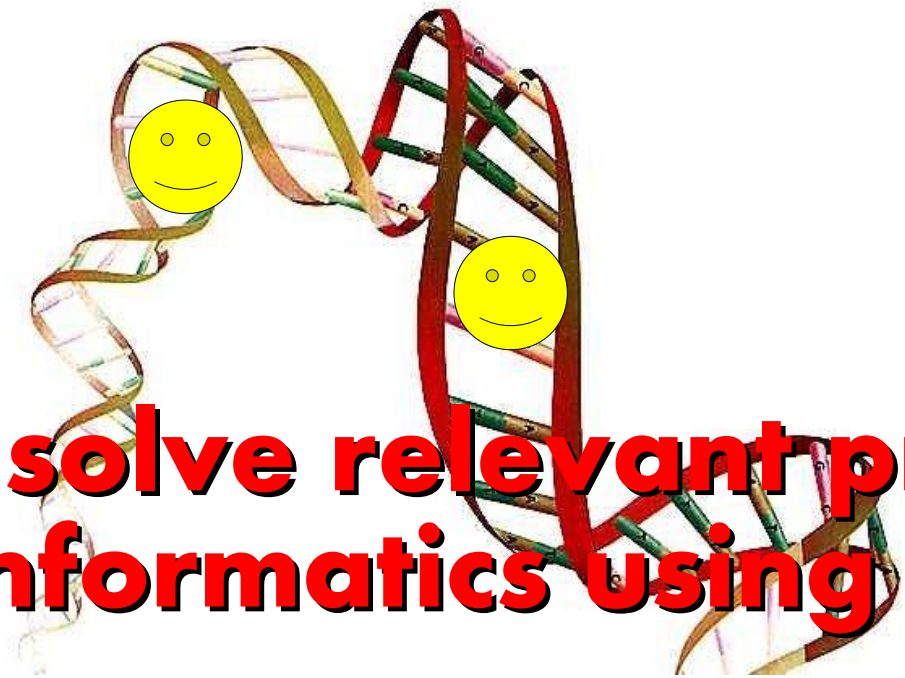


How to solve relevant problems in bioinformatics using agents

Eloisa Vargiu

Intelligent Agents and Soft-Computing Group

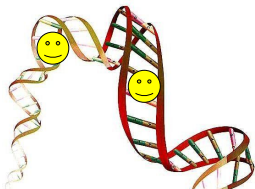
Dip.to di Ingegneria Elettrica ed Elettronica

Università degli Studi di Cagliari

vargiu@diee.unica.it

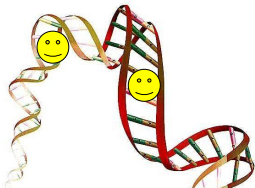
Outline

- Introduction
- An Introduction to Bioinformatics
- Open Problems in Bioinformatics
- Proposed Agent Solutions
- Conclusions
- References

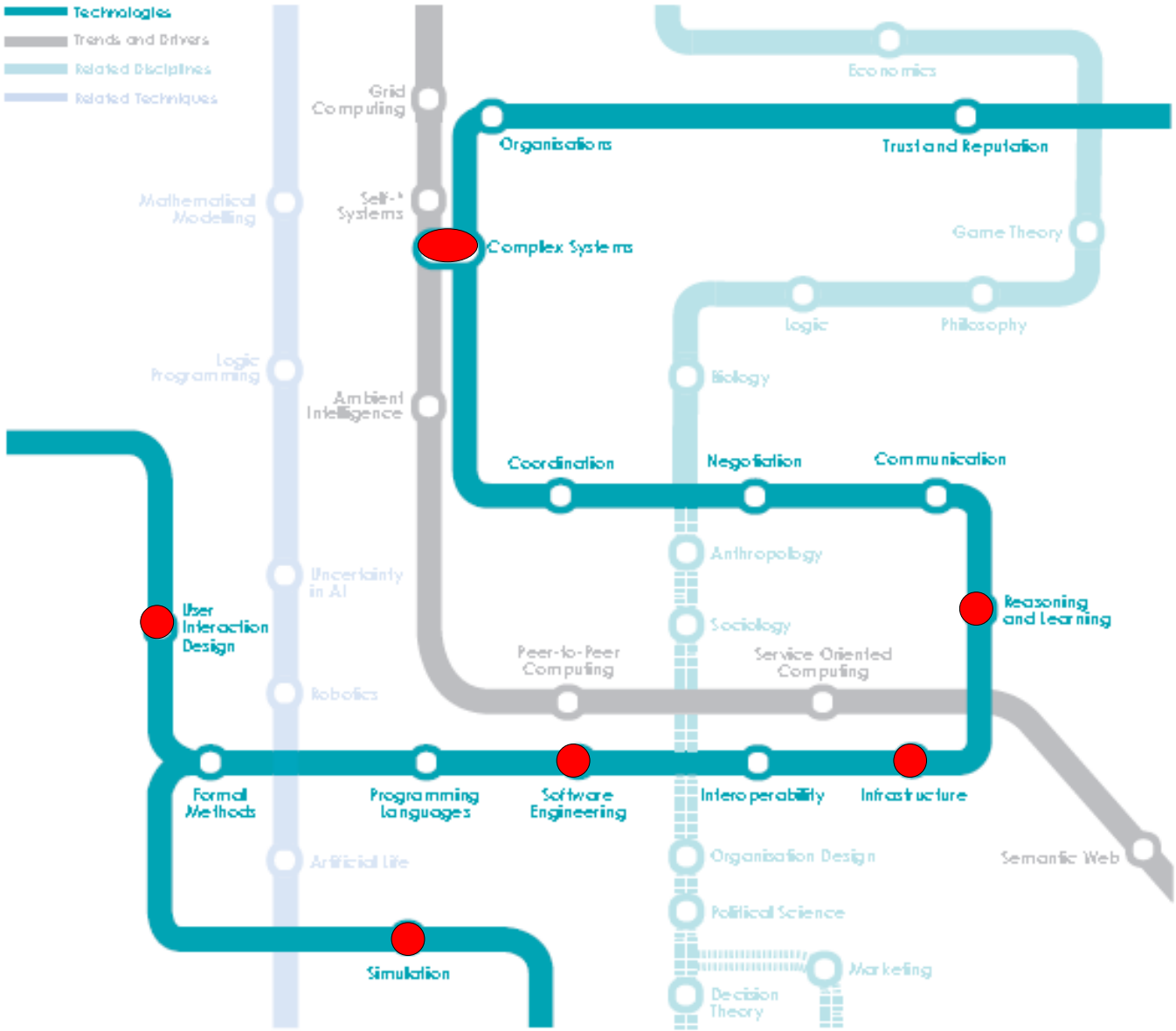


Introduction

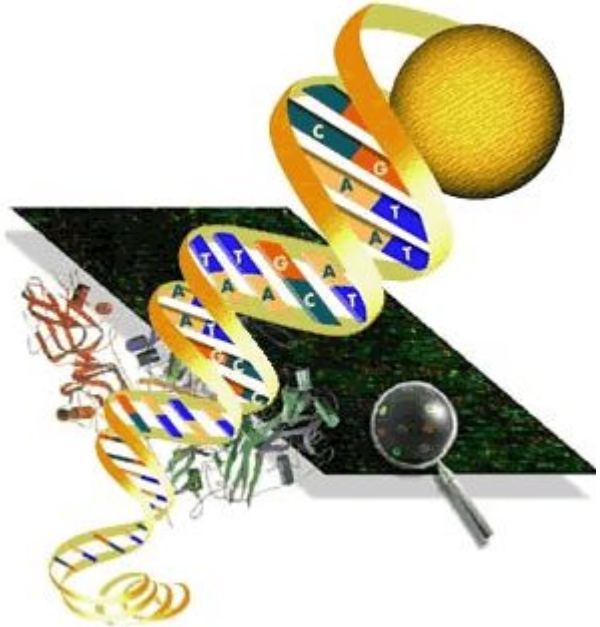
- Intended Audience
  - computer scientists
  - computer engineers
- Goal
  - To give an overview on bioinformatics issues
  - To provide agent solutions to bioinformatics

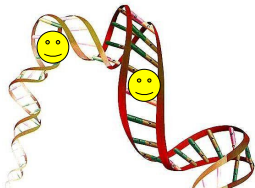


Introduction



An Introduction to Bioinformatics

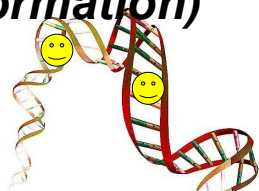




Aims and Scope of Bioinformatics

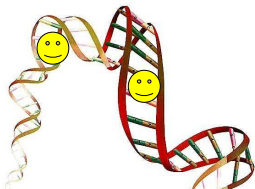
- Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline
- The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned

Definition by NCBI (National Center for Biotechnology Information)



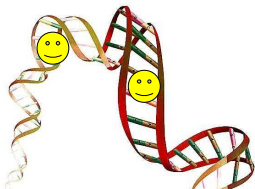
Aims and Scope of Bioinformatics

- Bioinformatics is aimed at solving biological problems using techniques from
  - applied mathematics
  - informatics
  - statistics
  - computer science



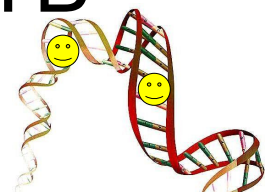
Aims and Scope of Bioinformatics

- Bioinformatics is concerned with researching, developing and applying tools and methods to acquire, analyze, organize and store biological and medical data



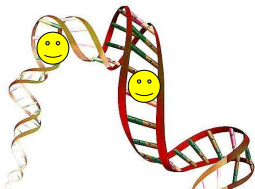
Bioinformatics Repositories

- Electronic stores of biological data
- Built using different systems
  - RDBMS
  - OODBMS
  - Flat files
- Distributed as
  - Flat files
  - Databases dumps
  - Web access only
- Size can range from a few hundreds MB to TB



Bioinformatics Repositories

- They are needed to
  - retrieve the information about published data
  - compare data against published data
  - discover the structure of a protein from a sequence



The Bioinformatics Repository Top Ten

- GenBank/DDJB/EMBL (Nucleotide sequences)
  http://www.ncbi.nlm.nih.gov
- Ensembl (Human/mouse genome)
  http://www.ensembl.org
- PubMed (Literature references)
  http://www.ncbi.nlm.nih.gov
- NR - entrez protein (Protein sequences)
  http://www.ncbi.nlm.nih.gov
- Swiss-Prot (Protein sequences)
  http://www.expasy.org

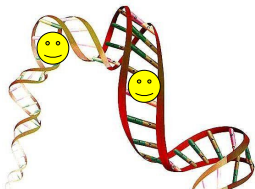
The Bioinformatics Repository Top Ten

- InterPro (Protein domains)
  http://www.ebi.ac.uk
- OMIM (Genetic diseases)
  http://www.ncbi.nlm.nih.gov
- Enzymes (Enzymes)
  http://www.expasy.org
- PDB (Protein structures)
  http://www.rcsb.org/pdb/
- KEGG (Metabolic pathways)
  http://www.genome.ad.jp

Bioinformatics Repositories

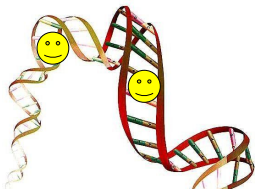
- The major sources of nucleotide sequence are the ones belonging to the International Nucleotide Sequence Database Collaboration:

- DDBJ (DNA DataBank of Japan)
  http://www.ddbj.nig.ac.jp/
- EMBL (European Molecular Biology Laboratory)
  http://www.ebi.ac.uk/emb
- GenBank (NIH genetic sequence database)
  http://www.ncbi.nlm.nih.gov/Genbank

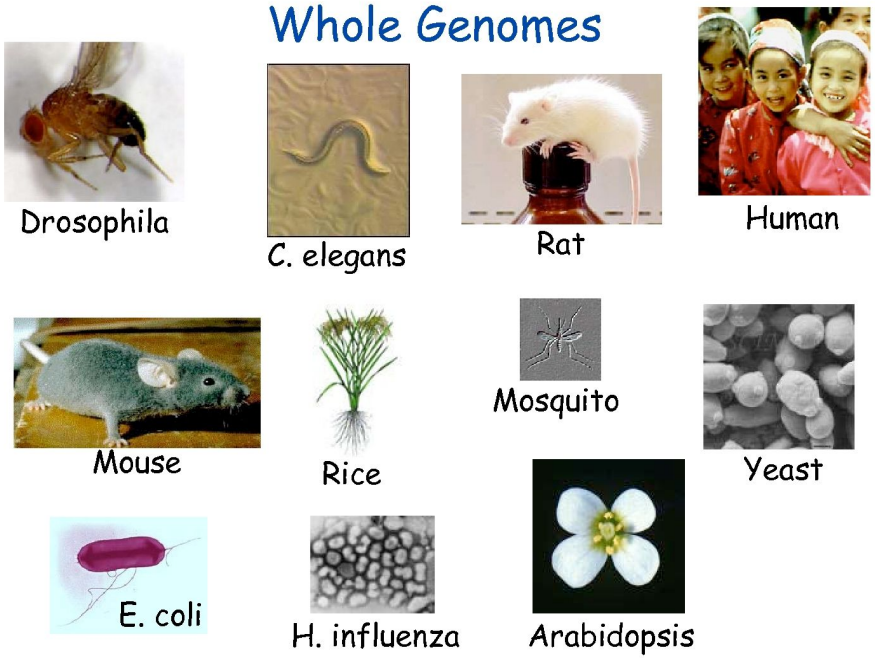


Main Disciplines

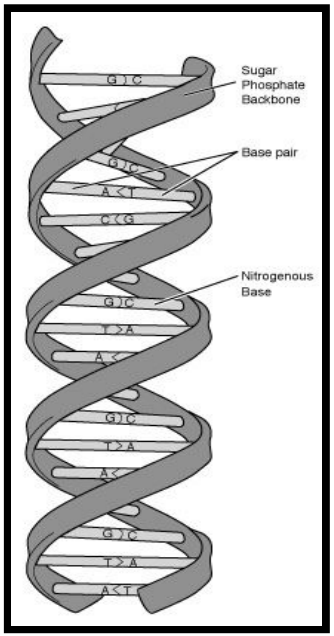
- Bioinformatics is a support to life science disciplines:
  - Genomics
  - Proteomics
  - Transcriptomics

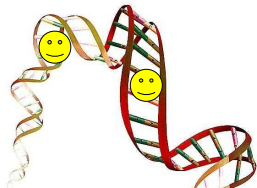


Genomics



- The study of an organism genome and the use of the genes

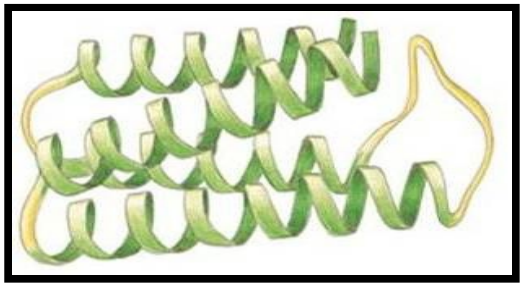


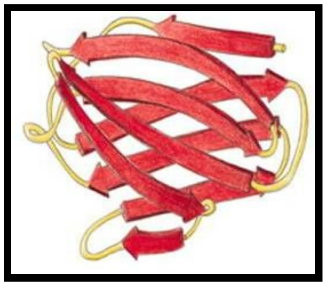


Proteomics

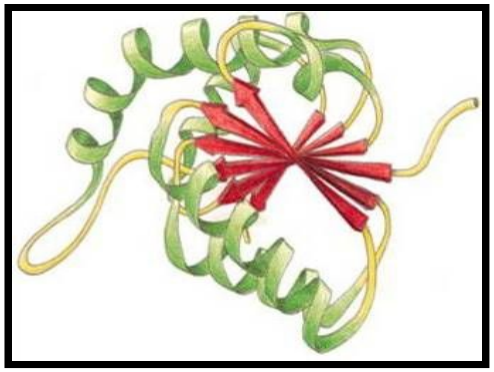
"IPLMTRWDQEQESDFGHKLPIYTREWCT"

primary structure

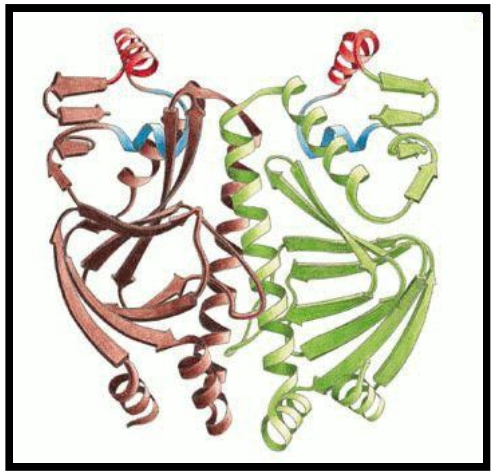




secondary structure

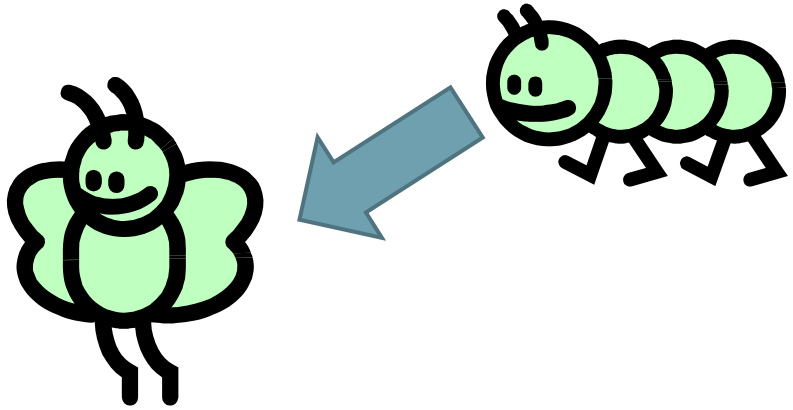


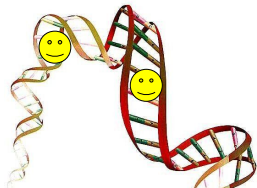
tertiary structure



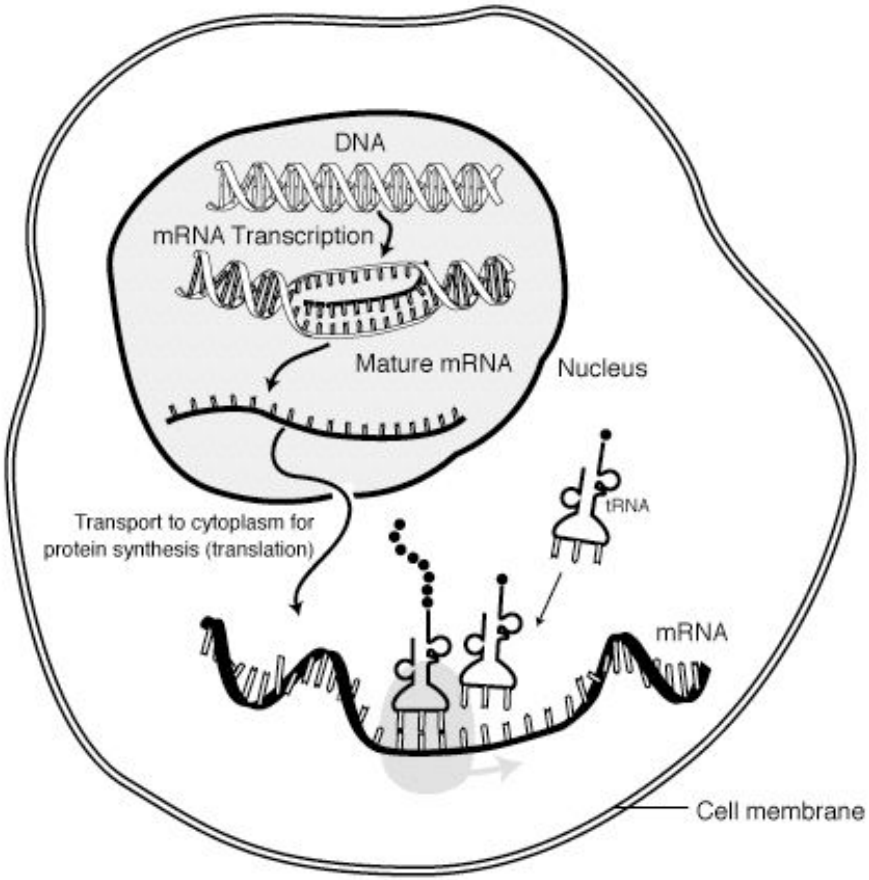
quaternary structure

- The large-scale study of proteins, particularly their structures and functions

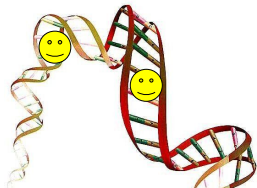




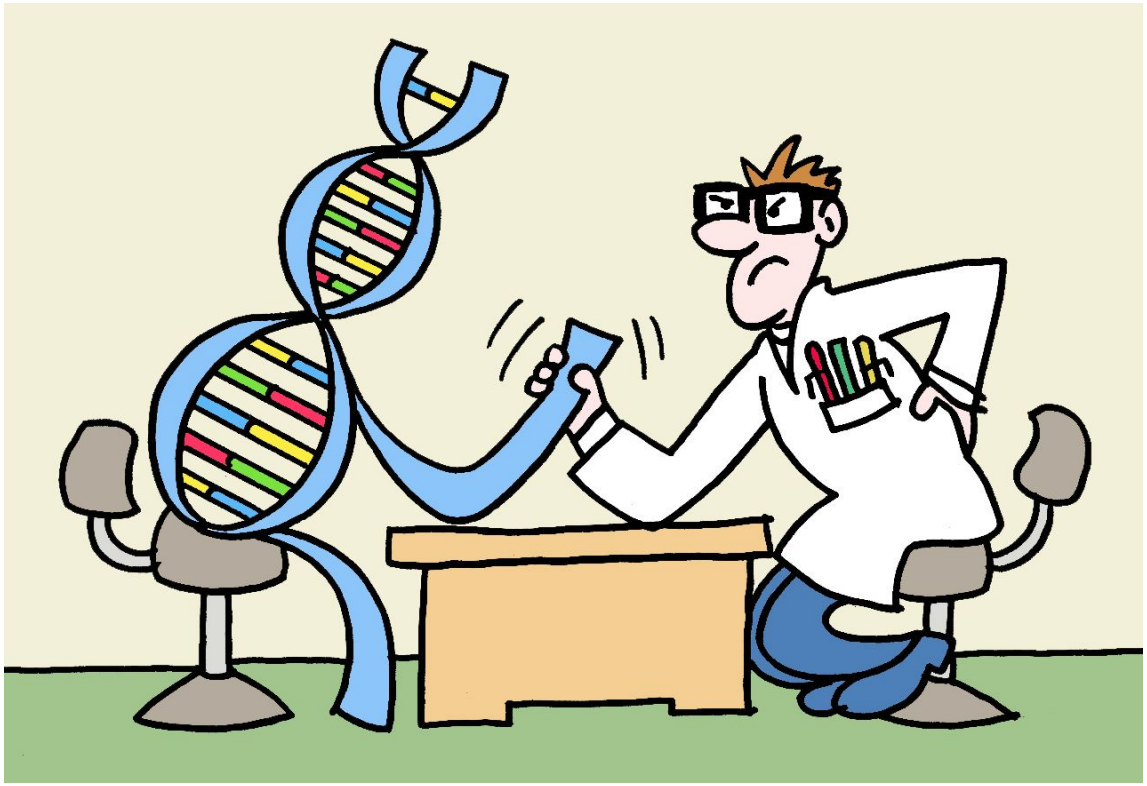
Transcriptomics

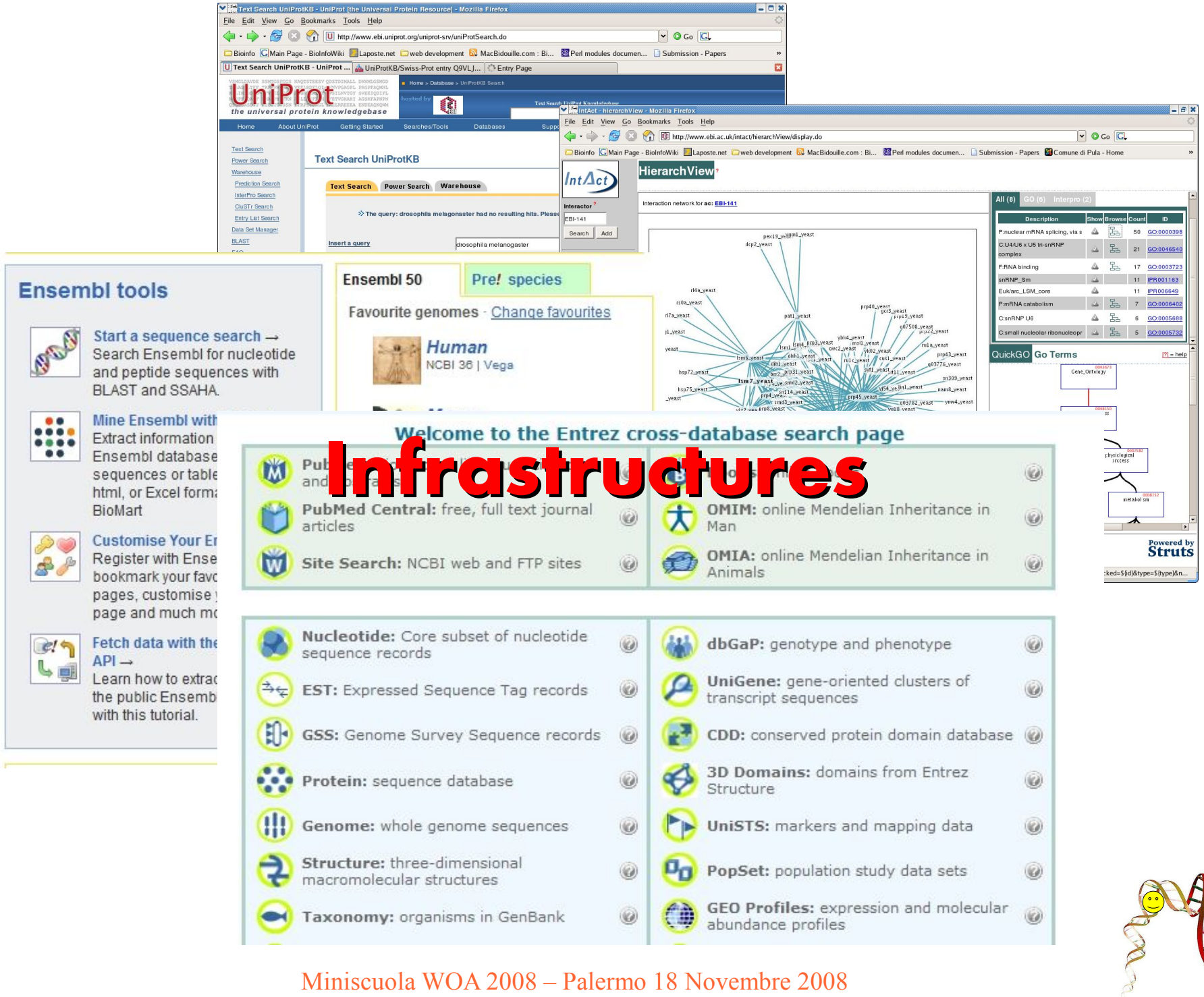


- The transcriptome is the set of all messenger RNA (mRNA) molecules, or "transcripts", produced in one or a population of cells



