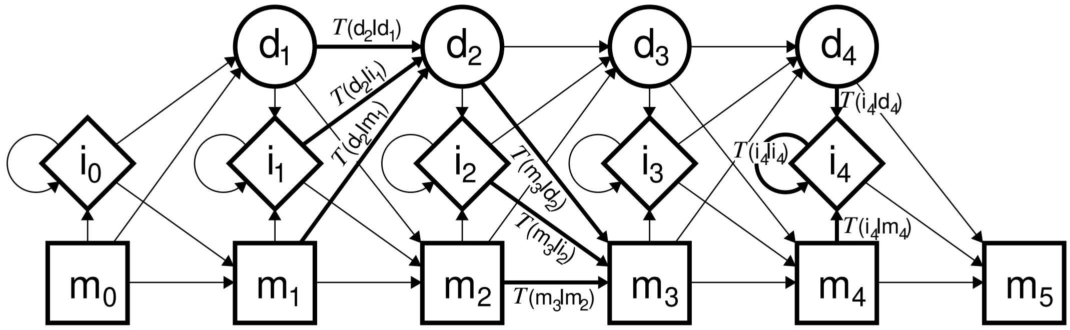


Figure 1: The model.

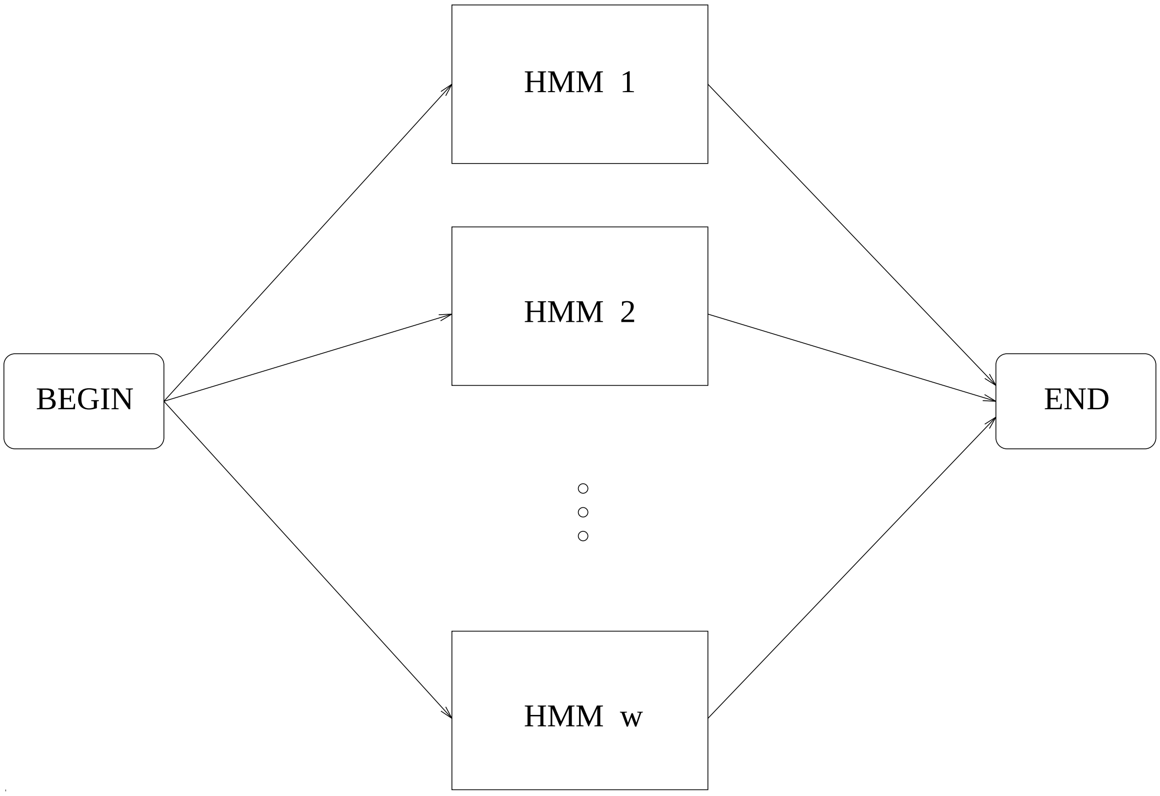


Figure 2: HMM architecture for discovering subfamilies.

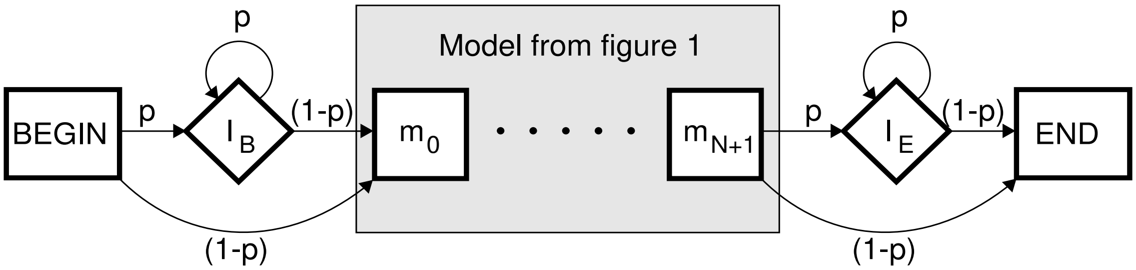


Figure 3: HMM architecture for modeling domains.

```
Helix                 AAAAAAAAAAAAAAAA   BBBBBBBBBBBBBBBBCCCCCCCCCCC    DDDDDDDEE
HBA_HUMAN  ---------VLSPADKTNVKAAWGKVGA--HAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSA
HBB_HUMAN  --------VHLTPEEKSAVTALWGKV----NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNP
MYG_PHYCA  ---------VLSEGEWQLVLHVWAKVEA--DVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE
GLB3_CHITP ----------LSADQISTVQASFDKVKG------DPVGILYAVFKADPSIMAKFTQFAG-KDLESIKGTA
GLB5_PETMA PIVDTGSVAPLSAAEKTKIRSAWAPVYS--TYETSGVDILVKFFTSTPAAQEFFPKFKGLTTADQLKKSA
LGB2_LUPLU --------GALTESQAALVKSSWEEFNA--NIPKHTHRFFILVLEIAPAAKDLFS-FLK-GTSEVPQNNP
GLB1_GLYDI ---------GLSAAQRQVIAATWKDIAGADNGAGVGKDCLIKFLSAHPQMAAVFG-FSG----AS---DP

Helix      EEEEEEEEEEEEEEEEEEE            FFFFFFFFFFFF  FFGGGGGGGGGGGGGGGGGGG
HBA_HUMAN  QVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKL--RVDPVNFKLLSHCLLVTLAAHLPAE
HBB_HUMAN  KVKAHGKKVLGAFSDGLAHL---D--NLKGTFATLSELHCDKL--HVDPENFRLLGNVLVCVLAHHFGKE
MYG_PHYCA  DLKKHGVTVLTALGAILKK----K-GHHEAELKPLAQSHATKH--KIPIKYLEFISEAIIHVLHSRHPGD
GLB3_CHITP PFETHANRIVGFFSKIIGEL--P---NIEADVNTFVASHKPRG---VTHDQLNNFRAGFVSYMKAHT--D
GLB5_PETMA DVRWHAERIINAVNDAVASM--DDTEKMSMKLRDLSGKHAKSF--QVDPQYFKVLAAVIADTVAAG----
LGB2_LUPLU ELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGSVHVSKG---VADAHFPVVKEAILKTIKEVVGAK
GLB1_GLYDI GVAALGAKVLAQIGVAVSHL--GDEGKMVAQMKAVGVRHKGYGNKHIKAQYFEPLGASLLSAMEHRIGGK

Helix       HHHHHHHHHHHHHHHHHHHHHHHHHH
HBA_HUMAN  FTPAVHASLDKFLASVSTVLTSKYR------
HBB_HUMAN  FTPPVQAAYQKVVAGVANALAHKYH------
MYG_PHYCA  FGADAQGAMNKALELFRKDIAAKYKELGYQG
GLB3_CHITP FA-GAEAAWGATLDTFFGMIFSKM-------
GLB5_PETMA -----DAGFEKLMSMICILLRSAY-------
LGB2_LUPLU WSEELNSAWTIAYDELAIVIKKEMNDAA---
GLB1_GLYDI MNAAAKDAWAAAYADISGALISGLQS-----
```

Figure 4: Seven representative globin sequences of known structure and their alignment taken from Bashford *et al.* (1987). The letters A to H in "Helix" denote the 8 different α-helices, Some regions, especially CD, D, and FG, are not well defined. The sequences and their SWISS-PROT identifiers are Human α (HBA_HUMAN), human β (HBB_HUMAN), sperm whale myoglobin (MYG_PHYCA), larval chironomous thummi globin (GLB3_CHITP), sea lamprey globin (GLB5_PETMA), lupinus luteus leghemoglobin (LGB2_LUPLU), and bloodworm globin (GLB1_GLYDI). (In SWISS-PROT 19 a "$" is used instead of an "_" in the identifiers.)

```
Helix                 AAAAAAAAAAAAAAAA   BBBBBBBBBBBBBBBBCCCCCCCCCCC    DDDDDDDEE
                      ***************+   ++++++++***************+++*            +
HBA_HUMAN  V.........LSPADKTNVKAAWGKVGA..HAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQ----
HBB_HUMAN  Vh........LTPEEKSAVTALWGKV--..NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNP
MYG_PHYCA  V.........LSEGEWQLVLHVWAKVEA..DVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE
GLB3_CHITP -.........LSADQISTVQASFDKV--..KGDPVG--ILYAVFKADPSIMAKFTQF-AGKDLESIKGTA
GLB5_PETMA PivdtgsvapLSAAEKTKIRSAWAPVYS..TYETSGVDILVKFFTSTPAAQEFFPKFKGLTTADQLKKSA
LGB2_LUPLU Ga........LTESQAALVKSSWEEFNA..NIPKHTHRFFILVLEIAPAAKDLF-SFLKGTSEVPQ-NNP
GLB1_GLYDI G.........LSAAQRQVIAATWKDIAGadNGAGVGKDCLIKFLSAHPQMAAVF-GF----SGASD---P

Helix      EEEEEEEEEEEEEEEEEEE            FFFFFFFFF   FFFFFGGG GGGGGGGGGGGGGGGG
           +******************            *********            ***************+
HBA_HUMAN  -VKGHGKKVADALTNAVAHVDD.....MPNALSALSDLHA...HKLRVDPV.NFKLLSHCLLVTLAAHLP
HBB_HUMAN  KVKAHGKKVLGAFSDGLAHLDN.....LKGTFATLSELHC...DKLHVDPE.NFRLLGNVLVCVLAHHFG
MYG_PHYCA  DLKKHGVTVLTALGAILKKKGH.....HEAELKPLAQSHA...TK-HKIPIkYLEFISEAIIHVLHSRHP
GLB3_CHITP PFETHANRIVGFFSKIIGELPN.....IEADVNTFVASHK...PR-GVTHD.QLNNFRAGFVSYMKAH--
GLB5_PETMA DVRWHAERIINAVNDAVASMDDtek..MSMKLRDLSGKHA...KSFQVDPQ.YFKVLAAVIADTVAA---
LGB2_LUPLU ELQAHAGKVFKLVYEAAIQLQVtgvvvTDATLKNLGSVHV...SK-GVADA.HFPVVKEAILKTIKEVVG
GLB1_GLYDI GVAALGAKVLAQIGVAVSHLGDegk..MVAQMKAVGVRHKgygNK-HIKAQ.YFEPLGASLLSAMEHRIG

Helix         HHHHHHHHHHHHHHHHHHHHHHHHHH
              ++**+******************
HBA_HUMAN  AEFTPAVHASLDKFLASVSTVLTSKY......R
HBB_HUMAN  KEFTPPVQAAYQKVVAGVANALAHKY......H
MYG_PHYCA  GDFGADAQGAMNKALELFRKDIAAKYkelgyqG
GLB3_CHITP TDF-AGAEAAWGATLDTFFGMIFSKM......-
GLB5_PETMA GD------AGFEKLMSMICILLRSAY......-
LGB2_LUPLU AKWSEELNSAWTIAYDELAIVIKKEMnda...A
GLB1_GLYDI GKMNAAAKDAWAAAYADISGALISGLq.....S
```

Figure 5: The alignment of the same seven globins as in figure 4, as obtained from our model trained on 400 randomly chosen globin sequences. The capital letters represent amino acids aligned to the main line of the model, "-" to deletions in the model, and lower-case letters to amino acids treated as insertions by the model. The "." is used as fill character to accommodate insertions. No attempt has been made to align the insertion regions. In the line above the alignments "*" indicates complete agreement of a column with the structural alignment (Figure 4) and "+" denotes a minor deviation (the only accepted difference is a reasonable displacements of a gap). The regions between the helices are not checked in this way. The training set contained five of the seven globins, not HBA_HUMAN and GLB5_PETMA.

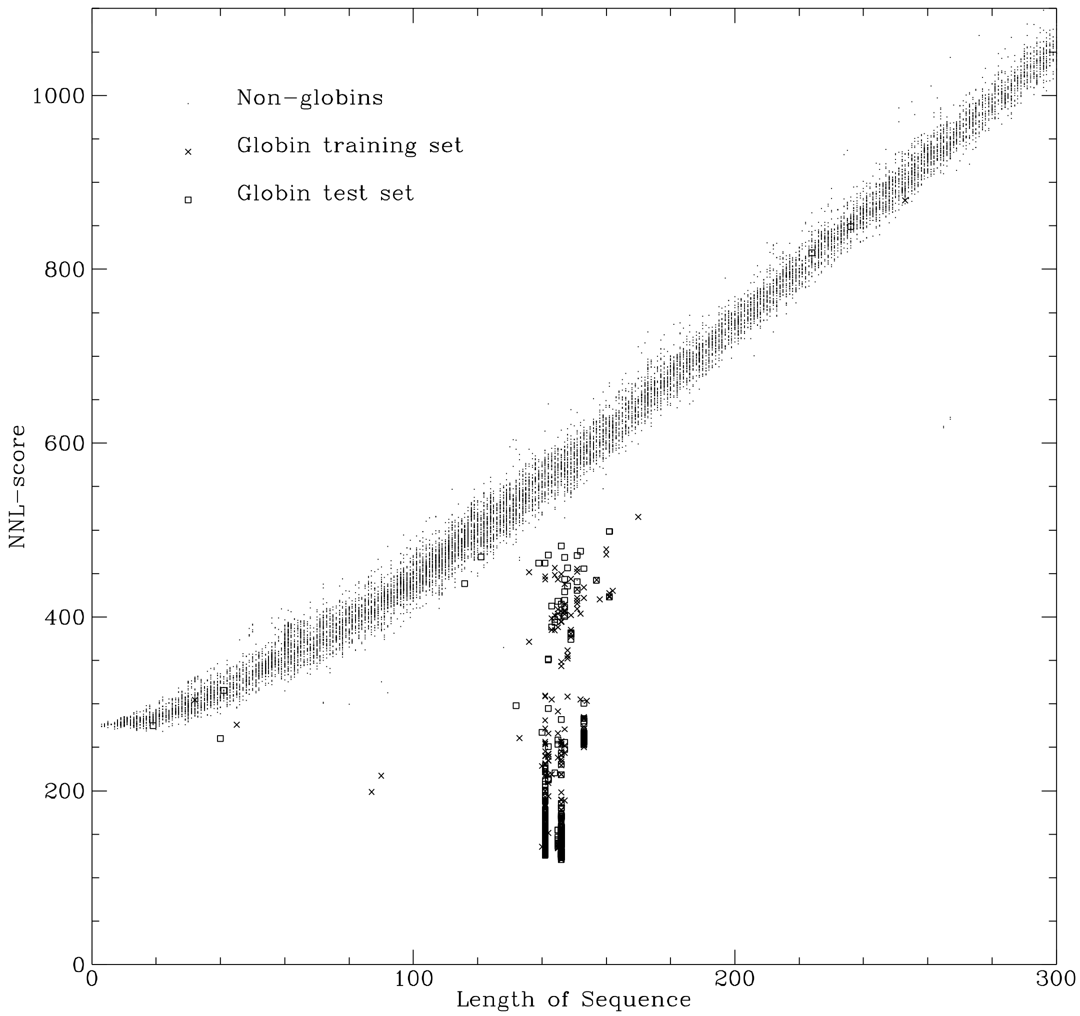
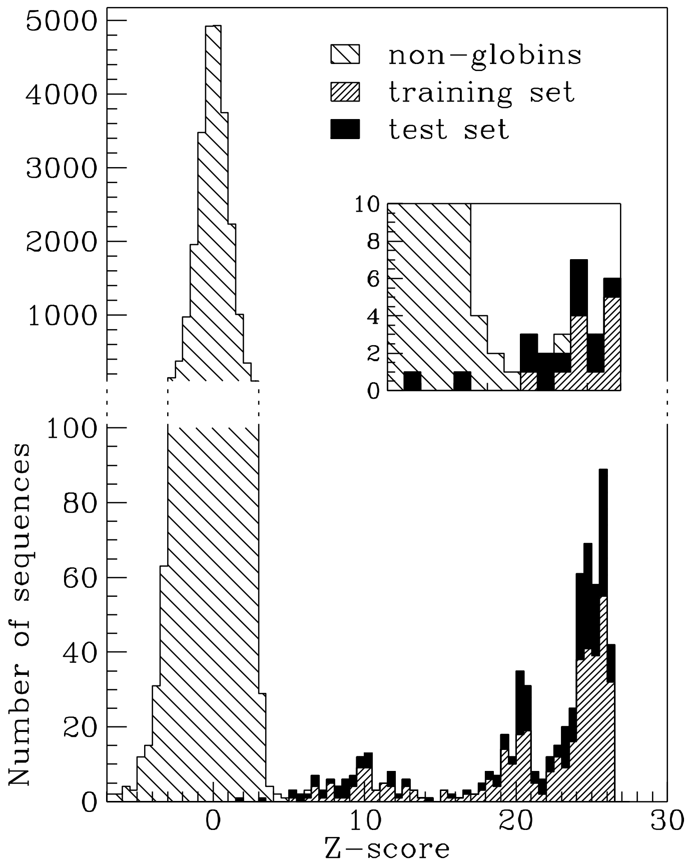


Figure 6: Plot of NLL-score vs. sequence length for globins and non-globins. All sequences of length less than 300 from the SWISS-PROT 22 database are shown, including partial sequences and three "false" globins from the globin file, and sequences from the database containing many Xs.

