

*Κωδικός πρωτεΐνης: 2RNM*

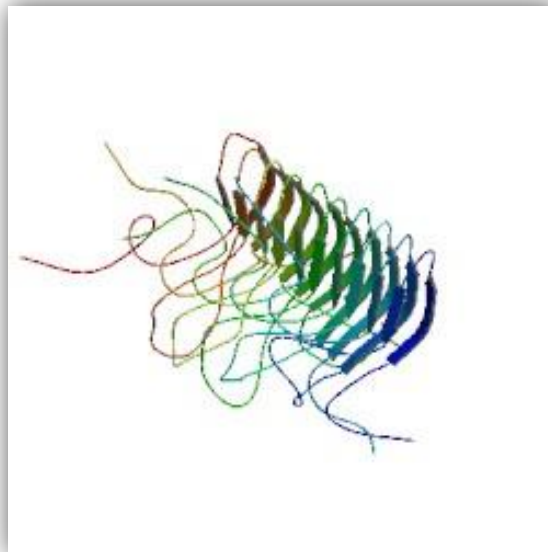
*ΔΟΜΙΚΗ ΒΙΟΧΗΜΕΙΑ ΚΑΙ ΣΤΟΙΧΕΙΑ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ*

*ΕΡΓΑΣΙΑ ΤΟΥ ΚΥΡΙΑΚΙΔΗ ΑΡΙΣΤΕΙΔΗ*

*A.E.M.:8326*

*ΚΩΔΙΚΟΣ ΠΡΩΤΕΪΝΗΣ: 2RNM (HET-S (218-289) PRION IN ITS AMYLOID FORM)*

*ΥΠΕΥΘΥΝΗ ΚΑΘΗΓΗΤΡΙΑ: ΧΟΛΗ-ΠΑΠΑΔΟΠΟΥΛΟΥ ΘΕΟΔΩΡΑ*



## **FASTA**

```
MSEPFGIVAGALNVAGLNFNNVDCFEYVQLGRPFGRDYERCQLRLDIAKARLSRWGEAVK
INDDPRFHSDAPTDKSVQLAKSIVEEILLLFESAQKTSKRYELVADQQDLVVFEDKDMKP
IGRALHRRLNDLVSRRQKQTSLAKKTAWALYDGKSLEKIVDQVARFVDELEKAFPIEAVC
HKLAEIEIEEVEDEASLTILKDAAGGIDAAMSDAAAQKIDAIVGRNSAKDIRTEERARVQ
LGNVVTAAALHGGIRISDQTTNSVETTVVGKGESRVLIGNEYGGKGFWDN
```

## **ProtParam**

### **User-provided sequence:**

```
      10      20      30      40      50      60
MSEPFGIVAG ALNVAGLNFNN CVDCFEYVQL GRPFGRDYER CQLRLDIAKA RLSRWGEAVK

      70      80      90     100     110     120
INDDPRFHSD APTDKSVQLA KSIVEEILLL FESAQKTSKR YELVADQQDL VVFEDKDMKP

     130     140     150     160     170     180
IGRALHRRLN DLVSRRQKQT SLAKKTAWAL YDGKSLEKIV DQVARFVDEL EKAFPIEAVC

     190     200     210     220     230     240
HKLAEIEIEE VEDEASLTIL KDAAGGIDAA MSDAAAQKID AIVGRNSAKD IRTEERARVQ

     250     260     270     280
LGNVVTAAAL HGGIRISDQT TNSVETTVVGK GESRVLIGNE YGGKGFWDN
```

**Number of amino acids:** 289

**Molecular weight:** 31978.2

**Theoretical pI:** 5.35

### **Amino acid composition:**

CSV format

Ala (A)	32	11.1%
Arg (R)	20	6.9%
Asn (N)	10	3.5%
Asp (D)	23	8.0%
Cys (C)	4	1.4%
Gln (Q)	12	4.2%
Glu (E)	24	8.3%
Gly (G)	20	6.9%
His (H)	4	1.4%
Ile (I)	18	6.2%
Leu (L)	25	8.7%
Lys (K)	20	6.9%
Met (M)	3	1.0%
Phe (F)	10	3.5%

