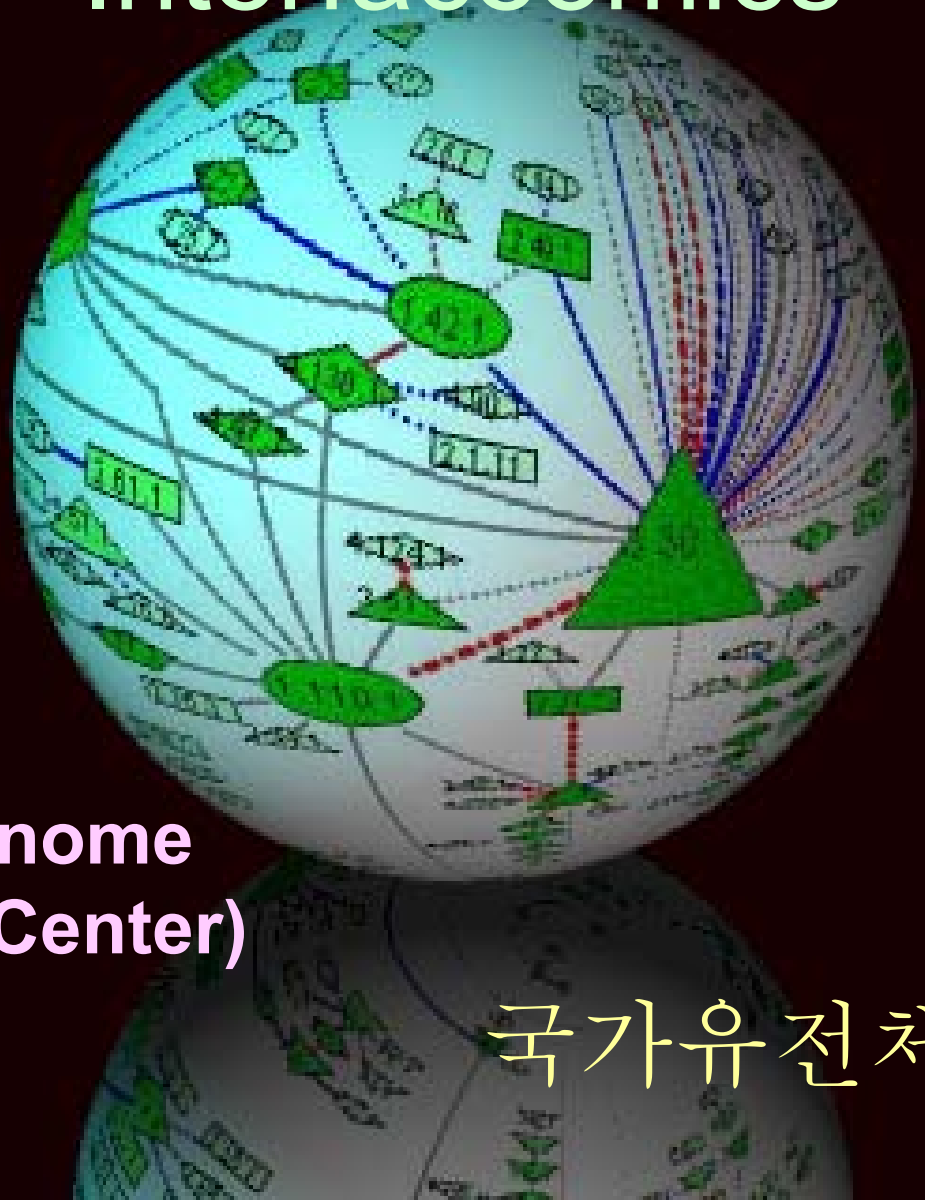
# Protein Interactomics & Interfaceomics




박종화

Jong Bhak

**NGIC**

**(National Genome Information Center)**

**KRIBB**

국가유전체정보센터

# Acknowledgement


- People who do science because science is interesting and fun.
- People who support science and technology by paying tax.

- MOST of Korea which funds NGIC.

- KRIBB for funding and support.



한국생명공학연구원
Korea Research Institute of Bioscience and Biotechnology

한국생명공학연구원 대전광역시 유성구 어은동 52번지 305-333
Korea Research Institute of Bioscience and Biotechnology(KRIBB)


Copy right by Korea Research Institute of Bioscience and Biotechnology (KRIBB)

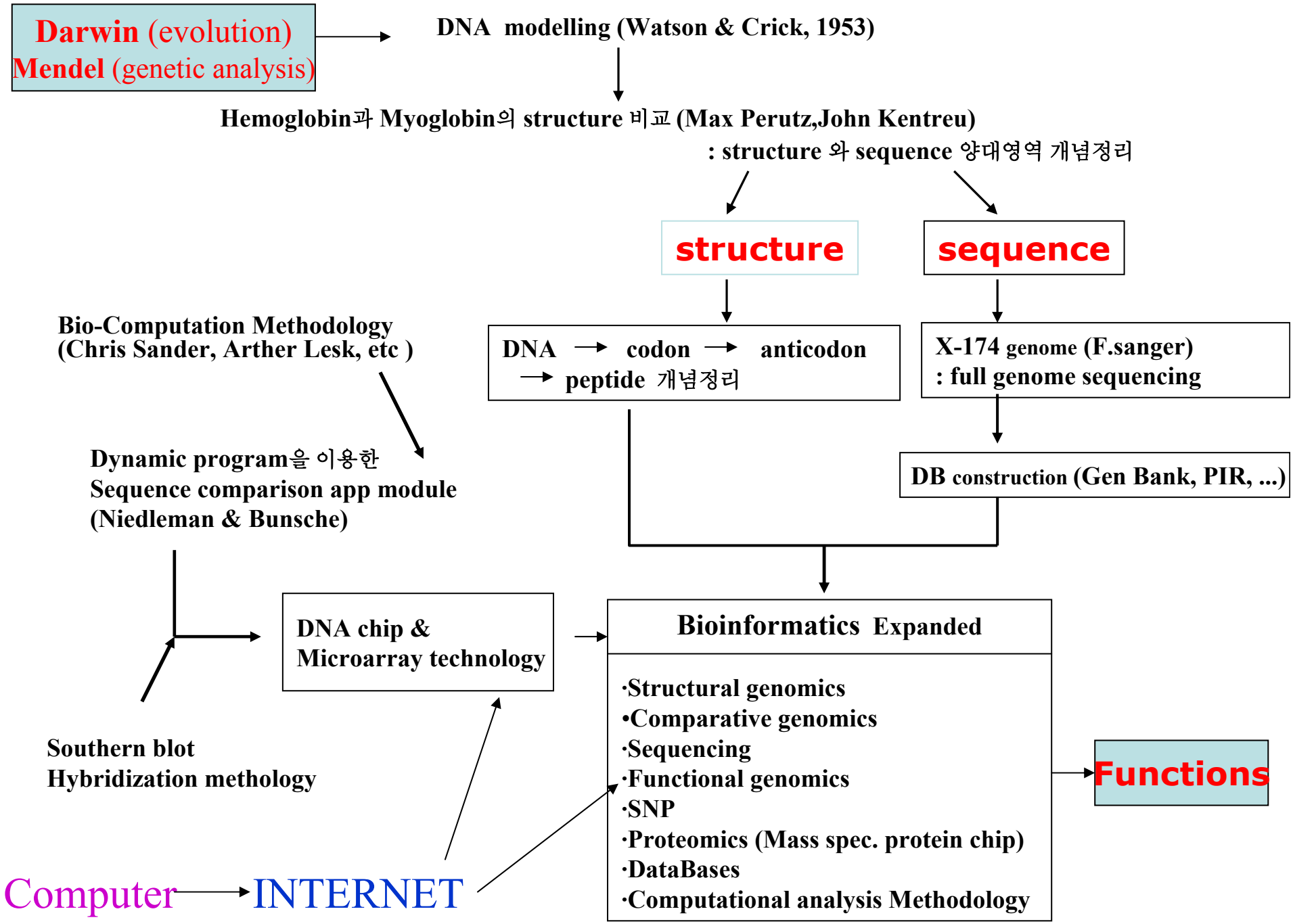

NGIC

# Content

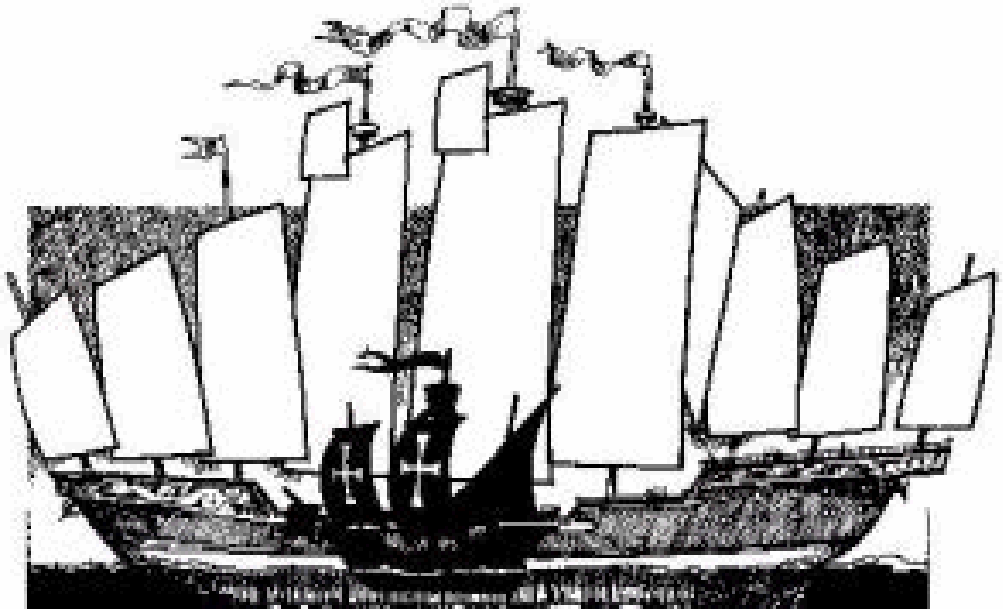
- Bioinformatics
- Protein Informatics
- Protein Interactomics & Interfaceomics

# Brief History of Bioinformatics

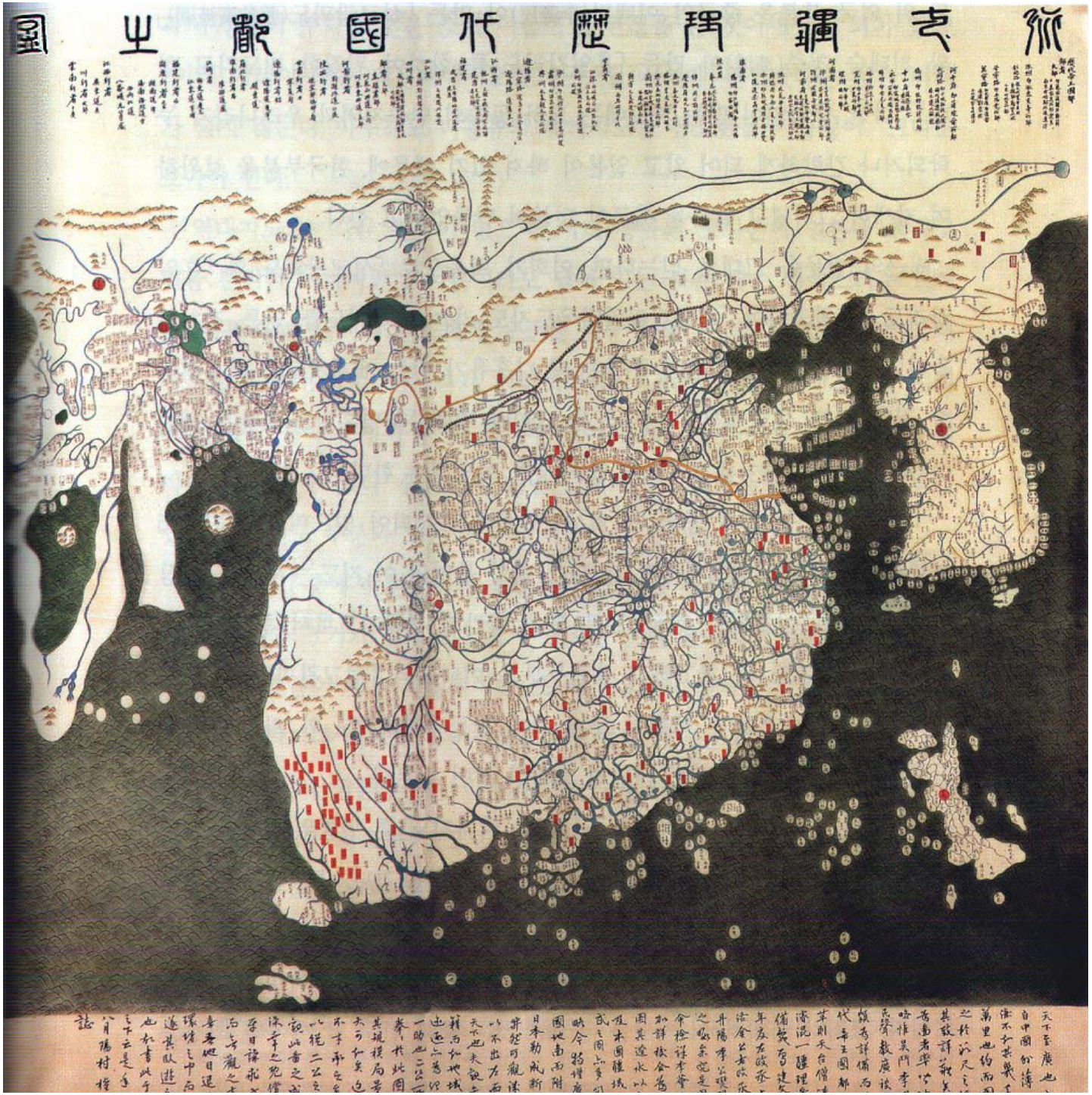
- **Darwin**: A theoretical biologist
- **Mendel**: A theoretical prediction and validation
- **Perutz** and **Kendrew**: structural biologists
- **Crick** and **Watson**: DNA modellers
- **Crick, Brenner,** : Codon
- **Sanger**: DNA sequencing: Genomics
- **Sanger**: Protein sequencing
- **Sanger**: Proteomics
- **Lesk**: Visualization of proteins
- **Needleman** & **Wunch**: Computer algorithms
- **Southern**: Hybridization➔ Functional Genomics
- Tim **Berners-Lee** 1990: HTTPD➔ Internet


Darwin (evolution)
Mendel (genetic analysis)
DNA modelling (Watson & Crick, 1953)
Hemoglobin과 Myoglobin의 structure 비교 (Max Perutz,John Kentreu)
: structure 와 sequence 양대영역 개념정리
structure
sequence
Bio-Computation Methodology
(Chris Sander, Arther Lesk, etc )
DNA → codon → anticodon
→ peptide 개념정리
X-174 genome (F.sanger)
: full genome sequencing
DB construction (Gen Bank, PIR, ...)
Dynamic program을 이용한
Sequence comparison app module
(Niedleman & Bunsche)
DNA chip &
Microarray technology
Bioinformatics Expanded
·Structural genomics
•Comparative genomics
·Sequencing
·Functional genomics
·SNP
·Proteomics (Mass spec. protein chip)
·DataBases
·Computational analysis Methodology
Functions
Southern blot
Hybridization methology
Computer
INTERNET

# Bioinformatics People ?



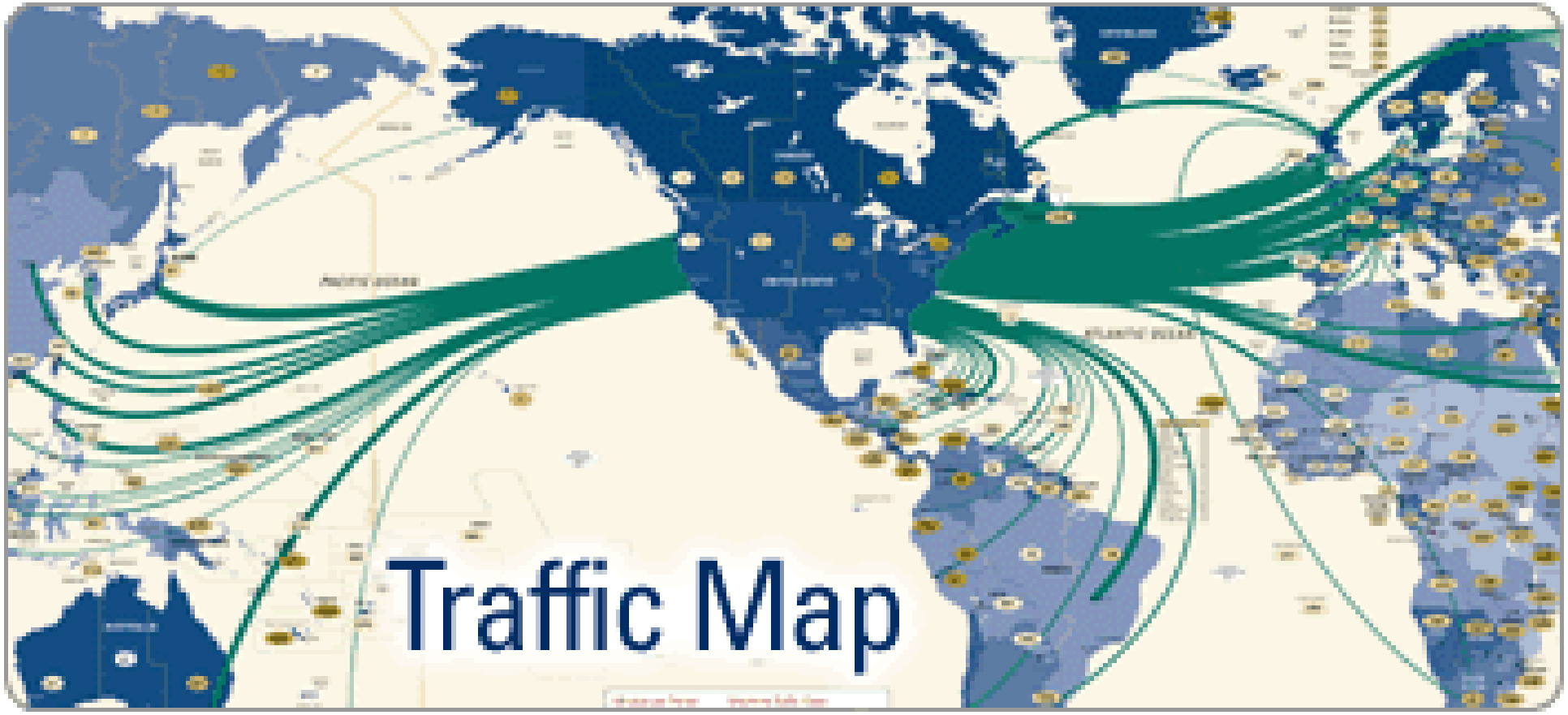
- **Explorers** of 1300-1400
- **Charting** the unknown territory of Biological Life.