Figure 7: Histogram showing the number of sequences with a certain Z-score. The training set of 397 globins, the test set of 231 globins, and the rest of the sequences from SWISS-PROT 22 after "filtering" are shown. The insert shows a expansion of the region around a Z-score of 5.



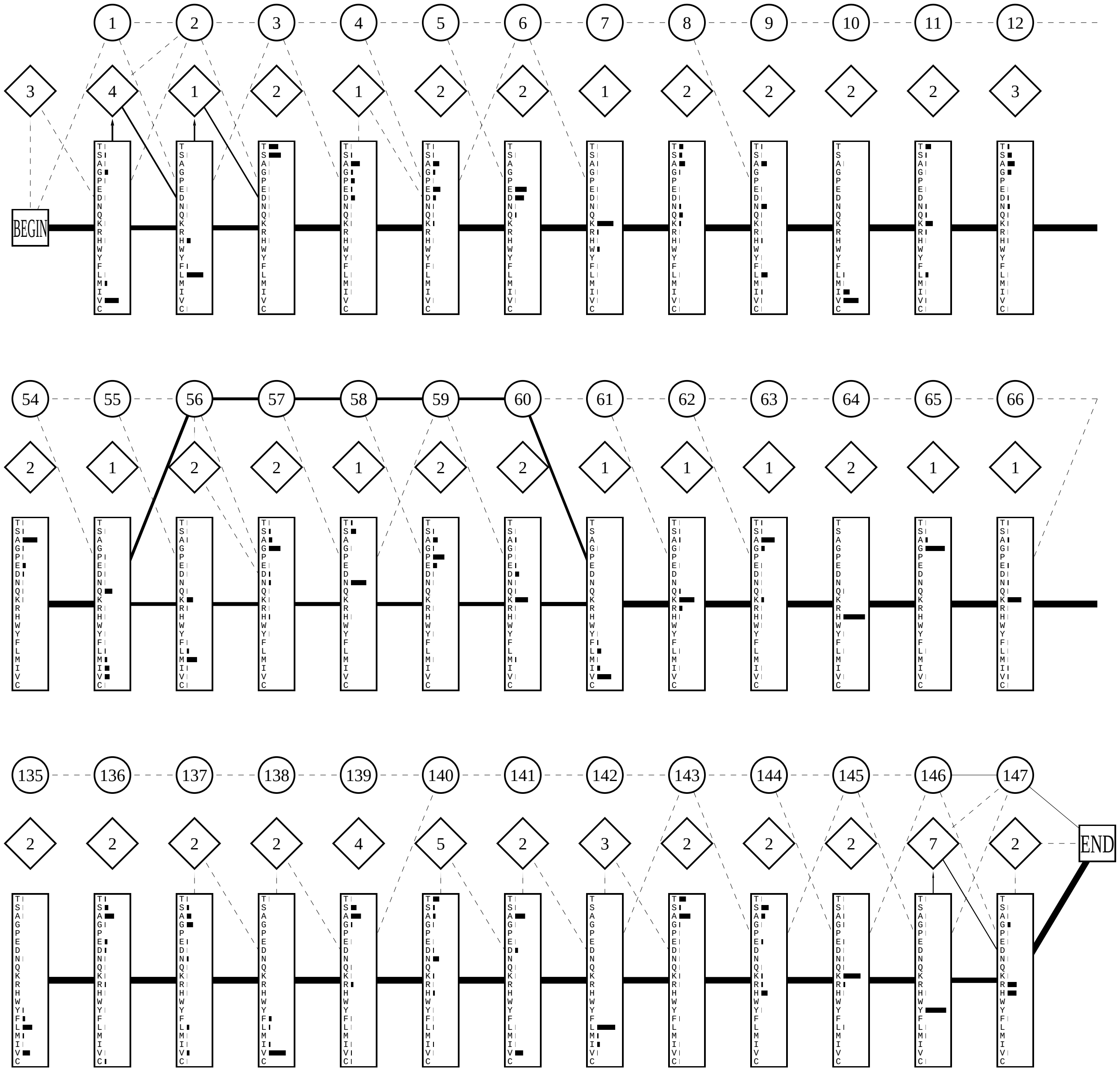


Figure 8: Parts of the final globin model. The position numbers are shown in the delete states.

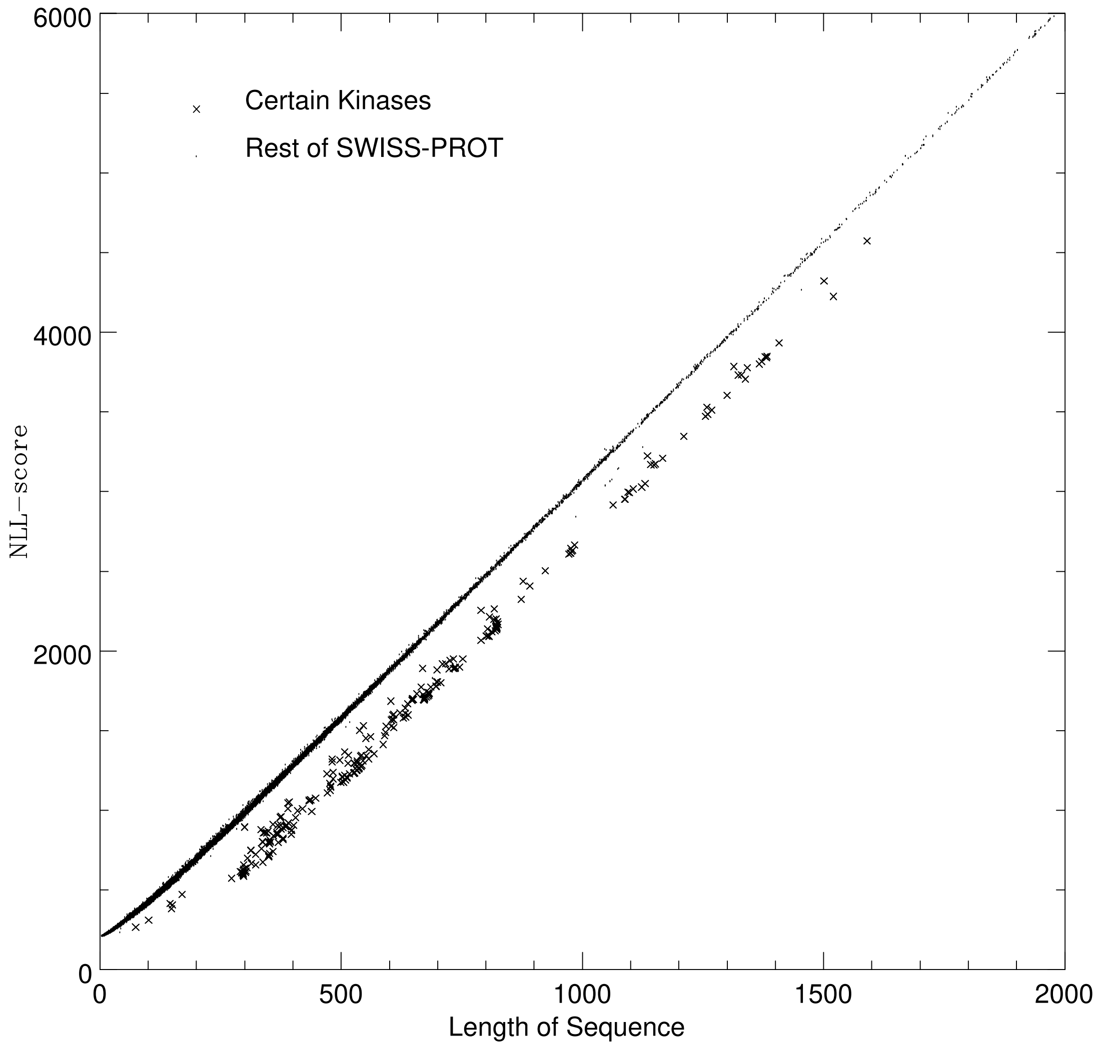


Figure 9: Scatter plot of NLL-score versus length for sequences in SWISS-PROT using the Kinase HMM.

Figure 10: Histogram showing the number of sequences with a certain Z-score relative to the kinase model.

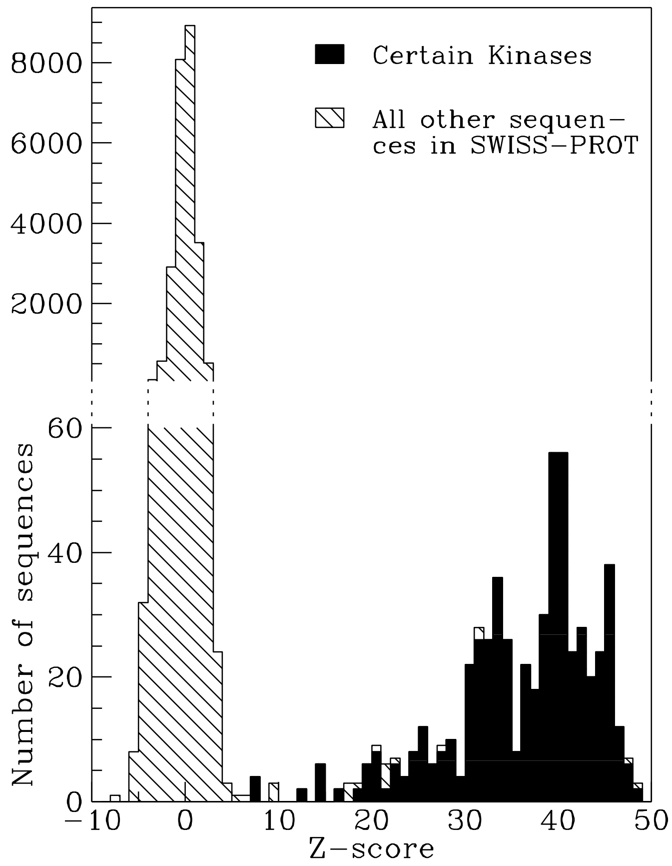


Figure 11:

A) Multiple sequence alignment generated by our kinase HMM of some of the sequences used to train the HMM (1-22) and "test" sequences from the SWISS-PROT 22 database (23-60) (see Section 3.2). Numerals appearing in the alignments indicate the number of amino acids to be inserted at that point, otherwise the notation follows the convention of Figure 5.

In **Subdomain**, the Roman numerals and "*" refer to the subdomains and residues conserved across 75 serine/threonine kinases given in Hanks and Quinn (1991). "A" and "B" in **PROSITE** refer to the ATP binding and catalytic regions respectively used to create two different signature patterns for kinases. **X-ray** identifies the location of the α-helices AA-AI and β-strands B1-B9 (read vertically) derived from the 2.7Åcrystal structure of the catalytic subunit of cAMP-dependent protein kinase (sequence 1) (Knighton *et al.*, 1991).

Sequences 1-22 are representative kinases taken from the March 1992 Protein Kinase Catalytic Domain Database (Hanks & Quinn, 1991). These are: **CAPK-ALPHA**, cAMP-dependent protein kinase catalytic subunit, α-form; **WEE1+**, "reduced size at division" mutant wild type allele gene product; **TIK**, mouse serine/threonine kinase; **SPK1**, *S. cerevisiae* kinase cloned with anti-p-tyr antibodies; **RSK1-N**, amino domain of type 1 ribosomal protein S6 kinase; **PYT**, putative serine/threonine kinase cloned with anti-p-tyr antibodies; **PKC-ALPHA**, protein kinase C, α-form; **PDGFR-B**, platelet-derived growth factor receptor B type; **PBS2**, polymixin B antibiotic resistance gene product; **MIK1**, *S. pombe mik1* acts redundantly with *wee1+*; **MCK1**, *S. cerevisiae* protein kinase; **INS.R**, insulin receptor; **HSVK**, Herpes simplex virus-US3 gene product; **ERK1**, rat insulin-stimulated protein kinase; **EGFR**, epidermal growth factor receptor (cellular homolog of *v-erbB*); **ECK**, receptor-like tyrosine kinase detected in epithelial cells; **DPYK1**, developmentally regulated tyrosine kinase in *D. discoideum*; **CLK**, mouse serine/threonine/tyrosine kinase; **CDC2HS**, human functional homolog of yeast cdc2+/CDC28; **CAMII-ALPHA**, calcium/calmodulin-dependent protein kinase II, α-subunit; **C-SRC**, cellular homolog of *v-src*; and **C-RAF**, cellular homolog of *v-raf/mil*. Sequences 2-4, 6, 10, 11, 14, 17 and 18 are the candidate dual-specificity protein kinases as defined by Lindberg (Lindberg *et al.*, 1992).

