Εργασία Δομικής Βιοχημείας

Ονόματα Φοιτητών:
Τσιούρη Αθηνά, ΑΕΜ: 8432
Παπαθανασίου Θανάσης, ΑΕΜ: 8356

Όνομα πρωτεΐνης: 1EFE

Ακολουθίες δύο ισομερών:

>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens GN=INS PE=1 SV=1
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

>sp|F8WCM5|INSR2_HUMAN Insulin, isoform 2 OS=Homo sapiens GN=INS-IGF2 PE=2 SV=1
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED
LQASALSLSSSTSTWPEGLDATARAPPALVVTANIGQAGGSSSRQFRQRALGTSDSPVLF
IHCPGAAGTAQGLEYRGRRVTTELVWEEVDSSPQPQGSESLPAQPPAQPAPQPEPQQARE
PSPEVSCCGLWPRRPQRSQN

Πρώτο Ισομερές

ProtParam

User-provided sequence:

```
        1Ω         2Ω         3Ω         4Ω         5Ω         6Ω
MALWMRLLPL LALLALWGPD PAAAFVNQHL CGSHLVEALY LVCGERGFFY TPKTRREAED

        7Ω         8Ω         9Ω        10Ω        11Ω
LQVGQVELGG GPGAGSLQPL ALEGSLQKRG IVEQCCTSIC SLYQLENYCN
```

References and documentation are available.

Number of amino acids: 110

Molecular weight: 11980.9

Theoretical pI: 5.22

CSV format

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: $C_{535}H_{841}N_{143}O_{153}S_8$
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

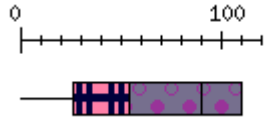
Grand average of hydropathicity (GRAVY): 0.193

ProDom

database: multiple alignments
Program: ncbi-blastp
Matrix: BLOSUM62
Expect: 0.01
Filter: 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

Domain ID	BEGIN	END	
PDC6E119	54	89	Submit
PD000491	26	53	Submit

PDB892H1	33	53	Submit
PDB2H9G3	90	109	Submit

Domain 3D modelling using Swiss-Model

Domain ID	BEGIN	END	
PDC6E119	54	89	Submit
PD000491	26	53	Submit
PDB892H1	33	53	Submit
PDB2H9G3	90	109	Submit

Domain 3D modelling using Geno3D

Domain ID	BEGIN	END	
PDC6E119	54	89	Submit
PD000491	26	53	Submit
PDB892H1	33	53	Submit
PDB2H9G3	90	109	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: unkwown
(110 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
26-53	#PD000491	143	2e-10
33-53	#PDB892H1	107	2e-05
54-89	#PDC6E119	175	3e-15
90-109	#PDB2H9G3	102	9e-05

>**PDC6E119** (Closest domain: INS_PONPY 54-89)
Number of domains in family: 38
Commentary (automatic):
RECNAME: FULL=INSULIN METABOLISM CHAIN CONTAINS: CARBOHYDRATE DISULFIDE BOND B A
Length = 36
Score = 175 (72.0 bits), Expect = 3e-15
Identities = 36/36 (100%), Positives = 36/36 (100%)

```
Query:     54 TRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR 89
              TRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR
Sbjct:     54 TRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR 89
```

>**PD000491** (Closest domain: INS_TRASC 2-29)
Number of domains in family: 272
Commentary (automatic):
GROWTH SECRETED FACTOR SUBNAME: FULL=INSULIN-LIKE RECNAME: FULL=INSULIN METABOLISM CONTAINS: CHAIN
Length = 28
Score = 143 (59.7 bits), Expect = 2e-10
Identities = 26/28 (92%), Positives = 27/28 (96%)

```
Query:     26 VNQHLCGSHLVEALYLVCGERGFFYTPK 53
              NQHLCGSHLVEALYLVCGERGFFY+PK
Sbjct:      2 ANQHLCGSHLVEALYLVCGERGFFYSPK 29
```