Gangnido 1402

THE
NOSAMO
WORLD

# World Traffic Map



# Long Definition of Bioinformatics

- Bioinformatics is a **discipline of science** that analyses, seeks understanding and models the **whole life** as an **information processing phoenomenon** utilizing **energy** with methods from **philosophy, mathematics and computer science** using **biological experimental data.**

-- Jong Bhak

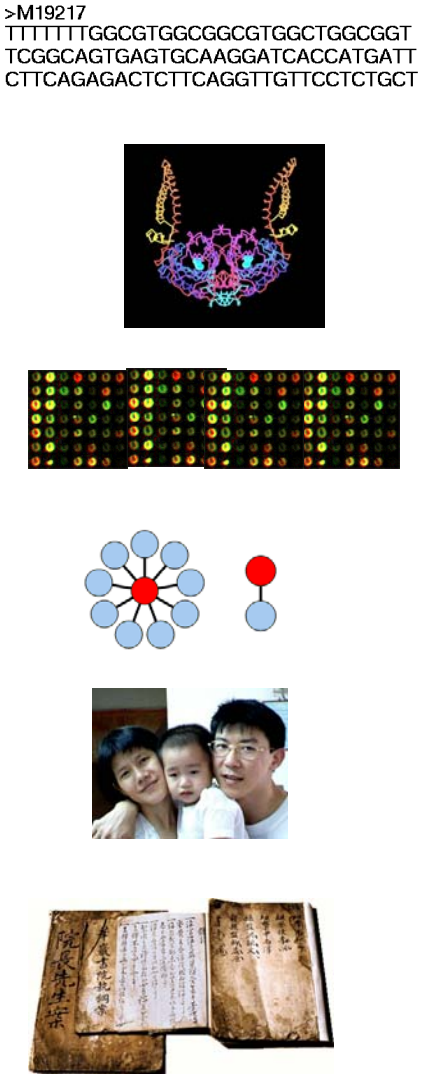
# A short definition of Bioinformatics

- **Biology is bioinformatics and bioinformatics is biology.**

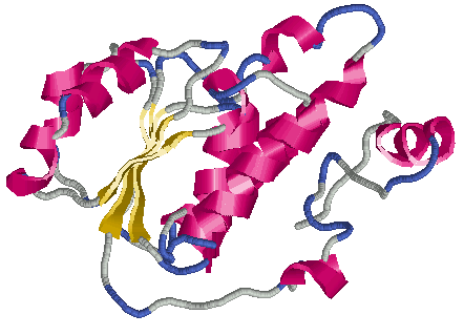
  -- Jong Bhak

# Research Domains of Bioinformatics

- Sequence
- Structure
- Expression
- **Interaction**
- Function
- Literature



# What is Protein?



- The most informative object in the biological universe.

- Protein level is efficient to work with.

  ➔ a Naturally distinct unit. So, favoured by bioinformatic computing.

# BioComplexity Density Chart

Total complexity

Relative complexity

Complexity degree

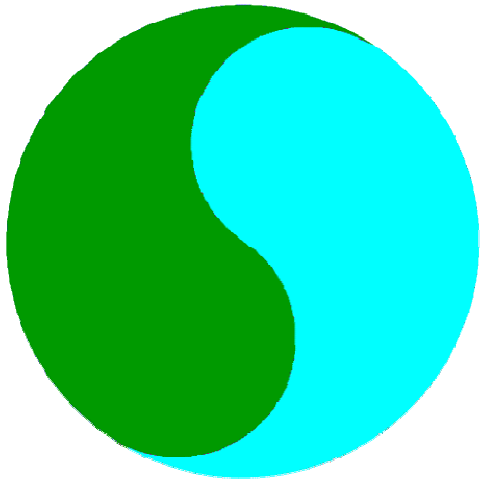
10-e10 10-e9 10-e6 10-e3 1 10

**Size Scale**

# Why Interactome and Interactomics?

- Context in Biology is always the problem of ‘interactions’

**Dan Bolser & Jong Bhak, *Genome informatics***



# Why Protein Intearctome ?

- Protein Intearctome can give us the most valuable insights and information about biological functions in cells.

(it is the best map you can have in biology)

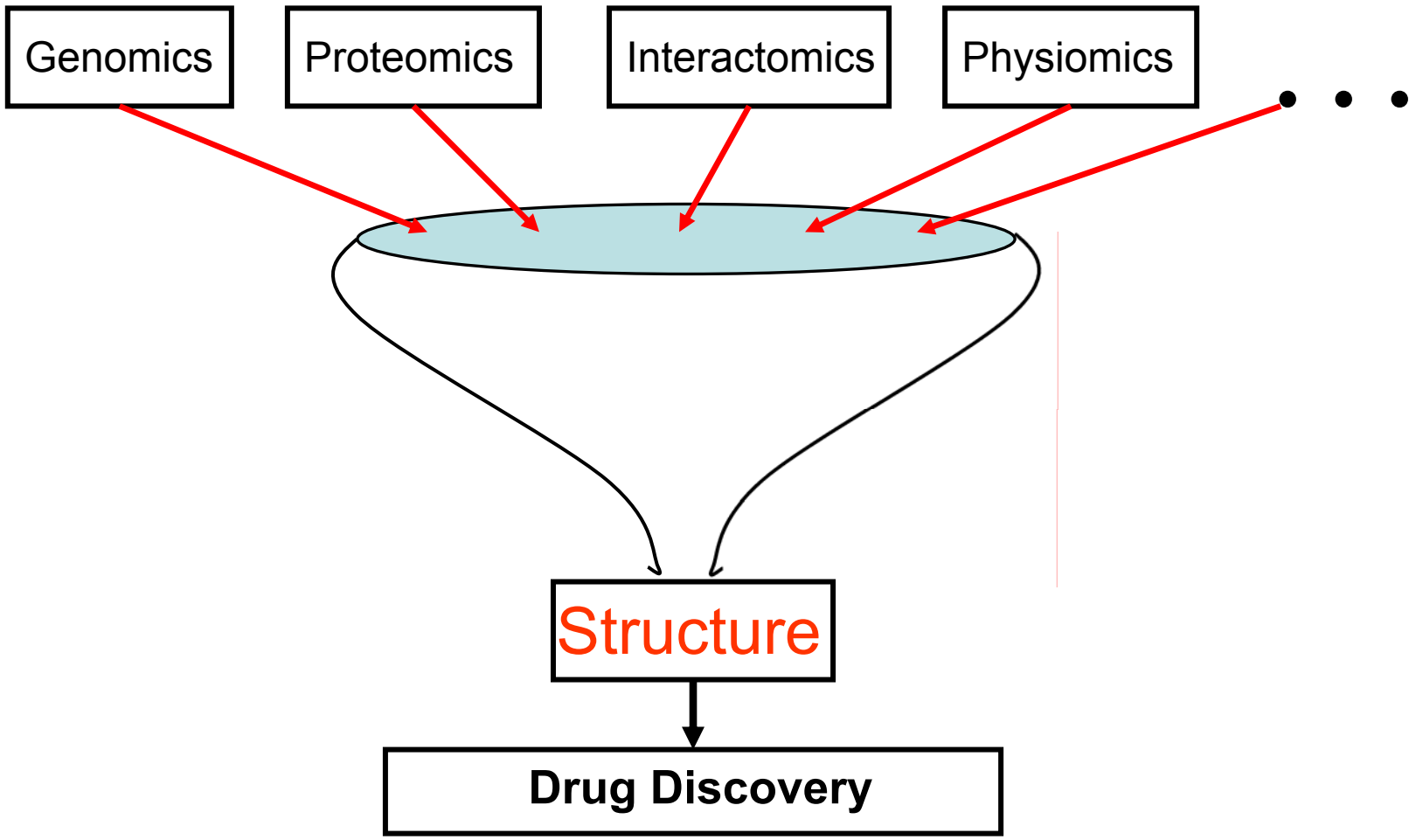
http://interactome.org

# Why Structural Interactomics ?

- Most definite.
- Most informative
- Most accurate
- Directly connected to drug discovery.
- Structural → certain well-known forces and physical rules are considered.

- Most fun and beautiful to work with ?