Sequences 23-40 are the SWISS-PROT 22 sequences designated as kinases by our HMM (Z-score > 6.0) but not by all three other methods, PROSITE, PROFILESEARCH and the keyword search. Sequences 41-50 are the top 10 sequences below our cutoff of 6.0 and 41-43 and 51-60 are sequences that were not classified as kinases by the HMM but were so by one or more (but not all) of the three other methods. Note that sequences identified as kinases by all four methods are not shown. All sequences that are less than 200 residues in length have been removed.

```
                 1         11        21        31        41        51                  71        81        91        101       111       121       131
   Subdomain     .........<-I-----*-*-----*------------->.........<-II----------*-----------------------.<-III--------------*--------------->....<-IV----
   PROSITE       .................AAAAAAAA.A.....................................................................................................
   X-ray         .........BBBB.BB.........BBBBB..BB..................B......BBBBBBB.BAAA..AA.A..........A..A.AAA.........A.A.AA......AAA..AA.A............
   X-ray         .........1111.11.........22222..22..................3......3333333.3BBB..BB.B..........B..C.CCC.........C.C.CC......CCC..CC.C............
 1 CAPK-ALPHA    .........FERI.KTLGTGSFGR.VMLVK..HK.....E..........TG.N......HYAMKIL.DKqk..VV.K..........LK.Q.IEH.........T.L.NE......KRI..LQ.AV....N...F...
 2 WEE1+         .........FRNV.TLLGSGEFSE.VFQVE..DP.....V..........EK.Tl.....KYAVKKL.KVk...FS.G..........PK.E.RNR.........L.L.QE......VSI..QR.AL....Kg..H...
 3 TIK           .........FEDI.EEIGLGGFGQ.VFKAK..HR.....I..........DG.K......RYAIKRV.KY....--.N..........TE.K.AE-.........-.-.HE......VQA..LA.EL....N...H...
 4 SPK1          df.......SIID.EVVGQGAFAT.VKKAI..ER.....T..........TG.K......TFAVKII.SKrk..VI.G..........NM.DgVT-.........-.-.RE......LEV..LQ.KL....N...H...
 5 RSK1-N        .........FELL.KVLGQGSFGK.VFLVR..KVtrp..D..........SG.H......LYAMKVL.KKa...TL.K..........VR.D.RVR.........T.K.ME......RDI..LA.DV....N...H...
 6 PYT           .........YSIL.KQIGSGGSSK.VFQVL..NE.....K..........K-.Q......IYAIKYV.NLe...EA.D..........NQ.T.LDS.........Y.R.NE......IAY..LN.KL....Qqh.S...
 7 PKC-ALPHA     .........FNFL.MVLGKGSFGK.VMLAD..RK.....G..........TE.E......LYAIKIL.KKdv..VI.Q..........DD.D.VEC.........T.M.VE......KRV..LA.LL....Dk..P...
 8 PDGFR-B       .........LVLG.RTLGSGAFGQ.VVEAT..AH.....G..........LS.Hsqatm.KVAVKML.KS....TA.R..........SS.E.KQA.........L.M.SE......LKI..MS.HL....Gp..H...
 9 PBS2          .........LEFL.DELGHGNYGN.VSKVL..HK.....P..........TN.V......IMATKEV.RL....EL.D..........EA.K.FRQ.........I.L.ME......LEV..LH.KC....N...S...
10 MIK1          .........FQQV.KPIHESDFSF.VYHVS..SInp...P..........TE.T......VYVVKML.KKn...AA.K..........FT.G.KER.........H.L.QE......VSI..LQ.RL....Qa..C...
11 MCK1          .........VKEY.RKIGRGAFGT.VVQAY..LT.....Q..........DK.Knwlg..PFAIKKV.PA....--.H..........TE.Y.KS-.........-.-.RE......LQI..LR.IA....D...H...
12 INS.R         .........ITLL.RELGQGSFGM.VYEGN..AR.....D..........II.Kgeaet.RVAVKTV.NE....SA.S..........LR.E.RIE.........F.L.NE......ASV..MK.GF....T...C...
13 HSVK          .........FTIH.GALTPGSEGC.VFDSS..HP.....D..........YP.Q......RVIVKAG.WY....-T.S..........TS.-.---.........-.-.HE......ARL..LR.RL....D...H...
14 ERK1          .........YTQL.QYIGEGAYGM.VSSAY..DH.....V..........RK.T......RVAIKKI.SP....FE.H..........QT.Y.CQR.........T.L.RE......IQI..LL.GF....R...H...
15 EGFR          .........FKKI.KVLGSGAFGT.VYKGL..WI.....P..........EG.Ekvki..PVAIKEL.RE....AT.S..........PK.A.NKE.........I.L.DE......AYV..MA.SV....D...N...
16 ECK           .........VTRQ.KVIGAGEFGE.VYKGM..LKt....S..........SG.Kkev...PVAIKTL.KA....GY.T..........EK.Q.RVD.........F.L.GE......AGI..MG.QF....S...H...
17 DPYK1         ne.......LEFG.QTIGKGFFGE.VKRGY..WR.....E..........T-.-......DVAIKII.YRdq..FK.T..........KS.S.LVM.........F.Q.NE......VGI..LS.KL....R...H...
18 CLK           ar.......YEIV.DTLGEGAFGK.VVECI..DH.....K..........VGgR......RVAVKIV.KN....--.V..........DR.Y.CEA.........A.Q.SE......IQV..LE.HL....Ntt.D...
19 CDC2HS        .........YTKI.EKIGEGTYGV.VYKGR..HK.....T..........TG.Q......VVAMKKI.RLe...SE.E..........EG.V.PST.........A.I.RE......ISL..LK.EL....R...H...
20 CAMII-ALPHA   .........YQLF.EELGKGAFSV.VRRCV..KV.....L..........AG.Q......EYAAKII.NTk...KL.S..........AR.D.HQK.........L.E.RE......ARI..CR.LL....K...H...
21 C-SRC         .........LRLE.VKLGQGCFGE.VWMGT..WN.....G..........T-.T......RVAIKTL.KP....--.G..........TM.S.PEA.........F.L.QE......AQV..MK.KL....R...H...
22 C-RAF         .........VMLS.TRIGSGSFGT.VYKGK..WH.....G..........D-.-......-VAVKIL.KVv...DP.T..........PE.Q.FQA.........F.R.NE......VAV..LR.KT....R...H...
23 KLSK_HUMAN    mgcgcss244LKLV.ERLGAGQFGE.VWMGY..YN.....G..........H-.T......KVAVKSL.KQ....--.G..........SM.S.PDA.........F.L.AE......ANL..MK.QL....Q...H...
24 KLSK_MOUSE    mgcvcss244LKLV.ERLGAGQFGE.VWMGY..YN.....G..........H-.T......KVAVKSL.KQ....--.G..........SM.S.PVP.........F.L.AE......ANL..MK.QL....Q...H...
25 ARKB_HUMAN    madleav190FSVH.RIIGRGGFGE.VYGCR..KR.....D..........TG.K......MYAMKCL.DKk...RI.K..........MK.Q.GET.........LaL.NE......RIM..LS.LVstg.D...C...
26 ARKB_BOVIN    madleav190FSVH.RIIGRGGFGE.VYGCR..KA.....D..........TG.K......MYAMKCL.DKk...RI.K..........MK.Q.GET.........LaL.NE......RIM..LS.LVstg.D...C...
27 BYR1_SCHPO    mfkrrrnp65LEVV.RHLGEGNGGA.VSLVK..HR.....-..........-N.I......FMARKTV.YV....GS.D..........SK.L.QKQ.........I.L.RE......LGV..LH.HC....R...S...
28 CYGR_ARBPU    mattrll579QQIF.ATIG------.----T..YR.....G..........T-.-......ICAIHAVhKN....--.H..........ID.L.TRA.........V.R.TE......LKL..MR.DM....R...H...
29 ANPA_RAT      mpgsrrv536GSLL.TT--EGQF-Q.VFAKT..AY.....Y..........KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
30 ANPA_HUMAN    mpgprrp471LEVL.ALVGSLS---.-LLGI..LI.....Vsffiyrkm70KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
31 ANPB_HUMAN    malpsll512SRLT.LSLRGSSYGS.LMTAH..GKyqifaN..........TG.Hfkgn..VVAIKHV.NK....-K.R..........IE.L.TRQ.........V.L.FE......LKH..MR.DV....Q...F...
32 ANPA_MOUSE    mprsrrv536GSLL.TT--EGQF-Q.VFAKT..AY.....Y..........KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
33 ANPB_RAT      malpsll512SRLT.LSLRGSSYGS.LMTAH..GKyqifaN..........TG.Hfkgn..VVAIKHV.NK....-K.R..........IE.L.TRQ.........V.L.FE......LKH..MR.DV....Q...F...
34 CYGS_STRPU    maharhl582QQIF.ATIG------.----T..YR.....G..........T-.-......VCALHAVhKN....--.H..........ID.L.TRA.........V.R.TE......LKI..MR.DM....R...H...
35 VPSF_YEAST    gaqlslvv34SRFL.KT--------.-CKAL..DP.....-..........NG.E......IV-IKVFiKP....--.K..........DQyS.LRP.........F.L.QR......IRA..QSfKL....Gq..L...
36 HSER_RAT      mtsllgl503----.---------R.VRQCK..YD.....-..........-K.K......KVILKDL.KHcd..GNfS..........EK.Q.KI-.........-.-.-E......LN-..-K.LL....Q...S...
37 HSER_HUMAN    mktllld445LLLV.ALL-------.-MLRK..YR.....Kdyelrqkk49DK.K......RVILKDL.KHnd..GNfT..........EK.Q.KI-.........-.-.-E......LNK..LL.QI....D...Y...
38 KR2_VZVD      mdaddtp138----.---GRGTYGR.V----..HI.....Y..........PS.S......KIAVKTM.DSr...VF.N..........RE.L.INA.........I.LaSE......GSIraGE.RL....G...I...
39 KR2_HSV11     mdesrrq157----.---GSGGYGD.VQLIR..EH.....K..........L-.-......--AVKTI.KEkew.FA.V..........EL.I.ATL.........L.V.GE......C-V..LR.AG....R...T...
40 KR1_HSVI1     m.........----.---------E.VYAWEtaHF.....L..........D-.-......--AAPKL.IEwe..VS.G..........TR.E.NR-.........-.-.--......---..--.--....-...-...
41 KR2_EBV       sgwrssv108----.-LLGRGSYGA.VY---..AH.....A..........DN.A......TV---KL.YD....SV.T..........EL.Y.HE-.........-.-.--lmvcdmIQI..GK.ATaedgQ...D...
42 KRB2_VACCV    mesfkycf14WIIG.NTLYSGN-SI.LYKVR..KN.....F..........TS.Sfy....NYVMKI-.--....--.D..........HK.S.HKP.........L.L.SE......IRF..YI.SVl...D...Plti
43 KRB2_VACCC    mesfkycf14WIIG.NTLYSGN-SI.LYKVR..KN.....F..........TS.Sfy....NYVMKI-.--....--.D..........HK.S.HKP.........L.L.SE......IRF..YI.SVl...D...Plti
44 AK3_ECOLI     mseivvsk32VRL-.---------V.VLSAS..AG.....I..........TN.L......LVALAEG.LEpcerFE.K..........LD.A.IRN.........I.Q.--......FAI..LE.RL....R...Y...
45 PSP_MOUSE     mfqlgslv18ESLL.GELGS-----.---AV..NN.....-..........--.-......---LKIL.NP....-P.S..........EA.V.PQN.........L.N.LD......VEL..LQ.QAt...Sw..-...
46 DHOM_BACSU    mkairvg...LLGL.GTVGSG---V.VKIIQ..DH.....Q..........DK.LmhqvgcPVTIKKV.LV....--.K..........DL.E.KK-.........-.-.RE......VDL..PK.EVlt..-...-...
47 FLIG_BACSU    marrd.....----.----------.----Q..DK.....L..........TG.Kq.....KAAILMI.SL....GL.DvsasvykhltDE.E.IER.........L.T.LE......ISG..VR.SV....D...H...
48 CALQ_RABIT    mnaadrmg12LLLL.LVLGSPQSG-.-VHGE..EG.....L..........D-.-......------F.PEyd..GV.D..........RV.I.NVN.........A.K.NY......KNV..FK.KYe...-...-...
49 NU1M_PODAN    my........YSII.ISL-----IE.VVLVL..VP.....A..........L-.-......-------.-L....--.G..........IA.Y.V--.........T.I.AE......RKT..MA.SM....QrrlG...
50 RUVA_ECOLI    migrlrgi15LVLI.EVGGVG----.-YEVH..MP.....M..........T-.-......-----CF.YE....LP.E..........AG.-.---.........-.-.QE......AIV..FT.HF....-...-...
51 U15R_HSV6U    mdngvet194KTMA.RVLGVGAYGK.VF---..DL.....D..........K-.-......-VAIKTA.NE....--.D..........ES.V.ISA.........F.I.A-......---..--.--....-...-...
52 KRF1_VACCC    mgvandss94----.----TGGYGI.VFKID..NY.....Vvkfvfeat79YS.S......KVFLKAF.NEr...KD.S..........IK.F.VK-.........-.-.--......--L..LS.HF....Y...P...
53 UL97_HCMVA    mssalrs335----.-RLGQGSFGE.VW---..--.....P..........LD.R......YRVVKVA.RK....--.H..........SE.T.VLT.........V.W.M-......---..--.--....-...-...
54 KKA6_ACIBA    melpniiq13SVLEpNKIGQ-SPSD.VYS-F..NR.....N..........N-.E......TFFLKRS.STly..TE.T..........TY.S.VS-.........-.-.RE......AKM..LS.WL....Se..K...
55 KKA8_ECOLI    mndidree27WARD.KV---GQSGCaVYRLH..SK.....S..........GG.S......DLFLKHG.KDafadDV.T..........DE.M.VR-.........-.-.--......---..--.-L....Rw..-...
56 KGPB_BOVIN    ..........----.----------.--MGT..LR.....-..........-D.L......QYA---L.QEk...IE.E..........LR.Q.RDA.........L.I.DE......LEL..--.--....-...-...
57 EGFR_CHICK    mgvrspls16VLVL.LLLG------.VALCS..AV.....E..........E-.-......----KKV.CQ....GT.N..........NK.-.---.........-.-.--......---..LT.QL....G...Hved
58 KKA1_ECOLI    mshiqret32----.-----VGQSG-.---ATi.YR.....L..........YG.Kpdape.LF----L.KH....--.G..........KG.S.VA-.........-.-.--......---..--.--....-...-...
59 KDTK_DROME    mpsarlf393NRLS.KELGK-EFGK.EF-GK..EF.....Gpapsmslq72PG.Gpp....PFLKKKM.PR....--.P..........KG.Q.HSAprggpprswT.N.TE......LT-..-E.AL....Q...H...
60 KPCG_HUMAN    maglgpg121----.-------YGL.VHQGM..KCsccemN..........VH.R......RC----V.RS....VP.S..........LC.G.VDH.........T.E.RR......GRL..QL.EI....R...A...
```

```
               141       151       161       171       181       191       201       211       221       231       241       251       261       271
   Subdomain   -------------------------------------------><-V---------------------------------------------------------------------->.........<-VIa-------------
   PROSITE     ..............................................................................................................................................
   X-ray       .........................BB....B....BBB.....B..........BB..BBB.................AAA......AA...AA...A.........................AA...A.....AAA.
   X-ray       .........................44....4....444.....5..........55..555.................DDD......DD...DD...D.........................EE...E.....EEE.
 1 CAPK-ALPHA  .......P......FL...V.....KLE....F....SFKD..NSN..........LY..MVM..EYVPGG.......E.MFS......HL...RR...I.....G..-..R..........FS.EP...H.....ARF.
 2 WEE1+       .......D......HI...V.....ELM....D....SWEH..GGF..........LY..MQV..ELCENG.......S.LDR......FL...EE...Q.....GqlS..R..........LD.EF...R.....VWK.
 3 TIK         .......V......NI...V.....QYHscweG....VDYD..PEHsmsdtsry12LF..IQM..EFCDKG.......T.LEQ......WM...RNr..N.....Q..S..K..........VD.KA...L.....ILD.
 4 SPK1        .......P......RI...V.....RLK....G....FYED..TES..........YY..MVM..EFVSGG.......D.LMD......FV...AA...H.....G..-..A..........VG.ED...A.....GRE.
 5 RSK1-N      .......P......FV...V.....KLH....Y....AFQT..EGK..........LY..LIL..DFLRGG.......D.LFT......RL...SK...E.....V..-..M..........FT.EE...D.....VKF.
 6 PYT         .......D......KI...I.....RLY....D....YEIT..DQY..........IY..MVM..ECGN-I.......D.LNS......WL...KK...K.....K..-..S..........ID.PW...E.....RKS.
 7 PKC-ALPHA   .......P......FL...T.....QLH....S....CFQT..VDR..........LY..FVM..EYVNGG.......D.LMY......HI...QQ...V.....G..-..K..........FK.EP...Q.....AVF.
 8 PDGFR-B     .......L......NV...V.....NLL....G....ACTK..GGP..........IY..IIT..EYCRYG.......D.LVD......YL...HR...N.....K..H..Tflqhhsdk98LS.YM...D.....LVG.
 9 PBS2        .......P......YI...V.....DFY....G....AFFI..EGA..........VY..MCM..EYMDGG.......S.LDK......IY...DE...Ssei..G..G..-..........ID.EP...Q.....LAF.
10 MIK1        .......P......FV...V.....NLV....N....VWSY..NDN..........IF..LQL..DYCENG.......D.LSL......FL...SE...L.....GllQ..V..........MD.PF...R.....VWK.
11 MCK1        .......P......NI...V.....KLQ....Y....FFTH..LSPqdnk......VYqhLAM..ECLP-E.......T.LQI......EI...NRyvtN.....K..L..E..........MP.LK...H.....IRL.
12 INS.R       .......H......HV...V.....RLL....G....VVSK..GQP..........TL..VVM..ELMAHG.......D.LKS......YL...RS...L.....R..P..Eaennpgrpp.PT.LQ...E.....MIQ.
13 HSVK        .......P......AI...L.....PLL....D....LHVV..SGV..........TC..LVL..PKYQA-.......D.LYT......YL...SR...R.....L..N..P..........LG.RP...Q.....IAA.
14 ERK1        .......E......NV...I.....GIR....D....ILRA..PTLeamrd.....VY..IVQ..DLME-T.......D.LYK......LL...KS...Q.....Q..-..-..........LS.ND...H.....ICY.
15 EGFR        .......P......HV...C.....RLL....G....ICLT..S-T..........VQ..LIT..QLMPFG.......C.LLD......YV...RE...H.....K..D..N..........IG.SQ...Y.....LLN.
16 ECK         .......H......NI...I.....RLE....G....VISK..YKP..........MM..IIT..EYMENG.......A.LDK......FL...RE...K.....D..G..E..........FS.VL...Q.....LVG.
17 DPYK1       .......P......NV...V.....QFL....G....ACTAggEDH..........HC..IVT..EWMGGG.......S.LRQ......FL...TD...H.....F..-..N..........LL.EQnphI.....RLK.
18 CLK         .......Phstf..RC...V.....QML....E....WFEH..RGH..........IC..IVF..ELL-GL.......S.TYD......FI...KEn..S.....F..L..P..........FR.MD...H.....IRK.
19 CDC2HS      .......P......NI...V.....SLQ....D....VLQD..S-R..........LY..LIF..EFLS-M.......D.LKK......YL...DSip.P.....G..Q..Y..........MD.SS...L.....VKS.
20 CAMII-ALPHA .......P......NI...V.....RLH....D....SISE..EGH..........HY..LIF..DLVTGG.......E.LFE......DI...VA...R.....E..-..Y..........YS.EA...D.....ASH.
21 C-SRC       .......E......KL...V.....QLY....A....VVSE..E-P..........IY..IVT..EYMSKG.......S.LLD......FL...KGe..T.....G..K..Y..........LR.LP...Q.....LVD.
22 C-RAF       .......V......NI...L.....LFM....G....YMTK..D-N..........LA..IVT..QWCEGS.......S.LYK......HL...HV...Q.....E..T..K..........FQ.MF...Q.....LID.
23 KLSK_HUMAN  .......Q......RL...V.....RLY....A....VVTQ..E-P..........IY..IIT..EYMENG.......S.LVD......FL...KTp..S.....G..I..K..........LT.IN...K.....LLD.
24 KLSK_MOUSE  .......P......RL...V.....RLY....A....VVTQ..E-P..........IY..IIT..EYMENG.......S.LVD......FL...KTp..S.....G..I..K..........LN.VN...K.....LLD.
25 ARKB_HUMAN  .......P......FI...V.....CMS....Y....AFHT..PDK..........LS..FIL..DLMNGG.......D.LHY......HL...SQ...H.....G..-..V..........FS.EA...D.....MRF.
26 ARKB_BOVIN  .......P......FI...V.....CMS....Y....AFHT..PDK..........LS..FIL..DLMNGG.......D.LHY......HL...SQ...H.....G..-..V..........FS.EA...D.....MRF.
27 BYR1_SCHPO  .......P......YI...V.....GFY....G....AFQY..KNN..........IS..LCM..EYMDCG.......S.LDA......IL...RE...G.....G..-..P..........IP.LD...I.....LGK.
28 CYGR_ARBPU  .......D......NI...C.....PFI....G....ACID..RPH..........IC..ILM..HYCAKG.......S.LQD......IM...EN...D.....D..I..K..........LD.SM...F.....LAS.
29 ANPA_RAT    .......E......HL...T.....RFV....G....ACTD..PPN..........IC..ILT..EYCPRG.......S.LQD......IL...EN...E.....S..I..T..........LD.WM...F.....RYS.
30 ANPA_HUMAN  .......E......HL...T.....RFV....G....ACTD..PPN..........IC..ILT..EYCPRG.......S.LQD......IL...EN...E.....S..I..T..........LD.WM...F.....RYS.
31 ANPB_HUMAN  .......N......HL...T.....RFI....G....ACID..PPN..........IC..IVT..EYCPRG.......S.LQD......IL...EN...D.....S..I..N..........LD.WM...F.....RYS.
32 ANPA_MOUSE  .......E......QL...T.....RFV....G....ACTD..PPN..........IC..ILT..EYCPRG.......S.LQD......IL...EN...E.....S..I..T..........LD.WM...F.....RYS.
33 ANPB_RAT    .......N......HL...T.....RFI....G....ACID..PPN..........IC..IVT..EYCPRG.......S.LQD......IL...EN...D.....S..I..N..........LD.WM...F.....RYS.
34 CYGS_STRPU  .......D......NI...C.....PFI....G....ACID..RPH..........IS..ILM..HYCAKG.......S.LQD......IL...EN...D.....D..I..K..........LD.SM...F.....LSS.
35 VPSF_YEAST  .......P......HV...L.....NYS....K....LIET..NRA..........GY..MIR..QHLKN-.......N.LYD......RL...S-...L.....R..P..Y..........LQ.DI...E.....LKF.
36 HSER_RAT    .......Dyy....NL...T.....KFY....G....TVKL..DTR..........IF..GVV..EYCERG.......S.LRE......VL...ND...TisypdG..T..F..........MD.WE...F.....KIS.
37 HSER_HUMAN  .......Y......NL...T.....KFY....G....TVKL..DTM..........IF..GVI..EYCERG.......S.LRE......VL...ND...TisypdG..T..F..........MD.WE...F.....KIS.
38 KR2_VZVD    .......S......SI...V.....CLL....G....FSLQ..TK-..........-Q..LLF..PAYD-M.......D.MDE......YI...VR...L.....S..R..R..........LT.IP...DhidrkIAH.
39 KR2_HSV11   .......H......NIrgfI.....APL....G....FSLQ..QR-..........-Q..IVF..PAYD-M.......D.LGK......YIgqlAS...L.....Rt.T..N..........PS.VS...T.....ALHq
40 KR1_HSVI1   .......-......--...-.....---....-....----..-KS..........AQ..MVTfsECGLYG.......S.MEG......YF...YR...E.....R..A..-..........-T.VD...I.....CAT.
41 KR2_EBV     .......K......AL...V.....DYL....S....ACTSc.H-A..........LF..M--p.QFRC--.......S.LQD......YG...HW...H.....D..G..-..........-S.IE...P.....LVR.
42 KRB2_VACCV  dnwtr14L......AI...P.....DLY....G....IGET..DDY..........MF..FV-..--IKN-.......-.-LG......RV...FA...P.....K..D..T..........ES.VF...E.....A--.
43 KRB2_VACCC  dnwtr14L......AI...P.....DLY....G....IGET..DDY..........MF..FV-..--IKN-.......-.-LG......RV...FA...P.....K..D..T..........ES.VF...E.....A--.
44 AK3_ECOLI   .......P......NV...IreeieRLL....E....NITV..LAEaaalats...PA..LTD..ELVSHG.......E.LMStllfveIL...RE...R.....D..V..Q..........AQ.WF...D.....VRK.
45 PSP_MOUSE   .......-......--...-.....---....-....----..---..........--..---..PLAKN-.......-.--S......IL...ET...L.....N..-..T..........AD.LG...N.....LKS.
46 DHOM_BACSU  .......-......--...T.....EVY....D....VIDD..P-D..........VD..VVI..EVIGGV.......EqTKQ......YL...VDa..L.....R..S..Kkhvvtank48LS.SD...R.....ITK.
47 FLIG_BACSU  .......Qkkd...--...-.....---....E....IIEE..FHN..........IA..IAQ..DYIS--.......-.---......--...--...Q.....G..G..-..........--.--...-.....-LS.
48 CALQ_RABIT  .......-.......-V...L.....ALL....Y....HEPP..ED-..........--..---..---DK-.......-.---......--...-A...S.....Q..R..Q..........FEmEE...L.....ILE.
49 NU1M_PODAN  .......P......NF...V.....GYY....G....LLQAf.ADA..........LK..LLLk.EYVA-P.......T.QAN......II...LFf..L.....G..P..V..........IT.LI...Fs....LLG.
50 RUVA_ECOLI  .......-......--...-.....---....-....VVRE..DAQ..........LL..---..----YGfnnkqerT.LFK......EL...IKt..N.....G..-..-..........VG.PK...-.....---.
51 U15R_HSV6U  .......-......--...-.....---....G....VIRA..K--..........--..---..---SGA.......D.LLS......H-...--...-.....-..-..Ecvinnlli30FE.DW...D.....VRN.
52 KRF1_VACCC  .......AvinsniNV...I.....NYF....NrmfhFFEH..EKRtnyeyerg..NI..IIF..PLALY-.......S.ADK......VD...TE...L.....A..I..K..........LG.FK...S.....LVQ.
53 UL97_HCMVA  .......S......GL...I.....RTR....A....AGEQ..QQP..........PS..LV-..--GTG-.......-.---......--...-V...H.....R..G..Lltatgccl32ID.SY...R.....RAF.
54 KKA6_ACIBA  .......L......KV...P.....ELI....M....TFQD..EQF..........EF..MIT..KAINA-.......-.--K......PI...SA...L.....F..-..-..........LT.DQ...E.....LLA.
55 KKA8_ECOLI  .......-......--...-.....-LA....G....HISV..P--..........--..---..------.......S.VVS......FV...RTp..N.....Q..Aw.L..........LT.TA...I.....HGK.
56 KGPB_BOVIN  .......-......--...-.....---....E....LDQK..D--..........--..---..ELIQK-.......-.LQN......EL...DK...Y.....R..S..-..........--.--...-.....---.
57 EGFR_CHICK  hftsl32-......--...L.....TFL....K....TIQEv.AGY..........VL..IAL..NMVD--.......-.---......--...--...-.....-..-..V..........IP.LE...N.....L--.
58 KKA1_ECOLI  .......-......--...-.....--N....D....VTDE..MVR..........LN..WLT..EFMPLP.......T.IKH......FI...RT...P.....D..DawL..........LT.TA...I.....PGK.
59 KDTK_DROME  .......Vwnkkm.--...-.....---....-....TTSQ..ASR..........IFg.---..--IPYN.......S.LLM......YV...RGk..Y.....G..K..S..........LK.LE...Q.....LRK.
60 KPCG_HUMAN  .......P......--...-.....---....-....---T..ADE..........I-..---..-HVTVG.......E.ARN......LI...--...-.....-..-..P..........MD.PN...G.....L--.
```

```
               281       291       301       311       321       331       341       351       361       371       381       391       401       411
Subdomain      -------------------------------------------->.<-VIb---*-*---*------------->..........<-VII--------*---*-------------------------------->.<-V
PROSITE        ...............................................B...BBBBBBBB.B.BB..B........................................................................
X-ray          AAA.........A.......AAA..AAAAA.A..........A...BB...BB.......B.BB..B.B.....B..........B.......BBB.........B....BBB..........B...............
X-ray          EEE.........E.......EEE..EEEEE.E..........E...66...66.......7.77..7.7.....8..........8.......888.........9....999..........9...............
 1 CAPK-ALPHA  YAA.........Q.......IVL..TFEYL.H..........S.L.DL...IYRDLKPE.N.LL..I.D...QQG..........Y.......IQVT.D..FGF.A....KRV..........KG...-...--R.T..
 2 WEE1+       ILV.........E.......VAL..GLQFI.H..........H.K.NY...VHLDLKPA.N.VM..I.T...FEG..........T.......LKIG.D..FGM.A....S--..........VW...P...VPR.G..
 3 TIK         LYE.........Q.......IVT..GVEYI.H..........S.K.GL...IHRDLKPG.N.IF..L.V...DER..........H.......IKIG.D..FGL.A....TAL..........EN...D...GKS.R..
 4 SPK1        ISR.........Q.......ILT..AIKVI.H..........S.M.GI...SHRDLKPD.N.IL..I.E...QDDpv........L.......VKIT.D..FGL.A....K-V..........QG...N...GSF.M..
 5 RSK1-N      YLA.........E.......LAL..GLDHL.H..........S.L.GI...IYRDLKPE.N.IL..L.D...EEG..........H.......IKLT.D..FGL.S....KEA..........ID...H...EKK.A..
 6 PYT         YWK.........N.......MLE..AVHTI.H..........Q.H.GI...VHSDLKPA.N.FL..I.V...-DG..........M.......LKLI.D..FGI.A....NQM..........QP...Dtt.SVV.K..
 7 PKC-ALPHA   YAA.........E.......ISI..GLFFL.H..........K.R.GI...IYRDLKLD.N.VM..L.D...SEG..........H.......IKIA.D..FGM.C....KEH..........MM...D...GVT.T..
 8 PDGFR-B     FSY.........Q.......VAN..GMEFL.A..........S.K.NC...VHRDLAAR.N.VL..I.C...EGK..........L.......VKIC.D..FGL.A....RDI..........MR...Ds..NYI.S..
 9 PBS2        IAN.........A.......VIH..GLKELkE..........Q.H.NI...IHRDVKPT.N.IL..C.Sa..NQG..........T.......VKLC.D..FGV.S....GNL..........VA...-...-SL.A..
10 MIK1        MLF.........Q.......LTQ..ALNFI.H..........L.L.EF...VHLDVKPS.N.VL..I.T...RDG..........N.......LKLG.D..FGL.A....TSL..........PV...-...SSM.V..
11 MCK1        YTY.........Q.......IAR..GMLYL.H..........G.L.GV...CHRDIKPS.N.VL..V.Dp..ETG..........V.......LKIC.D..FGS.A....KKL..........EH...N...QP-.S..
12 INS.R       MAA.........E.......IAD..GMAYL.N..........A.K.KF...VHRDLAAR.N.CM..V.A...HDF..........T.......VKIG.D..FGM.T....RDI..........YEt..D...YYR.K..
13 HSVK        VSR.........Q.......LLS..AVDYI.H..........R.Q.GI...IHRDIKTE.N.IF..I.N...TPE..........D.......ICLG.D..FGA.A....CFV..........QG...Srs.SPF.P..
14 ERK1        FLY.........Q.......ILR..GLKYI.H..........S.A.NV...LHRDLKPS.N.LL..I.N...TTC..........D.......LKIC.D..FGL.A....RIA..........DPeh.Dh..TGF.L..
15 EGFR        WCV.........Q.......IAK..GMNYL.E..........D.R.RL...VHRDLAAR.N.VL..V.K...TPQ..........H.......VKIT.D..FGL.A....KLL..........GA...Eek.EYH.A..
16 ECK         MLR.........G.......IAA..GMKYL.A..........N.M.NY...VHRDLAAR.N.IL..V.N...SNL..........V.......CKVS.D..FGL.S....RVL..........ED...Dp..EATyT..
17 DPYK1       LAL.........D.......IAK..GMNYL.Hgwtpp.....-.-.-I...LHRDLSSRnN.IL..L.D...HNIdpknplvs13K.......CKIS.D..FGL.S....RLK..........KE...Q...ASQ.M..
18 CLK         MAY.........Q.......ICK..SVNFL.H..........S.N.KL...THTDLKPE.N.ILf.V.K...SDYteaynpkm18D.......IKVV.D..FGS.A....T--..........YD...D...E-H.H..
19 CDC2HS      YLY.........Q.......ILQ..GIVFC.H..........S.R.RV...LHRDLKPQ.N.LL..I.D...DKG..........T.......IKLA.D..FGL.A....RAF..........GI...P...IRV.Y..
20 CAMII-ALPHA CIQ.........Q.......ILE..AVLHC.H..........Q.M.GV...VHRDLKPE.N.LL..L.A...SKL..........Kgaa....VKLA.D..FGL.A....IEV..........EG...E...QQA.W..
21 C-SRC       MAA.........Q.......IAS..GMAYV.E..........R.M.NY...VHRDLRAA.N.IL..V.G...ENL..........V.......CKVA.D..FGL.A....RLI..........ED...N...EYT.A..
22 C-RAF       IAR.........Q.......TAQ..GMDYL.H..........A.K.NI...IHRDMKSN.N.IF..L.H...EGL..........T.......VKIG.D..FGL.A....TVK..........SRws.G...SQQ.V..
