

## รายการอ้างอิง

- Altschul *et al.*, 1997. Gapped BLAST and PSI-BLAST : A new generation of protein database search programs. Nucleic Acids Res 25 : 3899-3402.
- Altschul *et al.*, 1990. Basic local alignment search tool. J. Mol. Bio 215 : 403-410.
- Andreas, et al.1987. SNNS User Manual. Version 3.3. University of Stuttgart.
- Baker D. and Riddle. 1997, Functional rapidly folding proteins from simplified amino acid sequences. Nature Structural Biology : 805.
- Beale M., Howard B., Demuth., and Martin T. Hagan. 1995. Neural Network Design. PWS Publishing Company.
- Benchetrit, T., Bissery, V., Gaboriaud, C. and Mornon, J.P. 1987. Hydrophobic cluster analysis : an efficient new way to compare and analyse amino acid sequence. FEBS Lett 224 : 149-155.
- Berzofsky J.A., Cease K.B., Comette J.L., Delisi C., Margalit H. and Spouge J.L. 1987. Hydrophobicity scales and computational techniques for detecting amphiphatic structure in protein. J. Mol. Biol 195 : 659-685.
- Bissery V., Gadorriaud C., Henrissat B., Lemesie L., Morgat V. and Mornon. J.P. 1990. Hydrophobic cluster analysis: to derive structural and functional information from 2-D representation of protein sequence. Biochimie 72 : 555-574.
- Biosym/MSI.1995. Homology 95.0 Protein Structure Prediction User Guide. San Diego.
- Cohen J.I., Moriuchi H., Moriuchi M., Pichyanghura R., Starus S.E. and Triezenberg S.J. 1995. Hydrophobic cluster analysis prediction an amino terminal domain of varicella-zoster ORF 10 required for transcriptional activation. National Academy of Science 92 : 9333-9337.

- Cox., Lehninger. and Nelson. 1993. Principles of Biochemistry. New York. :  
Worth Publishers.
- Creighton T. E.1984. Proteins , Structure Molecular principles. New York : W.H.  
Freeman and Company.
- Dudchak I., Muchnik I., And Mayoraz E.N. 1995. Relative between Protein  
Structure, Sequence Homology and Composition of amino Acid.  
DIMACS Technical Report 95-6.
- Eddy, S. 1995. Multiple alignment using hidden Markov models. ISMB 95 : 114-  
120.
- Edmunson A.B. and Schiffer M. 1967. Use of helical wheels to represent the  
structures of protein and to identify segments with helical potential.  
Biophys J. 7 : 121-135.
- Gregory K.1997. Special Edition Using Visual C++ 5. USA : Que Corporation.
- Henrissat B., Mornon JP., Raimbaud E. and Tran V. and 1990. HCABAND : A  
computer program for the 2D-Helical representation of protein  
sequences. CABIOS 6 : 3-5.
- Henrissat B., Mornon JP., and Woodcock S. 1992. Detection of secondary  
structure elements in proteins by hydrophobic cluster analysis. Protein  
Engineering 5 : 629-635.
- Hertz J., Krongh A., and Palmer R.G. 1991. Introduction to the Theory of Neural  
Computation. Addison-Wesley Publishing Company.
- Hinton, .G.E., Rumelhart, D.E. and Williams, R.J. 1986. Learning  
representations by back-propagating errors. Nature 323 : 533-536.
- Hooft R.W.W., Rodriguez R., Sander C. and Vriend G. Protein modelling  
homology, How good are the model. Germany : EMBL-EBI.
- Leinecker R.C. and Yao P., 1997. Visual C++ 5 Bible. USA : IDG Group  
Company.

- Needleman, SB. and Wunsch, CD.1969. A general method to search for a similarities in the amino acid sequence of two protein. J. Mol. Biol 48 : 443-453.
- Skapura D.N., and Freeman J.A.1991. Neural Networks Algorithms , Applications, and Programming Techniques. Addison-Wesley Publishing Company.

## ภาคผนวก ก

โปรตีนที่ใช้เป็น Training set and Test set จำนวน 5 กลุ่ม คือ

- |                               |              |
|-------------------------------|--------------|
| 1. Myoglobin                  | จำนวน 17 ตัว |
| 2. Fibronectin                | จำนวน 14 ตัว |
| 3. TBP                        | จำนวน 13 ตัว |
| 4. Transmembran               | จำนวน 13 ตัว |
| 5. Triose phosphate Isomerase | จำนวน 14 ตัว |

**This is a sequenced for Fibronectin****Structural class:****Sequence****14 sequence**

```

1. DEFINITION fibronectin 1.
ACCESSION   AAA41166
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha;

ORIGIN
    1  VPGSKSTATI NNIKPGADYT ITLYAVTGRG DSPASSKPVV INYQTEIDKP SQMQVTDVQD
    61  NSISVRWLPS TSPVTGYRVT TAPKNGLGPT KSQTVSPDQT EMTIEGLQPT VEYVVSVYAQ
   121  NRRNGESQPLV QTAVTTIPAP TNLKFTQVSP TTLTAQWTAP SVKLTGYRVR VTPKEKTGPM
   181  KEINLSPDST SVIYSGLMVA TKYEVSVYAL KDTLTSRPAQ GVVTTLENVS PRRRARTDA
   241  TETTITISWR TKTETITGFQ VDAIPANGQT PVQRTISPDV RSYTITGLQP GTDYKIHLYT
   301  LNDNARSSPV VIDASTAIDA PSNLRFLTTT PNSLLVSNQA PRARITGYII KYEKPGSPPR
   361  EYVPRPRPGV TEATITGLEP GTEYTIYVIA LKNNQKSEPL IGRKKTDELP QLVTLPHPNL
   421  HGPEILDVPS TVQKTPFVTN PGYDTENGIQ LPGTSHQQPS VGQQMIFEEH GFRRTTPPTA
   481  ATPVRLRPRP YLPNVDVEEQ IGHVPRGDVD YHLYPHVPGL NPNASTGQEA LSQTTISWTP
   541  FQESSEYIIS CQPVGTDDEP LQFQVPGTST SATLTGLTRG VTYNIIVEAL HNQRHVKVRE
   601  EVVTVGNTVN EGLNQPTDDS CFDPYTVSHY AVGEEWERLS DSAFKLTCQC LGFGSGHERC
   661  DSSKWHCHDNQ VNYKIGEKWD RQGENGQRMS CTCLGNGKGE FKCDPHEATC YDDGKTYHVG
   721  EQWQKEYLGA ICSETDFGGQ RGWRCDNCRP PGAAEPSDQ TTGHTYNQYT QRYHQRTNTN
   781  WNCPIECFMP LDVQADRODS RE

//
2. DEFINITION [Segment 1 of 2] FIBRONECTIN (FN).
ACCESSION   FINC_CHICK_1
SOURCE      chicken.
ORGANISM    Gallus gallus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
            Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

ORIGIN
    1  pldqcdset rtfyqigdsx ekynvgvryq cycygrgige whcqpqaya

//
3. DEFINITION [Segment 2 of 2] FIBRONECTIN (FN).
ACCESSION   FINC_CHICK_2
SOURCE      chicken.
ORGANISM    Gallus gallus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
            Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

ORIGIN
    1  PLSPTNLRL EPNPDTGILI VSWDRSTTPG ISGYRVTTAP TNGQQGSTLE EVWGADQTSK
    61  TFENLNPQVE YNVSVYAVKD DQESIPISKT ITQEVPLQTD LSFVDITDSS IGLRWTPLNA
   121  STIIGYRLTV VAAGESVPIF EDFVDSSVGY YTVTGLEPGI DYDISVITLI NGGESAPTTI
   181  TQQTAVPPPT DLRFTNVGPD TMRVTWTAPT SIVLSSFLVR YSPVKKEEDV AELTISPSDN
   241  YVLTNLLPG TEYLVRVYSV AEQHEAPLS GIQKTGLDSP TGLDFSDITA NSFVHWIAP
   301  RATITGYKIR HHPEHGVGRP KEDRVPPSRN SITLTNLLPG TEYVVSIIAV NGREESVPLV
   361  GQQTIVSDVP RDEVNPTSP TSLEISWDAP AVTVRYRIT YGETGGSSPV QEFTVPGTMS
   421  RATITGLKPG VDYITIVYAV TGRGDSPASS KPVTVYKTE IDTPSQMQVT DVQDNSISIR
   481  WLPSSSPVTG YRVTAVPKKG HGPTKTKNVP PDQTQVTIEG LQPTVEYMVS VYAQNQNGES
   541  LPLVETAVTN IDRPKGLTFT EVDVDSIKIA WESPQQQVTR YRVTYSSPED GIHELLPAPG
   601  GEEDTAEHLG LRPQSEYTN IVAIYDMES LPLTGTQSTA IPPPTNLKFT QVTPTSLTVN
   661  WMAPNVRTLG YRVRVNPKEK TGPMKEINLS PDSTSAVVSQ LMVATKYEVS VYALKDLSLTS
   721  RPAQGVVTTL ENVSPRRAR VDATETTIT ITWRKTETI TGFQIDAI PA ASGNPIQRT
   781  ISPDVRTYTI TGLQPGNDYK IYLYTLNENA RSSPVVIDAS TAIDAPSNLR FLTTTTNSLL
   841  ASWQPPRAKI TGYIIRYDKP GSPAPELLPR PRPGTTEATI TGLEPGTEYT IYIIAVKNNQ
   901  KSEPLVGRKR TDDLPTLITG PHPNQPDMLD VPSVDEGTPY LTNNRYDNGN GIQLPGTSGH
   961  PQTIGHQQQQ VFEEHGYRR VPPTATPLR PGRRRQPNV DEAEIPGYQ VPIIVVPSYP
  1021  HSREPRRNDT TGQEALSQTT ISWRPLLEST EYIISCQPVV QEDDTLQFRV PGTSSSATLT
  1081  GLTRGATYNI IVEALKDHRH QKVLEEVTV GNTVSEGLHQ PADDTCYDTY TGSFYSIGEE
  1141  WERLSETGFK LWCQCLGFGS GHFRCDSSKW CHDNGVNYKI GEKWRDQGEN GQMIDCTCLG
  1201  NGKGEF

//
4. DEFINITION fibronectin.
ACCESSION   AAD24483
SOURCE      Oryctolagus cuniculus.
ORGANISM    Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Lagomorpha; Leporidae; Oryctolagus.

ORIGIN

```

1 TSLGISWEAP AVTVRYRIT YGETGGNSPV QEFTVPGSKS TATISGLKPG ADYITIVYAA  
 61 TGRGDSPASS KPISIDYHTE IDKPSQMQVT DVQDNSISVR WLPSSSPVTG YRVTTTPKNG  
 121 AGPTKTKTAG PDQTEMTIEG LQPTVEYVVS VYAQRNGES QPLVQTAVTT IPAPTNLKFT  
 181 QVTPTSLSAQ WTPPNVQLTG YRVRVTPQEK TGPMEINLA PDSSSVVSG LMVATKYEVS  
 241 VYALKDTLTS RPVQGVITTL ENVSPRRAR VTDATETTIT ISWRKTETI TGF

//

## 5. DEFINITION FIBRONECTIN (FN).

ACCESSION Q91289

SOURCE Iberian ribbed newt.

ORGANISM Pleurodeles waltl

Eukaryotae; Metazoa; Chordata; Vertebrata; Amphibia; Batrachia;  
Caudata; Salamandroidea; Salamandridae; Pleurodeles.

## ORIGIN

1 HESEAPITRT VVTPISPTD LHLESPDFA TLTVSWRSR SPGITGYRIN TALLLGIRLH  
 61 SGYTLEEEVT ESQSRVICFD NLSPGVEYNV SVVSVKDDQE SEPIWKTITQ EVPSLTDLNF  
 121 VDVTDTSIDL RWTPLKGPTI IGYRVTVVAA GESVPIYEDK VGPTQGYKVV SGLEPGIDYD  
 181 ISVITLVTDG ESAPTTLTTA DCCPTATDLR FTNVGPDMSL VTWSAPPSMV LSSFLVRYVP  
 241 SKNEEDAEL TISPSDNMVV LTNLLPGTEY IVSVFAVYEE RESTPLTGVQ RTGLDSPTGL  
 301 DFSDTTSSEF TVYWMAPRAT VTGYKIYHP ETGGAGQKEE RCVPPSRNSL TLTNLTPTGE  
 361 YVVSIFAVNG RQESVPLVGQ QATVSDTPTN LEVTSSTPTS MSISWDAPPV GVRYYRITYT  
 421 ETGGETPVQE FTVPGDRSDA PIRGLKPGAE YIITVYAVTG RGDSPASSKP VTVTHKTVVD  
 481 KPTQLQVTDV QDHSIQVRWM PSSTPVTYGR VTSVPKSGVG PTVSHVVPD QTEMTIEGLE  
 541 PTVEYVVSYY AQGNGETEP LVETAVTNID RPKGLAFTEV DVDSLKLWVE SPKGQVTTYK  
 601 VTYSNPEDGI HELVPAPNGD EDTAQLHGLR PGAEYTVSVV ALHDDMESQP LIGTQVTAIP  
 661 PPTNLLFSQI TPTSVTVSWR PPNVQLTGYR VRVHPKEKAG PMKEINLSPD STSAVVTGLM  
 721 VATKYEVSYY ALKDSLTSRP AQGIVTTQEN VSPRRRRIT DVTETTITIT WRKTETITIG  
 781 FHIDAIPAGG QNPIQRTISP DLRTYVITGL QPGTDYKIH IYTLNDNARSS PVTIDATTAV  
 841 DSPSNLRFIT TTSNLLFSW QPPRSKLTGY IIKYEKPGGP VREVVRPHG GATESQOSQN  
 901 LEPGTEYTIY IIAVRSNYKS GPLVGGKRTD ELPRLVTLAQ PGQOGRILDV PSLVENTPFI  
 961 SQTSTFDNGNG IQLPGTSGQQ NIGHGQQQVF LEEHGFRSPV LPTTATPVKP GYQAPEQHTL  
 1021 DKYTPGQHPV TIREDIELAR FPPRQIDMDR PSHDSDGPPQ VDRTGQEAQT TISWRPLET  
 1081 TEYIITCHPV GNEETPQQFT VPGTSSSATL NGLTRGATYN IIVEALKGKN KHKSRELVTV  
 1141 TSAAHGSGVL HGLEDTCYDI TTGSSYSIGQ EWERMSESGF KLWCRCGLG SGHFKCDSSK  
 1201 WCHDSGLNYR VGEKWDRSGE NGQMSCTCL GMEWKGEFK EPHEATCYDD GKMYVGEQWQ  
 1261 KEYLGAICSC TCYGGQGGWR CDNCRPPGAG VTPSADGVVG QTLSHFSQRY QONANFNLKC  
 1321 PIECYLPL

//

## 6. DEFINITION FIBRONECTIN (FN).

ACCESSION Q28275

SOURCE dog.

ORGANISM Canis familiaris

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Carnivora; Fissipedia; Canidae; Canis.

## ORIGIN

1 YEVSVALYALD TLTSRPAQGV VTTLENSVPP RRARVTDAT TITISWRKT TETITGFQVD  
 61 AIPANGQMPI QRTIRPDVRS YTTIGLQPGT DYKIYLYTLN DNARSSPVVI DASTAIDAPS  
 121 NLRFLATTPN SLLVSWQPPR ARITGYIYK EKPSPPREV VPRPRPGVTE ATITGLEPGT  
 181 EYTIQYALK NNQKSEPLIG RKKTDELPLQ VTLPHPNLHG PEILDVPTV QKTPFITNPG  
 241 YDTGNGIQLP GTSQQQPSVG QQMIFEEHGF RRTTPTTAT PVRHRPRYP PNVNEEIQVG  
 301 HVPRGDDVHH LYPHVMGLNP NASTGQEALS QTTISWTFPQ ESSEYIISCH PVGIDEPLQ  
 361 FRVPGTSASA TLTGLTRGAT YNIVEALKD QKRHKVREEV VTVGNSVDQG LNQPTDDSCF  
 421 DPYTVSHYAI GEEWERLSES GFKLSCQCLG FSGHFRCD S KWCHDNGVN YKIGEKWDRQ  
 481 GENGQMSCT CLGNGKGEFK CDPHEATCYD DGKTYHVGEH WQ

//

## 7. DEFINITION fibronectin.

ACCESSION AAD00019

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

## ORIGIN

1 TFDNLSPLGLE YNVSVYTVKD DKESVPISDT IIPAVPPPTD LRFTNIGPDT MRVTWAPPPS  
 61 IDLTNLFVRY SPVKNEEDVA ELSISPSDNA VVLTNLLPGT EYVSVSSVY EQHESTPLRG  
 121 RQKTGLDSPT GIDFSDITAN SFTVHWIAPR ATITGYRIRH HPEHFSGRPR EDRVPHSRNS  
 181 ITLTNLTPTG EYVVSIVALN GREESPLLIG QQSTVSDVPR DLEVVAATPT SLLISWDAPA  
 241 VTVRYRITY GETGGNSPVQ EFTVPGSKST ATISGLKPGV DYTITVYAVT GRGDSPASSK  
 301 PISINYRTEI DKPSQMQVTD VQDNSISVKW LPSSSPVTGY RVTTTPKNGP GPPTKTKTAGP  
 361 DQTEMTIEGL QPTVEYVVSV YAQNPSGESQ PLVQTAVTTI PAPTDLKFTQ VTPPTSLSAQW  
 421 TPNVLTPTG RVRVTPKEKT GPMKEINLAP DSSSVVVSGL MVATKYEVSV YALKDTLTSP  
 481 PAQGVVTTLE NVSPRRRARV TDATETTITI SWRKTETIT GFQVDVAVPAN GQTPIQRTIK  
 541 PDVRSYITG LQPGTDYKIY LYTLNDNARS SPVIDASTA IDAPSNLRFI ATTPNSLLVS  
 601 WQPPRARITG YIIKYEKPGS PPREVVPRPR PGVTEATITG LEPGTEYTIY VIALKNNQKS  
 661 EPLIGRKKTV QKTPFVTHPG YDTGNGIQLP GTSQQQPSVG QQMIFEEHGF RRTTPTTAT

721 PIRHRPRPYP PNVGQEALSQ TTISWAPFQD TSEYIISCHP VGTDEEPLQF RVPGTSTSAT  
781 LT

//  
8. DEFINITION fibronectin.