Open Problems in Bioinformatics

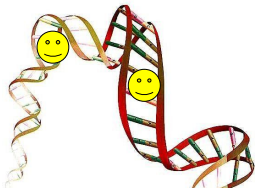




Infrastructures

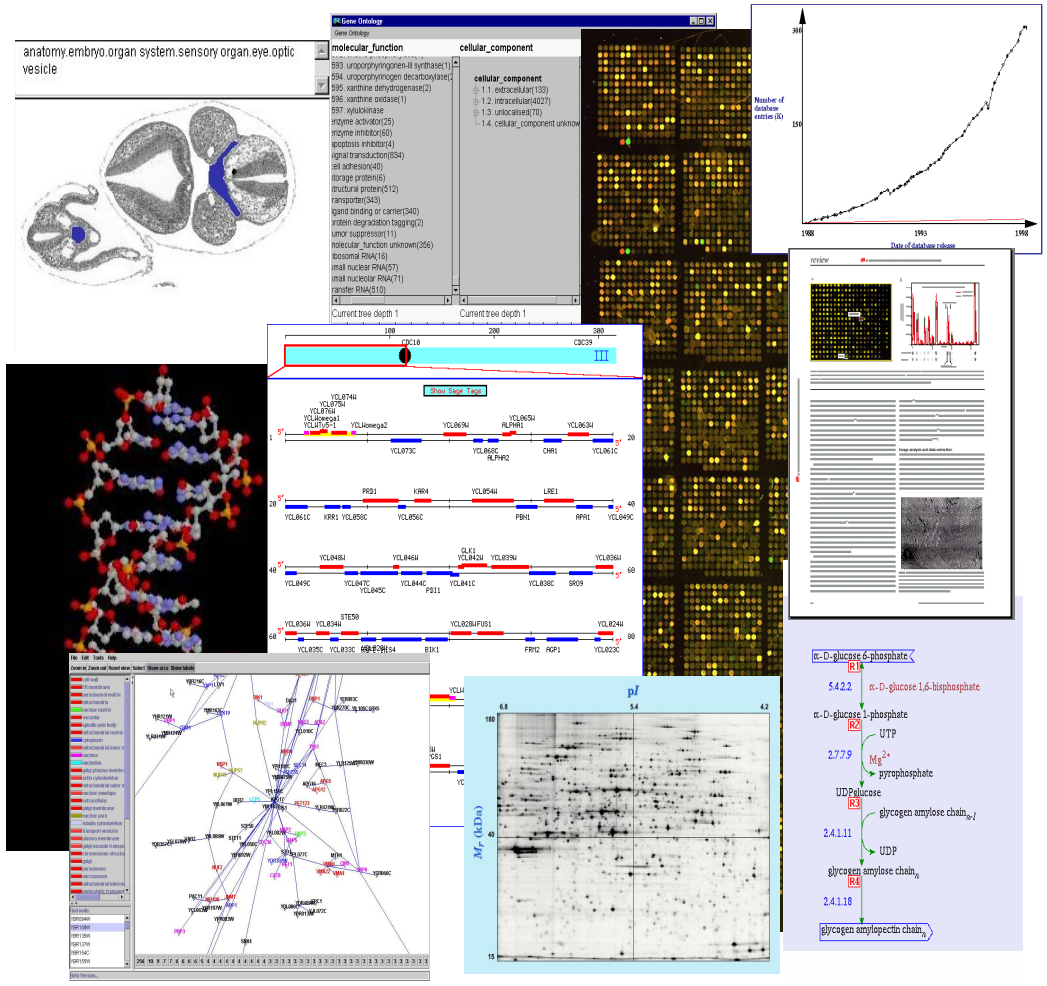
Infrastructures

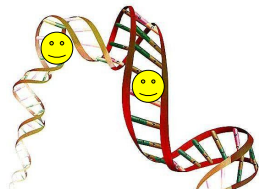
- Infrastructures are mainly concerned with the handling of data repositories and/or applications throughout suitable interfaces
- Users are allowed to access relevant data throughout
  - generic/specialized browsers
  - web services

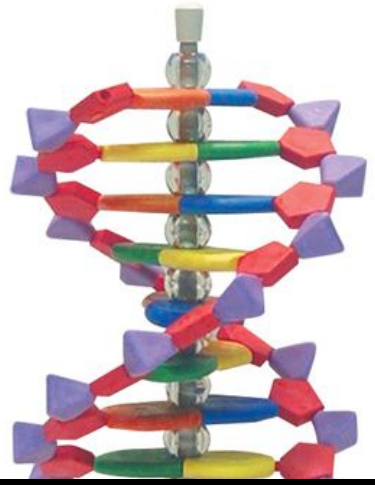


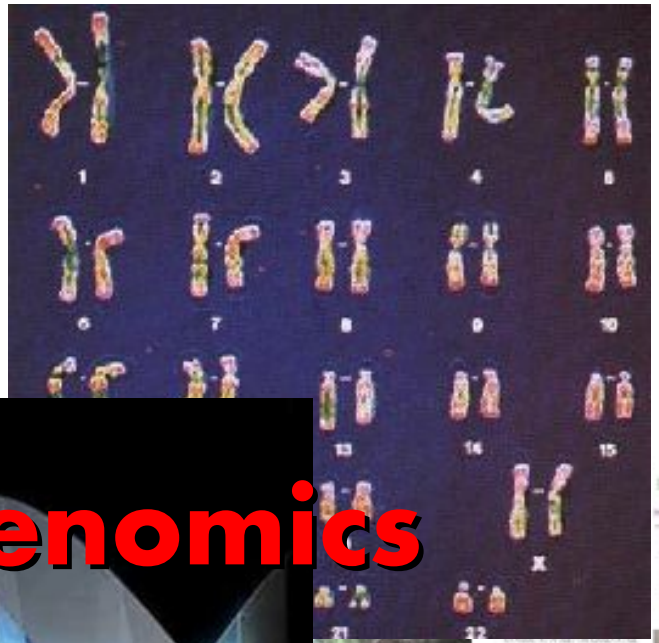
Infrastructures

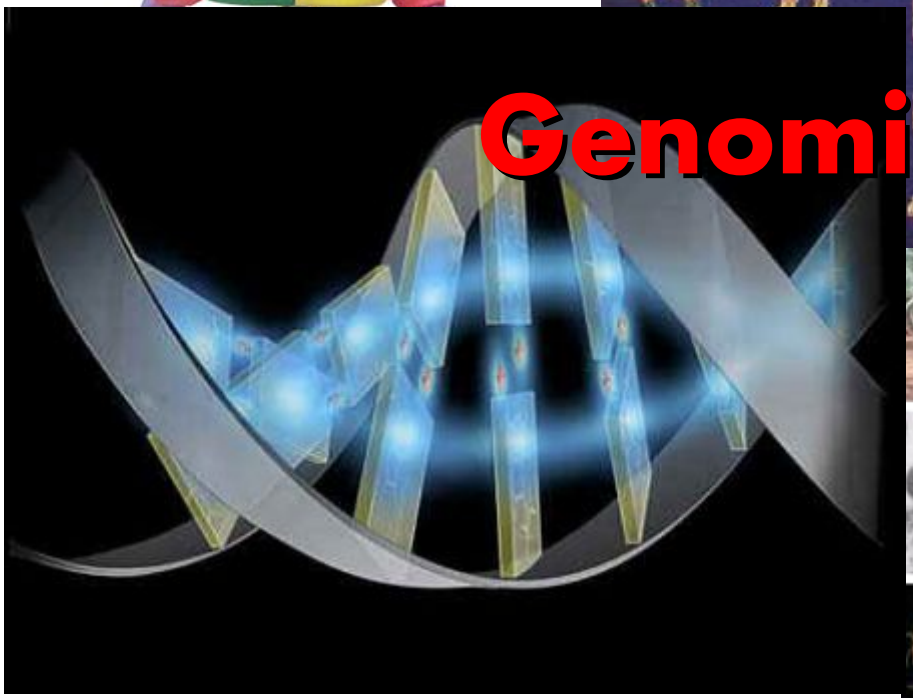
- Large amounts of data:
  - Highly heterogeneous
  - Highly complex and inter-related
  - Volatile



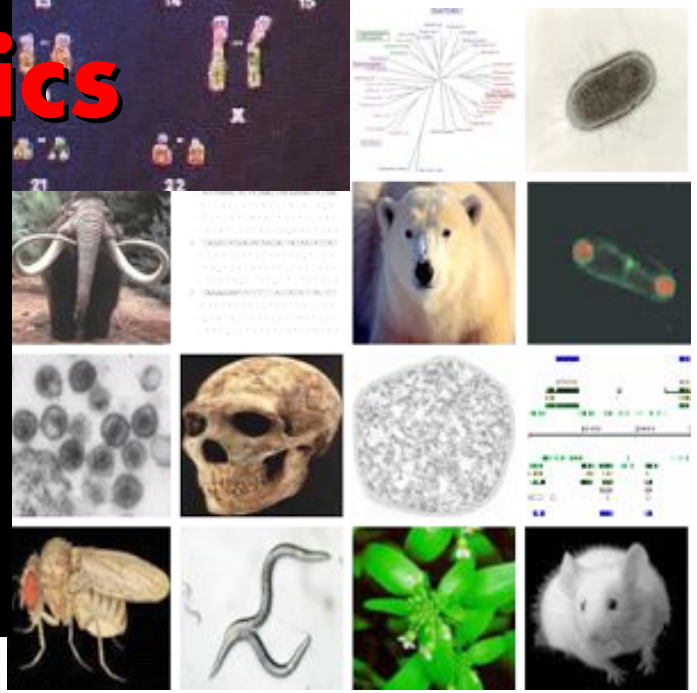


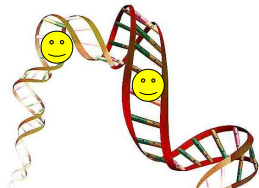






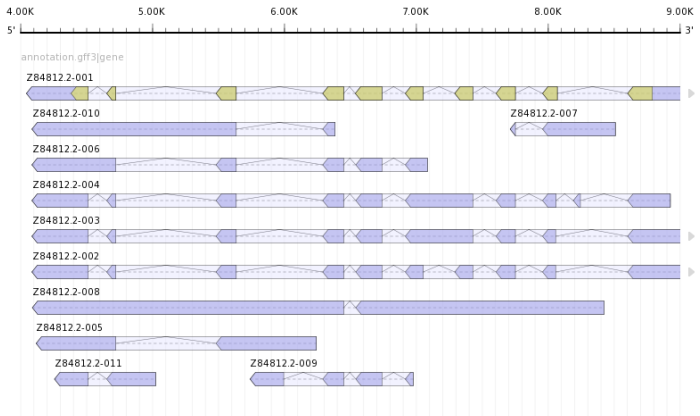
Genomics





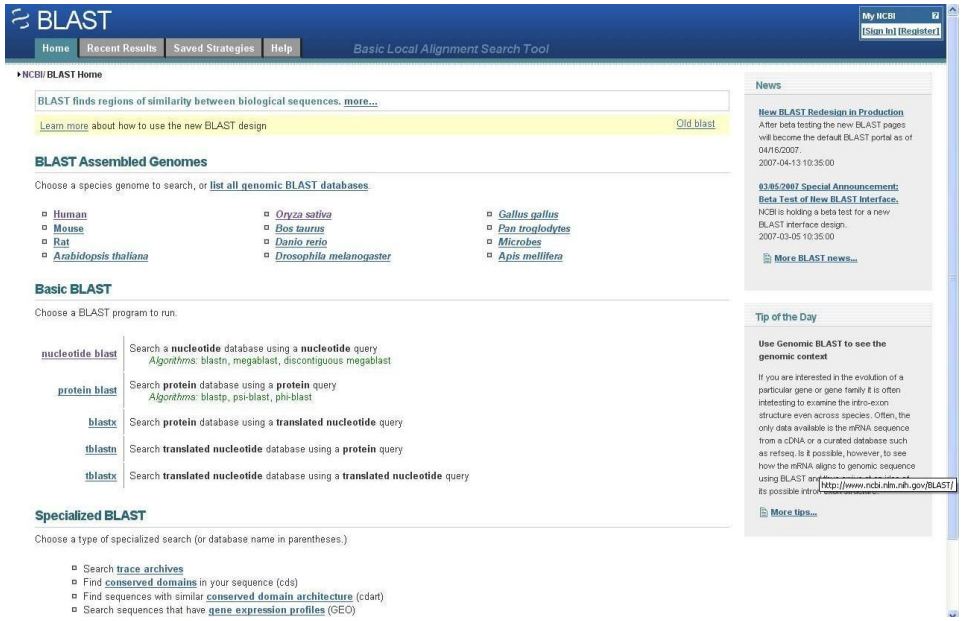
Genome Annotation

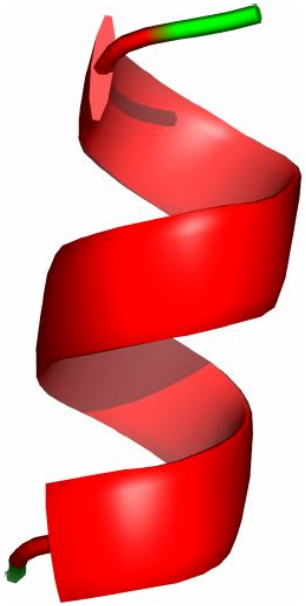
- Genome annotation is the next major challenge for the Human Genome Project
- Genome annotation is the process of attaching biological information to sequences
- It consists of two main steps
  - identifying elements on the genome (gene finding)
  - attaching biological information to these elements

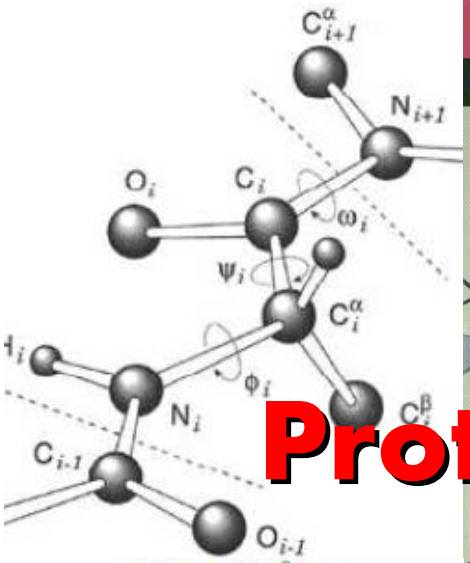


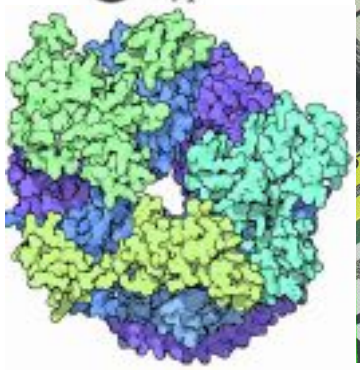
Genome Annotation

- Automatic annotation tools try to perform all this by computer analysis, as opposed to manual annotation (a.k.a. curation)
- The basic level of annotation is using BLAST **[Altschul90]** for finding similarities

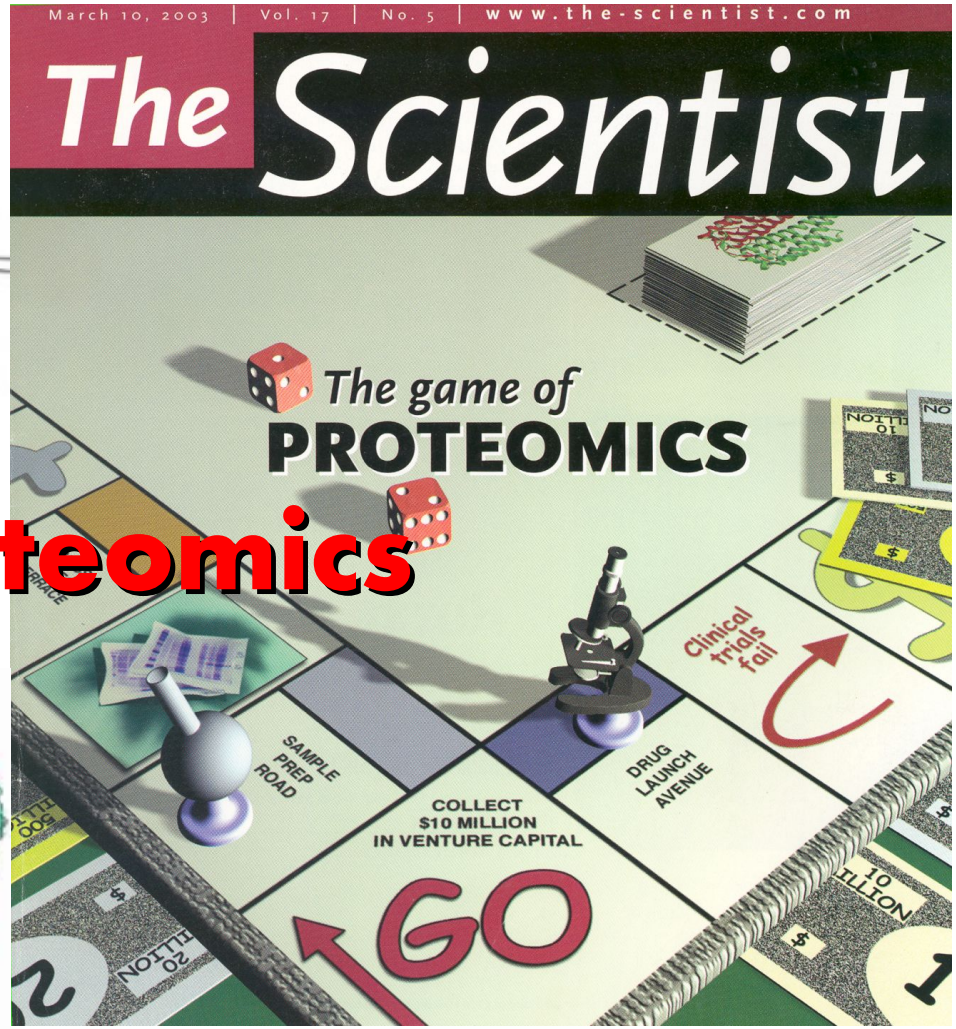


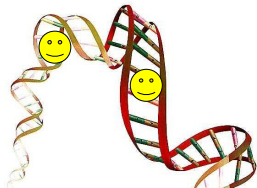



C^{α}_{i+1}
N_{i+1}
O_i
C_i
ω_i
ψ_i
C^{α}_i
ϕ_i
N_i
C^{β}_i
C_{i-1}
O_{i-1}



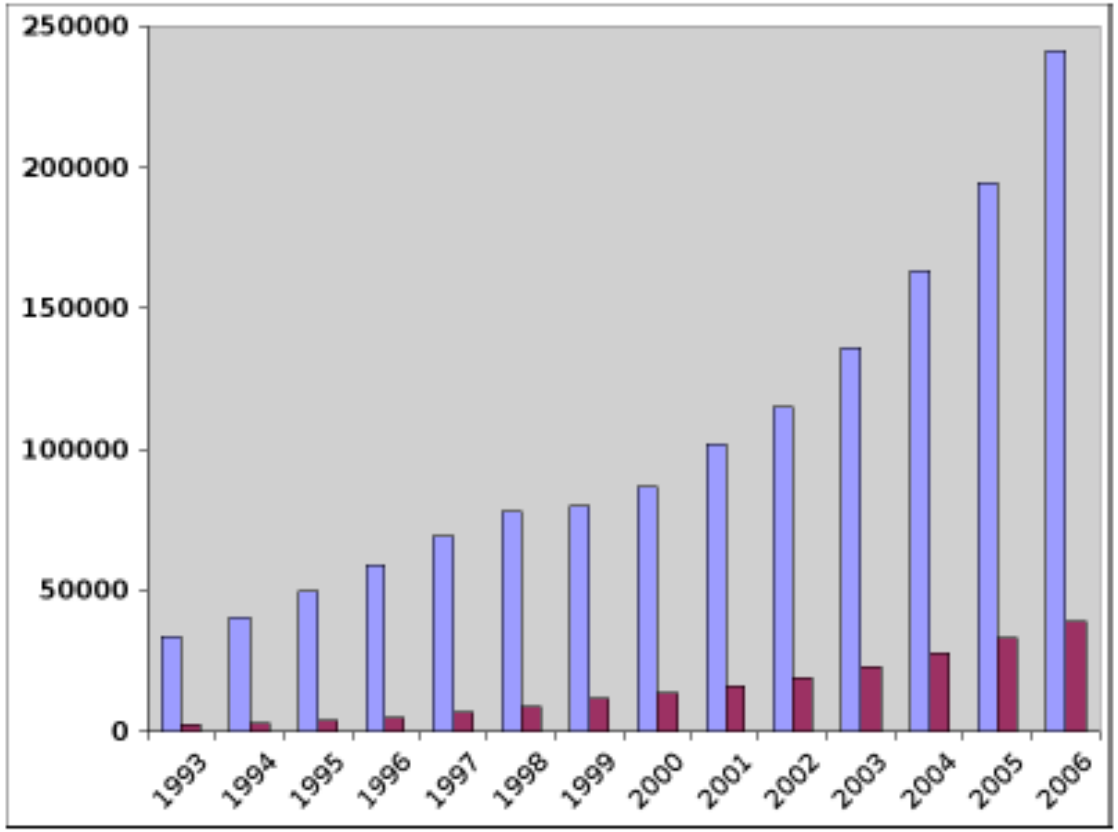
Proteomics


March 10, 2003 | Vol. 17 | No. 5 | www.the-scientist.com
The Scientist
The game of
PROTEOMICS
SAMPLE PREP ROAD
COLLECT $10 MILLION IN VENTURE CAPITAL
GO
DRUG LAUNCH AVENUE
Clinical trials fail

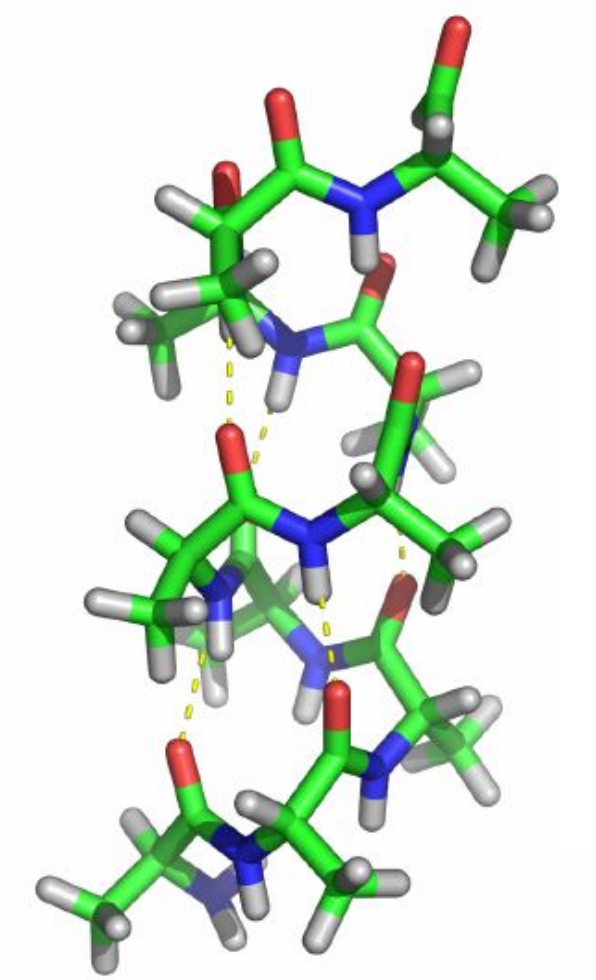


Structure Prediction: Motivation

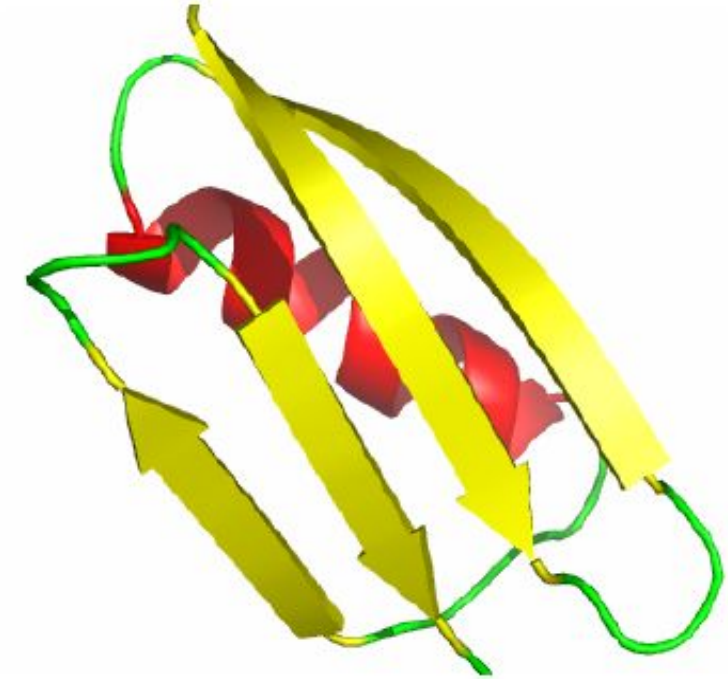
- Large gap between the number of known protein sequences and known protein structures



