

ΔΟΜΙΚΗ ΒΙΟΧΗΜΕΙΑ ΚΑΙ ΣΤΟΙΧΕΙΑ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Επιβλέπουσα Καθηγήτρια: Χολή-Παπαδοπούλου
Θεοδώρα
Όνοματεπώνυμο: Κοσμά Άννα
Α.Ε.Μ.: 8321

Κωδικός Πρωτεΐνης: 2HIU

Βήματα που ακολουθήθηκαν: Στην ιστοσελίδα με διεύθυνση <http://molwave.chem.auth.gr/pdb> (RCSB Protein Data Bank), εισάγουμε τον κωδικό της πρωτεΐνης και από την επιλογή Sequence κατεβάζουμε το PDB File (Text), καθώς και το FASTA file. Στη συνέχεια βρίσκουμε τον κωδικό UniProtKB (P01308) και μπαίνουμε στην ιστοσελίδα www.expasy.org. Από το Proteomics μπαίνουμε από τα Tools στο ProtParam και εκεί εισάγουμε την αμινοξική αλληλουχία που κατεβάσαμε σε FASTA file. Για την πρωτεΐνη μας προκύπτουν δύο ισομορφές, τα χαρακτηριστικά των οποίων δίνονται παρακάτω:

ISOFORM 1

```
>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens GN=INS PE=1 SV=1  
MALWMRLLPLLALLLWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED  
LQVGGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

ISOFORM 2

```
>sp|F8WCM5|INSR2_HUMAN Insulin, isoform 2 OS=Homo sapiens GN=INS-IGF2  
PE=2 SV=1  
MALWMRLLPLLALLLWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED  
LQASALSLSSSTSTWPEGLDARAPPALVVVITANIGQAGGSSSRQFRQALGTSDSPVLF  
IHCPGAAGTAQGLEYRGRVTTTELWVEEVDSSPQFPQGSSESLPAQPPAQPAPQPEPQQARE  
PSPEVSCCGLWPRRPQRSQN
```

ISOFORM 1

User-provided sequence:

 10 20 30 40 50 60
MALWMRLLP LALLALWGPD PAAAFVNQHL CGSHLVEALY LVCGERGFFY TPKTREAEED

 70 80 90 100 110
LQVGQVELGG GPGAGSLQPL ALEGLQKRG IVEQCCTSIC SLYQLENYCN

Number of amino acids: 110

Molecular weight: 11980.9

TheoreticalpI: 5.22

Amino acid composition:

Ala (A)	10	9.1%
Arg (R)	5	4.5%
Asn (N)	3	2.7%
Asp (D)	2	1.8%
Cys (C)	6	5.5%
Gln (Q)	7	6.4%
Glu (E)	8	7.3%
Gly (G)	12	10.9%
His (H)	2	1.8%
Ile (I)	2	1.8%
Leu (L)	20	18.2%
Lys (K)	2	1.8%
Met (M)	2	1.8%
Phe (F)	3	2.7%
Pro (P)	6	5.5%
Ser (S)	5	4.5%
Thr (T)	3	2.7%
Trp (W)	2	1.8%
Tyr (Y)	4	3.6%
Val (V)	6	5.5%
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): 10

Total number of positively charged residues (Arg + Lys): 7

Atomic composition:

Carbon	C	535
Hydrogen	H	841
Nitrogen	N	143
Oxygen	O	153

Sulfur S 8

Formula: C₅₃₅H₈₄₁N₁₄₃O₁₅₃S₈

Total number of atoms: 1680

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient 17335
Abs 0.1% (=1 g/l) 1.447, assuming all pairs of Cys residues form cystines

Ext. coefficient 16960
Abs 0.1% (=1 g/l) 1.416, 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).

Instability index:

The instability index (II) is computed to be 40.33

This classifies the protein as unstable.

Aliphatic index: 102.91

Grand average of hydropathicity (GRAVY): 0.193

ISOFORM 2

User-provided sequence:

```
      10      20      30      40      50      60
MALWMRLLP LALLALWGPD PAAAFVNQHL CGSHLVEALY LVCGERGFFY TPKTRREAED

      70      80      90     100     110     120
LQASALSLS STSTWPEGLD ATARAPPALV VTANIGQAGG SSSRQFRQRA LGTSDSPVLF

     130     140     150     160     170     180
IHCPGAAGTA QGLEYRGRRV TTELVWEEVD SSPQPQGSSE LPAQPPAQP PAQPEPQQARE

     190     200
PSPEVSCCGL WPRRPQRSQN
```

Number of amino acids: 200

Molecular weight: 21537.2

TheoreticalpI: 5.93

Amino acid composition:

Ala (A)	24	12.0%
Arg (R)	15	7.5%
Asn (N)	3	1.5%
Asp (D)	5	2.5%
Cys (C)	5	2.5%
Gln (Q)	15	7.5%
Glu (E)	13	6.5%
Gly (G)	15	7.5%
His (H)	3	1.5%
Ile (I)	2	1.0%
Leu (L)	23	11.5%
Lys (K)	1	0.5%
Met (M)	2	1.0%
Phe (F)	5	2.5%
Pro (P)	22	11.0%
Ser (S)	19	9.5%
Thr (T)	10	5.0%
Trp (W)	5	2.5%
Tyr (Y)	3	1.5%
Val (V)	10	5.0%
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): 18

Total number of positively charged residues (Arg + Lys): 16

Atomic composition:

Carbon	C	947
Hydrogen	H	1483
Nitrogen	N	275
Oxygen	O	287
Sulfur	S	7

Formula: C₉₄₇H₁₄₈₃N₂₇₅O₂₈₇S₇
Total number of atoms: 2999

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient 32220
Abs 0.1% (=1 g/l) 1.496, assuming all pairs of Cys residues form cystines

Ext. coefficient 31970
Abs 0.1% (=1 g/l) 1.484, 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).

Instability index:

The instability index (II) is computed to be 77.16
This classifies the protein as unstable.

Aliphatic index: 75.25

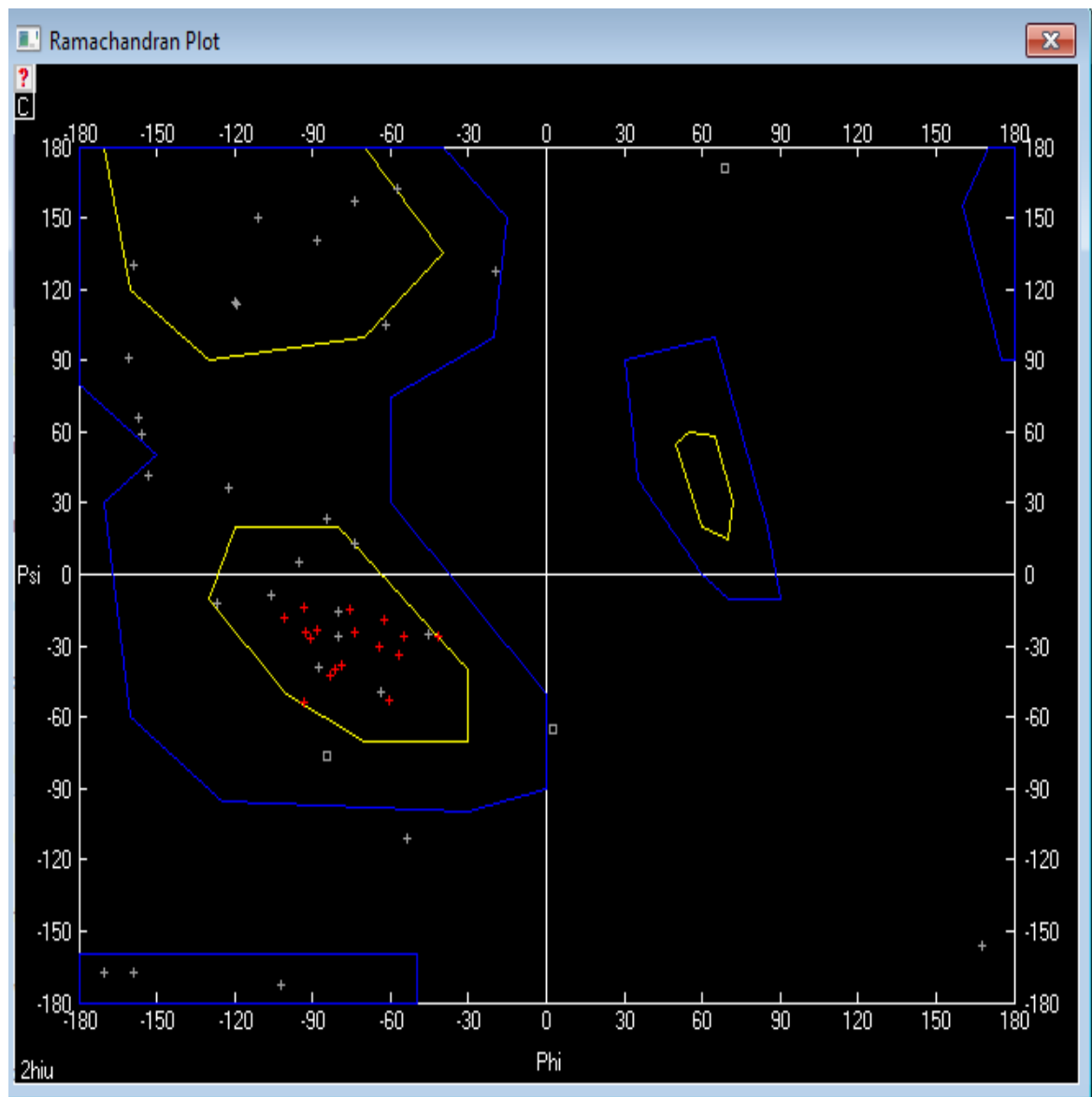
Grand average of hydropathicity (GRAVY): -0.335

Στη συνέχεια ανοίγουμε το πρόγραμμα `srdn`, επιλέγουμε από το File το αρχείο PDB που κατεβάσαμε για την πρωτεΐνη μας και επιλέγουμε το Window->Control Panel. Επιλέγουμε όλα τα αμινοξέα μέχρι αυτά να κοκκινίσουν και προβαίνουμε στις παρακάτω ενέργειες: Display->Render in solid 3D (εφαρμόζω δύο φορές) και έπειτα Color->By Secondary Structure. Οι εικόνες παρουσιάζονται γραφικά στη συνέχεια. Τέλος, αφού έχουμε επιλέξει τα αμινοξέα, κλικάρουμε Window->Ramachandran Plot και λαμβάνουμε το αντίστοιχο διάγραμμα, ερμηνεία του οποίου δίδεται εν συνεχεία.

Ερμηνεία του Ramachandran Plot:

Οι δύο γωνίες στρέψης μιας πολυπεπτιδικής αλυσίδας περιγράφουν τις περιστροφές του πολυπεπτιδικού σκελετού ανάμεσα στους δεσμούς μεταξύ των ατόμων N-Cα (γωνία φ) και Ca-C (γωνία ψ). Το διάγραμμα Ramachandran παρέχει έναν απλό τρόπο να οπτικοποιηθεί η κατανομή των γωνιών στρέψης μιας πολυπεπτιδικής αλυσίδας. Επίσης, είναι ένα εργαλείο το οποίο μας δείχνει τις επιτρεπτές και μη επιτρεπτές περιοχές των τιμών των γωνιών στρέψης. Έτσι, είναι ένας σημαντικός δείκτης της ποιότητας των τρισδιάστατων πρωτεϊνικών δομών. Οι γωνίες στρέψης είναι από τις σημαντικότερες παραμέτρους, οι οποίες ελέγχουν την αναδίπλωση μιας πολυπεπτιδικής αλυσίδας. Οι γωνίες στρέψης σε ένα τέτοιο διάγραμμα περιορίζονται σε συγκεκριμένες τιμές, επειδή μερικές άλλες οδηγούν σε στερεοχημικές παρεμπόδισεις ανάμεσα στην κύρια αλυσίδα και τις πλευρικές αλυσίδες. Στο διάγραμμα Ramachandran, το πλαίσιο στην κάτω αριστερά περιοχή δείχνει τις χαρακτηριστικές γωνίες στρέψης που απαντώνται στις δεξιόστροφες α-έλικες και πάνω αριστερά οι αντίστοιχες για τα β-ελάσματα. Στο πλαίσιο πάνω δεξιά δίνονται οι επιτρεπτές τιμές που απαντώνται στις αριστερόστροφες α-έλικες.

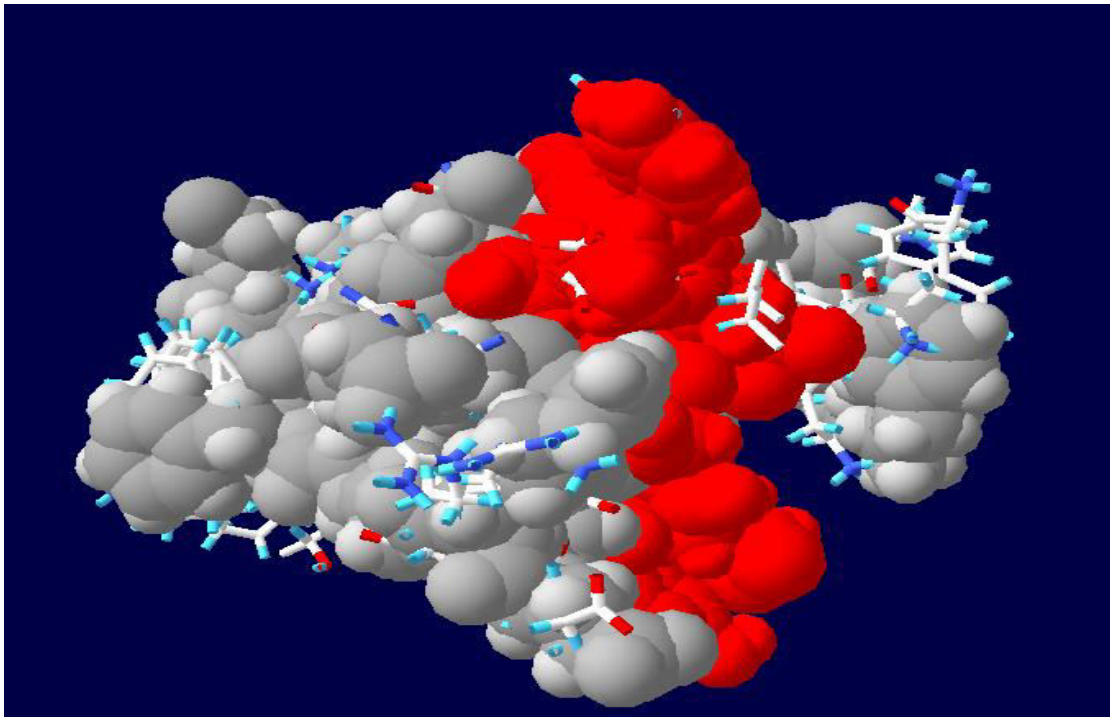
Κάποιες τιμές των γωνιών φ και ψ είναι απαγορευτικές, επειδή κάποια άτομα πρόκειται να έρθουν πάρα πολύ κοντά, με αποτέλεσμα στερεοχημικής παρεμπόδισης. Σε αυτή την περίπτωση, η ενέργεια του συστήματος αυξάνει υπερβολικά, γεγονός καθόλου ευνοϊκό. Αυτές οι τιμές απαντώνται στα αμινοξέα που βρίσκονται εκτός των επιτρεπτών περιοχών Ramachandran. Τα συγκεκριμένα αμινοξέα αναφέρονται εν συνεχεία.



Αμινοξέα εκτός του Ramachandran Plot:

- CYS20
- PHE24
- GLN4
- SER12
- SER9
- GLY23
- GLY8

(Render in solid 3D)



(Color by Secondary Structure)

