

# ΕΡΓΑΣΙΑ ΓΙΑ ΤΟ ΜΑΘΗΜΑ ΔΟΜΙΚΗ ΒΙΟΧΗΜΕΙΑ ΚΑΙ ΣΤΟΙΧΕΙΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

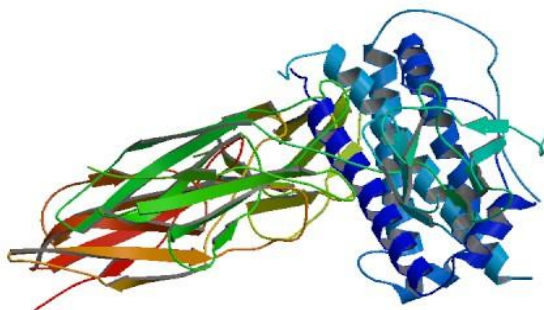
Ακαδημαϊκό έτος 2015-2016

ΥΠΕΥΘΥΝΗ ΚΑΘΗΓΗΤΡΙΑ: Θ.ΧΟΛΗ-ΠΑΠΑΔΟΠΟΥΛΟΥ

Καρακίτσου Κυριακή, Α.Ε.Μ.: 8409

Παλαιολόγου Αναστασία, Α.Ε.Μ.: 8353

HUMAN GROWTH HORMONE AND EXTRACELLULAR DOMAIN OF  
ITS RECEPTOR(**3HHR**)



## Fasta from UnitProtkb

1: This isoform has been chosen as the 'canonical' sequence

```
>sp|P01241|SOMA_HUMAN Somatotropin OS=Homo sapiens GN=GH1 PE=1 SV=2  
MATGSRTSLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEA  
YIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLR  
SVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDD  
ALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF
```

2

```
>sp|P01241-2|SOMA_HUMAN Isoform 2 of Somatotropin OS=Homo sapiens  
GN=GH1  
MATGSRTSLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFNPQ  
TSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSN  
VYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKD  
MDKVETFLRIVQCRSVEGSCGF
```

3

>sp|P01241-3|SOMA\_HUMAN Isoform 3 of Somatotropin OS=Homo sapiens  
GN=GH1  
MATGSRTSLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEA  
YIPKEQKYSFLQNPQTSLCFSESIPSPNREETQQKSNLELLRISLLLIQTLMGRLEDGS  
PRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF

4

>sp|P01241-4|SOMA\_HUMAN Isoform 4 of Somatotropin OS=Homo sapiens  
GN=GH1  
MATGSRTSLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEA  
YIPKEQKYSFLQNPQTSLCFSESIPSPNREETQQKSNLELLRISLLLIQSWLEPVQIFK  
QTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF

5

>sp|P01241-5|SOMA\_HUMAN Isoform 5 of Somatotropin OS=Homo sapiens  
GN=GH1  
MATGSRTSLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFNLE  
LLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPR  
TGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF

ProtParam

**Number of amino acids:** 217

**Molecular weight:** 24847.2

**Theoretical pI:** 5.29

**Amino acid composition:**

Ala (A)	10	4.6%	Pro (P)	9	4.1%
Arg (R)	12	5.5%	Ser (S)	21	9.7%
Asn (N)	9	4.1%	Thr (T)	12	5.5%
Asp (D)	11	5.1%	Trp (W)	2	0.9%
Cys (C)	5	2.3%	Tyr (Y)	8	3.7%
Gln (Q)	14	6.5%	Val (V)	7	3.2%
Glu (E)	15	6.9%	Pyl (O)	0	0.0%
Gly (G)	11	5.1%	Sec (U)	0	0.0%
His (H)	3	1.4%			
Ile (I)	8	3.7%	(B)	0	0.0%
Leu (L)	33	15.2%	(Z)	0	0.0%
Lys (K)	9	4.1%	(X)	0	0.0%
Met (M)	4	1.8%			
Phe (F)	14	6.5%			

**Total number of negatively charged residues (Asp + Glu): 26**

**Total number of positively charged residues (Arg + Lys): 21**

**Atomic composition:**

Carbon	C	1113
Hydrogen	H	1729
Nitrogen	N	293
Oxygen	O	334
Sulfur	S	9

**Formula:** C<sub>1113</sub>H<sub>1729</sub>N<sub>293</sub>O<sub>334</sub>S<sub>9</sub>

**Total number of atoms:** 3478

**Extinction coefficients:**

Extinction coefficients are in units of  $M^{-1} \text{ cm}^{-1}$ , at 280 nm measured in water.

