

# Supplementary Data

**Usage of a dataset of NMR resolved protein structures to test aggregation vs. solubility prediction algorithms**

Daniel B. Roche<sup>1,2</sup>, Etienne Villain<sup>1,2</sup> and Andrey V. Kajava<sup>1,2,\*</sup>

<sup>1</sup>Centre de Recherche en Biologie cellulaire de Montpellier, CNRS-UMR 5237, Montpellier, France

<sup>2</sup>Institut de Biologie Computationnelle, Université de Montpellier, Montpellier France

\*Corresponding author: Andrey V. Kajava email: [andrey.kajava@crbm.cnrs.fr](mailto:andrey.kajava@crbm.cnrs.fr)

>1ap8\_1-24  
MSVEEVSKKFEENVSVDDTTATPK  
>1b22\_1-15  
MAMQMQLLEANADTSV  
>1b22\_86-114  
FTTATEFHQRREIIQITTGSKELDKLL  
>1bq0\_77-103  
GMGGGGFGGGADFSDIFGDVFGDIFG  
>1ckv\_1-19  
MSVNSNAYDAGIMGLKGD  
>1cl4\_1-28  
GGSCFKCGKKGHFAKNCHEHAHNNAEPK  
>1clid\_43-61  
PQVVRTPLTRAHTEMENR  
>1d1r\_1-27  
MSDSNSRLVYSTETGRIDEPKAAPVRP  
>1e17\_125-150  
KSGKAPRRRAASMDSSSKLLRGRSK  
>1e17\_1-34  
GSSHHHHHHSSGLVPRGSHMLEDPGAVTGPRKGG  
>1ezy\_134-166  
PSSCGAEKQKGAUSSADCTSLVPCAHHHHH  
>1g9l\_1-44  
GPLGSAAAATPAVRTVPQYKYAAGVRNPQQHLNAQPQVTMQQPA  
>1hvx\_1-28  
MRGSHHHHHHGSEKLAKTKSTMVDVSDK  
>1ibi\_1-35  
MDRGERLGIKPESSPSHRPTTNPNTSKFAQKFGG  
>1ibi\_95-113  
FGPKGFGYGQGAGALVHA  
>1j8c\_104-125  
RDPNSSVDKYAAALEHHHHH  
>1j8c\_1-19  
MAENGESSGPPRPSRGPA  
  
>1jfn\_1-16  
ARIHHHHHHIEGRAPT  
>1jgg\_1-21  
MTVIKPSRRPRRKNKNIKVN  
>1k9c\_1-19  
SKKIKDPDAAKPEDWDERA  
>1kft\_1-22  
MGSSHHHHHHSSGLVPRGSHMN  
>1kkd\_81-102  
NDQANTLVDLAKTQLEHHHHH  
>1kq8\_82-100  
NSEYTFADGVFRRRRYRL  
>1l3g\_122-136  
KVDKLAAALEHHHHH  
>1lg4\_1-28  
VQTRGIKHRIKWNRKALPSTAQITEAQV  
>1lv3\_52-68  
SSGDLSESDDWSEEPKQ  
>1m9o\_44-77  
RHPKYKTELCHKFLQGRCPYGSRCHFIHNPTED  
>1n4c\_1-57  
GPLGSPEFSMPHSSPQNRPNYNVFSMPGGQNERGKAAANLEGKQKAADFEDLLSG  
>1nmv\_1-19  
MADEEKLPGWWEKRMSRSS

>1nmv\_24-46  
YFNHITNASQWERPSGNSSSGGK  
>1no8\_1-28  
SRPKQLPDKWQHDLFDSGFGGGAGVETG  
>1nwb\_102-124  
PNATGSCGCGSSFSCGLEHHHHH  
>1nyn\_116-131  
SSLKTKGGNAGTKAY  
  
>1ov2\_1-16  
YSREKNQPKPSPKRES  
>1owa\_1-46  
MEQFPKETVVESSGPKVLETAEEIQERRQEVTRYQSFKERVAERG  
>1pul\_1-17  
MGHHHHHHSHMAAAAGF  
>1pxe\_1-15  
MHVKKPYDPSRTEK  
>1pyc\_46-71  
AEEAEKELLKDNELKKLRERVKSLEK  
>1q3m\_1-15  
YLDHWLGAPAPYPDP  
>1qkl\_1-16  
MSDNEDNFDGDDFDDV  
>1qlz\_1-104  
GSKKRPKPGGWNTGGSRYPGQSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSK  
PKTNMKHMAGAAAAGAVVGG  
>1qu5\_1-17  
EAETREQKLLHSNNTEN  
>1r4t\_1-18  
GSASSAVVFKQMVLLQQAL  
>1ri9\_1-19  
MGSSHHHHHHSSGKLLKKQ  
>1rw2\_1-22  
MHHHHHHKLTEQGGAHFSVSS  
>1rwu\_1-22  
MGTSHHHHHHSSGRENLYFQGH  
>1s24\_1-31  
PTSLEKLPSADVKGQDLYKTQPPRSDAQGGK  
>1s3a\_1-17  
KAGMAAAAASRGVGAKL  
>1s6i\_167-188  
GLVDNGSNQVIEGYFKHHHH  
>1se9\_111-126  
KPKGDPKMNKVCSV  
>1sg7\_1-21  
MGSSHHHHHHHSSGFNPRGS  
>1sgo\_1-19  
METDCNPMELSSMSGFEEG  
>1sjr\_1-39  
MHHHHHHGSAQAALQAVNSVQSGNLALAASAAAVDAGMA  
>1sjr\_148-164  
SQPSLDQTMAAAFGL  
>1so9\_1-20  
MVPLYDMFCRVGTGYNGTTQR  
>1sr3\_115-136  
MEANHRRPASVYKDPASHHHHH  
>1ssf\_131-156  
AVTPLTKAADISLDNLVEGKRKRRS  
>1suh\_107-146

RPEFTQEVFEGSVAEGAVPGTSVMKVSATDADDDVNTYN  
>1t0y\_93-122  
FKDESMVEKEYEMSDDTYGKRTDSVRAWKK  
>1tp4\_69-97  
GELADISDVEGEVGAIGEEAPQMNYIQV  
>1ttn\_1-20  
MIEEKSDIETLDIPEPPPNS  
>1u6f\_1-31  
MSQIPLVSQYDPYGQTAQLQLLQQQQQQHIP  
>1umq\_1-30  
MGSSHHHHHSSGLVPRGSHMLAKGESLPP  
>1v9j\_1-19  
MKGSSHHHHHSSGASLVP  
>1vaz\_1-16  
MRGSHHHHHHGSEERR  
>1vg5\_1-18  
GSSGSSGRQAPIANAAV  
>1wfi\_112-131  
NSKLSLDSETRSMVSGPSS  
>1wfk\_73-88  
SKWSPPQNYKSGPSS  
  
>1wfo\_1-16  
GSSGSSGRIGDGSPSH  
>1wi5\_102-119  
ATEQQSWNLNLSGPSS  
>1wif\_1-16  
GSSGSSGSKNEKEQLS  
>1wiv\_1-20  
GSSGSSGLLSHMDDPDIDAP  
>1wj7\_86-104  
EMVGKKKGVSGQKSGPSS  
>1wki\_1-16  
MLMPRRMKYRKQRGR  
>1x37\_97-124  
GQAETEDQVGVVTGLAYTTVLRHHHHH  
>1x64\_1-16  
GSSGSSGVRAPVTKVH  
>1x6f\_64-88  
QLLSKQKYADGAFADFKQESGPSS  
>1xsf\_1-16  
NVVVTPAHEAVVRVGT  
>1xx3\_1-55  
MHHHHHHVKKVQEQPKRDVKPVESRPASPENTAPARLTSSTATAATSKPVTSVA  
>1y9j\_1-19  
GSPGISGGGGGILVGSKMA  
>1ydu\_1-44  
SDQIFNKVGSYWLQKANKQFDSVGNLNSVSTSIEGGTKWLVN  
>1yse\_105-141  
AASAMGPAPLISTPPSRPPQVKTATIATERNKGPEN  
>1yse\_1-18  
GSHVSRSMNKPLEQQVST  
>1ysm\_55-77  
ELDNEKPAAVVAPLTTGYTVKIS  
>1ywu\_1-27  
MGSSHHHHHSSGRENLYFQGHMSDQH  
>1z8r\_1-17  
MAHHHHHHERASLITG  
>1z9b\_1-24

NEFELGTRGSSRVDLQEQRSVKTR  
>1z9e\_86-127  
GEGDSVITQVLNKS LAEQRQHEEANKTKDQGGKGNKKLEK  
>1zec\_1-27  
XGGKWSKSSVIGWPAVRERMRAEPAX  
>1zr9\_1-35  
GHHHHHHLEKAKRRRDLDEIHRELRPQGSARPQP  
>1zr9\_91-124  
EAERAAGMGSYVPPRRLAVPTEVSTEVPEMDTS  
>1zu1\_72-127  
DSVPAKKFKAAPAEISGDERSKCCPVCNMTFSSPVVAESHYIGKTHIKNLRLRE  
>2a00\_1-20  
MGHHHHHHHHHSSGHGGRH  
>2a3j\_1-21  
MKETAAAKFERQHMDSPDLGS  
>2aby\_1-22  
MGTSHHHHHSSGRENLYFQSH  
>2adb\_1-23  
MGSSHHHHHSSGLVPRGSHMDA  
>2adz\_81-104  
SPYFKNSAGGTSVGWDSPPASPL  
>2afj\_1-15  
GSSARQSTPTSQALY  
>2akl\_1-56  
MGSSHHHHHSSGRENLYFQGHMVSTLPPCQCNSEYTYEDGALLVCPECAHEWSP  
>2ami\_1-20  
GSMSYKAKTVVSARRDQKKG  
>2ayx\_1-94  
MGSGSVEGLSGKRCWLAVRNASLQCFLETSLQRSGIVVTYEQEPTPEDVLITDEVVSKKWQGRAVVTFCRRHIGIPLEKAPGEVWHSVAAPH  
>2b1w\_100-127  
SPSLLTQSKVVVNTDPVSRYTQQLRHHL  
>2baf\_111-166  
TRSSHEFDGRTGLAPEFAALGESGSSSKTSTHKSQFVSSSTTVNRGGSIAIESK  
  
>2baf\_1-32  
MGTFREEGSVSSGTKQEFHTGKLVTTKGDKEL  
>2bgo\_124-140  
GLTIRSPANLEHHHHH  
>2bl5\_115-140  
LMELAILNGTYRDANLKSPALHHHHH  
>2cka\_1-37  
MGSSHHHHHSSHMLVPRGSLKLVLEVEECMEEPNH  
>2cqe\_84-98  
DEKEVEELKKS GPSS  
>2cqw\_67-84  
DFVVQHLKWTDPSPGPSS  
>2cru\_1-20  
GSSGSSGLRRQRLAELQAKH  
>2cs2\_116-134  
GDEV DGVDEVAKKKSGPSS  
>2ctf\_1-15  
GSSGSSGEPEKLGQA  
>2czo\_140-155  
GFDREFERDENQNNI  
>2d1u\_84-104  
TVVGDWLG DARENDLEHHHHH  
>2dba\_1-21  
GSSGSSGMTVSGPGTPEPRPA  
>2dip\_1-17

GSSGSSGLEEFKNSSKL  
>2e8m\_1-15  
GSSGSSGTIGRSAAQ  
>2elh\_1-15  
GSSGSSGMNIRMGTK  
>2evz\_1-25  
MGSSHHHHHSSGLVPRGSHMGRIA  
>2f05\_86-105  
RSLFTGNGSCEMNSGQKNEE  
>2f3j\_1-70  
MASMTGGQQMGRDPMADKMDMSLDDIIKLNRRNRGGGPRRNRPAIARGGRNRPAYSRPKLPDKW  
>2ffw\_1-19  
QKASVSGPNPSETRRERA  
>2fjl\_38-83  
DQGNDEEEEPKEASGSTEHLSSLEVLFGPNPAILEPEREHLDEN  
>2g0q\_150-173  
IMECKKKPQGGQNDISHVLR  
>2gppq\_1-27  
MATVEPETTPNPPTTEEEKTESNQE  
>2h7a\_95-110  
LKQHLSVMGQKTDDTN  
>2ha1\_101-119  
SNTVTGETTPFSPLVATSE  
>2hdm\_73-92  
IQTkPTGTQQSTNTAVTLT  
>2hep\_43-85  
SSMKNTLKSVKIIDPEGNDVTPEKLRQNNKLHLEHHHHH  
>2hfh\_94-109  
LRRRKRFRKRLQHHHHH  
>2hg7\_61-110  
PPDQVELLAQELSQEKLARKQLEELNKT LGNELSDIKLSLLSLEHHHHH  
>2hgc\_83-102  
YKTIKEIKDWIKLEHHHHH  
>2hgn\_1-44  
MGSSHHHHHSSGLVPRGSHMASMTGGQQMGRGSGDSEFTVQST  
>2hh8\_134-149  
KTTQIFPRLEHHHHH  
>2hji\_157-179  
FDNPEGWIAQFTGDDIASAYPR  
>2hlw\_1-23  
MPGEVQASYLKSQSKLSDEGRLE  
>2i83\_144-160  
RTNPEDIYPSNPTDDD  
>2if1\_1-30  
MRGSHHHHHHTDPMSAIQNLHSFDPFADAS  
  
>2jn6\_55-97  
GTTPSAAVSEAEQIRQLKKNALQRARTRHPAESCLEHHHHH  
>2jnb\_1-16  
MGHHHHHHSSGIEEGR  
>2jnc\_1-29  
GSGMKETAAAKFERQHMDSPDLGTTLLEQ  
>2jne\_1-30  
FCLTLRRRYTMGSSHHHHHSSGLVPRGSH  
>2jng\_81-105  
DIEDMVEADEYQGAVASRVLGRAL  
>2jnu\_137-154  
EGRPLREPGSSRLGSPDA  
>2joh\_1-35

MHQGGTHNQWGKPSKPKTSMKXVAGAAAAGAVVGG  
>2joo\_52-66  
NQGDFEPIPEDAYDE  
>2joz\_120-135  
EKDEKSKSLEHHHHH  
>2jp0\_1-21  
MGGSHHHHHSSGLVPRGSHM  
>2jq6\_1-36  
GPLGSESLMPSQVVKGGAFDGTMMNGPFGHGYGEGAG  
>2jqa\_1-19  
MGHTMPAHTPPAQTAPAAQ  
>2jqj\_130-151  
ENDDEKVSSERSYRSHHHHH  
>2jqq\_1-50  
GSDSLIRDLSGLSQKMQVTLLEQIRSNYDDYLTFSTNTYDDEENETLINLE  
>2jqv\_146-165  
NKNPNKGSMEEAVGSLLS  
>2jrb\_1-15  
MQVTYKGRPIRITPD  
>2jrf\_115-184  
TKAKTGGAVDRLDTSRYTGSHERFDESGKGIAGRQDILDDSGYVSAYKNAGTYDAKVKKLEHHHH  
>2jrs\_1-16  
MGHHHHHSHMAAAMA  
>2jtx\_1-49  
GSMAGSVGAAAPVTAANGDSESESESDDDSPRPAAVAVHKREEDDEE  
>2jtz\_54-69  
PEELQEMIDEVEDEGDS  
>2ju5\_1-19  
MAHHHHHSAARRRASGEN  
>2juf\_78-105  
PEEATEDKASAAVEKGAGATVLGTAFP  
>2jul\_1-75  
MQRTKEAVKASDGNLLGDPGRIPLSKRESIKWQRPRFTRQALMRCLIKWILSSAAPQGSDDSELELSTVRHQ  
>2jvl\_1-18  
GAMDPEFAGGTEGQRLTK  
>2jvo\_1-27  
MGSSHHHHHSSGLVPRGSHMHRQEG  
>2jvr\_1-24  
MGSSHHHHHSSGLVPRGSHMSKL  
>2jwy\_1-17  
MVQQSEVRQMKHSVSTL  
>2jx2\_1-34  
MGSSHHHHHSSGLVPRGSHMGPFRRSDSFPERR  
>2jx3\_111-131  
KPSGKPLPKSKKTCSKGSKKE  
>2jy9\_108-148  
ATRPTRASKERRLSSKAQKSSVKALRGKVRRLDLEHHHHH  
>2jya\_84-106  
TRKVVSYTDNFRFNRTQPWTHG  
>2jyv\_35-72  
ASCCEDRVHCCPHGAFCDLVHTRCIKAAALEHHHHHH  
>2jzc\_1-23  
MGSSHHHHHSSGLVPRGSHMLE  
>2k1d\_1-39  
GSDPGQGGGTHSQWNKPSKPKTNMKHMAGAAAAGAVVGG  
>2k2d\_1-32  
GSHMMHSALDMTRYWRQLDDEVAQTPMPSEYQ  
  
>2k31\_151-176

VLHNAQLYETSLENKRNLEHHHH  
>2k3a\_1-34  
MKKLVTTATTLTAGIGAAIVGLDHGNEADAAEQTQ  
>2k3j\_110-146  
PQEDEDEEEEREKKPAEQAEETAPEATATKEEEGS  
>2k3j\_1-44  
GSFTMSYCRQEGKDRIFVTKEDHETPSSAELVADDPNDPYEEH  
>2k3p\_1-29  
SGNYLGVSQNFGRIAPVTGGTAGISVGVV  
>2k4b\_1-30  
MSYYHHHHHHHDYDIPTTENLYFQGAMNEVE  
>2k4k\_93-130  
KPKAAQVSEEAETPQGFNTLKDLEEWIEMSNRKDLIK  
>2k5f\_89-105  
RRRHRFGKRALEHHHHH  
>2k5h\_1-25  
MGSSHHHHHSSGRENLYFQGHMAA  
>2k8p\_1-50  
GWQAFKNDATEIIPELGEYEPPELENNKTMNRAENGGRPPHPFETKD  
>2k8p\_169-189  
RKPRPRARSAKANQAELEN  
>2k8q\_109-134  
ALTENTDAKKTQKPLIQEVETDGVSN  
>2kav\_1-29  
MGSSHHHHHSSGLVPRGSHMASENFSVA  
>2kc5\_144-162  
DVPHAGRRALLFGRRSGE  
>2kco\_1-17  
MGFYQGPDPNRKITGGLK  
>2kd3\_1-16  
GSHMNGGRPPHPYDA  
>2ke7\_1-22  
GSDITSLYKKAGSEMNGPRCPV  
>2ke7\_86-103  
MRPIGHDGYHPTSVAEW  
>2keb\_1-23  
MGSSHHHHHSSGRENLYFQGPVS  
>2keo\_1-20  
MHHHHHSSGRENLYFQGN  
>2kfw\_161-196  
DHDHDGCCGGHGHGHEHGEGCCGGKNGGCG  
>2kgo\_1-31  
MGSSHHHHHSSGLVPRGSHMASGWANDDAV  
>2khc\_1-25  
AAGLPQFGSASALSTPLASVALSA  
>2khd\_1-16  
MSNQTCVENEVCEACG  
>2khi\_1-22  
MGSSHHHHHSSGLVPRGSHWV  
>2khn\_1-16  
MGHHHHHSHMAQFPT  
>2kkl\_1-20  
VFRADFLSELDAPAQAGTES  
>2kkt\_58-84  
EVESLDNDSFDMTDSQALISRLQWD  
>2kkw\_100-140  
QLGKNEEGAPQEGILEDMPVDPNEAYEMPSEEGYQDYEPE  
>2klo\_1-30



PAASPSALKGVSQALLERIRAKEVQQLAR  
>2kmv\_96-110  
NIKNASLVQIDASNE  
>2kmw\_117-150  
ETASDDESFAFVNQDSESSDDGLLYLPDLEKARN  
>2kn0\_45-66  
FPWIWVEQKLISEEDLHHHHHH  
>2kna\_89-104  
EISTEEQLRRLQEEK  
>2ko6\_70-89  
KDYEEDFKTALLRARGVIK  
>2kom\_1-20  
MRGSHHHHHHHHGSENLYFQ  
>2kpy\_84-108  
GGSPPPADGGSPVDGGSPPPPST  
>2kqb\_56-89  
RPECYPGPSCYRKNPQHKEYRHNTLPVRNVLD  
>2kql\_2-24  
DCLPHLKLCKENKDCCKKCKRR  
>2kqv\_130-198  
SSKTSEEHFVETVSLAGSYRDWSYSGQRTELGVFELKRGDKIVYHTLESPEVFHLDGEVLSLDKLSL  
>2kqz\_1-20  
STAASPTQPIQLSDLQSILA  
>2kr0\_1-19  
GPGSMTTSGALFPSLVPGS  
>2kr0\_156-251  
GEGGLQSLGNNMSHSQLMQLIGPAGLGLGGLGALTGPGLASLLGSSGPPGSSSSSSRSQSAAVTPSSTTSSTRATPAPSAPAAAASATSPSPAPS  
>2kr6\_1-41  
MVWLVNMAEGDPEAQRVSKNSKYNAESTERESQDTVAEND  
>2krq\_119-206  
TNGEIQKENSREALAALESPPALVRSASSDTSEELNSQDSPPKQDSTAPSSSTSSDPILDFNISLAMAKERAHQKRSSKRAPQM  
>2ksd\_90-115  
LVQKLEEMRERDLSLNVQLKDNIQA  
>2kse\_41-149  
VDELFDLQMLFAKRLSTLDLNEINAADRMAQTPNRLKHGHVDDALTAIFTHDGRMVLNDGDNGEDIPYSYQREGFADGQLVGEDDPWRFV  
WMTSPDGKYRIVVGQ  
>2ktl\_1-23  
MPRTASPGNPKSSLSGFVNPQSG  
>2kx4\_1-16  
MADFDNLFDAAIARAD  
>2kxv\_1-18  
MSVSLSKGGNVSLSKTAP  
>2ky5\_1-20  
GSSDVQYTEVQVSSAESHKD  
>2kz3\_67-83  
NGADLYEELKTSTAIL  
>2kzw\_1-48  
MNARDNKFNTWNSRGNYSYEGSDENGDGIGDSAYAVNPEAGSMDY  
>2l16\_53-78  
KEEKSAELTAVKQDKNAGLEHHHHHH  
>2l22\_105-123  
TAAPVQREP VATAPQPGAQ  
>2l2n\_1-18  
GHMMTSTDVSSGVSNCYV  
>2l4n\_71-113  
PSPQKPAQGCRCRKGASKTGKKGKSGKCKRTERSQTPKGP  
>2l4w\_1-16  
GSHMTKPAPDFGGRWK

>2I5v\_1-19  
GSQPIQLSDLQSIATMNV  
>2I6f\_140-154  
HGGSGGSGSGSGSGGS  
>2Ian\_1-17  
MGNVCVKYPLRNLSRKDR  
>2Ib5\_1-30  
MDTETWAAAARPSRDALINRITHQIRQSLE  
>2Ic0\_1-15  
MGSQKRLVQRVERKL  
>2Ic2\_1-34  
GSSGSSGNVDSNQNKASMLQARLNDEAGGTRLLR  
>2Ic5\_85-151  
DSTEDIKKAFFIDKEKNGYISASELKHVLTTLGEKLTEQEVDLLKEIGVEEGLINVDFFVKLITS  
>2Icu\_1-15  
SSGIEGCTEDEKRDS  
>2Idi\_72-106  
KSSVTLNGHKHPHSHREEGHSHSHGAGEFNKQE  
>2Iez\_121-145  
EFSLPQDSVILTAQGGMNLKGAVLT  
>2Ife\_1-32  
MHHHHHHSSQRENLYFQGLQRANSDDLVTSE  
>2Ilgq\_1-20  
GSFTMPGLVDSNPAPPESQE  
>2Igw\_72-99  
GLTGTGTGPSRAEAGSGGPGLEHHHHH  
>2Ihu\_1-81  
MGSSHHHHHHSSGLVPRGSHMHEAIGSGDLDRSAFRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPAE  
>2Ij6\_1-32  
MPSSKPLAEYARKRDFRQTPEPSGRKPRKDS  
>2Ij9\_1-77  
MGSSHHHHHHSSGLVPRGSHMAAPEGGISDVVEKSIKAEQETCAGDPVSGECVAAWDEVEELSAAASHARDKKKADG  
>2Ili\_90-124  
VDDNEKAKPKVLPFHITIYCNGGKGHFGDDCKE  
>2Im5\_1-30  
MGHHHHHHHHHHSSGHIDDDDKHMGGSGSR  
>2Imb\_16-43  
ENQEEEEERAELNQSEPEAGESSTGGP  
>2Iq3\_1-21  
MGQGQNVLGQDLEVCCCAPMT  
>2Iql\_97-113  
QPPRSPATVEAQPLPA  
>2Ite\_1-20  
MGSSHDHHHHSSGRENLYFQ  
>2Iu1\_115-149  
NWSFSSTYHTYEADYIKEQDSVYDRSPKKKYIKA  
>2Iu1\_1-25  
TSNKKILLNVDKLVDQYLLNLKNNH  
>2Iu2\_1-57  
RTMDTQNDVESAGRQSEPMEAADRQAEHPGAPTQSEMKEFQEEIKEGVEETKHEGDP  
>2Iut\_1-51  
GSKNKKKKNKGGKGGADCAEWLYGSCVANNGDCGQGMREGTCNEQTRKVKC  
>2Iuu\_1-54  
GSMKKKEKKEPKADAECSEWQYGKCVPSNGDCGNGIREATCNEQTKKTKCKVP  
>2Iuy\_74-97  
NEFSKAQRHVLDPRCQICVHSQR  
>2Iv2\_1-17  
MGHHHHHHSHMGDGEA

>2lv3\_1-16  
RRARLASPGTSRPSSE  
>2lv4\_1-23  
GPLGSGSKIKLEIYNETDMASAS  
>2lv7\_1-23  
MPFHPVTAALMYRGIYVFNLLS  
>2lva\_72-129  
PSEVEGSAANKEVLAKVIDLTHDNKDDLQAAIALSLLESPKIQADGRDLNRMHEATS  
>2lvt\_1-29  
MKPYVCIHCQRQFADPGALQRHVRIHTGE  
>2lvt\_59-83  
PYVCERCGRFVQSSQLANHIRHH  
>2lvv\_1-17  
MGCSGSKDTTNSKDGAA  
>2lw1\_1-25  
GQQEQYVALKQPAVKKTEEAAAAKA  
>2lv7\_1-61  
AERAALEELVKLQGERVRLKQKASAEIEEVAKLLKLAQLGPDESQKQFVLKTPKAL  
>2lxd\_99-125  
EFGDEDERLITRLENTQFDAANGIDD  
>2lxe\_1-28  
MGSSHHHHHSSGLVPRGSHMISLSGLT  
>2lxf\_1-24  
MAHHHHHHMGTLEAQTQGGSMQMG  
>2lxl\_164-183  
PQAGPVGIEEDNDIEENEDA  
>2lyh\_1-34  
MGSSHHHHHSSHMLVPRGSLEPSKNPLLKCKRA  
>2m03\_164-181  
NAAAESRKGQERLEHHHH  
>2m0d\_31-112  
EHKCPHCDKKNVQGNLKAHLKIHIADGPLKCRECGKQFTTSGNLKRHLRIHSGEKPVYVCIHCQRQFADPGALQRHVRIHT  
>2m28\_1-99  
QDAAQRTYGPLLNRMFQKDGRELPEEELQAAFEEDFTDQDGYIGYRELGDCMRTLGYMPTMELLEVSQHVKMRMGGFVDFEEFVELISPKLR  
EET  
>2m2b\_115-131  
NEGSRISAASVASTAI  
>2m2f\_1-26  
MGSSHHHHHSSGLVPRGSHMDDDDK  
>2m2k\_1-17  
KVQEQSVGAPPPGRADK  
>2m46\_1-24  
MHHHHHHSSGVDLGTENLYFQNSA  
>2m4v\_1-19  
SMADRVLRSRLGAVSYET  
>2m5y\_1-15  
GSSPNSEDDSSAIST  
>2m6n\_59-86  
SDGKLLKASCKIGPLPGTKKSKKNLRR  
>2m70\_1-29  
MGSSHHHHHSSGLVPRGSHMGSDLEDMK  
>2m74\_114-136  
NGGRCVAPNRCACYGFTGPQC  
>2m7a\_1-15  
GSMKRGVEMSIHDLCL  
>2m7b\_1-17  
GSMGGIKRLMEEEDAKY  
>2m7l\_98-116

STIAAGKDSITETDMQKAG  
>2m7n\_1-66  
MAEALFKEIDVNGDGAVSYEEVKAFVSKKRAIKNEQLQLIFKSIDADGNGEIDQNEFAKFGYSIQ  
>2m88\_1-16  
MPDELNVSNNPHYRPK  
>2m88\_157-194  
IIVGEGVSSKAISQKMPTDSGRNGPRSGKPLEHHHH  
>2m9h\_1-16  
GPGSLSNFANVGVGTS  
>2m9h\_82-106  
QQGYRGKNSVAARSPEKASNMENN  
>2ma0\_1-24  
MRGSHHHHHHGSHVISSIASRGSM  
>2max\_1-24  
MRGSHHHHHHGSLGVFERPIANT  
>2me8\_40-83  
APEGTESEMETSAINGNPSWHLADSPAVNGATGHSSSLDAREV  
>2mez\_88-106  
LGDVVNIKEGKLEHHHHH  
>2mf9\_1-26  
MGQPPAEEAEQPGALAREFLAAMEPE  
>2mgx\_112-131  
EEKSAATAEKNHTIPGFQQA  
>2mh9\_128-144  
NSSSVKKQKALVAKVS  
>2mii\_1-48  
GSHMVGQREPAPVEEVKPAPEQPAEQQPVPVPTVPSVPTIPQQPGPIEH  
>2miq\_1-30  
SHMSGGSPLATGTTANTRGASQKKGGEPPA  
>2miz\_1-23  
MASSRNDNESEKMQKEYKEKMKY  
>2miz\_180-200  
VPDLTPRRDEFYTKNESP  
>2mk0\_104-120  
PSAVTVPLTPTMLPSP  
>2mk0\_1-15  
SYHHHHHHTMMMPSP  
>2mk5\_1-27  
SNAKNYMDKGTSSSTVVKGKTSSAST  
>2mnw\_1-37  
GSTAIGMKETAATAAKFERQHMDSPDLGTGGGSGDDDDK  
>2mob\_1-34  
MSSAHNAYNAGIMQKTGKAFADFFAEENQVVHE  
>2moq\_1-15  
MAHHHHHHVDDDDKM  
>2mpf\_1-19  
GPSSPMELNGPLREEIVNF  
>2mpw\_95-115  
VAAGGVAGHTNGSGSENLYF  
>2mqa\_1-16  
MHHHHHHSSGLVPRGS  
>2mva\_1-19  
GPATKTEKDTLQSALPVIE  
>2mwq\_1-15  
EARPIVVGPPPPLSG  
>2mxx\_131-179  
NPAAEDKGATTASKEEAKKSEAAAKNAGKAAGKALPKTSAVKHHHHH  
>2myw\_1-24

APQDNTSMGSSHHHHHSSGRENL  
>2mz8\_1-21  
HHHHHHSSGTGSGENLYFQGH  
>2n11\_51-74  
GTRPKMTPEQMAKEMSEFLSRGPA  
>2n17\_1-17  
MAHHHHHHVDDDDKPED  
>2n1b\_1-41  
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSRSAAGT  
>2n1v\_1-22  
MGSSHHHHHHSQDPMSDQEAKP  
>2n1w\_1-18  
MGSSHHHHHHSQDPMAD  
>2n25\_1-29  
MVIHKCEDCGKEFTHTGNFKRHIRIHTGE  
>2n25\_59-112  
PYGCEECGKSYRLISLLNLHKKRHSGEARYRCEDCGKLFSTSGNLKRHQLVHS  
>2n2e\_1-21  
DKGAVKALRLQNFDVTSDISD  
>2n31\_28-58  
RITPTQQQRQVQLDAQAAQQLQYGGAVGTVG  
>2n51\_1-39  
GPGSEDDIDLFGSDNEEDKEAAQLREERLRQYAEKKA  
>2n53\_1-37  
MGSSHHHHHSSGLVPRGSHMSNKPSKPKTNMKHVAG  
>2n5x\_70-94  
GEEAGPGDPLLEAVPKTGDEKDV  
>2n6f\_113-136  
TKPKPQAESKKKKKEGKKQEKML  
>2na0\_186-205  
GEQDEEGASGRETEAAEA  
>2nbb\_92-118  
ADSSKGPDEAKIKALLERTGYTLDVTT  
>2o3d\_95-113  
PSYGRSRSRSRSRSRLE  
>2odx\_1-18  
MMADVFDTKPLDSSRKT  
>2out\_62-131  
EGGEGLSKEPAGSDEQQLRADPPSTDLNFTTVEQLKAQLTERGITFKQSATKAELIALFAPADGEKSE  
>2owi\_135-151  
PQRPTNLRRRSRSTCN  
>2rn7\_63-108  
GGLTTAERQRLKEPERENRELRNSDILRLASAYFAKAEFDRLWK  
>2rnj\_1-24  
GSSHHHHHHSSGLVPRGSHMKKRA  
>2roh\_1-24  
GSPFADPNSLALANVPLSRKRPD  
>2rop\_1-19  
SANQNFNNSETLGHQGSV  
>2rop\_91-119  
PDGAEGSGTDHRSSSHSPGPPRNQVQG  
>2rqm\_1-48  
GDIRDYNADMARLLEQWEKDDIEEGDLPEHKRPSAPIDFSKLDPGK  
>2rrl\_1-47  
HMASDDRATGPALTPLVVAAAATSAKVEVDSPPAPVTHGAAMPTLSS  
>2rsm\_1-29  
GSSGSSGKWGLRLQEKPALLPFGMAASTV  
>2rso\_1-21

GSHMESKSSSKLKENAKEEE  
>2rtx\_107-121  
RPTRSGPSSGENLYF  
>2ru8\_76-98  
VNTVSEPDSQIPPGFRGHHHHH  
>2ruj\_1-16  
GPGHMGSVSNKAPTS  
>2v6z\_1-24  
MGSSHHHHHSQDPNSSARLQVD  
>2vkc\_1-30  
MGSSHHHHHSSHMLVPRGSLQQGLHEQK  
>2vrg\_1-44  
GSHMEEPAASFQPGSMGLDKNTVHDQEHIMEHLEGVINKPEAE  
>2ytu\_1-25  
GSSGSSGMNYSYMDEKNGPPPPNM  
>3zfj\_1-16  
GGQVKKESNRVSYIDG  
>4aqz\_1-26  
MGSSHHHHHGLVPRGSHMASMTGGQ  
>4ar0\_1-24  
MKHHHHHHPMSDYDIPTTENLYFE  
>5a4g\_1-59  
MGSATVNGPHDGHDPAAAGAEKLGKRILYWRDPMVPGQRFDKPGKSPYMDMPLIPVYEE  
>5a4g\_160-180  
SLRGTVARMQETTSGLEVL  
>5abk\_68-84  
AKILKVIDKQKGPKA  
>5i1r\_1-41  
MQRGNFRNRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGK