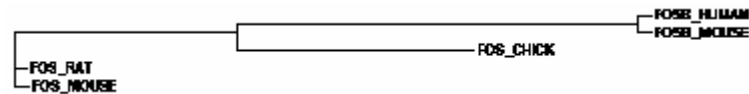




## ΒΙΟΜΕΤΡΙΑ-ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

```
FOS_RAT      KIPSGFNIDYELSSRCSSLRFLEDSLSTYRSPADSPSSDQSPVNTQDFCLDLVSVSSLNF 60
FOS_MOUSE   KIPSGFNIDYELSSRCSSLRFLEDSLSTYRSPADSPSSDQSPVNTQDFCLDLVSVSSLNF 60
FOS_CHICK   KIPSGFNIDYELSSRCSSLRFLEDSLSTYRSPADSPSSDQSPVNTQDFCLDLVSVSSLNF 60
FOS_HUMAN   -KIQGFPAGLYEAPSSRCSSLRFLEDSLSTYRSPADSPSSDQSPVNTQDFCLDLVSVSSLNF 54
FOS_MOUSE   -KIQGFPAGLYEAPSSRCSSLRFLEDSLSTYRSPADSPSSDQSPVNTQDFCLDLVSVSSLNF 54
```



ΗΛΙΑΣ ΖΙΝΤΖΑΡΑΣ, M.Sc., Ph.D.

ΚΑΘΗΓΗΤΗΣ

ΒΙΟΜΑΘΗΜΑΤΙΚΩΝ-ΒΙΟΜΕΤΡΙΑΣ

ΧΡΥΣΟΥΛΑ ΔΟΞΑΝΗ, MD, MSc, PhD.

Εντεταλμένος Διδάσκων

# ΠΕΡΙΕΧΟΜΕΝΑ

- Σχέση Βιοπληροφορικής και Βιομετρίας
- Βασική Βιολογία
- Πληροφοριακά συστήματα Γονιδιώματος
- Ζευγαρωτή αντιστοιχία αλληλουχιών
- Πολλαπλή αντιστοιχία αλληλουχίας
- Πηγές δεδομένων για πρωτεΐνες
- Δευτεροταγείς βάσεις δεδομένων
- Σύνθετες βάσεις δεδομένων αλληλουχιών πρωτεϊνών
- Σύγκριση δομών πρωτεϊνών
- Βάσεις δεδομένων δομών
- Γενετικές μελέτες συσχέτισης
- Φαρμακογενωμικές μελέτες

# Bioinformatics

- a scientific subdiscipline that involves using **computer technology** to collect, store, analyze and disseminate biological data and information.

*(retrieved by National Human Genome Research Institute)*

*“Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline.”*

National Center for Biotechnology Information  
(NCBI)

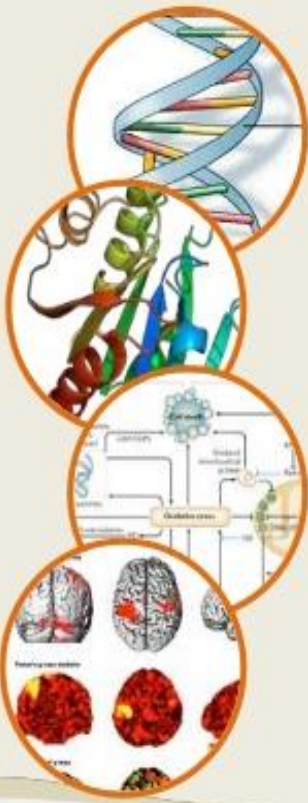
# Τί είναι Βιοπληροφορική

- Η βιοπληροφορική είναι η υπολογιστική διαχείριση και στατιστική ανάλυση βιολογικών αλληλουχιών (DNA, RNA, πρωτεΐνες) και δεδομένων που αναφέρονται σε τρισδιάστατες απεικονίσεις πρωτεϊνών.

Η Βιοπληροφορική ασχολείται με:

- Τη δημιουργία βάσεων δεδομένων για την αποθήκευση και διαχείριση μεγάλου όγκου δεδομένων.
- Την ανάπτυξη αλγορίθμων και βιομετρικών μεθόδων για τον προσδιορισμό σχέσεων μεταξύ αλληλουχιών ή δομών που ανήκουν σε μεγάλα σύνολα δεδομένων.
- Την στατιστική ανάλυση και ερμηνεία των διαφορετικών τύπων δεδομένων (DNA, RNA, πρωτεϊνική αλληλουχία και δομή).

# Areas where bioinformatics is applied



## Genomics

- Genomic feature prediction
- Sequencing data analysis

## Proteomics








- Protein 3D structure modeling
- Drug design

## Systems Biology

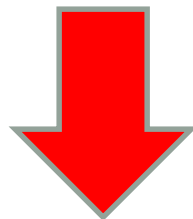
- Gene set enrichment
- Pathway analysis

## Phenotype

- Image analysis
- Integration

		Breadth: Homologs, Large-scale Surveys, Informatics—				
			pairwise comparison, sequence & structure alignment	multiple alignment, patterns, templates, trees	databases, scoring schemes, censuses	
		1	2	3-100	100+	
Depth: Rational Drug Design (physics) →		<b>Genome Sequence</b>	atc gatc gatattgggattgggga	atc gatc gatattgggattgggga atc gatc gatattgggattgggga	atc gatc gatattgggattgggga atc gatc gatattgggattgggga atc gatc gatattgggattgggga atc gatc gatattgggattgggga atc gatc gatattgggattgggga	atc gatc gatattgggattgggga atc gatc gatattgggattgggga atc gatc gatattgggattgggga atc gatc gatattgggattgggga atc gatc gatattgggattgggga
	gene finding	↓				
		<b>Protein Sequence</b>	ALMNAKKKPPQRT	ALMNAKKKPPQRT ALMNAKKKPPQRT	ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT	ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT ALMNAKKKPPQRT
	structure prediction	↓				
		<b>Protein Structure</b>				
	geometry calculation	↓				
		<b>Protein Surface</b>				
	molecular simulation	↓				
		<b>Force Field</b>				
	structure docking	↓				
	<b>Ligand Complex</b>					

- **Δεδομένα:** ακατέργαστα, μη οργανωμένα αποτελεσματα πρέπει να υποβληθούν σε επεξεργασία
- **Πληροφορίες:** Όταν τα δεδομένα υποβάλλονται σε επεξεργασία, οργανώνονται δομημένα ή παρουσιάζονται σε ένα δεδομένο πλαίσιο έτσι ώστε να είναι χρήσιμα




- **Βάση δεδομένων:**  
μια οργανωμένη συλλογή δεδομένων που αποθηκεύονται και έχει πρόσβαση ηλεκτρονικά (*retrieved from Wikipedia*)
- Αποτελείται από βασικές μονάδες που ονομάζονται *records or entries*



## **Βάσεις Δεδομένων**

Μια Βάση Δεδομένων (ΒΠ) είναι μια υπολογιστική αποθήκη από δεδομένα που παρέχει ένα προτυποποιημένο τόπο ανάκτησης, προσθήκης, διαγραφής και αλλαγής δεδομένων

- Biological databases----→ store of biological information
- Genetic database -----→ store of genetic data (genes, gene products, variants, phenotypes)
- enable users to retrieve data, to add data and to extract information from the data.

 developing databases to deal with gigantic volumes of biological data is a fundamentally essential task in bioinformatics.

>1552 databases that are publicly accessible online

(Nucleic Acids Research, 2015)

*Υπάρχουν δύο μεγάλες κατηγορίες βάσεων πληροφοριών :*

- οι σχεσιακές και
  - οι αντικειμενοστραφείς.
- 
- Οι σχεσιακές ΒΠ τοποθετούν τα δεδομένα σε πίνακες όπου κάθε γραμμή αναφέρεται σε συγκεκριμένο δεδομένο και κάθε στήλη αναφέρεται σε συγκεκριμένο χαρακτηριστικό των δεδομένων. Για τους πίνακες αυτούς δημιουργούνται δείκτες και συσχετίζονται μεταξύ τους, έτσι ώστε κάθε δεδομένο στη ΒΠ να έχει ένα μοναδικό σύνολο χαρακτηριστικών που να το προσδιορίζουν.

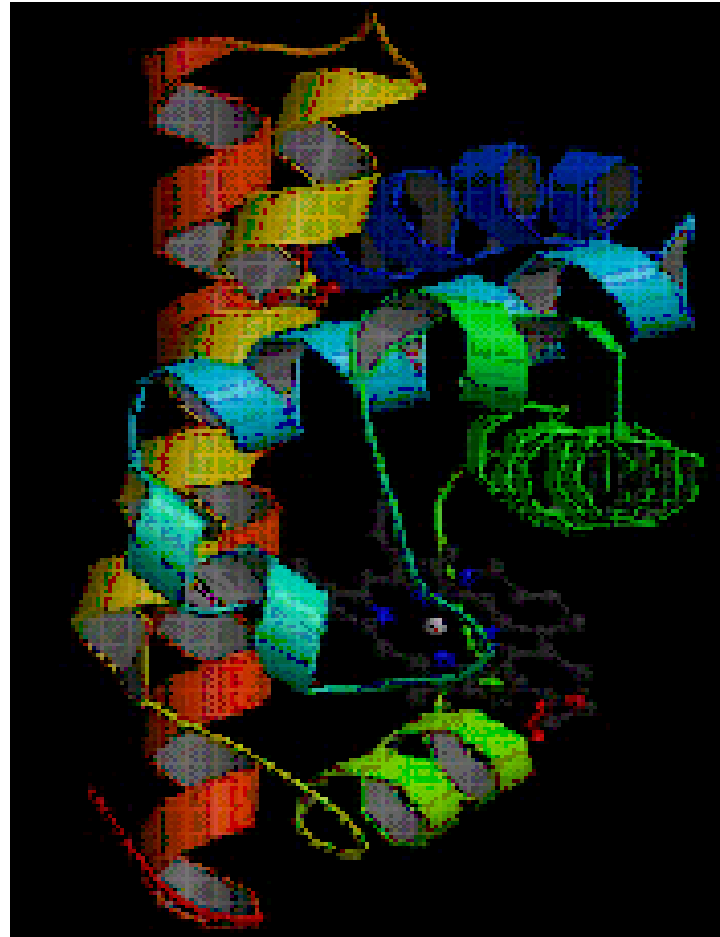
	Name
Protein1	
Protein2	
Protein3	
Protein4	

	Sequence	Polarity
Protein1		
Protein2		
Protein3		
Protein4		

	a-helices	b-strands	residues
Protein1			
Protein2			
Protein3			
Protein4			

- Η αντικειμενοστραφής ΒΠ αποτελείται από αντικείμενα (γονίδια ή πρωτεΐνες) που το καθένα έχει ένα συσχετιζόμενο σύνολο από τυποποιημένα εργαλεία για ανάλυση και αναπαράσταση του αντικειμένου και ένα σύνολο από χαρακτηριστικά όπως ένα όνομα που να το προσδιορίζει ή μια παραπομπή.

- Myoglobin



# Το Διαδίκτυο

- Το Διαδίκτυο είναι ένα παγκόσμιο δίκτυο υπολογιστών που συνδέει δημόσια, εκπαιδευτικά και ιδιωτικά ιδρύματα.
- Η μετάδοση της πληροφορίας επιτυγχάνεται με τη βοήθεια ενός πρωτοκόλλου γνωστού ως TCP/IP, το οποίο προσδιορίζει τον τρόπο με τον οποίον:
  - 1) τα δεδομένα χωρίζονται σε πακέτα,
  - 2) τα πακέτα φτάνουν στον προσδιορισμό τους και
  - 3) τα πακέτα συναρμολογούνται.
- Το είδος αυτής της επικοινωνίας επιτρέπει διαφορετικά είδη υπολογιστών και λειτουργικών συστημάτων το «μιλήσουν» μεταξύ τους με έναν κοινό τρόπο.

## Διευθύνσεις IP

- Για να μπορέσουν οι υπολογιστές σε ένα δίκτυο να επικοινωνήσουν, δίνεται σε κάθε υπολογιστή ένας μοναδικός αριθμός (IP διεύθυνση), ο οποίος κωδικοποιείται με μια μορφή δεκαδικής τελείας (π.χ. ένας υπολογιστής στο Διαδίκτυο μπορεί να έχει την IP Διεύθυνση 147.30.32.50).
- Επειδή η μορφή διευθύνσεων δεν είναι και πολύ φιλική στον άνθρωπο, σε κάθε τέτοια διεύθυνση έχει δοθεί ένα όνομα, το οποίο προσδιορίζει
  - έναν συγκεκριμένο υπολογιστή,
  - τον τόπο που ο υπολογιστής αυτός βρίσκεται, και τέλος
  - το διαδικτυακό μέρος στο οποίο ανήκει αυτός ο τόπος

(η παραπάνω διεύθυνση αντιστοιχεί στο όνομα ebi.ac.uk: Το Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής (European Bioinformatics Institute) είναι ένας ακαδημαϊκός τόπος του ευρύτερου διαδικτυακού τόπου UK).



## Παγκόσμιος Προσδιοριστής Πόρων (ΠΠΠ) (URL)

- Ο Παγκόσμιος Ιστός βασίζεται στο ότι κάθε ιστοσελίδα υπερκειμένου έχει έναν Παγκόσμιο Προσδιοριστή Πόρων (Uniform Resource Locator-URL). Ο ΠΠΠ περιέχει διάφορα μέρη, τα οποία προσδιορίζουν
  - το πρωτόκολλο επικοινωνίας (http – hypertext transfer protocol),
  - τον διακομιστή ιστοσελίδων,
  - τον κατάλογο και τέλος
  - την ιστοσελίδα.

Για παράδειγμα :

- Ο <http://www.ebi.ac.uk/hinxton/hinxton.html> προσδιορίζει σαν
  - πρωτόκολλο επικοινωνίας το http, σαν
  - Διακομιστή Ιστοσελίδων το (ebi) και σαν
  - κατάλογο το hinxton και τέλος σαν
  - ιστοσελίδα την hinxton.html.

# Διαδικτυακοί Τόποι Βιοπληροφορικής

- Τρεις τόποι για να αρχίσει κανείς την αναζήτηση βιοπληροφοριών στο Διαδίκτυο είναι οι εξής :
- <http://www.expasy.ch/>: The ExPASy (Expert Protein Analysis System) Molecular Biology Server. Ο συγκεκριμένος διαδικτυακός τόπος διατηρείται από το Ελβετικό Ίδρυμα Βιοπληροφορικής. Περιέχει συνδέσμους, ΒΠ και λογισμικό για την ανάλυση πρωτεϊνικών αλληλουχιών και δομών.
- <http://www.ebi.ac.uk/>: The EMBL European Bioinformatics Institute outstation. Ο τόπος αυτός αποτελεί μία συλλογή ΒΠ με βιολογικό περιεχόμενο και παρεμφερές λογισμικό.
- <http://www.ncbi.nlm.nih.gov/>: The National Center for Biotechnology Information. Ο τόπος περιέχει συλλογή δημοσίων ΒΠ , εργαλείων για βιοπληροφορική και διαφόρων εφαρμογών. Παρέχει επίσης συνδέσμους σε πολλούς χρήσιμους διαδικτυακούς τόπους και συνδέσμους για λογισμικό βιοπληροφορικής.

- Ένα άλλο παράδειγμα ιστοσελίδας που είναι η διεπαφή μιας ΒΠ για πρωτεΐνες και μπορούμε να προσπελάσουμε βρίσκεται στο <https://www.expasy.org/resources/uniprotkb-swiss-prot> .
- Επιλέγοντας το “at EBI” που είναι ένας υπερσύνδεσμος και επιτρέπει να συνδεθούμε στην ιστοσελίδα του Ευρωπαϊκού Ινστιτούτου Βιοπληροφορικής.



e.g. [BLAST](#), [UniProt](#), [MSH6](#), [Albumin](#)...

# UniProtKB/Swiss-Prot



UniProtKB/Swiss-Prot is the expertly curated component of UniProtKB (produced by the UniProt consortium). It contains hundreds of thousands of protein descriptions, including function, domain structure, subcellular location, post-translational modifications and functionally characterized variants.

💡 UniProt is one of the most widely used protein information resources in the world.

 [Watch the video](#)

[Browse the resource website](#)

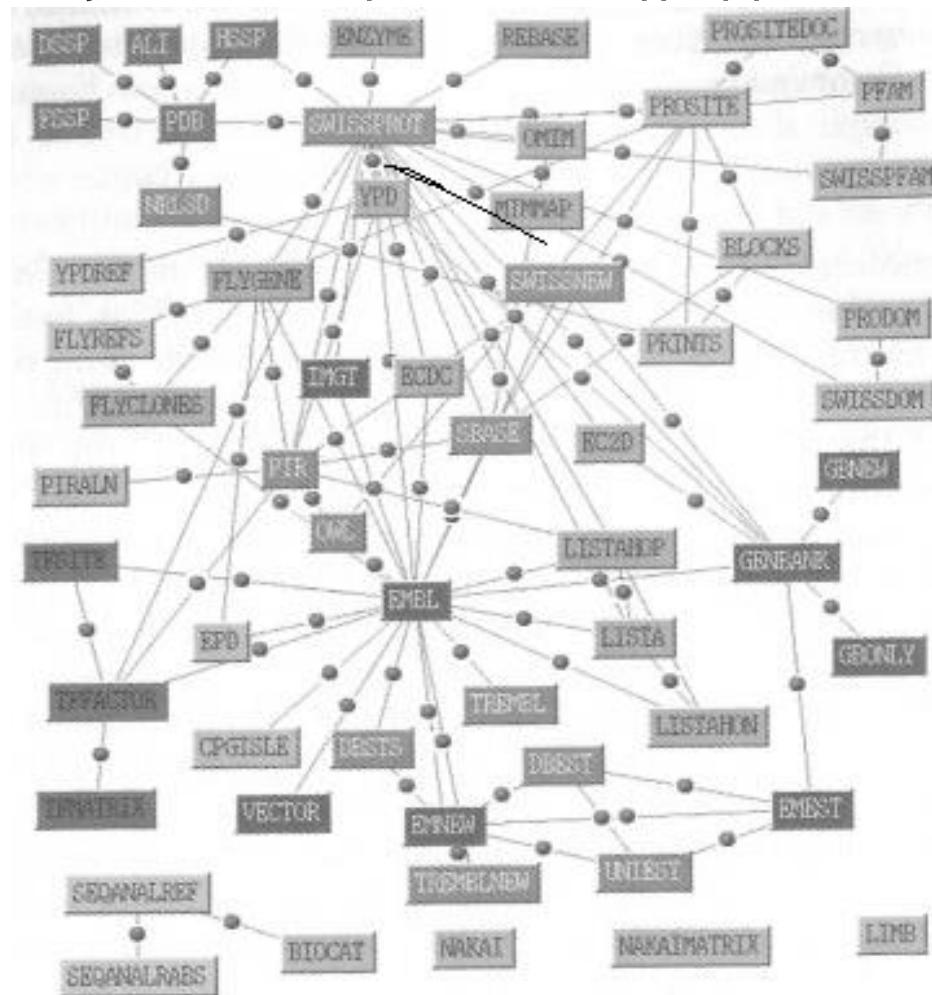
## What you can do with this resource

[Query and retrieval](#),  
[Protein feature detection](#)

## Browse these keywords in Expasy

[Protein sites, features and motifs](#),  
[Data submission, annotation and curation](#),  
[Proteomics](#),  
[Molecular interactions, pathways and networks](#),  
[Function analysis](#),

- Καθένας από τους παραπάνω διαδικτυακούς τόπους έχει πολλούς συνδέσμους σε άλλους παρόμοιους τόπους βιοπληροφορικής.
- Ένα δίκτυο από πηγές βιοπληροφόρησης παρουσιάζεται στο παρακάτω διάγραμμα



## Μηχανές Αναζήτησης (Search Engines)

- Για να βρούμε οποιοδήποτε είδος πληροφορίας βιολογικού ενδιαφέροντος στο Διαδίκτυο , θα πρέπει να χρησιμοποιήσουμε μηχανές αναζήτησης πληροφορίας. Οι μηχανές αναζήτησης ψάχνουν στις σελίδες του Διαδικτύου για συγκεκριμένες λέξεις κλειδιά και συγκεκριμένες φράσεις. Μια τέτοια μηχανή αναζήτησης είναι το Google. Για να χρησιμοποιήσουμε το Google, γράφουμε [www.google.com](http://www.google.com) στο Internet Explorer.
- Αν πληκτρολογήσουμε πχ τη λέξη κλειδί “protein structure”, εμφανίζονται σελίδες που περιέχουν πληροφορίες σχετικά με την λέξη κλειδί.