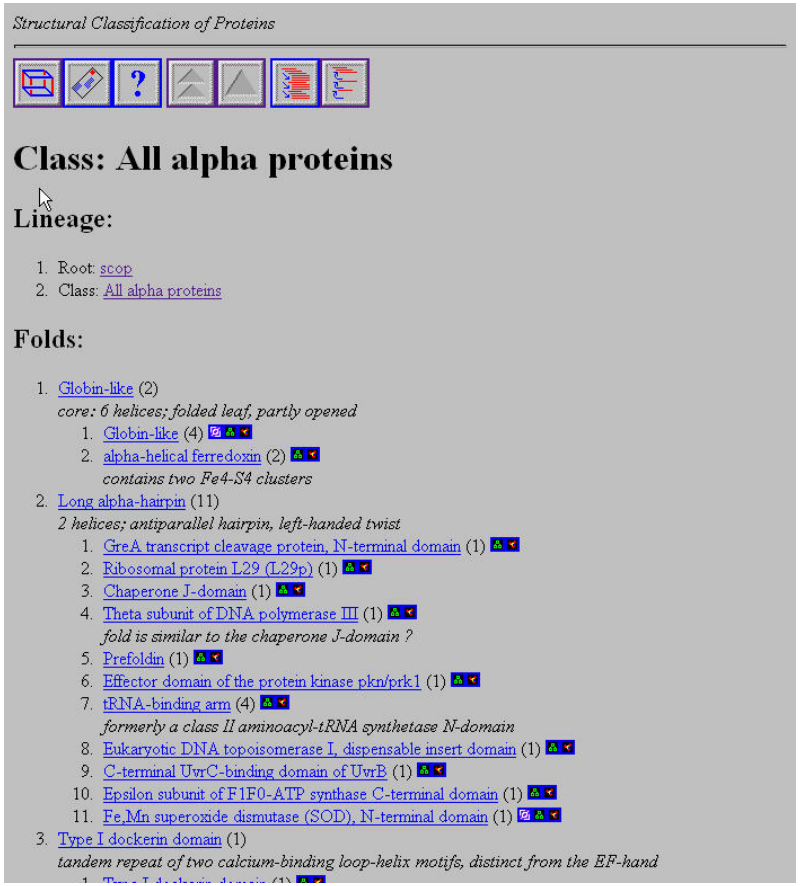
# Funnel of Biological Drug Discovery



# The unit of Interaction

## Protein Structural

# Domains



*Structural Classification of Proteins*

**Class: All alpha proteins**

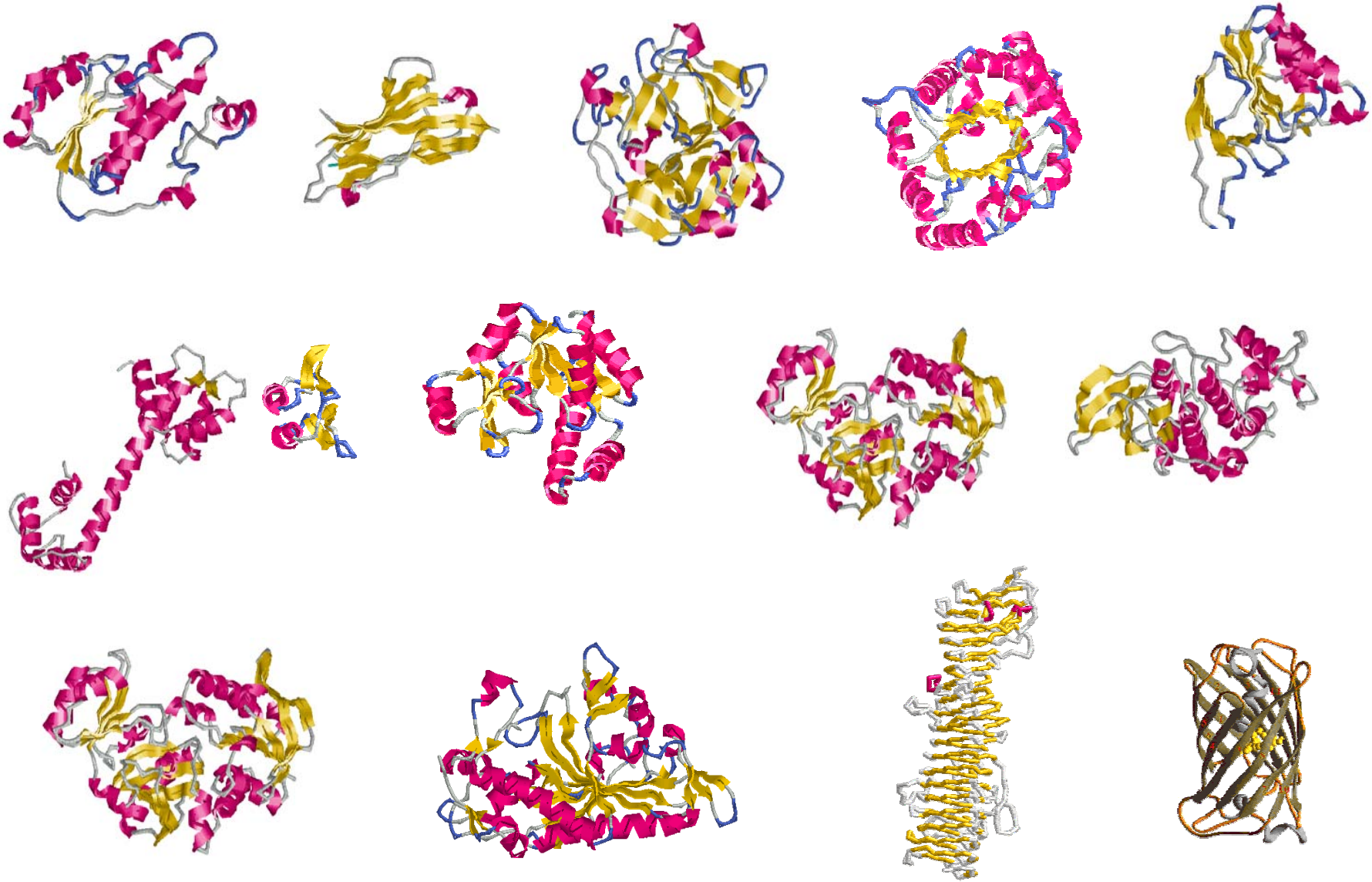
**Lineage:**

1. Root: scop
2. Class: All alpha proteins

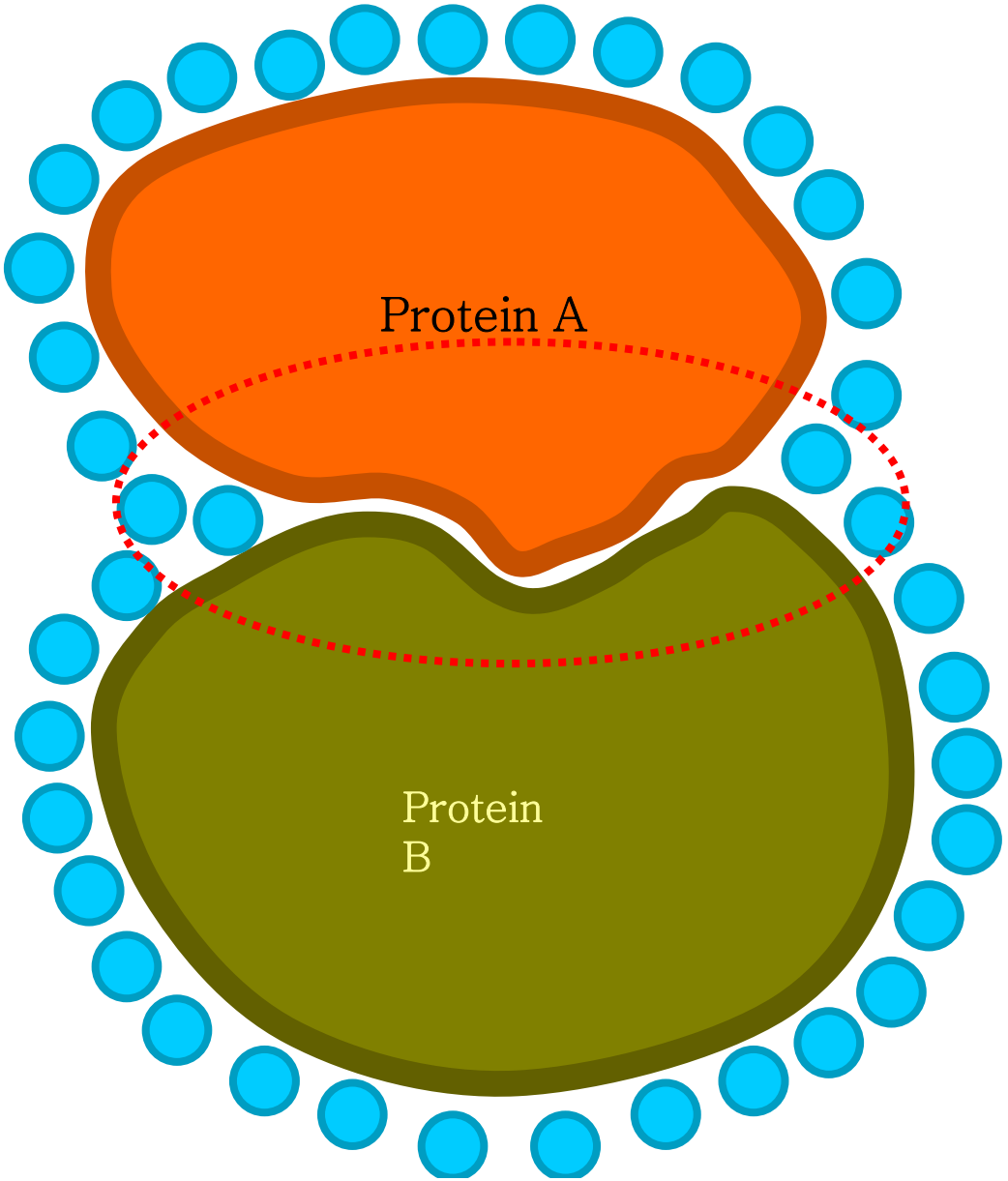
**Folds:**

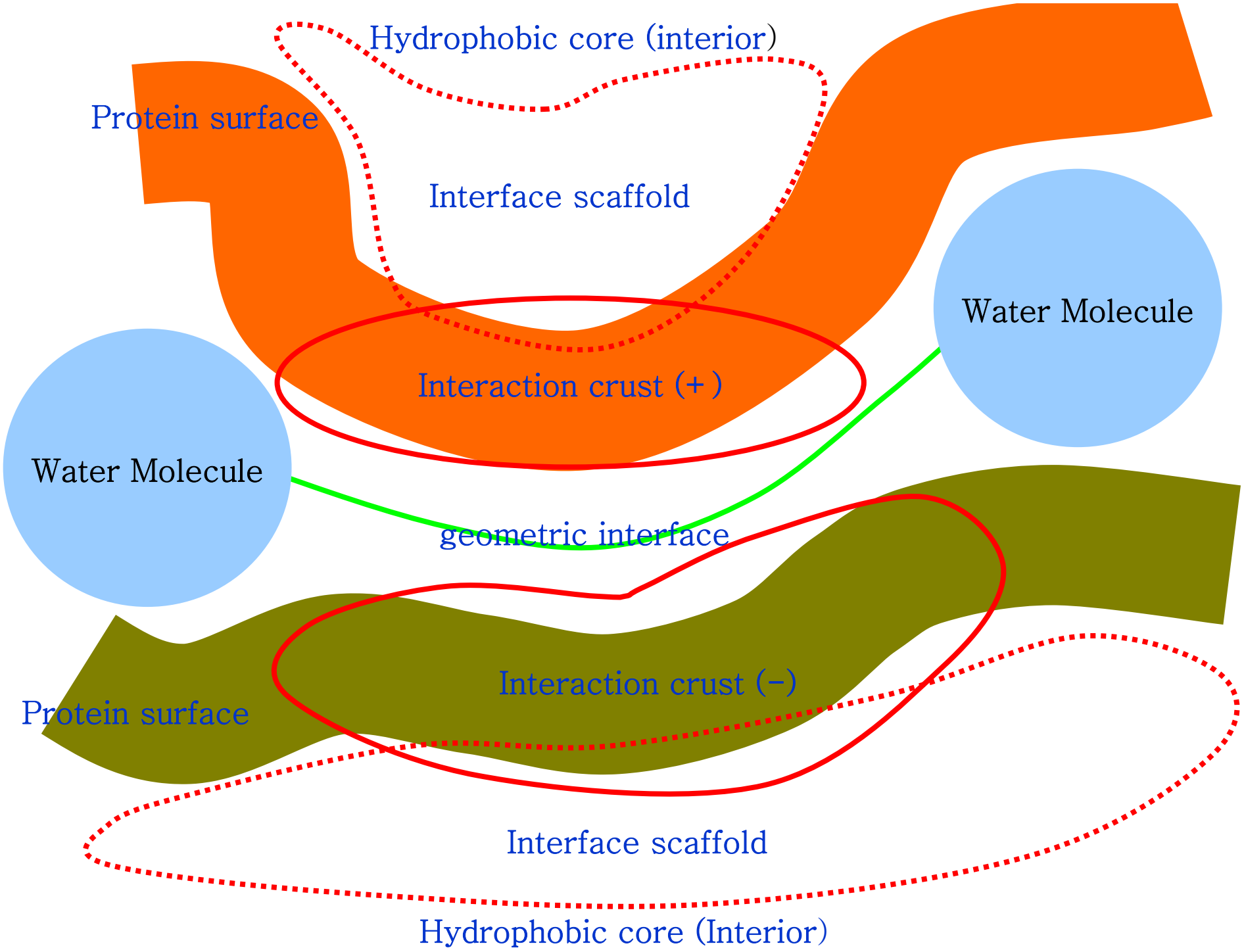
1. Globin-like (2)
   *core: 6 helices; folded leaf, partly opened*
   1. Globin-like (4)
   2. alpha-helical ferredoxin (2)
      *contains two Fe4-S4 clusters*
2. Long alpha-hairpin (11)
   *2 helices; antiparallel hairpin, left-handed twist*
   1. GreA transcript cleavage protein, N-terminal domain (1)
   2. Ribosomal protein L29 (L29p) (1)
   3. Chaperone J-domain (1)
   4. Theta subunit of DNA polymerase III (1)
      *fold is similar to the chaperone J-domain ?*