*Κωδικός πρωτεΐνης: 2RNM*

Pro (P)	6	2.1%
Ser (S)	16	5.5%
Thr (T)	10	3.5%
Trp (W)	3	1.0%
Tyr (Y)	5	1.7%
Val (V)	24	8.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

**Total number of negatively charged residues (Asp + Glu): 47**

**Total number of positively charged residues (Arg + Lys): 40**

**Atomic composition:**

Carbon	C	1403
Hydrogen	H	2262
Nitrogen	N	402
Oxygen	O	437
Sulfur	S	7

**Formula:** C<sub>1403</sub>H<sub>2262</sub>N<sub>402</sub>O<sub>437</sub>S<sub>7</sub>

**Total number of atoms:** 4511

**Extinction coefficients:**

Extinction coefficients are in units of M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm measured in water.

Ext. coefficient            24200  
Abs 0.1% (=1 g/l)        0.757, assuming all pairs of Cys residues form cystines

Ext. coefficient            23950  
Abs 0.1% (=1 g/l)        0.749, assuming all Cys residues are reduced

**Estimated half-life:**

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

*Κωδικός πρωτεΐνης: 2RNM*

**Instability index:**

The instability index (II) is computed to be 35.26  
This classifies the protein as stable.

**Aliphatic index:** 93.18

**Grand average of hydropathicity (GRAVY):** -0.314

**ProDom**

database: multiple alignments **Graphical results and forms to other applications**

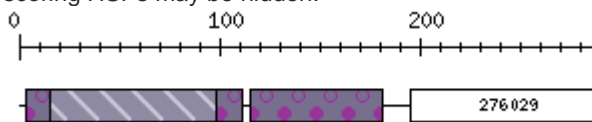
**Program:** ncbi-blastp

**Matrix:** BLOSUM62

**Expect:** 0.01

**Filter:** seg

The following is the graphical representation of the HSP found by BLAST. Please note that HSPs are sorted from highest to lowest scores, so that lower scoring HSPs may be hidden.



**Align subsequence with ProDom domains, using Multalin**

Domain ID	BEGIN	END	
PD276029	<input type="text" value="195"/>	<input type="text" value="287"/>	<input type="button" value="Submit"/>
PD127442	<input type="text" value="15"/>	<input type="text" value="97"/>	<input type="button" value="Submit"/>
PDB082Y3	<input type="text" value="115"/>	<input type="text" value="180"/>	<input type="button" value="Submit"/>
PDC6Y682	<input type="text" value="3"/>	<input type="text" value="39"/>	<input type="button" value="Submit"/>
PDC0C1P8	<input type="text"/>	<input type="text" value="110"/>	<input type="button" value="Submit"/>

**Domain 3D modelling using Swiss-Model**

Domain ID	BEGIN	END	
PD276029	<input type="text" value="195"/>	<input type="text" value="287"/>	<input type="button" value="Submit"/>
PD127442	<input type="text" value="15"/>	<input type="text" value="97"/>	<input type="button" value="Submit"/>
PDB082Y3	<input type="text" value="115"/>	<input type="text" value="180"/>	<input type="button" value="Submit"/>

## Domain 3D modelling using Geno3D

Domain ID	BEGIN	END	
PD276029	<input type="text" value="195"/>	<input type="text" value="287"/>	<input type="button" value="Submit"/>
PD127442	<input type="text" value="15"/>	<input type="text" value="97"/>	<input type="button" value="Submit"/>
PDB082Y3	<input type="text" value="115"/>	<input type="text" value="180"/>	<input type="button" value="Submit"/>



---

## HSP Results

Warning: Original output has been filtered to yield non-redundant similarities

blastp 2.2.26 [Sep-21-2011]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), *quot*;Gapped BLAST and PSI-BLAST: a new generation of protein database search programs\_quot;;, Nucleic Acids Res. 25:3389-3402.

Query: unknown  
(289 letters)

Database: prodom2010.1 multiple alignments  
45,292,438 sequences; 2,147,483,647 total letters

ProDom domains producing High-scoring Segment Pairs:

Position	ProDom domain	Score	E value
3-39	#PDC6Y682	107	0.0002
15-97	#PD127442	451	2e-53
56-110	#PDC0C1P8	103	0.0009
115-180	#PDB082Y3	346	5e-38
195-287	#PD276029	470	4e-56

>**PD276029** (Closest domain: Q03689\_PODAS 186-287)

Number of domains in family: 9

Commentary (automatic):

SUBNAME: S REFERENCE FULL=PUTATIVE UNCHARACTERIZED FULL=SMALL 3D-STRUCTURE FULL=BIG CDS FULL=PREDICTED

Length = 102

Score = 470 (185.7 bits), Expect = 4e-56

*Κωδικός πρωτεΐνης: 2RNM*

Identities = 93/93 (100%), Positives = 93/93 (100%)

Query: 195 ASLTILKDAAGGIDAAMSDAAAQKIDAIVGRNSAKDIRTEERARVQLGNVVTAALHGGI  
254

ASLTILKDAAGGIDAAMSDAAAQKIDAIVGRNSAKDIRTEERARVQLGNVVTAALHGGI  
Sbjct: 195 ASLTILKDAAGGIDAAMSDAAAQKIDAIVGRNSAKDIRTEERARVQLGNVVTAALHGGI  
254

Query: 255 RISDQTTNSVETVVGKGESRVLIGNEYGGKGF 287

RISDQTTNSVETVVGKGESRVLIGNEYGGKGF  
Sbjct: 255 RISDQTTNSVETVVGKGESRVLIGNEYGGKGF 287

>**PD127442** (Closest domain: Q03689\_PODAS 15-97)

Number of domains in family: 29

Commentary (automatic):

SUBNAME: FULL=PUTATIVE UNCHARACTERIZED REFERENCE S FULL=SMALL REPEAT FULL=PREDICTED  
3D-STRUCTURE ANK

Length = 83

Score = 451 (178.3 bits), Expect = 2e-53

Identities = 83/83 (100%), Positives = 83/83 (100%)

Query: 15 AGLFNNCVDCFEYVQLGRPFGRDYERCQLRLDIAKARLSRWGEAVKINDDPRFHS DAPT D  
74

AGLFNNCVDCFEYVQLGRPFGRDYERCQLRLDIAKARLSRWGEAVKINDDPRFHS DAPT D  
Sbjct: 15 AGLFNNCVDCFEYVQLGRPFGRDYERCQLRLDIAKARLSRWGEAVKINDDPRFHS DAPT D  
74

Query: 75 KSVQLAKSIVEEILLLFESAQKT 97

KSVQLAKSIVEEILLLFESAQKT  
Sbjct: 75 KSVQLAKSIVEEILLLFESAQKT 97

>**PDB082Y3** (Closest domain: Q03689\_PODAS 115-180)

Number of domains in family: 29

Commentary (automatic):

SUBNAME: FULL=PUTATIVE UNCHARACTERIZED REFERENCE S FULL=SMALL REPEAT FULL=PREDICTED  
3D-STRUCTURE TPR

Length = 66

Score = 346 (137.9 bits), Expect = 5e-38

Identities = 66/66 (100%), Positives = 66/66 (100%)

Query: 115 DKDMKPIGRALHRRRLNDLVSRRQKQTS LAKKTAWALYDGKSLEKIVDQVARFVDELEKAF  
174

DKDMKPIGRALHRRRLNDLVSRRQKQTS LAKKTAWALYDGKSLEKIVDQVARFVDELEKAF  
Sbjct: 115 DKDMKPIGRALHRRRLNDLVSRRQKQTS LAKKTAWALYDGKSLEKIVDQVARFVDELEKAF  
174

Query: 175 PIEAVC 180

PIEAVC  
Sbjct: 175 PIEAVC 180

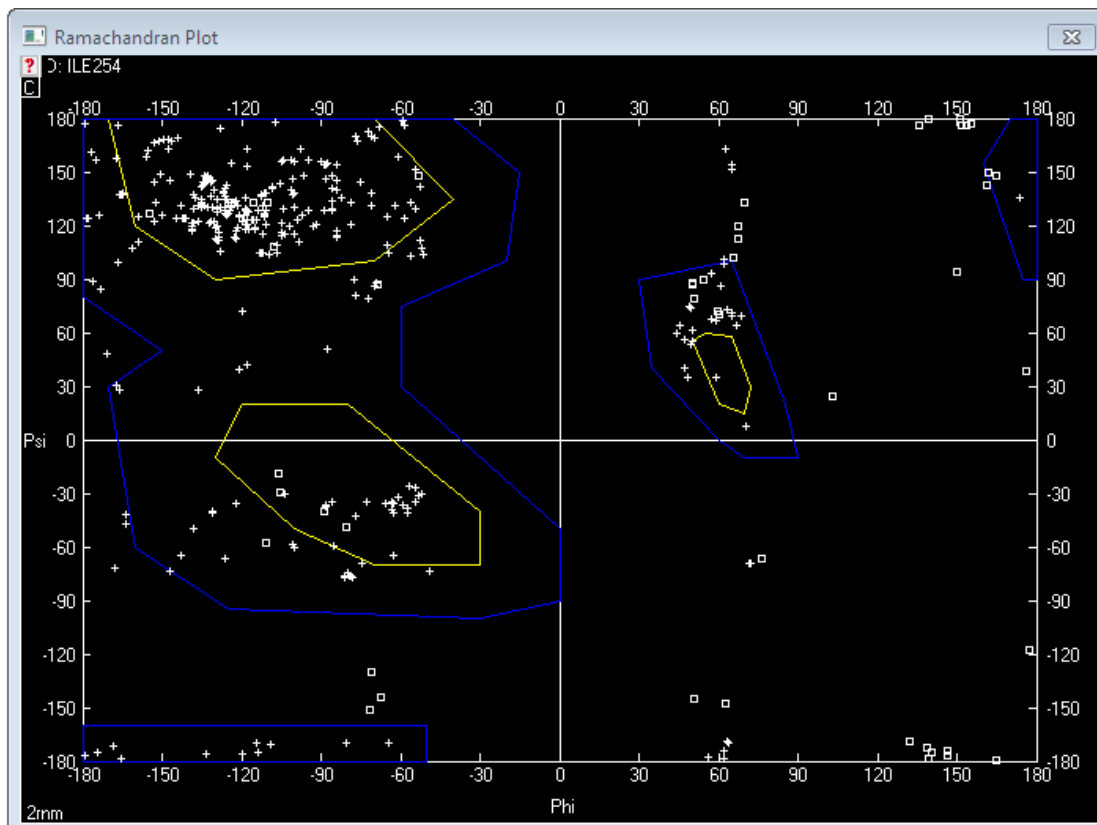
*Κωδικός πρωτεΐνης: 2RNM*

>**PDC6Y682** (Closest domain: C7ZM97\_NECH7 3-39)  
Number of domains in family: 14  
Commentary (automatic):  
SUBNAME: FULL=PUTATIVE UNCHARACTERIZED REPEAT ANK NUCLEOTIDE-BINDING ATP-BINDING  
REFERENCE S PRECURSOR  
Length = 37  
Score = 107 (45.8 bits), Expect = 0.0002  
Identities = 19/37 (51%), Positives = 21/37 (56%), Gaps = 3/37 (8%)

Query: 3 EPFGIVAGALNVAGLFNNCVDCFEYVQLGRPFGRDYE 39  
EP GI G L + G F NCV+ F Y R GRDYE  
Sbjct: 3 EPVGITLGVLGLLGTFENCVELFAYFSASRSLGRDYE 39

>**PDC0C1P8** (Closest domain: C7ZA77\_NECH7 1-59)  
Number of domains in family: 1  
Commentary (automatic):  
SUBNAME: REFERENCE FULL=PUTATIVE UNCHARACTERIZED  
Length = 59  
Score = 103 (44.3 bits), Expect = 0.0009  
Identities = 22/55 (40%), Positives = 30/55 (54%), Gaps = 7/55 (12%)

Query: 56 GEAVKINDDPRFHSDAPTDKSVQLAKSIVEEILLFFESAQKTSKRYELVADQDDL 110  
GEAV +N +PRF +D D + Q ++EE L FE + S RY AD + L  
Sbjct: 3 GEAVGVNTEPRFATDNSDDITAQRVCRVLEETRLCFEGVHRLSSRYSPPADSRGL 57

**RAMACHANDRAN PLOT (Μετά από χρήση του λογισμικού SPDB Viewer)**

Το Ramachandran Plot ,όπως γνωρίζουμε, είναι ένα διάγραμμα το οποίο δείχνει τις τιμές των γωνιών που μπορεί να υιοθετήσει ένα αμινοξύ ,κατά την περιστροφή μιας πολυπεπτιδικής αλυσίδας , είτε γύρω από το δεσμό **Ca-N (γωνία φ-οριζόντιος άξονας)** είτε γύρω από το δεσμό **Ca-C (γωνία ψ-κάθετος άξονας)**. Η περιστροφή γύρω από τον πεπτιδικό δεσμό είναι αδύνατη, καθώς αυτός διαθέτει *χαρακτήρα εν μέρει διπλού δεσμού*.

Στα τεταρτημόρια του RP απεικονίζονται οι τιμές των γωνιών φ και ψ για διάφορες επιτρεπόμενες δευτεροταγείς δομές: Το πρώτο άνω τεταρτημόριο από δεξιά αντιστοιχεί σε **αριστερόστροφη α-έλικα (παρότι θεωρητικά εφικτή, δεν έχει παρατηρηθεί στις πρωτεΐνες)** , το δεύτερο στη σειρά σε **β-έλασμα**, ενώ, το πρώτο κάτω τεταρτημόριο από αριστερά σε **δεξιόστροφη α-έλικα**. Το τελευταίο τεταρτημόριο περιλαμβάνει **μόνο γωνίες που συνεπάγονται απαγορευμένες κινήσεις**.

Οι περιοχές, λοιπόν, που βρίσκονται **εκτός των περιοχών των επιτρεπτών κινήσεων** (συμπεριλαμβανομένου και του τελευταίου τεταρτημορίου) αντιστοιχούν σε μη ορθές διαμορφώσεις στις οποίες τα άτομα πλησιάζουν πολύ κοντά μεταξύ τους. Συγκεκριμένα, πλησιάζουν περισσότερο από το άθροισμα των ακτίνων τους Van der Waals. Έτσι, σε αυτές τις περιοχές όλα τα αμινοξέα εμφανίζουν **στεreoχημικές παρεμποδίσεις, εκτός της Gly**, η οποία δε διαθέτει πλευρική αλυσίδα (R=H) και μπορεί να υιοθετήσει χωρίς πρόβλημα οποιαδήποτε από τις γωνίες φ και ψ σε κάθε τεταρτημόριο

του Ramachandran Plot. Γι' αυτό το λόγο, προτιμώνται στις στροφές των πολυπεπτιδικών αλυσίδων, καθώς οποιοδήποτε άλλο αμινοξύ θα προκαλούσε παρεμπόδιση.

Λαμβάνοντας υπόψη όλα τα παραπάνω και παρατηρώντας τη μορφή του RP μας, διαπιστώνουμε πως τα αμινοξέα εκτός των επιτρεπτών κινήσεων- περιστροφών είναι τα εξής:

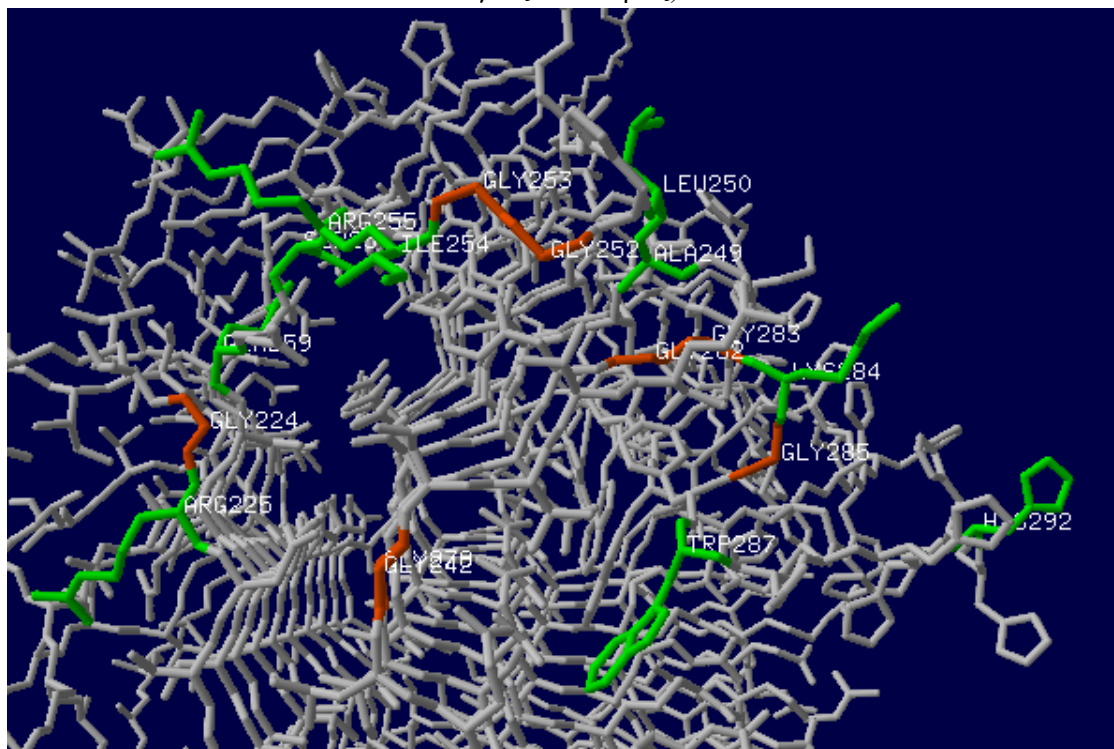
i)GLN -259	ii)SER- 257	iii)TRP- 287	iv)HIS -292	v)ARG- 225
vi)LEU- 250	vii)LYS -284	viii)ARG -255	ix)ILE- 254	x)ALA -249

**Ακολούθως παρατίθενται και οι Gly:**

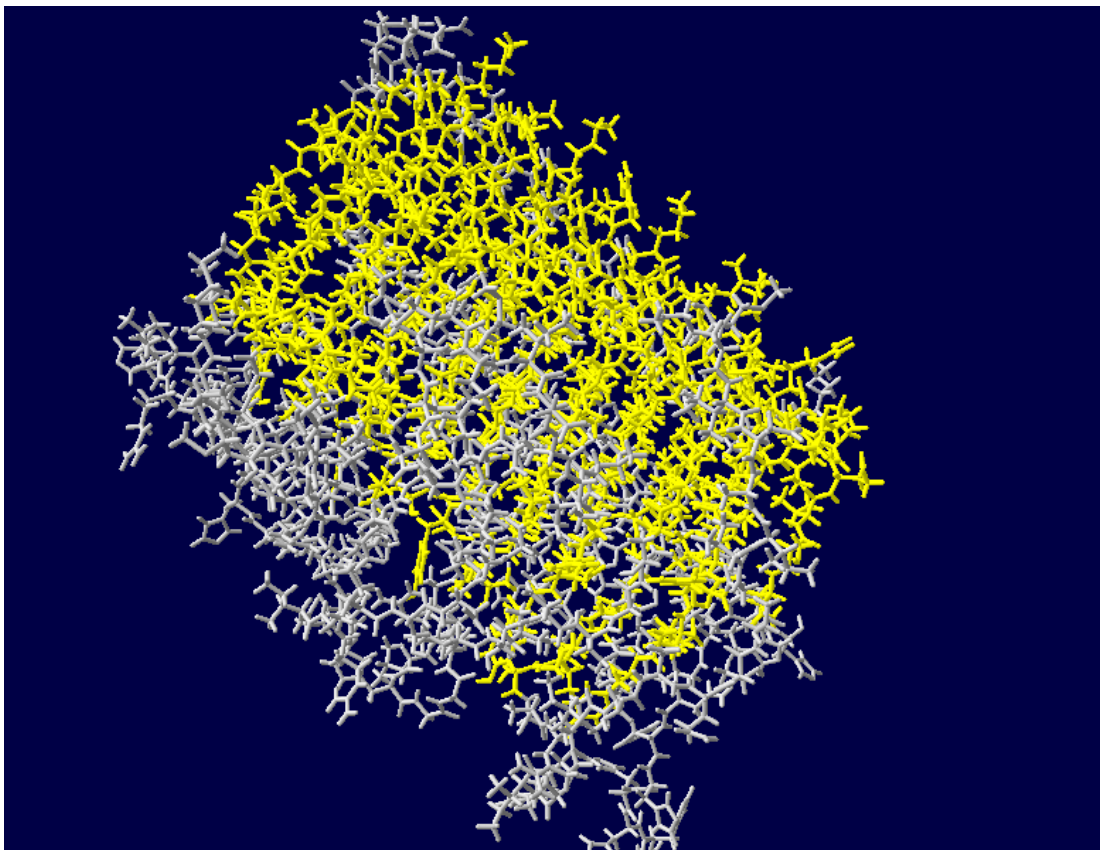
<b>Gly-224</b>	<b>Gly-242</b>	<b>Gly-252</b>	<b>Gly-253</b>	<b>Gly-278</b>	<b>Gly-282</b>	<b>Gly-283</b>	<b>Gly-285</b>
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------

**(Σημείωση: Οι Gly στο RP απεικονίζονται με τετράγωνο, ενώ τα υπόλοιπα αμινοξέα με σταυρό.)**

*Τα αμινοξέα που δε συμμετέχουν στις επιτρεπτές κινήσεις απεικονίζονται με πράσινο χρώμα, ενώ οι Gly απεικονίζονται με καφετί χρώμα. (Απεικονίζεται μόνο τα αμινοξέα στην αλυσίδα Α για λόγους απλότητας)*



**ΑΠΕΙΚΟΝΙΣΗ ΔΕΥΤΕΡΟΤΑΓΟΥΣ ΔΟΜΗΣ ΤΗΣ 2RNM**



*Με κίτρινο χρώμα απεικονίζονται οι β-πτυχωτές επιφάνειες (β-ελιάσματα). Το πρωτεϊνικό μόριο διαθέτει ,επίσης, στροφές και μη δομημένες περιοχές, αλλά δε διαθέτει α-έλικες.  
([www.rcsb.org/pdb](http://www.rcsb.org/pdb))*