKSAKNL HEKIATFIEC  
kthy\_helpy ..QQKLIKTHA .YALQ..... EKFGCEVLEL DA.KESVKNL HEKIAAFIKC  
kthy\_camje ..QDKLEKVL .HFLK..... EKISVEILT TL DA.KESKEKL HQQIKEFLQ~  
kthy\_arcfu ..RKCF...L .ENAD..... E..TT..VVV DASK.PLEEV KEEVRKVIES  
kthy\_xylfa ..RDIF...L .ALEL..... P..SK..RVI DATA.NAETV LSAALALVME  
kthy\_pyrab ..RMNY...L .KLAE..... M..FKEMRIV NAMA.SVEEV HEDIVALV.K  
kthy\_pyrho ..RHNY...L .KLAE..... M..FPEMRIV NALS.SIEDI HSDIVALV.K  
kthy\_metja ..QDKY...L .ELAE..... E..YNFIVID TTKK.SVEEV HNEIIGYL.K  
kth1\_sulso ..REKY...L .KLAK..... E..YNFY.VV DASK.DKNEV LEQAIKIIQK  
kthy\_metth ..RERY...L .ELAD..... K..NGFYTV. NAER.GVNLI QRDIRRILAP  
kthy\_aerpe ..KSMYEEVL .RLAR..... ARGVKVIRVE GVRG.GVERG IEDVQGEIAE  
kthy\_syny3 ..ISVHRNYD .RIFA..... HYQGQLCRL. DASL.PVEQL HQAIITKVEE  
kthy\_mycge ..DTVRSYYL .SLV..... .D.KKNCF LINGDNKQEH LEKFIELDTR  
kthy\_mycpn ..NTMRQYYH .SLV..... .DNKRVF LLDGQNQTGC LEQFIEQLSQ  
kthy\_urepa ..EKLLKQYP .SVFK..... VYNQPKKIIIF IDANKNENEV LCEVKEQLLK  
kthy\_haein ..HRTRARYL .ELVK..... D.NP.KAVVI NAE.QSIELV QADIESAVKN  
kthy\_pasmu ..HRTRQRYQ .ELVR..... H.NP.KAVTI DAS.QTMSKV AEDVESAIET  
kthy\_ecoli ..NRTRARYL .ELAA..... Q.DK.SIHTI DAT.QPLEAV MDAIRTTVTTH  
kthy\_yerpe ..ERTRARYL .ELAA..... S.DA.SIKTI DAS.QPIEQV SASISQALAQ  
kthy\_vibch .ELAN..... S.DD.SVMMI DAA.QSIEQV TADIRRALQD  
kthy\_bucaj ..KKTRRCYL .EKSK..... L.DK.KTIII NAN.LNIKKV TQNITKKMLN  
kthy\_chlmu ..RAAREGFI .SLAE..... R.SPDRYKIL DAL.LPTEVS VDQALLQIRA  
kthy\_chltr ..RSVREGFL .ALAE..... K.APDRYKVL DAL.LPTEAS VDQALLQIRA  
kthy\_chlpn ..NRIREGFL .SLAS..... A.DPSRYLVL DAR.ESLASL IDKVMLHTQL  
kthy\_neima ..MRVRSVYL .NRAA..... A.CPERYAVI DSN.LGLDEV RNSIEKVLDR  
kthy\_neimb ..MRVRGVYL .DRAA..... A.CPERYAVI DSN.RNLDEV RNSIEKVLDR  
kthy\_pseae ..EAVRQTYL .QRAA..... Q.APERYQVL DAG.LPLAEV QAGLDRLLPN  
kthy\_caucr ..ERLRAGYL .ETAR..... R.EPDRCVVI DAD.AELDAV TAAISDVVVQ  
kthy\_deira ..RRVRQGFL .DLAH..... A.EPQRFLVL DAT.RPEDEL EAEIWAAVSE  
kthy\_bachd ..QKVKEGYE .RIVE..... T.YPERVVEI DAN.QSFDQV VADAVRMIKQ  
kthy\_bacsu ..TKVQEGYQ .ELMK..... R.FPERFHSV DAG.QSKDLV VQDVLKVIDE  
kthy\_lacla ..RKVREGYL .EIVV..... K.EPERFVKI DAS.QPLEKV VADTSLVLKK  
kthy\_aquae ..EKVRKGFL .ELAK..... E.E.ENVVVI DAS.GEEVEV FKEILRALSG  
kthy\_thema ..ERVREGYL .VLAR..... E.HPERIVVL DGK.RSIEEI HRDVVREVKR  
kthy\_ricpr ..KKIYTCFK .ELSN..... R.FPERIKTI KASHLSPLEV HELIQKHL~~~  
kthy\_haln1 ..ADVRANYE .QLI..... DYTPERFVRI DAT.QSPEAV IADAEALAD  
kthy\_theac ..RQIYKGFQ WDLIV..... DGGRDKEAII SETFEKIIAR LRQEKT~~~~~  
kth2\_sulso PLEEGFLKYQ SRIVEIYDKI AKEESNFITI DGN.RPLKDV QIDIRKILGE