   5. Prefoldin (1)
   6. Effector domain of the protein kinase pkn/prk1 (1)
   7. tRNA-binding arm (4)
      *formerly a class II aminoacyl-tRNA synthetase N-domain*
   8. Eukaryotic DNA topoisomerase I, dispensable insert domain (1)
   9. C-terminal UvrC-binding domain of UvrB (1)
   10. Epsilon subunit of F1F0-ATP synthase C-terminal domain (1)
   11. Fe,Mn superoxide dismutase (SOD), N-terminal domain (1)
3. Type I dockerin domain (1)
   *tandem repeat of two calcium-binding loop-helix motifs, distinct from the EF-hand*

# What are Structural Domains ?



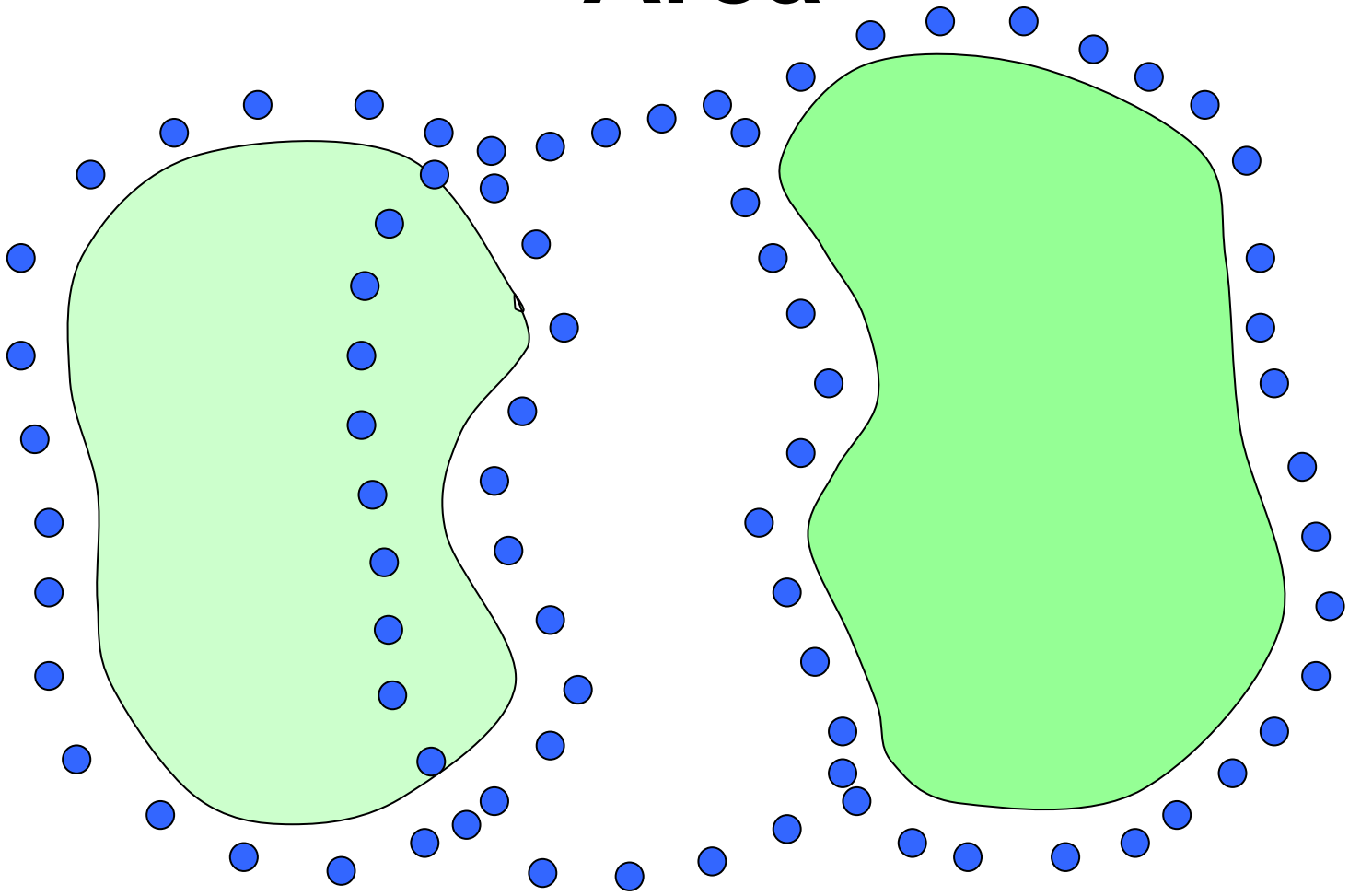
# Interaction among domains



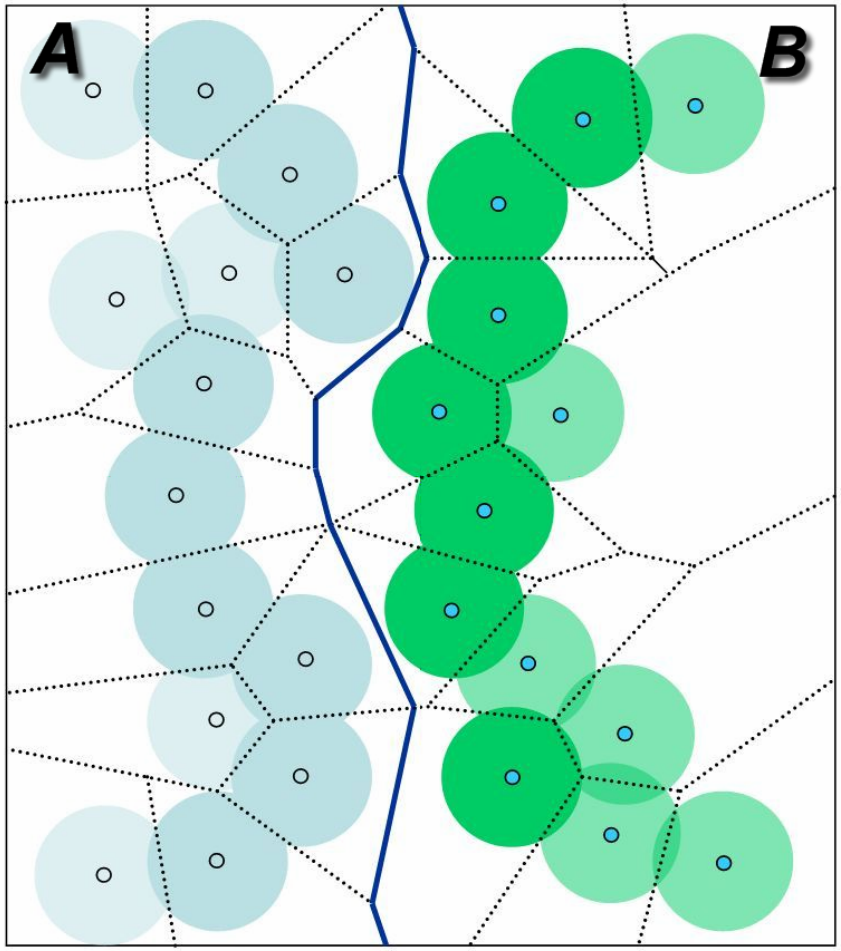

Hydrophobic core (interior)
Protein surface
Interface scaffold
Water Molecule
Interaction crust (+)
Water Molecule
geometric interface
Interaction crust (−)
Protein surface
Interface scaffold
Hydrophobic core (Interior)