ACCESSION CAA63654

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ORIGIN

1 GTPRSDNVPP PTDLQFVELT DVKVTIMWTP PDSVVSQYRV EVLPVSLPGE HGQRLPVNRN  
61 TFAEITGLSP GVTYLFKVFV VHOGRESNPL TAQQTTKLDA PTNLQFVNET DRTVLVWTWP  
121 PRARIAGYRL TAGLTRGGQP KQYNVGPLAS KYPLRNLQPG SEYTATLVAV KGNQOQSPKAT  
181 GVFTLQPLR SIPPYNTEVT ETTIVITWTP APRIGFKLGV RPSQGGAPR EVTSDSGSIV  
241 VSGLTPGVEY TYTIQVLRDQ QERDAPVNR VVTPLSPPTN LHLEANPTDG VLTVSWERST  
301 TPDITGYRIT TPTNGQQGT SLEEVHADQ SSCTFENLNP GLEYNVSVYT VKDDKESAPI  
361 SDTVVPEVPO LTDLSEFVDIT DSSIGLRWTP LMSSTIIGYR ITVVAAGEGI PIFEDFVDS  
421 VGYTVTGLE PGIDYDISVI TLINGGESAP TLTQQTAVP PPTDLRFTNI GPDTRMTWA  
481 NPPSPDSSVY VSGLMVATKY EDVAELSISP SDNAVVLNLT LPGTEYLVSV SSVYEQHESI  
541 PLRGRQKTGL DSPTGFDSO ITANSFTVHW VAPRAPITGY IIRHHAHSV GRPRQDRVPP  
601 SRNSITLTLN NPGEYVYSI IAVNGREES PLIGQQATVS DIPRDLVIA STPTSLLSW  
661 EPPAVSVRY RITYGETGGN SPVQEFVPG SKSTATINNI KPGADYTTITL YAVTGRGDSP  
721 ASSKPVSIYNY KTEIDKPSQM QVTDVQDNSI SVRWLPSTSP VTGYRVTTPK WLTQGNKQ  
781 TASPQTEMT IEGLOQTEY VSVYAQNRN GESQPLVQTA VTNIDRPKGL AFTDVEDVDSI  
841 KIAWESPQGO VSRYRVYSS PEDGIRELFP APDGEDDTAE LQGLRPPSEY TVSVVALHDD  
901 MESQPLIGIQ STAIAPATNL KLSQVTPTSF TAQWIAPSVQ LTGYRVRVNP KEKTGPMKEI  
961 NPPSPDSSVY VSGLMVATKY EVSVYALKDT LTRPAQFVI TLENVSPPR RARVTDATET  
1021 TITISWRKTK ETITGFQVTA IPANGQTPVQ RSISPDRSY TITGLQPGTD YKIHLYTLND  
1081 NARSSPVEID ASTAIDAPSN LRFLTTTPNS LLVSWQAPRA RITGYIIKYE KPGSPPREVV  
1141 PRPRPGVTEA TITGLEPGE TYIYVIALKN NQKSEPLIGR KKTDELPLQV TLPHPNLHGP  
1201 EILDVPSVY KTFPITNPGY DTENGIQLPG TTHQQPSVQ QMIFEEHGR RTTPPTAATP  
1261 VRLRPRPYLP NVDEEVQIGH VPRGDVDYHL YPHVPLNPN ASTGQEALSQ TTISWTFQE  
1321 SSEYIISQCP VGTDEEPLQF QVPGTSTSAT LTGLTRGVTY NIIVEALQNO RHHKVREEVV  
1381 TVGNAVSEGL NQPTDDSCFD PYTVSHYAIG EEWERLSDAG FKLTCQCLGF GSGHFRCDSS  
1441 KWCHDNGVNY KIGEKWDRQS ENGQRMSCCT LGNGKGEF

//

9. DEFINITION FIBRONECTIN (FN).

ACCESSION P07589

SOURCE Bos taurus.

ORGANISM Bos taurus

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Artiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae;

ORIGIN

1 QAQQIVQPS PLTVSQSKPG CYDNGKHYQI NQQWERTYLG SALVCTCYGG SRGFNCSEK  
61 EPEETCFDKY TGNTYRVGDT YERPKDSMIW DCTCIGAGRG RISCTIANRC HEGGQSYKIG  
121 DTWRRPHETG GYMLECVCGL NGKGWETCKP IAECFCFQAA GTSYVVGETW EKPVQGNMNV  
181 DCTCLGEGSG RITCTSPNRC NDQDTRTSYR IGDWTSKKN RGNLLQCIC TNGRGWEKCE  
241 IGRSLQTTSA GSGSFTDVRT AIYQOPHPQ PPPYGHCVTD SGVYVSVGMQ WLTQGNKQ  
301 LCTCLNGVVS CQETAVTQTY GGNSNGPCV LPFTYNGKTF YSCTTEGRQD GHLWCSTTSN  
361 YEQDQKYSFC TDHTVLVQTR GGNSNGALCH FPFLYNNHNY TDCTSEGRD NMKWCCTQ  
421 YDADQKFGFC PMAAHEEICT TNEGVMYRIG DQWDKQHDMG HMRCTCVGN GRGEWTCVAY  
481 SLDLQDCIVD GITYNVNDTF HKRHEEGHML NCTCFGQGRG RWKCDPVDQC QDSETRTFYQ  
541 IGDSEWEKYLQ GVRYQCYCYG RIGIEWACQP LQTYPDTSGP VQVIITETPS QPNSHPIQWS  
601 APESHISKY ILRWKPKNSP DRWKEATIPG HLNSYTIKGL RGVVYEGQL ISVQHYGQRE  
661 VTRFDFTTTS TSPAVTSTNTV TGETTPLSPV VATSESVTEI TASSFVVSUV SASDTVSGFR  
721 VEYELSEEGD EPQYLDLPST ATSVNIPDLL PGRKYTVNVY EISEEGEQL ILSTSQTTAP  
781 DAPPDPTVDQ VDDTSIVVRW SRPRAPITGY RIVYSPSVEG SSTEINLNPET ANSVTLSDLO  
841 PGVQYNITII AVEENQESTP VFIQOETTGV PRSDKVPPPR DLQFVEVTDV KITIMWTPPE  
901 SPVTGYRVDV EPVNLPGHEG QRLPVSRNTE AEVTGLSPGV TYHFKVFAVN QGRESKPLTA  
961 QATKLDAPT NLQFINETDT TVIVTWTTPR ARIVGYRLTV GLTRGGQPKQ YNVGPAASQY  
1021 PLRNLQPGSE YAVSLVAVKG NQSPRVTVG FTTLQPLGSI PHYNTEVET TIVITWTPAP  
1081 RIGFKLGVPR SQGGEAPREV TSESGSIVVS GLTPGVEYVY TISVLRDQGE RDAPIVKKVV  
1141 TPLSPPTNLH LEANPDTGVL TVSWERSTTP DITGYRITT PTNGQQYSL EEVVHADQSS  
1201 CTFENLSPGL EYNVSVYTVK DDKESVPISD TIIPAVPPPT DLRFNTVGPD TMRVWAPP  
1261 SIELTNLLVR YSPVKNEDV AELSISPSDN AVVLNLLPG TEYLVSVSSV YEQHESIPLR  
1321 GRQKTALDSP SGIDFSDITA NSFTVHWIAP RATITGYRIR HHPENMGRP REDRVPPSRN  
1381 SITLTLNLPNG TEYVVSIVAL NSKEESLPLV GQSTVSDVP RDLEVIATP TSLLSWDAP  
1441 AVTVRYRIT YGETGGSSPV QEFVPGSKS TATISGLKPG VDYITIVYAV TGRGDSPASS  
1501 KPVSIYRTE IDKPSQMVT DVQDNSISVR WLPSSSPVTG YRVTTAPKNG PGPSKTKTVG  
1561 PQTEMTIEG LQPTVEYVVS VYAQNQNGES QPLVQAVTT IPAPTNLKFT QVTPTSSTAQ  
1621 WAPNVQLTG YRVRVTPKEK TGPMKEINLA PDSSSVVVSQ LMVATKYEVS VYALKDITLS  
1681 RPAQGVVTTL ENVSPRRAR VTDATETTIT ISWRKTETI TGFQVDAIPA NGQTPIQRTI  
1741 RPDVRSYIT GLQPGTDYKI HLYTLNDNAR SSPVVIDAST AIDAPSNLRF LATTPNLSLLV

1801 SWQPPRARIT GYIIKYEKPG SPREVVRP RPVTEATIT GLEPGTEYTI QVIALKNNQK  
 1861 SEPLIGRKKT DELPQLVTL PHLNLHGPEIL DVPSTVQKTP FITNPGYDTG NEEQLPGTSS  
 1921 QQPSLGGQMI FEHGFRRRT PPTTATPVRH RRPYPNPNVN EEIQIGHVPR GDVDHLLYPH  
 1981 VVGLNPNAST GQEALSQTTI SWTPFQESSE YIISCHPVGI DEEPLQFRVP GTSASATLTS  
 2041 LTRGATYNI IVEAVKDDQQRQ KVREEVVTVG NSVDQGLSQP TDDSCFDPYT VSHYAIGEEW  
 2101 ERLSDSGFKL SCQCLGFGSG HERCDSSKWC HDNGVNYKIG EKWDRQGENG QMSCTCLGN  
 2161 GKGEFKCDPH EATCYDDGKT YHVGEQWQKE YLGAICSCTC FGGQRGWRC DRRRPGAEPG  
 2221 NEGSTAHSYN QYSQRYHQRT NTNVCNPIEC FMPLDVQADR EDSRE

//

10. DEFINITION fibronectin - African clawed frog.

ACCESSION A43908

SOURCE African clawed frog.

ORGANISM *Xenopus laevis*

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;

Pipidae; Xenopodinae; Xenopus.

ORIGIN

1 MRRGALTGLL LVLCLSVVLR AAPSATSKKR RQAQQQQQQQ VVQPOGTQDN HNGCYDNGK  
 61 YYQINQQWER TYLGNTLVCT CYGGGRGFCN ESKPESEETC FDKYTGVSYSR VSEYERPKD  
 121 NMIWDCTCIG AGRGRISCTI ANRCHEGGQS YKIGDTWRRP HETGGYMLEC VGLNGKGEW  
 181 TCKPVAERCY DNTAGTSYVV GQTWEKPYQG WMMVDCTCLG EGNRITCSS FNRCDNDQDTK  
 241 TSYRIGDTWS KTDTRGNLLQ CICTGNRGE WKCRHSSAQ ATGTGSNPIT NITALYQPD  
 301 SQLEPYGHCV TDNGVLYSLG MRWLRTQGSK QMLCTCLNG VSCEETVATI TFGNANGEP  
 361 CAIPFTHDGI TYYSCTSEGR QDGKLCWATT SNYDSDKKYS FCTEQLALVQ TRSGNSGAL  
 421 CNFPFLYNNL NYTDCSTSEGR QDSMKWCGTT ANYDADQKFG FCPMAAHEEI CTNEGVMYR  
 481 VGDQWDMQHD QGHMMRQTCV GNRGEWSSV AYSQLKDCQI VDGLTYNNS SFTKLHEEGH  
 541 MNMCTCFGGG RGRWKCDAID CQCDTETRQF YQIGDSWEKH LQGVQYQCYC YHFGIGEWHC  
 601 QPLSTSQAGT GPVQVIIIES ANEPTSHPIQ WNAPOASHIK NYILRWPKL FHPBPWKQATI  
 661 PSHLNSYTIIS GLKPGILYEG QLISILQYCN REVTTFDFTT TTIHRHSQT ESBETPLPP  
 721 LVSISESVTE ITASSFLVSW VSASDTVSGF RVEYELSEDG DEKRYLELPN TATSVINPDL  
 781 LPGRRYNMYV YQITEEGEKS LILSTTQTTA PDAPPERHVE NVEDTSEMIL WHPQAPITG  
 841 YRVVYSPSVE GSSTELNLP SANSVTLTEL LPGIEYNITI YAVEDSLSV PVEIQQGTG  
 901 TPQTVIVPSP TDLQLVEVTD VKIIMWTSP QSEVSGYRVV VKPVSPASRD VCNLPVNRNT  
 961 FAEVVNLQPG RTYSFEVYAV NRGQSEPLV GEFATKLDAP TDLQFTDTE STVVIWIIPP  
 1021 QAKIGRYLLS VQTRGGQPS QFPINPSVTN HKLDNLLPST ETVSVLWALK GNGQSASASG  
 1081 VFTLEPVGS IPHYNTEVT TTIIVTWTPI PRIGFKLDVR PQQGGEAPR VESESGSIV  
 1141 SGLTPGVEYH YSISVLTDGV EKDIPIKTY VTPLSPPNL RQPSRCSAT LTYWDRSIS  
 1201 PGITGYRIST TPTPMQVGN SLEEVGPSQT YCVFENLSPG VEYMVSVYAV KEEESAPLS  
 1261 QMFLQEIPLQ TDIKYDDVTD TSDLRWTP LNSSNIIGYRI TWAAGESVP IYEEFVGPTD  
 1321 GYKVSGLPE GIDYEISLIT LINGGESAPT TIIQHTAVPP PTNLRFTNIG PENIRVTWSP  
 1381 PPSIELSSYL VRYSPVKKPD DVTELSLSPS TNMVLNLL PTEYLVSYH STRESRESS  
 1441 LNGVAKTHLD SPTGIAFSEI TPNSFTVHWI APRGPITGYR IRIQLESAG RPEERVPPS  
 1501 RNSITLTHLI PGSEYLVSI AINGQOESLP LAGQOATVSD VPTDLEVTSS SFNTLTISWE  
 1561 APAVSVRYR ITYSQTGGHG PEKEFTVPGT SNTATIRGLN PGVSYTITVY AVTGRGDSPA  
 1621 SSKPLTIIHK TDVDQPIDMA VTDIQDHSIH VKWSPPPGPV TGYRVTSVPK SQGETFSQV  
 1681 ISPDQTEVTI VGLQPAVEYV VSIYSQGENG FSEPLVETAV TNIDNPKGLT FTVGVDSIR  
 1741 LAWEVDPGQV TRYRVTYSSP EDGVKELFPA PEGDDTAE LGLRPGTEYT VSIVALHDDM  
 1901 ESKPLIGIV TAIPAPTNLQ FSQVTPSGFS LSWHAPT VHL TGYLVRVMPK ERTSPTKEVR  
 1861 LSPGVAATTV TGLMVATKYE WNYALKDLS TSQPLQLLS TSNVSPFR RPKQDVTETT  
 1921 VTLSWRKTKE TITGFQIDAI PADGQNPIRR TVDADLRTFT ITGLQPGTDY KIYLYTLNDN  
 1981 ARSSPVTVDV TTAVDSPSNL RFLTTSNSL LFTWQPPRAR ITGYIIRYK AGLIKEHLP  
 2041 RLPAGTTEST LTNLEPGTEY IYIIAVRNN MKSEPLVGRK RTDELPLRVT LPHPGQPEI  
 2101 LDVPTDEENT PHITQTKLDN GNGIQLPGSN GQQPSSDEG QLIEEHGFRS PLAPTTAVPV  
 2161 RPKGFAPGRY PHERVDIELD TFPVQHGDFD GPYPHGLGPQ LNDSGVQEVA SHATTISWRPE  
 2221 LETTEYIISC HPIDHEEAPL QFRVPGTSSS ATLNGLTRGA TYNIVVEAQK GTRKHVLEK  
 2281 RVTVGSPPSP EGVLPQVEDT CYDTFSGAHY SVGQEWERMS ESGFRLWCKC LGYSGGHFRC  
 2341 DSSKWCHDNG VNHRIKWKWD RRGENGQMS CTCLNGKGE FKCEPPEATC YDESKMYNVG  
 2401 EQWQKEYLGA ICSCCTYGGG QGWRCDNCR R PAVSPDGTA GQTVSQFTQR YCCNYNLNCP  
 2461 IECYLPLGLQ ADTQHSQQTQ K

//

11. DEFINITION fibronectin.

ACCESSION 224863

SOURCE cattle.

ORGANISM *Bos taurus*

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora;

ORIGIN

1 GVPRSDKVPV PRDLQFVEVT DVKITIMWTP PESPVTGYRV DVIPVNLPE HQRLPVSRN  
 61 TFAEVTGLSP GVTYHFKVFA VNQGRESKPL TAQQATKLD PTNLQFINET DTTIVTWTPT  
 121 PRARIVGYRL TVGLTRGGQP KQYNVGPAAS QYPLRNLQPG SEYAVLSVN KGNQQSPRVT  
 181 GVFTTLQPLG SIPHYNTEVT ETTIVITWTP APRIGFKLVG RPSQGGEAPR EVTSESGSIV  
 241 VSGLTPGVEY VYIISVLRDG QERDAPIVKT VKDDKESVPI SDTIIPAVPP PDLRFTNVG  
 301 PDMRVTWAP PSSIELTNLL VRYSPVKNEE DVAELSSIPS DNAVVLTNLL PGTEYLVSVS



361 SVYEQEHESIP LRGRQKTALD SPSGIDFSDI TANSFTVHWI APRATITGYR IRHHPENMGG  
 421 RPREDRVPPS RNSITLTNLN PGTEYVVSIV ALNSKEESLP LVGQQSTVSD VPRDLEVIAA  
 481 TPTSELLISWD APAVTVRYR ITYGETGGSS PVQEFVPGS KSTATISGLK PGVDYITIVY  
 541 AVTGRGDSPA SSKPVSINR TEIDKPSQMT IPAPTNLKFT QVTPTSLTAQ WTPAPNVQLTG  
 601 YRVRVTPKEK TGPMEINLA PDSSSVVVS LGMVATKYEVV VYALKDRTLTS RPAQGVVTTL  
 661 ENVSPPRRAR VTDATETIT ISWRKTETI TGFQVDAIPA NGQTPIQRTI RPDVRSYIT  
 721 QLPQGTDYKI HLYTLNDNAR SSPVVIDAST AIDAPSNLRF LATPNSLLV SWQPPRARIT  
 781 GYIIKYKPG SPPREVVRP RPQVTEATIT GLEPGTEYTI QVIALKNNQK SEPLIGR

//

## 12. DEFINITION fibronectin.

ACCESSION AAA49707

SOURCE African clawed frog.

ORGANISM *Xenopus laevis*

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;

Pipidae; Xenopodinae; Xenopus.

## ORIGIN

1 MRRGALTGLL LVLCLSVVLR AAPSATSKKR RQAQQQQQQQ VVQPQGTQDN HQKGCYDNGK  
 61 YYQINQQWER TYLGNLTVCT CYGGGRGFNC ESKPESEETC FDKYTGVSYR VGETYERPKD  
 121 NIWWDCTCIG AGRGRISCTI ANRCHEGGQS YKIGDTRWRP HETGGYMLEC VCLGNKGWTC  
 181 TCKPVAERCY DNTAGTSYVV GQWWEKPYQG WMMVDCTCLG EGNRITCSS KNRCNDQDTK  
 241 TSYRIGDTSW KTDTRGNLLQ CICTGNRGE WKCERHSSAQ ATGTGSNPIT NIQTALYQPD  
 301 SQLEPYGHCV TDNGVLYSLG MRWLRTOGSK QMLCTCLGNG VSCEETVATI TFGGNANGEP  
 361 CAIPFTHDQG TYYSCTSEGR QDGLWCATT SNYDSDKYS FCTEQALALV TRGGNSNGAL  
 421 CNFPFLYNNL NYTDSSTSEGR QDSMKWCGTT ANYDADQKFG FCPMAAHEEI CTTNEGVMYR  
 481 VGDQWDKQHD QGHMRCCTCV GNGRGEWSCV AYSQLKDQCI VDGLTYNVNS SFTKLHEEGH  
 541 MNCTCFGQGG RGRWKCDAID QCQDTETROF YQIGDSWEKH LQGVQYQCYC YGKGIGEWHC  
 601 QPLSTSQAGT GPVQVITITES ANFPTSHPIQ WNAPOASHIK NYILRWKPKL KAGPWKQATI  
 661 PGHLNSYTIIS GLKPGILYEG QLISILQYGN REVTFDFETT TTTIHRHSQT ESGETPLPP  
 721 LVSISESVTE ITASSFLVSW VSASDTVSGF RVEYELSEGD DEKRYLELPN TATSVNIPDL  
 781 LPPRRYNNV YQITEEGEKS LILSTTQTTA PDAPPEHNV NVDTSIMIK WNKQPAPITG  
 841 YRVVYSPSVE GSSTELNLP TANSVTLTEL LPGIEYNITI YAVEDSLESV PVFIQQGTG  
 901 TPQTVIVPSP TDLQLEVTD VKIIMWTSQ QSEVSGYRVV VKPVSPAGRD VQNLVNRNT  
 961 FAEVVNLQPG RTYSFEVYAV NRGQSEPLV GEFATKLDAP TDLQFTDVTE STVVIWIPP  
 1021 QAKIGRYLLS VGQTRGGQPS QFPINPSVTN HKLDNLLPGT EYTVSLVALK GNQSSASASG  
 1081 VFTGLEPVS IPHYNTEVTE TTIVVTWTPV PRIGFKLDVR PSQGGAEFR VISESGSIVI  
 1141 SGLTPGVEYT YSISVLTGTV EKDIPITKT VTPLSPTNL RLQPSRDSAT LTVYWDRSIS  
 1201 PGITGYRIST TPTPMQVGNL LEEEVGPSQT YCVFENLSPG VEYNVSVYAV KEEESAPLS  
 1261 QMFLQEIPLQ TDIKYDDVTD TSI DLRWTP LNSSNIIGYRI TVVAAGESVP IYEEFVGPTD  
 1321 GYKVSGLPE GIDYEISLIT LINGGESAPT TIIQHTAVPP PTNLRFTNIG PDNIRVWSP  
 1381 PYSIELSSYL VRYSPVKKPD DVTELSLSPS TNMVLNLL PFTEYLVSVH SVYERESSS  
 1441 LNGVAKTHLD SPTGIAFSEI TPNSFTVHWI APRGPITGYR IRYQLESAG RPKEERVPPS  
 1501 RNSITLTHLI PGSEYLVSI AINGQQESLP LAGQQATVSD VPTDLEVTSS SPNTLTISWE  
 1561 APAVSVRYR ITYSQTGGHG PEKEFTVPGT SNTATIRGLN PGVSYTIVY AVTGRGDSPA  
 1621 SSKPLTIIHK TDVDQPIDMA VTDIQDHSIH VKWSPPPGPV TGYRVTSVPK SQQETFSQV  
 1681 ISPDQTEVTI VGLQPAVEYV VSIYSQGENG ESEPLVETAV TNIDNPKGLT FTDVGVDSIR  
 1741 LAWEVDPGQV TRYRYTYSSP EDGKVELFPA PEGDDDTAEL HGLRPGTEYT VSIVALHDDM  
 1801 ESKPLIGIQ TAIPAPTNLQ FSQVTPSGFS LSWHAPTVEL TGYLVRVNP KKTGPTKEVR  
 1861 LSPGVAATTV TGLMVATKYE VNVYALKDNL TSQPLQGLIS TLDNVSPRR PRIQDVTETT  
 1921 VTLRWRTKTE TITGFQIDAI PADGQNPIRR TVDADLRTFT ITGLQPGTDY KIYLYTLNDN  
 1981 ARSSPVTVDV TTAVDSPSNL RFLTTSNSL LFTWQPPRAR ITGYIIRYK AGGLIKEHLP  
 2041 RLPAGTTEST LTNLEPGTEY IYIIAVRNN MKSEPLVGRK RTDELRLVT LPHPGQGPEI  
 2101 LDVPTDEENT PHITQTKLDN NGIQLPGSN GQQPSSDHEG QLIIEHGFRS PRLPATAVPV  
 2161 RPKGFAPGRY PQERVDIELD TFPVQHGDFD GPYPHGLGPO LNDSGVQEVA SHTTISWRPE  
 2221 LETTEYIISC HPIDHEEAPL QFRVPGTSSS ATLNGLTRGA TYNIVVEAQK GTDKHKVLEK  
 2281 RVTVSGSPGSP EGVLPQVEDT CYDTFSGAHY SVGQEWERMS ESGFRLWCKC LYGSGGHFRC  
 2341 DSSKWCHDNG VNHRIKEDWD RRGENGQMS CTCLGNGKGE FKCEPPEATC YDEGKMYNVG  
 2401 EQWQKEYLGA ICSCTCYGGQ QGWRCDNCRR PGAVSPDGTA GQTVSQFTQR YQNYNLNCP  
 2461 IECYLPGLQ ADTQHSQQTQ K

//

## 13. DEFINITION FIBRONECTIN PRECURSOR (FN).

ACCESSION P04937

SOURCE Norway rat.

ORGANISM *Rattus norvegicus*

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

## ORIGIN

1 MLRGGPGRL LLLAVLCLGT SVRCTETGKS KRQAQQIVQP PSPVAVSQSK PGCDFNGKH  
 61 QINQQWERTY LGNALVCTCY GSRGFNCES KPEPEETCFD KYTGNTYKVG DTYERPKDSM  
 121 IWDCTCIGAG RGRISCTIAN RCHEGGQSYK IGDKWRPHE TGGYMLECLC LGNGKGWTC  
 181 KPFAEKCFDH AAGTSYVVEG TWEKPYQGM MVDCTCLGEG NGRITCTSRN RCNDQDTRTS  
 241 YRIGDTSWKK DNRGNLLQCV CTGNRGEWK CERHVLQSAS AGSGSFTDVR TAIYQPQTHP  
 301 QPAPYGHCVT DSGVVYSVGM QWLKSQGDQ MLCTCLNGV SCQETAVTQT YGGNSNGEPC

361 VLPFHYNERT FYSCTTEGRQ DGHLWCSTTS NYEQDQKYSF CTDHAVLVQT RGGNSNGALC  
421 HEPFLYSNRN YSDCTSEGRR DNMKWCGTTQ NYDADQKFGF CPMAAHEEIC TTNEGVMYRI  
481 GDQWDKQHDH GHMMRCTCVG NGRGQWACIP YSQLRDQCIV DDITYNVNDT FHKRHEEGHM  
541 LNCTCFGQGR GRWKCDPIDR CQDSETRTFY QIGDSWEKEFV HGVRYQCICY GRGIGEWHCQ  
601 PLQTYPGTTG PVQVIITETP SQPNSHPIQW NAPEPSHITK YILRWRPKTS TGRWKEATIP  
661 GHLNSYTIKG LTPGVIYEGQ LISIQQYGHQ EVTRFDFTTS ASTPVTSTNTV TGETAPFSPV  
721 VATSESVTEI TASSEFVSWV SASDTVSGFR VEYELSEEGD EPQYLDLPST ATSVNIPDLL  
781 PGRKYIVNVY QISEEGKQSL ILSTSQTTPAP DAPPDPTVDQ VDDTSIVVRW SRPQAPITGY  
841 RIVYSPFVEG SSTEINLPET ANSVTLSDLQ PGVQYNITII AVEENQESTP VFIQQETTGV  
901 PRSDVVPAPK DLQFVEVTDV KVTIMWTPPN SAVTGYRVVDV LPVNLPGEHG QRLPVNRTTF  
961 AEVTLGLSPGV TYLFKVFVAVH QGRESKPLTA QQTTKLDAPT NLQFVNETDR TVLVWTWPPR  
1021 ARIAGYALTV GLTRGGQPKQ YNVGPMASKY PLRNLQPGSE YTVTLMAVKG NQOSPATGV  
1081 FTTLQPLRSI PPNTEVTEI TIVITWTPAP RIGFKLGVRP SQGGEAPREV TSDSGSIVVS  
1141 GLTPGVEYTY TIQVLRDQGE RDAPIVNRVV TPLSPPTNLH LEANPDTGVL TVSWERSTTP  
1201 DITGYRITTT PTNGQQGTAL EEVHADQSS CTFENRNPLG EYNVSVYTVK DDKESAPISD  
1261 TVIPEVQLT DLSFVDITDS SIGLRWTPLN SSTITIGYRIT VVAAGEGPI FEDFVDSVVG  
1321 YTVTGLEPQ IDYDISVITL INGGESAPTT LTQQTAVPPP DLRFETNIGP DTMRVTWAPP  
1381 PSIELTNLLV RYSPVKNEED VAELSISPSD NAVVLTNLLP GTEYLVSVSS VYEQHESIPL  
1441 RGRQKTGLDS PTGFDDSDVT ANSFTVHWA PRAPITGYII RHAAEHSAGR PRQDRVPPSR  
1501 NSITLNLNPN GTEYIVTIIA VNGREESPL IGQOSTVSDV PRDLEVIAT PTSLLSWEP  
1561 PAVSVRYRYI TYGETGGNSP VQEFVPGSK STATINNIKP GADYITILYA VTGRGDSPAS  
1621 SKPVSYNYQT EIDKPSQMQV TDVQDNSISV RWLPSTSPVT GYRVTTAPKN GLGPTKSQTV  
1681 SPDQTEMTIE GLQPTVEYVV SVYAQNRRGE SQPLVQTAVT NIDRPKGLAF TDVDVDSIKI  
1741 AWESPQGVQS RYRVYSSPE DGIHELFPAP DGDEDTAELH GLRPGSEYTV SVVALHGGME  
1801 SQPLIGVQST AIPAPTNLKF TVQSPTTLTA QWTAPSVKLT GYRVVTPKE KTGPMKEINL  
1861 SPDSTSVIVS GLMVATKYEV SVYALKDILT SRPAQGVVTT LENVSPRRA RVTDATETTI  
1921 TISWRKTET ITGFQVDAIF ANGQTPVQRT ISPDVRSYTI TSLQPGTDYK IHLYTLNDNA  
1981 RSPVNELEAS TAIDAPSNLR FLTTTPNSLL VSWQAPRARI TGYIIKYEK GSPPREVVPR  
2041 PRPGVTEATI TGLEPGTEYT IYVIALKNNQ KSEPLIGRKK TDELPLQVLT PHPNLHGPEI  
2101 LDVPSTVQKT PFVTNPGYDT ENGIQLPGTS HQQPSVQQM IFEHGFRRT TPPTAATPVR  
2161 LRPRYLPNV DEEVQIGHVP RGDVDYHLYP HVPGLNPNAS TGQEALSQTT ISWTPFQESS  
2221 EYIISQCPVG TDEEPLQFQV PGTSTSATLT GLTRGVYNI IVEALHNQR HKVREVVTV  
2281 GNTVNELEAS PTDDSCFDY TVSHYAVGEE WERLSDSGFK LTCQCLGFS GHFRCDSSKW  
2341 CHDNGVNYKI GEKWDRQGEN GQRMSCTCLG NGKGEFKCDP HEATCYDDGK TYHVGEQWQK  
2401 EYLGATCSCT CFGGQRGWRC DNCRRPGAAE PSPDGTGHT YNQYTRQYHQ RTNTNVNCP  
2461 ECFMPLDVQA DRDSSRE

//

## 14. DEFINITION fibronectin - bovine.

ACCESSION FNBO

SOURCE Bos taurus.

ORGANISM Bos taurus

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia;

## ORIGIN

1 QAQQIVQPOS PLTVSOSKPG CYDNGKHYQI NQOWERTYLG SALVCTCYGG SRGENCESKP  
61 EPEETCFDKY TGNTYRVGDT YERPKDSMIW DCTCIGAGRG RISCTIANRC HEGGQSYKIG  
121 DTWRRPHETG GYMLECVCLE NGKGEWTCCK IAECQFDQAA GTSYVVGETW EKPQYQGMV  
181 DCTCLGEGSG RITCTSRNRC NQDTRTSYR IGDTWSKKNR RGNLLQCICG GNGRGEWKCE  
241 RHTSLQTTSA GSGSFTDVRT AIYQPQPHQP POPYGHCVTD SGVYYSVMQK WLKTOGNKQM  
301 LCTCLGNGVS CQETAVTQTY GGNSNGEPCV LPFTYNGKTF YSCTTEGRQD GHLWCSTTSN  
361 YEQDQKYSFC TDHTVLVQTR GGNSNGALCH PFFLYNNHNY TDCTSEGRRD NMKWCSTTQN  
421 YDADQKFGF PMAAHEEICT TNEGVMYRIG DQWDKQHDMG HMMRCTCVGN GRGEWTCVAY  
481 SQLRDQCIV DITYNVNDT FHKRHEEGHML NCTCFGQGR RWKCDPVDQC QDSETRTFYQ  
541 IGDSWEKYLQ GVRYQCICYG RIGGEWACQP LQTYPDTSGP VQVIITETPS QPNSHPIQWS  
601 APESHISKY ILRWKPKNSP DRWKEATIPG HLNSYTIKGL RGVVYEGQL ISVQHYGQRE  
661 VTRFDFTTS TSPAVTSNTV TGETTPLSPV VATSESVTEI TASSEFVSWV SASDTVSGFR  
721 VEYELSEEGD EPQYLDLPST ATSVNIPDLL PGRKYIVNVY EISEEGEQL ILSTSQTTPAP  
781 DAPPDPTVDQ VDDTSIVVRW SRPRAPITGY RIVYSPSVEG SSTEINLPET ANSVTLSDLQ  
841 PGVQYNITII AVEENQESTP VFIQQETTGV PRSDKVPVPPR DLQFVEVTDV KITIMWTPPE  
901 SPVTGYRVVDV IPVNLPGEHG QRLPVSRNTF AEVTLGLSPGV TYHFVAVAVH QGRESKPLTA  
961 QATKLDAPT NLQFINETDT TVIVTWTPPR ARIVGYRLTV GLTRGGQPKQ YNVGPAASQY  
1021 PLRNLQPGSE YAVSLVAVKG NQOSPRTGV FTTLQPLGSI PHYNTVTEI TIVITWTPAP  
1081 RIGFKLGVRP SQGGEAPREV TSESGSIVVS GLTPGVEYVY TISVLRDQGE RDAPIVKKVV  
1141 TPLSPPTNLH LEANPDTGVL TVSWERSTTP DITGYRITTT PTNGQQGYSL EEVHADQSS  
1201 CTFENLSPGL EYNVSVYTVK DDKESVPIST TIIPAVPPPT DLRFETNIGP TMRVTWAPP  
1261 SIELTNLLVR YSPVKNEEDVA AELSISPSDN AVVLTNLLP TEYLVSVSSV VYEQHESIPLR  
1321 GRQKTALDSP SGIDFSDITA NSFTVHWIAP RATITGYRIR HHPENMGGRP REDRVPPSRN  
1381 SITLNLNPNP TEYVVSIVAL NSKEESLPLV GQOSTVSDVP RDLEVIATP TSLLSWDAP  
1441 AVTVRYRIT YGETGGSSPV QEFVTPGSKS TATISGLKPG VDYITIVYAV TGRGDSPASS  
1501 KPVSYNYRTE IDKPSQMQVT DVQDNSISVR WLPSSSPVTV YRVTTAPKN GPGSKTKTVG  
1561 PDQTEMTIEG LQPTVEYVVS VYAQNQNGES QPLVQTAVT IPAPTNLKFQ QVTPTSLTAQ  
1621 WTAPNVQLTG YRVVTPKEK TGPMKEINLA PDSSSVVSG LMVATKYEVS VYALKDILT  
1681 RPAQGVVTTL ENVSPRRAR VTDATETIT ISWRKTETI TGQVDAIPA NGQTPIQRTI

1741 RPDVRSYTII GLQPSTDYKI HNYTLNINAA SSPNVIDAST AIDAPSNLAF LATTENSLIV  
 1801 SWQEPRAAIT EYIINYEKPD SPRENTSAP RPDVTEAITI GLEPSTETI QVIALHNNQK  
 1861 SEPLIGRKKI DELPQLVTLF HPNLHGPEIL DVPSTVQHTP FITNPGYDTG NBIQLPGTSG  
 1921 QQPSLGQQMI FEEHGFRRTT PPTTATEPQH RPRPYPPNVN EEIQIGHVPR GDVCHHLYPH  
 1981 VVGLNPNAST GQEALSQTTI SWTPFQESSE YIISCHPVGI DEEPLQFRVP GTSASATLIG  
 2041 LTAGATYHII YEAVKDQQRQ HVREEVTVG NSVDQGLSQP TDDSCFDPYT VSHYAIGBEW  
 2101 ERLSDSGFKL SQQLGFGSG HERODSPTWC HONGVNYKIG EKWDRQGENG QMMSCTCLGN  
 2161 EKSEPKDDEH EATCYDDGKT YAVGEQWLEH YLEAGSCTD FGGQRGWRDD NARRSJAEPJ  
 2221 NEGETARSHV QNSQRYHQT KQVNDSEED RMLDQADR EGGRE

**This is a sequenced for Myoglobin**  
**Structural class: All Alpha**  
**Sequence 17**

1. DEFINITION Chain A, Atomic Resolution Crystal Structure Aquomet-Myoglobin.  
 ACCESSION 3991450  
 SOURCE sperm whale.  
 ORGANISM *Physeter catodon*  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Cetacea; Odontoceti; Physeteridae; Physeter.