Ext. coefficient 23170

Abs 0.1% (=1 g/l) 0.932, assuming all pairs of Cys residues form cystines

Ext. coefficient 22920

Abs 0.1% (=1 g/l) 0.922, 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 42.75

This classifies the protein as unstable.

**Aliphatic index:** 87.65

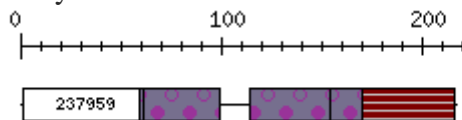
**Grand average of hydropathicity (GRAVY):** -0.269

*ProDom*

<b>database:</b>	multiple alignments
<b>Program:</b>	ncbi-blastp
<b>Matrix:</b>	BLOSUM62
<b>Expect:</b>	0.01
<b>Filter:</b>	seg

**Graphical results and forms to other applications**

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**

<b>Domain ID</b>	<b>BEGIN</b>	<b>END</b>
<b>PD237959</b>	1	58
<b>PD003663</b>	170	215
<b>PDA0L2H6</b>	114	153
<b>PDC67792</b>	174	210
<b>PDC650A7</b>	61	98
<b>PDC649R4</b>	59	89
<b>PDC6A2F2</b>	154	180
<b>PDD0P7B1</b>	60	98
<b>PDB001Y4</b>	141	211

## Domain 3D modelling using Swiss-Model

Domain ID	BEGIN	END
PD237959	1	58
PD003663	170	215
PDA0L2H6	114	153
PDC67792	174	210
PDC650A7	61	98
PDC649R4	59	89
PDC6A2F2	154	180

## Domain 3D modelling using Geno3D

Domain ID	BEGIN	END
PD237959	1	58
PD003663	170	215
PDA0L2H6	114	153
PDC67792	174	210
PDC650A7	61	98
PDC649R4	59	89
PDC6A2F2	154	180

## 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: unkwown  
(217 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
1-58	#PD237959	307	4e-33
59-89	#PDC649R4	161	2e-12
60-98	#PDD0P7B1	123	6e-07
61-98	#PDC650A7	196	4e-17
114-153	#PDA0L2H6	208	1e-18
141-211	#PDB001Y4	104	0.0004
154-180	#PDC6A2F2	144	4e-10
170-215	#PD003663	249	8e-25
174-210	#PDC67792	200	8e-18

>**PD237959** (Closest domain: SOMA\_HUMAN 1-58)

Number of domains in family: 256

Commentary (automatic):

HORMONE SECRETED FULL=GROWTH DISULFIDE BOND SUBNAME:

PRECURSOR ALTNAME: FLAGS: SIGNAL

Length = 58

Score = 307 (122.9 bits), Expect = 4e-33

Identities = 58/58 (100%), Positives = 58/58 (100%)

Query: 1

MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDT  
YQEFE 58

MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDT  
YQEFE

Sbjct: 1

MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDT  
YQEFE 58

>**PD003663** (Closest domain: Q6IYF1\_HUMAN 170-215)

Number of domains in family: 395

Commentary (automatic):

HORMONE SECRETED DISULFIDE BOND FULL=GROWTH SUBNAME:

PRECURSOR FLAGS: SIGNAL ALTNAME:

Length = 46

Score = 249 (100.5 bits), Expect = 8e-25

Identities = 46/46 (100%), Positives = 46/46 (100%)

Query: 170  
SKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSC 215  
SKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSC  
Sbjct: 170  
SKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSC 215

>**PDA0L2H6** (Closest domain: B1A4G7\_HUMAN 86-138)

Number of domains in family: 264

Commentary (automatic):

HORMONE SECRETED FULL=GROWTH DISULFIDE BOND SUBNAME:

ALTNAME: PRECURSOR FLAGS: SIGNAL

Length = 53

Score = 208 (84.7 bits), Expect = 1e-18

Identities = 40/40 (100%), Positives = 40/40 (100%)

Query: 114 EPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGR 153

EPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGR

Sbjct: 99 EPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGR 138

>**PDC67792** (Closest domain: B1A4G9\_HUMAN 79-115)

Number of domains in family: 82

Commentary (automatic):

HORMONE SECRETED DISULFIDE BOND SUBNAME: FULL=PROLACTIN

PRECURSOR FLAGS: SIGNAL RECNAME:

Length = 37

Score = 200 (81.6 bits), Expect = 8e-18

Identities = 37/37 (100%), Positives = 37/37 (100%)

Query: 174 TNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRS 210

TNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRS

Sbjct: 79 TNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRS 115

>**PDC650A7** (Closest domain: B1A4G6\_HUMAN 61-99)

Number of domains in family: 180

Commentary (automatic):

HORMONE FULL=GROWTH DISULFIDE BOND SECRETED SUBNAME:

ALTNAME: PRECURSOR FLAGS: RECNAME:

Length = 39

Score = 196 (80.1 bits), Expect = 4e-17

Identities = 38/38 (100%), Positives = 38/38 (100%)

Query: 61 YIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSN 98

YIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSN

Sbjct: 61 YIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSN 98

>**PDC649R4** (Closest domain: F6JY45\_PAPHA 59-89)

Number of domains in family: 185

Commentary (automatic):

HORMONE SECRETED FULL=GROWTH DISULFIDE BOND SUBNAME:

PRECURSOR FLAGS: SIGNAL ALTNAM:

Length = 31

Score = 161 (66.6 bits), Expect = 2e-12

Identities = 31/31 (100%), Positives = 31/31 (100%), Gaps = 1/31 (3%)

Query: 59 EAYIPKEQKYSFLQNPQTSLCFSESIPTPSN 89

EAYIPKEQKYSFLQNPQTSLCFSESIPTPSN

Sbjct: 59 EAYIPKEQKYSFLQNPQTSLCFSESIPTPSN 89

>**PDC6A2F2** (Closest domain: B1A4G7\_HUMAN 139-165)

Number of domains in family: 56

Commentary (automatic):

HORMONE SECRETED DISULFIDE BOND SUBNAME: FULL=GROWTH

ALTNAM: PRECURSOR FLAGS: FULL=SOMATOTROPIN

Length = 27

Score = 144 (60.1 bits), Expect = 4e-10

Identities = 27/27 (100%), Positives = 27/27 (100%), Gaps = 1/27 (3%)

Query: 154 LEDGSPRTGQIFKQTYSKFDTNSHNDD 180

LEDGSPRTGQIFKQTYSKFDTNSHNDD

Sbjct: 139 LEDGSPRTGQIFKQTYSKFDTNSHNDD 165

>**PDD0P7B1** (Closest domain: G5BDI6\_HETGA 59-99)

Number of domains in family: 2

Commentary (automatic):

HORMONE SUBNAME: DISULFIDE BOND SECRETED FULL=GROWTH

FULL=SOMATOTROPIN

Length = 41

Score = 123 (52.0 bits), Expect = 6e-07

Identities = 24/39 (61%), Positives = 31/39 (79%), Gaps = 1/39 (2%)

Query: 60 AYIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSN 98

AYIP+ Q+YS LQN Q + CFSSESIP P+ +EE QQ+S+

Sbjct: 60 AYIPEGQRYS-LQNTQAAFCFSSESIPAPTGKEEAQQRS 97

>**PDB001Y4** (Closest domain: Q75QK4\_COTCO 151-224)

Number of domains in family: 152

Commentary (automatic):

HORMONE SUBNAME: SECRETED DISULFIDE BOND FULL=PROLACTIN

RECNAME: ALTNAME: FLAGS: PRECURSOR

Length = 74

Score = 104 (44.7 bits), Expect = 0.0004

Identities = 23/76 (30%), Positives = 42/76 (55%), Gaps = 10/76 (13%)

Query: 141 KDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFD-----

TNSHNDDALLKNYGLLYCFRKD 195

K L E G++ ++GR+ G + YS++D + D L Y LL+CFR+D

Sbjct: 154 KRLLEGMEKIVGRVHSGDGGN-----

EIYSQWDGLPSLQLADED SRLFAFYNLLHCFRRD 208

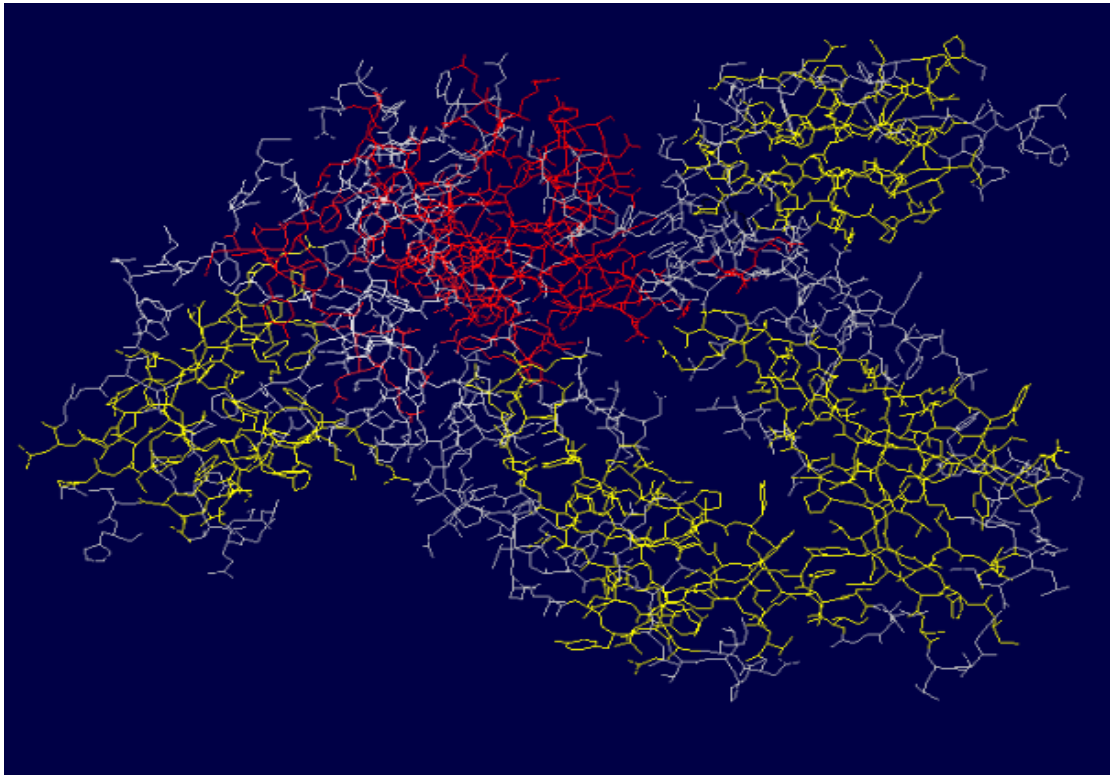
Query: 196 MDKVETFLRIVQCRSV 211

K++ +L++++CR +

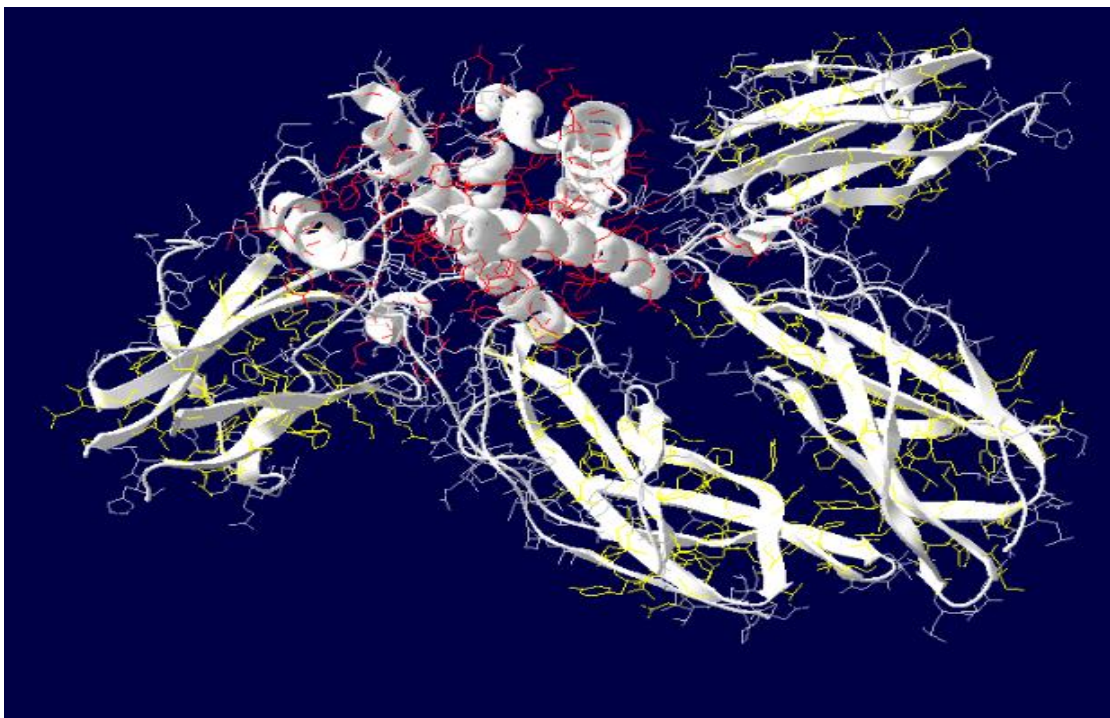
Sbjct: 209 SHKIDNYLKVLKCR LI 224

## ΕΠΕΞΕΡΓΑΣΙΑ ΤΗΣ ΠΡΩΤΕΪΝΗΣ ΜΕ ΤΟ ΠΡΟΓΡΑΜΜΑ SPDBV

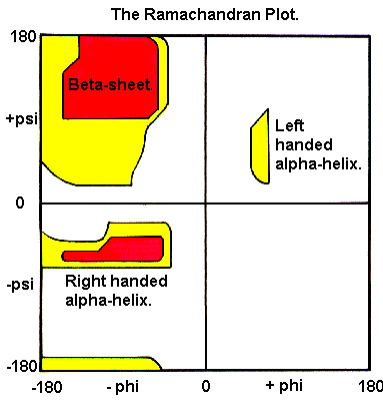
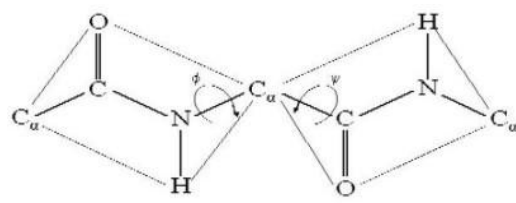
Αρχικά έχουμε μία εικόνα της πρωτεΐνης χρωματισμένη με βάση την δευτεροταγή δομή της. Το κόκκινο χρώμα υποδηλώνει την δομή α-έλικας, ενώ το κίτρινο την δομή β-κλώνου.



Στην επόμενη εικόνα βλέπουμε τις α-έλικες και τους β-κλώνους που εμφανίζονται στις αντίστοιχες χρωματισμένες περιοχές.

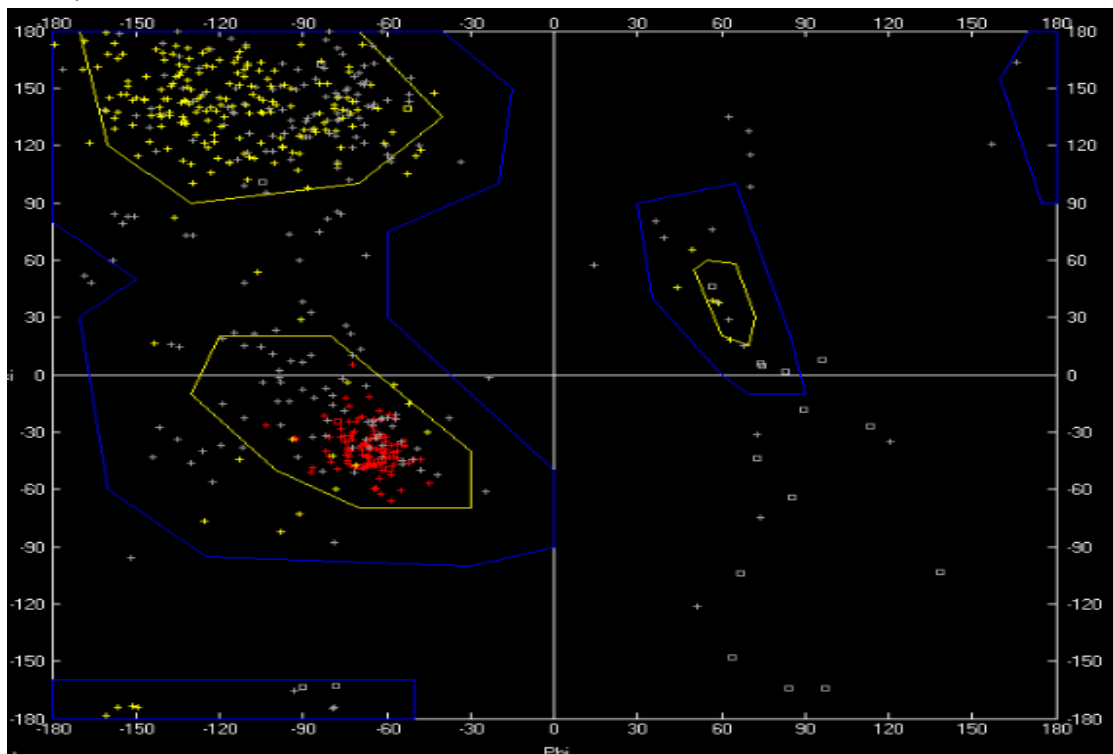


Στην συνέχεια, θα αναλύσουμε το διάγραμμα Ramachandran, όπου παρουσιάζονται οι επιτρεπτοί συνδιασμοί των γωνιών  $\phi$  και  $\psi$ . Αφού οι  $\phi$  και  $\psi$  αναφέρονται σε περιστροφές δύο άκαμπτων πεπτιδικών ομάδων γύρω από το ίδιο άτομο  $C_{\alpha}$ , οι περισσότεροι συνδιασμοί παράγουν στεreoχημικές παρεμποδίσεις είτε μεταξύ ατόμων σε διαφορετικές πεπτιδικές ομάδες είτε μεταξύ μίας πεπτιδικής ομάδας και της πλευρικής αλυσίδας που συνδέεται με το  $C_{\alpha}$ . Συνεπώς, αυτοί οι συνδιασμοί δεν επιτρέπονται. Να σημειωθεί ότι η γλυκίνη έχοντας μόνο ένα άτομο υδρογόνου σαν πλευρική αλυσίδα, μπορεί να υιοθετήσει ένα πιο ευρύ φάσμα στεροδιατάξεων απ'ότι τα άλλα κατάλοιπα.



Οι έγχωμες περιοχές δείχνουν στεreoχημικά επιτρεπτές περιοχές. Η περιοχή στο πάνω αριστερά τεταρτημόριο αντιστοιχεί σε γωνίες που συναντώνται σε β-κλώνους. Η περιοχή στο κάτω αριστερά τεταρτημόριο αντιστοιχεί σε δεξιόστροφες α-έλικες. Η περιοχή στο πάνω δεξιά τεταρτημόριο αντιστοιχεί στις πιο σπάνιες αριστερόστροφες έλικες.

Βλέπουμε το διάγραμμα Ramachandran και παρατηρούμε ότι οι κουκκίδες με κόκκινο χρώμα που αντιπροσωπεύουν την α-έλικα βρίσκονται στην αντίστοιχη περιοχή στο διάγραμμα Ramachandran. Το ίδιο ισχύει και για τις κίτρινες κουκκίδες που αντιστοιχούν στους β-κλώνους.



Καταγράφουμε τα αμινοξέα που βρίσκονται εκτός των περιοχών των επιτρεπτών γωνιών.

VAL 54	GLY 136	MET 170
TRP 76	ILE 138	GLY 187
GLY 90 (υπάρχουν δύο συνδιασμοί γωνιών)	SER 144	GLY 190
	LYS 145	GLY 220
GLY 104	PHE 146	ASN 221
GLY 115 (υπάρχουν δύο συνδιασμοί γωνιών)	LYS 167	TYR 222
GLY 131	TRP 169	PRO 234

#### BIBΛΙΟΓΡΑΦΙΑ

Βάσεις δεδομένων: PDB database, Prodom, Protparam, Uniprot

“Εισαγωγή στην δομή των πρωτεϊνών”, Carl Branden, John Tooze, Ακαδημαϊκές εκδόσεις  
Ι.Μπάσδρα και ΣΙΑ Ο.Ε., 2006

<http://www.cerm.unifi.it/ramachandran-potentials>

[http://www.cryst.bbk.ac.uk/PPS95/course/3\\_geometry/rama.html](http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html)