# Detecting (defining) Domain Interaction

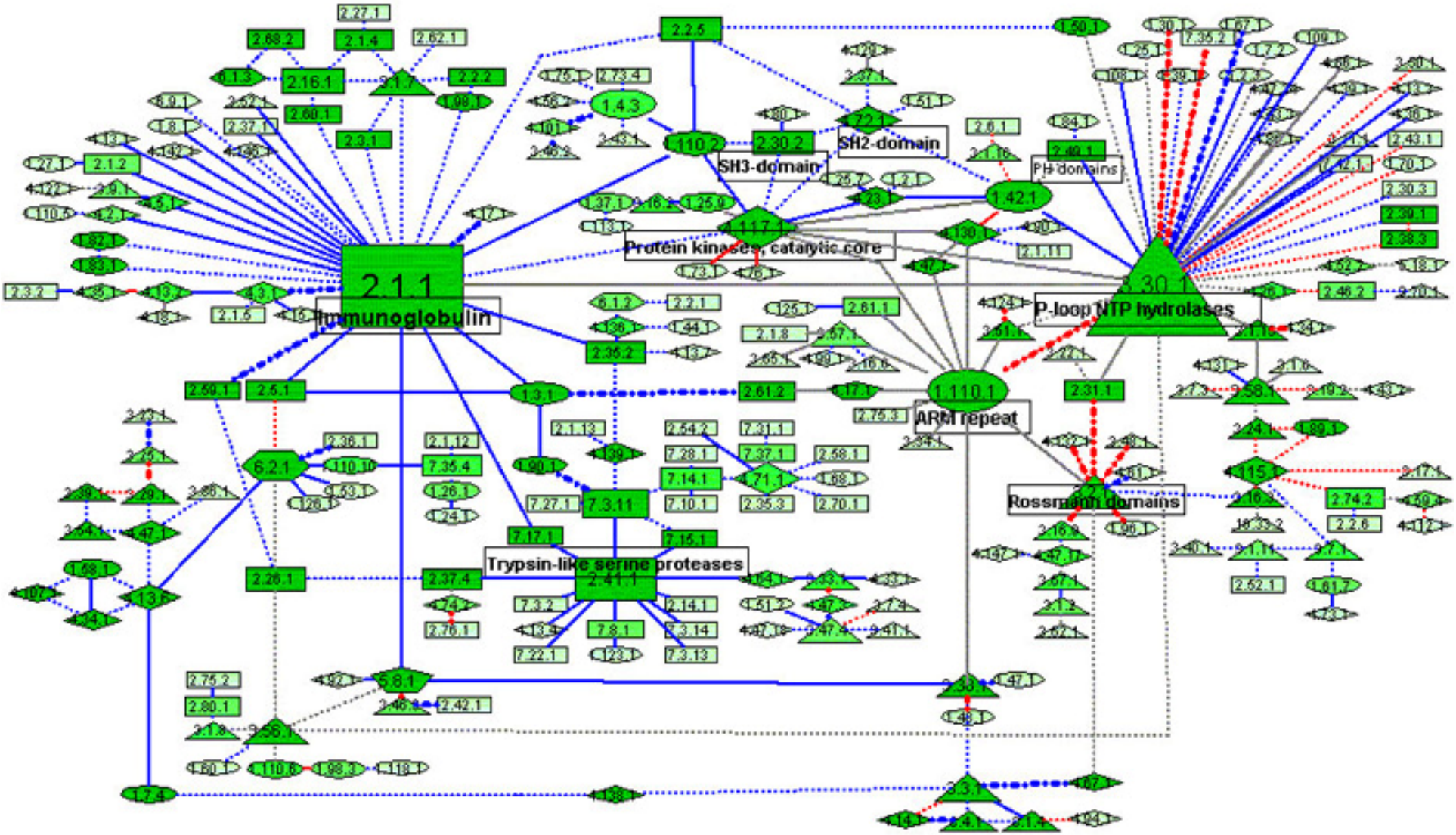
# ASA: Accessible Surface Area



# Voronoi Diagram



# A protein domain interaction map



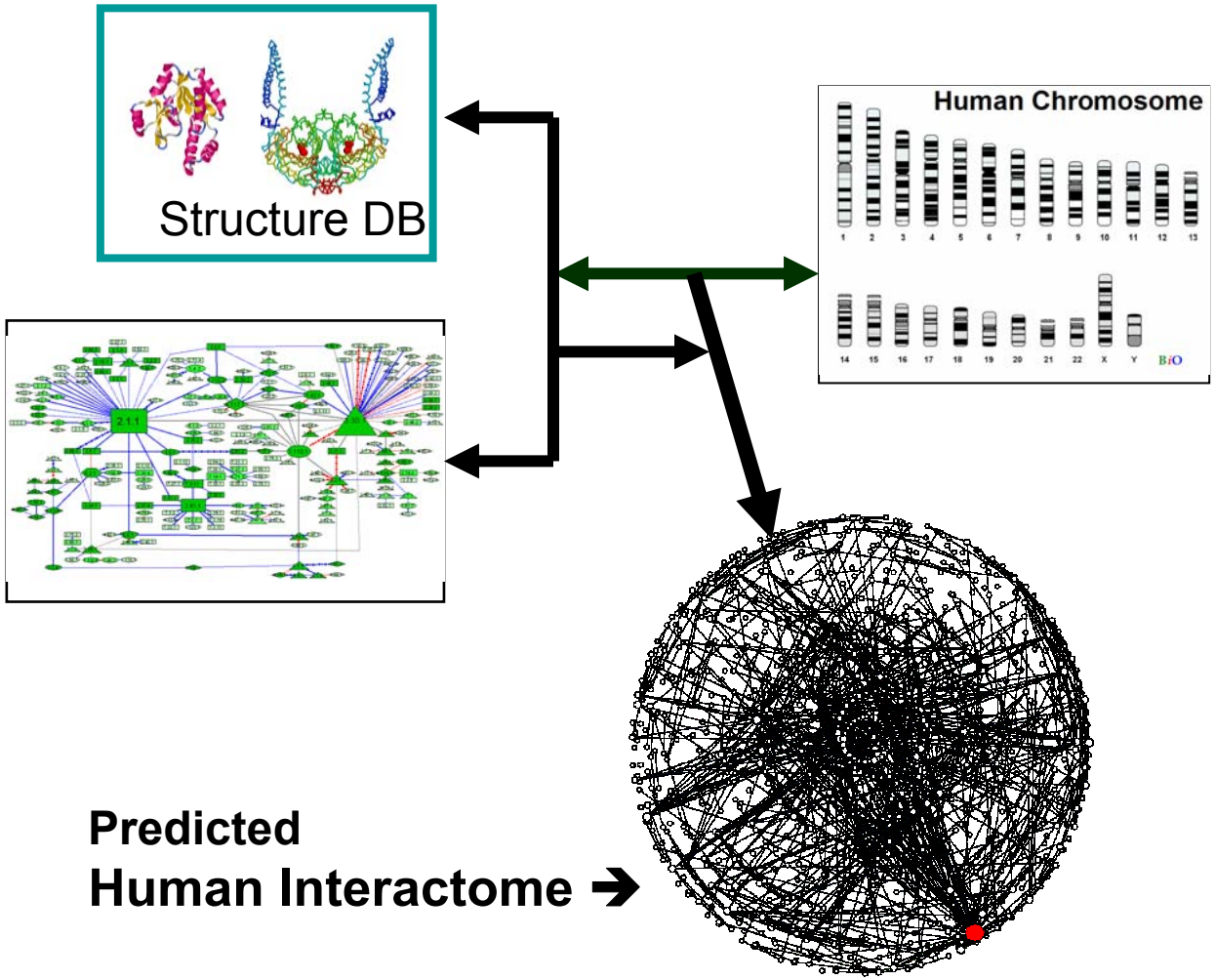
# What can we do with Interactomes?

- The main academic goal of Interactomics
  - Mapping all the molecular interactions in cells

# Computational Human Interactome

- Mapping all the protein-protein and protein-ligand interactions using computational and experimental methods.