23 KLSK_HUMAN  MAA.........Q.......IAE..GMAFI.E..........E.R.NY...IHRDLRAA.N.IL..V.S...DTL..........S.......CKIA.D..FGL.A....RLI..........ED...N...EYT.A..
24 KLSK_MOUSE  MAA.........Q.......IAE..GMAFI.E..........E.Q.NY...IHRDLRAA.N.IL..V.S...DTL..........S.......CKIA.D..FGL.A....RLI..........ED...N...EYT.A..
25 ARKB_HUMAN  YAA.........E.......IIL..GLEHM.H..........N.R.FV...VYRDLKPA.N.IL..L.D...EHG..........H.......VRIS.D..LGL.A....CDF..........SK...-...-KK.P..
26 ARKB_BOVIN  YAA.........E.......IIL..GLEHM.H..........N.R.FV...VYRDLKPA.N.IL..L.D...EHG..........H.......VRIS.D..LGL.A....CDF..........SK...-...-KK.P..
27 BYR1_SCHPO  IIN.........S.......MVK..GLIYL.Y..........NvL.HI...IHRDLKPS.N.VV..V.N...SRG..........E.......IKLC.D..FGV.S....GEL..........VN...-...-SV.A..
28 CYGR_ARBPU  LIA.........D.......LVK..GLVYL.H..........S.S.EIk..SHGHLKSS.N.CV..V.D...NRW..........V.......LQIT.D..YGL.H....EFR..........KGqk.E...DVD.L..
29 ANPA_RAT    LTN.........D.......IVK..GMLFL.Hngai......-.-.-C...SHGNLKSS.N.CV..V.D...GRF..........V.......LKIT.D..YGL.E....SFR..........DPe..P...EQG.H..
30 ANPA_HUMAN  LTN.........D.......IVK..GMLFL.Hngai......-.-.-C...SHGNLKSS.N.CV..V.D...GRF..........V.......LKIT.D..YGL.Esf.RDL..........DP...Eq..GHT.V..
31 ANPB_HUMAN  LIN.........D.......LVK..GMAFL.H..........N.S.IIs..SHGSLKSS.N.CV..V.D...SRF..........V.......LKIT.D..YGL.AsfrsTAE..........PD...D...SHA.L..
32 ANPA_MOUSE  LTN.........D.......IVK..GMLFL.Hngai......-.-.-C...SHGNLKSS.N.CV..V.D...GRF..........V.......LKIT.D..YGL.E....SFR..........DPe..P...EQG.H..
33 ANPB_RAT    LIN.........D.......LVK..GMAFL.H..........N.S.IIs..SHGSLKSS.N.CV..V.D...SRF..........V.......LKIT.D..YGL.AsfrsTAE..........PD...D...SHA.L..
34 CYGS_STRPU  LIA.........D.......LVK..GIVYL.H..........S.S.EIk..SHGHLKSS.N.CV..V.D...NRW..........V.......LQIT.D..YGL.N....EFK..........KG...Qkq.DVD.L..
35 VPSF_YEAST  IAF.........Q.......LLN..ALKDI.H..........N.L.NI...VHGDIKTE.N.IL..V.T...SWN..........W.......CILT.D..F--.A....AFI..........KPvylP...EDN.Pge
36 HSER_RAT    VLN.........D.......IAK..GMSYL.H..........S.S.KIe..VHGRLKST.N.CV..V.D...SRM..........V.......VKIT.D..FG-.C....NSI..........LP...-...--P.K..
37 HSER_HUMAN  VLY.........D.......IAK..GMSYL.H..........SsK.TE...VHGRLKST.N.CV..V.D...SRM..........V.......VKIT.D..FG-.C....NSI..........LP...-...--P.K..
38 KR2_VZVD    VFL.........D.......LAQ..ALTFL.N..........RtC.GL...THLDVKCG.N.IF..L.N...VDNfasleit...T.......AVIG.D..YSL.V....T--lntyslct26PR..Da..SQM.S..
39 KR2_HSV11   CFT.........E.......LAR..AVVFL.N..........TtC.GI...SHLDIKCA.N.IL..V.Mlr.SDA..........Vslrr..AVLA.D..FSL.V....T-L..........NS...N...STI.Arg
40 KR1_HSVI1   ILA.........D.......LTG..KLLAL.I..........R.K.GI...YHGDLKSE.NiIM..M.S...RSG..........P.......GKLI.D..F--.-....---ehshgpge14YP...D...RTF.F..
41 KR2_EBV     GFQ.........G.......LKD..AVYFL.N..........R.HcGL...FHSDISPS.N.IL..V.D...FTD..........Tmwgmr.LVLT.D..YGT.A....SLH..........DR...NkmlDVR.L..
42 KRB2_VACCV  -CV.........T.......MIN..TLEFI.H..........S.Q.GF...THGKIEPR.N.IL..I.R...NKR..........L.......S-LI.D..YSR.T....NKL..........YK...S...GNS.Hid
43 KRB2_VACCC  -CV.........T.......MIN..TLEFI.H..........S.R.GF...THGKIEPR.N.IL..I.R...NKR..........L.......S-LI.D..YSR.T....NKL..........YK...S...GNS.Hid
44 AK3_ECOLI   VMRtndrfgraepD.......IA-..ALAEL.A..........A.L.QL...LPR-----.-.--..L.N...EGL..........V.......I---.-..---.-....---tqgfigse42VP...G...IYT.T..
45 PSP_MOUSE   FTS.........-.......-LN..GL---.-..........-.-.--...--------.-.-L..L.K...INN..........-.......LKVL.D..F--qA....K-L..........SS...N...GN-.-..
46 DHOM_BACSU  MM-.........G.......IVN..G----.-..........-.-.--...------TT.N.FI..L.TkmiKEKspyee.....V.......LKEAqD..LGF.A....E-A..........DP...-...--T.S..
47 FLIG_BACSU  YAR.........Q.......VLE..--KAL.G..........Ed-.--...-----KAE.N.ILnrL.T...SSL..........Q.......VK--.-p.FDF.A....RKA..........EP...-...EQI.Lnf
48 CALQ_RABIT  LAA.........Q.......V--..----L.E..........D.K.GVgfg--------.-.-L..V.D...SEK..........D.......AAVA.-kkLGL.T....E--..........ED...S...IYV.F..
49 NU1M_PODAN  YAV.........-.......---..-----.-ipygpsla12S.L.GI...YY------.-.ILa.V.S...SLA..........Tyg.....ILLA.G..WGS.A....NSK..........YA...-...--F.L..
50 RUVA_ECOLI  LAL.........A.......ILS..GMS--.-..........A.Q.QF...V-------.-.--..-.-...---..........-.......----.-..---.-....NAV..........ER...E...E--.V..
51 U15R_HSV6U  VMN.........-yysvfckLAD..AVRFL.N..........L.KcRI...NHFDISPM.N.IF..L.N...HKK..........Eiifd...AVLA.D..YSL.S....E--mhpnyng113YP...-...--F.A..
52 KRF1_VACCC  YIK.........F.......IFLqmALLYI.Kiyelpccd..-.-.NF...LHADLKPD.N.IL..LfD...SNEpiii......H.......LK--.-..---.D....KKF..........VF...N...ERI.K..
53 UL97_HCMVA  CT-.........-.......LAD..AIKFLnH..........Q.C.RV...CHFDITPM.N.VL..I.D...VNP..........HnpseivrAALC.D..YSL.S....E--pypdyner46IC...D...PHA.R..
54 KKA6_ACIBA  IYK.........E.......ALN..LLNSI.Aiidcpfisn.-.-.--...--------.-.--..I.D...HRL..........Ke......----.-..---.S....KFF..........ID...N...Q-L.L..
55 KKA8_ECOLI  TAY.........Q.......VLK..-----.-..........-.-.--...--------.-.--..-.-...---..........-.......---S.D..FG-.A....RLV..........VV...D...ALA.Afm
56 KGPB_BOVIN  ---.........-.......---..-----.-..........-.-.--...--------.-.--..-.-...---..........-.......----.-..---.-....-VI..........RPa..T...QQA.Q..
57 EGFR_CHICK  ---.........Q.......IIR..G----.-..........-.-.--...--------.N.VL..Y.D...NSF..........A.......LAV-.-..--L.S....NYH..........MN...-...--K.T..
58 KKA1_ECOLI  TAF.........Q.......VLE..-----.-..........-.-.--...--------.-.--..-.-...---..........-.......----.-..---.-....-EY..........PDs..G...ENI.V..
59 KDTK_DROME  ---.........-.......---..-----.-..........-.-.--...--------.-.--..-.Dc..ISGppiemlqm28-.......----.-..---.-....KEK..........DK...N...SMS.S..
60 KPCG_HUMAN  ---.........-.......---..-----.-..........-.-.--...--------.-.--..-.-...SDP..........Y.......VKL-.-..---.-....KLI..........PD...P...RNL.T..
```

```
                  421       431       441       451       461       471       481       491       501       511       521       531       541       551
    Subdomain     III--------------------**---><-IX------------------------------*----*---------------------------------------------->...<-X-------------------
    PROSITE       ............................................................................................................................................
    X-ray         ............................................................A.AAAAAAAAAAAAAAA..................................................A....A.......
    X-ray         ............................................................F.FFFFFFFFFFFFFFF..................................................G....G.......
  1 CAPK-ALPHA    ........WTLCGT.P..EY.LAPE..IIL..........SK..........G-YNK...A.VDWWALGVLIYEMAA.G......YP...P..........F.......F....A..-DQP......I....Q.......
  2 WEE1+         ........MEREGD.C..EY.IAPE..VLA..........NH..........L-YDK...P.ADIFSLGITVFEAAAnI......VL...Pdngqswqk17-.......P....R..LSST......D....N.......
  3 TIK           ........TRRTGT.L..QY.MSPE..QLF..........LK..........H-YGK...E.VDIFALGLILAELL-.-......HT...C..........F.......T....E..-SEK......I....K.......
  4 SPK1          ........KTFCGT.L..AY.VAPE..VIR..........GKdtsvsdpe12NEYSS...L.VDMWSMGCLVYVILT.G......HL...P..........F.......S....G..-STQ......D....Q.......
  5 RSK1-N        ........YSFCGT.V..EY.MAPE..VVN..........RQ..........G-HTH...S.ADWWSYGVLM-----.-......--...-..........-.......-....G..KDRK......E....T.......
  6 PYT           ........DSQVGT.V..NY.MPPE..AIKdmsssrengkSK..........SKISP...K.SDVWSLGCILYYMTY.G......KT...P..........F.......Q....Q..IINQi.....S....K.......
  7 PKC-ALPHA     ........RTFCGT.P..DY.IAPE..IIA..........YQ..........P-YGK...S.VDWWAYGVLLYEMLA.G......QP...P..........F.......D....G..-EDE......D....E.......
  8 PDGFR-B       ........KGSTFL.P..LK.WMAP..ESI..........FN..........SLYTT...L.SDVWSFGILLWEIFTlG......GT...P..........Y.......P....E..LPMN......E....Q.......
  9 PBS2          ........KTNIGC.Q..SY.MAPE..RIK..........SLnpdr......ATYTV...Q.SDIWSLGLSILEMAL.G......RY...P..........Y.......P....P..-ETY......DnifsQ.......
 10 MIK1          ........D-LEGD.R..VY.IAPE..ILA..........SH..........N-YGK...P.ADVYSLGLSMIEAATnV......VL...Pengvewqr16L.......P....N..LKDL......L....L.......
 11 MCK1          ........ISYICS.R..FY.RAPE..LII..........GC..........TQYTT...Q.IDIWGLGCVMGEMLI.G......KA...I..........F.......Q....G..QEPL......L....Q.......
 12 INS.R         ........GGKGLL.P..VR.WMAP..ESL..........KD..........GVFTT...S.SDMWSFGVVLWEITSlA......EQ...P..........Y.......Q....G..LSNE......Q....V.......
 13 HSVK          ........YGIAGT.I..DT.NAPE..VLA..........GD..........P-YTT...T.VDIWSAGLVIFETAVhN......AS...L..........FsaprgpkR....G..-PCD......S....Q.......
 14 ERK1          ........TEYVAT.R..WY.RAPE..IML..........NS..........KGYTK...S.IDIWSVGCILAEMLS.N......RP...I..........F.......P....G..KHYL......D....Q.......
 15 EGFR          ........E-GGKV.P..IK.WMAL..ESI..........LH..........RIYTH...Q.SDVWSYGVTVWELMTfG......SK...P..........Y.......D....G..IPAS......E....I.......
 16 ECK           ........TSGGKI.P..IR.WTAP..EAI..........SY..........RKFTS...A.SDVWSFGIVMWEVMTyG......ER...P..........Y.......W....E..LSNH......E....V.......
 17 DPYK1         ........TQSVGC.I..PY.MAPE..VFK..........GD..........S-NSE...K.SDVYSYGMVLFELLT.S......DE...P..........Q.......Q....D..MKPM......K....M.......
 18 CLK           ........STLVST.R..HY.RAPE..VIL..........AL..........G-WSQ...P.CDVWSIGCILIEYYL.G......FT...V..........F.......S....T..HDSR......E....H.......
 19 CDC2HS        ........THEVVT.L..WY.RSPE..VLL..........GS..........ARYST...P.VDIWSIGTIFAELAT.K......KP...L..........F.......H....G..DSEI......D....Q.......
 20 CAMII-ALPHA   ........FGFAGT.P..GY.LSPE..VLR..........KD..........P-YGK...P.VDLWACGVILYILLV.G......YP...P..........F.......W....D..-EDQ......H....R.......
 21 C-SRC         ........RQGAKF.P..IK.WTAP..EAA..........LY..........GRFTI...K.SDVWSFGILLTELTTkG......RV...P..........Y.......P....G..MVNR......E....V.......
 22 C-RAF         ........EQPTGS.V..LW.MAPE..VIR..........MQdnn.......P-FSF...Q.SDVYSYGIVLYELMT.G......EL...P..........Y.......S....H..INNR......D....Q.......
 23 KLSK_HUMAN    ........REGAKF.P..IK.WTAP..EAI..........NY..........GTFTI...K.SDVWSFGILLTEIVThG......RI...P..........Y.......P....G..MTNP......E....V.......
 24 KLSK_MOUSE    ........REGAKF.P..IK.WTAP..EAI..........NY..........GTFTI...K.SDVWSFGILLTEIVThG......RI...P..........Y.......P....G..MTNP......E....V.......
 25 ARKB_HUMAN    ........HASVGT.H..GY.MAPE..VLQ..........KG..........VAYDS...S.ADWFSLGCMLFKLLR.G......HS...P..........F.......R....Q..HKTK......Dkh..E.......
 26 ARKB_BOVIN    ........HASVGT.H..GY.MAPE..VLQ..........KG..........VAYDS...S.ADWFSLGCMLFKLLR.G......HS...P..........F.......R....Q..HKTK......Dkh..E.......
 27 BYR1_SCHPO    ........QTFVGT.S..TY.MSPE..RIR..........GG..........K-YTV...K.SDIWSLGISIIELAT.Q......EL...Pws........F.......S....N..IDDSigil..D....L.......
 28 CYGR_ARBPU    ........GEHAKL.A..RKlWTAP..EHL..........REgksmhp....G-GTP...K.GDIYSFSIILTEMYS.R......QE...P..........F.......Hen..D..LELA......D....I.......
 29 ANPA_RAT      ........TLFAKK.L..--.WTAP..ELLrmaspp....AR..........G--SQ...A.GDVYSFGIILQEIALrS......GV...F..........Yveg....L....D..LSPK......E....I.......
 30 ANPA_HUMAN    ........Y-AKKL.-..--.WTAP..ELL..........RMasppv.....R-GSQ...A.GDVYSFGIILQEIAL.R......SG...V..........F.......HveglD..LSPK......E....I.......
 31 ANPB_HUMAN    ........Y--AKK.L..WT.-APE..LLS..........GN..........PLPTTgmqK.ADVYSFGIILQEIAL.R......SG...P..........F.......YleglD..LSPK......E....I.......
 32 ANPA_MOUSE    ........TLFAKK.L..--.WTAP..ELLrmaspp....AR..........G--SQ...A.GDVYSFGIILQEIALrS......GV...F..........Yveg....L....D..LSPK......E....I.......
 33 ANPB_RAT      ........Y--AKK.L..WT.-APE..LLS..........GN..........PLPTTgmqK.ADVYSFAIILQEIAL.R......SG...P..........F.......YleglD..LSPK......E....I.......
 34 CYGS_STRPU    ........GDHAKLaR..QL.WTSP..EHL..........RQegsmpta...G--SP...Q.GDIYSFAIILTELYS.R......QE...P..........F.......H....EneMDLA......D....I.......
 35 VPSF_YEAST    flfyfd..TSKRRT.-..CY.LAPE..RFN..........SKlyqdgksnn.GRLTK...E.MDIFSLGCVIAEIFAeG......RP...I..........F.......-....-..--NL......S....Q.......
 36 HSER_RAT      ........KDL---.-..--.WTAP..EHL..........RQ..........ATISQ...K.GELYSFSIIAQEIIL.R......KE...T..........F.......Y....T..LSCR......D....Qnek....
 37 HSER_HUMAN    ........KDL---.-..--.WTAP..EHL..........RQ..........ANISQ...K.GDVYSYGIIAQEIIL.R......KE...T..........F.......Y....T..LSCRdrn...E....K.......
 38 KR2_VZVD      ........FRLVLS.H..GT.NQPP..EIL..........LDyingtglt15QRVGL...A.IDLYALGQALLEVILlG......RL...Pgqlpisvh14Y.......Y....G..HKLS......P....D.......
 39 KR2_HSV11     qfclqe28HTLVG-.H..GY.NQPP..ELLvkylnner15LK..........HDVGL...A.VDLYALGQTLLELVVsV......YV...Apslgvpvtr.F.......P....G..----......-....-.......
 40 KR1_HSVI1     ........WNPIGT.E..AY.ASPE..RSR..........DRvpdrpdsa12GTHGA...G.I--------------.-......RE...P..........Hli.....K....G..DGYR......A....H.......
 41 KR2_EBV       ........KSSKGR.Q..LY.R--L..YCQ..........RE..........P-FSI...A.KDTY-----------.-......KP...Lcllskcyi24-.......-....G..AQTA......L....R.......
 42 KRB2_VACCV    ynedmi19NHLGAT.V..SR.RGDL..EML..........GY..........C----...-.---------MIEWFG.G......KL...P..........W.......-....-..-KNE......S....S.......
 43 KRB2_VACCC    ynedmi19NHLGAT.V..SR.RGDL..EML..........GY..........C----...-.---------MIEWFG.G......KL...P..........W.......-....-..-KNE......S....S.......
 44 AK3_ECOLI     ........DPRVVS.A..AK.RIDE..IAF..........AE..........A----...-.----------AEMATfGakvlhpAT...L..........L.......P....A..-VRS......D....I.......
 45 PSP_MOUSE     ........--GIDL.T..VP.LAGE..ASL..........VL..........PFIGK...T.VDI-SVSLDLINSLS.I......KT...Naqtglpev14-.......-....-..-SNT......D....K.......
 46 DHOM_BACSU    ........DVEGLD.A..AR.KMA-..ILA..........RL..........G-FSM...N.VDLE-----------.-......--...-..........-dvkvkgiS....Q..ITDE......D....I.......
 47 FLIG_BACSU    iqqehpq.T-----.-..--.-MAL..ILSy.........LD..........PVQ--...-.-----AGQILSELN-.-......--...-..........-.......P....E..-VQA......E....V.......
 48 CALQ_RABIT    ........KED---.-..--.---E..VIE..........YD..........GEFSAd..-.--------TLVEFL-.-......--...-..........-.......-....-..LDVL......E....D.......
 49 NU1M_PODAN    ........GSLRST.A..QL.ISYE..LVL..........SS..........A----...-.-------ILLVIMLT.G......SL...Nlsvniesq14F.......P....L..-LPV......F....I.......
 50 RUVA_ECOLI    ........GALVKL.PgiGK.KTAE..RLIvemkdrfk11GD..........L-FTP...A.ADL--------VLTSpA......SP...A..........T.......D....D..AEQE......A....V.......
 51 U15R_HSV6U    ........YRDACC.-..-K.VLAEhvVLL..........GL..........L----...-.---------------.-......--...-..........F.......Y....R..-DVV......E....I.......
 52 KRF1_VACCC    ........S-----.-..--.----..--A..........LN..........D-FDF...S.---------------.-......--...-..........-.......-....-..-QVA......G....I.......
 53 UL97_HCMVA    ........FPVAGL.R..RY.CMSE..LSA..........LG..........NVLGF...-.-------CLM-----.-......--...-..........-.......-....R..LLDR......R....G.......
 54 KKA6_ACIBA    ........DD----.-..--.----..--I..........DQ..........DDFDT...E.---------------.-......--...L..........W.......G....D..HKTY......LslwnE.......
 55 KKA8_ECOLI    rrlhai18TTHAGL.P..ER.GSIE..AGVvdvddfdke.RE..........G-WTA...Eq---------VWEAMH.R......LL...P..........L.......A....P..DPVV......T....H.......
 56 KGPB_BOVIN    ........KQSAST.-..--.----..-LQ..........GE..........P-RTK...R.---------------.-......--...-..........-.......Q....A..ISAE......P....T.......
 57 EGFR_CHICK    ........QGLREL.P..MK.RLSE..ILN..........GG..........VKIS-...-.---------------.-......--...-..........-.......-....-..-NNPklcnmdT....V.......
 58 KKA1_ECOLI    ........DALAVF.L..RR.L---..---..........--..........--HSI...P.V--------------.-......--cncP..........F.......N....S..-DRV......F....R.......
 59 KDTK_DROME    ........NGSGGS.-..--.----..-AN..........SQ..........G----...-.---------------.-......GA...P..........T.......S....G..-SGP......M....Qhsgelgp
 60 KPCG_HUMAN    ........KQKTRT.-..--.----..-VK..........AT..........L----...-.---------------.-......--...-..........-.......-....-..-NPVwn....E....T.......
```

```
                561       571       581       591       601       611       621       631       641       651       661       671
   Subdomain    ------------------------------------><-X1------------------------------------*--------------------------->...........
   PROSITE      ....................................................................................................................
   X-ray        A.AAAAA...A...............................AA........AA......A..AAAAA.......................AAAAAA...................
   X-ray        G.GGGGG...G...............................HH........HH......H..HHHHH.......................IIIIII...................
 1 CAPK-ALPHA   I.YEKIV...S...GK.V.RF..........P....SH.F.SSD........LK......D..LLRNL.LQVD.LTKRF.Gnlkng.....VNDIKNHK....WF-..........
 2 WEE1+        G.SSLTSs..S...RE.T.PA..........N....SI.I.GQG........GL......Dr.VVEWM.LSPE.PRNRP.T..........IDQILATD....EV-cw........
 3 TIK          F.FESLR...K...GD.F.SN..........-....DI.F.DNK........EK......S..LLKKL.LSEK.PKDRP.E..........TSEILKT-....-L-aewrnise18
 4 SPK1         L.YKQIGr..GsyhEG.P.LK..........D....FR.I.SEE........AR......D..FIDSL.LQVD.PNNRS.T..........AAKALNHP....WI-..........
 5 RSK1-N       M.TLILKa..K...LG.M.--..........P....QF.L.STE........AQ......S..LLRAL.FKRN.PANRL.Gsgpdg.....AEEIKRH-....-I-fystidwn20
 6 PYT          L.HAIIDp..N...HE.I.EF..........P....DI.P.EKD........LQ......D..VVKCC.LKRD.PKQRI.S..........IPELLAHP....YV-..........
 7 PKC-ALPHA    L.FQSIM...E...HN.V.SY..........P....KS.L.SKE........AV......S..ICKGL.MTKH.PAKRL.Gcgpeg.....ERDVREHA....FF-..........
 8 PDGFR-B      F.YNAIKr..G...YR.M.AQ..........P....AH.A.SDE........IY......E..IMQKC.WEEK.FEIRP.P..........FSQLVLL-....-L-..........
 9 PBS2         L.SAIVD...G...PP.P.RL..........P....SDkF.SSD........AQ......D..FVSLC.LQKI.PERRP.T..........YAALTEHP....WL-..........
10 MIK1         S.KEKVQ...I...NK.V.R-..........-....--.C.AES........LQ......C..LLQRM.THPY.VDCRP.T..........TQDLLAMP....EM-if........
11 MCK1         L.REIAK...L...LG.P.PDkrfiffsn37-....--.-.-PD........GI......D..LLMKI.LVYE.PQQRL.S..........PRRILAHQ....FF-nelrnddt11
12 INS.R        L.KFVMD...G...GY.L.DQ..........P....DN.C.PER........VT......D..LMRMC.WQFN.PNMRP.T..........FLEIVNL-....-L-..........
13 HSVK         I.TRIIRqaqV...HV.D.EFsphpesrl35-....--.-.DID........VE......Y..LVCKA.LTFD.GALRP.S..........AAELLCLP....LF-..........
14 ERK1         L.NHILG...I...LG.S.PSqedlncii31-....PK.S.DSK........AL......D..LLDRM.LTFN.PNKRI.T..........VEEALAHP....YL-eqyydptd49
15 EGFR         S.SILEK...G...ER.L.PQ..........P....PI.C.TID........VY......M..IMVKC.WMID.ADSRP.K..........FRELIIE-....-F-..........
16 ECK          M.KAIND...G...FR.L.PT..........P....MD.C.PSA........IY......Q..LMMQC.WQQE.RARRP.K..........FADIVSI-....-L-..........
17 DPYK1        A.HLAATe..S...YR.P.PI..........P....LT.T.SSK........WK......E..ILTQC.WDSN.PDSRP.T..........FKQIIVH-....-L-kemedqgv..
18 CLK          L.AMMER...I...LG.P.LPkhmiqktr48-....--.-.-EL........LF......D..LIGKM.LEYD.PAKRI.T..........LKEALKHP....FF-yplkkht...
19 CDC2HS       L.FRIFR...A...LG.T.PNnevwpeve29-....KN.L.DEN........GL......D..LLSKM.LIYD.PAKRI.S..........GKMALNHP....YF-..........
20 CAMII-ALPHA  L.YQQIK...A...GA.Y.DF..........PspewDT.V.TPE........AK......D..LINKM.LTIN.PSKRI.T..........AAEALKHP....WI-..........
21 C-SRC        L.DQVER...G...YR.M.PC..........P....PE.C.PES........LH......D..LMCQC.WRKE.PEERP.T..........FEYLQAF-....-L-..........
22 C-RAF        IiFMVGR...G...YA.S.PDlskly.....-....KN.C.PKA........MK......R..LVADC.VKKV.KEERP.L..........FPQILSS-....-I-..........
23 KLSK_HUMAN   I.QNLER...G...YR.M.VR..........P....DN.C.PEE........LY......Q..LMRLC.WKER.PEDRP.T..........FDYLRSV-....-L-edfftate15
24 KLSK_MOUSE   I.QNLER...G...YR.M.VR..........P....DN.C.PEE........LY......H..LMMLC.WKER.PEDRP.T..........FDYLRSV-....-L-ddfftate15
25 ARKB_HUMAN   I.DRMTL...T...MA.V.EL..........P....DS.F.SPE........LH......S..LLEGL.LQRD.VNRRL.Gclgrg.....AQEVKESP....FF-rsldwqm236
26 ARKB_BOVIN   I.DRMTL...T...MA.V.EL..........P....DS.F.SPE........LR......S..LLEGL.LQRD.VNRRL.Gclgrg.....AQEVKESP....FF-rsldwqm236
27 BYR1_SCHPO   L.HCIVQ...E...EP.P.RL..........P....SS.F.PED........LR......L..FVDAC.LHKD.PTLRA.S..........PQQLCAMP....YF-qqalminv20
28 CYGR_ARBPU   I.ARVSK...G...EV.P.PYrpvlnavn..-....EA.A.PDC........VL......T..AIRAC.WVED.PMERP.N..........IIEVRTM-....-L-aplqkgl150
29 ANPA_RAT     I.ERVTR...G...EQ.P.PFrpsmdlqshl-....--.-.-EE........LG......Q..LMQRC.WAED.PQERP.P..........FQQIRLA-....-L-rkfnken260
30 ANPA_HUMAN   I.ERVTR...G...EQ.P.PFrpslalqshl-....--.-.-EE........LG......L..LMQRC.WAED.PQERP.P..........FQQIRLT-....-L-rkfnren260
31 ANPB_HUMAN   V.QKVRN...G...QR.P.YFrpsidrtq..-....--.L.NEE........LV......L..LMERC.WAQD.PAERP.D..........FGQIKGF-....-I-rrfnkeg261
32 ANPA_MOUSE   I.ERVTR...G...EQ.P.PFrpsmdlqshl-....--.-.-EE........LG......Q..LMQRC.WAED.PQERP.P..........FQQIRLA-....-L-rkfnken260
33 ANPB_RAT     V.QKVRN...G...QR.P.YFrpsidrtq..-....--.L.NEE........LV......L..LMERC.WAQD.PTERP.D..........FGQIKGF-....-I-rrfnkeg261
34 CYGS_STRPU   I.GRVKS...G...EV.P.PYrpilnavn..-....AA.A.PDC........VL......S..AIRAC.WPED.PADRP.N..........IMAVRTM-....-L-aplqkgl286
35 VPSF_YEAST   L.FKYKSn..S...YD.V.NReflmeemn..-....--.-.STD........LR......N..LVLDM.IQLD.PSKRL.S..........CDELLNK-yrgiFF-pdyfyt1155
36 HSER_RAT     I.FRVEN...S...YG.T.KPfrpdlfle15-....--.-.--E........VY......L..LVKSC.WEED.PEKRP.D..........FKKIEST-....-L-akifglf328
37 HSER_HUMAN   I.FRVEN...S...NG.M.KPfrpdlfle15-....--.-.--E........VY......L..LVKNC.WEED.PEKRP.D..........FKKIETT-....-L-akifglf328
38 KR2_VZVD     L.ALDTL...A...YRcV.LA..........PyilpSD.I.PGD........LNynpfih-..-----.-AGE.LNTRI.S..........RNSLRRI-....-F-qchavryg58
39 KR2_HSV11    -.YQYFN...N...QL.S.PDfalallay66-....VA.L.PPE........LKpll...V..LVSRL.CHTN.PC---.-..........---ARHA-....-L-s.........
40 KR1_HSVI1    V.LKVIK...A...RG.T.LDlrggartw11-....--.-.-DE........LI......G..LVARC.LERD.PAMRP.S..........LETLVDE-....-F-ski.......
41 KR2_EBV      L.DLQSL...G...YS.L.LYgimhlads38-....--.-.---........LL......E..VLSQM.WNLN.LDMGL.Tscgespcv49VAELLADD....FF-gpdgrrg...
42 KRB2_VACCV   I.KVIKQ...K...KE.Y.KKfiatffed15-....--.-.PLE........LV......R..YIELV.YTLD.YSQTP.N..........YDRLRKL-....-F-iqd.......
43 KRB2_VACCC   I.KVIKQ...K...KE.Y.KKfiatffed15-....--.-.PLE........LV......R..YIELV.YTLD.YSQTP.N..........YDRLRRL-....-F-iqd.......
44 AK3_ECOLI    P.VFVGS...S...KD.P.RAggtlvcnk79-....TL.L.TQS........LL......M..ELSAL.CRVE.VEE--.-..........GLALVAL-....-I-gndlskac57
45 PSP_MOUSE    I.SISLL...G...RR.L.PIinsildgvs.-....TL.L.TST........LS......T..VLQNF.LC--.----P.L..........LQYVLST-....-L-npsvlqgl21
46 DHOM_BACSU   S.FSKRL...G...YT.M.KLigiaqrdg56-....--.-.-PT........AT......S..VVSDL.VAVM.KNMRL.Gvtgnsfvg17PSDIYAQQ....FL-rihvkdev82
47 FLIG_BACSU   A.RRIAV...M...DR.T.--..........-....--.-.SPE........II......N..EVERI.LEQK.LSSAF.Tqdytqtgg..IEAVVEV-....-L-ngvdrgt130
48 CALQ_RABIT   P.VELIE...G...EReL.QA..........F....EN.I.EDE........IK......-..LIGYF.KNKD.SEHYK.A..........FKEAAEE-....-F-hpyipff197
49 NU1M_PODAN   I.FFIGS...V...AE.TnRA..........P....FD.L.AEA........ES......E..LVSGF.MTEH.AA---.-..........---VVFV-....-F-fflaeyg137
50 RUVA_ECOLI   A.RLVAL...G...YK.-.--..........-....--.-.PQE........AS......R..MVSKI.ARPD.AS---.S..........ETLIREA-....-L-raal......
51 U15R_HSV6U   Y.EKLY-...D...FL.D.ER..........G....EF.G.SRD........LF......EatFLNN-.--SK.LTRRQ.P..........IREGLAS-....-L-qsseygek35
52 KRF1_VACCC   I.NKKIKn..N...FK.V.KHnwyy......-....--.-.--D........FH......F..FVHTL.LKTY.PEIEK.Die........FSTALEE-....-F-imctktdc41
53 UL97_HCMVA   L.DEVRM...G...TE.A.LL..........F....KH.A.GAA........CR......A..LENGK.LTH-.-----.-..........CSDACLL-....-I-laaqmsyg97
54 KKA6_ACIBA   L.TETRV...E...ER.L.VFshgditds25-....AG.L.ADE........FV......DisFVERC.LRED.ASE--eT..........AKIFLKH-....-L-kndrpdkr18
55 KKA8_ECOLI   G.DFSLD...N...LL.I.VEgkvvgcid28-....EE.F.EPS........LQe.....R..LVAQY.GIAD.PDRRK.Lq.........FHLLLDE-....-L-f.........
56 KGPB_BOVIN   A.FDIQD...L...SH.V.TL..........P....FY.PkSPQ........SK......D..LIKEAiLDND.FMKNL.E..........LSQIQEI-....-V-dcmypve162
57 EGFR_CHICK   LwNDIIDt..S...RK.P.LTvldfasn193-....SI.L.PVA........FLg.....D..AFTKT.LPLD.PKK-L.Dvfrt......VKEISGF-....-L-liqawpd291
58 KKA1_ECOLI   L.AQAQSrm.N...NG.L.VD..........A....SD.F.DDErngwpveqVW......K..EMHKL.LPFS.PDS--.-..........---VVTHG....DF-sldnlifd72
59 KDTK_DROME   M.GQLDL...D...LG.L.PLgppggprs39-....PG.L.PLS........ML......N..LL---.----pPAERH.H..........AAAAMHH-....-L-gvrwmrta37
60 KPCG_HUMAN   F.VFNLK...P...GD.V.E-..........-....RR.L.SVE........V-......-..-----.WDWD.RTSRN.Dfmgamsfg..VSELLKAP....VD-gwykllnq46
```