251

285

kthy\_vaccv LWM~~~~~  
kthy\_varv LWM~~~~~  
kthy\_human LWK~~~~~  
kthy\_mouse FWFCKRLVEG SHLYTISRS~  
kthy\_schpo ~~~~~~  
kthy\_yeast FF~~~~~  
kthy\_caeel ~~~~~~  
kthy\_asfb7 YLNVQDLNLF DFDWQKCIED NNDKEEYDDD DGFII  
kthy\_helpj VV~~~~~  
kthy\_helpy AV~~~~~  
kthy\_camje ~~~~~~  
kthy\_arcfu FLNLKKNNSN~



kthy_xylfa	VLRVRLGALG	AVVLRRLAG~	~~~~~	~~~~~
kthy_pyrab	...HELLGL~	~~~~~	~~~~~	~~~~~
kthy_pyrho	...HELLGL~	~~~~~	~~~~~	~~~~~
kthy_metja	NIPH~~~~~	~~~~~	~~~~~	~~~~~
kth1_sulso	NLF~~~~~	~~~~~	~~~~~	~~~~~
kthy_metth	HFGICSGGIM	~~~~~	~~~~~	~~~~~
kthy_aerpe	RVFEALGLAR	A~~~~~	~~~~~	~~~~~
kthy_syny3	ML~~~~~	~~~~~	~~~~~	~~~~~
kthy_mycge	CLQQPTHY~~	~~~~~	~~~~~	~~~~~
kthy_mycpn	CLTQPTLS~~	~~~~~	~~~~~	~~~~~
kthy_urepa	IFKEHKYI~~	~~~~~	~~~~~	~~~~~
kthy_haein	WVKSNK~~~~	~~~~~	~~~~~	~~~~~
kthy_pasmu	WLTR~~~~~	~~~~~	~~~~~	~~~~~
kthy_ecoli	WVKELDA~~~	~~~~~	~~~~~	~~~~~
kthy_yerpe	WLTNQEPV~~	~~~~~	~~~~~	~~~~~
kthy_vibch	WLSQVNRV~~	~~~~~	~~~~~	~~~~~
kthy_buca	WLNKQVI~~~	~~~~~	~~~~~	~~~~~
kthy_chlmu	LI~~~~~	~~~~~	~~~~~	~~~~~
kthy_chltr	LI~~~~~	~~~~~	~~~~~	~~~~~
kthy_chlpn	GLCT~~~~~	~~~~~	~~~~~	~~~~~
kthy_neima	HFGC~~~~~	~~~~~	~~~~~	~~~~~
kthy_neimb	HFGC~~~~~	~~~~~	~~~~~	~~~~~
kthy_pseae	LLERLNG~~~	~~~~~	~~~~~	~~~~~
kthy_caucr	RLGL~~~~~	~~~~~	~~~~~	~~~~~
kthy_deira	R.GH~~~~~	~~~~~	~~~~~	~~~~~
kthy_bachd	RLSL~~~~~	~~~~~	~~~~~	~~~~~
kthy_bacsu	ALKKIQL~~~	~~~~~	~~~~~	~~~~~
kthy_lacla	RFVSEF~~~~	~~~~~	~~~~~	~~~~~
kthy_aquae	VLRV~~~~~	~~~~~	~~~~~	~~~~~
kthy_thema	RWKLDV~~~~	~~~~~	~~~~~	~~~~~
kthy_ricpr	~~~~~	~~~~~	~~~~~	~~~~~
kthy_haln1	ALPDDAWA~~	~~~~~	~~~~~	~~~~~
kthy_theac	~~~~~	~~~~~	~~~~~	~~~~~
kth2_sulso	YIDNSL~~~~	~~~~~	~~~~~	~~~~~

## Appendix 4. Blocks from Block-Maker

Your blocks are being made, please wait ...