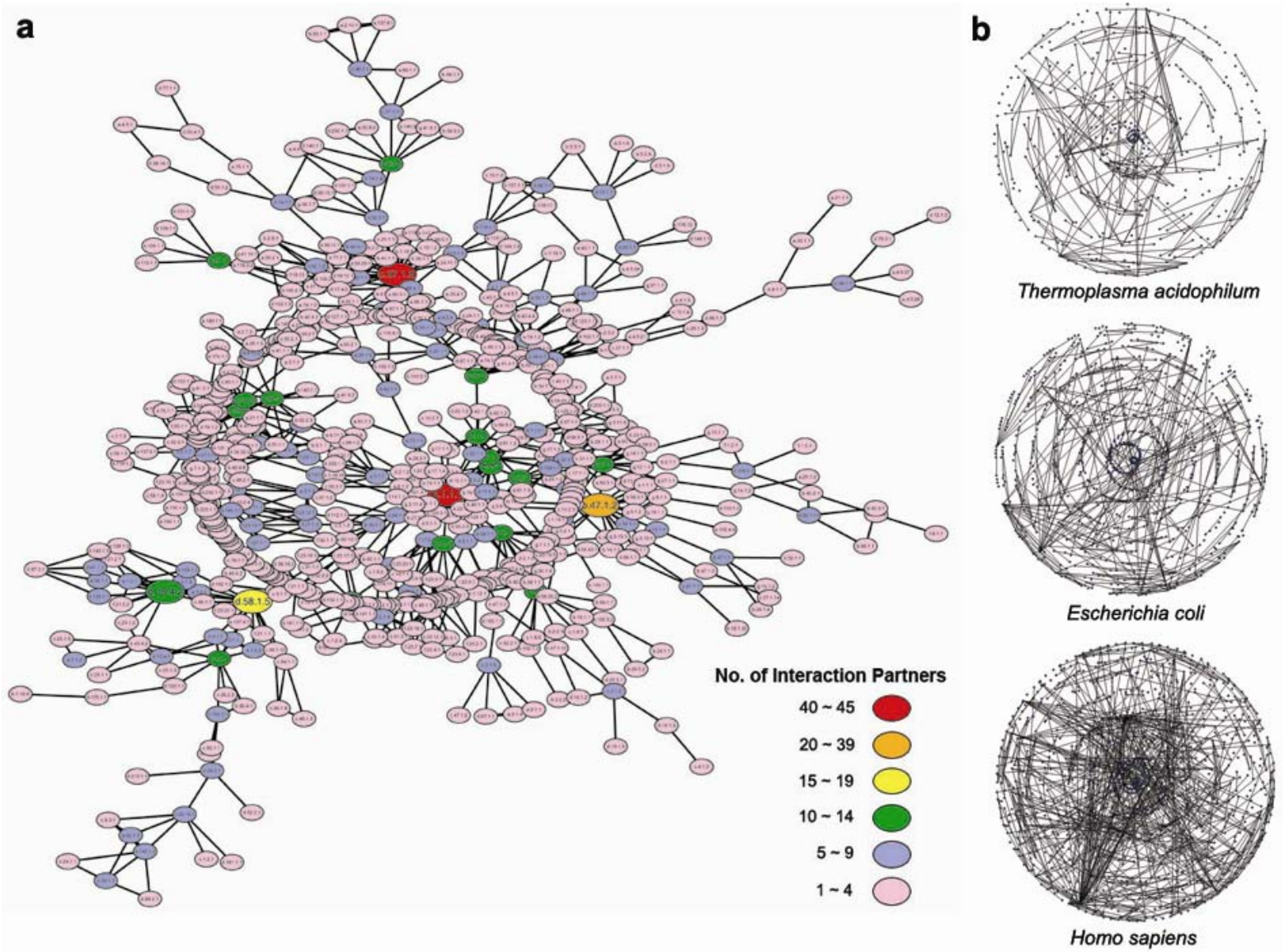
# Practical Steps of Complete Human Interactome



# Comparative Interactomics

- A large scale analysis of comparing interactomes of species to understand the evolution and function of cells.

# Global view of protein family interaction networks for **146** genomes



# Comparative Interactomics

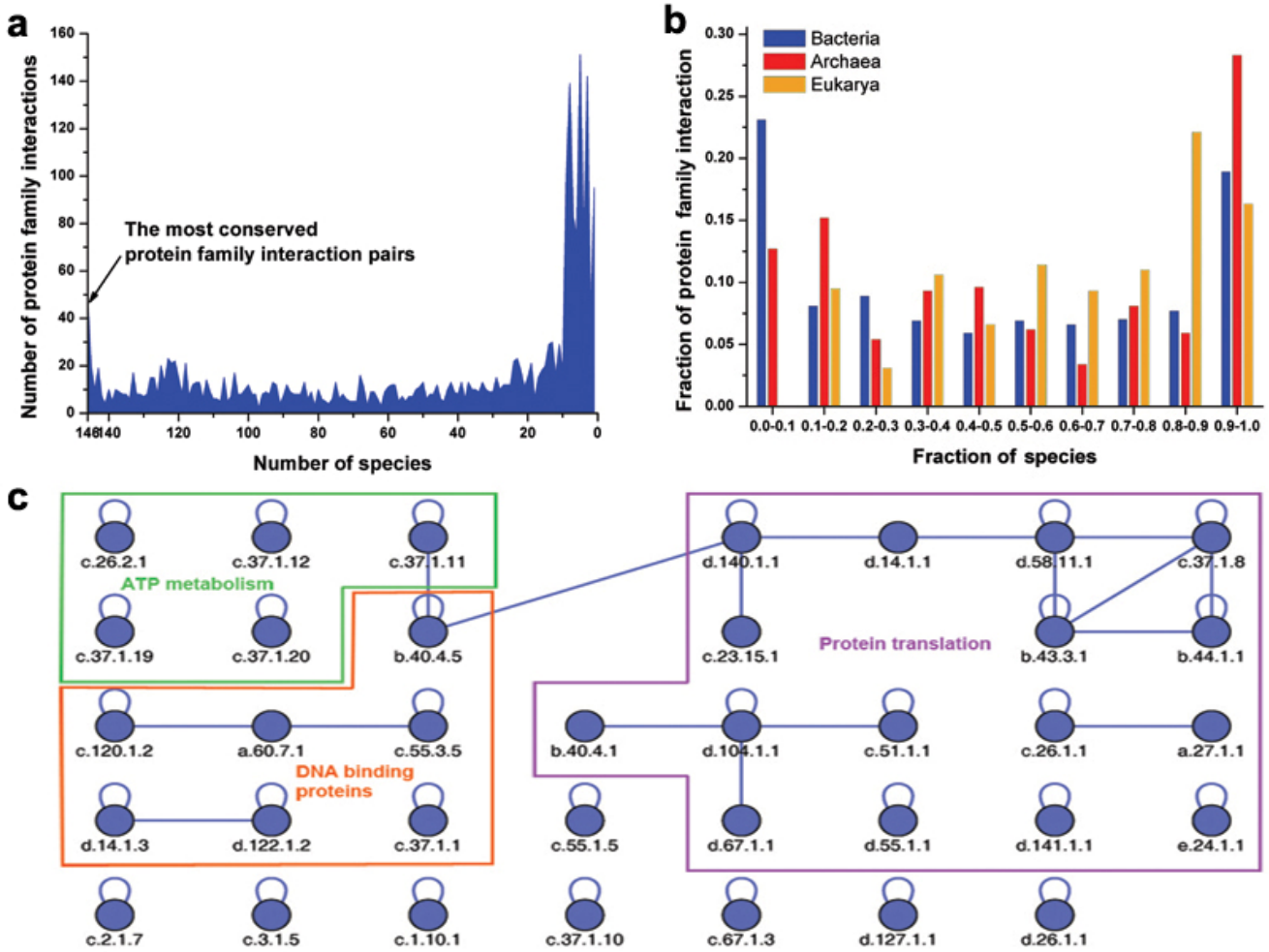
- We found that all 146 interactomes are scale-free networks, and they share a core protein-interactome comprising 36 protein families related to indispensable functions in a cell.

- *Daeui Park, Semin Lee, Dan Bolser, Michael Schroeder, Michael Lappe, Donghoon Oh, & Jong Bhak. Bioinformatics, 2005.*

# Eukarya and Prokarya

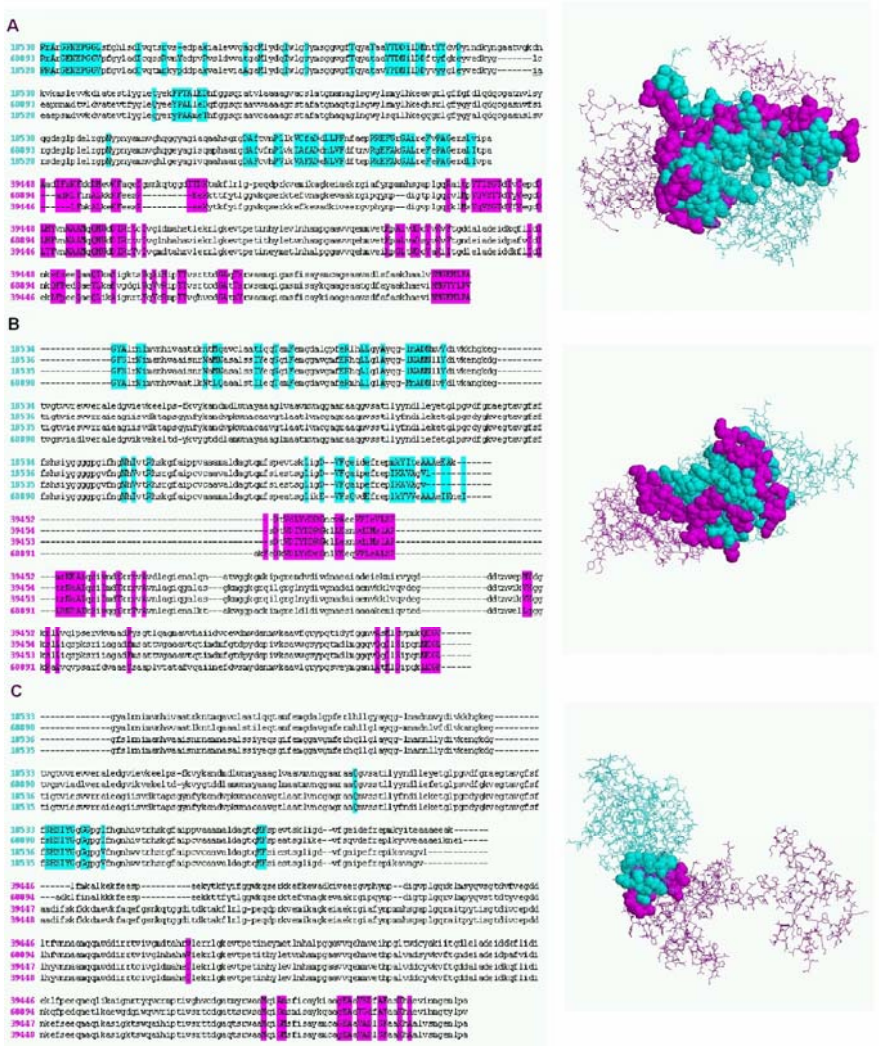
- We found two fundamental differences among prokaryotic and eukaryotic interactomes:
  - 1) eukarya had significantly more hub families than archaea and bacteria, and
  - 2) certain special hub families determined the shapes of the eukaryotic interactomes.

# Core Interactome



# Interfaceomics

- http://interpare.net/
- Analyzing the actual interaction interfaces of protein-protein interactions.
- Interfaceome: the whole set of interacting molecular pairs in cells.



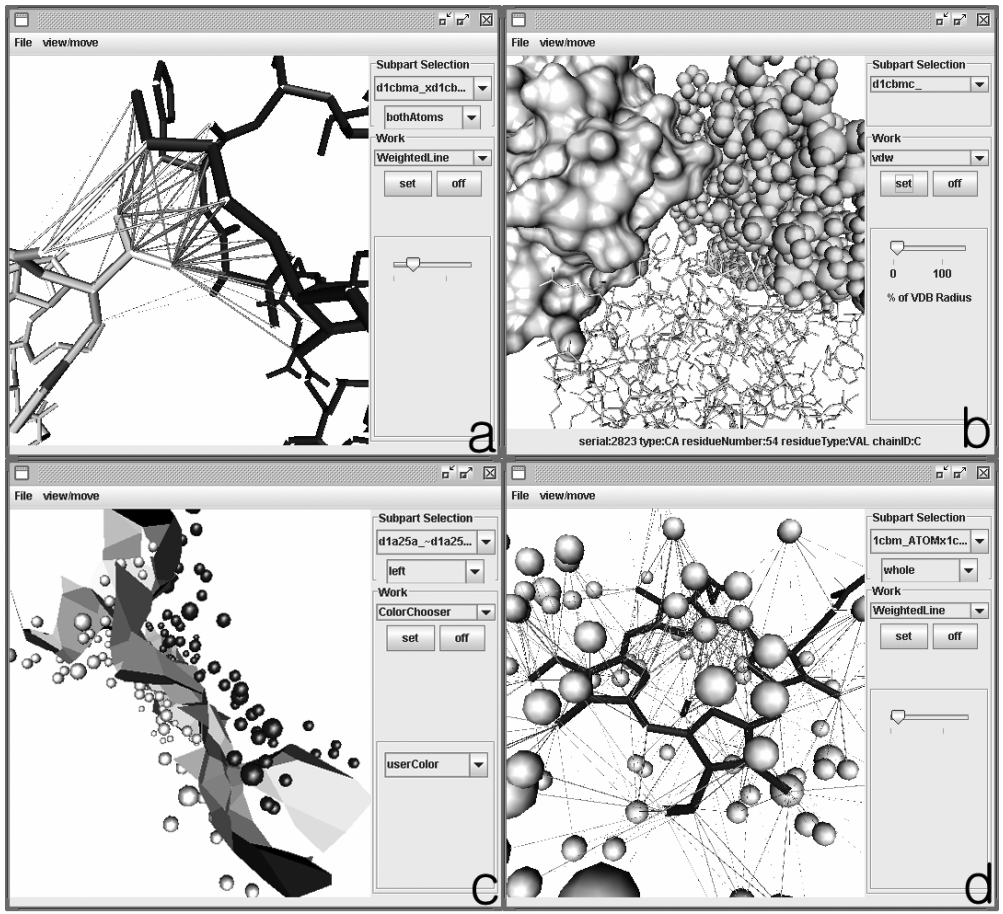
Wankyu Kim, Dan M. Bolser and Jong Park, *Bioinformatics*, 2004, 20(7):1138-1150.

# InterFacer

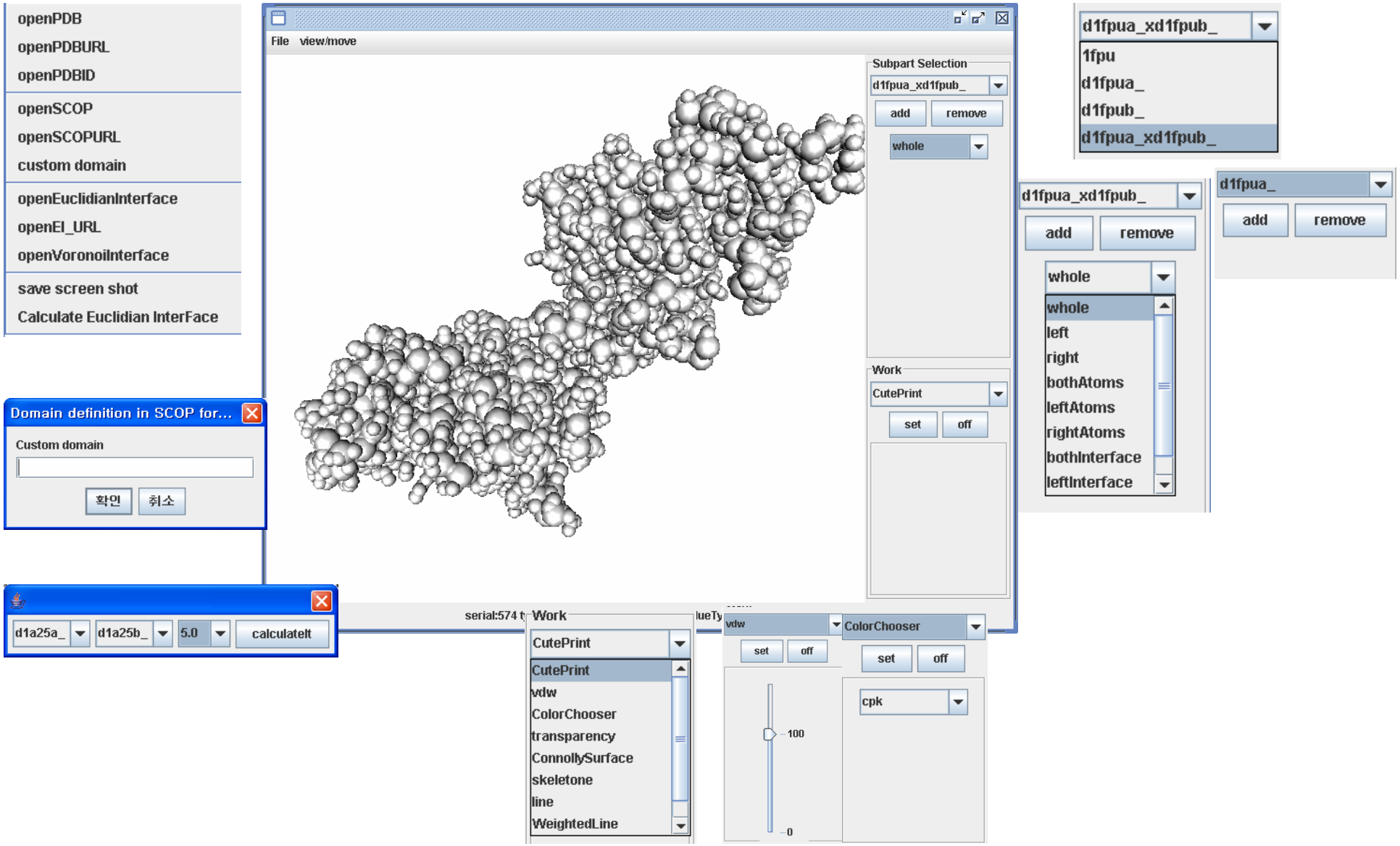
Choi HanSol

# InterFacer

- http://www.interfacer.org/



# Graphic User Interface of InterFacer.



# Goal of Interactomics & Interfaceomics

- Fastest Drug Discovery Possible

# References


**Large scale co-evolution analysis of Protein Structural Interlogues using the global Protein Structural Interactome Map(PSIMAP).**

Wankyu Kim, Dan M. Bolser and **Jong Park,** *Bioinformatics*, 2004, 20(7):1138-1150.

**Visualisation and Graph-theoretic Analysis of a Large-scale Protein Structural Interactome**

Dan M Bolser , Panos Dafas, Richard Harrington , Jong H Park and Michael Schroeder

*BMC Bioinformatics* **2003 4:**45

**Conservation of protein interaction network in evolution.**

Park J, Bolser D.
*Genome Informatics*. **2001** ;12:135-40.

**Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast.**
Park J, Lappe M, Teichmann SA.

*J Mol Biol*. **2001** Mar 30;307(3):929-38.

# Brief Introduction on NGIC

- 國家有傳體情保 (National Genome Information Center http://ngic.re.kr/ )

NGIC was established with the vision to become the hub of Korean bioinformatics effort:

1) collects and distributes genomic and proteomic data,

2) provides bioinformaatic databases and analysis platforms

3) promotes domestic and international research collaborations in bioinformatics.

Since 2001