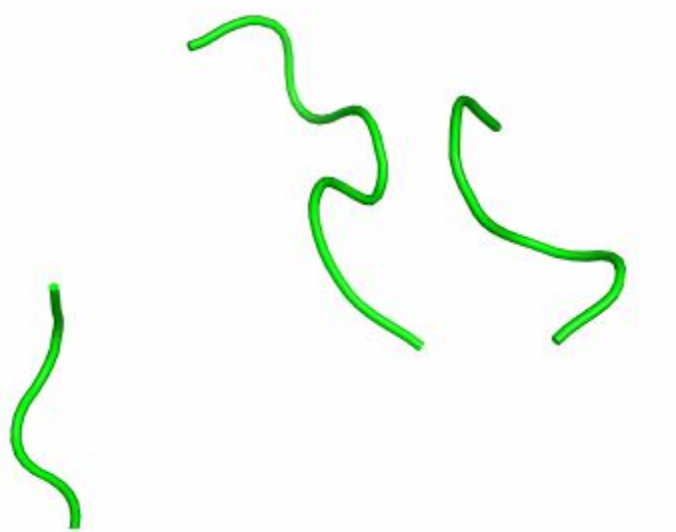
Secondary Structure Prediction



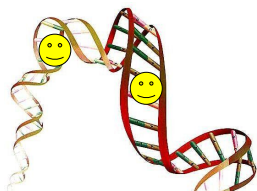
α-helix



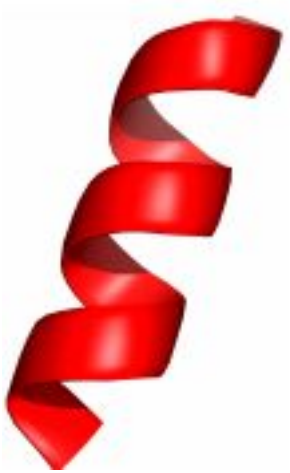
β-sheet

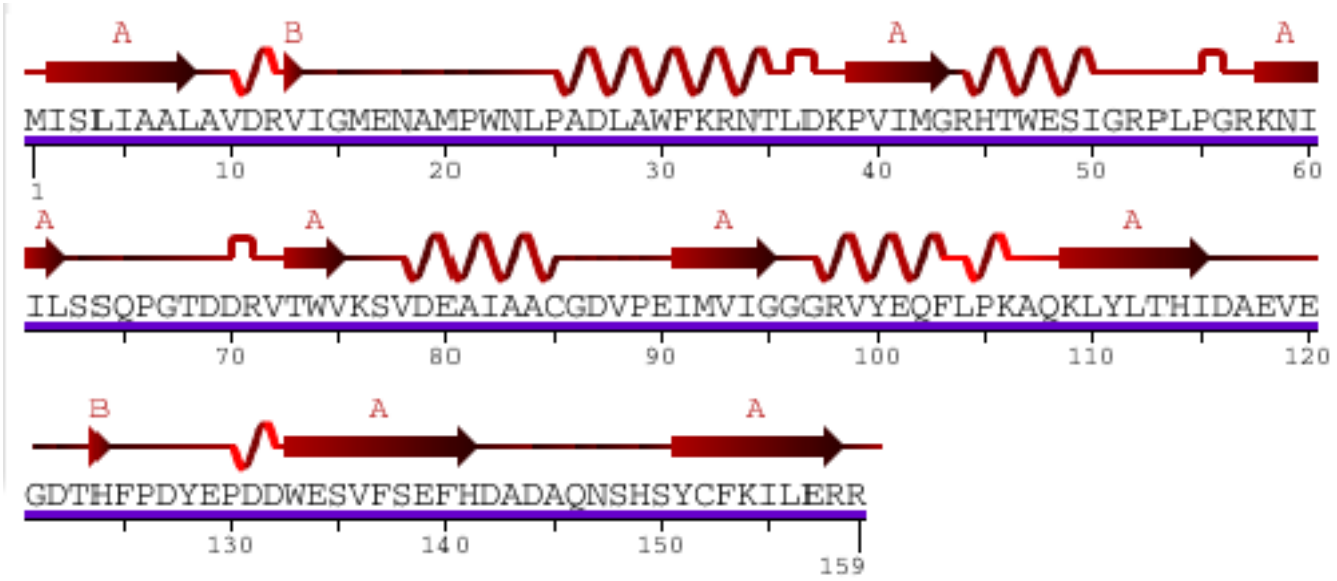


coil

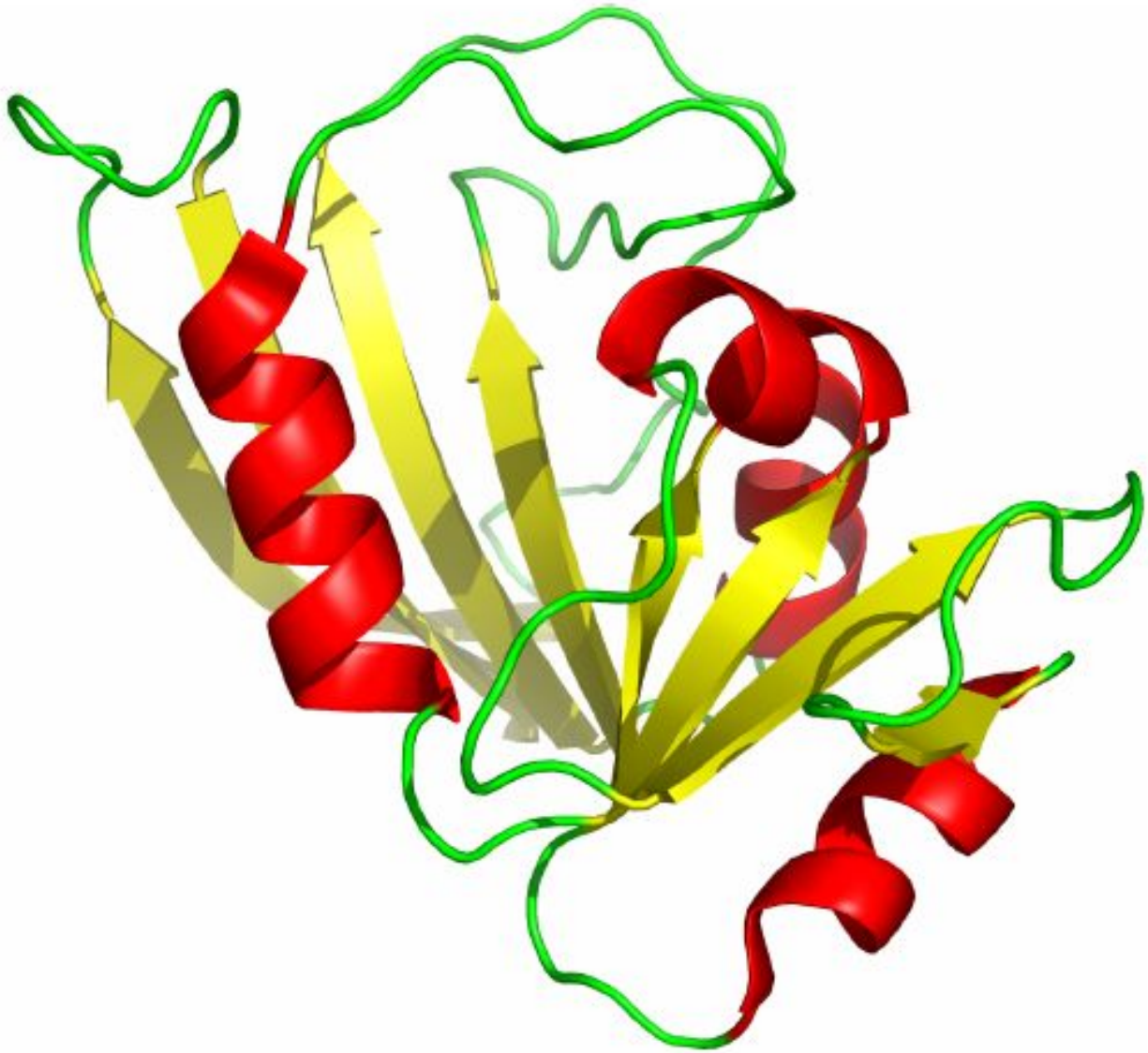


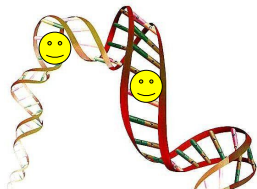
Secondary Structure Prediction



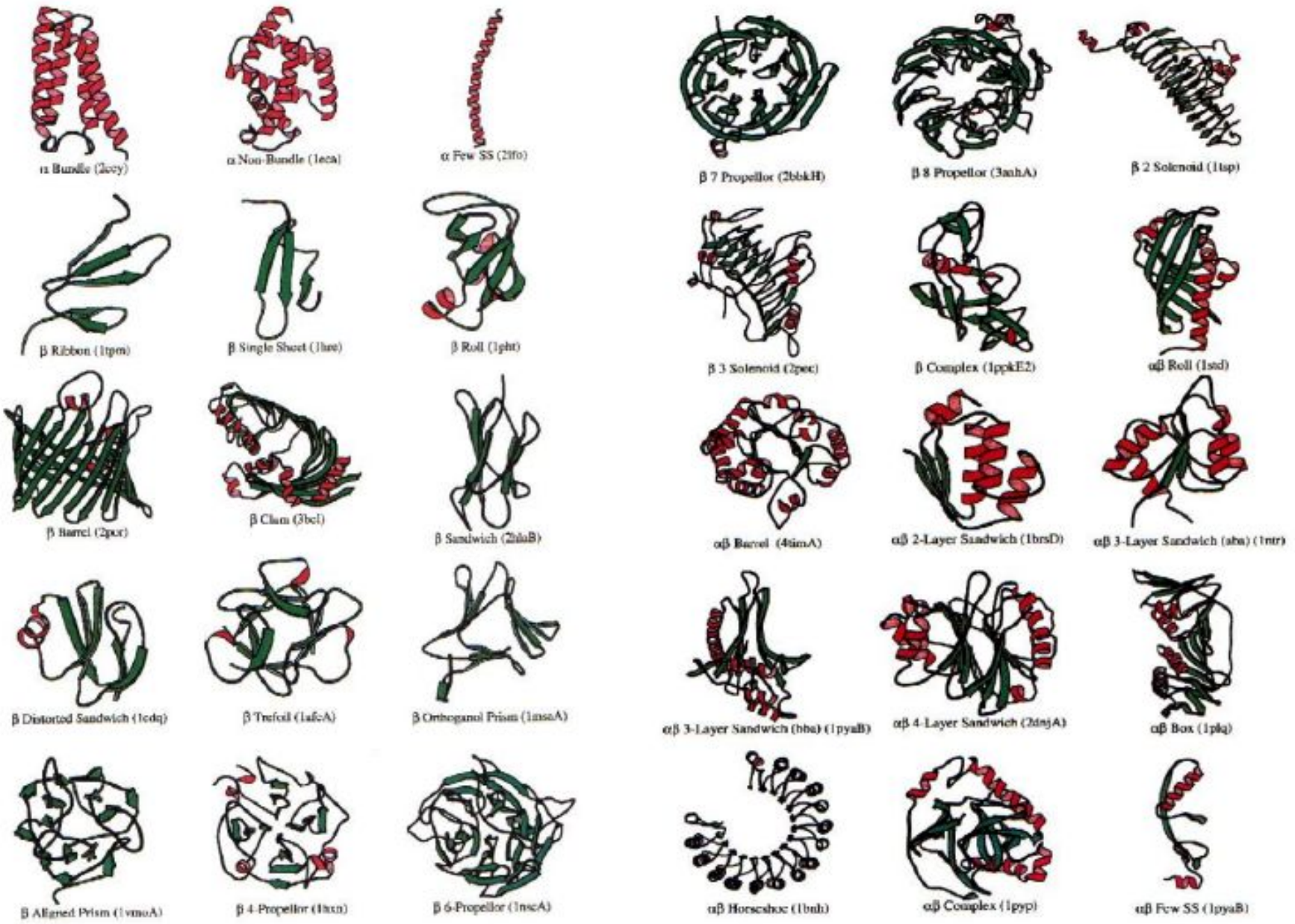


Tertiary Structure Prediction



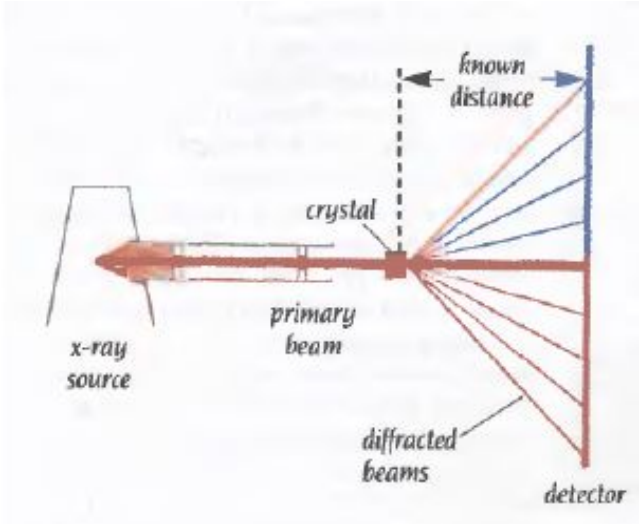


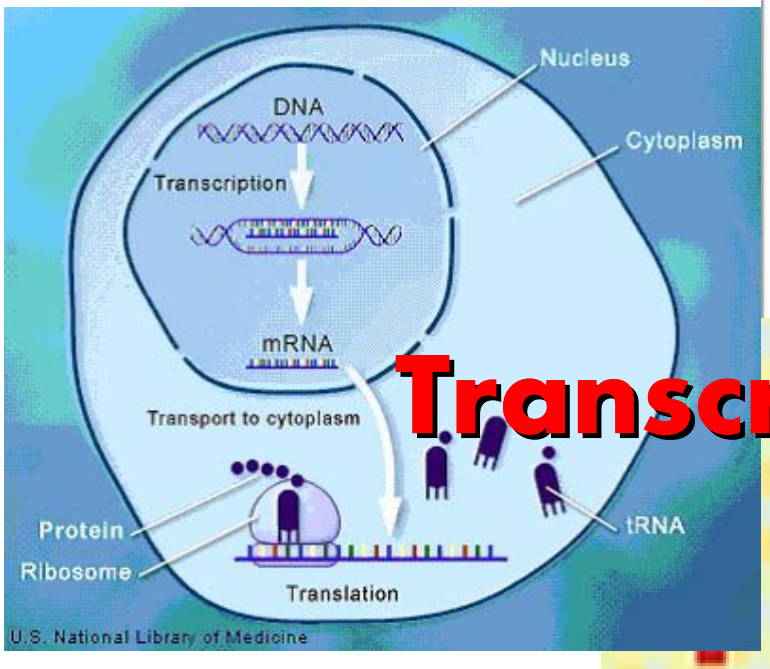
Tertiary Structure Prediction

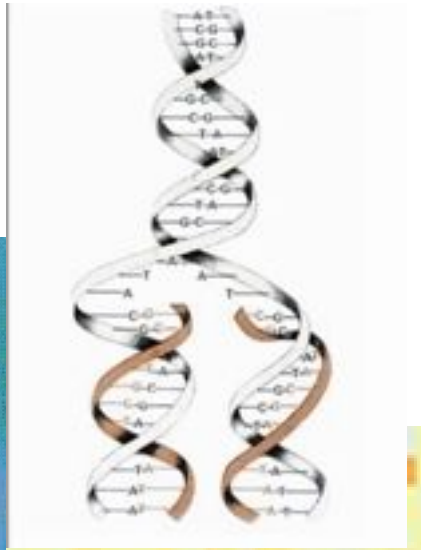


Tertiary Structure Prediction

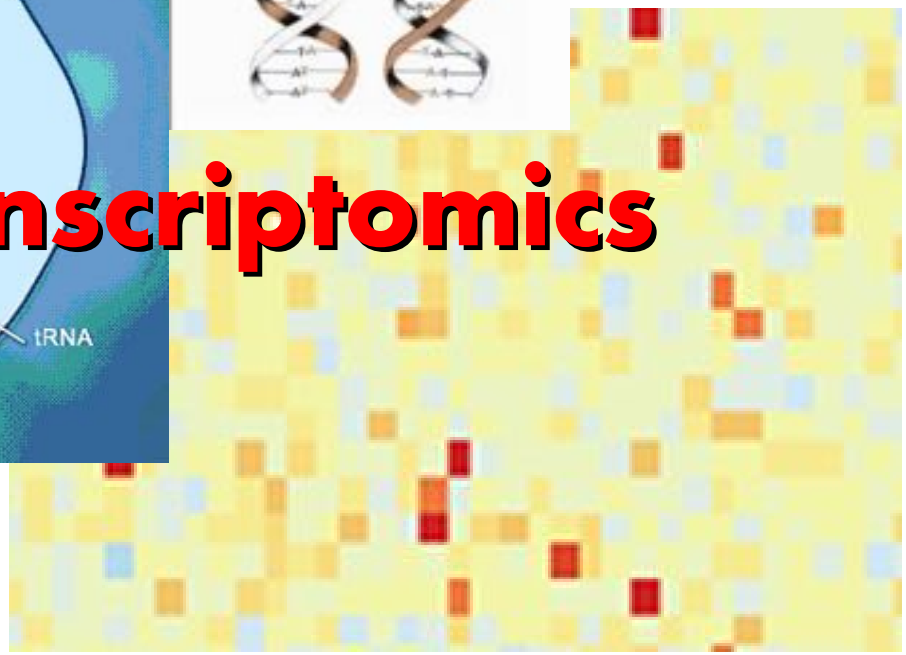
- Two main methods
  - X-ray crystallography
    - Principle: Scattering of x-rays by electrons
    - Application: small molecules, proteins, viruses, ribosome
  - NMR
    - Principle: nuclear spin transition of isotropes
    - Application: small, soluble proteins (monomeric)





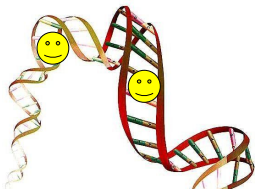


Transcriptomics



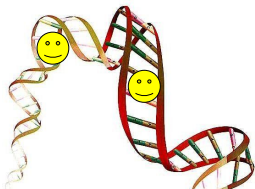
Functional Genomics

- Functional genomics attempts to make use of the vast wealth of data produced by genomic projects to describe gene (and protein) functions and interactions
- Unlike genomics and proteomics, functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein-protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures



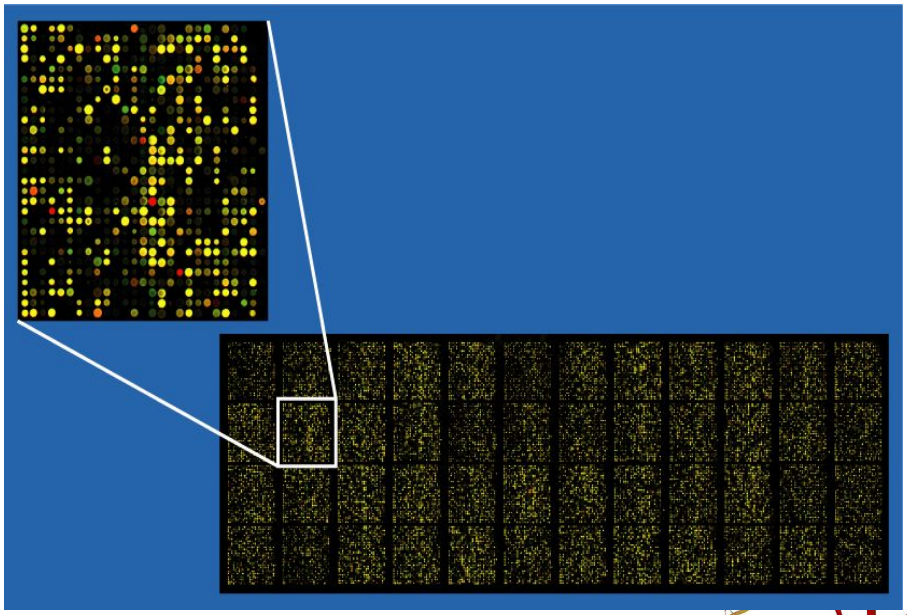
Functional Genomics

- Functional genomics includes
  - function-related aspects of the genome itself such as mutation and polymorphism (such as SNP) analysis
  - measurement of molecular activities (gene/protein expression)



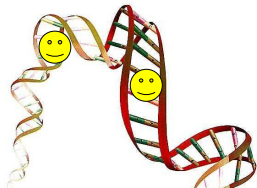
DNA Microarrays

- A DNA microarray is a multiplex technology used in molecular biology and in medicine
- It consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, each containing picomoles of a specific DNA sequence



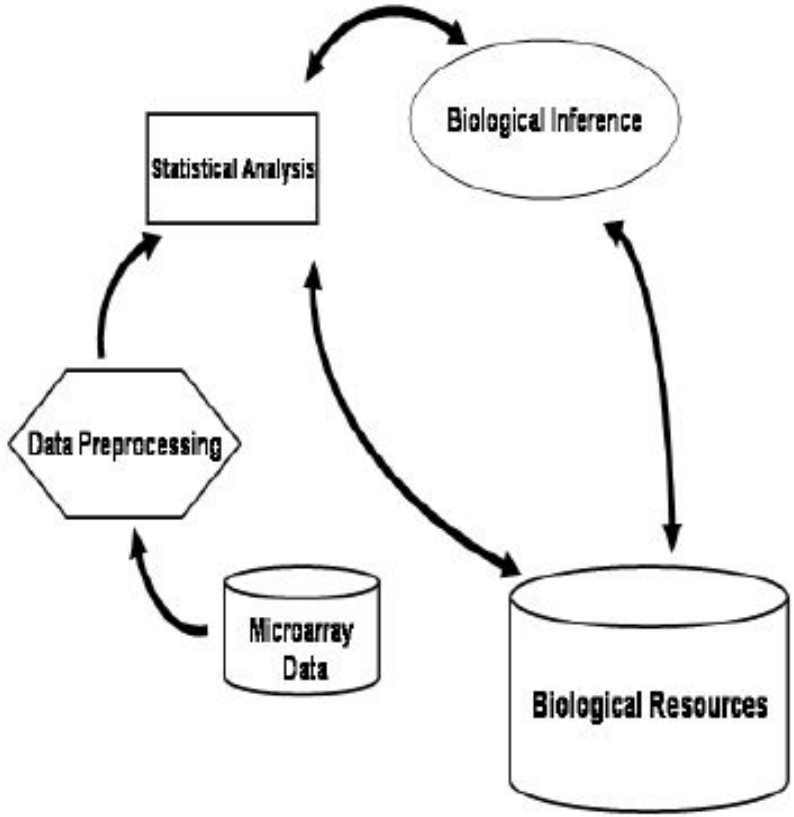
Gene Expression

- DNA microarrays can be used to measure changes in expression levels or to detect single nucleotide polymorphisms (SNPs)
- Microarray technology has been very effective in producing large amount of gene expression data by measuring the transcription levels of thousands of genes simultaneously
- A typical microarray experiment can generate up to millions of data points for analysis

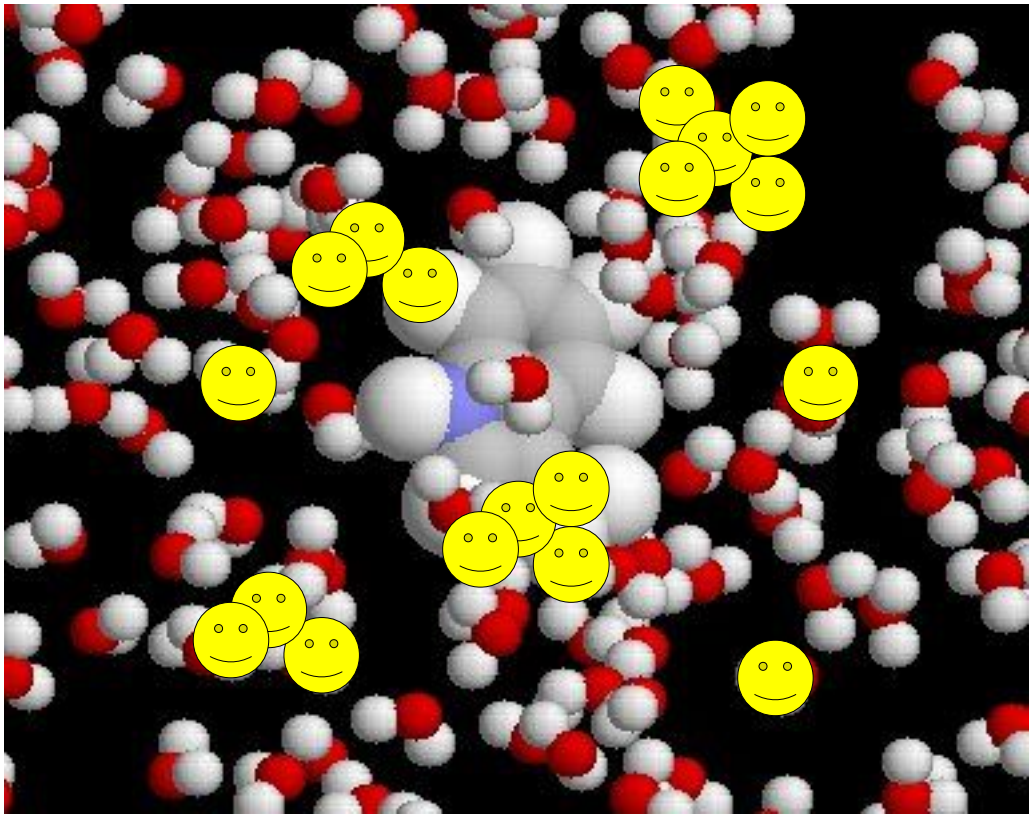


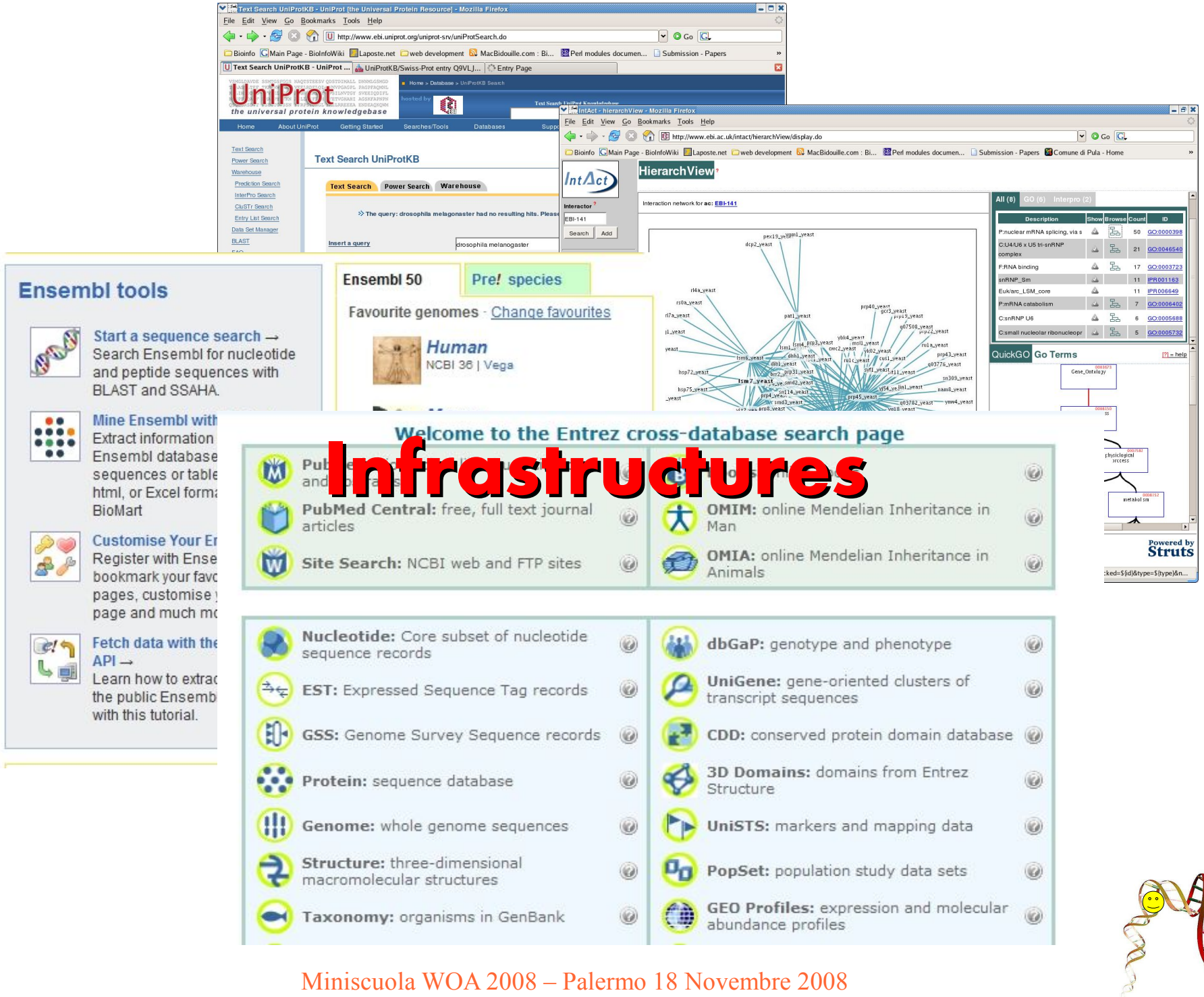
Gene Expression

- There are three data processing steps in gene expression analysis
  - data preprocessing
  - statistical analysis
  - biological inference



