Datasets used in the present work

1. Amyl139 (139 amyloid fibril forming hexapeptides)

RVFNIM
HYFNIF
GTFNII
HLVYIM
SQFYIT
HGWLIM
AQFIIS
NQFIIS
RQVLIF
HQLIIM
GTFFIT
HFVWIA
GTFFIN
WVFWIG
KLFIIQ
SFFFIQ
WIVIFF
NYVWIV
GTVLFM
IEMIFV
MIENIQ
IFDFIQ
YAELIV
QLENYC
MLVLFV
DCVNIT
LAVLFL
VTSTFS
ALEEYT
VSFEIV
EYSNFS
GEWTYD
EGVLYV
TYVEYI
YVEYIG
KVQIIN
AGVNYF
GVNYFL
SFQIYA
YLNWYQ
KLLIYE
FESNFN
SMVLFS
VILLIS
ISFLIF
SFLIFL
LIFLIV
NQQNQY
YYQNYQ
ATVIIE
DTVIIE

ETVIIE
FLVHSS
FTVIIE
FYLLYY
GTVIIE
ITVIIE
KDWSFY
KIVKWD
LLYYTE
LTVIIE
LVEALY
LYQLEN
MTVIIE
NFGAIL
NHVTLS
NTVIIE
QTVIIE
SAVIIE
SDVIIE
SEVIIE
SFVIIE
SGVIIE
SIVIIE
SLVIIE
SMVIIE
SNVIIE
SQVIIE
SSVIIE
STEIIE
STFIIE
STLIIE
STNIIE
STQIIE
STSIIE
STVEIE
STVFIE
STVIFE
STVIIA
STVIID
STVIIE
STVIIF
STVIII
STVIIL
STVIIM
STVIIN
STVIIQ
STVIIS
STVIIT
STVIIV
STVIIW
STVIIY
STVIYE
STVLIE
STVNIE
STVTIE
STVWIE

STVYIE
SVVIIE
SYVIIE
TTVIIE
VEALYL
VQIVYK
VTVIIE
WTVIIE
YTVIIE
YLVNFT
FIVNIV
LYVLIV
GLCYFN
NFVNYS
VVFFIC
MCSIFQ
GIFNIK
HICNFF
KCWCFT
QQSLFQ
FVFYIF
MHVNIQ
NHFCIQ
VNQLYQ
KCLNYL
VIFTIM
AVFIIY
GLSIFV
YLLWYC
VTFTIQ
WTVNYS
GVWWFF

2. Amor168 (168 amorphous β-aggregating hexapeptides)

RMFNII
GMFNIQ
KMFFIQ
WYFYIQ
FMFFII
HAFLII
MYWIIF
HIFIIM
MYFFIF
NIVLIM
KAFIIQ
EECLFL
KAILFL
ECFFFE
DECFFF
NYNTYR
KAILFL
TNELYM
YLEIII
FGELFE

DADLYL
KENIIF
DDSLFF
ILENIS
TYQIIR
AINKIQ
FFWRFM
AEKLFD
AAELRN
EKNLYL
YASEIE
KECLIN
AARRFF
RLVFID
HKALFW
TAELIT
FLKYFT
TFWEIS
WVENYP
AEMEYL
EYLKIA
RETWFF
RVAFFE
KLLEIA
NYFAIR
VAWLKM
AEVLAL
ETWFFG
TAWYAE
LMSLFG
ICSLYQ
RGFFYT
KGENFT
SSNNFG
VTQEFW
YTFTIS
YAELIV
NIQYQF
QAQNQW
DILTYT
EECLFL
LAEAIG
TLKNYI
QTNLYG
KAILFL
YLEIII
KENIIF
DDSLFF
ILENIS
TYQIIR
RLVFID
AENGKS
DIEVDL
DLLKNG
EALYLV
EHSDLS

EKDEYA
ERGFFY
ERIEKV
EVDLLK
FFYTPK
FHPSDI
FSKDWS
FTPTEK
FVNQHL
FYTPKT
GERGFF
GFFYTP
GSHLVE
HLVEAL
HPAENG
IEKVEH
IQRTPK
IQVYSR
KSNFLN
KTVIIE
KTVIIT
KTVIVE
KTVIYE
KTVLIE
KVEHSD
KWDRDM
LIAGFN
LKNGER
LSFSKD
LSQPKI
NGERIE
NGKSNF
NLGPVL
PKIQVY
PSDIEV
PTEKDE
PTVIIE
QPKIVK
RGFFYT
RTPKIQ
RVNHVT
SDLSFS
SGFHPS
SHLVEA
SLYQLE
SPVIIE
SRHPAE
STAIIE
STDIIE
STGIIE
STIIIE
STMIIE
STPIIE
STTIIE
STVAIE
STVDIE