>**PDB892H1** (Closest domain: Q7LZN0_POLSP 9-29)
Number of domains in family: 52
Commentary (automatic):
RECNAME: FULL=INSULIN METABOLISM CHAIN CONTAINS: CARBOHYDRATE DISULFIDE BOND GLUCOSE A
Length = 21

```
Score = 107  (45.8 bits), Expect = 2e-05
Identities = 20/21 (95%), Positives = 20/21 (95%), Gaps = 1/21 (4%)

 Query:     33 SHLVEALYLVCGERGFFYTPK 53
               SHLVEALYLVCGERGFFYTP
 Sbjct:      9 SHLVEALYLVCGERGFFYTPN 29
```

```
>PDB2H9G3 (Closest domain: INS_BALPH 31-50)
Number of domains in family: 123
Commentary (automatic):
FULL=INSULIN RECNAME: METABOLISM CHAIN CONTAINS: CARBOHYDRATE
DISULFIDE BOND GLUCOSE HORMONE
Length = 20
Score = 102  (43.9 bits), Expect = 9e-05
Identities = 20/20 (100%), Positives = 20/20 (100%), Gaps = 1/20 (5%)

 Query:     90 GIVEQCCTSICSLYQLENYC 109
               GIVEQCCTSICSLYQLENYC
 Sbjct:     31 GIVEQCCTSICSLYQLENYC 50
```

Δεύτερο ισομερές

ProtParam

User-provided sequence:

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

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

       130        140        150        160        170        180
IHCPGAAGTA QGLEYRGRRV TTELVWEEVD SSPQPQGSES LPAQPPAQPA PQPEPQQARE
```

190 200 PSPEVSCCGL

WPRRPQRSQN

References and documentation are available.

Number of amino acids: 200

Molecular weight: 21537.2

Theoretical pI: 5.93

Amino acid composition: CSV format

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

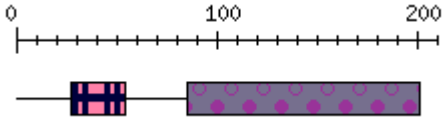
Grand average of hydropathicity (GRAVY): -0.335

ProDom

database: multiple alignments
Program: ncbi-blastp
Matrix: BLOSUM62
Expect: 0.01
Filter: 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

Domain ID	BEGIN	END	
PDB54014	85	200	Submit
PD000491	27	53	Submit
PDB892H1	33	53	Submit

Domain 3D modelling using Swiss-Model

Domain ID	BEGIN	END	
PD000491	27	53	Submit
PDB892H1	33	53	Submit

Domain 3D modelling using Geno3D

Domain ID	BEGIN	END	
PD000491	27	53	Submit

PDB892H1 33 53 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: unkwown
(200 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
27-53	#PD000491	149	8e-11
33-53	#PDB892H1	108	5e-05
85-200	#PDB54014	427	1e-50

>**PDB54014** (Closest domain: Q1WM24_HUMAN 85-200)
Number of domains in family: 3
Commentary (automatic):
METABOLISM SUBNAME: GLUCOSE CARBOHYDRATE TRANSCRIPT VARIANT FULL=INSIGF FULL=PUTATIVE UNCHARACTERIZED LONG
Length = 116
Score = 427 (169.1 bits), Expect = 1e-50
Identities = 88/116 (75%), Positives = 88/116 (75%)

```
 Query:     85
APPALVVTANIGQAGGSSSRQFRQRALGTSDSPVLFIHCPGAAGTAQGLEYRGRRVTTEL 144

APPALVVTANIGQAGGSSSRQFRQRALGTSDSPVLFIHCPGAAGTAQGLEYRGRRVTTE
 Sbjct:     85
APPALVVTANIGQAGGSSSRQFRQRALGTSDSPVLFIHCPGAAGTAQGLEYRGRRVTTEP 144
```

```
 Query:    145
VWEEVDSSPQPQGXXXXXXXXXXXXXXXXXXXXXXXXXXXVSCCGLWPRRPQRSQN 200
               VWEEVDSSPQPQG
VSCCGLWPRRPQRSQN
 Sbjct:    145
VWEEVDSSPQPQGSESLPAQPPAQPAPQPEPQQAREPSPEVSCCGLWPRRPQRSQN 200
```

```
>PD000491 (Closest domain: INS2_XENLA 25-52)
Number of domains in family: 272
Commentary (automatic):
GROWTH SECRETED FACTOR SUBNAME: FULL=INSULIN-LIKE RECNAME:
FULL=INSULIN METABOLISM CONTAINS: CHAIN
Length = 28
Score = 149  (62.0 bits), Expect = 8e-11
Identities = 25/27 (92%), Positives = 26/27 (96%), Gaps = 3/27 (11%)

 Query:     27 NQHLCGSHLVEALYLVCGERGFFYTPK 53
               NQHLCGSHLVEALYLVCG+RGFFY PK
 Sbjct:     26 NQHLCGSHLVEALYLVCGDRGFFYYPK 52
```

```
>PDB892H1 (Closest domain: Q7LZN0_POLSP 9-29)
Number of domains in family: 52
Commentary (automatic):
RECNAME: FULL=INSULIN METABOLISM CHAIN CONTAINS: CARBOHYDRATE
DISULFIDE BOND GLUCOSE A
Length = 21
Score = 108  (46.2 bits), Expect = 5e-05
Identities = 20/21 (95%), Positives = 20/21 (95%), Gaps = 1/21 (4%)

 Query:     33 SHLVEALYLVCGERGFFYTPK 53
               SHLVEALYLVCGERGFFYTP
 Sbjct:      9 SHLVEALYLVCGERGFFYTPN 29
```

ΑΜΙΝΟΞΕΑ ΠΟΥ ΕΙΝΑΙ ΕΚΤΟΣ:

A: GLY8

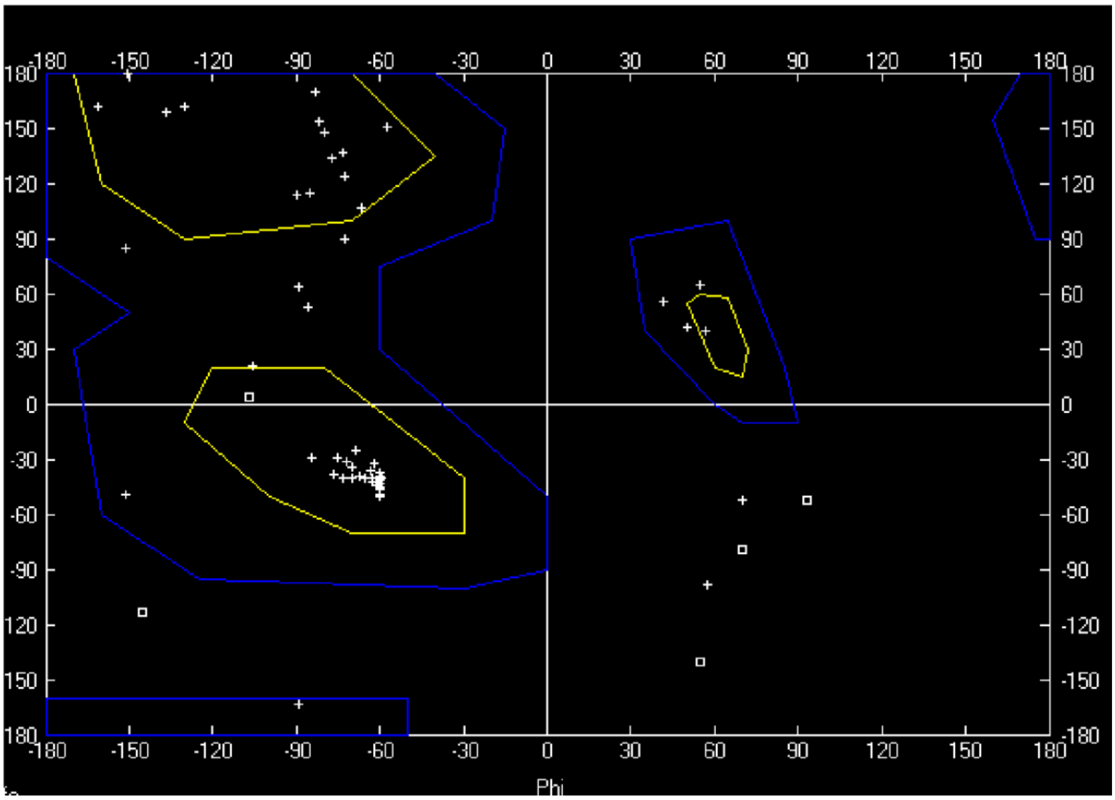
A:ARG32

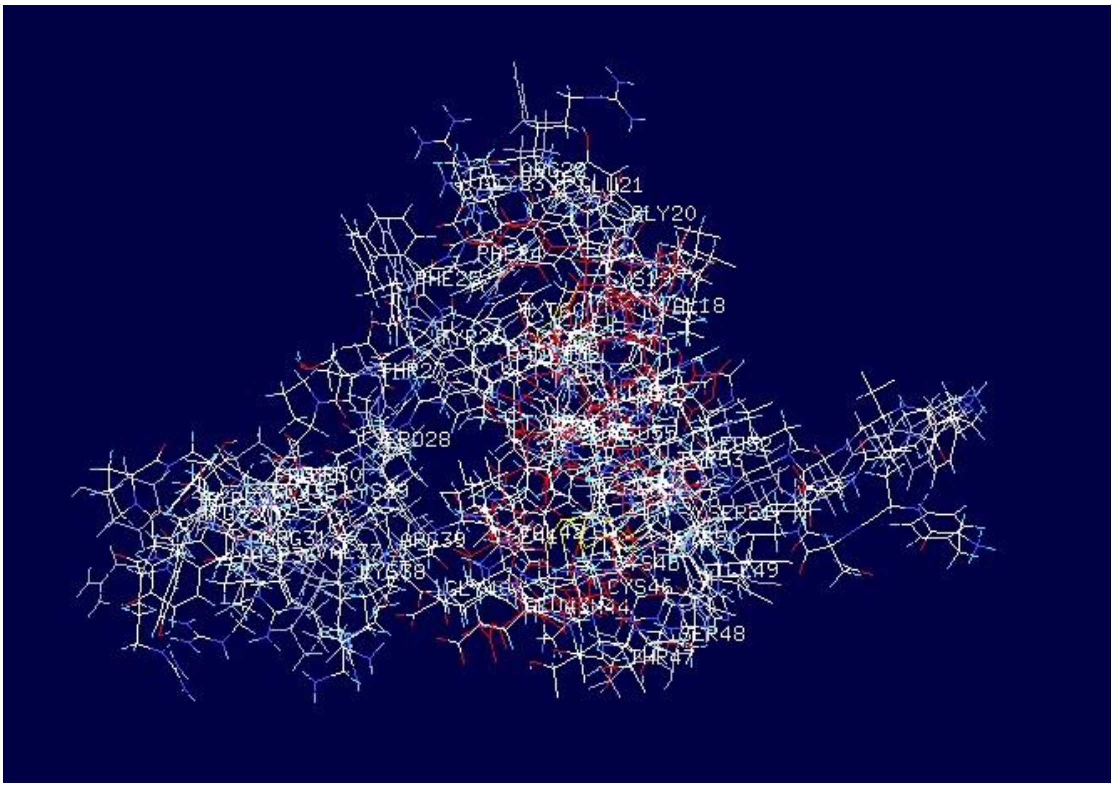
A:GLY20

A:THR30

A:GLY40

A:GLY35


-180 -150 -120 -90 -60 -30 0 30 60 90 120 150 180
180 150 120 90 60 30 0 -30 -60 -90 -120 -150 -180
Phi


GLY20
SER48