Proposed Agent Solutions

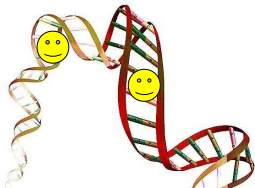




Infrastructures

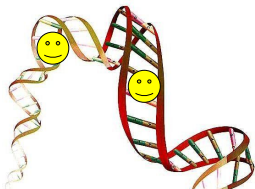
Infrastructures

- The multiplicity of data repositories and analysis tools naturally promotes the adoption of a multiagent approach
- In fact, the corresponding environment is
  - open
  - dynamic
  - distributed



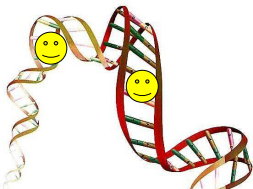
Infrastructures

- MAS infrastructures
  - MyGrid **[Stevens03]**
  - GeneWeaver **[Bryson01]**
  - BioAgent **[Corradini05]**
  - A MAS for Retrieving Bioinformatics Publications **[Armano07]**



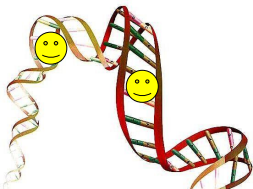
myGrid

- myGrid supports:
  - e-Scientists
    - Lab book application to show off myGrid core components
  - Developers
    - myGrid-in-a-Box developers kit
    - Incorporating third party tools and services

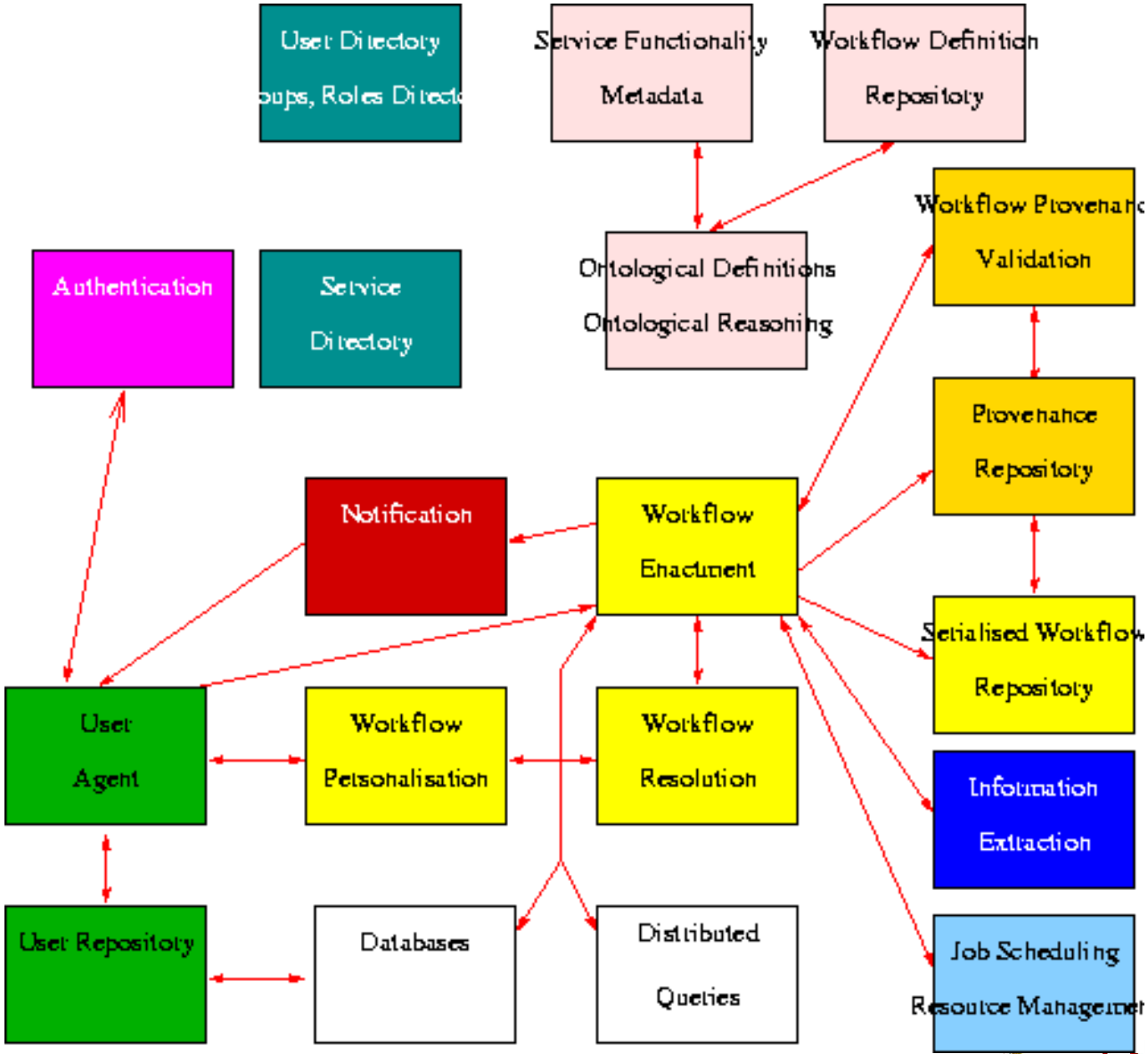


myGrid

- Agents in myGrid:
  - User agents
    - able to customize and personalize data
  - Agent communication languages
    - offering a generic and portable communication medium
  - Negotiation
    - allowing multiple distributed entities to reach service level agreements

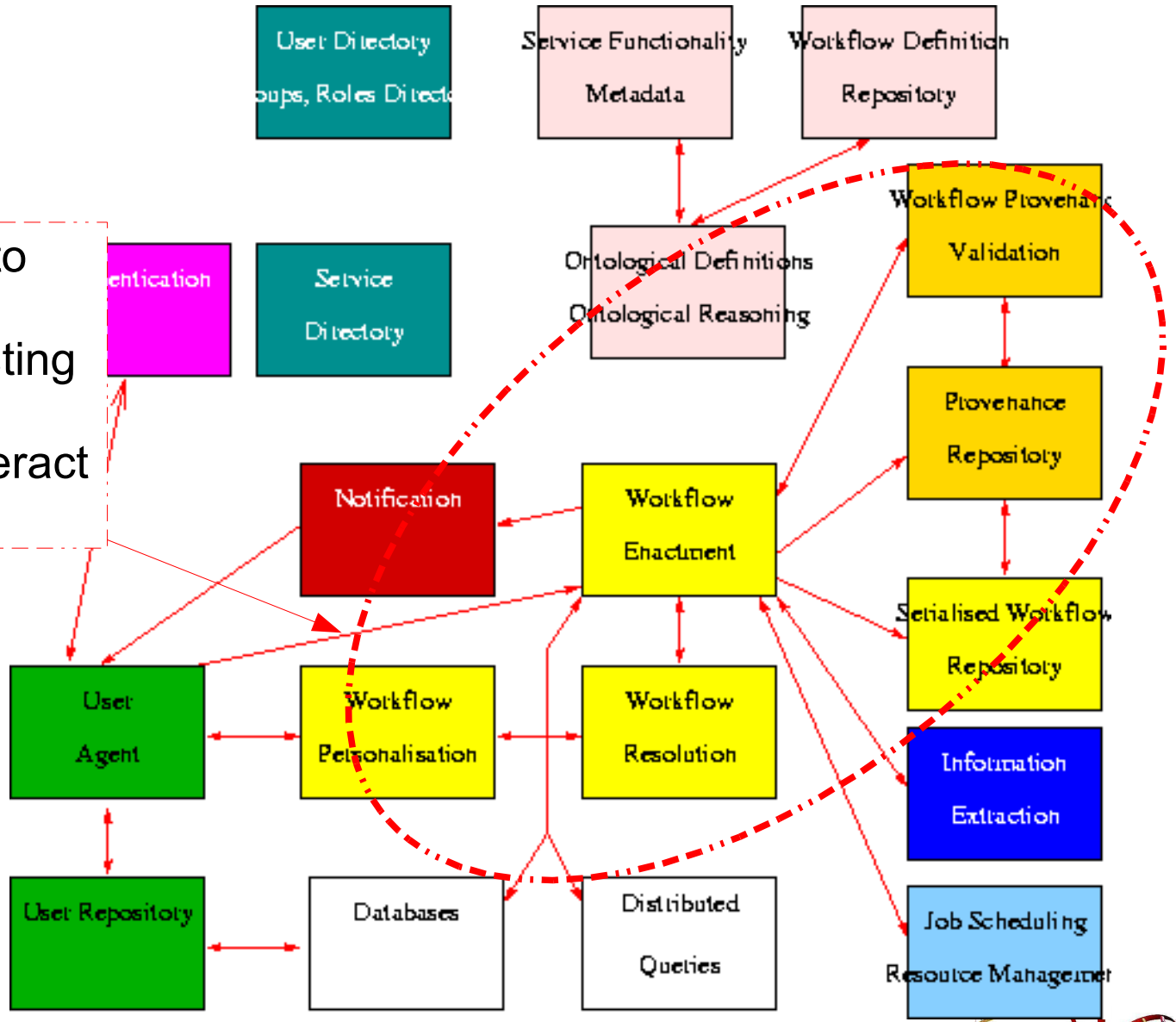


myGrid

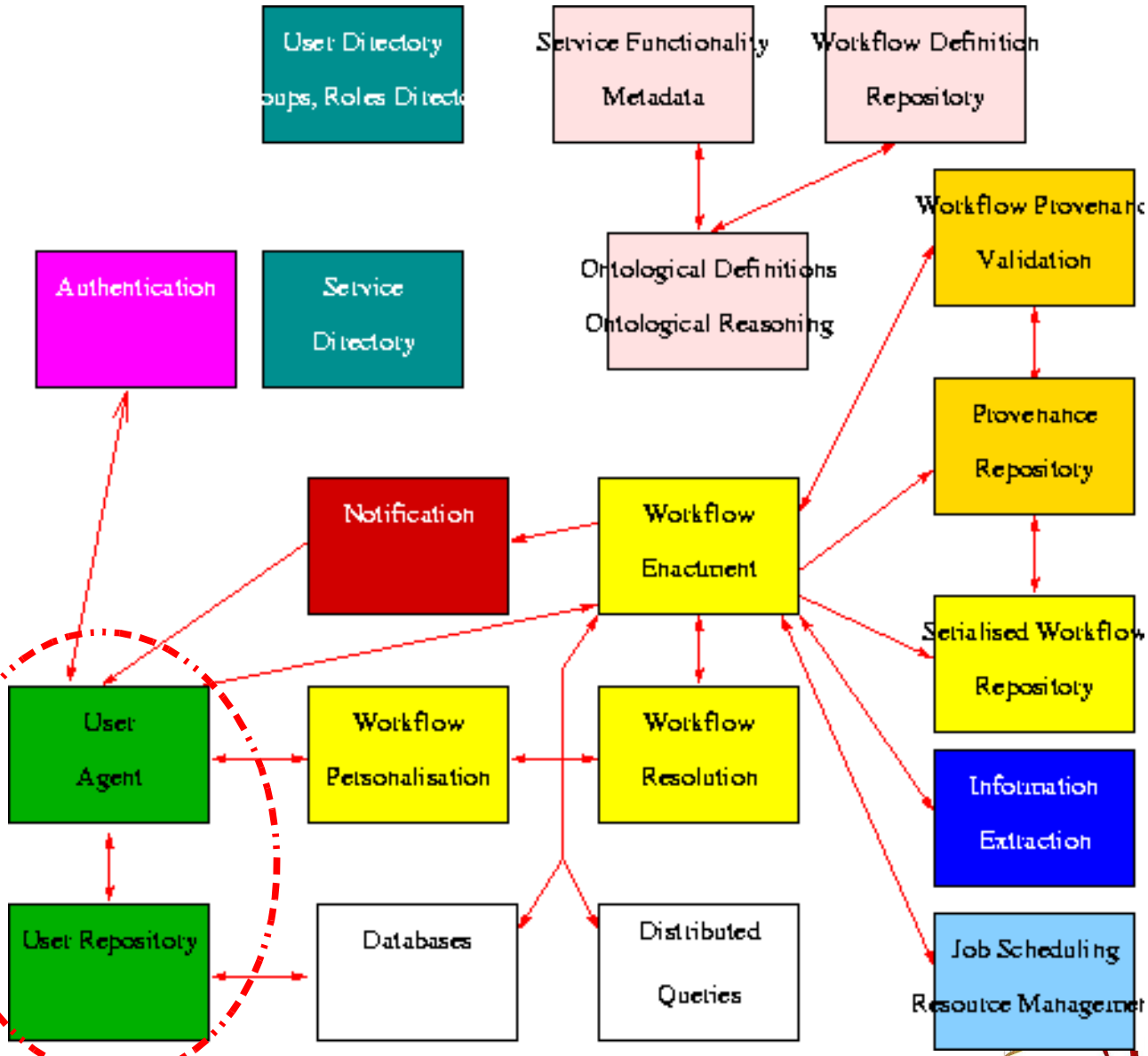

User Directory
oups, Roles Direct
Service Functionality
Metadata
Workflow Definition
Repository
Workflow Provenanc
Validation
Authentication
Service
Directory
Ontological Definitions
Ontological Reasoning
Provenance
Repository
Notification
Workflow
Enactment
Serialised Workflow
Repository
User
Agent
Workflow
Personalisation
Workflow
Resolution
Information
Extraction
User Repository
Databases
Distributed
Queries
Job Scheduling
Resource Manageme

myGrid

- Given a workflow, is able to execute the script
- Can send requests to existing running services
- Can activate tools and interact with them



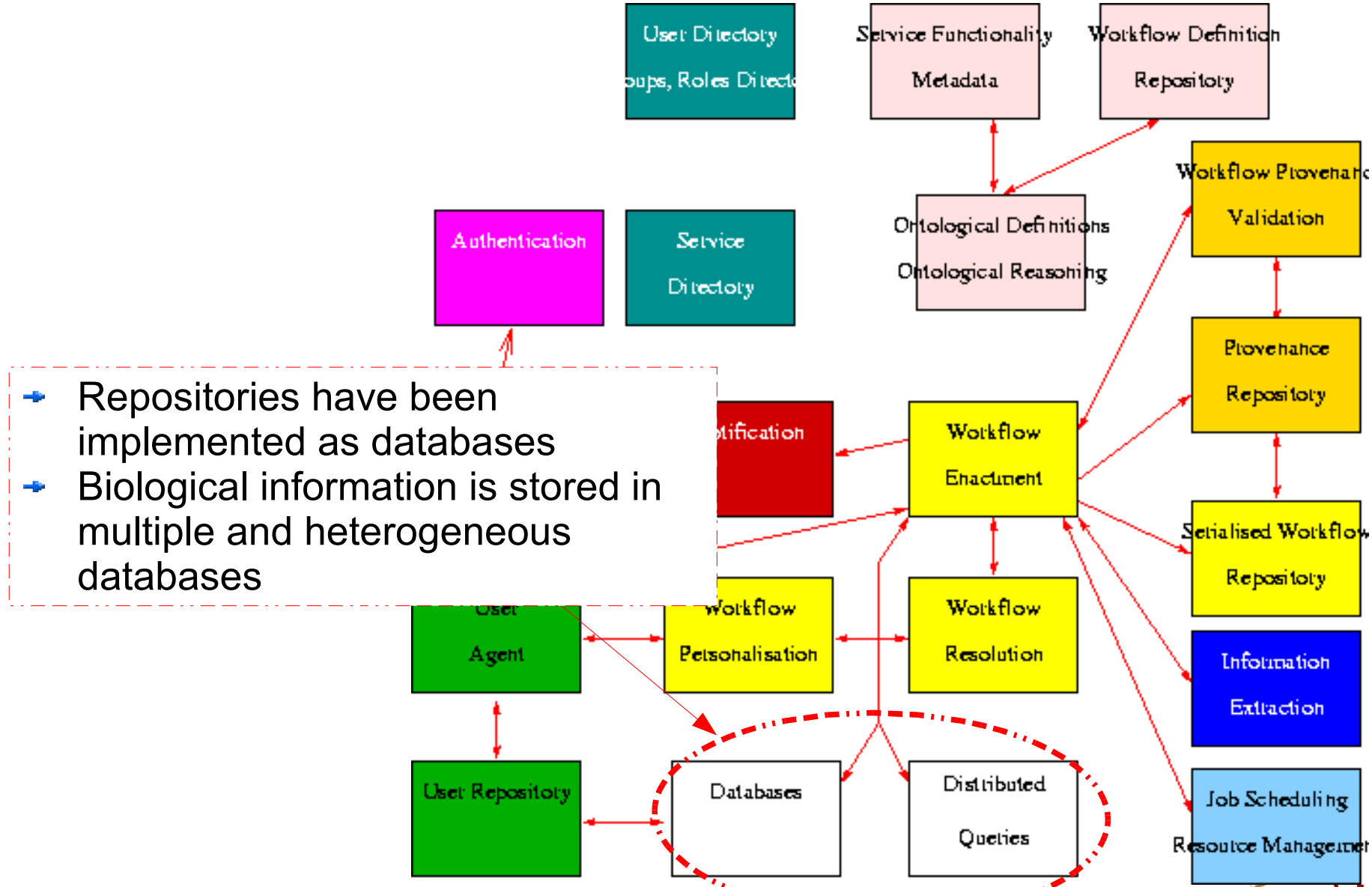
myGrid



- Interacts with the workflow enactment engine, suspending and resuming workflows

- All information about workflows is structured according to a set of ontologies

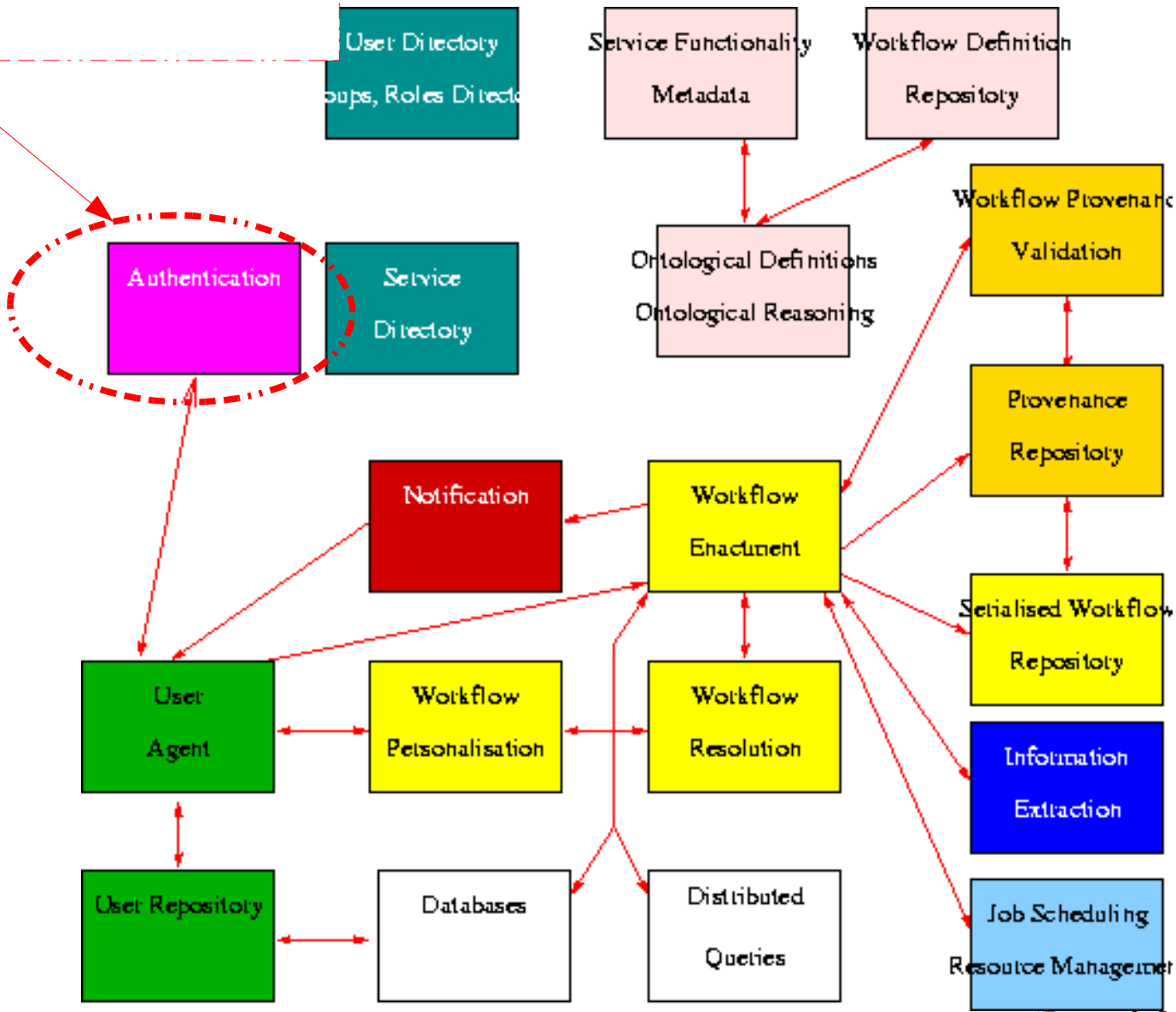
myGrid



- Repositories have been implemented as databases
- Biological information is stored in multiple and heterogeneous databases

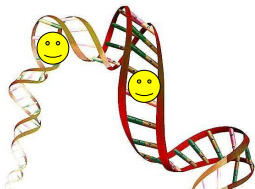
myGrid

- Provides X.509 certificates for users and objects needing verification



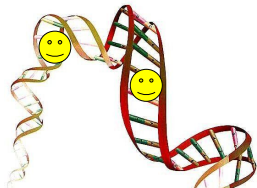
GeneWeaver

- GeneWeaver is a multiagent system aimed at addressing many of the problems concerning the management of data and analysis techniques in the bioinformatics domain



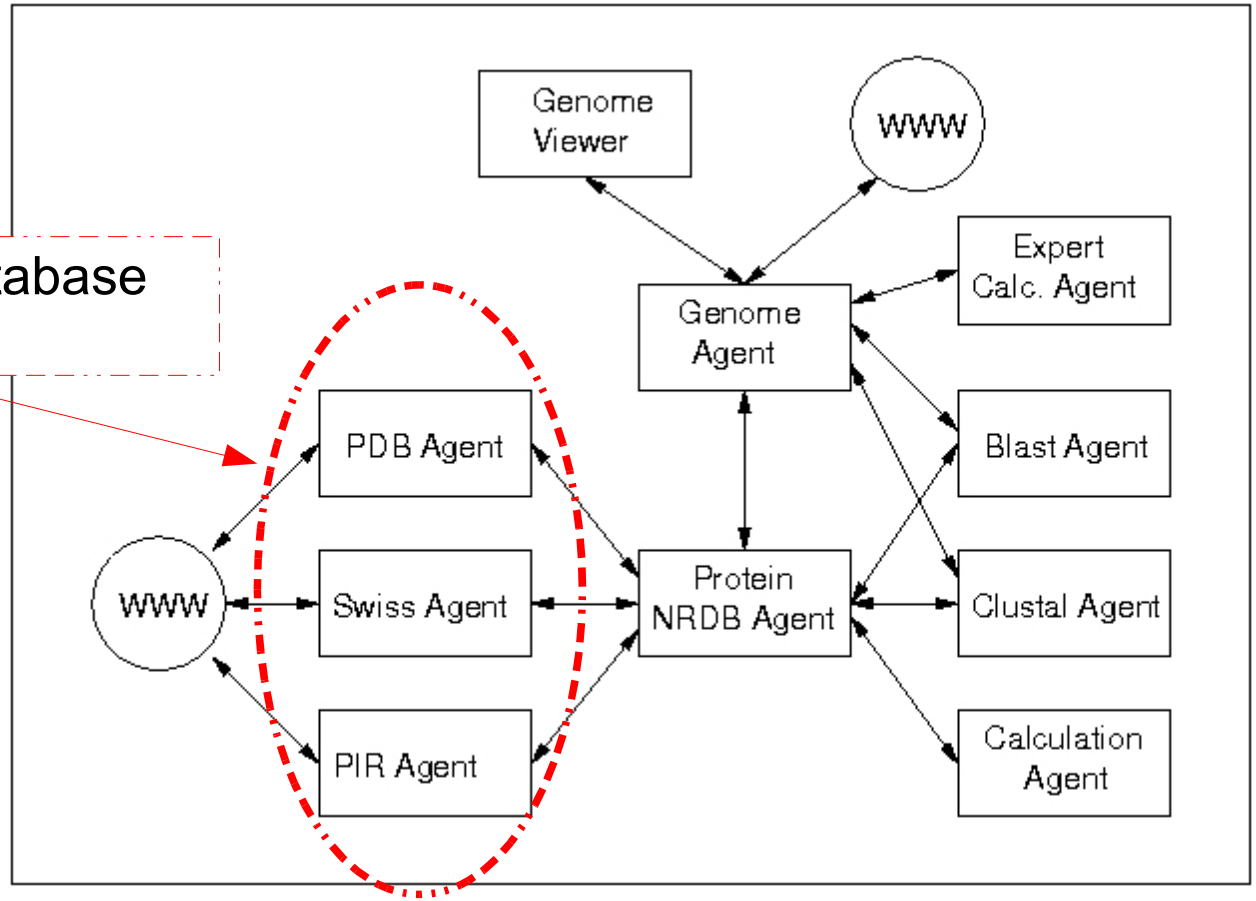
GeneWeaver

- Agents in GeneWeaver:
  - Broker agents
    - to register information about other agents
  - Primary database agents
    - to manage remote primary sequence databases
  - Non-redundant database agents
    - to construct and maintain non-redundant databases
  - Calculation agents
    - to emoby methods or tools for analysis of sequence data
  - Genome agents
    - to manage genomic information