STVGIE
STVIAE
STVIDE
STVIEE
STVIGE
STVIIP
STVILE
STVIME
STVINE
STVIPE
STVIQE
STVISE
STVITE
STVIVE
STVIWE
STVMIE
STVPIE
STVQIE
STVSIE
STVVIE
STYIIE
SWVIIE
TEFTPT
VKWDRD
VTLSQP
VYSRHP
WSFYLL
YQLENY
YVSGFH
YYTEFT
VFFFIG
SFFCII
YLVLIM
NEFIIT
SLFLIG
MIFFIY

3. Hex40 (40 amyloid fibril forming hexapeptides that are different from those in Amyl139).

NNQQNY
NNQNTF
SNQNNF
GGVVIA
MVGGVV
VQIINK
SSTNVG
SSTSAA
GVATVA
GSIAAT
KAKEGV
KTKQGV
KTKEGV
KTKEQV
KTVEGA

KNEEGA
STVIIG
STWIIE
KTVIIE
RTVIIE
HTVIIE
GDCFIL
GNCFIL
GYCFIL
NFGSVQ
QGVCFR
SQAIIH
DIDLHL
IFQINS
TFQINS
ILQINS
GNVCIN
GNTCVN
GTLVFM
ASSSNY
QEMRHF
SYSTMS
KHIIVA
STMSIT
QANKHI

4. Amyl310: Amyloid fibril forming peptides of lengths 7–71 residues.

AEVSIVV
AIAAIVF
DAQAIAL
DFVIFLD
EFQEICA
FNGLIVV
GILNIVV
GLINIAT
GNNQQNY
GQVIWVN
GTYSFYT
GVVAIGC
IIHFGSD
KLVFFAE
LASLIYR
MIHFGND
MMHFGND
NFLVHSS
NHVTLSQ
NLSLFDQ
QQEVINK
SLLNIVV
TAFTILA
TLSIYQY
TVIAFLA
VGDIVIF
VQQNYQA

YGNLISL
AFGAILSS
AGAAAAGA
DLSFMKGE
FLIFLIVG
GKVQIVYK
NAGDVAFV
NFAAILSS
NFGAALSS
NFGAIASS
NFGAILSS
NFGSVQFV
SELNIYQY
SNNFGAIL
SNTFINNA
TNVGSNTY
YHIRLSFN
GETYVVTL
KKLVFFAED
LVEALYLVC
NFLVHSSNN
PAGGYYQNY
PQGGYQQYN
QQGGYQQYN
QRLANFLVH
RQGNINIVA
SLYQLENYC
VGGAVVTGV
CPLMVKVLDA
DIQKVAGTWY
DLSFSKDWSF
DVKVGDIVIF
GGYYQNYQGY
GNNQQNYQQY
GYLTVAAVFR
KDWSFYLLYY
KKLVFFPED
KYLLFCMENS
LANWMCLAKW
NDFQKQQKQA
NFNYNNNLQG
NLQGYQAGFQ
PDAGYQQQYN
PQGGRGNYKN
PQGGYQQQFN
RCELARTLKR
RLANFLVHSS
RYQGYQAYNA
SMVLFSSPPV
SNNFGAILSS
SSPPVILLIS
SSTSAASSNY
TEFTPTEKDE
VHDCVNITIK
YLLYYTEFTP
YNNNLQGYQA

YQNYQGYSGY
YQQGGYQQYN
YSGYQQGGYQ
YSQNGNQQQG
AIIGLMVGGVV
CATQRLANFLV
CPLMVKVLDAV
FSMLVAAIQSA
GAVVTGVTAVA
GYANNFLFKQG
HHQKLVFFAED
ILLISFLFIFL
MSTYTGIFTDQ
RTVKKNIIEEN
SFFSFLGEAFD
SFNNGDCFILD
SFNNGNCFILD
SFNNGYCFILD
VALRIQNVRFS
VVLDDKDYFLF
YTIAALLSPYS
AASIKVAVSADR
DISTKYFQMSLE
GQQQQQQQQQQG
KFFEAAAKKFFE
LSNIDYILIKAS
MKVIFLKDVKG
NFGAILSSTNVG
SAKVDAIGLEIV
SLVCQCLVRTPW
VILQQSAADIAR
VTGVTAVAQKTV
VTVKVDAVKVTV
VTVKVEAVKVTV
VTVKVNAVKVTV
WQQVLNVWGKVE
DWSFYLLYYTEFT
DYFLFRDGDILGK
EDVAVYYCHQYYS
IPAGVDAVKISM
KLKLKLELELELG
LGVVGSTTTQLYT
NHAIVQTLVNSVN
VHHQKLVFFAEDV
AEFHRWSSYMVHWK
AVVGGLGGYMLGSA
DCVNITIKQHTVTT
DFVIFLDVKHFSPE
DNNNSANYQVSTSK
DWSFYLLYYTEFT
EQVTNVGGAVVTG
ISKLEYSNFSVRY
MKHMAGAAAAGAVV
PGGGKVQIVYKPV
RVEKVAILGLMVLA
SFFSFLGEAFDGAR

VGDIVIFNDGYGVK
VNITIKQHTVTTTT
AGAAAAGAVVGGLGG
FNNGDCFILDLGNNI
FNNGNCFILDLGNNI
FNNGYCFILDLGNNI
MLSNTTAIAEAWARL
RSFFSFLGEAFDGAR
VHHQKLVFFAEDVGS
VPHQKLVFFAEDVGS
AVVGGLGGYMLGSAMS
DCVNITIKQHTVTTTT
DVAGHGQDILIRLFKS
DWWKVEVNDRQGFVPA
DYFLFRDGDILGKYVD
EVHHQKLVFFAEDVG
GSTDYGILQINSRWWS
GSTDYGTLQINSRWWS
LEVLLGSGDGSLVFV
VHHPKLVFFAEDVGS
VHHQEKLVFFAEDVPS
VHHQPLVFFAEDVGS
VHPQKLVFFAEDVGS
VTNVGGAVVTGVTAVA
AKKENIIAAAQAGASGY
DIKIMERVVEQMCTTQY
GEWTYDDATKTFTVTE
LSQTFVYGGSRAKRNN
RLLIVVVVVVLVVVVIV
TKRPRFLYEIAMALNSD
VHHQEKLVFFAEDPGS
VHHQEKLVFFAEPVGS
YEVHHQKLVFFAEDVGS
AGAVVGGLGGYMLGSAMS
DFVIFLDVKHFSPEDLTV
FSMLVAAIQSAGLTETLN
FVIFLDVKHFSPEDLTVK
FVNVEAVKAFLEAHGIAY
FVNVQAVKVFLESQGIAY
GNLPFLGTAGVAGEFPTA
GSTTTQLYTDCTEKLRPE
GSTTTQLYTDRTEKLRPE
GVGKSALTIQLIQNHFVY
IKYLEFISQAIIHVLHSR
KFVIFLDVKHFSPEDLTV
MPEEELLNAPGETYVVTL
RGGRLCYCRRRFCVCVGR
TCFFRDHSYQEEAAVELK
TVNGVGEVTATAVQGVAV
DDQQQQQQQQQQQQQQKK
EQVTNVGGAVVTGVTAVA
NLKHQPGGGKVQIVYKEVD
STNVKTAFEMVILDIYNNV
VLSEGEWQLVLHVWAKVEA
EQVTNVGGAVVTGVTAVAQK
EVVPHKKMHKDFLEKIGGL

GEKIVFKNNAGFPHNVVFDE
GYEVHHQKLVFFAEDVGSN
KETAAAKFERQHMDSSTSAA
RQGVEDAFYTLVREIRQHK
TESKEKITQYIYHVLNGEIL
WNFAGIEAAASAIQGNVTSI
DILTLLNSTNKDWWKVEVND
DWSFYLLYYTEFTPTGKDEYA
GEYTVDVADKGYTLNIKFAGD
GVVGWVKNTSKGTVTGQVQG
PFTAATLEEKLNKIFEKLGMY
QISFADYNLLDLLRIHQVLN
SGYEVHHQKLVFFAEDVGSNK
STAAMSTYTGIFTDQVLSVLK
TNVKHVAGAAAAGAVVGGLGG
VTIKANLIFANGFTQTAEFKG
DDNNQQNYQQYSQNGNQQQGKK
DWSFYLLYYTEFTPTGKDEYA
ELEAQVINTFDGVADYLQTYHK
ELSAQVINTFDGVADYLQTYHK
ESLAQVINTFDGVADYLQTYHK
GTVSFVTSPHQGAGMVGKVTVN
KLSAQVINTFDGVADYLQTYHK
KSEAQVINTFDGVADYLQTYHK
LGVVGSTTTQLYTDCTEKLRPE
LGVVGSTTTQLYTDHTEKLRPE
LGVVGSTTTQLYTDRTEKLRPE
LGVVGSTTTQLYTDSTEKLRPE
LLSAQVINTFDGVADYLQTYHK
SLEAQVINTFDGVADYLQTYHK
SLSAQVINTFDGVADYLQTYHK
SNFLNCYVSGFHPSDIEVDLLK
YEQLDEEHKKIFKGIFDCIRD
KSSTEEKYNEKGELSEKKITRAD
QNNFVHDCVNITIKQHTVTTTTK
AEAIGNVAVAGELPVAGKTAVAGR
DSGYEVHHQKLVFFAEDVGSNKG
GAAVPGVLGGLGALGGVGIPGGVV
SAPNLATLVKVTTNHFTHEEAMMD
GLVGAAGLGGLGVGGLGVPGVGGLG
IKNESKYSNTFINNAYNMSIRRSMA
KGTFEKATSEAYAYADTLKKDNGEY
AVRLSLGAQNLASSQIQRNPVLITLG
CHQKLVFFAEDYNGKDEAFFVLKQHW
CHQKLVFFAEDYNGKHQKLVFFAEDC
CHQKLVFFAEDYNGKHQKLVFFAEDW
HDSGYEVHHQKLVFFAEDVGSNKGA
DILTLLNSTNKDWWKVEVNDRQGFVPA
CHQKLVFFAEDDYNGKHQKLVFFAEDSC
DAEFRHDSGYEVHHQKLVFFAEDVGSNK
ETDVKMMERVVEQMCVTQYQKESQAYYD
KPFIARFEGRLFSRSDELKAIIKELTGE
KPFLARFRGRIFSRSDELRTLIAAFTGE
KPFLARVEGRIFSRSDELRAYITAYTGE
KPFTARFEGRIFSRSDELRALITEITGE
KPFTARISGRLFSRSDELKTIIATITGE

KPYIARFEGRLFSRSDELRAVIEAHTGE
VHHQKLVFFAEDVGSNKAIIGLMVGGVV
GLSDGEQQVLVWIVNVWGKEADAGHGQEL
GLSDGEQQVLWINVWGKVEADAGHGQEVL
GLSDGEWQQVENVWGKVEADIAGHGQLVL
GLSDGEWQQVLENVWGKVEADIAGHGQVL
GLSDGEWQQVLNVWGKVEADIAGHGQEVL
HSQGTFTSDYSKYLDSRRAQDFVQWLMNT
EGEWQLVLHVWAKVEADVAGHGQDILIRLFK
YHSRALQVVRARKQIVAGVNYFLDVELGRTTCT
YHSRALQVVRARKQIVAGVNYFLDVEQGRTTCT
EASNCFAIRHFENKFAVETLICFNLFLNSQEKHY
EASNCFAIRHFENKFAVETLICSRTVKKNIIEEN
DAEFRHDSGYEVHHLKLVFFAEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIVLMVGGVA
DAEFRHDSGYEVHHQKLVFFAQDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQTLVFFVEDVGSNKGAIIGLMVGGVV
DAEPRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
DAAFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFAADVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFADDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFAEDMGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLLVGGVV
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMPGGVV
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGPVV
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVPGVV
DAEFRHDSGYEVHHQKLVFFAEDVGSSKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFAEYVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFAPDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVFFMEDVGSNKGAIIGLLVGGVV
DAEFRHDSGYEVHHQKLVVFVEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKPVFFAEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQTLVFFAEDAGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHRKLVFFAEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHPQKLVFFAEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHYQKLVFFAEDVGSNKGAIIGLMVGGVV
DAEFRHNSGYEVHYQKLVFFAEDVGSNKGAIIGLMVGGVM
DVEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
DVEFRHDSGYEVHHQKLVFFAEDVGSNKGALIGLLVGGVV
DVEFRHDSGYEVHHQKLVFFAGDVGSNKGAIIGLMVGGVV
GAEFRHDSGYEVHHQKLVFFVEDVGSNKGAIIGLMVGGVV
DAEFRHDSGYEVHHQKLVAFFEDVGSNKGAIIGLLGGVVVIA
DAEFRHDSGYEVHHQKLVAFFEDVGSNKGAIIGLLVGGVVIA
DAEFRHDSGYEVHHQKLVFFAEDVGSNDGAIIGLLVGFVVIA
DAEFRHDSGYEVHHQKLVFFAEDVGSNDGAIIGLLVGGVVIA
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLLVGGVVIA
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA
DAEFRHDSGYEVHHQKLVFFAEKVGSNDGAIIGLLVGGVVIA
DAEFRHDSGYEVHHQKLVFFAEKVGSNKGAIIGLLVGGVVIA
DAEFRHDSGYEVHHQKLVGFAEDVGSNKGAIIGLLVGFVVIA
DAEFRHDSGYEVHHQKLVGFAEDVGSNKGAIIGLLVGGVVIA
NLKHQPGGGKVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVE
KKQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQKK
QINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNA
SYGGEGIGNVAVAGELPVAGKTAVAGRVPIIGAVGFGGPAGAAGAVSIAGR

GQGGGTHNQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPMMHFGNDW
GQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDY
E
MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPP
PPPPPKK

5. AmylCryst15 (15 hexapeptides whose amyloid microcrystal structures are available at PDB.

AIIGLM
AILSST
GAIIGL
GDVIEV
GGVLVN
GYMLGS
GYVLGS
HSSNNF
IIHFGS
KLVFFA
LSFSKD
MIHFGN
MMHFGN
NFLVHS
NKGAII