B) Details on sequences 23-60 shown in the alignment (arranged in order of decreasing Z-score). **NLL-score** and **Z-score** are measures of how well the kinase HMM fits these SWISS-PROT 22 "test" sequence that were not present in the training set (see Section 3.2 for more details). In **HMM**, **PROFILESEARCH** and **Keyword**, "+" denotes sequences that are classified as containing a kinase domain and "-" those that do not. For PROFILESEARCH, "-$" identifies sequences that do not appear in the results obtained from searching SWISS-PROT 25 (not 22 as in HMM, Keyword and PROSITE) provided to us by M. Gribskov (personal communication).

Two PROSITE signature patterns for eucaryotic protein kinases have been derived and these are labelled "A" and "B" in the alignment. "A" is the region believed to be involved in ATP binding (PROSITE entry PROTEIN_KINASE_ATP) while "B1" and "B2" indicate the area important for catalytic activity in serine/threonine kinases (PROTEIN_KINASE_ST) and tyrosine kinases (PROTEIN_KINASE_TYR) respectively. In all instances, "T" signifies a true positive; "N" a false negative (a sequence which belongs to the set under consideration but which is not picked up by the pattern); "P" a "potential" hit (a sequence that belongs to the set but which is not picked up because the region that contains the pattern is not yet available in the data bank i.e. a partial sequence); and "?" an unknown (a sequence which possibly could belong to the set). "*" indicates SWISS-PROT files which contain a cross reference to the specified PROSITE pattern, but these PROSITE entries do not contain a corresponding pointer to the SWISS-PROT file. "-" signifies sequences that do not satisfy the kinase patterns and those followed by "%" denote particulate forms of guanylyl cyclase receptors which contain an intracellular protein kinase-like domain but which have not been shown to possess kinase activity to date (reviewed in (Garbers, 1992)).

ID	Length	NLL-score	Z-score	HMM	PROFILE-	Keyword	PROSITE		
					SEARCH		A	B1	B2
23 KLSK_HUMAN	509	1188.032	48.056	+	-	+	T	-	T
24 KLSK_MOUSE	509	1193.879	47.376	+	-	+	T	-	T
25 ARKB_HUMAN	689	1826.919	31.781	+	-	+	*	*	-
26 ARKB_BOVIN	689	1827.514	31.720	+	-	+	*	*	-
27 BYR1_SCHPO	340	808.153	27.540	+	+	-	N	T	-
28 CYGR_ARBPU	986	2839.392	22.121	+	+	-	%	-	-
29 ANPA_RAT	1057	3062.107	21.418	+	+	-	%	-	-
30 ANPA_HUMAN	1061	3072.615	21.390	+	+	-	%	-	-
31 NPB_HUMAN	1047	3033.232	21.220	+	+	-	%	-	-
32 ANPA_MOUSE	1057	3065.181	21.042	+	+	-	%	-	-
33 ANPB_RAT	1047	3038.053	20.633	+	+	-	%	-	-
34 CYGS_STRPU	1125	3277.621	18.745	+	+	-	%	-	-
35 VPSF_YEAST	1454	4263.173	17.896	+	-	+	N	T	-
36 HSER_RAT	1075	3143.529	17.681	+	-	-	%	-	-
37 HSER_HUMAN	1073	3139.039	17.552	+	-	-	%	-	-
38 KR2_VZVD	510	1521.597	9.615	+	-	+	N	T	-
39 KR2_HSV11	518	1548.949	9.042	+	-	+	N	-	-
40 KR1_HSVI1	230	710.448	6.773	+	-	+	N	T	-
41 KR2_EBV	455	1393.761	4.935	-	-	+	T	-	T
42 KRB2_VACCV	283	880.650	4.848	-	+	+	N	N	-
43 KRB2_VACCC	283	880.753	4.838	-	+	+	N	N	-
44 AK3_ECOLI	449	1385.412	3.900	-	-	-	-	-	-
45 PSP_MOUSE	235	754.545	3.804	-	-	-	-	-	-
46 DHOM_BACSU	433	1340.413	3.706	-	-	-	-	-	-
47 FLIG_BACSU	338	1055.096	3.699	-	-	-	-	-	-
48 CALQ_RABIT	395	1229.120	3.487	-	-	-	-	-	-
49 NU1M_PODAN	368	1149.759	3.415	-	-	-	-	-	-
50 RUVA_ECOLI	203	667.519	3.413	-	-	-	-	-	-
51 U15R_HSV6U	562	1728.770	3.171	-	-	+	T	-	T
52 KRF1_VACCC	439	1366.011	2.900	-	-	+	N	T	-
53 UL97_HCMVA	707	2165.296	2.854	-	-	+	N	-	T
54 KKA6_ACIBA	259	838.469	2.370	-	-	-	-	-	T
55 KKA8_ECOLI	271	885.548	1.182	-	-	-	-	-	T
56 KGPB_BOVIN	293	953.735	0.684	-	-	+	P	P	-
57 EGFR_CHICK	703	2179.703	0.065	-	-	+	P	-	P
58 KKA1_ECOLI	271	902.461	-0.467	-	-	-	-	T	-
59 KDTK_DROME	753	2334.760	-0.523	-	-	+	N	-	N
60 KPCG_HUMAN	318	1051.016	-1.486	-	-	+	P	P	-

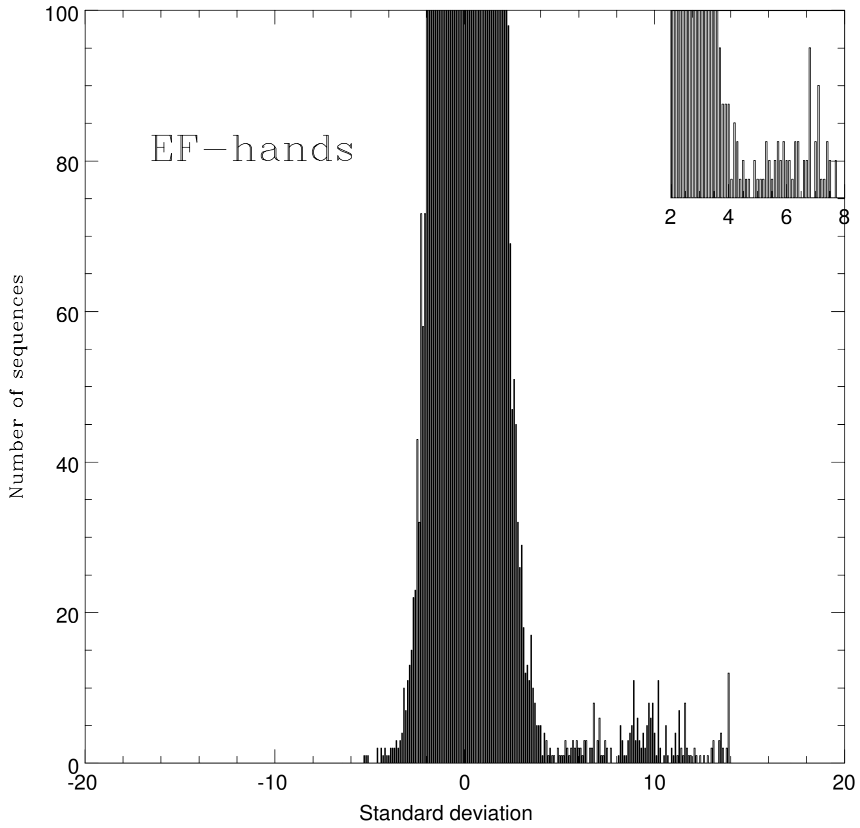


Figure 12: Histogram showing the number of sequences with a certain Z-score relative to the EF-hand model.

Figure 13:

A) Multiple sequence alignment generated by our EF-hand HMM of some of the sequences used to train the HMM (1-27) and "test" sequences from the SWISS-PROT 22 database (28-137) (see Section 3.3).

In **Structure**, "H" and "L" denote residues in an α-helical or loop conformation based upon EF-hands of known structure (Nakayama *et al.*, 1992). **PROSITE** denotes the positions used to generate the pattern "EF-HAND". **Ca-binding** identifies the EF-hand motif sequence preferences at different positions for *most* domains known to bind calcium: "E", glu; "n", hydrophobic; "*", variable; "G", gly; "I", ile, leu or val (Nakayama *et al.*, 1992). The six residues involved in octahedrally coordinating the calcium ion are denoted by "X", "Y", "Z", "x", "z" and "y" and the first five are usually Asx (D or N), Glx (E or Q), Ser (T), Thr (T), Cys (C) or Gly. The oxygen atom at position "y" comes from the main chain and so can be supplied by any amino acid.

Sequences 1-27 are representatives of the various EF-hand subgroups in the June 1992 database of EF-hand sequences maintained by Kretsinger and co-workers (Nakayama *et al.*, 1992). These sequences are: **CAMHS**, *Homo sapiens* calmodulin; **aACTGG**, *Gallus gallus* α-actinin; **VISININ**, *G. gallus* visinin; **TPP24CF**, *Canis familiaris* p24 thyroid protein; **TPHUCS**, *H. sapiens* skeletal troponin-C; **TPAP1**, *Astacus pontasticus* troponin-C-1; **TCBP25**, *Tetrahymena thermophila* TCBP-25; **SPEC2A**, *Strongylocentrotus purpuratus* spec2a; **SCBPBL1**, *Branchiostoma lanceolatum* SARC1; **QUIDLN**, *Loligo pealei* squidulin; **MOHSCR**, *H. sapiens* myosin (RLC-ventricle); **MOHSA1**, *H. sapiens* myosin (ELC-L1-skeletal); **LPS1A**, *Lytechinus pictus* α-Lps1; **LAV1**, *Physarum polycephalum* LAV1-2; **EFH5**, *Trypanosoma brucei* putative calcium binding protein; **CVP**, *B. lanceolatum* calcium vector protein; **CRGHS**, *H. sapiens* calmodulin-related gene; **CMSE**, *Saccharopolyspora erythraea* bacterial-CAM; **CDPK**, *Glycine max* calcium dependent protein kinase; **CDC31**, *Saccharomyces cerevisiae* cell division control protein 31; **CALPLHS**, *H. sapiens* calpain (light); **CALCIB**, *Bos taurus* calcineurin-B; **CALBNGG**, *G. gallus* calbindin; **CAL1CE**, *Caenorhabditis elegans* cal 1 gene; **BCHS**, *H. sapiens* β S-100 protein; **AEQAV1**, *Aequorea victoria* aequorin-1; and **1F8**, *Trypanosoma cruzi* flagellar calcium binding protein.

28-96 are the SWISS-PROT 22 sequences designated as EF-hands by our HMM (Z-score > 4.75) but not by all three other methods, PROSITE, PROFILESEARCH and the keyword search Note that sequences identified as EF-hands by all four methods are not shown. 97-116 are the top 20 sequences below our cutoff of 4.75; 117-137 are sequences that were not classified as EF-hands by the HMM but were so by one or more (but not all) of the three other methods.

```
                 1         11        21        31        41        51        61        71
   Structure     ..........H..H.HHHHH.H.H..........H....LL...LLL.LLLL....H.HH.HHHHHHH...........
   PROSITE       ..................................*....**...***.****....*.**...................
   Ca-binding    ..................................X.....Y....Z..y.x.......z....................
 1 CAMHS         ..........E..F.KEAFS.L.F..........D....KD...GDG.TITT....K.EL.GTVMRSL-..........
 2 aACTGG        ..........E..F.RASFN.H.F..........D....RK...KTG.MMDC....E.DF.RACLISM-..........
 3 VISININ       ..........E..L.SRWYE.G.F..........Qr...QC...SDG.RIRC....D.EF.ERIYGNF-..........
 4 TPP24CF       ..........G..L.ARFFR.R.L..........D....RD...RSR.SLDS....R.EL.QRGLAEL-..........
 5 TPHUCS        ..........E..F.KAAFD.M.F..........D....AD...GGG.DISV....K.EL.GTVMRML-..........
 6 TPAP1         ..........A..L.QKAFD.S.F..........D....TD...SKG.FITP....E.TV.GIILRMM-..........
 7 TCBP25        ..........V..A.RRIFE.N.Y..........D....KG...RKG.RIEN....T.DC.VPMITEA-..........
 8 SPEC2A        ..........L..F.KSSFK.S.E..........D....TD...GDG.KITS....E.EL.RAAFKSI-..........
 9 SCBPBL1       ..........K..I.KFTFD.F.Fl.........D....YN...KDG.SIQW....E.DF.EEMIKRY-..........
10 QUIDLN        ..........E..I.KDAFD.M.F..........D....ID...GDG.QITS....K.EL.RSVMKSL-..........
11 MOHSCR        ..........E..F.KEAFT.I.M..........D....QN...RDG.FIDK....N.DL.RDTFAAL-..........
12 MOHSA1        ..........E..F.KEAFL.L.F..........D....ST...GDS.KIIL....S.QV.GDVLRAL-..........
13 LPS1A         e.........A..L.KQEFK.DnY..........D....TN...KDG.TVSC....A.EL.VKLMNWT-..........
14 LAV1          ..........A..L.VADFR.K.I..........D....TN...SNG.TLSR....K.EF.REHFVRL-..........
15 EFH5          ..........E..L.AEGFR.V.L..........-....SN...GQK.TISIpm..K.EV.SALMASV-..........
16 CVP           ..........E..C.MKIFD.I.F..........D....RN...AEN.IAPV....S.DT.MDMLTKL-..........
17 CRGHS         l.........Q..L.-HYFK.M.H..........D....YD...GNN.LLDG....L.EL.STAITHV-..........
18 CMSE          ..........R..L.KKRFD.R.W..........D....FD...GNG.ALER....A.DF.EKEAQHI-..........
19 CDPK          ..........G..L.KELFK.M.I..........D....TD...NSG.TITF....D.EL.KDGLKRV-..........
20 CDC31         ..........E..I.YEAFS.L.F..........D....MN...NDG.FLDY....H.EL.KVAMKAL-..........
21 CALPLHS       ..........T..C.RSMVA.V.M..........D....SD...TTG.KLGF....E.EF.KYLWNNI-..........
22 CALCIB        ..........R..L.GKRFK.K.L..........D....LD...NSG.SLSV....E.EF.MS-LPEL-..........
23 CALBNGG       ..........Q..F.FEIWH.H.Y..........D....SD...GNG.YMDG....K.EL.QNFIQEL-..........
24 CAL1CE        ..........E..F.REAFM.M.F..........D....KD...GNG.TIST....K.EL.GIAMRSL-..........
25 BCHS          ..........A..L.IDVFH.Q.Y..........Sg...RE...GDKhKLKK....S.EL.KELINNE-..........
26 AEQAV1        ..........R..H.KHMFN.F.L..........D....VN...HNG.KISL....D.EM.VYKASDI-..........
27 1F8           ..........R..R.IELFK.K.F..........D....KN...ETG.KLCY....D.EV.HSGCLEV-..........
28 CALM_ASPNI    adslteeqvsE..Y.KEAFS.L.F..........D....KD...GDG.QITT....K.EL.GTVMRSL-gqnpses109
29 MLE1_HUMAN    apkkdvk129D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
30 MLE1_RABIT    apkkdvk127D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
31 MLEV_HUMAN    apkkpep130D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gerltede35
32 MLEC_CHICK    ppkkpep129D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLRTL-gerlteee35
33 MLEV_RAT      apkkpep135D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gerltede35
34 MLE1_CHICK    pkkdvkk126D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmteee35
35 MLE1_RAT      apkkdvk124D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
36 MLE1_MOUSE    apkkdvk123D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
37 MLEF_HUMAN    apkkpep132D..F.VEGLR.V.F..........D....KE...SNG.TVMG....A.EL.RHVLATL-gekmteae35
38 MLEF_RAT      ppkkpep128D..F.VEGLR.V.F..........D....KE...SNG.TVMG....A.EL.RHVLATL-gekmseae35
39 MLEF_MOUSE    ppkkpep128D..F.VEGLR.V.F..........D....KE...SNG.TVMG....A.EL.RHVLATL-gekmseae35
40 MLEX_CHICK    mplkkpd121D..F.VEGLR.V.F..........D....KE...GNG.LVMG....A.EL.RHVLVTL-gekmtese35
41 MLE3_HUMAN    sfsadqia85D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
42 MLEY_HUMAN    mppkkdv144D..Y.LEGFR.V.F..........D....KE...GNG.KVMG....A.EL.RHVLTTL-gekmteee35
43 MLE3_RABIT    sfsadqia85D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
44 MLE3_RAT      sfsadqia85D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
45 MLE3_MOUSE    sfsadqia85D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmkeee35
46 MLE3_CHICK    sfspdqid85D..F.VEGLR.V.F..........D....KE...GNG.TVMG....A.EL.RHVLATL-gekmteee35
47 AACT_HUMAN    mdhydsq749E..F.RASFN.H.F..........D....RD...HSG.TLGP....E.EF.KACLISL-gydignd114
48 MLE_HALRO     adfsddri86D..F.VEGLR.V.F..........D....KE...NNG.KIMG....A.EL.RHVLSTL-gekmseee36
49 MLES_HUMAN    mcdftedqtaE..F.KEAFQ.L.F..........D....RT...GDG.KILY....S.QC.GDVMRAL-gqnptna112
50 MLEN_HUMAN    mcdftedqtaE..F.KEAFQ.L.F..........D....RT...GDG.KILY....S.QC.GDVMRAL-gqnptna112
51 MLEN_CHICK    cdfseeqta.E..F.KEAFQ.L.F..........D....RT...GDG.KILY....S.QC.GDVMRAL-gqnptna112
52 MLEM_CHICK    cdfseeqta.E..F.KEAFQ.L.F..........D....RT...GDG.KILY....S.QC.GDVMRAL-gqnptna112
53 MLEG_HUMAN    eemnvkml30D..F.VEGLR.V.F..........D....KE...SNG.TVMG....A.EL.RHVLATL-gekmteae35
54 MLE_PATYE     pklsqdei84D..Y.MEAFK.T.F..........D....RE...GQG.FISG....A.EL.RHVLTAL-gerltdee43
55 MLE_AEQIR     pklsqdei84D..Y.MEAFK.T.F..........D....RE...GQG.FISG....A.EL.RHVLTAL-gerlsded43
56 AACT_DROME    mmmengl752E..F.RSSFN.H.F..........D....KN...RTG.RLSP....E.EF.KSCLVSL-gysigke114
57 RECO_CHICK    mgnsrss100K..L.EWAFS.L.F..........D....VD...RNG.EVSK....S.EV.LEIITAI-fkmipeee63
58 MLE_DICDI     msasadqi78E..M.LDAFK.A.L..........D....KE...GHG.TIQG....A.EL.RQLLTTL-gdylstae59
59 SPCA_DROME    menftp2268E..F.SMMFK.H.F..........D....KD...KSG.KLNH....Q.EF.KSCLRAL-gydlpmv118
60 MLR_DICDI     mastkrrl23E..L.KEAFE.L.F..........D....KD...RTG.FIKK....D.AL.KTTCKQF-gvfvmed109
61 MLE_TODPA     sqltkdei85E..F.MEAFK.T.F..........D....RE...GQG.LISS....A.EI.RNVLKML-geritedq45
62 SPCN_CHICK    mdpsgv2331E..F.SMMFK.H.F..........D....KD...KSG.RLNH....Q.EF.KSCLRSL-gydlpmv117
63 CL1L_MOUSE    psqmeham49A..V.DKIMK.D.L..........D....QC...RDG.KVGF....Q.SF.LSLVAGL-tiacndyf18
64 AACS_CHICK    mnsmnqi795E..F.ARIMS.L.V..........D....PN...GQG.TVTF....Q.SF.IDFMTRE-tadtdtae73
65 CL1L_RAT      psqmeham49A..V.DKIMK.D.L..........D....QC...RDG.KVGF....Q.SF.LSLVAGL-iiacndyf16
66 LAV1_PHYPO    msyqeaw220A..L.VADFR.K.I..........D....TN...SNG.TLSR....K.EF.REHFVRL-gfdkksv106
67 CAP3_RAT      mptvisp695S..C.RSMIA.L.M..........D....TD...GSG.RLNL....Q.EF.HHLWKKI-kawqkifk97
68 MLEP_DROME    mvdvpkre83D..F.IECLK.L.Y..........D....KE...ENG.TMLL....A.EL.QHALLAL-geslddeq43
69 MLEL_DROME    mvdvpkre83D..F.IECLK.L.Y..........D....KE...ENG.TMLL....A.EL.QHALLAL-geslddeq43
70 SP2D_STRPU    maanllf111K..I.KEMIE.K.A..........D....FP...NDG.KCSL....E.EF.VKMVMNF-c.........
71 CL1L_BOVIN    psqmeham49A..V.DKIMK.D.L..........D....QC...RDG.KVGF....Q.SF.FSLIAGL-tiacndyf18
72 EHF5_TRYBB    mkdkapv122E..M.RGAFL.H.Y..........D....KQ...KTG.FVTK....K.QF.TELFATG-gecstpee41
73 CL1L_PIG      psqmeham49A..V.DKIMK.D.L..........D....QC...RDG.KVGF....Q.SF.FSLIAGL-tiacndyf17
74 FCAB_TRYBB    mgcsgsk170D..A.TTVFN.E.I..........D....TN...GSG.VVTF....D.EF.SCWAVTK-klqvsgdp34
75 SCP1_ASTPO    ayswdnrv59L..W.NEIAE.L.A..........D....FN...KDG.EVTI....D.EF.KKAVQNV-cvgkafa104
76 CAP2_RABIT    qklirir297T..C.KIMVD.M.L..........D....SD...GTG.KLGL....K.EF.YVLWTKI-qkyqkiyr96
77 CAP3_HUMAN    saiisrn652S..C.RSMIA.L.M..........D....TD...GSG.KLNL....Q.EF.HHLWNKI-kawqkifk97
78 CAPS_HUMAN    mflvnsf142T..C.RSMVA.V.M..........D....SD...TTG.KLGF....E.EF.KYLWNNI-krwqaiyk97
```

```
 79 CAP2_HUMAN   magiaak575T..C.KIMVD.M.L..........D....SD...GSG.KLGL....K.EF.YILWTKI-qkyqkiyr96
 80 KDGL_PIG     mskergl157I..L.QEMMK.E.I..........D....YD...GSG.SVSL....A.EW.LRAGATT-vpllvll548
 81 SCPA_PENSP   ayswdnr103F..I.ANQFK.A.I..........D....VN...GDG.KVGL....D.EYrLDCITRS-afaevkei59
 82 SCPB_PENSP   ayswdnrv59L..W.NEIAE.L.A..........D....FN...KDG.EVTV....D.EF.KQAVQKN-ckgkafa104
 83 IPYR_ARATH   maeikde169E..I.RRFFE.D.Y..........K....KN...ENK.KVDV....E.AF.LPAQAAI-daikdsmd65
 84 SCP1_BRALA   glndfqk105R..I.PFLFK.G.M..........D....VS...GDG.IVDL....E.EF.QNYCKNF-qlqcadvp51
 85 SCP2_BRALA   glndfqk105R..I.PFLFK.G.M..........D....VS...GDG.IVDL....E.EF.QNYCKNF-qlqcadvp51
 86 PIP3_RAT     mdsgrdf143W..I.HSCLR.K.A..........D....KN...KDN.KMNF....K.EL.KDFLKEL-niqvddg584
 87 AACT_CHICK   mdhhydp786E..F.ARIMS.I.V..........D....PN...RMG.VVTF....Q.AF.IDFMSRE-tadtdtad73
 88 CAB_MOUSE    marpleea53A..F.QKVMS.N.L..........D....SN...RDN.EVDF....Q.EY.CVFLSCI-ammcneff19
 89 TEGU_SCHMA   matetkls11E..F.IRAFL.E.I..........D....AD...SNE.MIDK....Q.EL.IKYCQKY-rldmkli150
 90 CAB_RAT      marpleea53A..F.QKLMN.N.L..........D....SN...RDN.EVDF....Q.EY.CVFLSCI-ammcneff19
 91 G19P_HUMAN   mllplll211E..LaADAFK.E.L..........D....DD...MDG.TVSV....T.EL.QTH-PEL-dtdgdga287
 92 TCH2_ARATH   ..........-..-.-----.-.-..........D....KN...GDG.KISV....D.EL.KEVIRAL-sptaspee25
 93 KDGL_HUMAN   makergl158I..L.QEMMK.E.I..........D....YD...GSG.SVSQ....A.EW.VRAGATT-vpllvll548
 94 PIP3_BOVIN   pesqlfsi82W..I.HSCLR.K.A..........D....KN...KDN.KMSF....K.EL.QNFLKEL-niqvdds584
 95 CALM_LYTPI   kkmkdtdseeE..I.REAFR.V.F..........D....KD...GNG.FI--....-.--.-----RL-a.........
 96 CAP1_HUMAN   mseeiit588S..C.RSMVN.L.M..........D....RD...GNG.KLGL....V.EF.NILWNRI-rnylsifr97
 97 CIC1_CYPCA   mesgsg1421E..F.KKIWA.E.Y..........D....PE...ATG.RIKH....L.DV.VTLLRRI-qpplgfg402
 98 GUNF_CLOTM   mkkilaf699E..H.QKFIA.A.A..........D....VD...GNG.RINS....T.DL.YVL--NR-yilkliek13
 99 CIC1_RABIT   mepssp1400E..F.KAIWA.E.Y..........D....PE...AKG.RIKH....L.DV.VTLLRRI-qpplgfg444
100 V57A_BPT4    mseqtveq40T..L.AEIAR.A.V..........G....IT...GD-.TIKV....E.EI.VEAVKNL-taesadea12
101 CALG_CHICK   sllavfqr43V..V.DRMMK.R.L..........D....IN...SDG.QLDF....Q.EF.--------..........
102 NIFH_NOSCO   mdhyvprd56E..L.EELLI.E.F..........-....--...--G.ILES....D.EW.TAMLVGK-tateap....
103 ARFL_DROME   mggvlsyf85A..I.IYVVD.S.A..........D....RD...RIG.-ISK....D.EL.LYMLREE-elagailv67
104 AROA_KLEPN   mesltlq151R..L.RGGFT.G.G..........D....VE...VDG.SVSS....Q.FL.TALLMAS-plapqdt247
105 REL1_HUMAN   mprlflfh96E..L.KAALS.E.Rqpslpelq11P....AL...KDS.NLSF....E.EF.KKLIRNR-qseaadsn49
106 H11_BOVIN    setapaap41L..I.TKAVA.As-..........-....KE...RSG.-VSL....A.AL.KKALAAA-gydveknn36
107 YCSX_CHLPY   malsnil155K..L.IEFLD.N.Y..........K....VE...KAK.SITL....Q.QL.QSVLQNI-klnnsqks26
108 DP3X_ECOLI   msyqvla212S..L.RDALS.L.T..........D....QAiasGDG.QVST....Q.AV.SAMLGTL-dddqals399
109 AROA_SALTY   mesltlq151R..L.RGGFT.G.G..........D....IE...VDG.SVSS....Q.FL.TALLMTA-plapkdt247
110 ANX1_CAVCU   msmvsef102H..L.EEVVL.A.L..........L....KT...PA-.QLDA....D.EL.RAAMKGL-gtdedtl216
111 CICC_RAT     mirafa1524E..F.KRIWA.E.Y..........D....PE...AKG.RIKH....L.DV.VTLLRRI-qpplgfg616
112 CICC_RABIT   mlralv1525E..F.KRIWA.E.Y..........D....PE...AKG.RIKH....L.DV.VTLLRRI-qpplgfg617
113 LACA_LACLA   maivvgad21L..V.EEGFE.V.I..........D....VT...KDG.Q-DF....V.DV.TLAVASE-vnkdeqnl92
114 AROA_BORPE   msglayld31L..L.LAALA.E.G..........S....TE...ITG.LLDS....D.DT.RVMLAAL-rqlgvsv382
115 AROA_SALTI   mesltlq151R..L.RGGFI.G.G..........D....IE...VDG.SVSS....Q.FL.TALLMTA-plapedt247
116 AROA_SALGL   mesltlq151R..L.RGGFI.G.G..........D....IE...VDG.SVSS....Q.FL.TALLMTA-plapkdt247
117 CAP1_CHICK   mpfggia608S..W.LTIFR.Q.Y..........D....LD...KSG.TMSS....Y.EM.RMALESA-gfklnnkl67
118 PR10_CAVPO   ..........-..-.-EIWK.H.F..........D....AD...ENG.YIEG....K.FM.QKY--DK-nsdqhvgs68
119 SC1_RAT      mkavlll593EhcI.TRFFE.E.C..........D....PN...KDK.HITL....K.EW.GHCFGIK-eedidenllf
120 QR1_COTJA    mktvlll635EhcI.TRFFQ.E.C..........D....GD...QDK.LITL....K.EW.CHCFAIK-eedinenllf
121 RS37_NEUCR   gkkrkkkv22K..L.-AVLK.Y.Y..........K....VD...SDG.KIER....-.--.---LRRE-cpnetcga34
122 YTR1_SPIAU   mvmdhdin49P..T.KFVAS.I.A..........D....--...--G.RVTF....R.FF.VPLGLRL-daktplav66
123 SPCB_HUMAN   kfedflg225E..L.GELFA.Q.V..........PsmgeEG...GDA.DLSI....EkRF.LDLLEPL-grrkkqle15
124 OTNC_MOUSE   mrawiff261EhcT.TRFFE.T.C..........D....LD...NDK.YIAL....E.EW.AGCFGIK-eqdinkdlvi
125 CALG_RABIT   dgh.......S..V.-----.-.-..........-....--...---.TLSK....T.EF.LSFMNTE-laaftkdp16
126 SPCA_MOUSE   rvcdgde199A..V.QNVLD.T.A..........E....SL...RDKaAVGK....E.EI.QERLAQF-vqhweklk24
127 OTNC_HUMAN   mrawiff262EhcT.TRFFE.T.C..........D....LD...NDK.YIAL....D.EW.AGCFGIK-qkdidkdlvi
128 OTNC_BOVIN   mrawiff263EhcT.TRFFE.T.C..........D....LD...NDK.YIAL....D.EW.AGCFGIK-ekdidkdlvi
129 Y493_BPT4    mi........E..L.NEQIIfL.G..........Dg...TE...GDL.EYKL....Y.EY.MIWLAKA-egidfvvs69
130 KDGL_ECOLI   anttgft48D..V.DAITR.V.L..........-....--...---.LISS....V.ML.VMIVEIL-nsaieavv50
131 SPCA_HUMAN   etvvess128E..L.RHLWD.L.L..........-....--...--L.ELTL....E.KG.DQLLRAL-kfqqyvq443
132 IMMC_ECOLI   mglklhih15E..F.KG--G.E.Y..........S....KDf..GDD.----....-.--.GSVIESL-gmplkdni49
133 DGAL_ECOLI   mnkkvlt283Q..A.KATFD.L.A..........Knl..AD...GKG.AADG....T.NW.KID--NK-vvrvpyvg20
134 SPCB_MOUSE   lqaflqdl89A..L.RRMWE.S.R..........G....NT...LTQ.CLGF....Q.EF.QKDAKQA-eailsnq118
135 SP10_YEAST   mkftsvla51T..T.TTLFN.St-..........-....--...--S.TLNI....T.QL.YQIATQV-nqtlqse251
136 SRCH_HUMAN   mghhrpw592R..R.EEAGG.A.S..........S....EE...ESG.EDTGpqdaQ.EY.GNYQPGS-lcgycsfc74
137 SRCH_RABIT   mgcrgpw144D..L.AEHGS.H.Ghghee.....-....ED...ED-.VISS....E.RP.RHVLRRA-prghgge676
```

B) Details on sequences 28-137 shown in the alignment (arranged in order of decreasing Z-score). **NLL-score** and **Z-score** are measures of how well the EF-hand HMM fits these database "test" sequence that were not present in the training set (see Section 3.3 for more details). In **HMM**, **PROFILESEARCH**, **Keyword** and **PROSITE** "+" and "-" denote sequences that are and are not, respectively, classified as containing an EF-hand motif by the four specified methods. For **PROFILESEARCH**, "Gribskov" and "HMM" indicate results based upon profiles generated from four EF-hand sequences and our HMM alignments. "T", "N", "P" and "?" in **PROSITE** have the same meaning as in Figure 11. "

ID	Length	NLL-score	Z-score	HMM	PROFILESEARCH		Keyword	Prosite
					Gribskov	HMM		
28 CALM_ASPNI	148	398.961	12.975	+	-	-	+	T
29 MLE1_HUMAN	193	542.924	11.662	+	+	+	-	%
30 MLE1_RABIT	191	537.011	11.661	+	+	+	-	%
31 MLEV_HUMAN	194	546.027	11.631	+	+	+	-	%
32 MLEC_CHICK	193	543.095	11.605	+	+	+	-	%
33 MLEV_RAT	199	561.007	11.561	+	+	+	-	%
34 MLE1_CHICK	190	534.042	11.516	+	+	+	-	%
35 MLE1_RAT	188	528.051	11.262	+	+	+	-	%
36 MLE1_MOUSE	187	525.056	11.224	+	+	+	-	%
37 MLEF_HUMAN	196	554.316	11.005	+	+	+	-	%
38 MLEF_RAT	192	542.332	10.892	+	+	+	-	%
39 MLEF_MOUSE	192	542.332	10.892	+	+	+	-	%
40 MLEX_CHICK	185	521.797	10.342	+	+	+	-	%
41 MLE3_HUMAN	149	411.100	10.201	+	+	+	-	%
42 MLEY_HUMAN	208	588.847	10.194	+	+	+	-	%
43 MLE3_RABIT	149	411.179	10.177	+	+	+	-	%
44 MLE3_RAT	149	411.207	10.169	+	+	+	-	%
45 MLE3_MOUSE	149	411.208	10.169	+	+	+	-	%
46 MLE3_CHICK	149	411.206	10.169	+	+	+	-	%
47 AACT_HUMAN	892	2642.237	9.957	+	-	+	+	T
48 MLE_HALRO	151	418.497	9.918	+	+	+	-	%
49 MLES_HUMAN	151	418.627	9.879	+	+	+	-	%
50 MLEN_HUMAN	151	418.627	9.879	+	+	+	-	%
51 MLEN_CHICK	150	415.631	9.798	+	+	+	-	%
52 MLEM_CHICK	150	415.631	9.798	+	-	-	-	%
53 MLEG_HUMAN	94	248.725	9.735	+	+	+	-	%
54 MLE_PATYE	156	433.703	9.629	+	+	+	-	%
55 MLE_AEQIR	156	433.703	9.629	+	+	+	-	%
56 AACT_DROME	895	2653.286	9.130	+	-	+	+	T
57 RECO_CHICK	192	548.396	8.848	+	-	-	+	T
58 MLE_DICDI	166	465.170	8.834	+	+	+	-	T
59 SPCA_DROME	2415	7205.568	8.787	+	-	+	+	T
60 MLR_DICDI	161	451.967	8.678	+	+	+	+	-
61 MLE_TODPA	159	446.406	8.616	+	+	+	-	%
62 SPCN_CHICK	2477	7392.895	8.157	+	-	+	-	T
63 CL1L_MOUSE	96	263.095	7.516	+	+	+	-	%
64 AACS_CHICK	897	2663.548	7.446	+	-	-	+	-
65 CL1L_RAT	94	257.103	7.423	+	+	+	-	%
66 LAV1_PHYPO	355	1039.236	7.298	+	-	+	-	T
67 CAP3_RAT	821	2436.445	7.150	+	-	+	+	T
68 MLEP_DROME	155	439.713	7.053	+	+	+	-	%
69 MLEL_DROME	155	439.713	7.053	+	+	+	-	%
70 SP2D_STRPU	141	397.689	6.990	+	+	+	+	-
71 CL1L_BOVIN	96	265.582	6.819	+	-	-	-	%
72 EHF5_TRYBB	192	554.482	6.797	+	+	+	-	-
73 CL1L_PIG	95	262.586	6.763	+	+	+	-	%
74 FCAB_TRYBB	233	676.012	6.684	+	-	-	+	T
75 SCP1_ASTPO	192	554.824	6.681	+	-	+	+	T
76 CAP2_RABIT	422	1242.278	6.589	+	-	+	+	T
77 CAP3_HUMAN	778	2307.499	6.577	+	-	+	+	T
78 CAPS_HUMAN	268	782.852	6.383	+	-	+	+	T
79 CAP2_HUMAN	700	2074.486	6.305	+	-	-	+	T
80 KDGL_PIG	734	2176.760	6.160	+	-	-	+	T

ID	Length	NLL-score	Z-score	HMM	PROFILESEARCH		Keyword	Prosite
					Gribskov	HMM		
81 SCPA_PENSP	192	556.636	6.071	+	-	+	+	T
82 SCPB_PENSP	192	557.071	5.924	+	-	+	+	T
83 IPYR_ARATH	263	769.241	5.909	+	-	-	-	-
84 SCP1_BRALA	185	535.787	5.827	+	-	+	+	T
85 SCP2_BRALA	185	535.816	5.818	+	-	+	+	T
86 PIP3_RAT	756	2244.255	5.713	+	-	-	-	?
87 AACT_CHICK	888	2641.411	5.684	+	-	-	+	N
88 CAB_MOUSE	101	284.695	5.589	+	-	-	+	-
89 TEGU_SCHMA	190	552.242	5.469	+	-	+	-	?
90 CAB_RAT	101	285.488	5.369	+	-	-	-	-
91 G19P_HUMAN	527	1560.198	5.330	+	-	-	-	T
92 TCH2_ARATH	45	116.235	5.321	+	-	-	+	T
93 KDGL_HUMAN	735	2182.343	5.301	+	-	-	+	T
94 PIP3_BOVIN	695	2063.206	5.034	+	-	-	-	?
95 CALM_LYTPI	30	67.341	4.942	+	-	-	+	P
96 CAP1_HUMAN	714	2120.342	4.924	+	-	+	+	T
97 CIC1_CYPCA	1852	5530.321	4.714	-	+	-	-	-
98 GUNF_CLOTM	739	2196.618	4.602	-	-	-	-	?
99 CIC1_RABIT	1873	5593.640	4.550	-	+	-	-	-
100 V57A_BPT4	80	224.359	4.470	-	-	-	-	-
101 CALG_CHICK	65	178.908	4.438	-	+	+	+	T
102 NIFH_NOSCO	86	243.556	4.347	-	-	-	-	-
103 ARFL_DROME	180	524.609	4.300	-	-	-	-	-
104 AROA_KLEPN	427	1264.280	4.296	-	-	-	-	-
105 REL1_HUMAN	185	540.676	4.249	-	-	-	-	-
106 H11_BOVIN	104	298.227	4.240	-	-	-	-	-
107 YCSX_CHLPY	110	316.022	4.210	-	-	-	-	-
108 DP3X_ECOLI	643	1910.667	4.186	-	-	-	-	-
109 AROA_SALTY	427	1264.760	4.130	-	-	-	-	-
110 ANX1_CAVCU	346	1022.514	4.043	-	-	-	-	-
111 CICC_RAT	2169	6481.468	4.011	-	+	-	-	-
112 CICC_RABIT	2171	6487.460	4.010	-	+	-	-	-
113 LACA_LACLA	141	407.967	3.986	-	-	-	-	-
114 AROA_BORPE	442	1310.475	3.985	-	-	-	-	-
115 AROA_SALTI	427	1265.295	3.945	-	-	-	-	-
116 AROA_SALGL	427	1265.295	3.945	-	-	-	-	-
117 CAP1_CHICK	704	2093.590	3.888	-	-	-	+	T
118 PR10_CAVPO	92	267.751	2.866	-	+	+	+	P
119 SC1_RAT	634	1888.351	2.662	-	-	-	-	T
120 QR1_COTJA	676	2015.770	1.941	-	-	-	-	T
121 RS37_NEUCR	78	229.363	1.766	-	+	-	-	-
122 YTR1_SPIAU	140	412.844	1.753	-	+	-	-	-
123 SPCB_HUMAN	274	814.811	1.610	-	-	-	+	-
124 OTNC_MOUSE	302	899.470	1.146	-	-	-	+	T
125 CALG_RABIT	35	106.946	1.126	-	+	+	+	P
126 SPCA_MOUSE	253	753.490	1.101	-	-	-	+	-
127 OTNC_HUMAN	303	902.914	0.988	-	-	-	+	T
128 OTNC_BOVIN	304	905.856	0.983	-	-	-	+	T
129 Y493_BPT4	102	305.597	0.603	-	+	-	-	-
130 KDGL_ECOLI	121	362.137	0.547	-	-	-	+	-
131 SPCA_HUMAN	595	1779.087	0.039	-	-	-	+	-
132 IMMC_ECOLI	85	257.069	0.025	-	+	-	-	-
133 DGAL_ECOLI	332	992.734	-0.028	-	+	-	-	-
134 SPCB_MOUSE	236	706.853	-0.161	-	-	-	+	-
135 SP10_YEAST	326	978.184	-1.203	-	-	-	+	-
136 SRCH_HUMAN	699	2098.086	-2.613	-	-	-	+	-
137 SRCH_RABIT	852	2556.715	-3.145	-	-	-	+	-

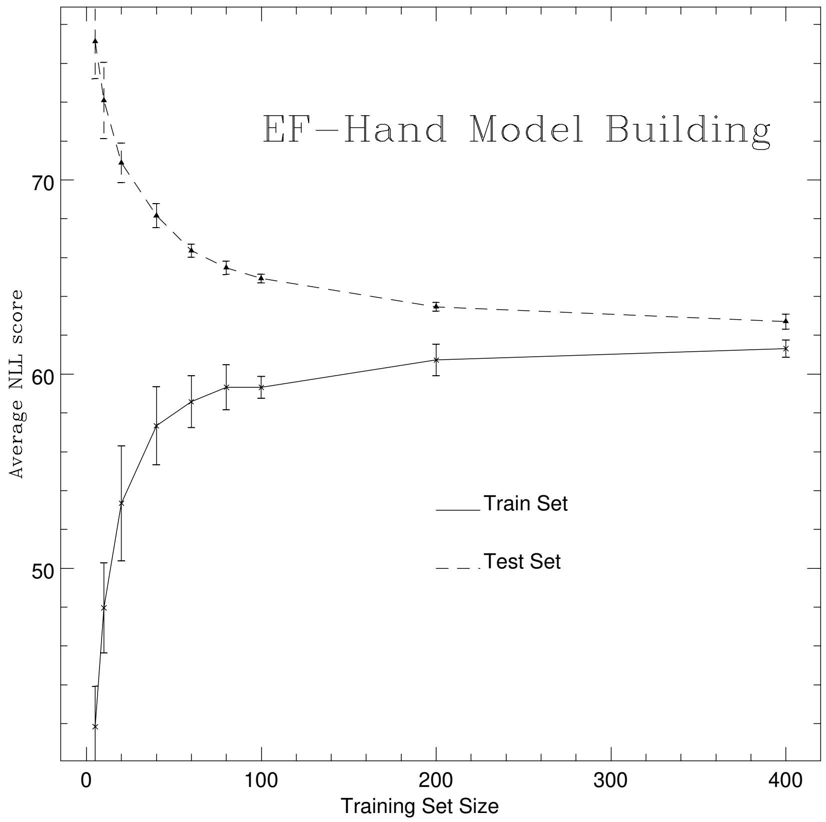


Figure 14: Average NLL scores for test and train sets for models with training sets of size 5, 10, 20, 40, 80, 100, 200, and 400. Error bars represent one standard deviation.

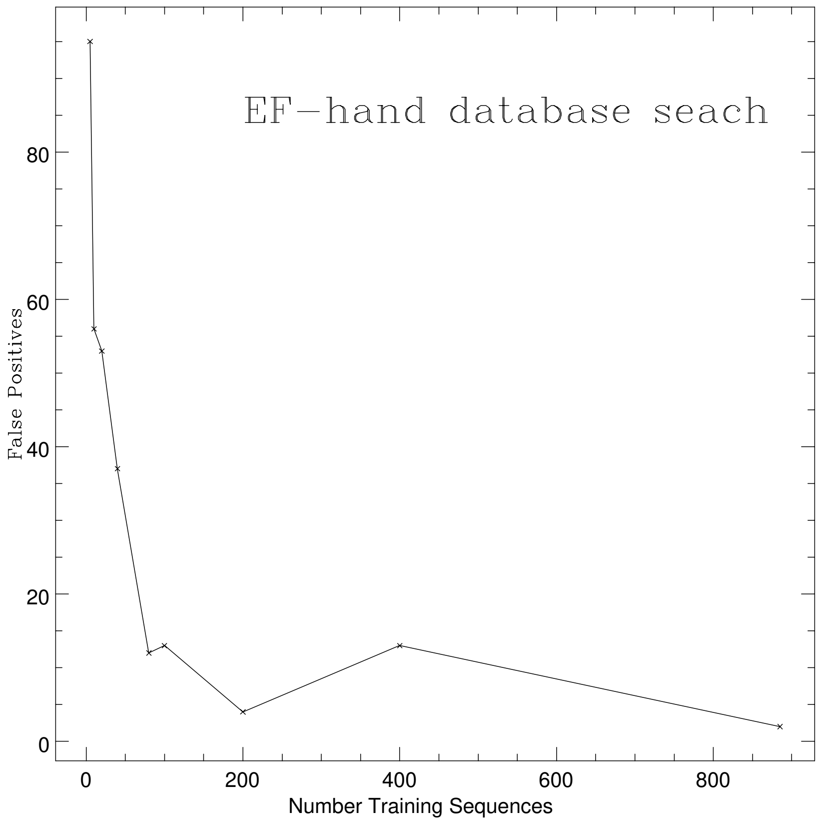


Figure 15: EF-hand database search false positives for models trained with 5, 10, 20, 40, 80, 100, 200, 400, and 885 sequences.