GeneWeaver

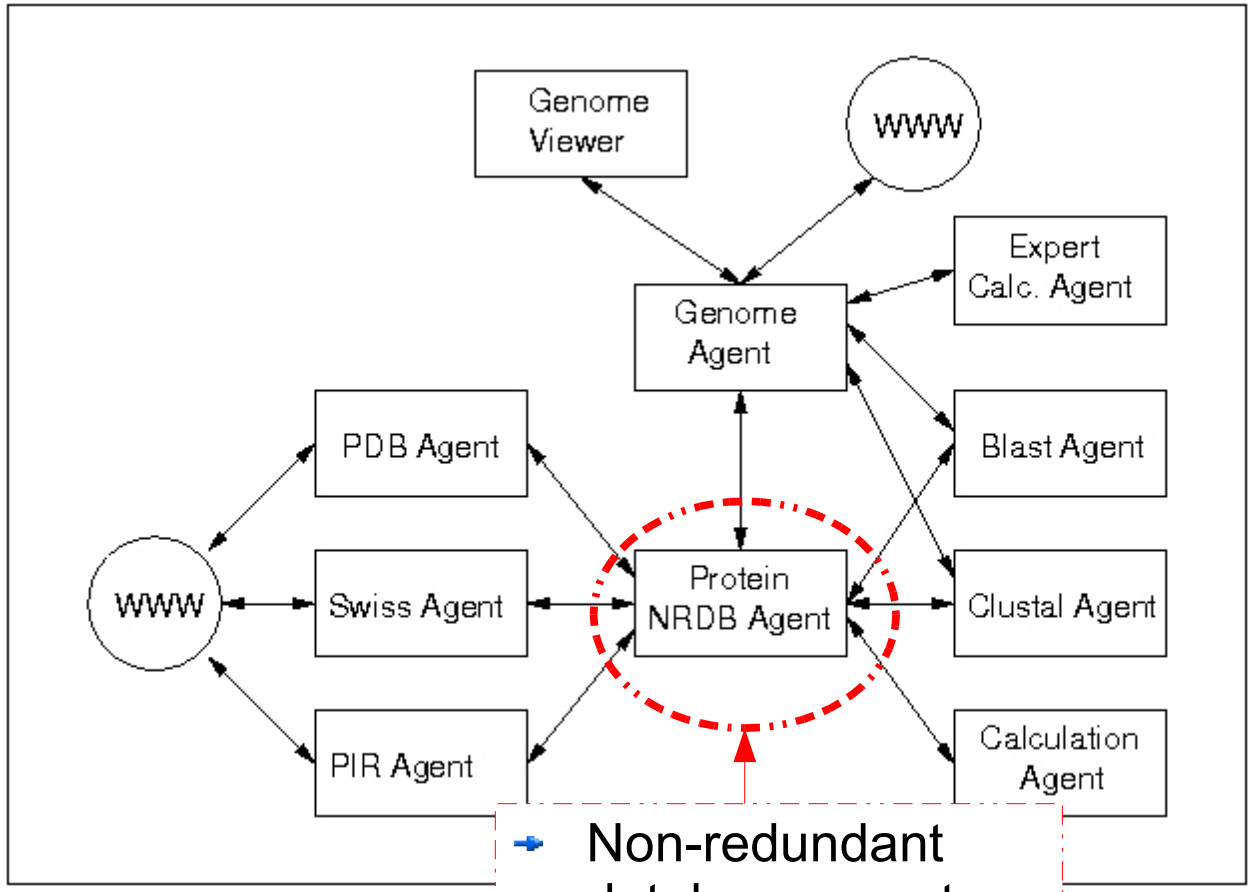
- Macro-architecture



- Protein database agents

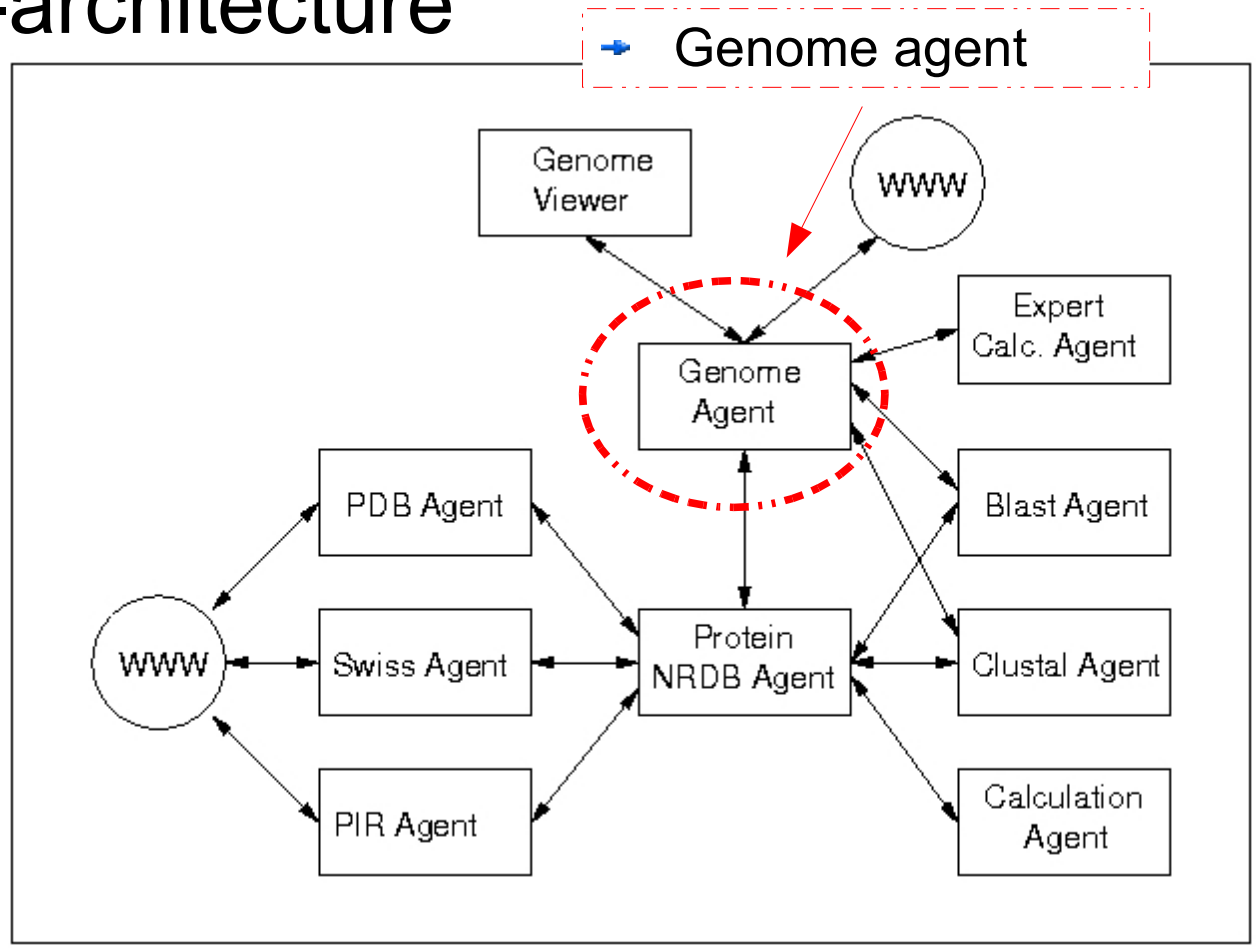
GeneWeaver

- Macro-architecture



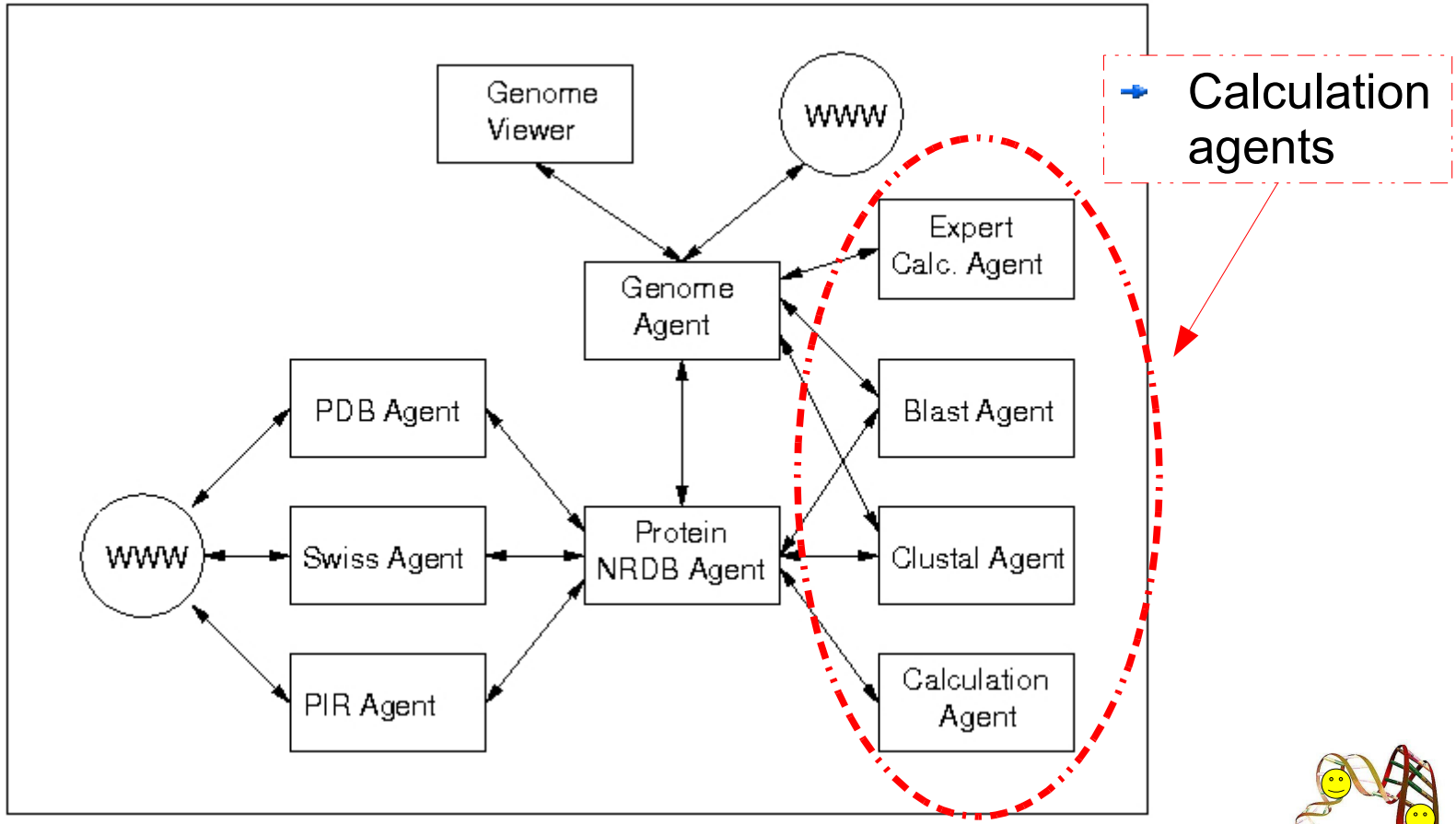
GeneWeaver

- Macro-architecture



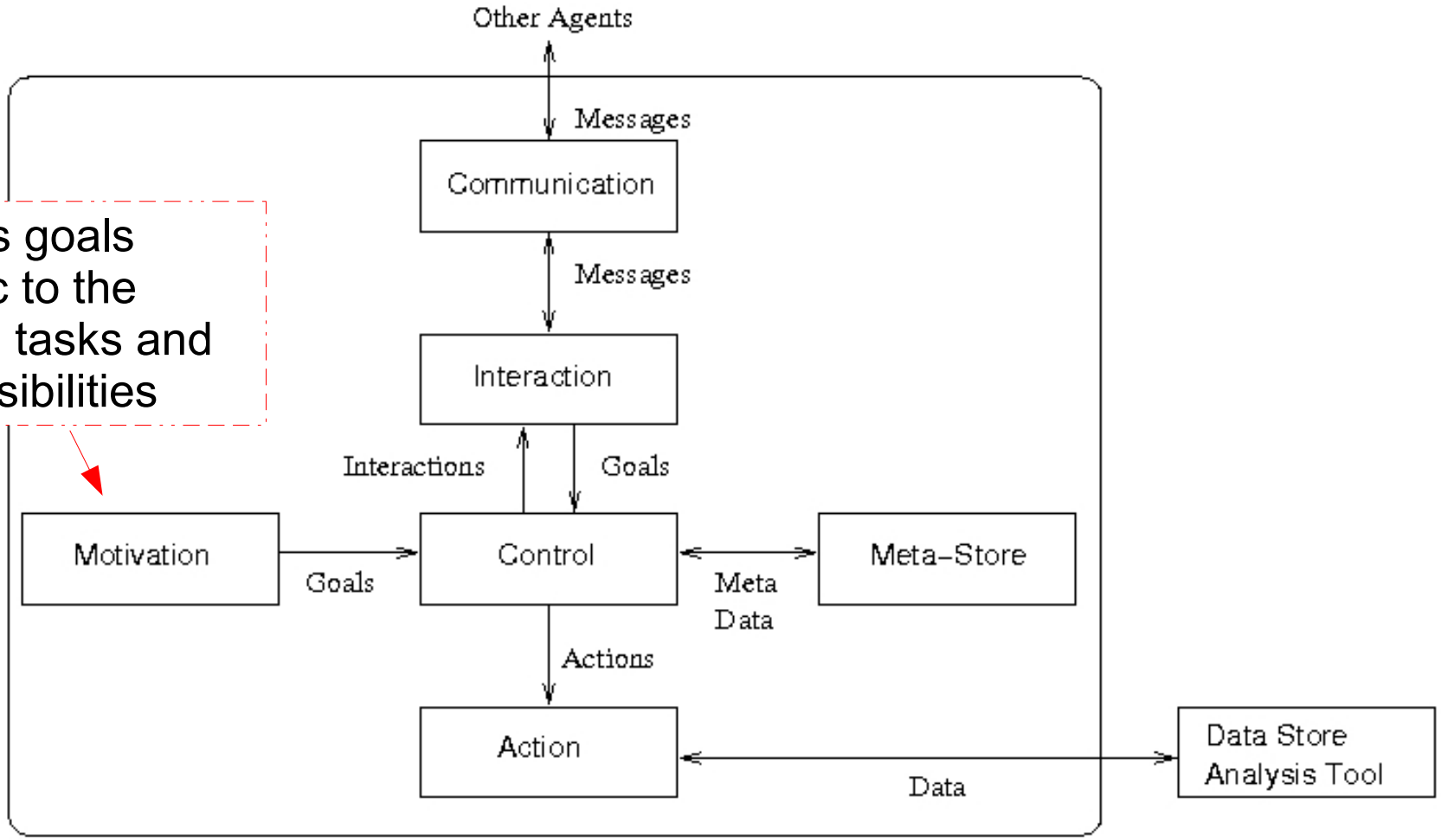
GeneWeaver

- Macro-architecture



GeneWeaver

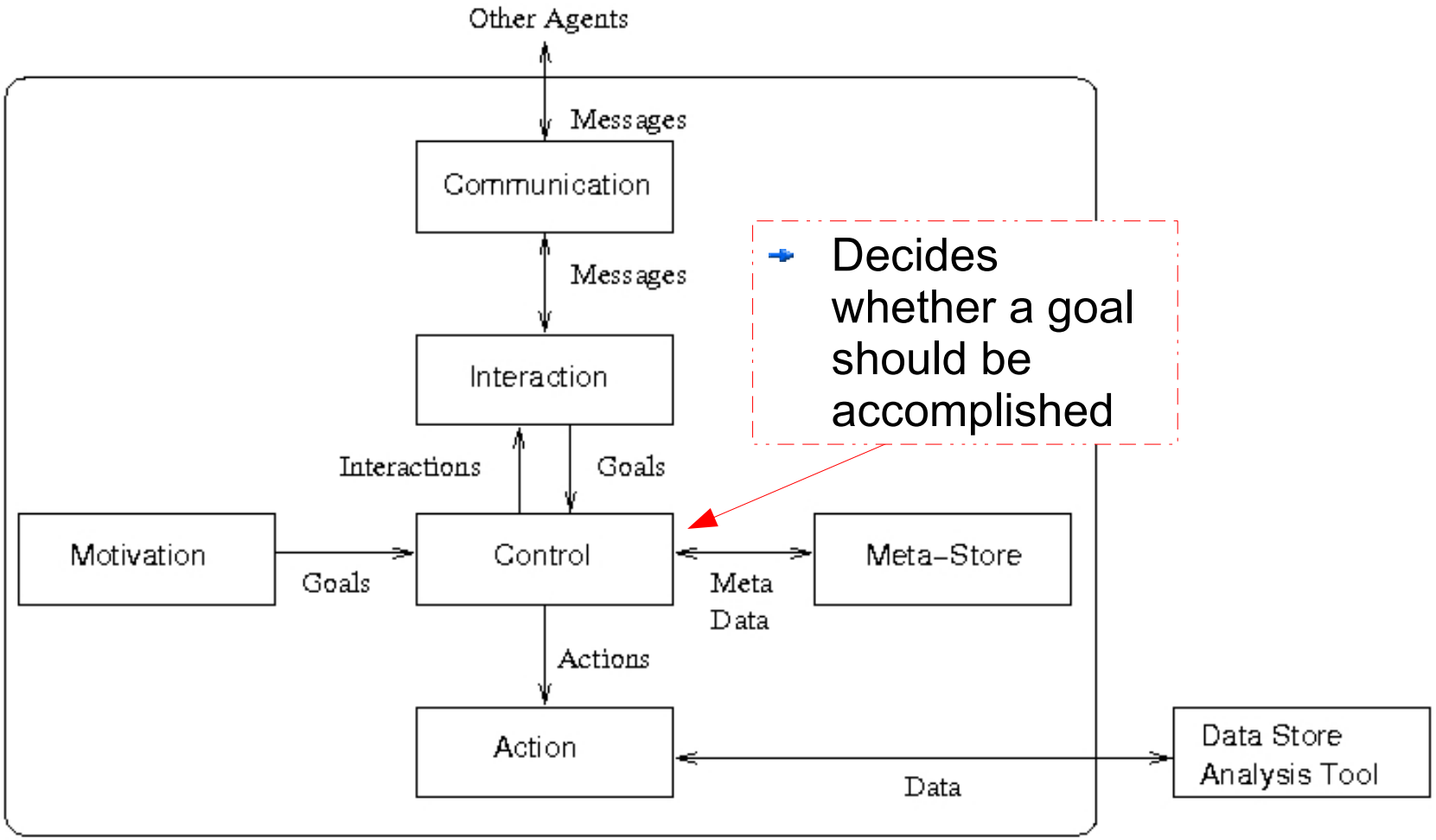
- Micro-architecture



- Causes goals specific to the agent's tasks and responsibilities

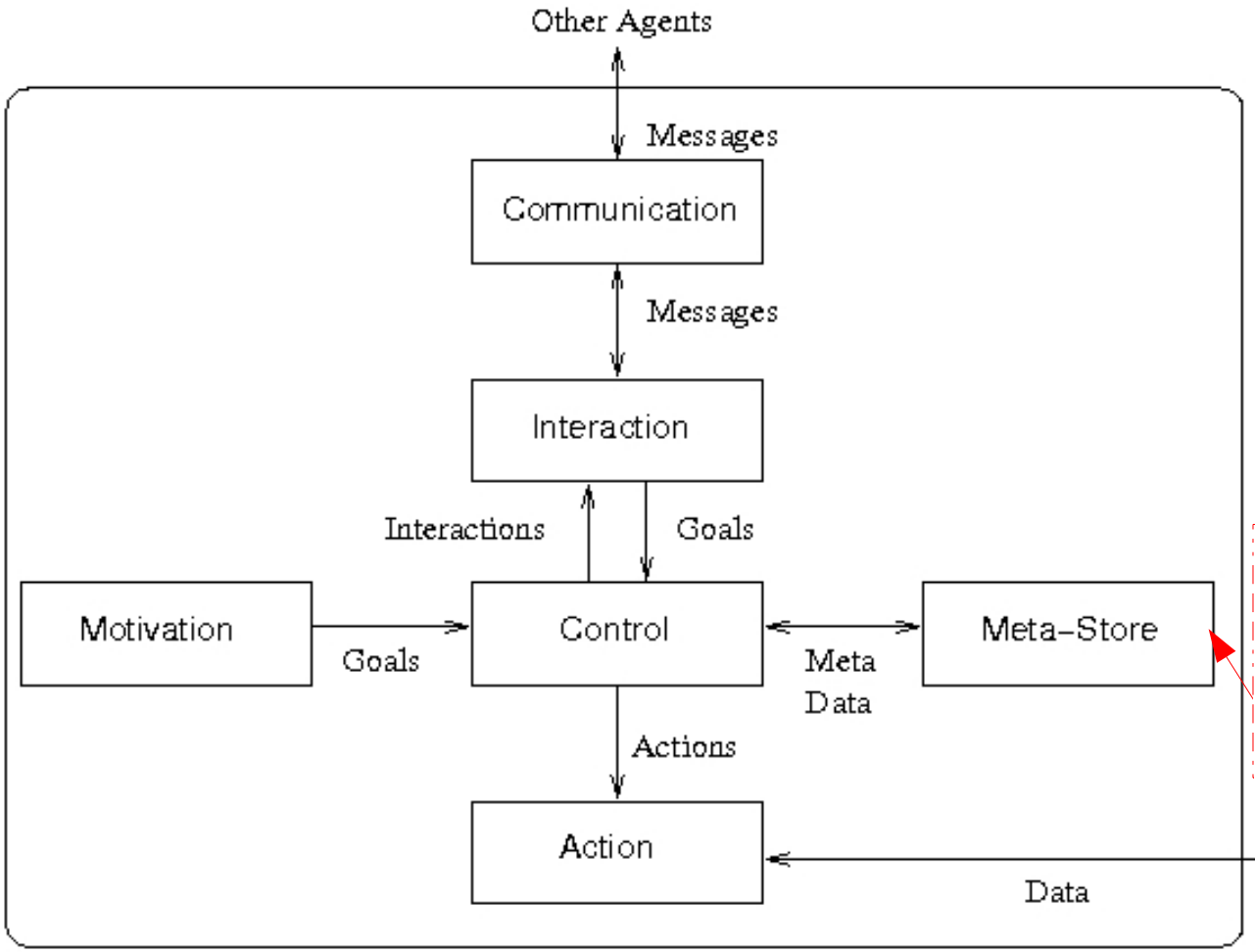
GeneWeaver

- Micro-architecture



GeneWeaver

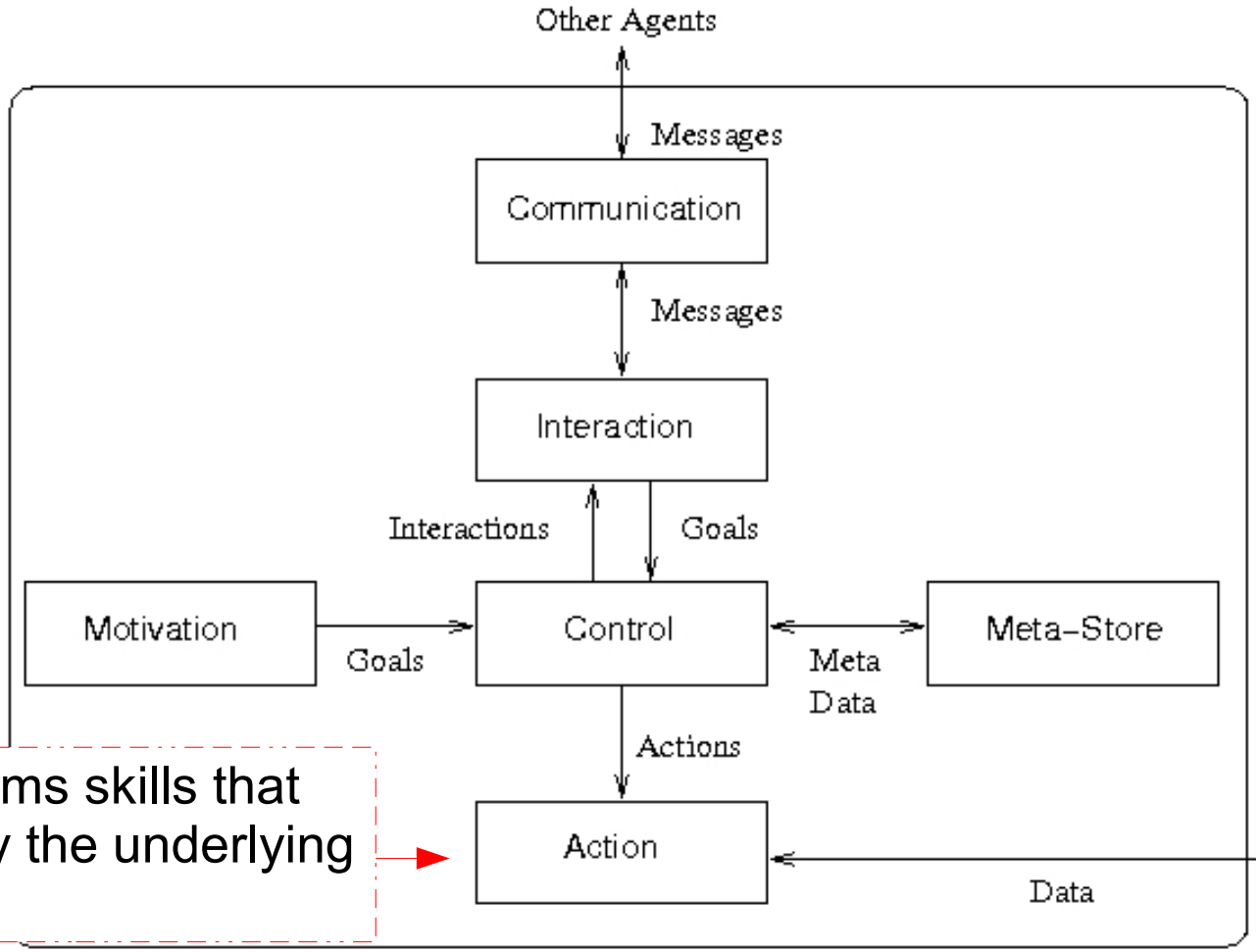
- Micro-architecture



- Provides a repository for the required information

GeneWeaver

- Micro-architecture

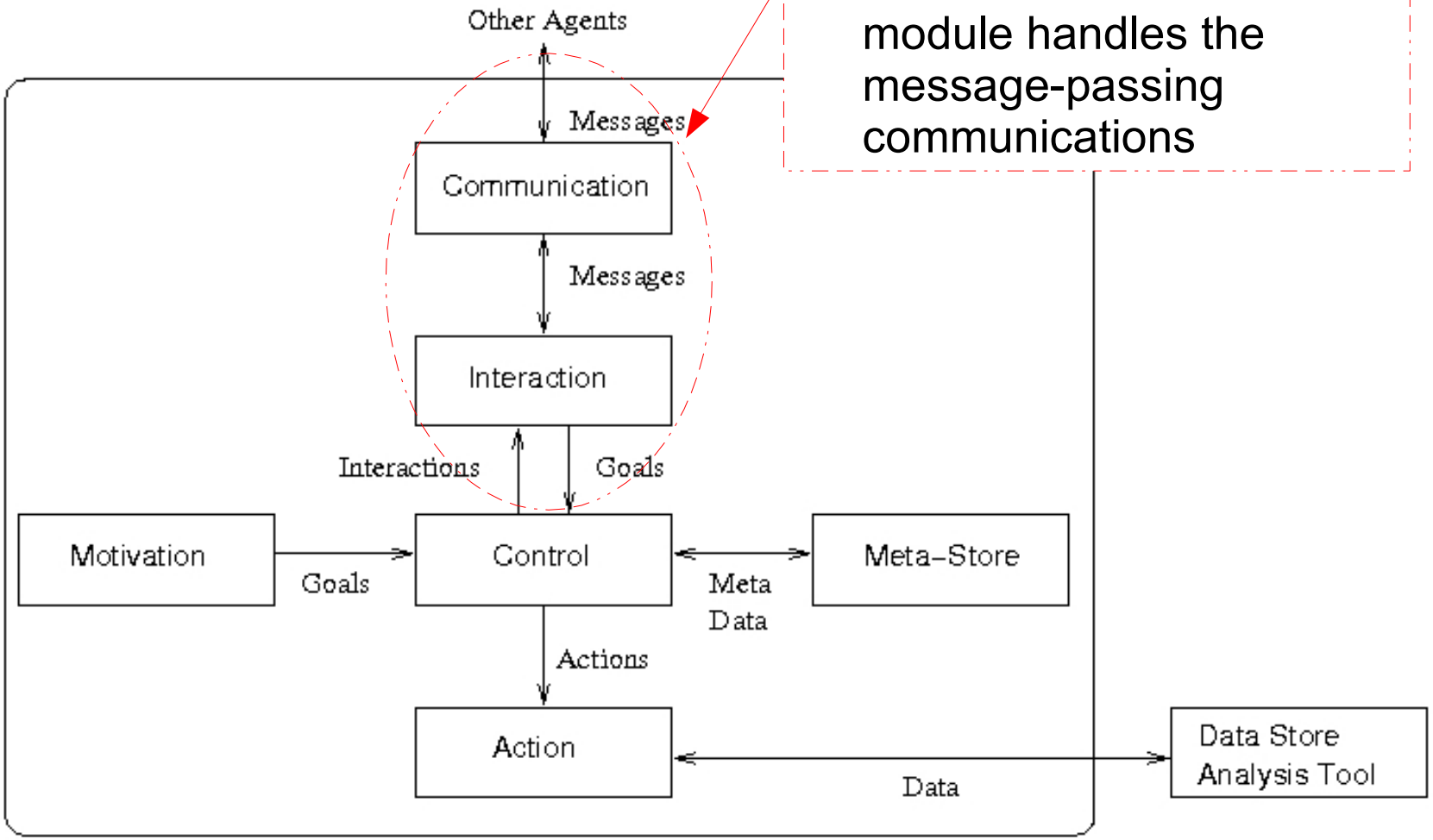


- Performs skills that modify the underlying data

GeneWeaver

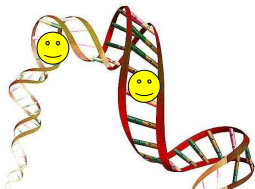
- Micro-architecture

- The interaction module handles the higher level interactions
- The communication module handles the message-passing communications