Check for Warnings before re-submitting  
Block Maker Results

Introduction  
Warnings  
Block Maps [ About Block Maps]  
BLOCKS from MOTIF  
COBBLER sequence from MOTIF [ About COBBLER]  
BLOCKS from GIBBS  
COBBLER sequence from GIBBS [ About COBBLER]  
Blocks Formatted for Searching Programs  
BLIMPS Searchable BLOCKS from MOTIF [ About BLIMPS] [Download BLIMPS]  
BLIMPS Searchable BLOCKS from GIBBS  
MAST Searchable PSSMS from MOTIF [ About MAST]  
MAST Searchable PSSMS from GIBBS

-----  
BLOCKS from MOTIFLogos: [Postscript][PDF][GIF] [About Logos] [Map]Tree:  
[Data][Xbitmap][Postscript][PDF][GIF][Newick] [About Trees]Search: [LAMA]  
[About LAMA] [MAST] [About MAST]Primers:[CODEHOP] [About  
CODEHOP]Substitutions:[SIFT] [About SIFT on Blocks]  
\*\*BLOCKS from MOTIF\*\*

>unknown Q9UXG7 Probable thymidylate kinase 1 (EC 2.7.4.9) ... family  
40 sequences are included in 4 blocks

```
unknownA, width = 14unknownB, width = 7
KTH1_SULSO      4 LIAIEGIDGSGKTT ( 19)  37 TREPFSE
KTHY_AERPE      3 IVALEGIDGSGVST ( 22)  39 TEGPVGR
KTHY_AQUAE      2 LIAFEGIDGSGKTT ( 18)  34 YREPGGT
KTHY_ARCFU      2 LIAVEGIDGAGKTT ( 18)  34 LKEPGDS
KTHY_ASFB7      5 LITIEGINGVGKST ( 24)  43 DTTTGGI
KTHY_BACHD      6 FITVEGEGEGAGKTS ( 18)  38 TREPGGI
KTHY_BACSU      5 FITFEGPEGAGKTT ( 18)  37 TREPGGI
KTHY_BUCAI      6 FIVIEGLEGAGKTN ( 19)  39 VRQPGST
KTHY_CAMJE      2 YVVFEGIDCVGKST ( 14)  30 TLEPGGT
KTHY_CAUCR      6 FISFEGEGEGAGKST ( 18)  38 TREPGGS
KTHY_CHLMU      2 FIVVEGEGEGAGKTQ ( 18)  34 TREPGGS
KTHY_CHLPN      2 FIVIEGEGEGSGKSS ( 18)  34 TREPGGC
KTHY_CHLTR      2 FIVVEGEGEGAGKTQ ( 18)  34 TREPGGC
KTHY_DEIRA      6 FITLEGPEGAGKTT ( 18)  38 TREPGGT
KTHY_ECOLI      5 YIVIEGLEGAGKTT ( 19)  38 TREPGGT
KTHY_HAEIN      5 FIVIEGLEGAGKSS ( 19)  38 TREPGGT
KTHY_HALN1      2 LVTLEGLDGSKTT ( 15)  31 TAEPTDS
KTHY_HELPJ      2 YVVLEGVDGAGKST ( 14)  30 TKEPGGT
KTHY_HELPY      2 YVVLEGVDGAGKST ( 14)  30 TKEPGGT
KTHY_LACLA      5 LISLEGPDGAGKTT ( 18)  37 TREPGGV
KTHY_METJA      6 FIVFEGIDGSGKTT ( 13)  33 TYEPSNS
KTHY_METHH      2 YICFEGIDGSGKTT ( 18)  34 VREPTDS
KTHY_MOUSE      8 LIVLEGVDGAGKTT ( 24)  46 STEIGKL
KTHY_MYCGE      6 FVVIEGVDGAGKTA ( 20)  40 TREPGGT
KTHY_MYCPN      6 FVAIEGVDGAGKTV ( 20)  40 SREPGGT
KTHY_NEIMA      5 FITLDGIDGAGKST ( 18)  37 TREPGGT
KTHY_NEIMB      5 FITLDGIDGAGKST ( 18)  37 TREPGGT
KTHY_PASMU      6 FIVLEGIEGAGKTT ( 19)  39 TREPGGT
```

KTHY_PSEAE	5	FVTLEGPEGAGKST ( 18)	37	TREPGGT
KTHY_PYRAB	5	FVVLEGIDGSGKTT ( 18)	37	TKEPTDT
KTHY_PYRHO	5	FIVLEGIDGSGKTT ( 18)	37	TKEPTDS
KTHY_RICPR	9	FITFEGVDGIGKST ( 18)	41	TREVGGT
KTHY_SYNY3	5	FIVLEGIDGSGKTT ( 22)	41	TNGPVGR
KTHY_THEAC	2	FIAIEGIDGAGKTT ( 18)	34	TKEPTDG
KTHY_THEMEA	2	FITFEGIDGSGKST ( 18)	34	KREPGGT
KTHY_UREPA	18	FIVFEGIDGAGKTS ( 20)	52	TREPGGK
KTHY_VIBCH	5	FIVIEGLEGAGKST ( 19)	38	TREPGGT
KTHY_XYLFA	2	LVAIEGIDGAGKTT ( 22)	38	TNGPWGM
KTHY_YEAST	7	LILIEGLDRTGKTT ( 21)	42	STRIGGL
KTHY_YERPE	5	FIVIEGLEGAGKTT ( 19)	38	TREPGGT

	unknownC, width = 15		unknownD, width = 17	
KTH1_SULSO	( 40)	84	IILDRYFSSIA YQG ( 18)	117 PDMVILLDLPIEVAISR
KTHY_AERPE	( 43)	89	LVVDRYKYSSLAYQG ( 17)	121 AEILVYIDVPTEVALRR
KTHY_AQUAE	( 45)	86	VILDRFVLSTIAYQG ( 22)	123 PDITLLLDIPVDIALRR
KTHY_ARCFU	( 43)	84	VVMDRYYFSSNIAYQS ( 19)	118 PDLTILLDVEPEIALER
KTHY_ASF7	( 40)	90	VIVDHYIWSGLAYAQ ( 13)	118 PDYFFFLLSSKKPLNEKP
KTHY_BACHD	( 48)	93	VLCDRFIDSSLAYQG ( 22)	130 PDLTLLFRVDPDVGLSR
KTHY_BACSU	( 48)	92	VLCDRFIDSSPLAYQG ( 22)	129 PHVTVYFSDPPEGLKR
KTHY_BUCAI	( 49)	95	VISDRHDLSSLAYQG ( 22)	132 PDLTIYLDVSPDIGLAR
KTHY_CAMJE	( 45)	82	IISDRSFISGMAYAK ( 19)	116 PQKIIIFLKGDNLIQER
KTHY_CAUCR	( 48)	93	VLCDRFADSTRAYQG ( 22)	130 PVLTLILDLPAEVLQR
KTHY_CHLMU	( 47)	88	VICDRFHDSTIVYQG ( 24)	127 PDITFLLDLPEKGLLR
KTHY_CHLPN	( 48)	89	VICERFHDSTIVYQG ( 24)	128 PNFVLLLDIPADIGLQR
KTHY_CHLTR	( 47)	88	VISDRFHDSTIVYQG ( 24)	127 PDITFLLDIPAREGLLR
KTHY_DEIRA	( 47)	92	VLCDRYADSSLAYQG ( 22)	129 PGLTVLLDLDPALGLQR
KTHY_ECOLI	( 51)	96	VIGDRHDLSTQAYQG ( 22)	133 PDLTLYLDVTPVEVLKR
KTHY_HAEIN	( 48)	93	VVGDRHDMSSQAYQG ( 22)	130 PDLTIYLDIDPSVGLAR
KTHY_HALN1	( 45)	83	VISDRYSDRYAYQG ( 24)	122 PDVTLYFDVDPDTGAAR
KTHY_HELPJ	( 45)	82	IISDRSLISGMAYSQ ( 15)	112 PEKIIILLINKENLQQR
KTHY_HELPY	( 45)	82	IISDRSLISGMAYSQ ( 15)	112 PAKIIILLIDKEGLKQR
KTHY_LACLA	( 48)	92	VIVDRFIDSSVAYQG ( 22)	129 PDLTLYFDVDTDVALER
KTHY_METJA	( 44)	84	VVCDRYLYSSIA YQS ( 18)	117 PDIVFLLIVDIETALKR
KTHY_METH	( 47)	88	VVSDRCYSSMVYQG ( 14)	117 PDVVILLDIDVEVAMER
KTHY_MOUSE	( 40)	93	LVLDRYAFSGVAFTG ( 19)	127 PDLILFLQLQLLDAAR
KTHY_MYCGE	( 45)	92	VISDRFVFSFAYQG ( 22)	129 PNFVFFILDCNFKEALQR
KTHY_MYCPN	( 45)	92	VIVDRFVWSSYAYQG ( 22)	129 PDFVFFIVDCDFETALNR
KTHY_NEIMA	( 47)	91	VVSDRFTDATFAYQG ( 22)	128 PDLTLLLDVPLEVSMAR
KTHY_NEIMB	( 47)	91	VVSDRFTDATFAYQG ( 22)	128 PDLTLLLDVPLEVSMAR
KTHY_PASMU	( 48)	94	VIGDRHDMSSQAYQG ( 22)	131 PDLTLYLDIDPVLGLSR
KTHY_PSEAE	( 48)	92	VLCDRFTDATYAYQG ( 22)	129 PDLTLVFDLPVEIGLAR
KTHY_PYRAB	( 51)	95	VICDRYLYSSLAYQW ( 18)	128 PDLAILLDLPVKESIRR
KTHY_PYRHO	( 51)	95	VICDRYFYSSLAYQW ( 18)	128 PDLAILLDLPVKESLRR
KTHY_RICPR	( 45)	93	VICDRFIDSTACYQG ( 22)	130 PDITFFIDVEPHTAIKR
KTHY_SYNY3	( 51)	99	VITTRYFSSLAYNC ( 18)	132 PDWVIYLDLPVDLALQR
KTHY_THEAC	( 35)	76	VICDRYIRSSYAYQF ( 27)	118 PDVQIYVDVDEETAMER
KTHY_THEMEA	( 44)	85	VLLDRYTDSSVAYQG ( 22)	122 PDLTFYIDVDVETALKR
KTHY_UREPA	( 49)	108	VISDRFVHSSYIYQG ( 22)	145 IDYVFFYFDVNVNVALNR
KTHY_VIBCH	( 49)	94	VVGDRHDMSSQAYQG ( 22)	131 PDLTLYLDIDPKLGLER
KTHY_XYLFA	( 41)	86	VILDRYFPSMVAYQG ( 18)	119 PDVLLLLDVPPAIGLQR
KTHY_YEAST	( 41)	90	IVMDRYVYSGVAYSA ( 21)	126 PDLTLFLSTQDVDNNAE
KTHY_YERPE	( 49)	94	VVGDRHDLSSQAYQG ( 22)	131 PDLTLYLDLPPAVGLAR

[return to top]

COBBLER sequence from MOTIF>unknown KTHY\_ECOLI from 1 to 213 with embedded consensus blocks  
 mrskFIVIEGIDGAGKTTarnvvvetleqlgirdmvfTREPGGTqlaeklrslvldiksvgdevitdkaevlmfyaarv  
 q

lvetvikpalangtwVICDRYYYSSLAYQGggrgidqhmlatlrдавlgdfrPDLTIYLDVDPEVALQRarargeldrie  
gesfdffnrtrarylelaaqdksihtidatqpleavmdairttvthwvkelda

[Blast Search] of COBBLER sequence vs non-redundant protein database using default parameters  
[GAP-Blast Search] of COBBLER sequence vs non-redundant protein database using default parameters  
[[PSI-Blast Search] of COBBLER sequence vs non-redundant protein database using default parameters  
Structures found by BLAST: [3D Blocks] [About 3D Blocks]  
[return to top]

BLOCKS from GIBBSLogos: [Postscript][PDF][GIF] [About Logos] [Map]Tree:  
[Data][XBitmap][Postscript][PDF][GIF][Newick] [About Trees]Search: [LAMA]  
[About LAMA] [MAST] [About MAST]Primers:[CODEHOP] [About CODEHOP]Substitutions:[SIFT] [About SIFT on Blocks]  
\*\*BLOCKS from GIBBS\*\*

>unknown Q9UXG7 Probable thymidylate kinase 1 (EC 2.7.4.9) ... family  
44 sequences are included in 4 blocks

```
unknownA, width = 14unknownB, width = 14
KTH1_SULSO      4 LIAIEGIDGSGKTT ( 19)   37 TREPFSEDIKILIE
KTHY_AERPE      3 IVALEGIDGSGVST ( 18)   35 WKEPTEGPVGRLLR
KTHY_AQUAE      2 LIAFEGIDGSGKTT ( 18)   34 YREPGGTKVGEVLR
KTHY_ARCFU      2 LIAVEGIDGAGKTT ( 18)   34 LKEPGDSKFGKLIK
KTHY_BACHD      6 FITVEGEGEGAGKTS ( 18)   38 TREPGGIPIAEQIR
KTHY_BACSU      5 FITFEGPEGAGKTT ( 18)   37 TREPGGIDIAEQIR
KTHY_BUCAI      6 FIVIEGLEGAGKTN ( 19)   39 VRQPGSTPIAEDIR
KTHY_CAEEL     10 LIVFEGLDKSGKST ( 27)   51 QAFPDRSSSIGKLI
KTHY_CAMJE      2 YVVFEGIDCVGKST ( 14)   30 TLEPGGTELKHLR
KTHY_CAUCR      6 FISFEGEGEGAGKST ( 18)   38 TREPGGSPGAEAIR
KTHY_CHLMU      2 FIVVEGEGEGAGKTQ ( 18)   34 TREPGGALGEQLR
KTHY_CHLPN      2 FIVIEGEGEGSGKSS ( 18)   34 TREPGGCLIGERLR
KTHY_CHLTR      2 FIVVEGEGEGAGKTQ ( 18)   34 TREPGGCSLGDVSR
KTHY_DEIRA      6 FITLEGPEGAGKTT ( 18)   38 TREPGGTPLGTRVR
KTHY_ECOLI      5 YIVIEGLEGAGKTT ( 19)   38 TREPGGTQLAEKLR
KTHY_HAEIN      5 FIVIEGLEGAGKSS ( 19)   38 TREPGGTPLAEKLR
KTHY_HALN1      2 LVTLEGLDGSKTT ( 15)   31 TAEPTDSQYQAVR
KTHY_HELPJ      2 YVVEGVDGAGKST ( 14)   30 TKEPGGTRMGESLR
KTHY_HELPY      2 YVVEGVDGAGKST ( 14)   30 TKEPGGTRMGESLR
KTHY_HUMAN      8 LIVLEGVDRAGKST ( 18)   40 LRFPERSTEIGKLL
KTHY_LACLA      5 LISLEGPDGAGKTT ( 18)   37 TREPGGVVRVAEER
KTHY_METJA      6 FIVFEGIDGSGKTT ( 13)   33 TYEPSNSLVGKIIR
KTHY_METH       2 YICFEGIDGSGKTT ( 18)   34 VREPTDSNIGSLIR
KTHY_MOUSE      8 LIVLEGVDRAGKTT ( 18)   40 LRFPERSTEIGKLL
KTHY_MYCGE      6 FVVEGVDGAGKTA ( 20)   40 TREPGGTLLAEKIR
KTHY_MYCPN      6 FVAIEGVDGAGKTV ( 20)   40 SREPGGTPLAEKIR
KTHY_NEIMA      5 FITLDGIDGAGKST ( 18)   37 TREPGGTPVGEALR
KTHY_NEIMB      5 FITLDGIDGAGKST ( 18)   37 TREPGGTPVGEALR
KTHY_PASMU      6 FIVLEGIEGAGKTT ( 19)   39 TREPGGTPLAEKLR
KTHY_PSEAE      5 FVTLEGPEGAGKST ( 18)   37 TREPGGTPLAERIR
KTHY_PYRAB      5 FVVEGIDGSGKTT ( 18)   37 TKEPTDTEFGRLIR
KTHY_PYRHO      5 FIVLEGIDGSGKTT ( 18)   37 TKEPTDSELGKILR
KTHY_RICPR      9 FITFEGVDGIGKST ( 18)   41 TREVGGTTVAEKMR
KTHY_SCHPO      9 LIVIEGLDRSGKST ( 18)   41 FKFPDRRTTAIGKKI
KTHY_SYNY3      5 FIVLEGIDGSGKTT ( 18)   37 SPEPTNGPVGRLLR
KTHY_THEAC      2 FIAIEGIDGAGKTT ( 18)   34 TKEPTDGMENYAGD
KTHY_THEMEA     2 FITFEGIDGSGKST ( 18)   34 KREPGGTETGEKIR
KTHY_UREPA     18 FIVFEGIDGAGKTS ( 20)   52 TREPGGKNNNAEEM
KTHY_VACCV      6 LIVFEGLDKSGKTT ( 34)   54 TRKKTYNDHIVNLL
```

KTHY_VARV	6	LIVFEGLDKSGKTT ( 34)	54	TRKKTYNDHIVNLL
KTHY_VIBCH	5	FIVIEGLEGAGKST ( 19)	38	TREPGGTLLAEKLR
KTHY_XYLFA	2	LVAIEGIDGAGKTT ( 18)	34	SKEPTNGPWGMLLR
KTHY_YEAST	7	LILIEGLDRTGKTT ( 15)	36	LKFFPERSTRIGGLI
KTHY_YERPE	5	FIVIEGLEGAGKTT ( 19)	38	TREPGGTPLAEKLR

unknownC, width = 25

KTH1_SULSO ( 23)	74	WLSKIKDADLIILDRYFSSIAAYQG
KTHY_AERPE ( 30)	79	VERCGGSPEVLVVDYRYKYSSLAYQG
KTHY_AQUAE ( 28)	76	IIPDLKRDKVVILDRFVLSTIAYQG
KTHY_ARCFU ( 26)	74	ILPALQSGYAVVMDRYFVSNIAYS
KTHY_BACHD ( 31)	83	VLPALAEAGHVLCDFRIDSSLAYQG
KTHY_BACSU ( 31)	82	VKPALEQGFIVLCDFRIDSPAYQG
KTHY_BUCAI ( 32)	85	IKPALKNGIIVISDRHDLSSLAYQG
KTHY_CAEL ( 29)	94	IRDNIAKIDVICDRYCYSGVAYSL
KTHY_CAMJE ( 28)	72	EILKTNQNKLIISDRSFISGMAYAK
KTHY_CAUCR ( 31)	83	IRPGLARGAVVLCDFRFDSTRAYQG
KTHY_CHLMU ( 30)	78	ILPALESGKTVICDRFHDSTIVYQG
KTHY_CHLPN ( 31)	79	IIPALRDGYIVICERFHDSTIVYQG
KTHY_CHLTR ( 30)	78	IIPALKSGKTVISDRFHDSTIVYQG
KTHY_DEIRA ( 30)	82	LRPALERGETVLCDRYADSSLAYQG
KTHY_ECOLI ( 34)	86	IKPALANGTWVIGDRHDLSTQAYQG
KTHY_HAEIN ( 31)	83	IKPALMQGKVVVDRHDMSSQAYQG
KTHY_HALN1 ( 28)	73	VSPALDRGDVVISDRYSDRYAYQG
KTHY_HELPJ ( 28)	72	IKPALKEKKLIISDRSLISGMAYSQ
KTHY_HELPY ( 28)	72	IKPALKEKKLIISDRSLISGMAYSQ
KTHY_HUMAN ( 29)	83	IKEKLSQGVTLVVDYAFSGVAFTG
KTHY_LACLA ( 31)	82	MLPALQAGKVVIVDRFIDSSVAYQG
KTHY_METJA ( 27)	74	LIKEELKKRDVVCDRYLYSSIAYS
KTHY_METH ( 30)	78	KIEGDWAEDVVSDRCYYSMVYQG
KTHY_MOUSE ( 29)	83	IKAKLNQGVTLVLDYAFSGVAFTG
KTHY_MYCGE ( 28)	82	IKPAIEKEQLVIVDRFVSSFAAYQG
KTHY_MYCPN ( 28)	82	IQPALQKQLVIVDRFVWSSYAYQG
KTHY_NEIMA ( 30)	81	ILPALSDGIHVSDRFTDATFAYQG
KTHY_NEIMB ( 30)	81	ILPALSDGIHVSDRFTDATFAYQG
KTHY_PASMU ( 31)	84	IKPALAQKVVIGDRHDMSSQAYQG
KTHY_PSEAE ( 31)	82	IRPALARGAVVLCDFRFDSTRAYQG
KTHY_PYRAB ( 34)	85	ILPALEKGVVICDRYLYSSLAYQW
KTHY_PYRHO ( 34)	85	ILPALSEGKVVICDRYFYSSLAYQW
KTHY_RICPR ( 28)	83	IIPALKDGIVICDFRIDSTACYQG
KTHY_SCHPO ( 29)	84	IYEQINKGVTCILDRYAFSGIAFSA
KTHY_SYNY3 ( 38)	89	VEAKLAQQCHVITTRYFSSLAYNC
KTHY_THEAC ( 18)	66	IDRHIKNGEIVICDRYIRSSYAYQF
KTHY_THEMEA ( 27)	75	IKQYLSEGYAVLLDRYTDSSVAYQG
KTHY_UREPA ( 32)	98	INPHLEKDHIVISDRFVHSSYIYQG
KTHY_VACCV ( 11)	79	IQEQLQGITLIVDRYAFSGVAYAA
KTHY_VARV ( 11)	79	IQEQLQGITLIVDRYAFSGVAYAT
KTHY_VIBCH ( 32)	84	IKPALARGEVWVDRHDMSSQAYQG
KTHY_XYLFA ( 28)	76	IVPMIGRGAIVILDRYFVPSMVAYQG
KTHY_YEAST ( 30)	80	IKKDLLEGKNIVMDRYVYSGVAYSA
KTHY_YERPE ( 32)	84	IKPALARGSWVDRHDLSSQAYQG

unknownD, width = 18

KTH1_SULSO ( 18)	117	PDMVILLDLPIEVAISRI
KTHY_AERPE ( 17)	121	AEILVYIDVPTVEVALRRI
KTHY_AQUAE ( 22)	123	PDITLLLDIPVDIALRRL
KTHY_ARCFU ( 19)	118	PDLTILLDVEPEIALERV
KTHY_BACHD ( 22)	130	PDLTLLFRVDPDVGLSRI
KTHY_BACSU ( 22)	129	PHVTVYFSDPEEGLKRI
KTHY_BUCAI ( 22)	132	PDLTIYLDVSPEIGLARA
KTHY_CAEL ( 18)	137	PDAVLFDFVSPEVAAQRG
KTHY_CAMJE ( 19)	116	PQKIIFLKGDNLIQERL
KTHY_CAUCR ( 22)	130	PVLTLILDLPAEVGLQRA

KTHY_CHLMU	( 24)	127	PDITFLLDLPEKEGLLRK
KTHY_CHLPN	( 24)	128	PNFVLLLDIPADIGLQRK
KTHY_CHLTR	( 24)	127	PDITFLLDIPAREGLLRK
KTHY_DEIRA	( 22)	129	PGLTVLLDLDPALGLQRA
KTHY_ECOLI	( 22)	133	PDLTLYLDVTPPEVGLKRA
KTHY_HAEIN	( 22)	130	PDLTIYLDIDPSVGLARA
KTHY_HALN1	( 24)	122	PDVTLYFDVDPDTGAARS
KTHY_HELPJ	( 15)	112	PEKIILLINKENLKQRL
KTHY_HELPY	( 15)	112	PAKIILLIDKEGLKQRL
KTHY_HUMAN	( 19)	127	PDLVLFQLQLADAAKRG
KTHY_LACLA	( 22)	129	PDLTLYFDVDTVALERI
KTHY_METJA	( 18)	117	PDIVFLLIVDIETALKRV
KTHY_METH	( 14)	117	PDVVILLDIDVEVAMERC
KTHY_MOUSE	( 19)	127	PDLILFLQLLLDAAARG
KTHY_MYCGE	( 22)	129	PNFTFILDNCFKEALQRM
KTHY_MYCPN	( 22)	129	PDFTFIVDCDFETALNRM
KTHY_NEIMA	( 22)	128	PDLTLLLDVPLEVSMARI
KTHY_NEIMB	( 22)	128	PDLTLLLDVPLEVSMARI
KTHY_PASMU	( 22)	131	PDLTLYLDIDPVLGLSRA
KTHY_PSEAE	( 22)	129	PDLTLVFDLPVEIGLARA
KTHY_PYRAB	( 18)	128	PDLAILLDLPVKESIRRT
KTHY_PYRHO	( 18)	128	PDLAILLDLPVKESLRRRI
KTHY_RICPR	( 22)	130	PDITFFIDVEPHTAIKRV
KTHY_SCHPO	( 18)	127	PDLVIFLNVDPRIAATRG
KTHY_SYNY3	( 18)	132	PDWVIYLDLPVDLALQRL
KTHY_THEAC	( 27)	118	PDVQIYVDVDEETAMERI
KTHY_THEMEA	( 22)	122	PDLTFYIDVDVETALKRK
KTHY_UREPA	( 22)	145	IDYVFYFDVNVNNAALNRM
KTHY_VACCV	( 18)	122	PDLVIFLESGSKEINRNV
KTHY_VARV	( 18)	122	PDLVIFLESGSKEINRNV
KTHY_VIBCH	( 22)	131	PDLTLYLDIDPKLGLERA
KTHY_XYLFA	( 18)	119	PDVLLLLDVPPAIGLQRI
KTHY_YEAST	( 21)	126	PDLTLFLSTQDNDNAEK
KTHY_YERPE	( 22)	131	PDLTLYLDLPPAVGLARA

[return to top]

COBBLER sequence from GIBBS>unknown KTHY\_PASMU from 1 to 209 with embedded consensus blocks  
mttgkFIVIEGIDGAGKTTardsivralhahgihdivfTREPGGTPVGEKLRqlikheteeptvtdkaellmlyariql  
v  
envIKPALKQGVVICDRYYYSSIAAYQGggrqldqhlhltlktilgefepDLTIFLDVDPEVALQRVkgrgaldrieq  
q  
nldffhrtrqryqelvrhnpkavtidasqtmskvaedvesaietwltr

[Blast Search] of COBBLER sequence vs non-redundant protein database using default parameters  
[GAP-Blast Search] of COBBLER sequence vs non-redundant protein database using default parameters  
[[PSI-Blast Search] of COBBLER sequence vs non-redundant protein database using default parameters  
Structures found by BLAST: [3D Blocks] [About 3D Blocks]  
[return to top]

[BLOCKS home] [return to top]