ORIGIN  
 1 VLSEGEWQLV LHVWAKVEAD VAGHGQDILI RLFKSHPETL EKFDREFKHLK TEAEMKASED  
 61 LKKHGVTVLT ALGAILKKGK HHEAELKPLA QSHATKHKIP IKYLEFISEA IHHVLHSRHP  
 121 GDFGADAQGA MNKALELFRK DIAAKYKELG YQG

//  
 2. DEFINITION predicted using GeneFinder; Weak similarity to Shark myoglobin  
 (SW:P02206).

ACCESSION CAB04152  
 SOURCE *Caenorhabditis elegans*.  
 ORGANISM *Caenorhabditis elegans*  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia;

ORIGIN  
 1 MYRQLIEQHI EEVNVQRDAR PQNSLKEPTI IEHKEQKKSS EKPVTIIVEK PAKVAEKPKT  
 61 TSSEKENAVP VDPLARKIID ETSRLSDRQR DVLQKTFAPI LQDCVRNGLK IEFVLFSEYP  
 121 RYKLIWPQFR AIPDSSLNNA VELRRHASVY LNGLGKIIDS MRDEEALGKS MSRIAWAHIK  
 191 WNFQRNHYIV SDRATYFEQB FHVGAKMPAV PCSGLRLPRV ASLSAEMRRG TGRLEPEALM  
 241 PRLEPEALWG CLMLKQTSID NLLKFATFQH MIEPVLEVVK EONGYQLDDE TRQAWTLYQ  
 301 VIADLIEVFR CRALND

//  
 3. DEFINITION GLOBIN (MYOGLOBIN).

ACCESSION P02210  
 SOURCE slug sea hare.  
 ORGANISM *Aplysia limacina*  
 Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia;  
 Anaspidea; Aplysiidae; Aplysia.

ORIGIN  
 1 MSLSAEADL AGKSWAPVFA NKDANGDAFL VALFEKFPDS ANFFADFKGK SVADIFKSPK  
 61 LRDVSSRIFT RLNEFVNNA DAGKMSAMLS QFAKEHVGFG VGSAQFENVR SMFPGFVASV  
 121 AAPFAGADAA WTKLFGLIID ALKAAGK

//  
 4. DEFINITION MYOGLOBIN.

ACCESSION P02189  
 SOURCE pig.  
 ORGANISM *Sus scrofa*  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Artiodactyla; Suliformes; Suina; Suidae; Sus.

ORIGIN  
 1 MGLSDGEWQL VLNWVGKVEA DVAGHGQEV LIRLFKGH PET LEKFDKFKHL KSEDEMGASE  
 61 DLKKGHTVLT TALGGILKKT GHHEAELTPL AQSHATKHKI PVKYLEFISE AIIQVLQSKH  
 121 PGDFGADAQG AMSKALELFR NDMAAKYKEL GFQG

//  
 5. DEFINITION MYOGLOBIN.

ACCESSION P02185  
 SOURCE sperm whale.  
 ORGANISM *Physeter catodon*  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Cetacea; Odontoceti; Physeteridae; Physeter.

ORIGIN  
 1 VLSEGEWQLV LHVWAKVEAD VAGHGQDILI RLFKSHPETL EKFDREFKHLK TEAEMKASED  
 61 LKKHGVTVLT ALGAILKKGK HHEAELKPLA QSHATKHKIP IKYLEFISEA IHHVLHSRHP  
 121 GDFGADAQGA MNKALELFRK DIAAKYKELG YQG

//  
 6. DEFINITION MYOGLOBIN.

ACCESSION P02176  
 SOURCE *Phocoenoides dalli dalli*.  
 ORGANISM *Phocoenoides dalli dalli*  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Cetacea; Odontoceti; Phocoenidae; Phocoenoides.

ORIGIN

1 GLSEGEWQLV LNVWGKVEAD LAGHGQDILI RLFKQHPETL EKFDKFKHLK TEAEMKASED  
61 LKKGHTVLT ALGGILKKKG HHAELKPLA QSHATKHKIP IKYLEFISEA IIVLHLSRHP  
121 AEFGADAQGA MNKALELFRK DIATKYKELG FHG

//

## 7. DEFINITION MYOGLOBIN.

ACCESSION P02178

SOURCE humpback whale.

ORGANISM Megaptera novaeangliae

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Cetacea; Mysticeti; Balaenopteridae; Megaptera.

ORIGIN

1 VLSDAEWQLV LNIWAKVEAD VAGHGQDILI RLFKQHPETL EKFDKFKHLK TEAEMKASED  
61 LKKGHTVLT ALGGILKKKG HHAELKPLA QSHATKHKIP IKYLEFISEA IIVLHLSRHP  
121 AEFGADAQGA MNKALELFRK DIAAKYKELG FQG

//

## 8. DEFINITION MYOGLOBIN.

ACCESSION P02144

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

ORIGIN

1 MGLSDGEWQL VLNWVGKVEA DIPGHGQEVLI IRLFQHPETL LEKFDKFKHLK KSEDEMKASED  
61 LKKGHTVLT ALGGILKKKG HHAELKPLA QSHATKHKIP IKYLEFISEA IIVLHLSRHP  
121 AEFGADAQGA MNKALELFRK DIAAKYKELG FQG

//

## 9. DEFINITION MYOGLOBIN.

ACCESSION P02174

SOURCE long-finned pilot whale.

ORGANISM Globicephala melas

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Cetacea; Odontoceti; Delphinidae; Globicephala.

ORIGIN

1 GLSDGEWQLV LNVWGKVEAD LAGHGQDILI RLFKQHPETL EKFDKFKHLK TEAEMKASED  
61 LKKGHTVLT ALGAILKKKG HHAELKPLA QSHATKHKIP IKYLEFISEA IIVLHLSRHP  
121 AEFGADAQGA MNKALELFRK DIAAKYKELG FHG

//

## 10. DEFINITION MYOGLOBIN.

ACCESSION P02158

SOURCE bat-eared fox.

ORGANISM Otocyon megalotis

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Carnivora; Fissipedia; Canidae; Otocyon.

ORIGIN

1 GLSDGEWQIV LNIWQKVETD LAGHGQEVLI RLFKNHPETL DKFDKFKHLK TEDEMKGSED  
61 LKKGHTVLT ALGGILKKKG HHAELKPLA QSHATKHKIP IKYLEFISEA IIVLHLSRHP  
121 GDFHADTEAA MKKALELFRN DIAAKYKELG FQG

//

## 11. DEFINITION MYOGLOBIN.

ACCESSION P02199

SOURCE emperor penguin.

ORGANISM Aptenodytes forsteri

Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;

Neognathae; Sphenisciformes; Spheniscidae; Aptenodytes.

ORIGIN

1 GLNDQEWQV LTMWGKVESD LAGHGAVLM RLFKSHPETM DRFDKFKHLK TPDEMKGSED  
61 MKKHGTVLT LGQILKKKG HHAELKPLS QSHATKHKIP IKYLEFISEA IIVLHLSRHP  
121 MFGADAQEA MKKALELFRN MASKYKEFG QG

//

## 12. DEFINITION MYOGLOBIN.

ACCESSION P02188

SOURCE common zebra.

ORGANISM Equus burchelli

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Perissodactyla; Equidae; Equus.

ORIGIN

1 GLSDGEWQV LNVWGKVEAD IAGHGQEVLI RLFTGHPETL EKFDKFKHLK TEAEMKASED  
61 LKKGHTVLT ALGGILKKKG HHAELKPLA QSHATKHKIP IKYLEFISEA IIVLHLSRHP  
121 GDFGADAQGA MKKALELFRN DIAAKYKELG FQG

//

## 13. DEFINITION Myoglobin (Horse Heart) Recombinant Wild-Type.

ACCESSION 2914628  
 SOURCE Equus caballus.  
 ORGANISM Equus caballus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Perissodactyla; Equidae; Equus.

## ORIGIN

1 ELSDGSENIQV LNVWGHVEAD IAGHGQSEVLI RLFTGHPETL EKFDKFKHLK TEAEMKASED  
 61 LKKHSTTTLT ALGGILKPKG RHEAELKPLA QSHATKHKIP VKYLEFISDA IIVVLSWHP  
 121 SDFGADAQA MSHALELFR DIAAKYKELG EQG

## 14. DEFINITION Myoglobin - bovine.

ACCESSION MYBG  
 SOURCE Bos taurus.  
 ORGANISM Bos taurus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia

## ORIGIN

1 MGLSDGSEWQL VLNWGHKVEA DVAGHGQSEVL IRLFTGHPET LEKFDKFKHL KTEAEMKASE  
 61 LKKHSTTTLT ALGGILKPKG RHEAELKPLA QSHATKHKIP VKYLEFISDA IIVVLSWHP  
 121 SDFGADAQA MSHALELFR NDMAAQYKVL GFHG

## 15. DEFINITION Myoglobin.

ACCESSION 368949  
 SOURCE cattle.  
 ORGANISM Bos taurus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora;

## ORIGIN

1 MGLSDGSEWQL VLNWGHKVEA DVAGHGQSEVL IRLFTGHPET LEKFDKFKHL KTEAEMKASE  
 61 LKKHSTTTLT ALGGILKPKG RHEAELKPLA QSHATKHKIP VKYLEFISDA IIVVLSWHP  
 121 SDFGADAQA MSHALELFR NDMAAQYKVL GFHG

## 16. DEFINITION Myoglobin - common carp.

ACCESSION MYCA  
 SOURCE common carp.  
 ORGANISM Cyprinus carpio  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 Ostariophysii; Cypriniformes; Cyprinoidea; Cyprinidae;

## ORIGIN

1 HDAELVLEKQV GGVADFEET GGEVLTRELK QHPETQKLEP KFYGIASNEL AGNAAVKANG  
 61 ATVLKKELEL LKARGDHAAL LKPLATTHAN THKIALNNFR LITEVLVRYM ASKAGLDAGG  
 121 QSALRRNMCV VIGDIDTYYN EIGFAG

## 17. DEFINITION MYOGLOBIN.

ACCESSION P00197  
 SOURCE chicken.  
 ORGANISM Gallus gallus  
 Eukaryotae; Metazoa; Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

## ORIGIN

1 GLSDQEWQQV LTIWGHVEAD IAGHGHEVLM RLFDHPETL DREDFKFKGLK TPIQMKGSED  
 61 LKKHGATVLT QIGKILKQKG NHESLKLPLA QTHATKHKIP VKYLEFISEV IIVVIAEKHA  
 121 ADFGADQAA MSHALELFR DMASKYKEFG EQG

This is a sequenced for TBP

Structural class:

Sequence 13 sequence

1. DEFINITION TATA box binding protein.

ACCESSION NP\_003185

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

ORIGIN

1 MDQNSLPPY AQGLASPQGA MTPGIPIFSP MMPYGTGLTP QPIQNTNSLS ILEEQQRQQQ  
61 QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQAVAAA AVQQSTSQQA TQGTSGQAPQ  
121 LFHSQTLTTA PLPGTTPLYP SPMTMPITPATPASESSG IVPQLQNIVS TVNLGCKLDL  
181 KTIALRARN A EYNPKRFAAV IMRIREPRTT ALIFSSGKMV CTGAKSEEQS RLAARKYARV  
241 VQKLGFPAPK LDFKIQNMV SCDVKFPIRL EGLVLTHQQF SSYEPELFPG LIYRMKIPRI  
301 VLLIFVSGKV VLTGAKVREI YEAFENIYP ILKGFRTT

//

2. DEFINITION TBP.

ACCESSION CAA11000

SOURCE domestic silkworm.

ORGANISM Bombyx mori

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Lepidoptera; Bombycoidea; Bombycidae; Bombyx.

ORIGIN

1 MDHMLPSPYN IPGIGTPLHQ PEEDQQILPN AMQQQQQQQQ QSQAQPSLAA LGSSPIVGFQ  
61 AIMGTPQRSM HTYAPTASYA TPQQMMQPQT PQNMSPMIA AGNLSQQML SQASPAPMT  
121 LTPLSADPGI LPQLQNVST VNLDCCKLDL KIALHARNAE YNPKRFAAVI MRIREPRTTA  
181 LIFSSGKMVC TGAKSEEDSR LAARKYARII QKLGFTAKFL DFKIQNMVGS CDVKFPIRLE  
241 GLVLTHGQFS SYEPELFPGL IYRMVKPRIV LLIFVSGKVV LTGAKVREEI YEAFDNIYPI  
301 LKSFKKQ

//

3. DEFINITION TATA-binding protein.

ACCESSION AAC49986

SOURCE Candida albicans.

ORGANISM Candida albicans

Eukaryota; Fungi; Ascomycota; Hemiascomycetes;

ORIGIN

1 MDLKLPTNP TNPQAKTFM KSIEEDEKNK AEDLDIIKKE DIDEPKQEDT TDGNGGGGIG  
61 IVPTLQNIVA TVNLDCRLDL KTIALHARNA EYNPKRFAAV IMRIRDPKTT ALIFASGKMV  
121 VTGAKSEDDS KLASRYARL IQKLGFAKF CDFKIQNIVG STDVKFAIRL EGLAFHGTFL  
181 SSYEPELFPG LIYRMVKPKI VLLIFVSGKI VLTGAKKREE IYDAFESIYP VLNEFRKN

//

4. DEFINITION TATA-binding protein.

ACCESSION AAC35362

SOURCE green urchin.

ORGANISM Lytechinus variegatus

Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;  
Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;  
Lytechinus.

ORIGIN

1 MYNPSQQHQ QIVPVSVHKM QENQDEGQQ RSHYPQISSQ QSQSYLSVPS IGTSPFGSVG  
61 SVTSLAPGSS FIPSPMAPL TPATPASSES SGIVPQLQNI VSTVNLCKL ELKIALHAR  
121 NAEYNPKRFA AVIMRIREPR TTALIFSSGK MVCTGAKSED DSRLAARKYA RVVQKLGFAA  
181 KFLDFKIQNM VGSCDVKFPI RLEGLVLTHG QFSSYEPELF PGLIYRMVKP RIVLLIFVSG  
241 KVVLTGAKVR QEIYDAFNNI YPILKSFKKT T

//

5. DEFINITION TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR) (TATA SEQUENCE-BINDING PROTEIN) (TBP).

ACCESSION O13270

SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;  
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

ORIGIN

1 MDQNSLPPY AQGLASPQGA MTPGIPIFSP MMPYGTGLTP QPVQSTNSLS ILEEQQRQQQ  
61 QQQAAQSSTS QQATQQTSGQ TPQLFHSQTL TTAPLPGTTP LYPSMTPMT PITPATPASE  
121 SSGIVPQLQN IVSTVNLGCK LDLKTIALRA RNAEYNPKRF AAVIMRIREP RTTALIFSSG  
181 KMVCTGAKSE EQSRLAARKY ARVVQKLGFP AKFLDFKIQN MVGSCDVKFP IRLEGLVLTH  
241 QFSSYEPEL FPGLIYRMK PRIVLLIFVS GKVVLTGAKV RAEIYEAFEN IYPILKGFRT  
301 TT

//

6. DEFINITION transcription factor - African clawed frog (fragment).  
 ACCESSION I51648  
 SOURCE African clawed frog.  
 ORGANISM *Xenopus laevis*  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea;  
 Pipidae; Xenopodinae; Xenopus.

## ORIGIN

1 MDQNSIPPF QGLASPOGSL TPGINIFSPL ITYGTGLTPQ PVQTTNSLSI LEEQQRQQQQ  
 61 TQOSTLQQGN QGSGQTPQLF HPQTLTTAPL PGNTPLYPSP ITPMTPISPA TPASESSGIV  
 121 PQLQNIIVSTV NLGCKLDLKT IALRARNAEY NPKRFAAVIM RIREPRTAL IFSSGKMVCT  
 181 GAKSEEQSR L AARKYARVVQ KLGFPKFLD FKIQNMVWSC DVKFPPIRLEG LVLTHQQFSS  
 241 YEPELFPGLI YRMKPRIVL LIFVSGKVVL TGAKVRAEY EAFENIYPIL KGFRTT

//

7. DEFINITION transcription initiation factor IID - fission yeast  
 (*Schizosaccharomyces pombe*).

ACCESSION A35873  
 SOURCE fission yeast.  
 ORGANISM *Schizosaccharomyces pombe*  
 Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;  
 Archaeascomycetes; Schizosaccharomycetales;  
 Schizosaccharomycetaceae; Schizosaccharomyces.

## ORIGIN

1 MDFALPTTAS QASAFMNSS LTFPVLNAN NEATNETADS GDAEVSKNEG VSGIVPTLQN  
 61 IVATVNLDCR LDLKTIALHA RNAEYNPKRF AAVIMRIREP KSTALIFASG KMVVLGGKSE  
 121 DDSKLASKRY ARIIQKLGFN AKFTDFKIQN IVGSCDVKFP IRLEGLAYSH GTFSSYEPEL  
 181 FPGLIYRMVK PKVLLIFVS GKIVLTGAKV REEIYQAFEI IYPVLSEFRK H

//

8. DEFINITION Spt15p: TATA-box binding protein tfiId.

ACCESSION AAB64675  
 SOURCE baker's yeast.  
 ORGANISM *Saccharomyces cerevisiae*  
 Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;  
 Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;  
 Saccharomyces.

## ORIGIN

1 MADEERLKEF KEANKIVFDP NTRQWENQN RDGTPATTF QSEEDIKRAA PESEKDTSAT  
 61 SGIVPTLQNI VATVTLGCR L DLKTVALHAR NAEYNPKREA AVIMRIREPK TTALIFASGK  
 121 MVTGAKSED DSKLASRKYA RIIQKIGFAA KFTDFKIQNI VGSCDVKFP I RLEGLAFSHG  
 181 TFSSYEPELF PGLIYRMVKP KIVLLIFVSG KIVLTGAKQR EEIYQAFEI YPVLSEFRKM  
 241

//

9. DEFINITION 43453.

ACCESSION AAD10645  
 SOURCE thale cress.  
 ORGANISM *Arabidopsis thaliana*  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta;

## ORIGIN

1 MADQGTEGSQ PVDLTKHPSG IVPTLQNIIVS TVNLDCKLDL KAIALQARNA EYNPKRFAAV  
 61 IMRIREPKTT ALIFASGKMV CTGAKSEHLS KLAARKYARI VQKLGFPKPF KDFKIQNIIVG  
 121 SCDVKFP IRL EGLAYSHSAF SSYEPELFPG LIYRMKLPKI VLLIFVSGKI VITGAKMREE  
 181 TYTAFENIYP VLREFRKVQQ

//

10. DEFINITION TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR) (TATA  
 SEQUENCE-BINDING PROTEIN) (TBP).

ACCESSION P26357  
 SOURCE potato.  
 ORGANISM *Solanum tuberosum*  
 Eukaryota; Viridiplantae; Charophyta/Embryophyta group;  
 Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;  
 Magnoliophyta; eudicotyledons; Asteridae; Solananae; Solanales;  
 Solanaceae; Solanum clade.

## ORIGIN

1 MADQGLEGSQ PVDLTKHPSG IVPTLQNIIVS TVNLDCKLDL KAIALQARNA EYNPKRFAAV  
 61 IMRIREPKTT ALIFASGKMV CTGAKSEQQS KLAARKYARI IQKLGFPKPF KDFKIQNIIVG  
 121 SCDVKFP IRL EGLAYAHGAF SSYEPELFPG LIYRMKQPKI VLLIFVSGKI VITGAKVRDE  
 181 TYTAFENIYP VLTEFRKNQQ

//

11. DEFINITION TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR) (TATA  
 SEQUENCE-BINDING PROTEIN) (TBP).

ACCESSION Q27850  
 SOURCE *Tetrahymena thermophila*.  
 ORGANISM *Tetrahymena thermophila*



Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;  
Hymenostomatida; Tetrahymenina; Tetrahymena.

## ORIGIN

1 MSSDKTSQQT EKLPNNSVA QSNSIDQNKV KNNILSTIET MDKSISEDLY PKLQNIIVSTV  
61 NLSTKLDLQ IALRARNAY NPKRFAAVIM RLRDPKTTAL IFASGKMVCT GAKTEEDSNR  
121 AARKYAKIIQ KIGFPVQFKD FKTQNIIVGST DVKFPINLDH LEQDHKKFVQ YEPEIFPGKI  
181 YREFNTKIVL LIEVSGKIVL TGAKTRENIN KAFQKIYWVL YNYQKKDYRG ANLHNQNLNI  
241 KPSIKN

//

12. DEFINITION TATA-BOX BINDING PROTEIN (TATA-BOX FACTOR) (TATA SEQUENCE-BINDING  
PROTEIN) (TBP).

ACCESSION Q55031

SOURCE Sulfolobus shibatae.

ORGANISM Sulfolobus shibatae  
Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.

## ORIGIN

1 ISNSAVSYKP IVNIENIVAT VLEQSLDLY AMERSIPNIE YDPDQFPGLI FRLEQPKVTA  
61 LIFKSGKMV TGAKSTEELI KAVKRIIKTL KKYGIKIMGK PKIQIQNIVA SANLHVNVNL  
121 DKAFFLENN MYEPEQFPGL IFRMDDPRVV LLIFSSGKMV ITGAKREDEV SKAVKRIFDK  
181 LAELDCVKPI EEEEELEL

//

13. DEFINITION TATA-BOX BINDING PROTEIN (TATA-BOX FACTOR) (TATA SEQUENCE-BINDING  
PROTEIN) (TBP).

ACCESSION Q56253

SOURCE Thermococcus celer.

ORGANISM Thermococcus celer  
Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;  
Thermococcus.

## ORIGIN

1 MSNVKLIEN IVASVDLFTQ LNLERVIEMC PHSKYNPEEF PGIICRFDEP KVALIFSSG  
61 KLVVTGAKSV EDIERAVNKL IQMLKKIGAK FSRAPQIDIQ NMVFSGDIGM EFNLDAVALS  
121 LPNCEYEPEQ FPGVIYRVKE PRAVILLFSS GKIVCSGAKS EHDWEAVRK LLRELEKYDL  
181 IGEGEEEN

//

**This is a sequenced for Transmembran**

- **Hatpase**
- **Preseni**
- **Transfer**

**Structural class:**

**Sequence**

1. DEFINITION plasma membrane ATPase.

ACCESSION AAA33561

SOURCE Neurospora crassa.

ORGANISM Neurospora crassa

Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;

Euascomycetes; Pyrenomyces; Sordariales; Sordariaceae;

Neurospora.

ORIGIN

```

1 MADHSASGAP ALSTNIESGK FDEKAAEAAA YQPKPKVEDD EDEDICALIE DLESHDGHDA
61 EEEEEATPG GGRVVPEDML QTDRVGLTS EEVQRRRKY GLNQMKEEKE NHFLKFLGFF
121 VGPIQFVMEG AAVLAAGLED WVDGFGVICGL LLLNAVVG FV QEFQAGSIVD ELKKTALKA
181 VVLRDGLTKE IEAPEVVPGD ILQVEEGTII PADGRIVTDD AFLQVDQSAL TGESLAVDKH
241 KGDQVFASSA VKRGEAFVVI TATGDNTFVG RAAALVNAAS GSGHFTEVL NGIGTILLIL
301 VIFTLIVVWV SSFYRSNPV QILEFTLAI IIGVPGVGLPA VTTTMAVGA AYLAKKKAIV
361 QKLSAIESLA GVEILCSDKT GTLTKNKL LSL HDPYTVAGVD PEDLMLTACL AASRKKKGD
421 AIDKAFLKSL KYYPRAKSVL SKYKVLQFHP FDPVSKKVA VVESPOGERI TCVKGAPLFV
481 LKTVEEDHPI PEEVDQAYKN KVAEFATRGF RSLGVARKRG EGSWEILGIM PCMDPPRHDT
541 YKTVCEAKTL GLSIKMLTGD AVGIARETSR QLGLGTNIYN AERLGLGGGG DMPGSEVYDF
601 VEAADGFAEV FPQHKYNVVE ILQQRGYLVA MTGDGVNDAP SLKKADTGIA VEGSSDAARS
661 AADIVFLAPG LGAIIDALKT SRQIFHRMYA YVYRIALSI HLEIFLGLWI AILNRSNLIE
721 LVVFI AIFAD VATLAIAYDN APYSQTPVKW NLPKLWGM SV LLGVVLAVGT WITVTTMYAQ
781 GENGGIVQNF GNMDEVLFLO ISLTENWLIF ITRANGPFWS SIPSWQLSGA IFLVDILATC
841 FTIWGWFEHS DTSIVAVVRI WIFSGIFCI MGGVYIQLQD SVGFDNLMHG KSPKGNQKQR
901 SLEDFVVS LQ RVSTQHEKSQ

```

//

2. DEFINITION H<sup>+</sup>-transporting ATPase (EC 3.6.1.35) 1, plasma membrane - yeast  
(*Saccharomyces cerevisiae*).

ACCESSION PXBY1P

SOURCE baker's yeast.

ORGANISM *Saccharomyces cerevisiae*

Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;

Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;

*Saccharomyces*.

ORIGIN

```

1 MTDTS SSSSSS SSASSVSAHQ PTQEKPAKTY DDAASESSDD DDIDALIEEL QSNHGVDDED
61 SDNDGPVAA G EARPVPEEYL QTDP SYGLTS DEVLKRRK KY GLNQMADEKE SLVVKFV MFF
121 VGPIQFVME A AAILAAGLSD WVDGFGVICGL LMLNAGVGFV QEFQAGSIVD ELKKTALANTA
181 VVIRDGQLV E IPANEVVPGD ILQLEDGTVI PTDGRIVTED CFLQIDQSAI TGESLAVDKH
241 YGDQTFSSS T VKRGE GFMVV TATGDNTFVG RAAALVNKAA GGQGHFTEVL NGIGIILLVL
301 VIATLLVW T ACFYRTNGIV RILRYTLGIT IIGVPGVGLPA VTTTMAVGA AYLAKKKAIV
361 QKLSAIESL A GVEILCSDKT GTLTKNKL LSL HEPYTVGV S PDDLMLTACL AASRKKKGLD
421 AIDKAFLKSL KQYPKAKDAL TKYKVL E FHP FDPVSKKVA VVESPEGERI VCVKGAPLFV
481 LKTVEEDHPI PEDVHENYEN KVAELASRGF RALGVARKRG EGHWEILGVM PCMDPPRDT
541 AQTVSEARHL GLRVKMLTGD AVGI AKETCR QLGLGTNIYN AERLGLGGGG DMPGSELADF
601 VENADGFAEV FPQHKYRVVE ILQNRGYLVA MTGDGVNDAP SLKKADTGIA VEGATDAARS
661 AADIVFLAPG LSAIIDALKT SRQIFHRMYS YVYRIALSL HLEIFLGLWI AILDNSLDID
721 LIVFIAIFAD VATLAIAYDN APYSPKPVKW NLPRLWGM SI ILGIVLAIGS WITLTTMELP
781 KGGIIQNF GA MNGIMFLQIS LTENWLIFIT RAAGPFWSSI PSWQLAGAVE AVDIIATMFT
841 LFGWWS ENWT DIVTVVRVWI WSGIFCVLG GFYYEMSTSE AFDRLMNGKP MKEKKSTRSV
901 EDEMAAMQRV STQHEKET

```

//

3. DEFINITION H<sup>+</sup>-ATPASE.

ACCESSION AAA35325

SOURCE FISSION YEAST.

ORGANISM SCHIZOSACCHAROMYCES POMBE

EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; FUNGI; ASCOMYCOTA;

ARCHAEASCOMYCETES; SCHIZOSACCHAROMYCETALES;

SCHIZOSACCHAROMYCETACEAE; SCHIZOSACCHAROMYCES.

ORIGIN

```

1 MQRNNGEGR P EGMHRISRFL HGNPFKNAS PQDDSTRTE VYEEGGVEDS AVDYDNASGN
61 AAPRLTAAPN THAQANLQS GNTSITHETQ STSRGQEATT SP SLSASHEK PARPQTGEGS
121 DNEDEDEDID ALIEDLYSQD QEEEQVEEEE SPGPAGA AKV VPELLETDP KYGLTESEVE
181 ERK KKYGLNQ MKEEKTNNIK KFLSFFVGP I QFVME LAAAL AAGLRD WDF GVICALLLN
241 ATVGFVQEYQ AGSIVDELKK TMALKASVLR DGRVKEIEAS EIVPGDILHL DEGTICPADG

```

301 RLITKDCFLO VDQSAITGES LAVDKHQNDT MYSSSTVCRG EAFMVVTATA DSTFVGRAAS  
 361 LVGAAGQSQG HFTEVLNGIG TILLVLVILT LLCIYTAIFY RSVRLAALLE YTLAITIIGV  
 421 PVGLPAVVTT TMAVGAAYLA KKKAIYQKLS AIESLAGVEI LCSDKTGTLT KNRLSLGEPY  
 481 CVEGVSPDDL MLTACCLASSR KKKGLDAIDK AFLKALRNYP KAKDQLSKYK VLDHFHPFDPV  
 541 SKKITAYVEA PDGQRITCVK GAPLWVFKTV QDDHEVPEAI TDAYREQVND MASRGFRSLG  
 601 VARKADGKQW EILGIMPCSD PPRHDTARTI HEAIGLGLRI KMLTGDVAVGI AKETARQLGM  
 661 GTNVYNAERL GLSGGGDMPG SEVNDVEEAA DGFAEVFPQH KYAVVDILQQ RGYLVAMTGD  
 721 GVNDAPSLKK ADAGIAVEGA SDAARSAADI VFLAPGLSAI IDALKTSRQI FHRMYAYVVY  
 781 RIALSLHLEI FLGLWLIIRN QLLNLELIVF IAI FADVATL AIAYDNAPYA MKPVKWNLPK  
 841 LWGLATIVGI LLAIGTWIVN TTMAAQQQNR GIVQNFVQD EVLFLQISLT ENWLIIFITRC  
 901 S3PFWSF3PS WQLSGAVLV DILATLFCIF GWFKGGHQT3 IVAVIRIWMY SFGIFCLIAI  
 961 VYIILSESS3 FDRWMHGKHK ERGTTRKLED FVMQLQRTST HHEAEGKVT3

//

4. DEFINITION H<sup>+</sup>-transporting ATPase (EC 3.6.1.35) AHA10 - Arabidopsis thaliana.

ACCESSION S66367

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryotae; mitochondrial eukaryotes; Viridiplantae;

Charophyta/Embryophyta group; Embryophyta; vascular plants; seed

plants; Magnoliophyta; Magnoliopsida; Capparales; Brassicaceae;

Arabidopsis.

ORIGIN

1 MAEDLDKPLL DPDTFNRKGI DLGILPLEEV FEYLRTSPQG LLSGDAEERL KIFGPNRLEE  
 61 KQENRFVKFL GFMWNPLSWV MEAAALMAIA LANSQSLGPD WEDFTGIVCL LLINATISFF  
 121 EENNAGNAAA ALMARLALKT RVLRDGQWQE QDASILVPGD IISIKLGDII PADARLEGGD  
 181 PLKIDQSVLT GESLPVTKKK GEQVFSGSTC KQGEIEAVVI ATGSTTFFGK TARLV DSTDV  
 241 TGHFQQVLT3 IGNFICISIA VGMVLEIIM FPVQHR3YRI GINLLVLLI GGPIAMPTV  
 301 LSVTLAIGSH RLSQQGAIK RMTAIEEMAG MDVLCCKDTG TLTLSLTVD KNLIEVVDY  
 361 MCKDTILLLA GRASRLNQD AIDAAIVSML ADPREARANI REIHFLPNP VDKRTAITYI  
 421 DSDGKWRAT KGAPEQVNLN CQKNEIAQR VYAIIDRFAE KGLRSLAVAY QEIKE3S3NS  
 481 PGGPWRFCGL LPLFDPPRH DSGE3TLRALS LGVCVKMITG DQLAIK3ETG RRLGMGTNMY  
 541 PSS3LLGH3NN DEHEAIPVDE LIEMADGFAG VPEHKY3EIV KILQEMKHV3Y GMTGDGVNDA  
 601 PALKKADIGI AVADATDAAR SSADIVLTD3P GLSVIISAVL T3RAIFQ3RMR NYTVYAVSIT  
 661 IRILGFTLLA LIWEYDFPPF MVLIIAILND GTIMTISKDR VRPSPTPE3W KLNQIFATGI  
 721 VIGTYLALVT VLFYWIIVST TFFEKHF3VK SIANNSEQV3S SAMYLQV3SII 3QALIFVTR3S  
 781 R3W3SFFER3P3 TLLIFAFILA QLAATLI3AVY ANISFAKITG IGWRWAGVIW LYSLIFYIPL  
 841 DVIKFV3FH3YA LSGEAWN3LV DRKTAFTYK3K DY3KDDG3SN VTISQ3RS3A EELR3S3R3A  
 901 SWIAEQTRRR AEIARLLE3VH SVSRHLESVI K3KQIDQ3MI RAAHTV

//

5. DEFINITION PLASMA MEMBRANE ATPASE (PROTON PUMP).

ACCESSION Q07421

SOURCE Ajellomyces capsulatus.

ORGANISM Ajellomyces capsulatus

Eukaryotae; Fungi; Ascomycota; Euascomycetes; Plectomycetes;

Onygenales; Onygenaceae; Ajellomyces.

ORIGIN

1 MAHSAASGAA SAAHFEK3TP EVAHEEK3PP LPEEDED3ED MDALIEELES QDGHIDIEDD  
 61 EDGEPGGARP VPDELLT3DT RHGLTDAEVV ARK3KYGLNQ MKEEKENLVL KFLSYFVGPI  
 121 QFVMEAAAIL AAGLEDW3DF GVICALLLN ACVGFVQ3EQ AGSIVDELKK TLALKAVVLR  
 181 NGR3L3EVEAP EVVPGDILQV EEGTIIPADG RIVTEEAFLQ VDQSAITGES LAVDKHK3GDT  
 241 CYASSAVKRG EAFMVITATG DNTFVGRGPA LVNAASAGTG HFTEVLNGIG TVLLILVILT  
 301 LLVVV3S3SFY R3NSIVTILE FTLAITIIGV PVGLPAVVTT TMAVGAAYLA KKKAIYQKLS  
 361 AIESLAGVEI LCSDKTGTLT K3K3LSL3EPY CVSGVDPE3DL MLTACLAASR KKKGDAIDK  
 421 AFLKSLR3YP RAKSVLTQYK VLEFHPFDPV SKK3S3AVVLS PQGERITCVK GAPLSVLKTV  
 481 EEDHPI3PDEV DSAYK3KVAE FATRGFRSLG VARKRGE3SW EILGIMPCSD PPRHDTAKTI  
 541 NEAKTLGLSI KMLTGDVAVGI ARETSRQLGL GTNVYNAERL GLGGGGT3MPG SEVYDFVEAA  
 601 DGFAEVFPQH KYNVVEILQQ RGYLVAMTGD GVNDA3PSLKK ADTGIAVEGA SDAARSAADI  
 661 VFLAPGLSAI IDALKTSRQI FHRMYAYVVY RIALSLHLEI FLGLWIAILN T3SLNLQ3LVV3F  
 721 IAI3FADIATL AIAYDNAPFS K3PVKWNLPK LWGMSVLLGI VLAVGTWITL TMLVGS3ENG  
 781 GIVQNFGRTH PVLFL3EISLT ENWLI3FITRA NGPFW3S3IPS WQLSGA3ILLV DIIATLFTIF  
 841 GWFVGGQ3TSI VAVVRIW3VS FGCFCV3LGG3L Y3LLQGS3AGF DNMMHGK3SPK KNQKQ3R3LED  
 901 FV3SLQ3RVST QHEK3SS

//

6. DEFINITION proton motive ATPase H1B.

ACCESSION AAA29228

SOURCE Leishmania donovani.

ORGANISM Leishmania donovani

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Leishmania.