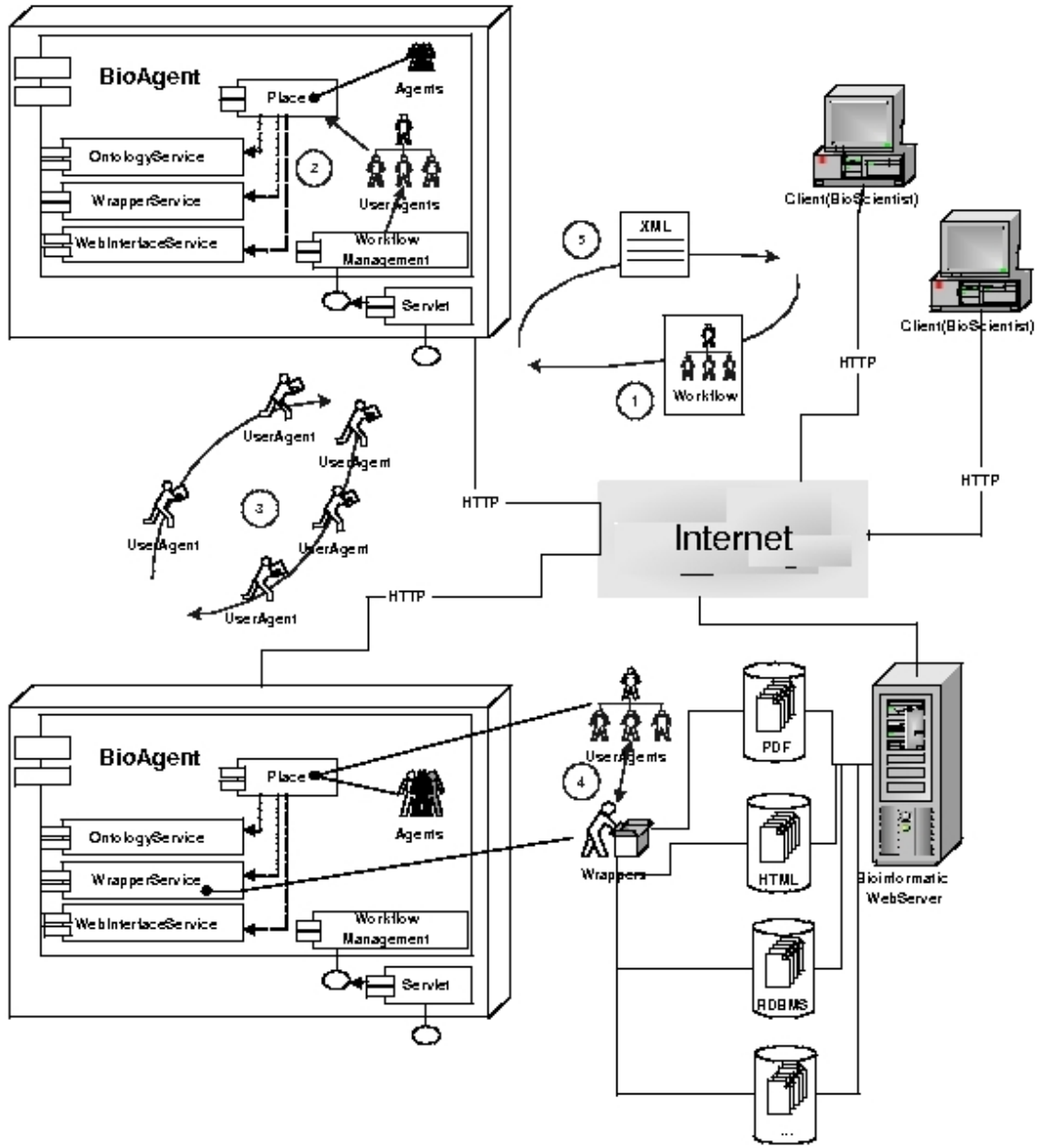
BioAgent

- Is aimed at
  - supporting bioscientists during the genome data analysis
  - decentralizing the coordination of local tasks processing that characterize a workflow of an experiment
  - allowing the use of
    - remote data
    - experiments that contain relevant information



BioAgent

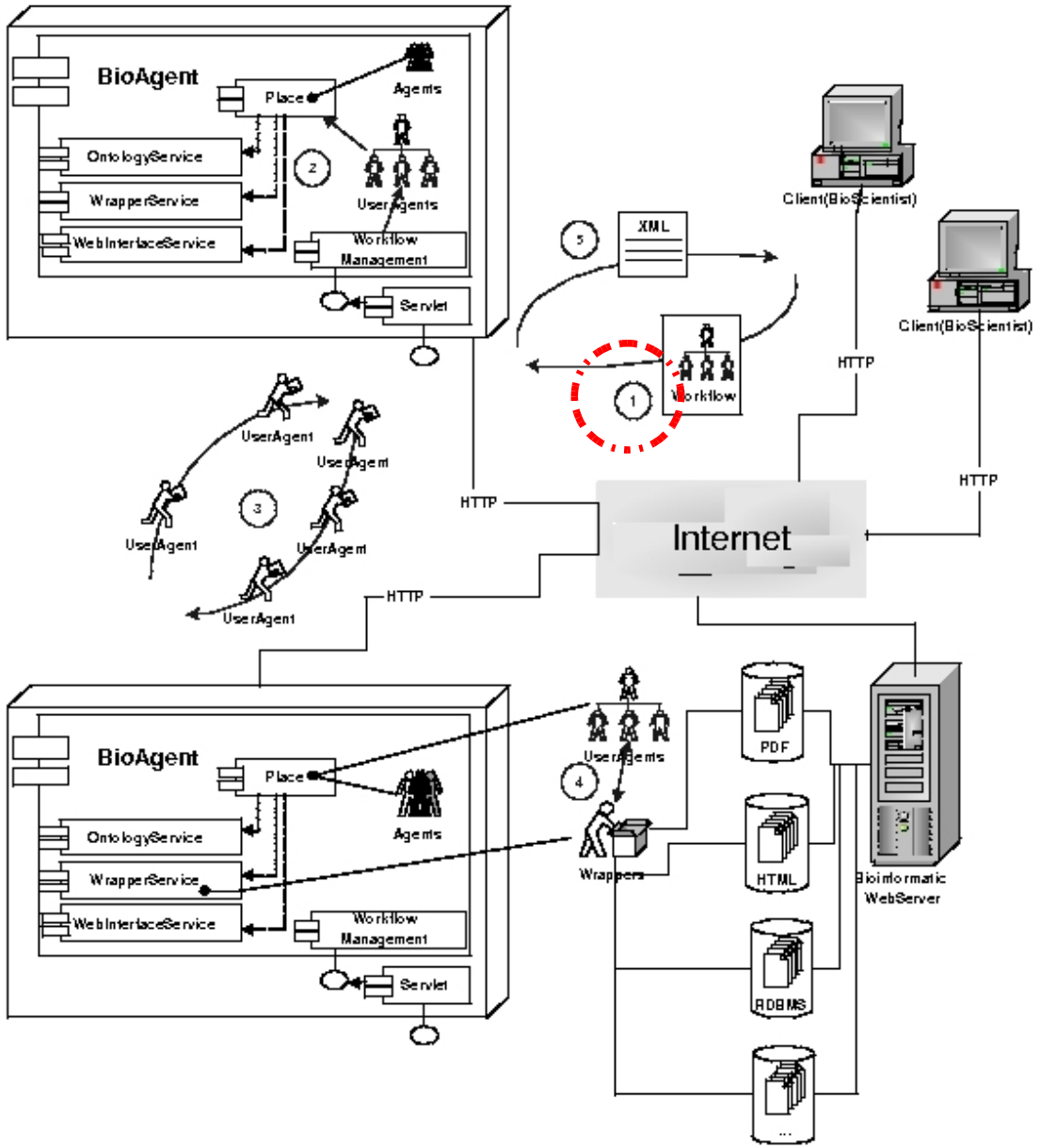
- HERMES **[Corradini05b]** customized for biologists



BioAgent

- HERMES **[Corradini05b]** customized for biologists

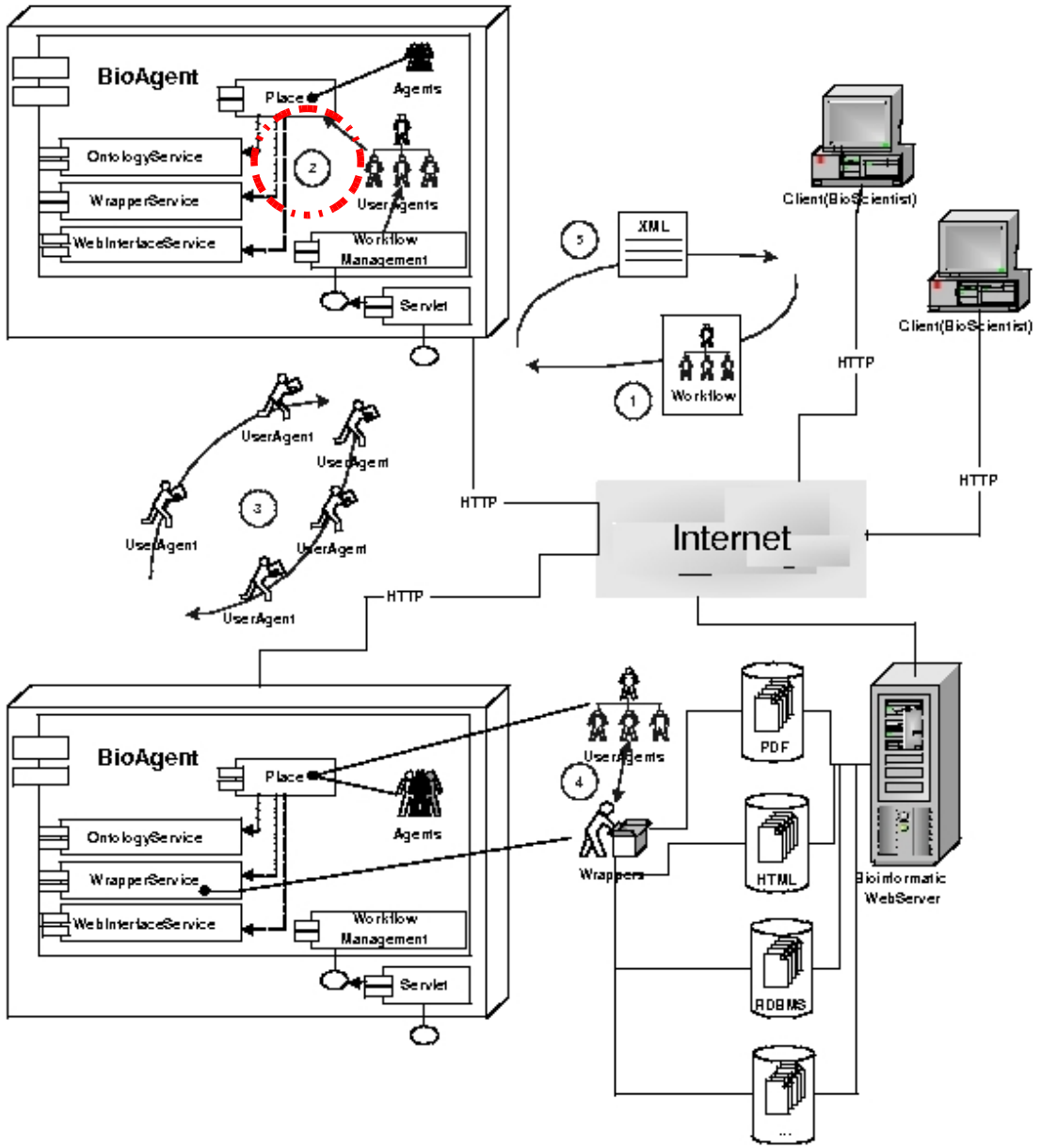
1. A bioscientist specifies the set of activities to be performed



BioAgent

- HERMES **[Corradini05b]** customized for biologists

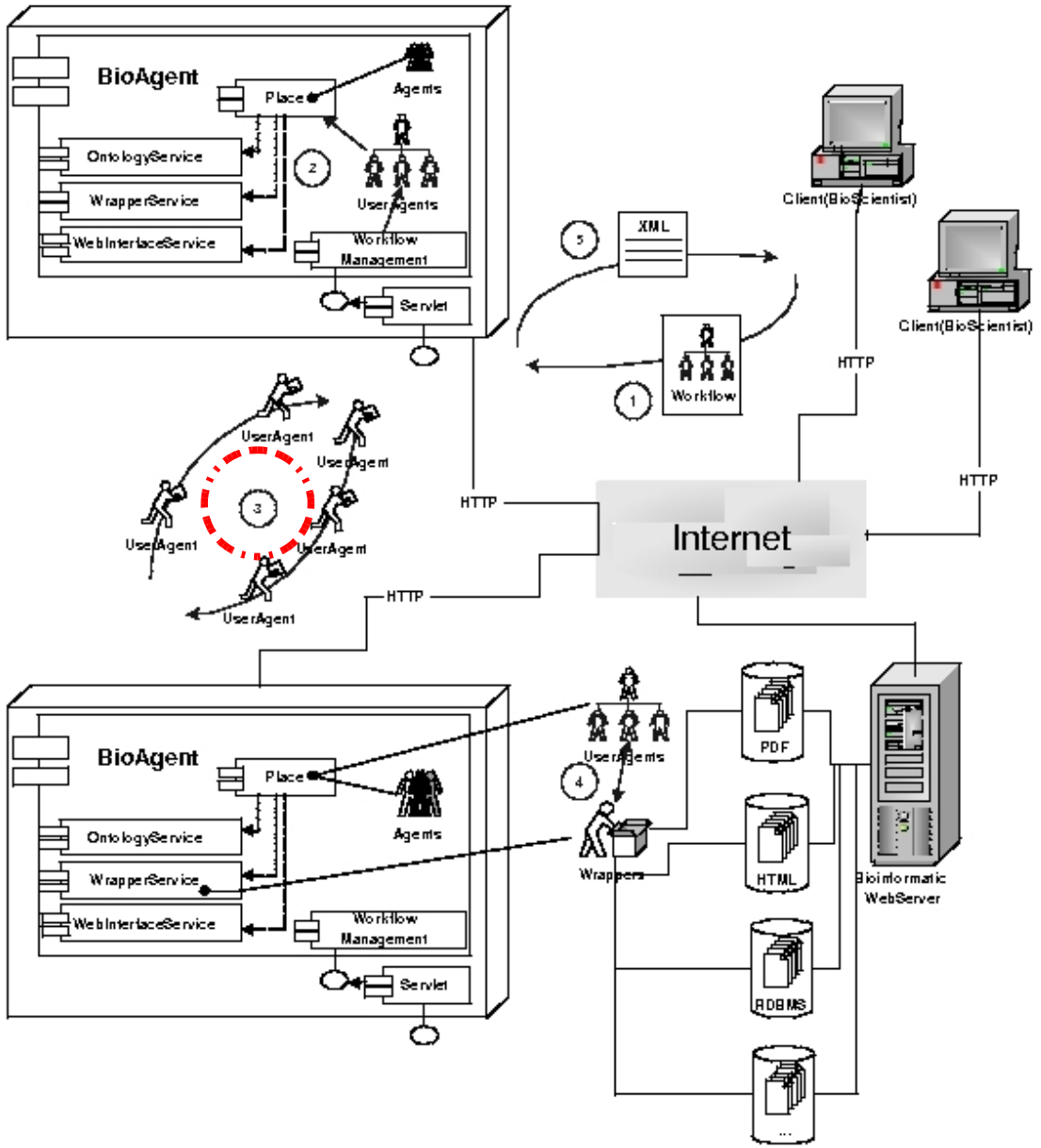
2. The compiler system generates a pool of user agents to execute the activities



BioAgent

- HERMES **[Corradini05b]** customized for biologists

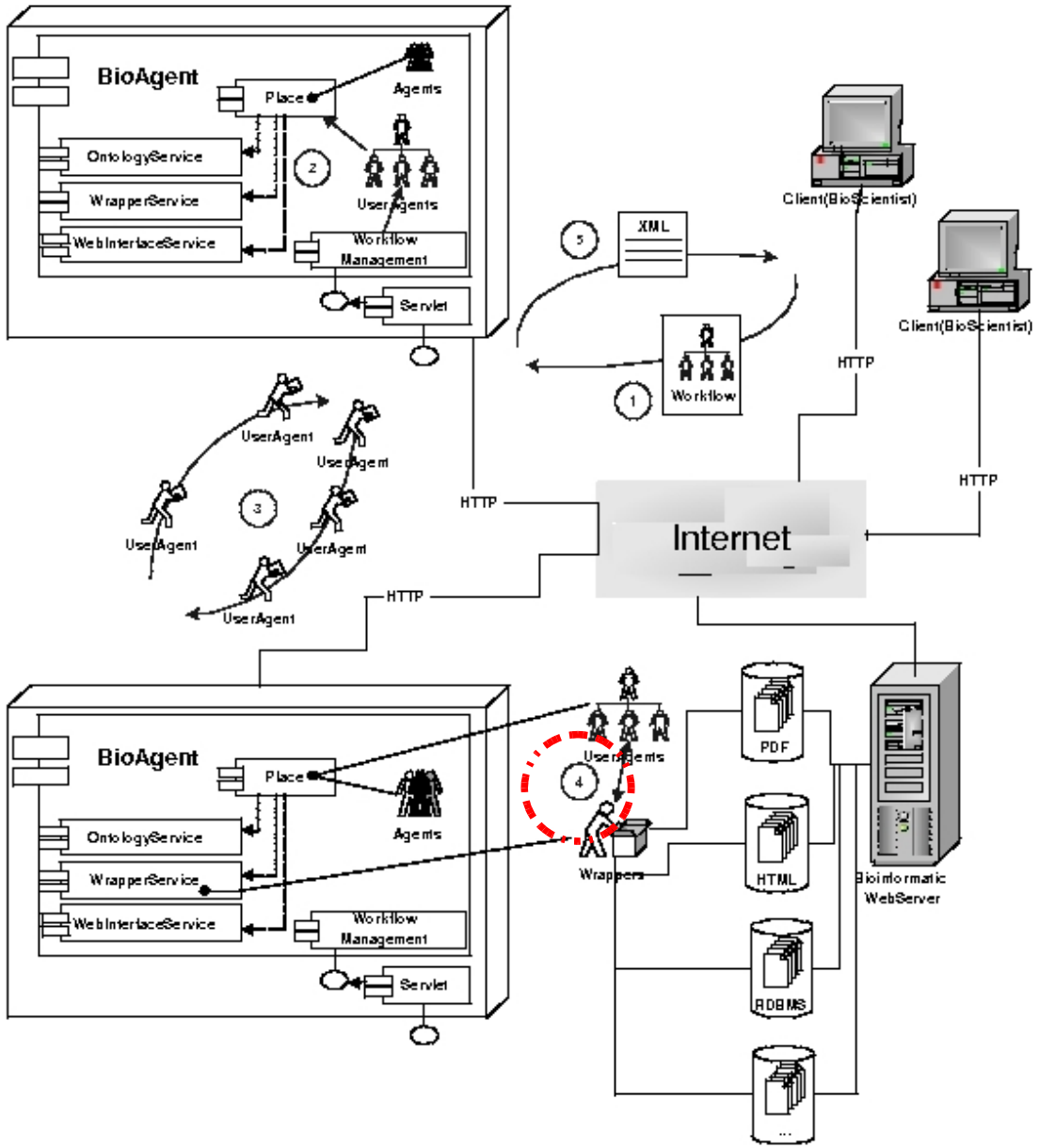
3. User agents migrate and clone in order to efficiently accomplish the activities



BioAgent

- HERMES **[Corradini05b]** customized for biologists

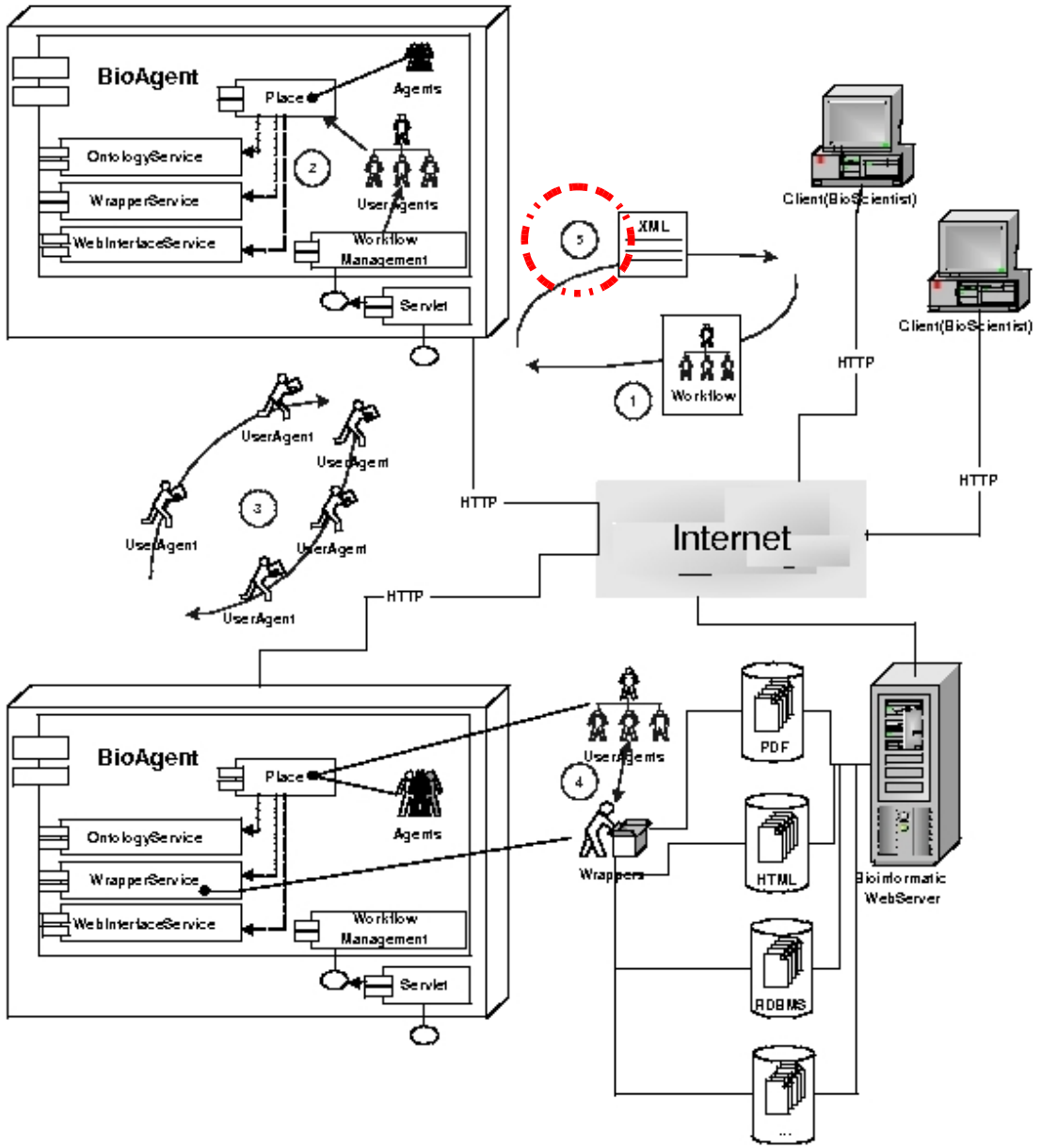
4. Agents query resources by interacting with local service agents. Service agents map the query to local schema



BioAgent

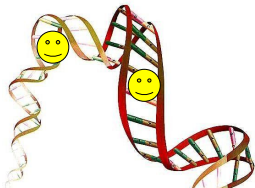
- HERMES **[Corradini05b]** customized for biologists

5. User agents merge results and provide data to the bioscientists

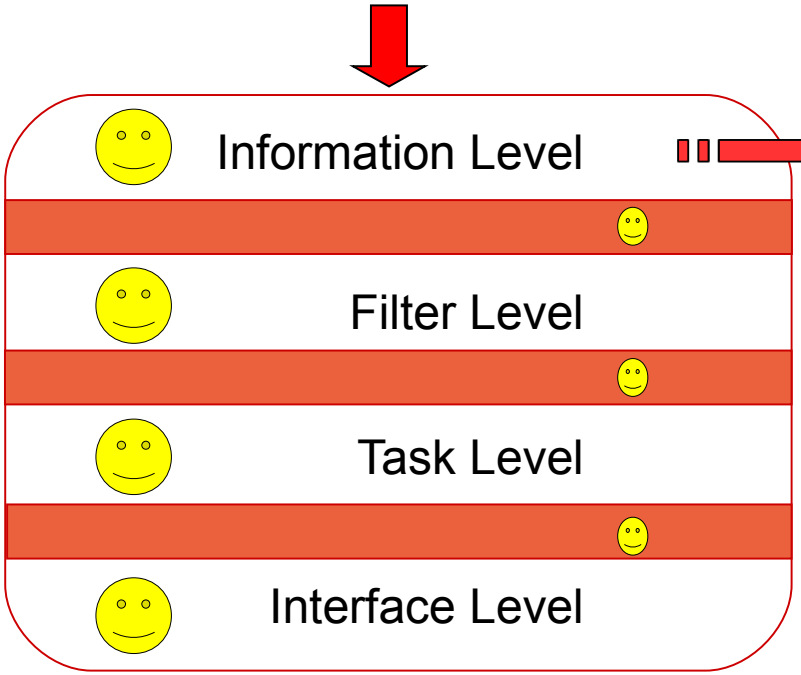


A MultiAgent System for Retrieving Bioinformatics Publications

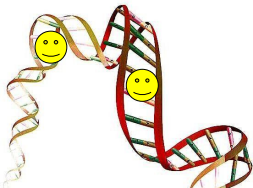
- A generic multiagent architecture has been customized to retrieve scientific publications from the web throughout a text categorization process.
- Experiments have been performed on publications extracted from BMC Bioinformatics and PubMed digital archives.



