

Table S2: The GASA protein sequences of Pineapple, *Arabidopsis*, Maize, Tobacco, Grape, and Rice.

Pineapple GASA proteins

>AcGASA1

MKLLPYTLALILLLLLTSSYLQATMALSDFCNSKQCGRCAKAGKTGRCMRYCGMCCEQCK
CVPSGTYG NKDECPCYRDMVSWKTKRPKCP*

>AcGASA2

MANKARFLSPVLLLLLLLVEMNIFDGPLVQGVGIDCAAKCDYRCSKAGWHKACIRACNTC
CERCHCVPPGTSGNKDVCPCYANMTTHGGRPKCP*

>AcGASA3

MRDDCPAKCAYRCSKSGRPKMCLRACNTCCQRCHCVPPGTSGNTEKCPCWAHMKTHGG
RPKCP*

>AcGASA4

MAMAYKLFILVAFALLAISMAEDKVLAQGTDHVEGEYHLEGSNGQGLRSYQCPGQCSRR
CAQTQYKKPCLFFCQKCCMKCLCVPSGYGNGKGEPCYNNWTKRGGPKCP*

>AcGASA5

MVVAAGFSRFSCLLLLLLLALSFSESMGGNLNPSCNPRCSYRCSGTSHKKPCMSFCMQC
CAQCLCVPPGTYG NRQVCACYNWTKRGGPKCP*

>AcGASA6

MINCAGLCAGRCSQH SRPNLCARACGTCCSRCKCVPPGTSGNREMGTCYTNMTTHGN
RTKCP*

>AcGASA7

MALKLALFLLACFLLA STRVSSDSEEAFTAA MAPYVPATPPALAPAPHKIIDTKECVGACK
ERCSLH SRKNVCSRA CLTCCSVCKCVPAGTAGNKETCGKCYTDWQTHGNRTKCP*

>AcGASA8

MATTKQTKEAVHQTLHMDLERLLLQYLQQRANATRR AQKTTQQMFP EAQLRDNGMLIA
KHCPSP TSQP PAISPTPQPIKFPMYGVTPGSLHPQEC PDRCAGRCSATAYKKPCMF FCQK
CCEKCLCVPPGTYG NKQFCPCYNDWTKRGGPKCP*

>AcGASA9

MKLLYTALPLLLLLLLLLLASDHFQATMAGSAFCNGKCKVRCSKASEQERCLKYCGLCCEE
CRCVPPGTYG NKDRCPCYRDKFTGHGRRRRPKCP*

>AcGASA10

MKPLYITTTTTLAFLLLLLLLASSYLQTS MAGSAFCDGKCKVRCSKASVQDRCLRYCGLCCA
ECNCVPSGTYG NKDECPCYRDKYTG TGLRRRSKCP*

>AcGASA11

MKLFSFSLALMMLLLASAYLQTFINASEWCEGKCSERCSKAGARRRCVEYCNLCCKE CNC
VPSGTAGNKDECPCYRDKRNSK GKPKCP*

>AcGASA12

MDSGRNKKLVSWFIVLFLFMQVVAEASLHNFIGLEDDKEGNATPADGFSRRPIINCNFACA
RRCMKASRKNVCARACGTCCIKCQCVAGTSGNKTMCPCYANLRTHGLKPKCP*

>AcGASA13

MNRDRSPKISKPNRNSLA EIIGGANASHYLP AEKAQFLQKYRCSATSHKKPCLFFCNECCQ
KCHCVPSGTYG HKEEPCYNNWTKEGKPKCP*

>AcGASA14

MATTQLILLTLLAFSFTLHHLAVEAHEEVEHQLASKGARTLLATMDCNSACSGRCQLSSR
LNYCMRACKTCCARCSCVPPGTSGNYDQCPCYASLTTRSSKKKCP*

>AcGASA15

MNEGSAIDCGGACAERCKESSRPNLCKRACGTCCRRCSVPPGTYGNYEVCPCYAAITTR
GGRKKCP*

***Arabidopsis thaliana* GASA proteins**

>AtSA1

MAISKALIASLLISLLVLQLVQADVENSQKKNGYAKKIDCGSACVARCRLSRRPRLCHRACTCCYRCNCVP
PGTYGNYDKCQCYASLTTHGGRRKCP

>AtSA2

MAVFRSTLVLLLIIVCLTTYELHVHAADGAKVGEGVVKIDCGGRCKDRCSKSSRTLCLRACNSCCSRCNCV
PPGTSGNTHLCPCYASITTHGGRLKCP

>AtSA3

MAIFRSTLVLLILFCLTTFELHVHAAEDSQVGEGVVKIDCGGRCKGRCSKSSRPNLCLRACNSCCYRCNCV
PPGTAGNHHLCPCYASITTRGGRLKCP

>AtSA4

MAKSYGAIFLLTLIVLFMLQTMVMASGSGSNVKSQKRYGPGSLKRTQCPSECDRRCKKTQYHKACITFCN
KCCRKCLCVPPGYGNKQVCSCYNNWKTQEGGPKCP

>AtSA5

MANCIRRNALFFLTLLFLLSVSNLVQAARGGGKLPQQCNSKCSFRCSATSHKKPCMFFCLKCKKCLCV
PGTFGNKQTCPCYNNWKTKEGRPKCP

>AtSA6

MAKLITSFLLLTILFTFVCLTMSKEAEYHPESYGPSLSYQCGGQCTRRCSNTKYHKPCMFFCQKCCAKCL
CVPPGTYGNNKQVCPCYNNWKTQGGGPKCP

>AtSA7

MKIIVSILVLASLLLISSSLASATISDAFGSGAVAPAPQSKDGPALKWCGQKCEGRCKEAGMKDRCLKYCGI
CCKDCQCVPSGTYGNNKHECACYRDKLSSKGTPKCP

>AtSA8

MKLVVVQFFIISLLLTSSFSVLSSADSSCGGKCNVRCSKAGQHEECLKYCNICQKCNVPSGTFGHKDECP
CYRDMKNSKGGSKCP

>AtSA9

MKKMNVVAFVTLIISFLLLSQVLAELSSSSNNETSSVSQTNDENQTAAFKRTYHHRPRINCGRHACARRCSKT
SRKKVCHRACTCCAKCQCVPPGTSGNTASCPCYASIRTHGNKLKCP

>AtSA10

MKFPVAVKVLIIISLLITSSLFILSTADSSPCGGKCNVRCSKAGRQDRCLKYCNICCEKCNVPSGTYGNNKDEC
PCYRDMKNSKGTSKCP

>AtSA11

MAVFRVLLASLLISLLVLDFVHADMTSNDAPKIDCNSRCQERCSLSSRPNLCHRACTCCARCNCVAPG
TSGNYDKCPCYGLTTHGGRRKCP

>AtSA12

MMKLIVVFVISSLLFATQFSNGDELESQAQAPAIHKNNGGEGSLKPEECPKACEYRCSATSHRKPCLFFCNKC
CNKCLCVPSGTYGHKEECPCYNNWTTKEGGPKCP

>AtSA13

MALSLLSVFIFHHVFTNVVFAASNEESNALVSLPTPTLPSPSPATKPPSPALKPPTPSYKPPTLPPTPIKPPTTKP
PVKPPTIPVTPVKPPVSTPPIKLPPVQPPTYKPPTPTVKPPSVQPPTYKPPTPTVKPPTTSPVKPPTTPPVQSPP
VQPPTYKPPTSPVKPPTTTPPVKPPTTTPPVQPPTYNPPTTPVKPPTAPPVKPPTPPPVRTRIDCVPLCGTRC
GQHSRKNVCMRACVTCCYRCKCVPPGTYGNKEKCGSCYANMKTRGGKSKCP
>AtSA14
MKMPVVVVQFFIISLLLTSSFYVLSSADSSACGGKCSVRCSKADRTHEECLEDCDICCQKNCNCPVSGTYGNKD
ECPCYRDMKNKSGGSKCP
>AtSA15
MATKLSIIVFSIVVLHLLLSAHMHFLINVCAECETKSAIPPLLECGPRCGDRCSNTQYKKPCLFFCNKCCNKCL
CVPPGTYGNKQVCPCYNNWTKSGGPKCP

Maize GASA proteins

>ZmGASA1
MTMTTMKKQQQLLLLLSLMFLVAVTAAAVAADPHPQQVQVQQQQQAQMRINRATRSL
LPQ
PPPKLDCPSTCSVRGNNWKNQMCNKM CNVCCNKCSCVPPGTGQDTRHLCPCYDTML
NPH
TGKLKCP*
>ZmGASA2
MAQASSFSIVLLFLALVLVVEVSAGTANEELYRPAGAEGSVPIEQCPEKCDYRCSATSY
KKPCLFFCNKYCCNKCLCVPSGTYGNKEECPCYDNMKTQGGPKCP*
>ZmGASA3
MCMCSAATLPCCHVGRHGHQLHCTAPPHSYIVYCCCAHSLYICITACFWNQERRSPLHSR
RHPPLPLL FLEMTMTTMKKQQQLLLLLSLMFLVAVTAAAVAADPHPQQVQVQQQQQAQ
M
RINRATRSLLPQPPPKLDCPSTCSVRGNNWKNQMCNKM CNVCCNKCSCVPPGTGQDTR
H
LCPCYDTMLNPHTGKLKCP*
>ZmGASA4
MAKPPLQTAAILLVLLAAASCLHTVDAATLGFCWGKCSVR CAGATARQARAACMSSCGL
CCEACNCVPHDIHDCPCYRNMLTAGPKKRPKCP*
>ZmGASA5
MAAASGRAPSACALLLLFLLLVVGAAAAAVIVVDANRGEQE QDWDWEQLSAA SPSPWSP
A
PAPAPSPVSFIDCGSACGAR CALSSRWNL CRRACGSCC ARCN CVPPGTAGNH DVCPCYAA
ITTRGGRPKCP*
>ZmGASA6
MAMAKPPLQTVTIILLVLLTAASWLHTVDASALGFCWGKCSVR CVQATAGQARAACMS
Y
GLCKACNCVPHDIHDCPCYRNMP TVDPKKRPKCP*
>ZmGASA7
MAKASSRLLFSLSLVLLLLVETTTSPHGQADAIDCGASCSYRCSKSGRPKMCLRACGTC
CQRGCVPPGTSGNEDVCPCYANMKTHDGQHKCP*
>ZmGASA8

MAVAKPPLQTA AVL L L L L L L V V A A S W L Q T V D A A S G F C S S K C S V R C G R A A S A R A R G A C M R S
C G L C C E E C N C V P T R P P R D V N E C P C Y R D M L T A G P R K R P K C P *

>ZmGASA9

MAKPPLQTAAI L L L L L L A A S C L H T V D A A A L G F C W G K C S V R C A G A T A R Q A R A A C M S S C G
L
C C E A C N C V P R D I H D C P C Y R N M L T A G P K K R P K C P *

>ZmGASA10

MRINRATRSLPQ P P P K L G T T E N C R Q Q D T R S H F I D A C F S R N T D C P S T C S V R C G N N W K N Q M
C N K M C N V C C N K C S C V P P G T G Q D T R H L C P C Y D T M L N P H T G K L K C P *

>ZmGASA11

MTMTTMKKQ Q Q L L L L L S L M F L V A V T A A A V A A D P H P Q Q V Q V Q V Q Q Q Q A Q M R I N R A T R S
L L
P Q P P P K L D C P S T C S V R C G N N W K N Q M C N K M C N V C C N K C S C V P P G T G Q D T R H L C P C Y D T
M L N P H T G K L K C P *

>ZmGASA12

MTMTTMKKQ Q Q L L L L L S L M F L V A V T A A A V A A D P H P Q Q V V Q V Q Q Q Q A Q M R I N R A T R S L
L P
Q P P P K L D C P S T C S V R C G N N W K N Q M C N K M C N V C C N K C S C V P P G T G Q D T R H L C P C Y D T M
L N P
H T G K L K C P *

>ZmGASA13

MTMTTMKKQ Q Q L L L L L S L M F L V A V T A A A V A A D P H P Q Q V Q V Q V Q Q Q Q A Q M R I N R A T R S
L L
P Q P P P K L G T T E N C R Q Q D T R S H F I D A C F S R N T D C P S T C S V R C G N N W K N Q M C N K M C N V C C
N K
C S C V P P G T G Q D T R H L C P C Y D T M L N P H T G K L K C P *

>ZmGASA14

MASRNKAAALLCFLFLA A V A A S A A E M I A G S G I G D G E G E E L D K G G G G G G G H H K H E G Y K
N K
D G K G N L K P S Q C G G E C R R R C S K T H H K P C L F F C N K C C A K C L C V P P G T Y G N K E T C P C Y N N W
K
T K K G G P K C P *

>ZmGASA15

MKPAATARVAGLLFFLV L L L A L P S L R V S M A G S G F C D G K C A V R C S K A S R H D D C L K Y C G I C C
A T C N C V P S G T A G N K D E C P C Y R D M T T G H G N R T R P K C P *

>ZmGASA16

MAPSKLAVVVALVAS L L L L T S N T K L G L F V L G Q A A P G A Y P P R A P P P H Q I V D L A K D C G G A C
D V R C G A H S R K N I C T R A C L K C C G V C R C V P A G T A G N Q Q T C G K C Y D W T T H G N K T K C P *

>ZmGASA17

MAKPPLQTATI L L L L L V A A S C L H T V D A A A L G F S W G K C C V R C A R T T T R R A R E A C M S S C G L
K C E F C K C G P A M P P R D I H D C P C Y R N M L T A S P K K R P K Y P *

>ZmGASA18

MASRNKAAALLCFLFLA A V A A S A A E M I A G S G I G D G E G E E L D K G G G G G G G H H K H E G Y K
N K

DGKGNLKP SQCGGECRRRCSKTHHKPCLFFCNKCCAKCLCVPPGTYGNKETCPCYNNW
KTKKG GPKCP*

Tobacco GASA proteins

>NtGASA1

MKSNPFKLIK YFMKSNFIDLHYFIHMKILNFSCYTSNAISEAAYSYPKIDCGGACKARCRLSS
RPRLCKRACGTCCARCNCVPPGTSG>NtETCPCYANMTTHGNRRKCP

>NtGASA2

MAISKTLFVSLVLSLLLLDQVQSIQTDQVTSNAISEAAYSYPKIDCGGACKARCRLSSRPRLC
KRACGTCCARCNCVPPGTSG>NtETCPCYANMTTHGNRRKCP

>NtGASA3

MAISKLILVAMVFFSLLVLHLVEADNQLVV>NtDATESFYTPKLDCGAACEARCLASRQKIC
KRACGTCCGRCNCVPPGTSGNQELCPCYFAMTTHGGKRKCP

>NtGASA4

MAISKLILVAMVFFSLLVVHLVEADNQVVVNKD ATKSSYTPKLDCGAACEARCLASRQKI
CKRACGTCCARCNCAPPGTSGNLELCPCYFAMTTHGGKRKCP

>NtGASA5

MDISKSKALIVSLIFLLFLLSAAENHELSSNNIDAISKALKARPNYTINCGKECTRRCKLASR
QKMCMRACGTCCARCNCVPPGTSGNENICPCYSTMTTHGNRRKCP

>NtGASA6

MPIKRPALPHFNHILNPQENMKFFTLVFIAILLIQVFTEAVSINNAEDTAAQIEKAGNDGAL
FKKSHHHPIRKINCGYACARRCRKSSRKNVCKRACKSCCARCHCVPRGTYGNKEACPCYA
RLKTHGNRPKCP

>NtGASA7

MKIFTLVFIAILLIQVFAEAVSFNNAEDTAAQIEKAGNEGTLFKKIHHPIRKINCGHACARR
CRKSSRKNVCKRACKSCCARCHCVPPGTYGNKEACPCYARLKTHGNRPKCP

>NtGASA8

MAFQKAFAALLIASLVLVHFTHALQQGNNSKPPAPSPQAPKPLGNITFTTYIFPPRLIYNHITI
ILNSLTNLSEFDYVADCTGACEYRCSESRPNLCNRACGSCCRTCHCVPPGTSGNYEACPCY
FNLTHNDTRKCP

>NtGASA9

MAMAIRLVFVMALLLFLGVKAEVSLTDPKVEEDKSQHFGLSQAFRVFTRGANRRLVQGV
VLKLVKYLNNGDLAVAPAPAPHPSQLDCGGLCKYRCSLHSRPKVCIRACGTCCLRCKCVPP
GTFGNREMCGKCYTEMTHGNKTKCP

>NtGASA10

ARQHFDASFLAKPPVGPTTCPVPTAQCSSACDQRCSATSHKNNCLMFCNMCCNWCQCVP
PGTFGQKECCSCYNDWKTEQGTPKCP

>NtGASA11

MAHSSLFECSSACDQRCSATSHKNNCLMFCNICCNWCQCVP PGTFGQKECCSCYNDWKT
EQGTPKCP

>NtGASA12

MAYNARLLFLSMFLVLITFSNVVEGYKKLRPEDCEPKCKYRCSATSHKKPCLFFCKKCCAKC
LCVPPGTHGNKETCPCYNNWKTKEGGPKCP

>NtGASA13

MEKMPCLMLLPLLIIMLLLVGTHAKITESAPQPQPP>NtFPMNGTTPGSLHPQECLPRCTYR

CSKTQYKKPCMFFCQKCCAKCFCVPPGTYGNKQFCPCYNNWTKRGGPKCP
>NtGASA14
MKLCFATLLVVTLVLTSSFIQTTVAGSDFCDKCKIRCSKAGRQDRCLKYCGICCNQCQVPS
GTYGNKDECPCYRDKKNSKGKPKCP
>NtGASA15
MKLLLATLLLFTLVLTSPFIQTTMAGSSYCDKCKLRCAKAGVMDRCLKYCGICCEECKCVP
SGTYGNKHECPCYRDKKNNKGKPKCP
>NtGASA16
MKRIFVAFMLVFALLLTSSFLETATAKSVYCARKCKARCSKAGVKDRCVKYCELCCAKCKC
VPTGTYGNKHQCPCYRDMKNFKGKPKCP
>NtGASA17
SCDSKCAVRRGKAGIAKRCLTYCGIYCNKCNVPSGNYGNKSECPCYRDMLNSKGKSKCP
>NtGASA18
MKLAMITLLIVSLVLTSSFLQSAVAYDDPSSCDKCAVRCGKAGIKKRCLTYCGICCNKCN
VPSGNYGNKSECHCYRDMLNSKGKSKCP

Grape GASA proteins

>VvGASA1
MGIRVLLLVTIMLFCAIEASSDDNNIFEHQHNQAVKGSAGRLLPFLDCGGLCKERC SLHSR
PNVCTRACGTCCVRCKCVPPGTYGNREMCCTCYTEMTHGNKPKCP
>VvGASA2
MKHLFPTLLLLSLLHSCFSQPTTDGAGFCGLKCSKRC SQA AVLDRCMKYCGICCQECKCV
PSGTYGNKHECPCYRDKKNSKGKPKCP
>VvGASA3
MPAPMKGFCDKCGVRCANAGVYDRCVKYCGICCQECKCVPSGTYGNKSECPCYRDKLN
SKGKPKCP
>VvGASA4
MLSLSMMLLLLQNNATITEAPTPQPQQSTNGFPMHGV TQGS LHPQECAPRCTTRCSKTA
YKKPCMFFCQKCCAKCLCVPPGTYGNKQFCPCYNNWTKRGGPKCP
>VvGASA5
MDLFIGHLTEFHGNLSSILDILEREKQVKTELHRYLSPILNHILERGVNYLQLLTPGLFFSP
CSAMALKILLLLLASYLLVTKRVSANDEEFGVQATYAKAPVPAPVKTPIPAPPVKPPTVTPNP
PAP>VvKPPT>VvPKPPAPPVSPPTVAPKPPAP>VvPKPPTPPANPPTVAPKPPAPPVKPPT>Vv
PKPPSPPPVNPPTPKPPASPVKPPTVAPKPPVPPVTPPTAPMPPVRARLDCIPLCDQRCKAHS
RKNICVRACMTCCDRCKCVPPGTYGNREKCGKCYTDMTHGNKPKCP
>VvGASA6
MAKVFAFLLLALLAISMLHTTVLASHGHGGHHYDQKNYGPSLKS FQCPSQCSRRCGKTQ
YHKPCMFFCQKCKKCLCVPPGYGNKAVCPCYNNWKTKEGGPKCP
>VvGASA7
MAKLVLPLLLALFSISMVATKVMMAKEAQYHLDSGSYGPSLKS NQCPSQCTRRCSKTQYHK
PCMFFCQKCCAKCLCVPPGYGNKAVCPCYNNWKTKEGGPKCP
>VvGASA8
MATLRLLLVFVALLVFLAQISSDFNIEGEEMPVFSQ>VvVRGGNRRMLQDIDCGGLCKDRCS
LHSRPNVCVRACGTCCVRCKCVPPGTSGNREL CGKCYTDMTHGNKTKCP

>VvGASA9

MKPLLATFLLVFLVLSSSFVQNAMAGSSFCDISKCAARCSKAGMKDRCLKYCGICCEECKCV
PSGTYGNKHECPCYKDKKNSKGQPKCP

>VvGASA10

MAISKTLIASLLISLLVYQITEAATTSGDGASSPTEKMDCGGAC SARCLSSRPNL CNRACGT
CCARCNCVPPGTSGNQEICPCYANMTTRGNERKCP

>VvGASA11

MKLFSVFIISILLQAF AEASLVISNAEHSLSVDES RDEVALHKKSHPRKINCSYACSRRCRK
ASRKNVCS RACKTCCRCHCVPPGT YGNKNMCPCYASLKTHGHKPKCP

>VvGASA12

MQVLFNSIKLYKLS DCKSKCAYRCSKAGWHKLCLRACNTCCERCNCVPPGTAGNEDVCP
CYAKMTTHGGRHKCP

>VvGASA13

MTYIYFLVSKYRCS DTQYRNACLEFCNLCKKCLCVPSGTYGHKEECPCYNNWKTKEGG
PKCP

>VvGASA14

MWVLGCKFLTYISKKK VSKMAQSSNLQSIFLLLLVAFMLLDV SIVKCPKACNYRCS DTQYL
NACLEFCNLCCQKCLCVPSGTYGHKEECPCYNNWKTKEGGPKCP

Rice GASA proteins

>OsGASA1

MKTRRAALLMLLLL VVAAASWPQPCDAASGFCGSKCAVRCGRGRGRSGCLRSCGLCC
EECNCVPTGSGSTRDECPCYRDMLTAGPRKRPKCP

>OsGASA2

MKLNTTTTLALLLLLL ASSSLQVSMAGSDFCDGKCKVRCSKASRHDDCLKYCGVCCASC
NCVPSGTAGNKDECPCYRDMTTGHGARKRPKCP

>OsGASA3

MAPGKLAVFALLAS LLLLNTIKAADYPPAPPLGPPPHKIVDPGKDCVGACDARCSEHSHK
KRCRSCLTCCSACRCVPAGTAGNRETCGRCYTDWVSHNNMTKCP

>OsGASA4

MEGVGVGV RIRALLCCIAMAA MLLSSYQQGQAEASYMPWP PATPPPPAAAAANSTSTAA
ANNSSSSSSTAPPQQPTAFPMYGVTPGSLRPQECGGRCAYRCSATAYRKPCMFFCQKCCAS
CLCVLPGT YGNKQSCPCYNDWKT KRGGPKCP

>OsGASA5

MASMAKSLLCISLVA ILLLVETTAPHGQAY AIDCGAKCGYRCSKSGRPKMCLRACGTCCQR
CGCVPPGTSGNENVCPCYANMTTHNGRHKCP

>OsGASA6

MASSTKIPFLLLAV LLLLSIAFPSEVMAGGRGRGGGGGGVAGGGNLRPWECSPKCAGRCS
NTQYKKACLTF CNKCCAKCLCVPPGT YGNKGACPCYNNWKTKEGGPKCP

>OsGASA7

MASSGSKSKTNNGFFV P SPATAMAPCFLLLLIFFFLHVDASAAA SSSSHPQLQVQQMQVK
RARSLLQAPKIDCQGTCSGRCAN NWKKEMCNKMCNVCCNRCNCVPPGSGQDTRHLCPC
YDTMVNPHNGK LKCP

>OsGASA8

MRVPPLRATTALLATLLVAASFQDLTVAADGGGGVVPVPSVCDAKCQKRCSLKVAGRCM
GLCKMCCHDCGGCVPSGPYASKDECPCYRDMVSPKSRRPKCP

>OsGASA9

MALAGRLLVLFAIALLAISIAEHKALAKGSTSEHDDNVYQVSKGGQGSLSYQCSPQCAYR
CSQTQYKKPCLFFCNKCCNACLCVPSGLYGNKGECPCYNNWTKRGGPKCP

>OsGASA10

MDPASRSLSIIFFLVAVTFVVEVSGQKNEAVYHLFGGEGSLTKNECPGKCSYRCSATSHTTVC
MTYCNYYCCERCLCVPSGT