ORIGIN

1 MSSK3KYELDA AAFEDK3PESH SDAEMTPQKP QRRQSVLSKA VSEHDERATG PATDLLPPSK  
 61 GLT3EEAEEL LK3YGRNELP EKKTP3SWLIY VRGLWGPMPA ALWIAIIEF ALENWPDGAI  
 121 LFAIQIANAT IGWYETIKAG DAVAALKNSL KPTATVYRDS KWQQIDA3AVL VPGDLVKLAS

181 GSAVPADCSI NEFVIDVDEA ALTGESLPVT MGPEHMPKMG SNVVRGEVEG TVQYTGSLTF  
 241 FGKTAALLQS VESDLGNIHV ILRRVMLALC AISFILCMCC FIYLLARFYE TFRHALQFAV  
 301 VLVVSIPIA LEIVVTTTLA VGSKHLEKHK IIVTKLSAIE MMSGVNMLOS DKTGTLLTNK  
 361 MEIQEQCFTF EEGNDLKSTL VLAALAAKWR EPPRDALDTM VLGAADLDEC DNYQQLNFPV  
 421 FDPPTKRTAA TLVDRRSGEK FDVTKGAPHV ILQMVYNQDE INDEVVDIID SLAARGVRCL  
 481 SVAKTDQQGR WEMAGILTEL DPPRPDTKDT IRRSKEYGVD VKMITGDHLL IAKEMCRMLD  
 541 LDPNILTADK LPQIKDANDL PEDLGEHTGD MMLSVGGFAQ VFPEHKFMYC ETLRQRGYTC  
 601 AMTGGDGNDA PALKRADVGI AVHGATDAAR AAADMVLTEP GLSVVVEAML VSREVFQRML  
 661 SFLTYRISAT LQLVQFFPIA CFSLTPHRYG SVDPNQFFFH LPVLMFMLIT LLNDGCLMTI  
 721 GYDHWIPSER PQWNLPVVF VSASILAAVA CGSSLMLLWI GLEGYSSQYY ENSWFHRLGL  
 781 AQLPQGHVLT MMLKISISD FLTLFSSRTG GHFFFYVPPS PILECGAIIS LLVSTMAASF  
 841 WHKSRPNTL TEGLAWGQTN AEKLLPLVWV IYCVVWVWVQ DVVKVLAHIC MDAVDLFGCV  
 901 SDASGSGPIK PYSDDMKVNG FEPVKKPAEK STEKALNLSV SSGPHKALEG LREDTHVLNE  
 961 STSPVNAFSP KXXX

//

7. DEFINITION PROBABLE E1-E2 TYPE CATION ATPASE 1B.

ACCESSION P12511

SOURCE Leishmania donovani.

ORGANISM Leishmania donovani

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.

ORIGIN

1 MSSKKYELDA AAFEDKPESH SDAEMTPQKP QRRQSVLSKA VSEHDERATG PATDLLPPSK  
 61 GLTTEEAEL LKGTGRNELP EKKTPSWLIY VRGLWGPMPA ALWIAIIIEF ALENWPDGAI  
 121 LFAIQIANAT IGMETIKAG DAVAALDNL KPTATVYRDS KWQQIDAAVL VPGDLVKLAS  
 181 GSAVPADCSI NEFVIDVDEA ALTGESLPVT MGPEHMPKMG SNVVRGEVEG TVQYTGSLTF  
 241 FGKTAALLQS VESDLGNIHV ILRRVMLALC AISFILCMCC FIYLLARFYE TFRHALQFAV  
 301 VLVVSIPIA LEIVVTTTLA VGSKHLEKHK IIVTKLSAIE MMSGVNMLOS DKTGTLLTNK  
 361 MEIQEQCFTF EEGNDLKSTL VLAALAAKWR EPPRDALDTM VLGAADLDEC DNYQQLNFPV  
 421 FDPPTKRTAA TLVDRRSGEK FDVTKGAPHV ILQMVYNQDE INDEVVDIID SLAARGVRCL  
 481 SVAKTDQQGR WEMAGILTEL DPPRPDTKDT IRRSKEYGVD VKMITGDHLL IAKEMCRMLD  
 541 LDPNILTADK LPQIKDANDL PEDLGEHTGD MMLSVGGFAQ VFPEHKFMYC ETLRQRGYTC  
 601 AMTGGDGNDA PALKRADVGI AVHGATDAAR AAADMVLTEP GLSVVVEAML VSREVFQRML  
 661 SFLTYRISAT LQLVQFFPIA CFSLTPHRYG SVDPNQFFFH LPVLMFMLIT LLNDGCLMTI  
 721 GYDHWIPSER PQWNLPVVF VSASILAAVA CGSSLMLLWI GLEGYSSQYY ENSWFHRLGL  
 781 AQLPQGHVLT MMLKISISD FLTLFSSRTG GHFFFYVPPS PILECGAIIS LLVSTMAASF  
 841 WHKSRPNTL TEGLAWGQTN AEKLLPLVWV IYCVVWVWVQ DVVKVLAHIC MDAVDLFGCV  
 901 SDASGSGPIK PYSDDMKVNG FEPVKKPAEK STEKALNLSV SSGPHKALEG LREDTHVLNE  
 961 STSPVNAFSP KXXX

//

8. DEFINITION proton motive ATPase.

ACCESSION AAB70152

SOURCE Trypanosoma cruzi.

ORGANISM Trypanosoma cruzi

Eukaryotae; mitochondrial eukaryotes; Euglenozoa;

ORIGIN

1 MNQNKDKSVL NNSNNGNFNE QHPQHFQKR QSVLSKAISE HKEDGVDEVP MLPPSKGLTT  
 61 AEAELLAKY GRNELPEKKT PSWLIFVNL WGRMPFALAV AIIIEFALEN WPDGAILLAI  
 121 QLANATIGWY ETIKAGDAVA ALKNSLHPVA TVHRDGAWE LDAALLVPGD LVKLAGSVA  
 181 PADCSINEGV IDVDEAALTG ESLPVTMTD HMPKMGSNVY RGEVGGTVQY TGQNTFFGKT  
 241 AVLLQSVESD LGNIHVILSR VMVVLTSFSE TLGLICFIYL MVKFKESFRR SLQFSVVVLV  
 301 VSIPIALEIV VTTTLAVGSK KLSRHKIIVT KLTAIMMSG VNMLOSCKYT TLTLNKHEIQ  
 361 DQCFTFEKGH DLRSVLVLA LAKWREPPR DALDTMVLGA ADLDECDNYT QTEFVFPDPT  
 421 TKRTAATLVD KRTNEKFSVT KGAPHVILQL VYNQDEINQ VVEIIDS LAA RGVRCILSVAK  
 481 TDSQGRWHLC GILTFDPPR PDTKETIARR KQYGVVVKMI TGDHVLIAKE MCRMLNLDPN  
 541 ILTADKLPKV DVNDLPDDLK EKYGENMLGV GGFQVVFPEH KEMIVEALRQ YGFTCAMTGD  
 601 GVNDAPALKR ADVGIAVQGA TDAARAAAM VLTGPGLSVV VEAMLVSRQV FQCMLSFIYV  
 661 RISATLQLVC FFFIACFSLT PRNYGSACAD FQFFHLPVLM FMLITLLNKG CLMTIGYDRV  
 721 WPSKLPQKWN LPVFTTAAI LAAVACGSSL MLLNIALESW SDEYTPNSWF KALGLAQLKQ  
 781 GKVVTLLYLK ISISDELTLF SSRTGGRWF TMAPGLVILI GAIISLFEVSS MVASFHWFSR  
 841 PDGLLTEGLA WGDTSERLL PLWVWYICIV WWLIQDAVAV GTHKLMEMMD LFGCVSRAYG  
 901 GKVVEQYEDE TPNSGNKNES KEPTV

//

9. DEFINITION H<sup>+</sup>-transporting ATPase (EC 3.6.1.35) (clone HAA13) - golden alga  
(Heterosigma akashiwo).

ACCESSION S53302

SOURCE Heterosigma akashiwo.

ORGANISM Heterosigma akashiwo

Eukaryotae; mitochondrial eukaryotes; stramenopiles;  
Raphidophyceae; Heterosigma.

ORIGIN

1 MQKPLNRKA SDRAMLNKQ SPATGGPYLF DRKRSMSIDD VMEEREKQDS SGGPIKLGDI  
 61 RSEDFRRSFH SGRSHDLIE DPSSAAAAS FEPSTGLTTE EAEILLKQWG KNELIEKTKS

121 KLEIFIEQFT APMPIMIWA ILIEAVLENW PDMYILCGLQ AINGGVGFYE MYKAGNAVAA  
 181 LFASLQPKAI CHRDGQFKNM NATLLVPGDL VILGAGAAMP ADCMINEGQI EVDQAALTGE  
 241 SLPVTMLKGD NPKMGSTVAR GEVEATVTAT GMNTFFGKTA NLIQSVDELG HLQKILLYIM  
 301 AFLIVLSFLL CGITLWYLLD QGEDFKESIS FVVVLLVASI PIAIEVVVTA TMLGSRELA  
 361 KMDAIVARLS AIEELAGMNM LCSDKTGTLT LNKMVIQDDC PMFVDGITPE DVILHAALAA  
 421 KWKEPKDAL DTMVLGACDV SLCNPFQTLD YTPFDPTLKR TEAELKGPDG KTFKVTKGAP  
 481 HIVLDDLCHK KRIEEAVDFK VLELAERGIR SLAVARTNAK GQWFMGLILT FLDPFRPDTK  
 541 LTIERARVHG VEVKMTGSDH QVIAKETARV LDMGTNILGC DGLPTLDAEG KLPSPAEMAD  
 601 ECQRVVDGNG FAQVFPEHKF VIVEAVRMGG FEVGMTGDGV NDAPALKRAD IGIIVQCATD  
 661 AAAAAADIVL TSPGLTVVVE AIVARKIFA RMKSFIVYRV ACTLQLLVFF FVGVLWLHPQ  
 721 DYNSEFPREW GMPVIALIMI TLLNDGTIIS IAYDNVQSSK NPEVWNLPLV YVSTVLGMI  
 781 ACVSSILLHL WALDSTSPTS LFNKFGVELE YAEVMAMVYL KVSLSDFLTL FASRTHGPEW  
 841 VQKPGKLLAA AFLFAVGLST ANSLTWPFGG GMAALPPVAP PPSGHTVWCG SSSRTLARPE  
 901 CTCGCTSTTS LGSTTPRVSS LNTTAENMLM RMRRLEEGMV VSQQMILLVL QVLERMVRRG  
 961 ICSSNLSGSS LLSRCLY

//

## 10. DEFINITION P-ATPase.

ACCESSION AAC27991

SOURCE Emericella nidulans.

ORGANISM

Emericella nidulans  
 Eukaryotae; Fungi; Ascomycota; Euascomycetes; Plectomycetes;  
 Eurotiales; Trichocomaceae; Emericella.

ORIGIN

1 MAERKISYFA DVENGDSRP TDVNDAGLD EYGPLNRYIS TARDNRCGST SSAGALSMMQ  
 61 KKKPWYKFA KAGGENGEEG FVAPEDWLET DLNGLPSSQI EPRRKRGGWN ELTTEKTNFF  
 121 VQFIGYFRGP ILYVMELAVL LAAGLRDWD LGVIIIGLML NAVVGWYQEK QAGNVVASLK  
 181 GDIAMKAVVK RDGQEQEILA RELVTGDIVV IEEGTIVPAD VRLICDYDKP ETYETYKEYL  
 241 ATANDDTLKE NDDDDDDHGI EARLGVSLVA VDQSAITGES LAVDKYMATD CYYTTGCKRG  
 301 KAYAVTATA KHSEVKGTA LVOGAQDQGH FKAVMDNIGT SLLVLVMEWI LAAWIGGFYR  
 361 HLKIATPEHS DNTLLHWTLI LLIIGVPVGL PVVTTTTLAV GAAYLAEQKA IVQKLTAIES  
 421 LAGVDILCSD KTGTLTANQL SIREPYVNEG VDVNMMMAVA AIASNNHVKN LDPIDKVTIL  
 481 TLRRYPKARE ILARNWVTEK YTPFDPVSKR ITTICTDGV RYPCAKGAP AILAMSECS  
 541 EEAQKREKA SEFARRGERS LGVAVQKEGE PWQLLGMYPM FDPREDTAH TIAEAQHLGL  
 601 SVKMLTGDAL AIKETCHML ALSTKVYDSE RLIHGLLAGS AQHDLVEKAD GFAEVFPEHK  
 661 YQVVEMLQCC GHLTAMTGDG VNDAPSLKKA DCGIAVEGST EAAQAAADIV FLAPGLSTIV  
 721 DAIKLARQIF QRMKAYIQYR IALCIHLELY LVTSMIINE TIKADLIVFI ALFADLATIA  
 781 VAYDNAHFEA RPVEWQLPKI WVIVSVLVGL LAAGTWIMRA SLFLENGGII QNFGSPQML  
 841 FLEVSLTENW LIFVTRGGKT WPSWQLVGAJ FVVDVLTATF CVFGWLAGDY VETSPPSQAT  
 901 FSTNNDTIDV TVVVIWAYS I GVTIIIAVY YLLTIIPALD NLGRKNRSVV DTKVENLLNH  
 961 LSKLAIEHEV DANGKSRYTL GARAPEPDE

//

11. DEFINITION H<sup>+</sup>-transporting ATPase (EC 3.6.1.35) 2, plasma membrane - yeast  
(*Saccharomyces cerevisiae*).

ACCESSION PXBY2P

SOURCE baker's yeast.

ORGANISM *Saccharomyces cerevisiae*

Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;  
 Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;  
 Saccharomyces.

ORIGIN

1 MSSTEAKQYK EKPSKEYLHA SDGDDPANNS AASSSSSSST STSASSSAAA VPRKAAAASA  
 61 ADDSDSDEDI DQLIDELQSN YGEGDESSEE EVRTDGVHAG QRVVPEKDL S TDPAYGLTSD  
 121 EVARRRKYK LNQMAEENES LIVKFLMFFV GPIQFVMEAA AILAAGLSDW VDVGVICALL  
 181 LLNASVGFQI EFQAGSIVDE LKKTLANAT VIRGQLIEI PANEVVPGEI LQLESGTIAP  
 241 ADGRIVTEDC FLQIDQSAIT GESLAAEKHY GDEVSSSTV KTGEAFMVVT ATGDNTFVGR  
 301 AAALVQASG VEGHFTEVLN GIGIILLVLV IATLLLVWTA IATLLLVWTA IRYTLGITI  
 361 IGVVGLPAV VTTMVAAGAA YLAKKQAIQV KLSAIESLAG VEILCSDKTG TLTKNKLHLH  
 421 EPTYVEGVSF DDLMLTACLA ASRKKKGLDA IDKAFKSLI EYPKAKDAL KYKVFHFHF  
 481 DPVSKKVTAV VESPEGERIV CVKGAPLFLV KTVEEDHPIP EDVHENYENK VAEASRGFR  
 541 ALGVARKRGE GHWEILGVMP CHDPPRDDTA QTINEARNLG LRIKMLTGA VGIKETCRQ  
 601 LGLGTNIYNA ERLGLGGGD MPGSELADFV ENADGFAEV PQHKYRVVEI LQNRGYLVAM  
 661 TGDGVNDAPS LKKADTGIAV EGATDAARSA ADIVFLAPGL SAIDALKTS RQIFHRMYSY  
 721 VVYRIALSLH LEIFLGLWIA ILNNSLDINL IVFIAIFADV ATLTIAIDNA PYAPEPVKWN  
 781 LPLRWGMSII LGIVLAIGSW ITLTTMFLPN GGIIQNFQAM NGVMFLQISL TENWLVFVTR  
 841 AAGPFWSSIP SWQLAGAVFA VDIIATMFTL FGWWSNWTD IVSVVRVWIV SIGIFCVLGG  
 901 FYYIMSTSQA FDRLMNGKSL KEKKSSTRSVE DFMAAMQRVS TQHEKSS

//

## 12. DEFINITION PLASMA MEMBRANE ATPASE 2 (PROTON PUMP).

ACCESSION P19657

SOURCE baker's yeast.

ORGANISM *Saccharomyces cerevisiae*

Eukaryotae; Fungi; Ascomycota; Hemiascomycetes;

ORIGIN

1 MSSTEARQYK EKPSKEYLHA SDGDDPANN S AASSSSSSST STSASSSAAA VPRKAAAASA  
 61 ADDSDSEEDI DQLIDELQSN YGEGDESSEE EVRTDGVHAG QRVVPEKDL TDPAYGLTSD  
 121 EVARRRKYG LNQMAEENES LIVKFLMEFV GPIQFVMEAA AILAAGLSDW VDVGVICALL  
 181 LLNASVGFQ EFQAGSIVDE LKKTLANAT VIRDGQLIEI PANEVVPGEI LQLESGTIAP  
 241 ADGRIVTEOC FLQIDQSAIT GESLAAEKHY GDEVFSSSTV KTGEAFMVVT ATGDNTFVGR  
 301 AAALVGCASG VEGHFTEVLN GIGIILLVLV IATLLLVWTA CFYRTVGIVS ILRYTLGITI  
 361 IGVVGLPRAV VTTTMAVGAA YLAKKQAIQV KLSAIESLAG VEILCSDKTG TLTKNKLSLH  
 421 EPYTVEGWS P DDMLTACL A ASRKKKGLDA IDKAFKLSLI EYPKAKDAL KYKVFLEFHPF  
 481 DPVSKHTTAV VESPEGERIV CVKGAPLFLV KTVEEDHPIP EDVHENYENK VAELASRGER  
 541 ALGVARYRGE GHWEILGVMP CMDPPRDDTA QTINEARNLG LRIKMLTGDA VGIKETCRQ  
 601 LGLGTNIYNA ERLGLGGGD MPGSELADFV ENADGFAEVF PQHKYRVVEI LQNRGYLVAM  
 661 TGDGVNDCAPS LKKADTGIAV EGATDAARSA ADIVFLAPGL SAIDALKTS RQIFHRMYSY  
 721 VVYRIALSLH LEIFLGLWIA ILNNSLDINL IVFIAIFADV ATLTIAYDNA PYAPEPVKWN  
 781 LPRLWQMSII LGIVLAIGSW ITLTTMFLPN GGIIQNFAM NGVMFLQISL TENWLIETVR  
 841 AAGPWSSIP SWQLAGAVFA VDIATMFTL FGWSENWTD IVSVRVVWIW SIGIFCVLGG  
 901 FYYIMSTEQA FDRLMNGKSL KEKKSTRSVE DFMAAMQRVS TQHEKSS

//

13. DEFINITION plasma membrane H(+)-ATPase.

ACCESSION AAD11605

SOURCE Emericella nidulans.

ORGANISM Emericella nidulans

Eukaryota; Fungi; Ascomycota; Euscomycetes; Plectomycetes;  
 Eurotiales; Trichocomaceae; Emericella.

ORIGIN

1 MAERKIEYFA DVENGDSRSP TDVNDASGLD EYGALNRYIS TARDNRRGST SSAGALSMKQ  
 61 KKKPWYKFWA KAGGENGEEG FVAPEDWLET DLNGLPSSQI EPRRKRGGWN ELTTEKTNFF  
 121 VQFIGYFRSP ILYVMELAVL LAAGLRDWID LGVIIGILML NAVVGWYQEK QAADVVASLK  
 181 GDIAMKAVK RDGQEQEILA RELVTGDIVV IEEGTIVPAD VRLICDYDKP ETYETYKEYL  
 241 ATANDDTLKE NDDDDDDHGI EARLGVSLVA VDQSAITGES LAVDKYMAT CYTTGCKRG  
 301 KAYAIVTATA KHSEVGTAA LVQGAQDQGH FKAVMDNIGT SLLVLVMEWI LAAWIGGFYR  
 361 HLKIATPEHS DNTLLHWTLI LLIIGVPVGL PVVTTTTLAV GAAYLAEQKA IVQKLTAIES  
 421 LAGVDILCSD KTGTLTANQL SIREPYVNEG VDVNWMMAVA AIASNHNKVN LDPIDKVTIL  
 481 TLRRYPKARE ILARNWVTEK YTPFDPVSKR ITTICTDGV RYTCAGKAPK AILAMSECS P  
 541 EEAQKREKA SEFARRGFRS LGVAVQKEGE PWQLLGMYPM FDPREDTAH TIAEAQHLGL  
 601 SVKMLTGDAI AIAKETCKML ALSTKVYDSE RLIHGGLAGS AQHDLVEKAD GFAEVFPEHK  
 661 YQVVEMLQOR GHLTAMTGDG VNDAPSLKKA DCGIAVEGST EAAQAAADIV FLAPGLSTIV  
 721 DAIKLARQIF QRMKAYIQYR IALCIHLELY LVTSMIINE TIKADLIVFI ALFADLATIA  
 781 VAYDNAHFEA RPVEWQLPKI WVISVVLGVL LAAGTWIMRA SLFLENGGII QNFGSPQML  
 841 FLEVSLTENW LIFVTRGGKT WPSWQLVGAI FVVDVLATLF CVFGWLAGDY VETSPPSQAT  
 901 FSTNNDTIV TVVVIWAYS I GVTIIIIVVY YLLTIIIPALD NLGRKNRSVV DTKVENLLNH  
 961 LSKLAIHEV DANGKSRYL GARAEPEDDE

//

**Triose phosphate Isomerase**  
**Structural class: Alpha/Beta**  
**Similar to Chitinase**  
**14 sequence**

1. DEFINITION triosephosphate isomerase 1.

ACCESSION NP\_000356  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

ORIGIN

1 MAPSRKFFVG GNWFMNGRKK SLGELIGTLN AAKVPADTEV VCAPPTAYID FARQKLDPKI  
 61 AVAAQNCYKV TNGAFTGEIS PGMIKDCGAT WVLGHSERR HVGESDELI GQKVAHALAE  
 121 GLGVIACIGE KLDEREAGIT EKVVFEQTKV IADNVKDWK VVLAYEPVWA IGTGKTATPQ  
 181 QAQEVHEKLR GWLKSINVSDA VAQSTRIIYG GSVTGATCKE LASQPDVDGF LVGGASLKPE  
 241 FVDIINAKQ

//

2. DEFINITION TRIOSEPHOSPHATE ISOMERASE (TIM).

ACCESSION P48500  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

ORIGIN

1 MAPSRKFFVG GNWFMNGRKK SLGELICTLN AAKLPADTEV VCAPPTAYID FARQKLDPKI  
 61 AVAAQNCYKV TNGAFTGEIS PGMIKDLGAT WVLGHSERR HVGESDELI GQKVNHALSE  
 121 GLEVIACIGE KLDEREAGIT EKVVFEQTKA IADNVKDWCK VVLAYEPVWA IGTGKTATPQ  
 181 QAQEVHEKLR GWLKSINVSEB VAQSTRIIYG GSVTGATCKE LASQPDVDGF LVGGASLKPE  
 241 FVDIINAKQ

//

3. isomerase, triosephosphate.

ACCESSION 223374  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

ORIGIN

1 APSRKFFVGG GNWFMNGRKK SLGELITLNA AAKVPADTEV VCAPPTAYID FARQKLDPKIA  
 61 VAAQNCYKVT NGAFTGEISP GMIKDCGATW WVLGHSERRH VVGESDELIG QKVAHALSEG  
 121 LGVIACIGEK LDEREAGITE KVVFEQTKVI ADNVKDWKSV VLAYEPVWAI GTGKTATPQQ  
 181 AQEVHEKLRG WLKSINVSDAV AQSTRIIYGG SVTGATCKEL ASQPDVDGFL VGGASLKPEF  
 241 VDIINAKQ

//

4. DEFINITION triose-phosphate isomerase

ACCESSION AAA35438  
 SOURCE chimpanzee.  
 ORGANISM Pan troglodytes  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Pan.

ORIGIN

1 MAPSRKFFVG GNWFMNGRKK SLGELIGTLN AAKVPADTEV VCAPPTAYID FARQKLDPKI  
 61 AVAAQNCYKV TNGAFTGEIS PGMIKDCGAT WVLGHSERR HVGESDELI GQKVAHALAE  
 121 GLGVIACIGE KLDEREAGIT EKVVFEQTKV IADNVKDWK VVLAYEPVWA IGTGKTATPQ  
 181 QAQEVHEKLR GWLKSINVSDA VAQSTRIIYG GSVTGATCKE LASQPDVDGF LVGGASLKPE  
 241 FVDIINAKQ

//

5. DEFINITION TRIOSEPHOSPHATE ISOMERASE (TIM).

ACCESSION P00939  
 SOURCE Oryctolagus cuniculus.  
 ORGANISM Oryctolagus cuniculus  
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Lagomorpha; Leporidae; Oryctolagus.

ORIGIN

1 APSRKFFVGG NWKMNGRKKN LGELITLNA AKVPADTEVV CAPPTAYIDF ARQKLDPKIA  
 61 VAAQNCYKVT NGAFTGEISP GMIKDCGATW VVLGHSERRH VFGESDELIG QKVAHALSEG  
 121 LGVIACIGEK LDEREAGITE KVVFEQTKVI ADNVDWSKV VLAYEPVWAI GTGKTATPQQ  
 181 AQEVHEKLRG WLKSNVSDAV AQSTRIIYGG SVTGATCKEL ASQPDVDGFL VGGASLKPEF  
 241 VDIINAKQ

//

## 6. DEFINITION TPI.

ACCESSION AAC36016  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Rodentia; Sciurograthi; Muridae; Murinae; Mus.

## ORIGIN

1 MAPTRKFFVG GNWKMNGRKK CLGELICTLN AANVPAGTEV VCAPPTAYID FARQKLDPKI  
 61 AVAAQNCYKV TNGAFTGEIS PGMIKDLGAT WVVLGHSERR HVFGESDELI GQKVAHALAE  
 121 GLGVIACIGE KLDEREAGIT EKVVFEQTKV IADNVKDWSK VVLAYEPVWA IGTGKTATPQ  
 181 QAQEVHEKLR GWLKSNVNDG VAQSTRIIYG GSVTGATCKE LASQPDVDGF LVGGASLKPE  
 241 FVDIINAKQ

//

## 7. DEFINITION triosephosphate isomerase.

ACCESSION CAA30823  
 SOURCE rhesus monkey.  
 ORGANISM Macaca mulatta  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.

## ORIGIN

1 MAPSRKFFVG GNWKMNGRKQ NLGELIGTLN AAKVPADTEV VCAPPTAYID FARQKLDPKI  
 61 AVAAQNCYKV TNGAFTGEIS PGMIKDCGAT WVVLGHSERR HVFGESDELI GQKVAHALAE  
 121 GLGVIACIGE KLDEREAGIT EKVVFEQTKV IADNVKDWSK VVLAYEPVWA IGTGKTATPQ  
 181 QAQEVHEKLR GWLKSNVSEA VAQSTRIIYG GSVTGATCKE LASQPDVDGF LVGGASLKPE  
 241 FVDIINAKQ

//

## 8. DEFINITION TIM.

ACCESSION AAA49094  
 SOURCE chicken.  
 ORGANISM Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Archosauria; Aves; Neognathae; Galliformes;  
 Phasianidae; Phasianinae; Gallus.

## ORIGIN

1 MAPRKFFVGG NWKMNKDCKS LGELIHTLNG AKLSADTEVV CGAPSIYLD F ARQKLDKIG  
 61 VAAQNCYKVP KGFTGEISP AMIKDIGAAW VILGHSERRH VFGESDELIG QKVAHALAEG  
 121 LGVIACIGEK LDEREAGITE KVVFEQTKAI ADNVDWSKV VLAYEPVWAI GTGKTATPQQ  
 181 AQEVHEKLRG WLKSHVSDAV AQSTRIIYGG SVTGGNCKEL ASQHDVDGFL VGGASLKPEF  
 241 VDIINAKH

//

## 9. DEFINITION triosephosphate isomerase.

ACCESSION AAC37247  
 SOURCE Anopheles sp.  
 ORGANISM Anopheles sp.  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;  
 Tracheata; Insecta; Pterygota; Diptera; Nematocera; Culicoidea;  
 Culicidae; Anopheles.

## ORIGIN

1 VGGNWKMGND KASITELCKT LSAGPLDPNT EVVVGCPAPY LSLARSLPE TIGVAAQNCY  
 61 KVAKGFTGE ISPAMLKDLG LGWVVLGHSE

//

## 10. DEFINITION triosephosphate isomerase.

ACCESSION AAC37246  
 SOURCE Spodoptera littoralis.  
 ORGANISM Spodoptera littoralis  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;  
 Tracheata; Insecta; Pterygota; Lepidoptera; Noctuoidea;

## ORIGIN

1 VGGNWKMGND KKQVTEIVET LKKGPLDSNV EVVVGVPAYI LEYVQSIVPN TINVAAQNCW  
 61 KSPKGFTGE ISPAMIKDIG ANWVVLGHSE R

//

## 11. DEFINITION TRIOSEPHOSPHATE ISOMERASE (TIM).

ACCESSION P50921  
 SOURCE Vibrio marinus.  
 ORGANISM Vibrio marinus



Eubacteria; Proteobacteria; gamma subdivision; Vibrionaceae;  
Vibrio.

## ORIGIN

1 MRHPVVMGNW KLNGSKEMVV DLLNGLNAEL EGVTVGDVAV APPALFVDLA ERTLTEAGSA  
61 IILGAQNTDL NNSGAFTGDM SPAMLKEFGA THIIIGHSER REYHAESDEF VAKKFAFLKE  
121 NGLTPVLCIG ESDAQNEAGE TMAVCAQOLD AVINTQGVEA LEGAIIAYEP IWAIGTGKAA  
181 TAEDAQRHA QIRAHIAEKS EAVAKNVIQ YGGSVKPENA AAYFAQPDID GALVGGALD  
241 AKSFAAIAKA AA EAKA

//

12. DEFINITION TRIOSEPHOSPHATE ISOMERASE (TIM).

ACCESSION P48499

SOURCE Leishmania mexicana.

ORGANISM Leishmania mexicana

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.

## ORIGIN

1 MSAKPQPIAA ANWKCNGTTA SIEKLVQVFN EHTISHDVQC VVAPTFVHIP LVQAKLRNPK  
61 YVISAENAI A KSGAFTGEVS MPILKDIGVH WVILGHSERR TYYGETDEIV AQKVSEACKQ  
121 GFMVIACIGE TLQOREANQT AKVLSQTS A IAAKLTDAW NOVVLAYEPV WAIGTGKVAT  
181 PEQAQEVHLL LRKWSSENI TDVAAKLRL YGGSVNAANA ATLYAKPDIN GLVGGASLK  
241 PEFRDIIDAT R

//

13. DEFINITION TRIOSEPHOSPHATE ISOMERASE (TIM).

ACCESSION P04790

SOURCE Escherichia coli.

ORGANISM Escherichia coli

Eubacteria; Proteobacteria; gamma subdivision;

## ORIGIN

1 MRHPLVMGNW KLNGSRHMVH ELVSNLRKEL AGVAGCAVAI APPEMYIDMA KREAEGSHIM  
61 LGAQNVLDNL SGAFTGETSA AMLKDIGAQY IIGHSERRT YHKESDELIA KKFVAVLKEQG  
121 LTPVLCIGET EAENEAGKTE EVCARQIDAV LKTQGA AAFE GAVIAYEPVW AIGTGKSATP  
181 AQAQAVHKFI RCHIAKVDAN IAEQVVIQYG GSVNASNAAE LFAQPDIDGA LVGGASLKAD  
241 AFAVIVKAAE AAKQA

//

14. DEFINITION triose phosphate isomerase.

ACCESSION AAC39065

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;  
Drosophilidae; Drosophila.

## ORIGIN

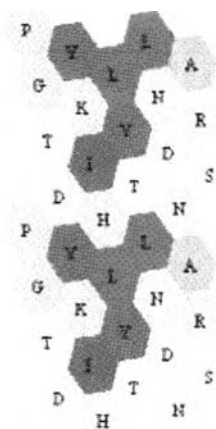
1 MSRKFCVGGN WRMNGDQKSI AEIAKTLSSA ALDPNTEVVI GCPAIYLYA RNLLPCELGL  
61 AGQNAVYKVAK GAFTGEISPA MLKDIGADWV ILGHSERRAI FGESDALIAE KAEHALAEG  
121 KVIACIGETL EEREAGKTNE VVARQMCAYA QKIKDWKVV VAYEPVWAIG TGKTATPDQA  
181 QEVHAFRLQW LSDNISKEVS ASLRIQYGG S VTAANAKELA KKPDIIDGFLV GGASLKPEFV  
241 DIINARQ

//

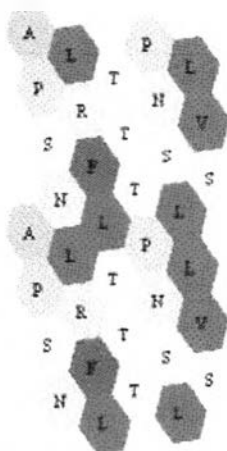
ภาคผนวก ข

รูปแบบ (Pattern) ที่ใช้ในการ Train โครงข่ายประสาทเทียม (บางส่วน)

12 PGTDYKI HLVTLENRAS



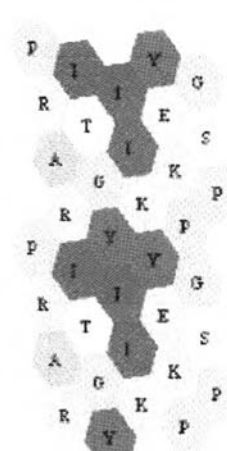
13 APSNLRFLTITDILLVS



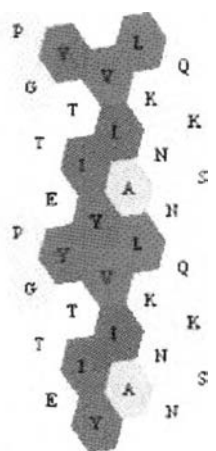
14 PMSLLWSMQAPRARI TGY



15 FRAKI TGVII KYKPKGEP



16 PGTETIYVI ALKNQFS



17 KIDELPQLVTLPHENLHG



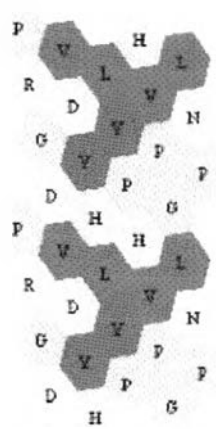
18 HGPFI LDWFS TVQRTPFV



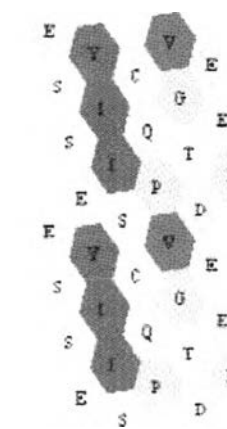
19 QQPS VGGQM FEEHGFRR



20 PRGVDVHLYPHWGLNS



21 ESSEVI SCOPWIDEED



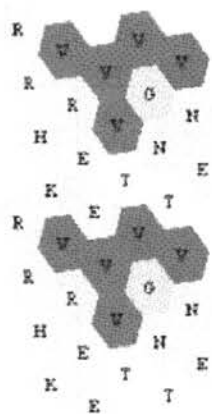
22 TS ALLTGLTRGVTYNI IV



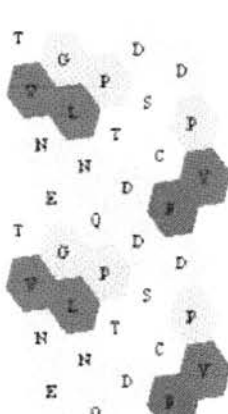
23 TYN IVEALHQQ&H&NK



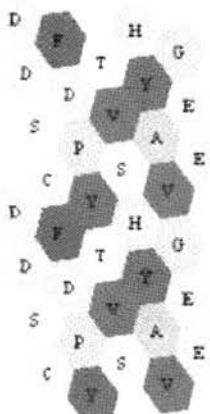
24 RRRHVREEVVTVORTVRE



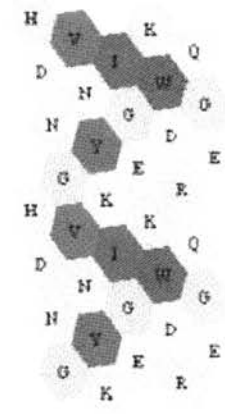
25 TVNEGLNQPTDSDCFDPV



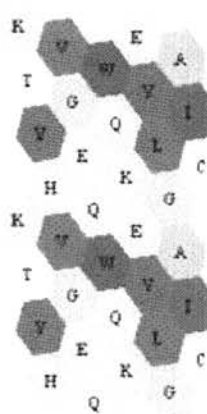
26 DSDCFDPVTVSHVAAGEE



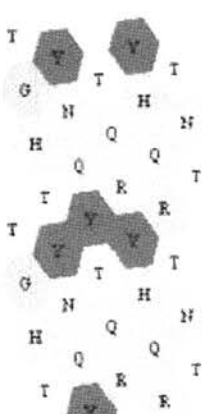
27 HGNQVYKI QEVDROGE



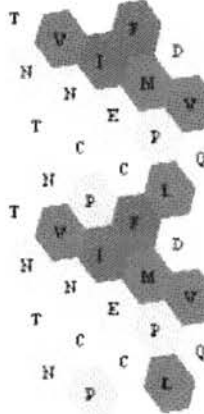
28 KTYHVGEOQVQKENVLGAIC



29 TGHYNOVYQRYVHQRTHT



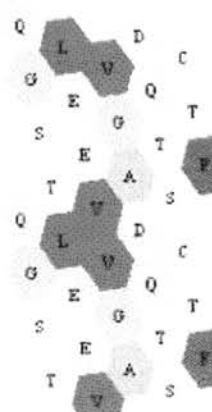
30 THTNWNQPI ECFNPLDQO



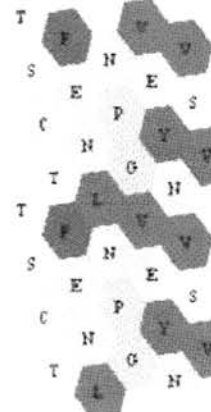
31 PVTGR LIMS WDRSTTPR



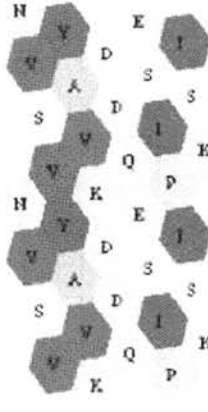
32 QOSTLEEVVAGADQTS CTF



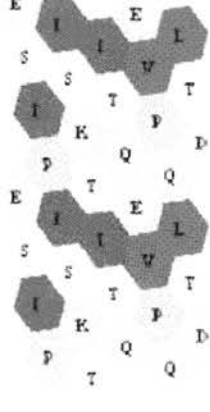
33 TSCTFENLNDGVEVNSV



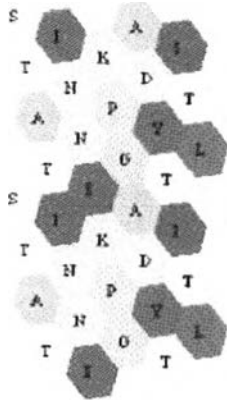
34 NUSVVAWDDQESI PLSK



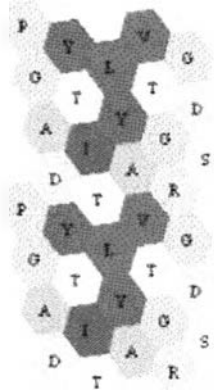
35 ESIPLSKITQEMVQLTD



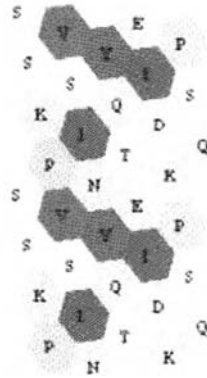
0 STATI NŌE KPOGALYTI TL



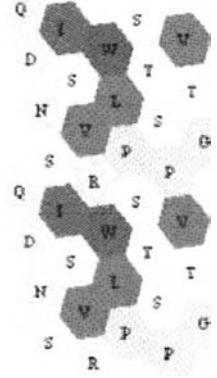
1 PGADYTI TLVANTORGES



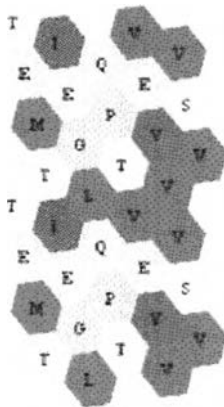
2 SSEPUSI NYOTEI DAPSO



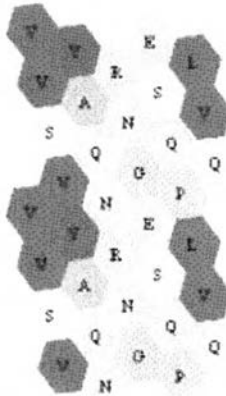
3 QONISURVLPSTSPWTO



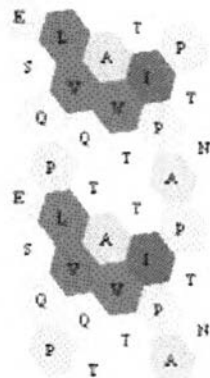
4 IENTI EQL OPTIENQNS V



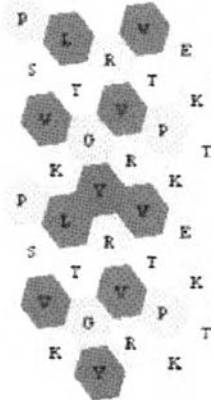
5 VUSVYAKNRNGES OPLVO



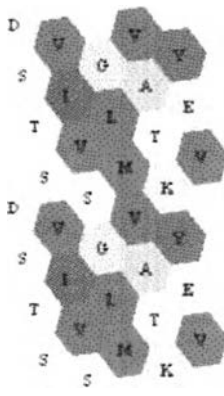
6 ES OPLVOTAVITIPARTN



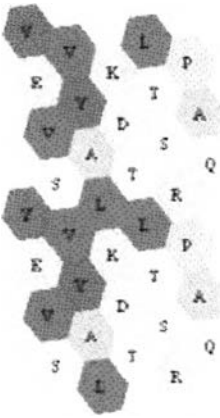
7 DSUULTOVKRVKPTMNT



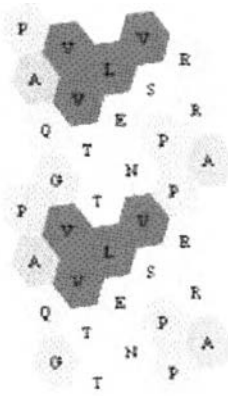
8 IGTSMEVS OLIVATKVEV



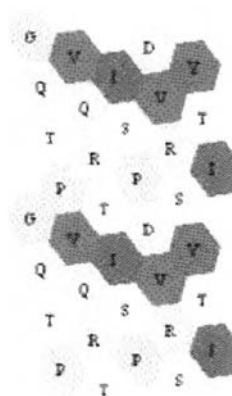
9 YENS VYALKDILTISPPAQ



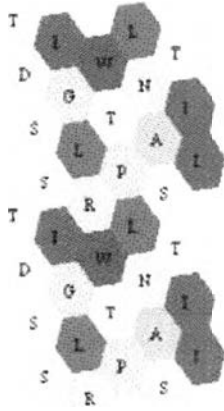
10 PAGOVTILENUS PFERA



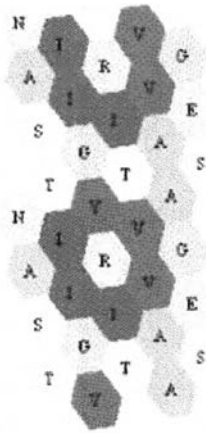
11 QOTPVORTI SPENUSYTI



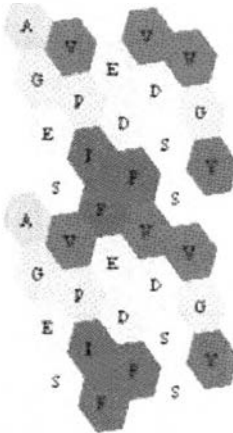
36 TINSI GLRPTFLNASTII



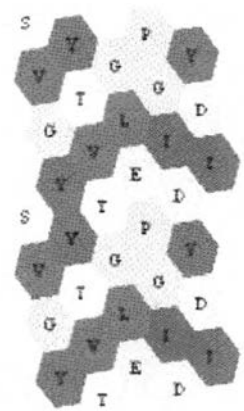
37 NASTII GYRI TVVAAGES



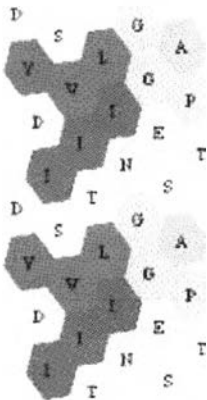
38 AGES VPI FEEFVDS SUGV



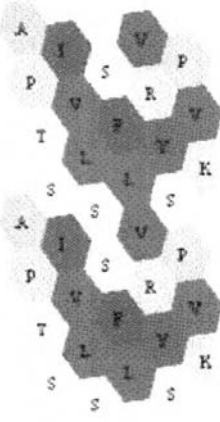
39 SWSYTYTGLEP G DYN



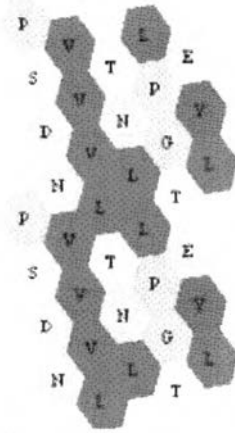
40 DVIISMTII NNGESAPT



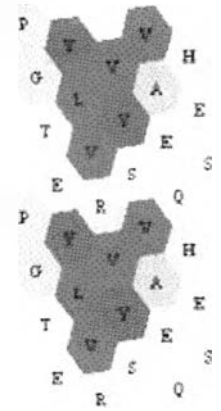
41 APTSI VLSSFLVRYSPVK



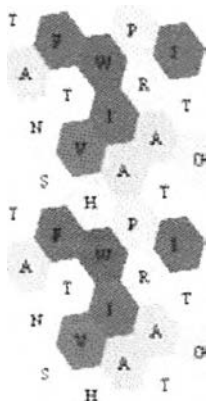
42 PGINWVLTNLLDQTEYL



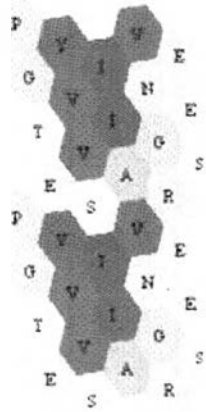
43 PGTETLVRYVSVAQKHS



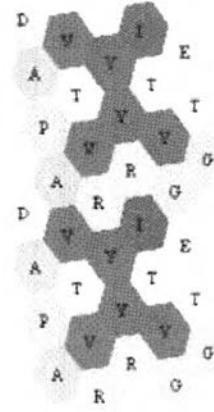
44 TANFTVHAI AFRATIIG



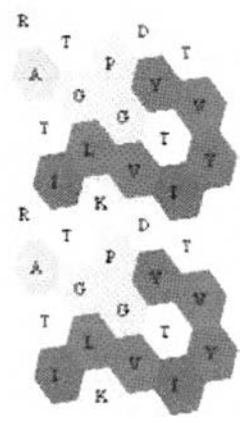
45 PGTETVWSII AAVGKES



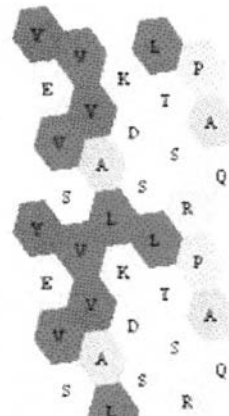
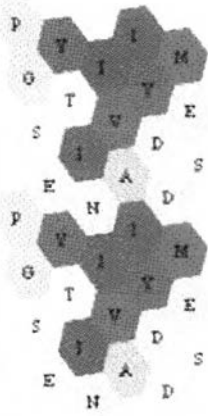
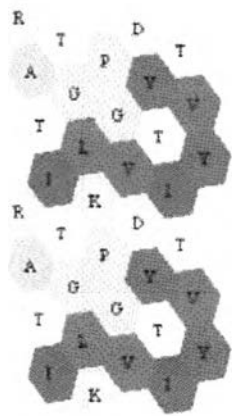
46 DAPAVTVRYVRI TVGETG



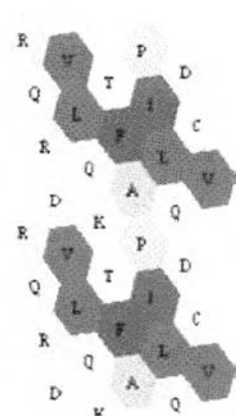
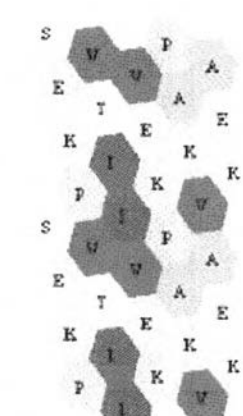
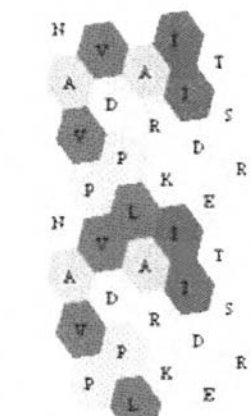
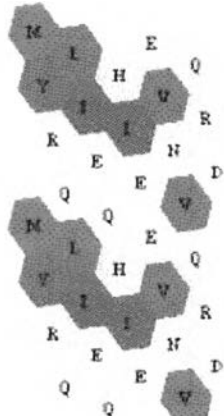
47 EATITGLKPGVDYTI TVY



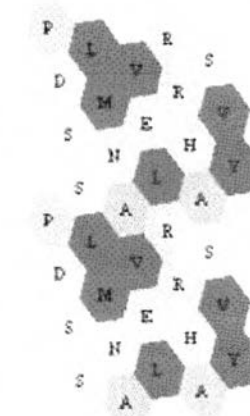
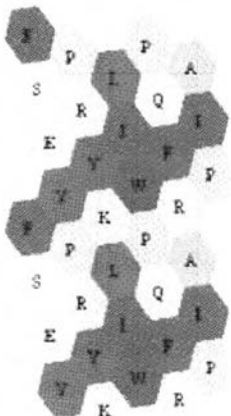
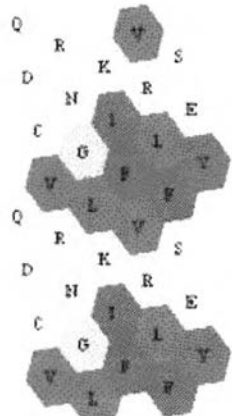
48 KATI TGLKPGVDATI TVY 49 PGEVTI NI VAI YDDMES 50 STS AVRS GLMVAIKYENS 51 VEVS VZALYIS LTS PDAQ



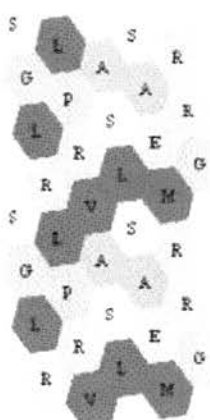
52 MYRQLI EQHI EEWVQRD 53 NAMPVDPLARKI I DETS R 54 SEKPVTI I VEKPAKVAEK 55 QRDVLQKTFAPILQDCV



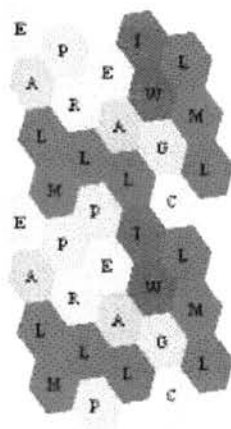
56 QDCVWGLKIFVRLFSEY 57 FSEVPRVYLI VDPQRAIP 58 PDS LINAVELRRHAS VY 59 ASVVLNGLGFI I DGMRE



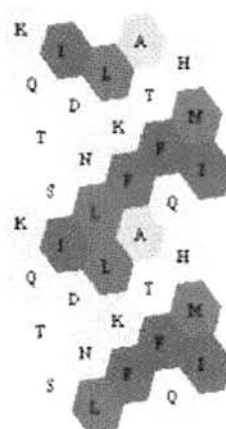
60 SGLRLPRVASLSAEMRRG



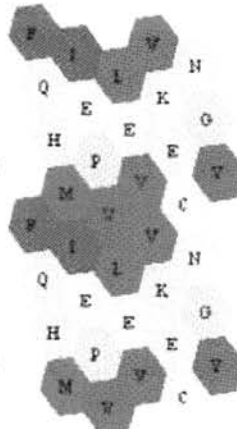
61 EALMPELPEALI WGLML



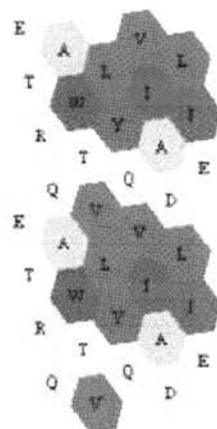
62 KQTSI DNLLEKFAFPQHNI



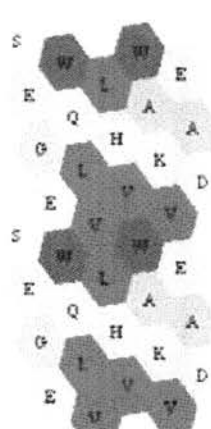
63 FQHM EPVLEVVYKECNGY



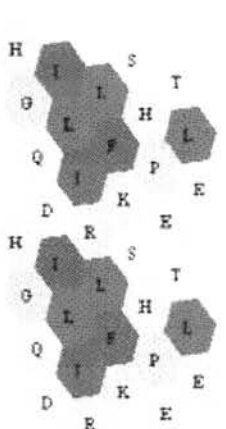
64 ETRQANTLYQMI ADLIE



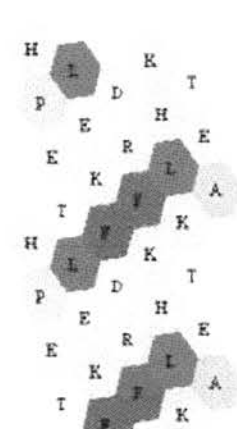
65 SEGEVQLVLFHWAKVEAD



66 HGQDI LIRLFKSHPETLE



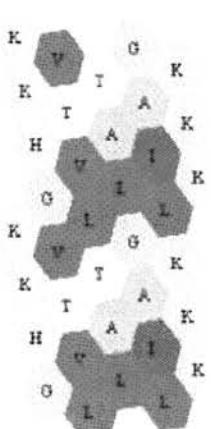
67 HPTLEKFDKFKGLTEA



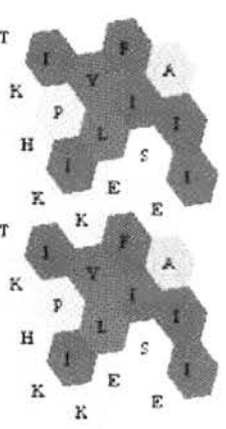
68 HTAEAKAS EDLKKHGVV



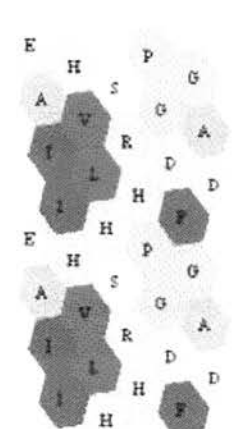
69 KKHGVVLTALGAILKKK



70 TPKKI PLYLEKI SEAI I



71 EAI I HVLEK RMD GDFGAD





## ประวัติผู้วิจัย

นายวิฑูร วิชัยพัฒน์ เกิดวันที่ 11 พฤษภาคม พ.ศ. 2517 ที่อำเภอทุ่งสง จังหวัด นครศรีธรรมราชสำเร็จการศึกษาปริญญาตรีวิทยาศาสตร์บัณฑิต สาขาคณิตศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2539 และ เข้าศึกษาต่อในหลักสูตร วิทยาศาสตร์มหาบัณฑิต ที่จุฬาลงกรณ์มหาวิทยาลัยในปี 2540 ปัจจุบันเป็นอาจารย์ประจำ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสยาม