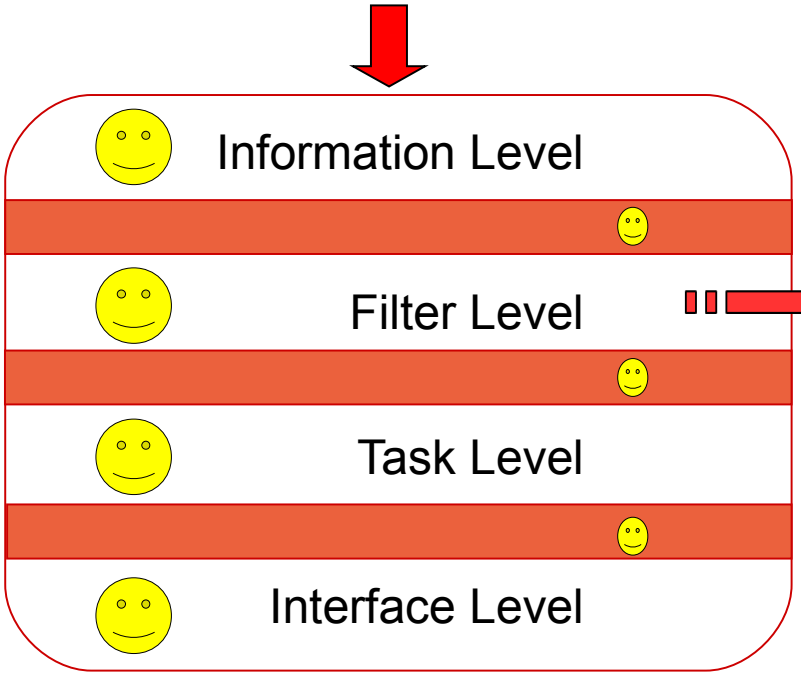
A MultiAgent System for Retrieving Bioinformatics Publications



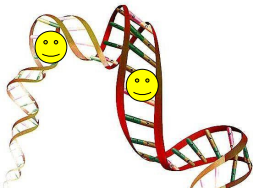
Agents are aimed at extracting data from the information sources



A MultiAgent System for Retrieving Bioinformatics Publications



Agents are aimed at selecting information deemed relevant to the users



A MultiAgent System for Retrieving Bioinformatics Publications

BMC Bioinformatics

PubMed A service of the National Library of Medicine and the National Institutes of Health www.pubmed.gov

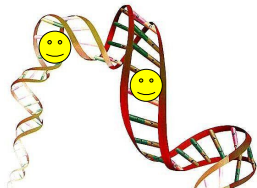
Information Level

Filter Level

Task Level

Interface Level

Agents are aimed at processing data according to users personal needs and preferences



A MultiAgent System for Retrieving Bioinformatics Publications

BMC Bioinformatics

PubMed A service of the National Library of Medicine and the National Institutes of Health www.pubmed.gov

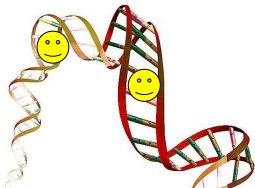
Information Level

Filter Level

Task Level

Interface Level

Agents are aimed at interacting with the users according to their preferences



A MultiAgent System for Retrieving Bioinformatics Publications

BMC Bioinformatics

PubMed A service of the National Library of Medicine and the National Institutes of Health www.pubmed.gov

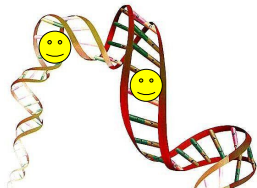
Information Level

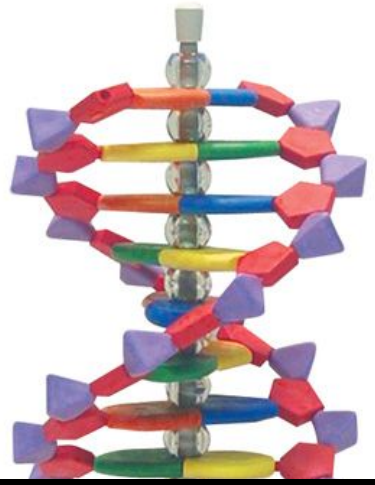
Filter Level

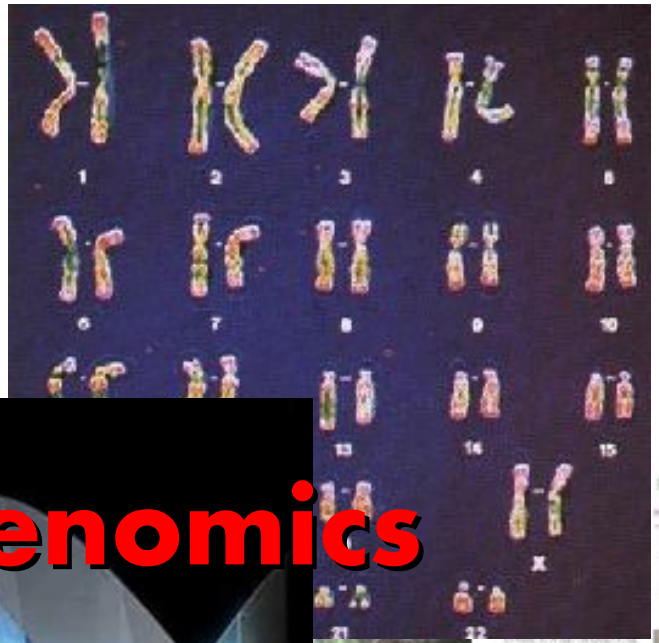
Task Level

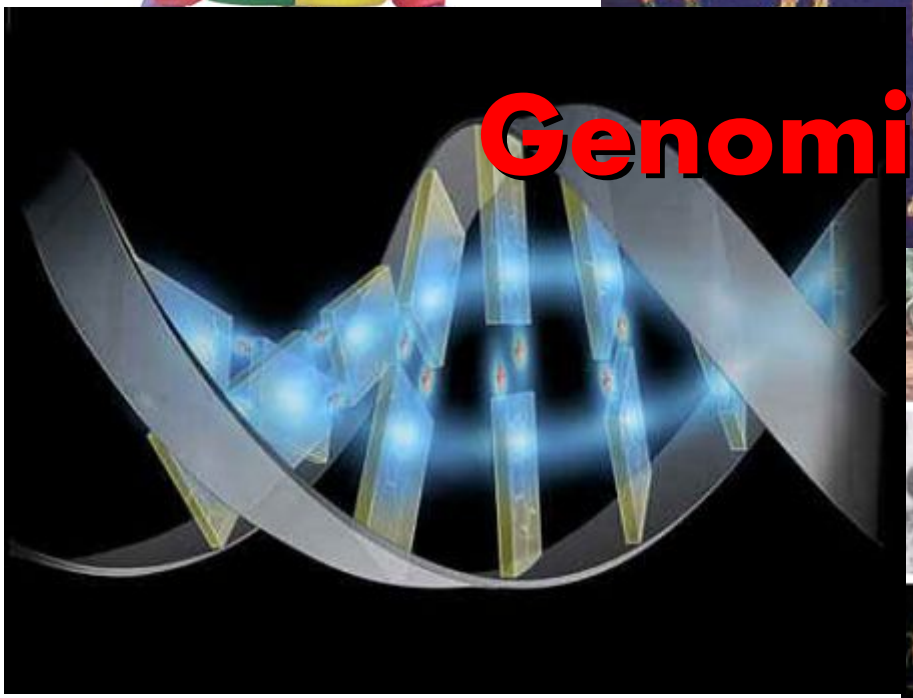
Interface Level

Agents are aimed at at establishing communication among requesters and providers

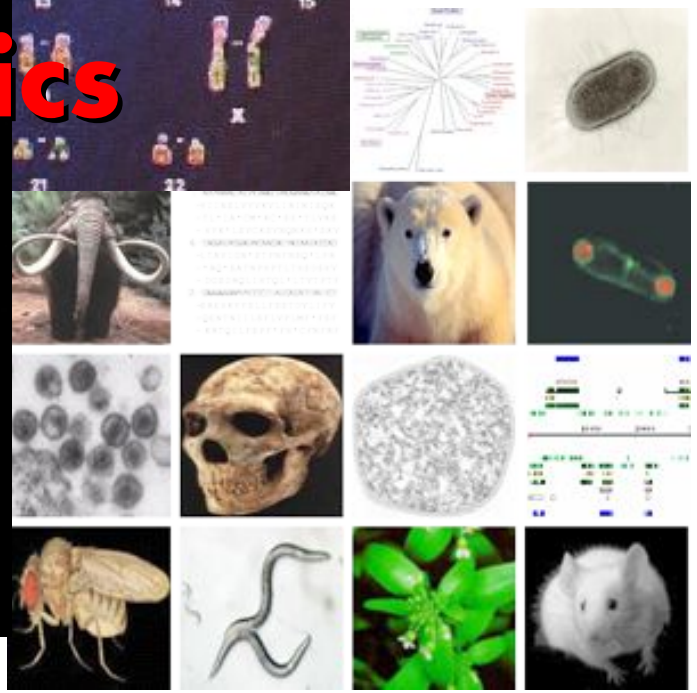


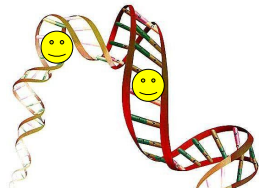






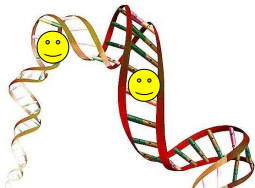
Genomics

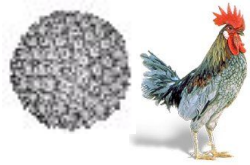




Genomics

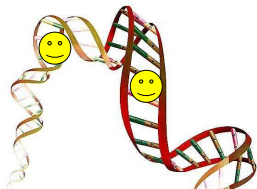
- Genome annotation
  - BioMAS **[Decker02]**





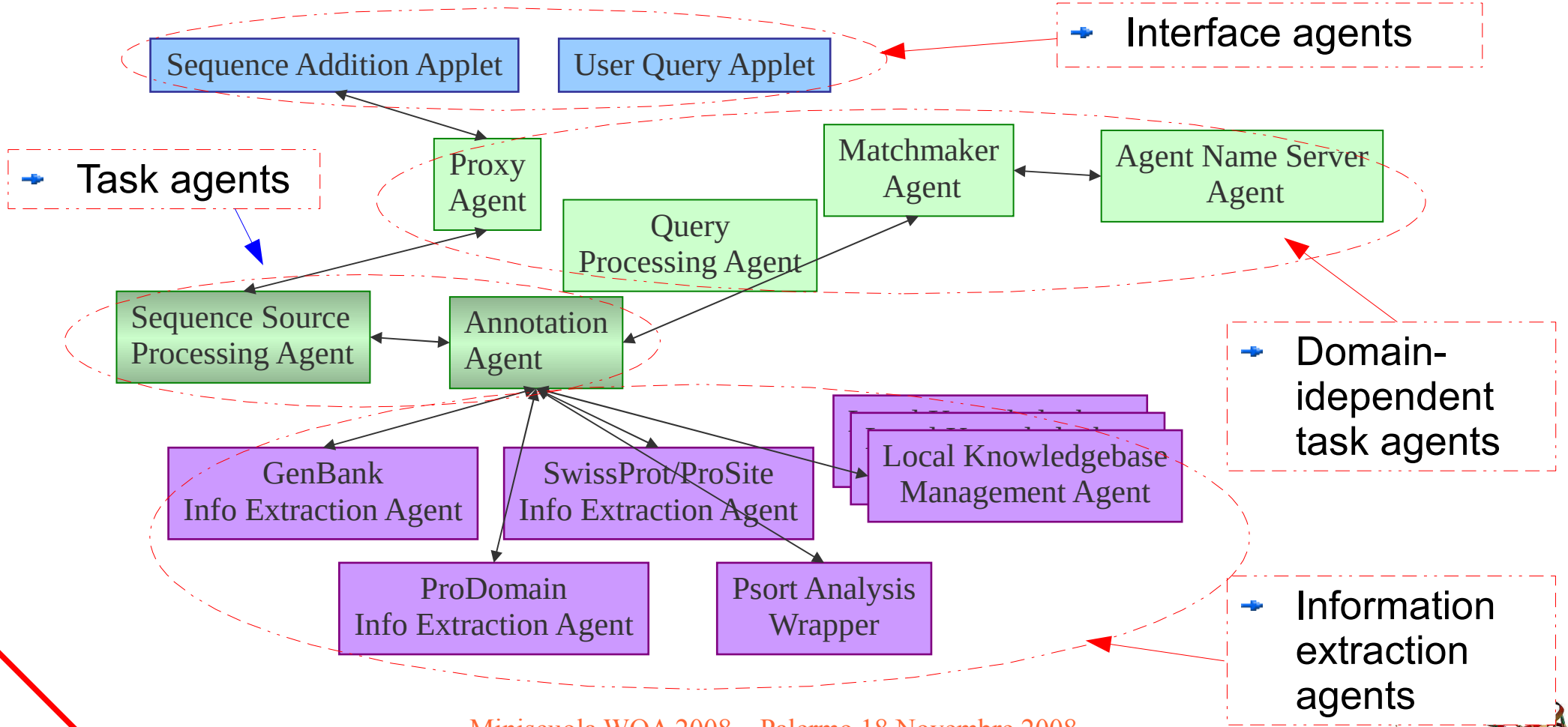
BioMAS

- Automates collection of information from various primary source databases
- Allows various analysis to be done automatically
- New data sources, annotation, analysis can be applied as they are developed, automatically (open system)
- Much more sophisticated queries (across multiple databases) than keyword search



