ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ AGRICULTURAL UNIVERSITY OF ATHENS

Βιοπληροφορική

Ενότητα 19:

Υπολογιστικός Προσδιορισμός Δομής (1/3), 2 ΔΩ

Τμήμα: Βιοτεχνολογίας

Όνομα καθηγητή: **Τ. Θηραίου**

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Μαθησιακοί Στόχοι

- κατανόηση της αναγκαιότητας και των εφαρμογών της υπολογιστικής πρόγνωσης δομής.
- επισκόπηση των μεθόδων πρόγνωσης δευτεροταγούς δομής.
- αναφορά στα μέτρα εκτίμησης της ακρίβειας πρόγνωσης.

Λέξεις Κλειδιά

- Λέξεις κλειδιά: Υπολογιστικός προσδιορισμός δομής, Πρόγνωση δευτεροταγούς δομής.
- Key words: Υπολογιστικός προσδιορισμός δομής, Πρόγνωση δευτεροταγούς δομής.

Υπολογιστικός Προσδιορισμός Δομής 1/4

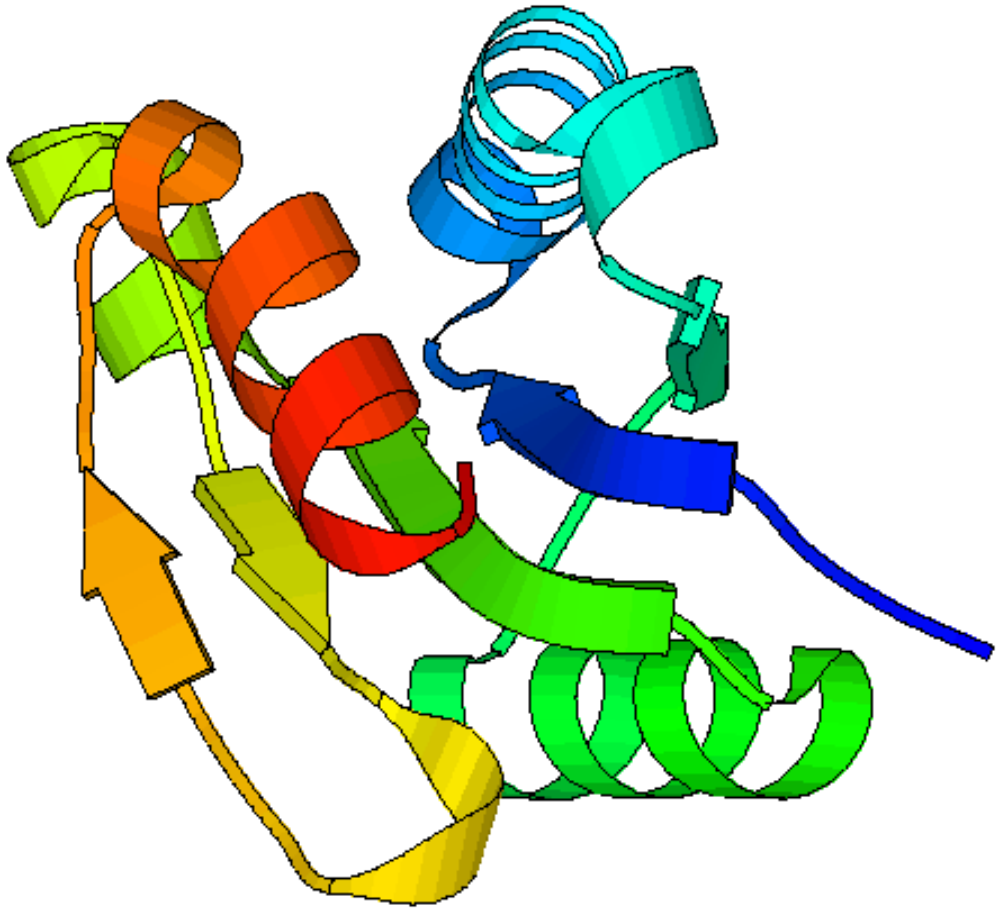
- πειραματικός προσδιορισμός δομών
  - κρυσταλλογραφία ακτίνων Χ
  - πυρηνικός μαγνητικός συντονισμός (NMR)
  - χρόνος / κόστος / περιορισμοί

- sequence - structure gap
  - Ο αριθμός των πειραματικά προσδιορισμένων δομών είναι πολύ μικρότερος του αριθμού των πρωτεϊνικών ακολουθιών.

Υπολογιστικός Προσδιορισμός Δομής 2/4

EKGPDLYLIPLTEEAVAEAF
YLAEALRPRLRAEYALAPRK
PAKGLEEALKRGAAFAGFLG
EDELRAGEVTLKRLATGEQV
RLSREEVPGYLLQALG

+ computer =



Υπολογιστικός Προσδιορισμός Δομής 3/4

- σχεδιασμός πειραμάτων για εύρεση λειτουργίας
- κατανόηση μηχανισμών λειτουργίας / επίδρασης μεταλλάξεων / γενετικών ασθενειών
- πρόβλεψη πρωτεϊνικών αλληλεπιδράσεων
- σχεδιασμός φαρμάκων
- μελέτη τρόπου αναδίπλωσης πρωτεϊνών / κατανόηση των αρχών της πρωτεϊνικής αρχιτεκτονικής
- σχεδιασμός πρωτεϊνών με προκαθορισμένες ιδιότητες / λειτουργία (protein design - protein engineering)

Υπολογιστικός Προσδιορισμός Δομής 4/4

- secondary structure (δευτεροταγής δομή)
- solvent accessibility (προσβασιμότητα του διαλύτη)
- protein disorder prediction
- transmembrane segments (διαμεμβρανικά τμήματα)
- inter-residue/strand contacts (επαφές μεταξύ καταλοίπων)

- homology or comparative modeling (προτυποποίηση πρωτεϊνών με ομολογία)
- fold recognition (αναγνώριση διπλώματος)
- ab initio prediction (απ' αρχής πρόγνωση)

} 3D

Πρόγνωση Δευτεροταγούς Δομής 1/20

- βελτίωση στοίχισης ακολουθιών
- χρήσιμη στην αναγνώριση διπλώματος
- διαχωρισμός στροφών / πυρήνα πρωτεϊνών

- ανάθεση στοιχείων δευτεροταγούς δομής (ΣΔΔ) σε πειραματικά προσδιορισμένες δομές
  - DSSP
    - δεσμοί υδρογόνου
  - STRIDE
    - δεσμοί υδρογόνου και γωνίες της κύριας αλυσίδας

Πρόγνωση Δευτεροταγούς Δομής 2/20

- Εκτίμηση Ακρίβειας Πρόγνωσης
  - Εφαρμογή των μεθόδων σε πειραματικά λυμένες δομές
  - Σύγκριση των αποτελεσμάτων της πρόγνωσης με τα ΣΔΔ
  - Qindex: (Qhelix, Qstrand, Qloop, Q3)
  - Matthews Correlation Coefficient (MCC)
  - Segment Overlap (SOV)

Πρόγνωση Δευτεροταγούς Δομής 3/20

- Εκτίμηση Ακρίβειας Πρόγνωσης
  - Qindex: (Qhelix, Qstrand, Qloop, Q3)
  - ποσοστό σωστά προβλεφθέντων καταλοίπων
  - π.χ.
    - ΣΔΔ L H H H H H H H H H H L
    - πρόβλεψη L H H H L H H H L H H L
    - Q3 = (10/12)*100% = 83.3%

Πρόγνωση Δευτεροταγούς Δομής 4/20

- Εκτίμηση Ακρίβειας Πρόγνωσης
  - Segment Overlap (SOV)
    - Είδος και θέση στοιχείων δευτεροταγούς δομής vs ανάθεση διαμόρφωσης ανά κατάλοιπο
    - Φυσική μεταβλητότητα των ορίων των ΣΔΔ σε οικογένειες ομόλογων πρωτεϊνών
    - Ασάφεια στον προσδιορισμό των άκρων των ΣΔΔ λόγω διαφοροποιήσεων στα κριτήρια ανάθεσής τους

Πρόγνωση Δευτεροταγούς Δομής 5/20

- Εκτίμηση Ακρίβειας Πρόγνωσης
  - Segment Overlap (SOV)

		Sov	Q_3
Observed	CHHHHHHHHHC		
Prediction 1	CHCHCHCHCCC	12.5	58.3
Prediction 2	CCCHHHHHCCCC	63.2	58.3
Prediction 3	CHHHCHHHCHHC	40.6	83.3
Prediction 4	CHHCCHHHHHCC	52.3	75.0
Prediction 5	CCCHHHHHHCCC	80.6	66.7

Πρόγνωση Δευτεροταγούς Δομής 6/20

- 1ης γενεάς

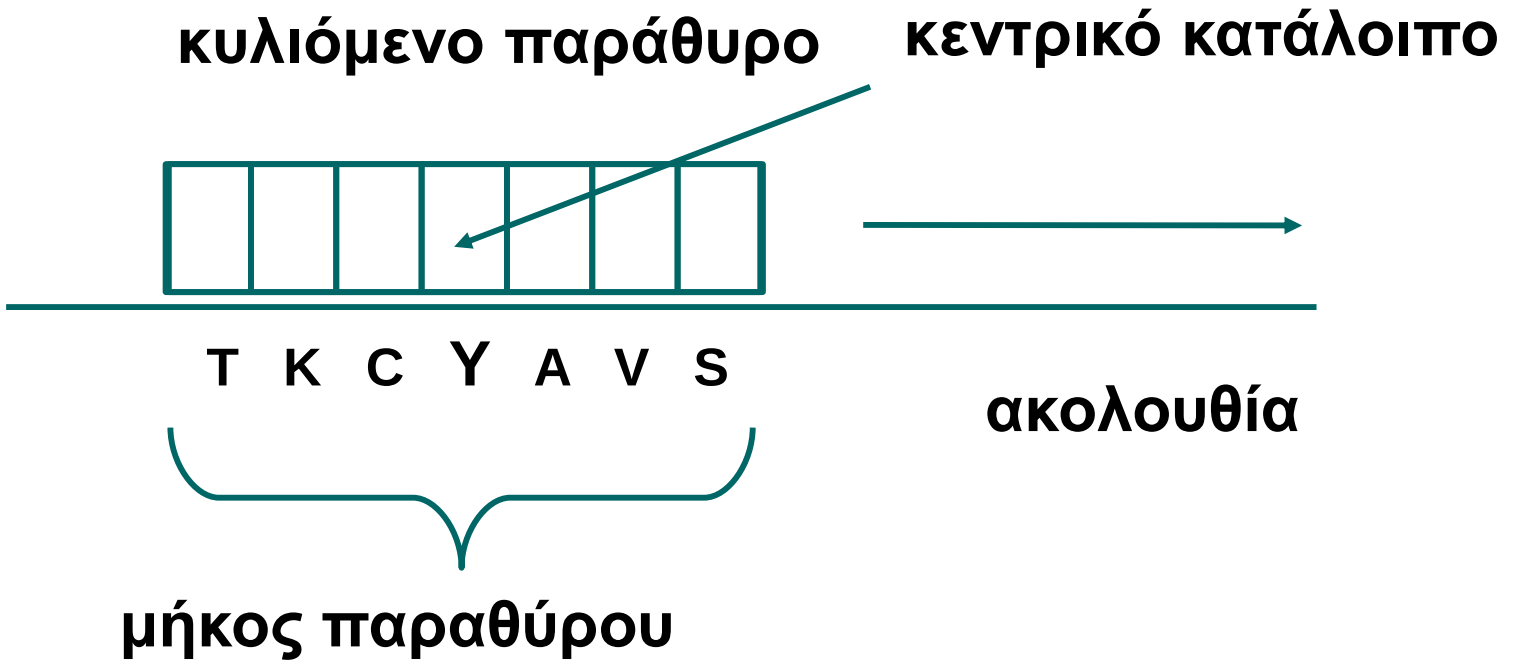
Name	P(H)	P(E)	P(turn)	f(i)	f(i+1)	f(i+2)	f(i+3)
Alanine	142	83	66	0.06	0.076	0.035	0.058
Arginine	98	93	95	0.07	0.106	0.099	0.085
Aspartic Acid	101	54	146	0.147	0.11	0.179	0.081
Asparagine	67	89	156	0.161	0.083	0.191	0.091
Cysteine	70	119	119	0.149	0.05	0.117	0.128
Glutamic Acid	151	37	74	0.056	0.06	0.077	0.064
Glutamine	111	110	98	0.074	0.098	0.037	0.098
Glycine	57	75	156	0.102	0.085	0.19	0.152
Histidine	100	87	95	0.14	0.047	0.093	0.054
Isoleucine	108	160	47	0.043	0.034	0.013	0.056
Leucine	121	130	59	0.061	0.025	0.036	0.07
Lysine	114	74	101	0.055	0.115	0.072	0.095
Methionine	145	105	60	0.068	0.082	0.014	0.055
Phenylalanine	113	138	60	0.059	0.041	0.065	0.065
Proline	57	55	152	0.102	0.301	0.034	0.068
Serine	77	75	143	0.12	0.139	0.125	0.106
Threonine	83	119	96	0.086	0.108	0.065	0.079
Tryptophan	108	137	96	0.077	0.013	0.064	0.167
Tyrosine	69	147	114	0.082	0.065	0.114	0.125
Valine	106	170	50	0.062	0.048	0.028	0.053

Πρόγνωση Δευτεροταγούς Δομής 7/20

- 1ης γενεάς: Q3 = 50-55%
- Καταγραφή πιθανοτήτων εύρεσης κάθε αμινοξέος σε συγκεκριμένο ΣΔΔ σε πειραματικά λυμένες δομές
- Δημιουργία κανόνων πρόγνωσης
- Ακρίβεια μεθόδου
  - αριθμός και ποιότητα των πειραματικά λυμένων δομών
  - ποιότητα κανόνων
- π.χ.
  - Chou and Fasman, 1974
  - GOR-1: Garnier, Osguthorpe, and Robson, 1978

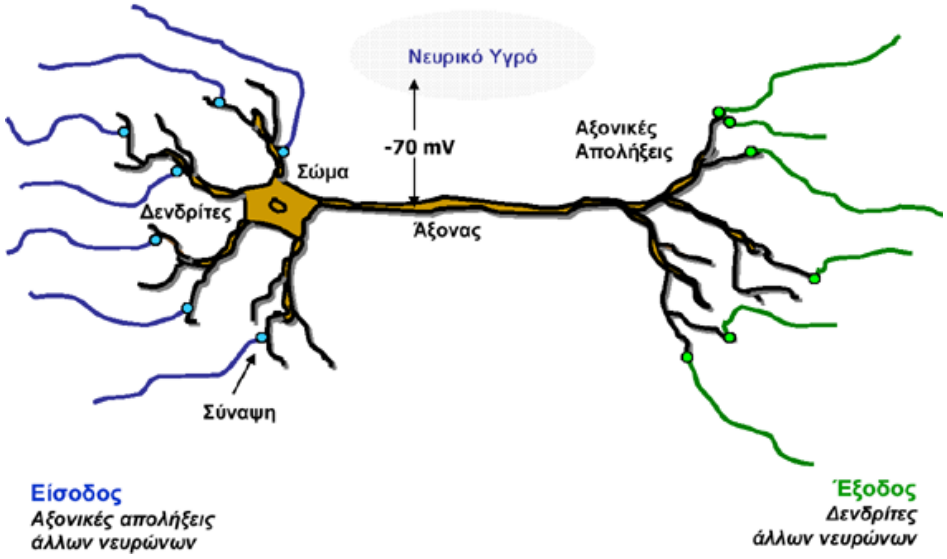
Πρόγνωση Δευτεροταγούς Δομής 8/20

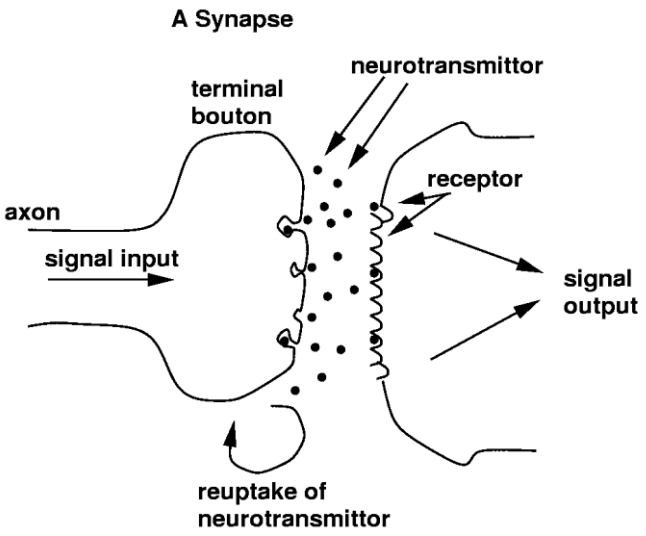
- 2ης γενεάς
  - Χρήση ενός κυλιόμενου παραθύρου κατά μήκος της ακολουθίας και πρόγνωση του ΣΔΔ στο οποίο βρίσκεται το κεντρικό κατάλοιπο του παραθύρου



Πρόγνωση Δευτεροταγούς Δομής 9/20

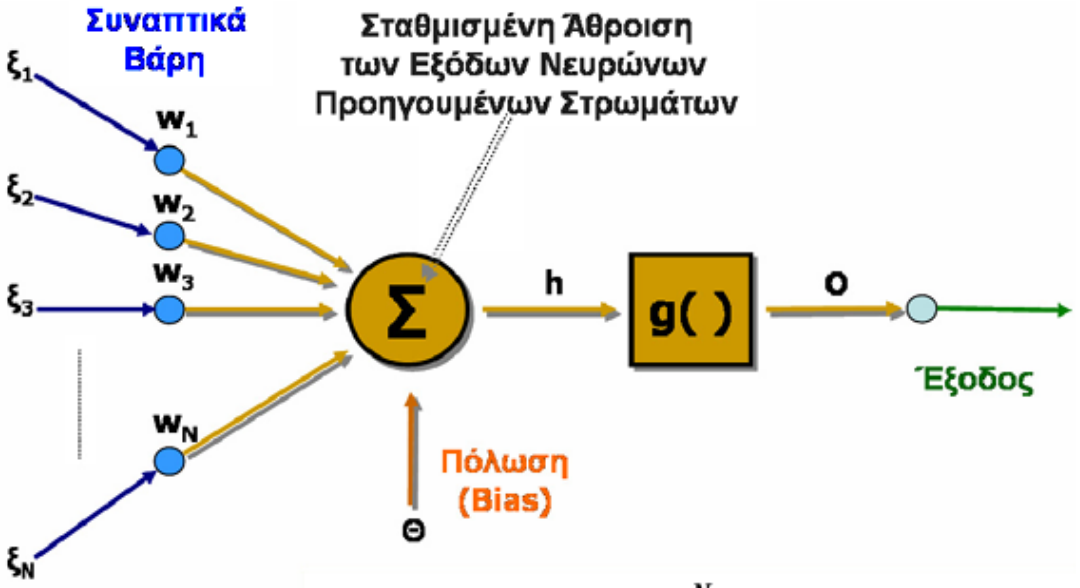
- 2ης γενεάς
  - Τεχνητά Νευρωνικά Δίκτυα
    - μάθηση βάσει εμπειρίας
      - μεταβολή της ισχύος των συνάψεων
      - προσθήκη / διαγραφή συνάψεων





Πρόγνωση Δευτεροταγούς Δομής 10/20

- 2ης γενεάς
  - Τεχνητά Νευρωνικά Δίκτυα



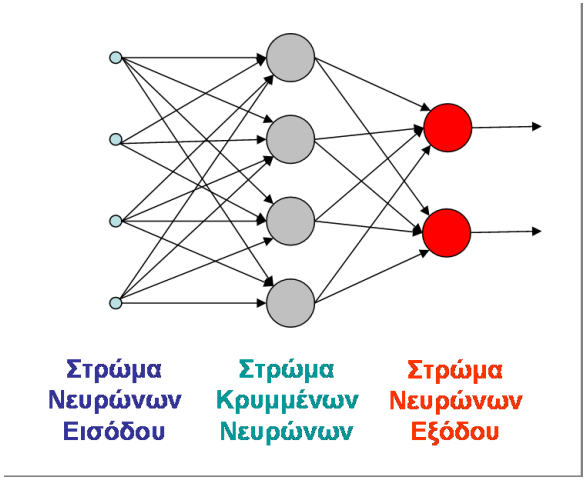
Ενεργοποίηση : $h = \sum_{k-1}^{N} w_k \xi_k + \theta$

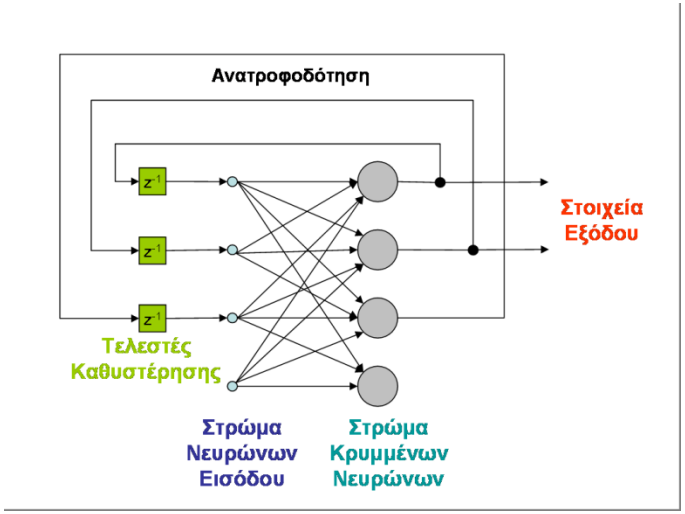
Συνάρτηση ενεργοποίησης: $g(\ )$

Έξοδος : $O = g(h) = g\left(\sum_{k=1}^{N} w_k \xi_k + \theta\right)$

Πρόγνωση Δευτεροταγούς Δομής 11/20

- 2ης γενεάς
  - Τεχνητά Νευρωνικά Δίκτυα
    - αρχιτεκτονική
      - πλήθος νευρώνων / διασυνδέσεις μεταξύ τους
      - οργάνωση νευρώνων κατά στρώματα
      - τρόπος διάδοσης σήματος





Πρόγνωση Δευτεροταγούς Δομής 12/20

- 2ης γενεάς
  - Μάθηση υπό Εποπτεία (Supervised Learning)
    - ΤΝΔ καθορισμένης αρχιτεκτονικής
    - σύνολο δεδομένων εκπαίδευσης γνωστής τιμής εξόδου (training set)
    - προσαρμογή των συναπτικών βαρών ώστε για συγκεκριμένη είσοδο, η απόκριση του δικτύου να ταυτίζεται με τη γνωστή τιμή εξόδου
    - στόχος:
      - γενίκευση σε σύνολο δεδομένων άγνωστης τιμής εξόδου

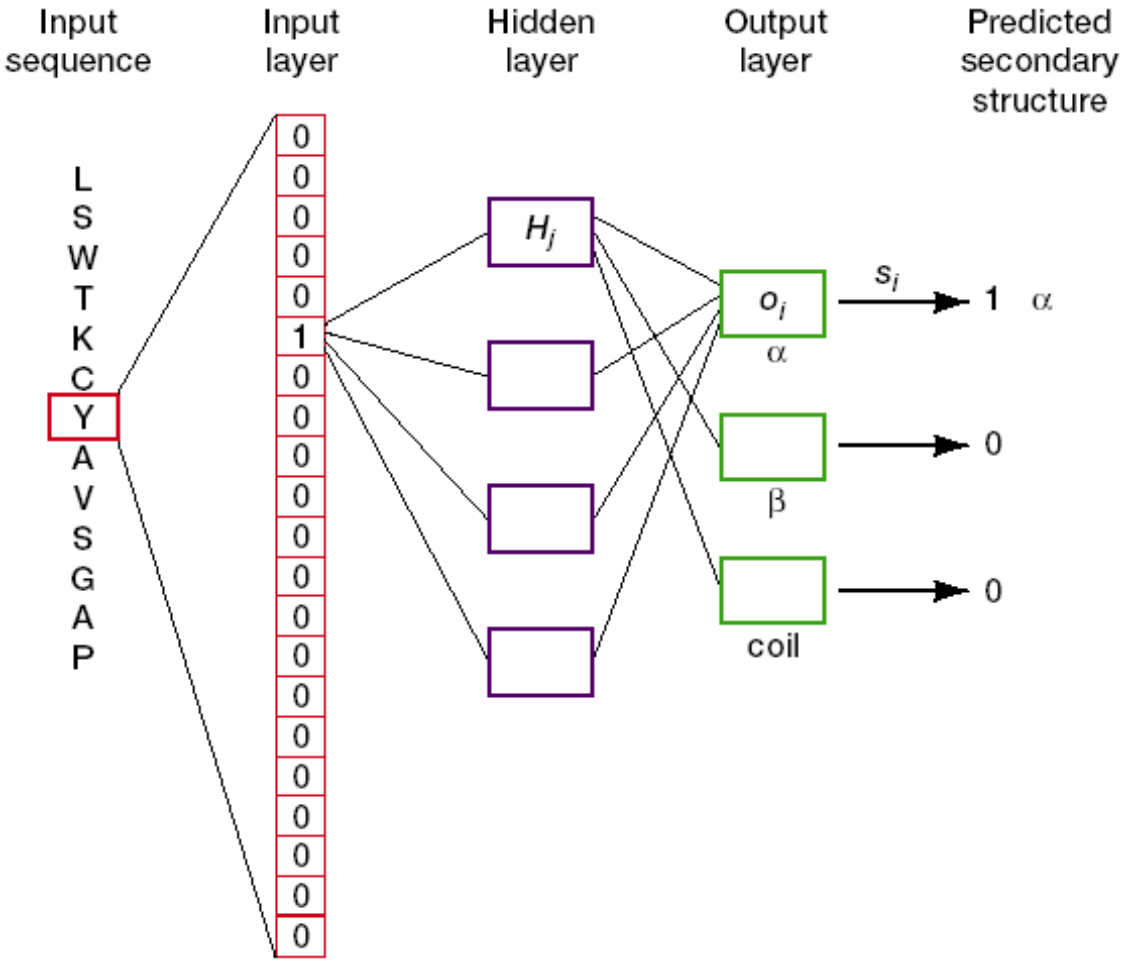
Πρόγνωση Δευτεροταγούς Δομής 13/20

- 2ης γενεάς
  - Κωδικοποίηση Δεδομένων Εισόδου
    - 1-out-of-21 (20 αμινοξέα και 1 τέλος ακολουθίας)
    - π.χ. LVA$ ($ = τέλος ακολουθίας)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	$
L	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

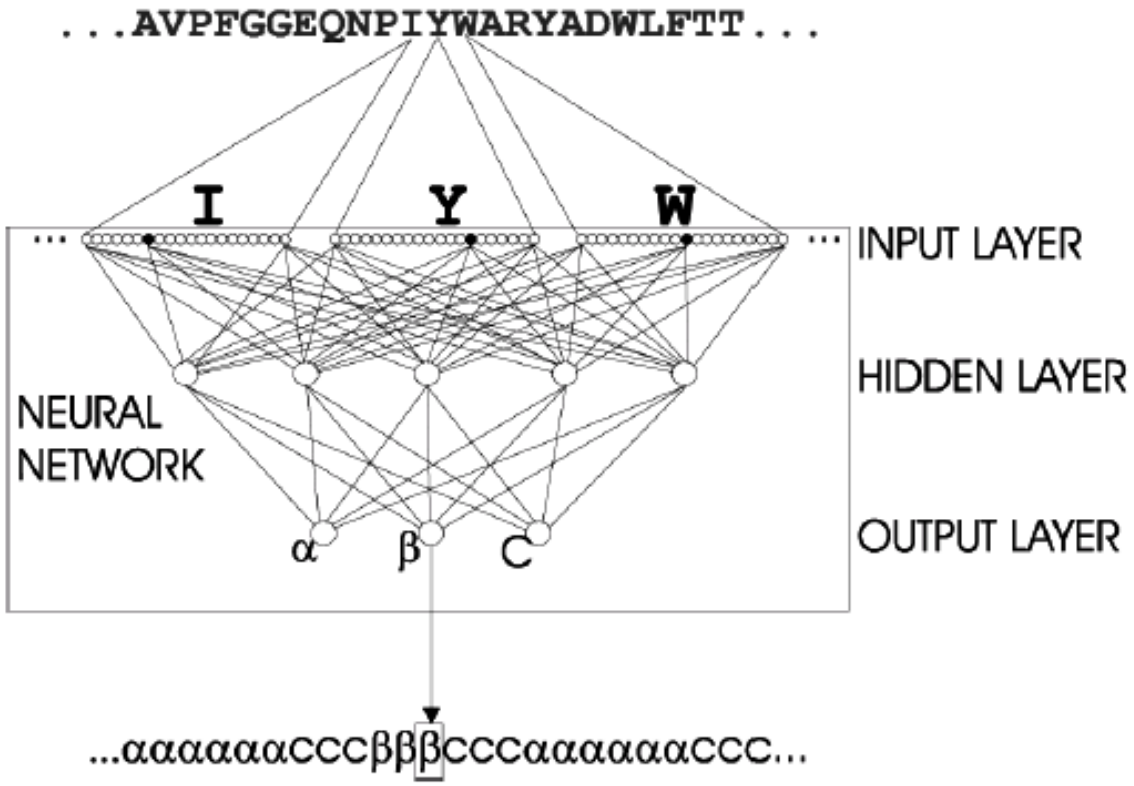
Πρόγνωση Δευτεροταγούς Δομής 14/20

- 2ης γενεάς



Πρόγνωση Δευτεροταγούς Δομής 15/20

- 2ης γενεάς: Q3 < 70%
  - GORIII
  - COMBINE



Πρόγνωση Δευτεροταγούς Δομής 16/20

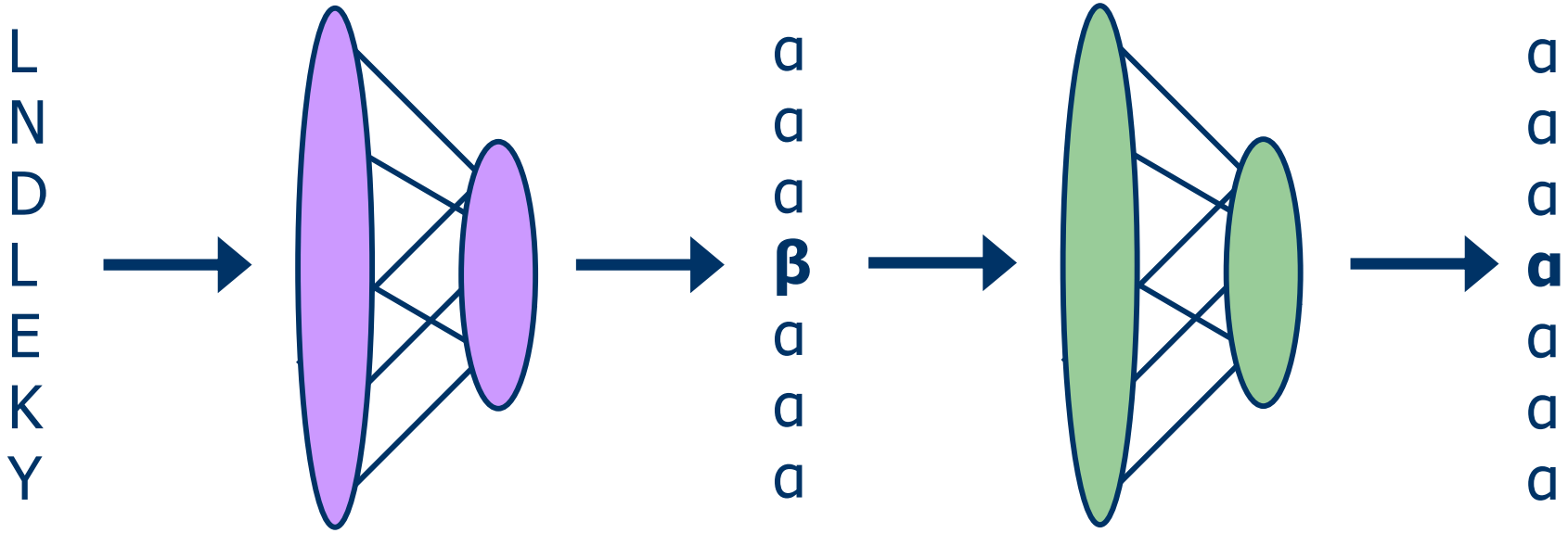
- 3ης γενεάς
  - Κωδικοποίηση Δεδομένων Εισόδου
    - βάσει πολλαπλής στοίχισης ακολουθιών
    - Η δομή είναι καλύτερα συντηρημένη από την ακολουθία.
    - Για ένα σύνολο ομόλογων ακολουθιών
      - Μεταλλάξεις αμινοξέων
      - Συντήρηση στοιχείων δευτεροταγούς δομής
    - Κωδικοποίηση αμινοξέων βάσει της συχνότητας εμφάνισής τους στο προφίλ της πολλαπλής στοίχισης.

Πρόγνωση Δευτεροταγούς Δομής 17/20

- 3ης γενεάς
  - βήματα
    - Αναζήτηση ομόλογων ακολουθιών σε βάσεις δεδομένων
      - PSI-BLAST, SAM-T2K
    - Πολλαπλή στοίχιση ακολουθιών και κωδικοποίηση δεδομένων
      - PSSMs, HMM models
    - Πρόγνωση δευτεροταγούς δομής βάσει της πολλαπλής στοίχισης
      - Single methods
      - Consensus methods

Πρόγνωση Δευτεροταγούς Δομής 18/20

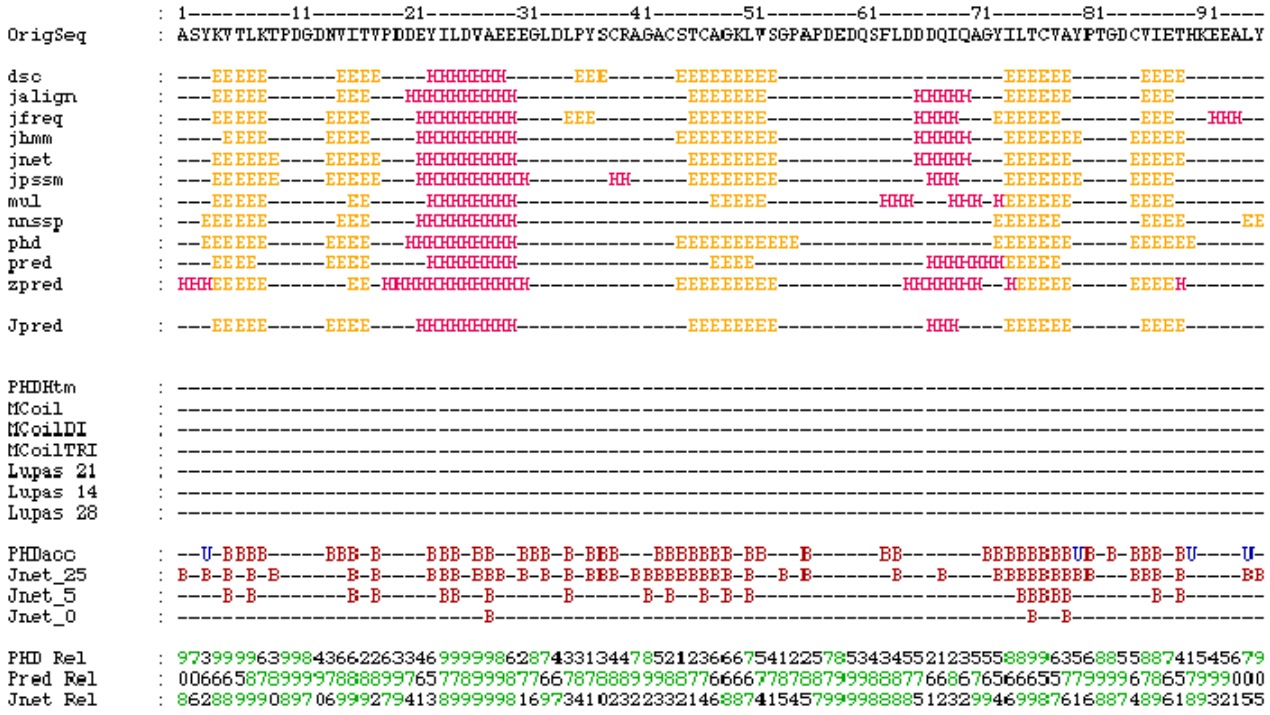
- 3ης γενεάς
  - PHD (Rost et. al.): Q3 = 72 - 76 %



Sequence to Structure Net **Structure to Structure Net**

Πρόγνωση Δευτεροταγούς Δομής 19/20

- 3ης γενεάς
  - JPRED: Consensus Prediction (Συναινετική Πρόβλεψη)



Πρόγνωση Δευτεροταγούς Δομής 20/20

- Q3 = 72-77% ± 11 %

- PredictProtein http://www.predictprotein.org/
- Jpred http://www.compbio.dundee.ac.uk/www-jpred/
- PSIPRED http://bioinf.cs.ucl.ac.uk/psipred/
- SOPMA http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
- SAM-T08 http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html

Protein Disorder Prediction

- Database of Protein Disorder (DisProt) http://www.disprot.org/index.php
- PONDR http://www.pondr.com/
- DISOPRED2 http://bioinf.cs.ucl.ac.uk/disopred/
- RONN http://www.strubi.ox.ac.uk/RONN
- metaPrDOS http://prdos.hgc.jp/cgi-bin/meta/top.cgi

Διαμεμβρανικές Πρωτεΐνες

- TMHMM http://www.cbs.dtu.dk/services/TMHMM/
- MetaTM http://metatm.sbc.su.se/
- TMBpro http://tmbpro.ics.uci.edu/

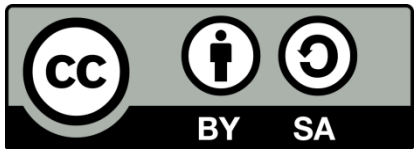
Βιβλιογραφία

- David Mount, "Bioinformatics: Sequence and Genome Analysis", Cold Spring Harbor Laboratory Press; 2nd edition (March 12, 2013)
- Jonathan Pevsner, "Bioinformatics and Functional Genomics", Wiley-Blackwell; 2nd edition (May 4, 2009)
- Andreas D. Baxevanis, B. F. Francis Ouellette, "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins", Wiley-Interscience; 3rd edition (October 29, 2004)
- Jenny Gu, Philip E. Bourne, "Structural Bioinformatics", Wiley-Blackwell; 2nd edition (March 16, 2009)

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.

- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα Γεωπονικού Πανεπιστημίου Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.