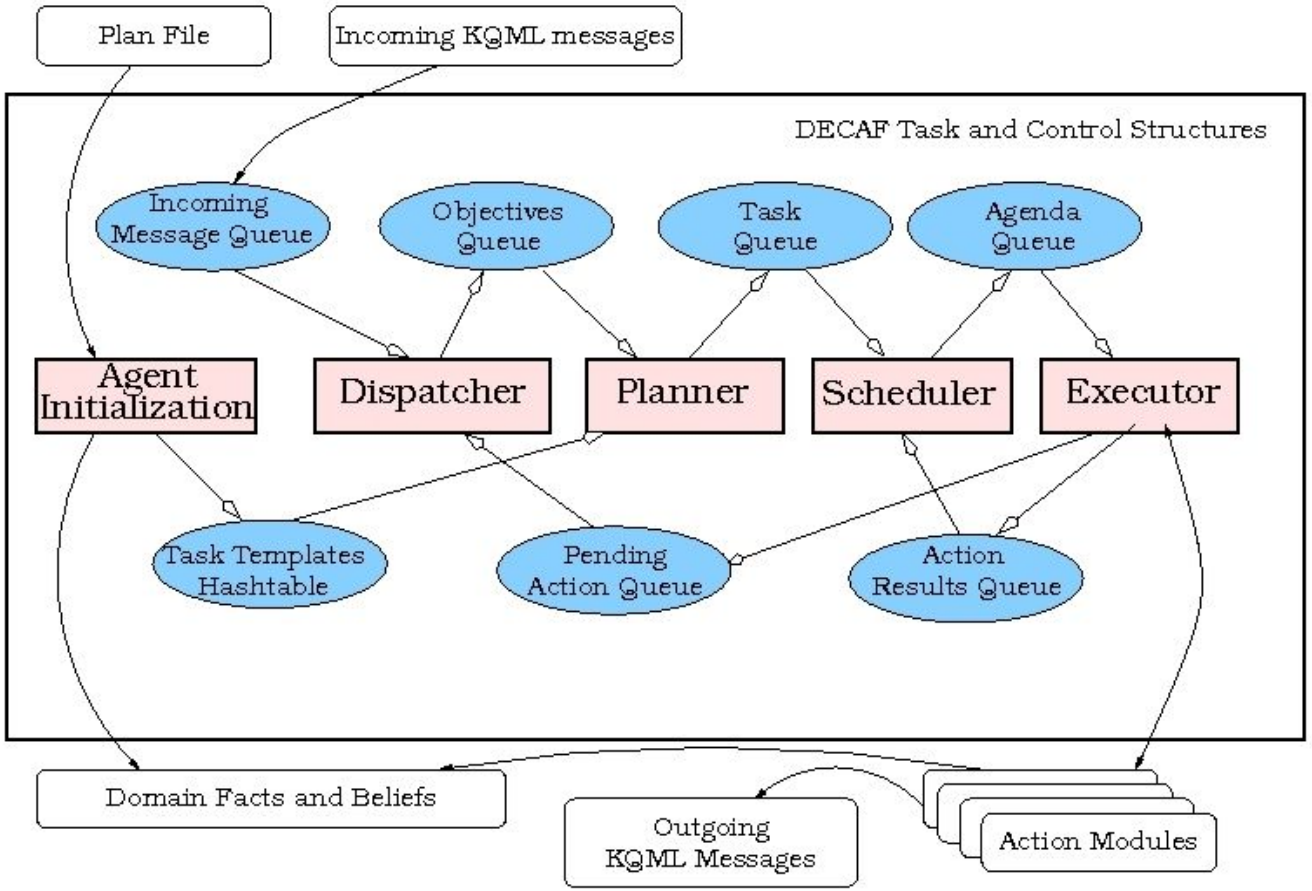
BioMAS

- RETSINA **[Decker97]** model of information gathering



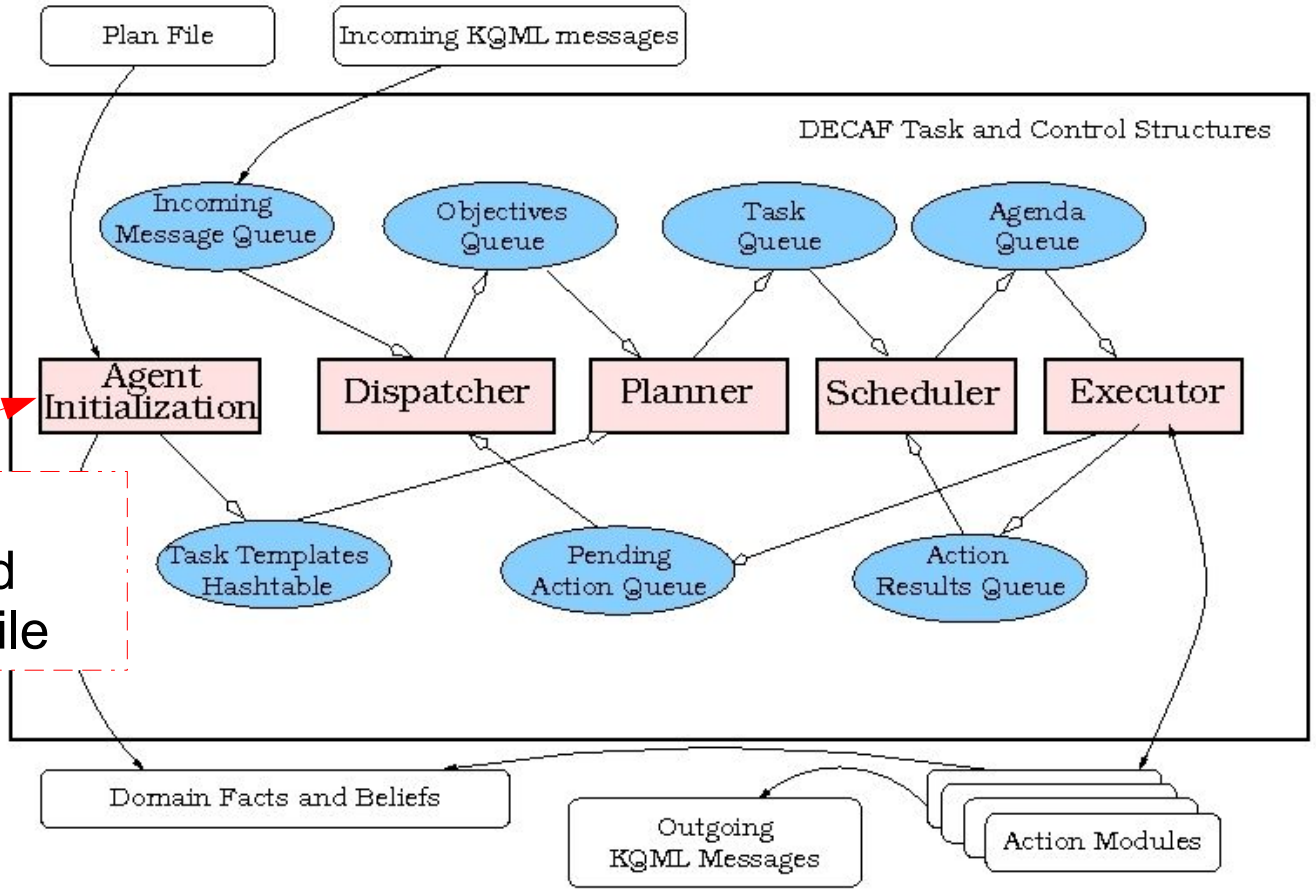
BioMAS

- DECAF **[Graham00]** architecture



BioMAS

- DECAF architecture

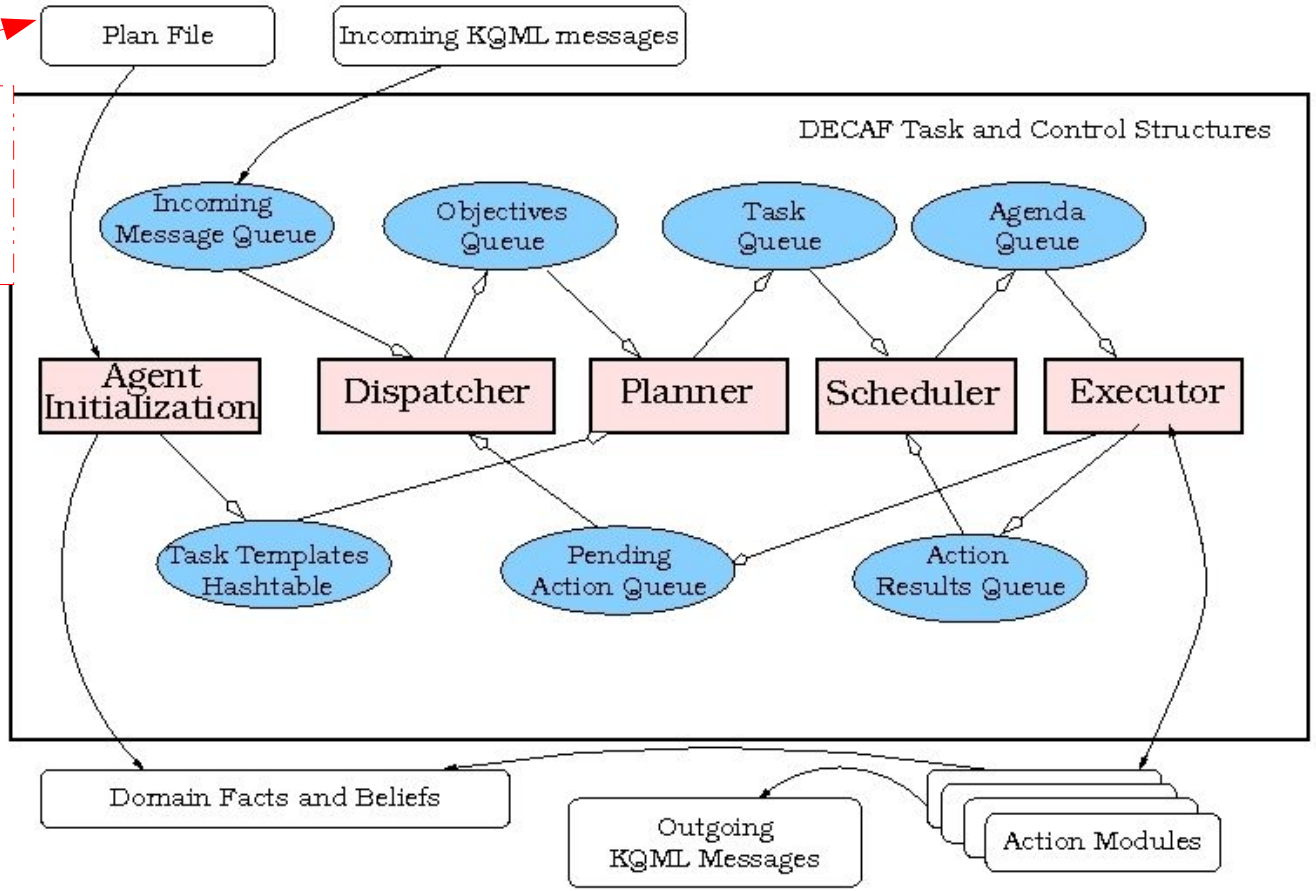


- Runs when an agent is started
- Reads a plan file

BioMAS

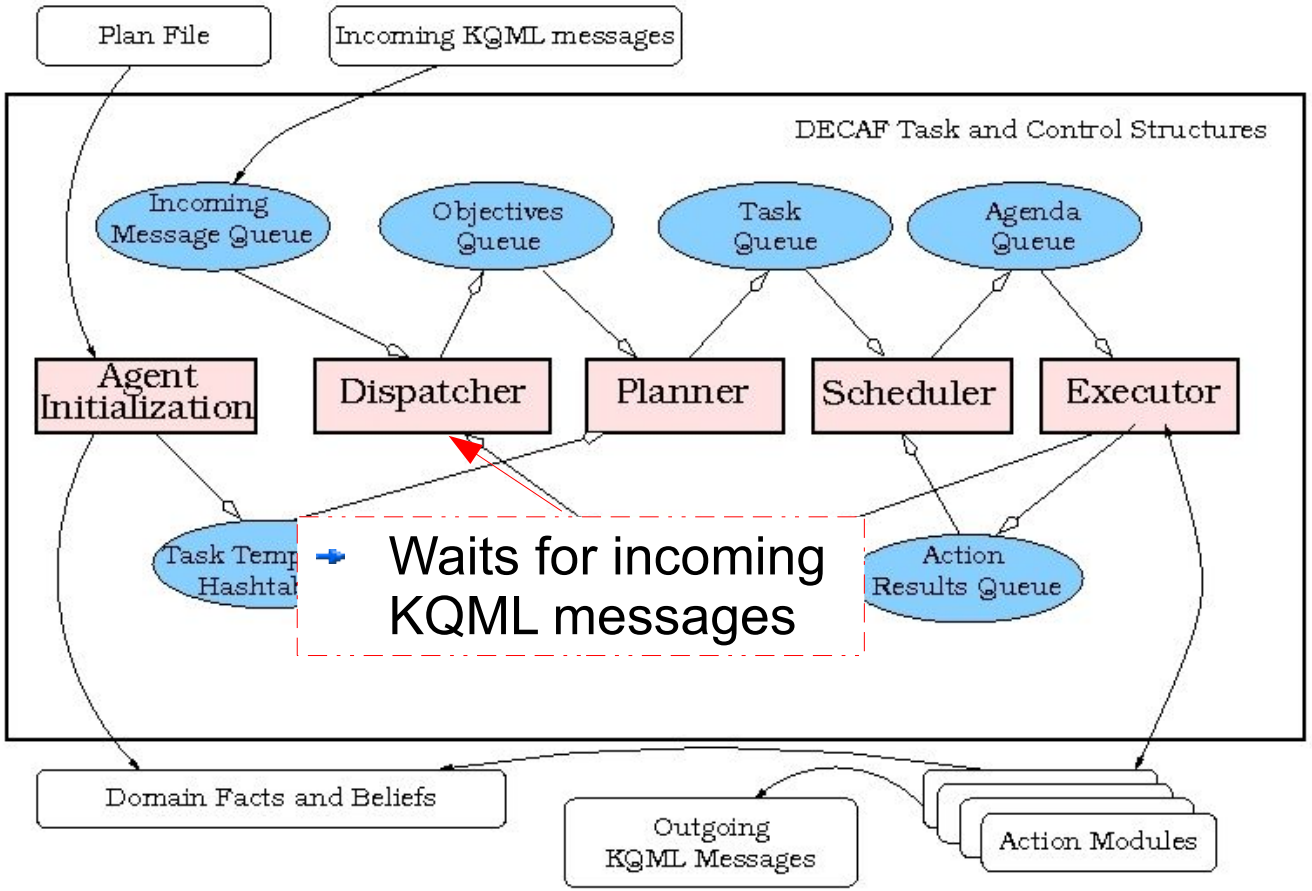
- DECAF architecture

- Describes the agents capabilities



BioMAS

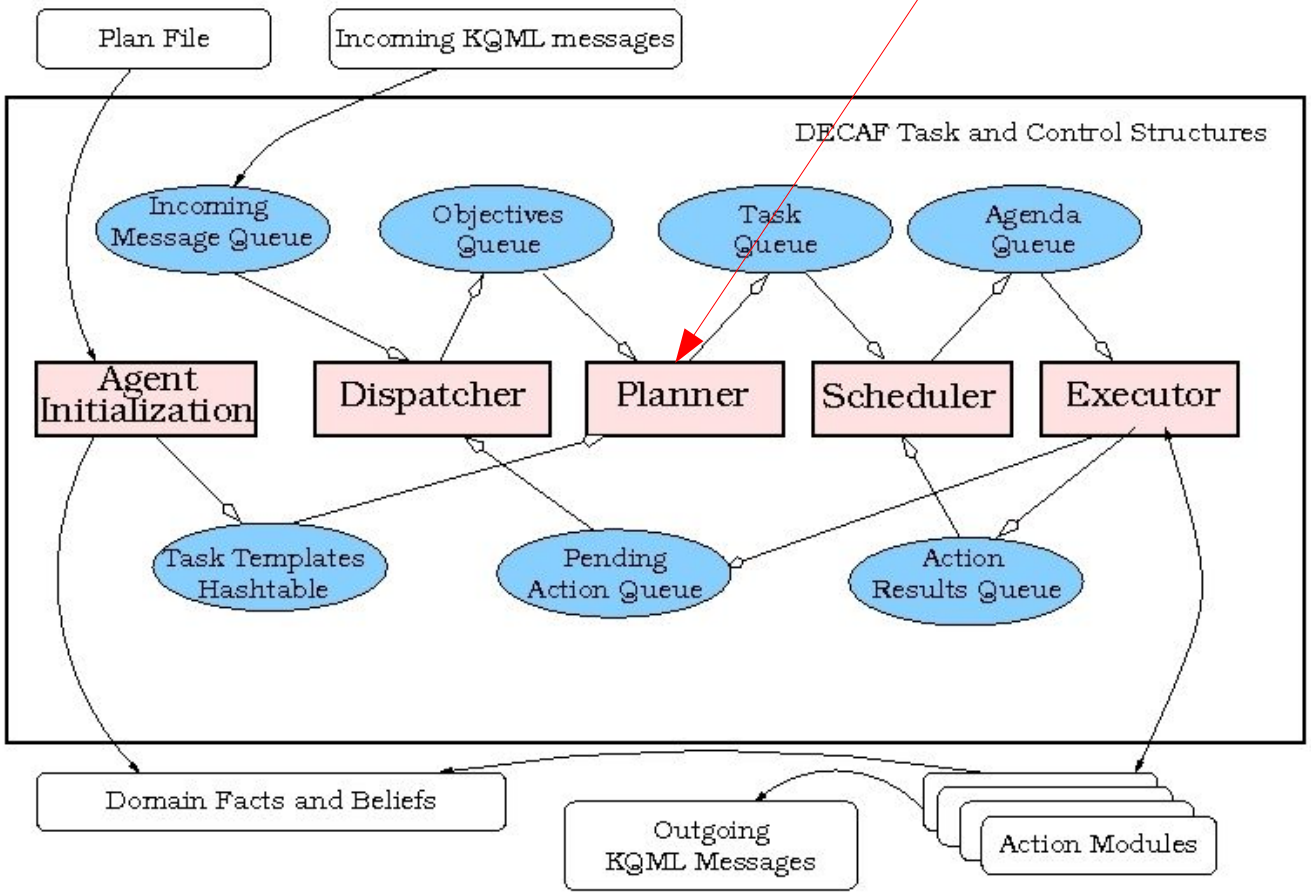
- DECAF architecture



BioMAS

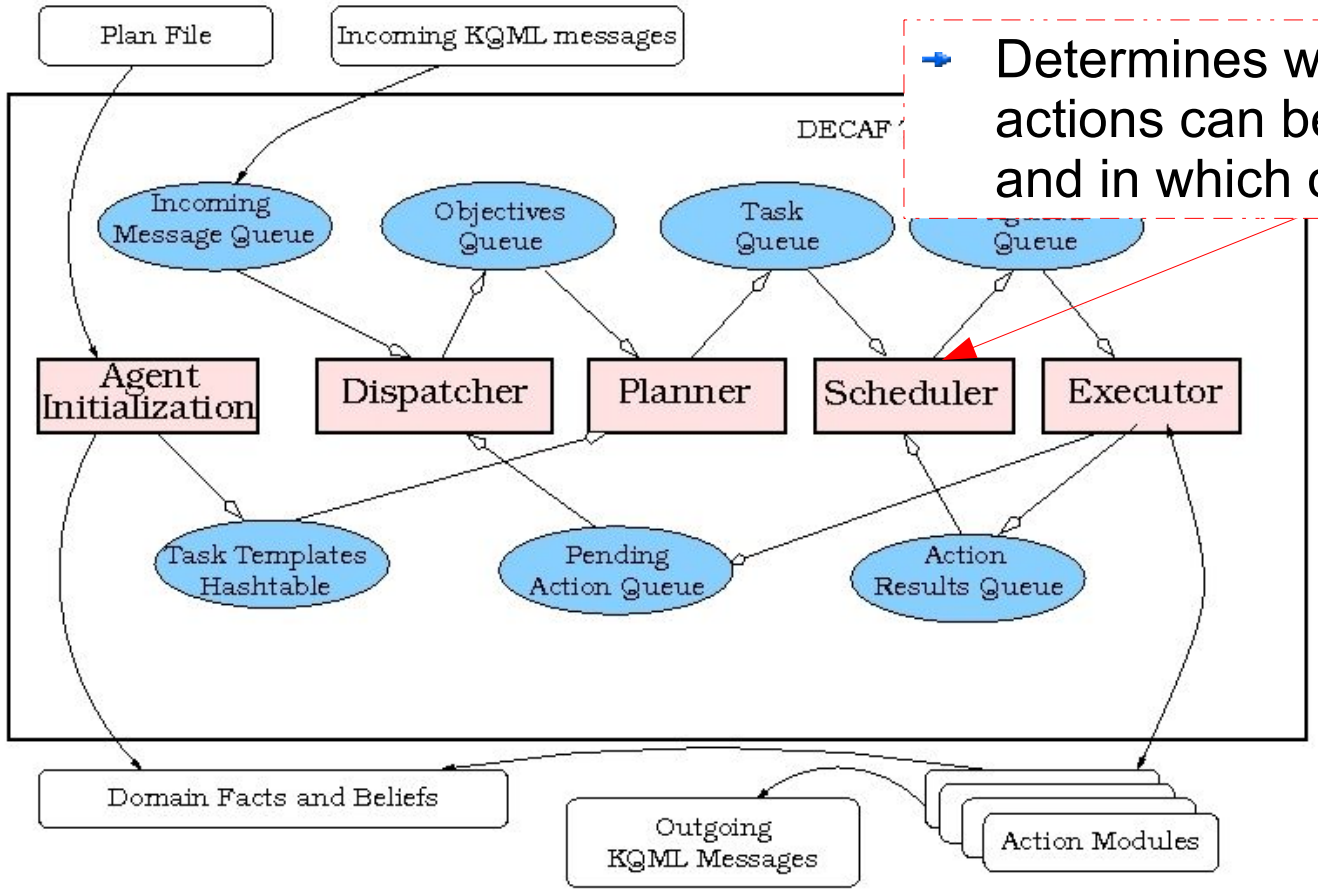
- DECAF architectur

- Monitors the Objectives Queue
- Matches new goals to an existing task template



BioMAS

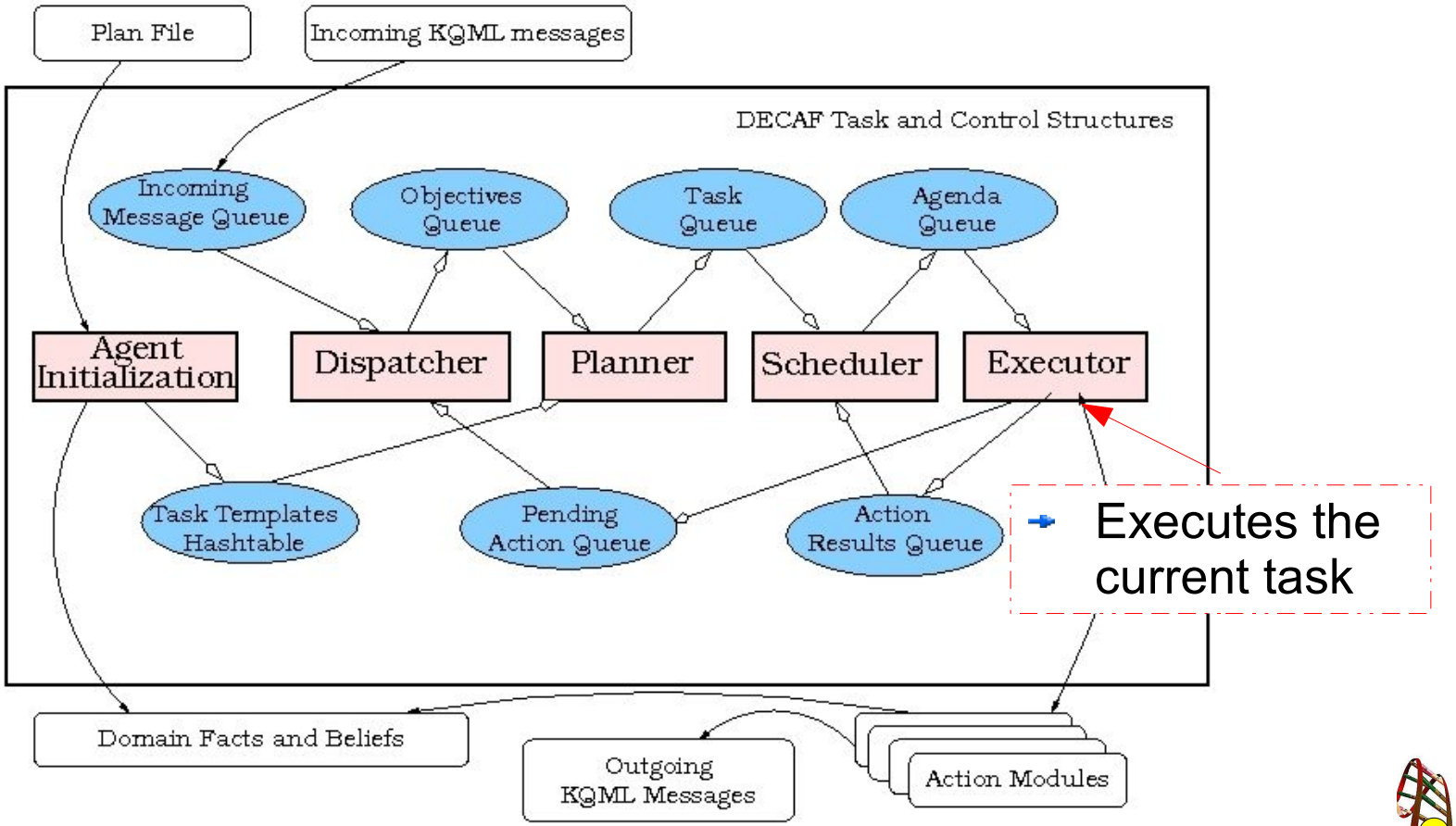
- DECAF architecture



- Determines which actions can be executed and in which order

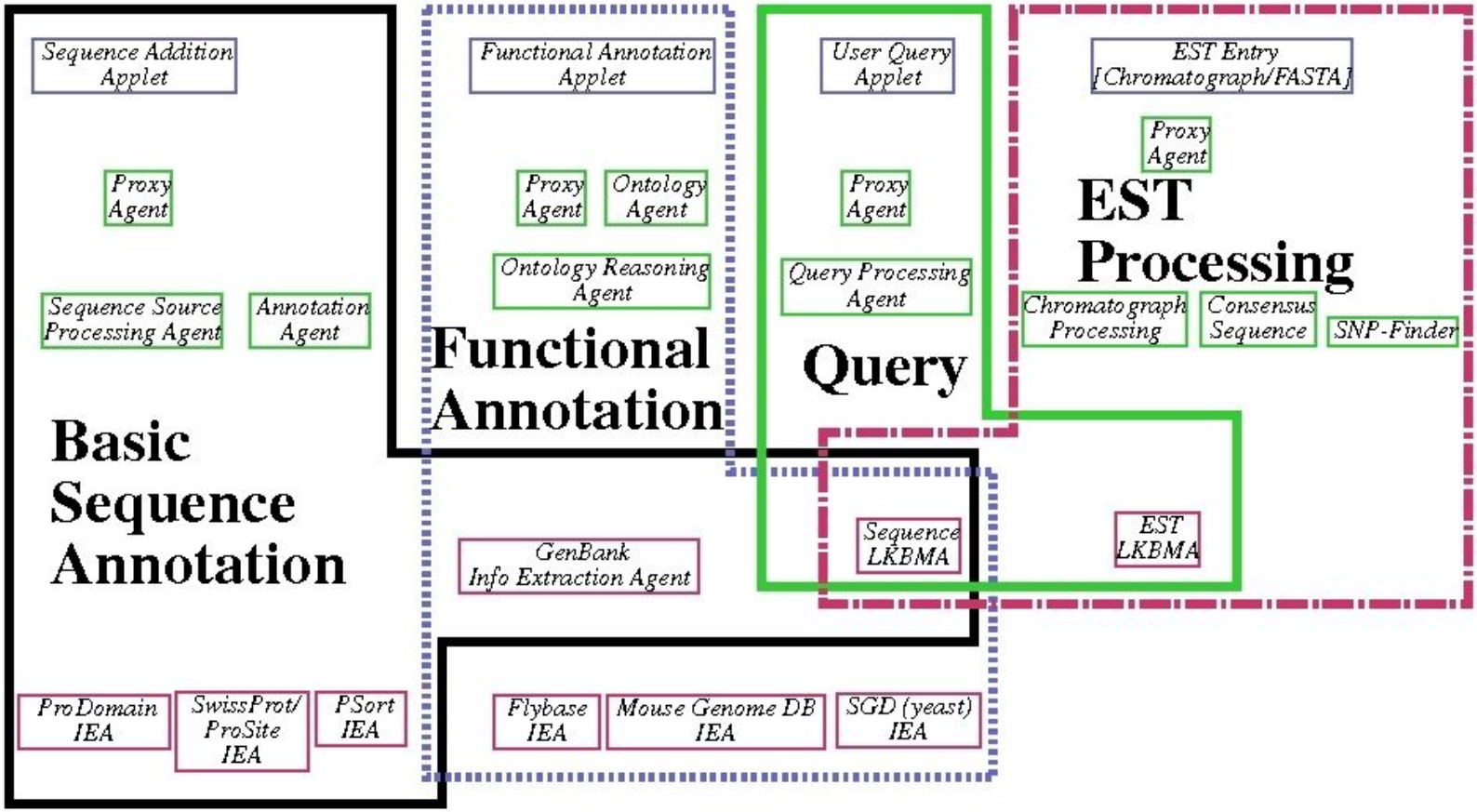
BioMAS

- DECAF architecture



BioMAS

- A multi-agent system for genomic analysis

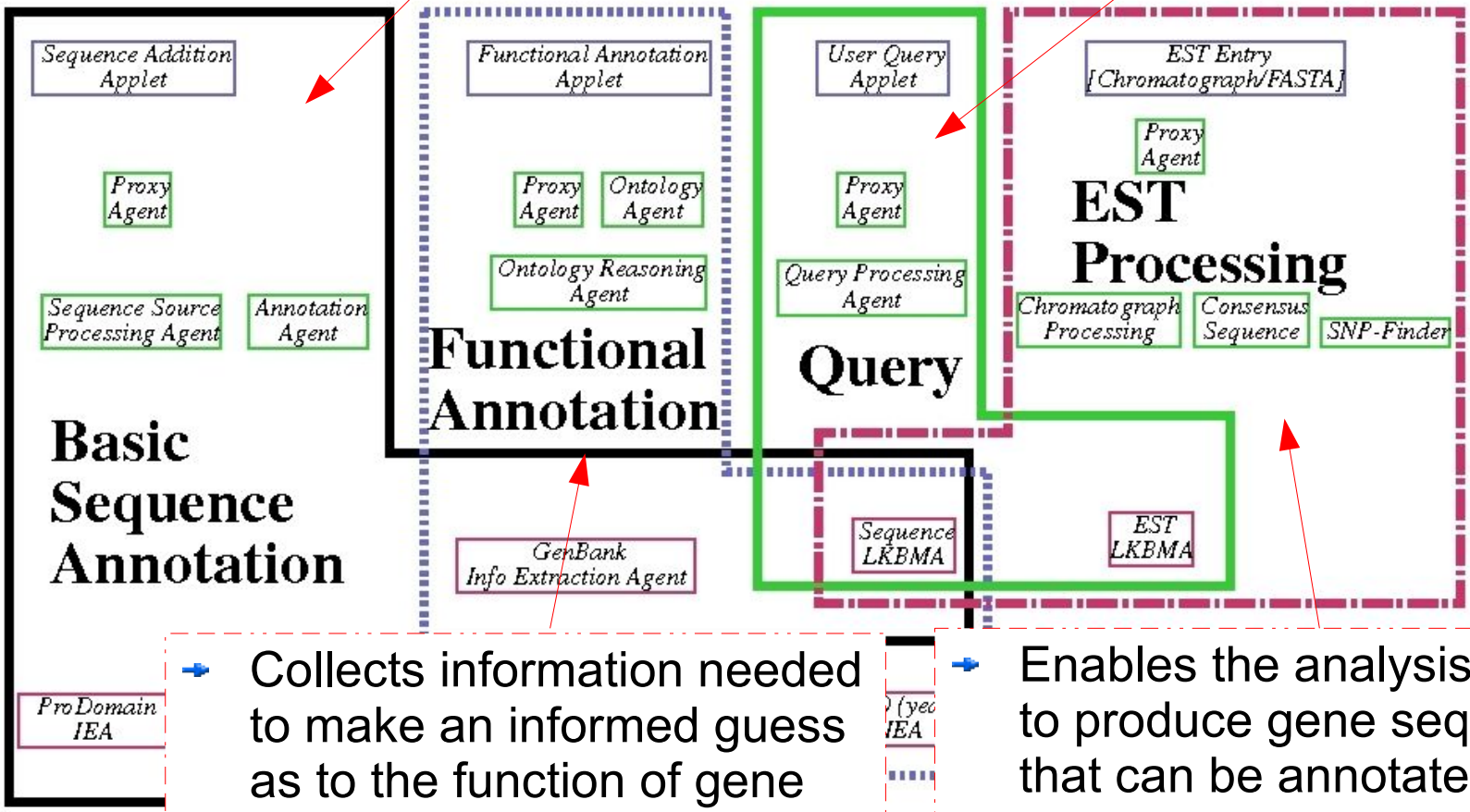


BioMAS

- A mul

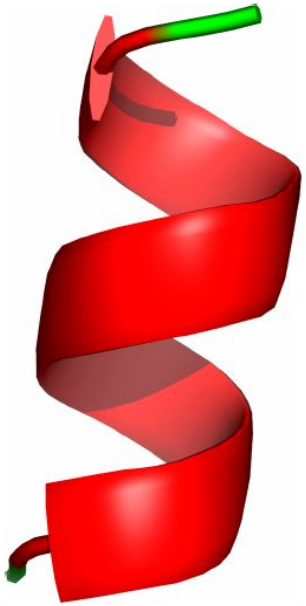
- Integrates remote gene sequence annotation from various sources with the gene sequences at the LKBMA

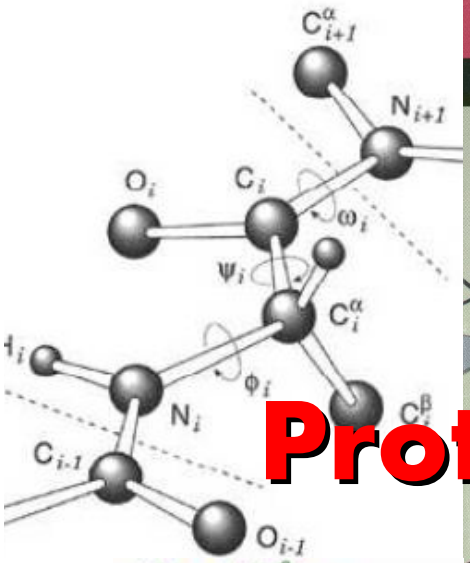
- Allows complex queries on the LKBMA via a web interface

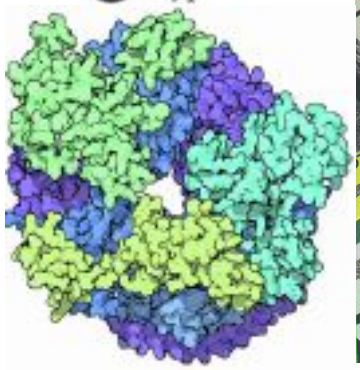


- Collects information needed to make an informed guess as to the function of gene

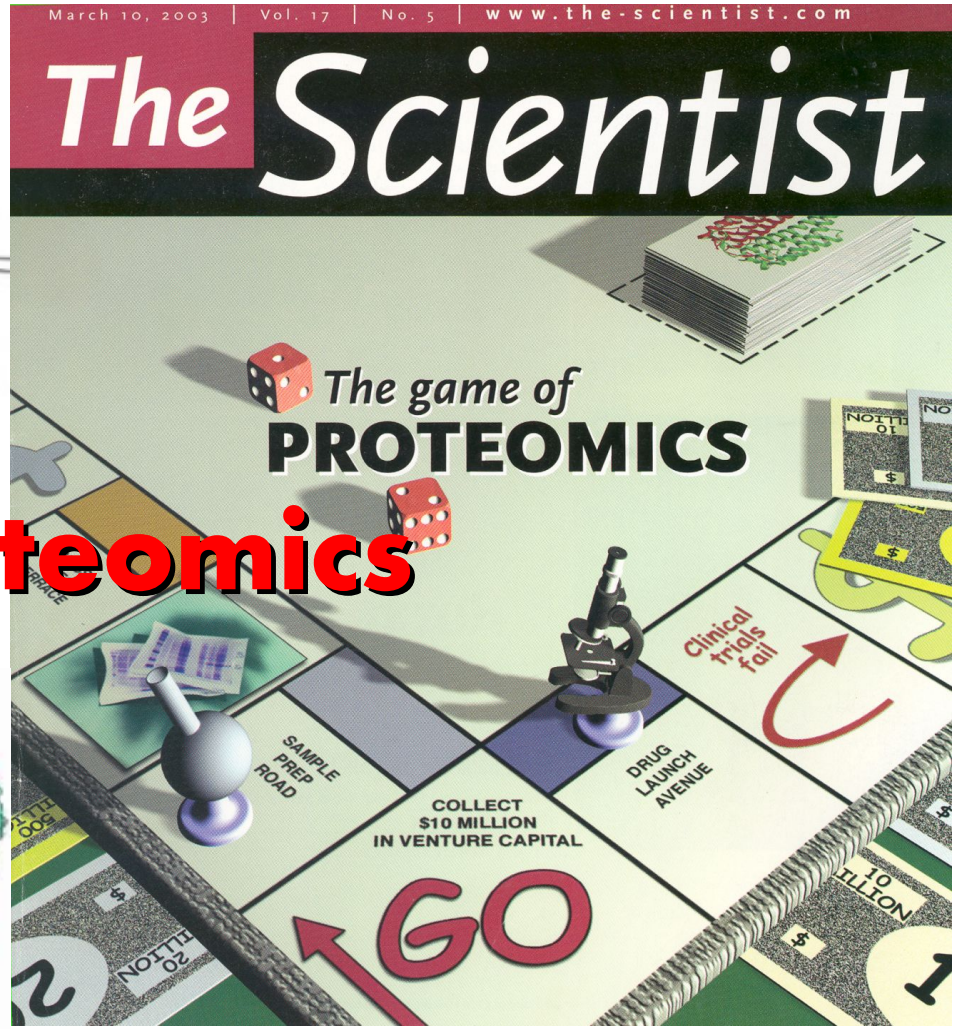
- Enables the analysis of EST to produce gene sequences that can be annotated by the other organizations

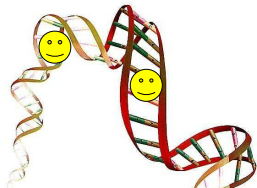



C^{α}_{i+1}
N_{i+1}
O_i
C_i
ω_i
ψ_i
C^{α}_i
ϕ_i
N_i
C^{β}_i
C_{i-1}
O_{i-1}



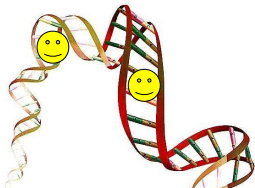
Proteomics


March 10, 2003 | Vol. 17 | No. 5 | www.the-scientist.com
The Scientist
The game of
PROTEOMICS
SAMPLE PREP ROAD
COLLECT $10 MILLION IN VENTURE CAPITAL
GO
DRUG LAUNCH AVENUE
Clinical trials fail



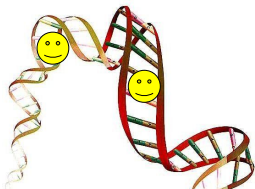
Proteomics

- MAS for Secondary Structure Prediction
  - MASSP3 **[Armano06]**
  - A Cooperative MultiAgent Learning Approach **[Addis07]**
- MAS for Tertiary Structure Prediction
  - X-MACoP **[Garro04]**
  - An Agent-Based Protein Structure Predictor **[Bortolussi07]**
- MAS for classifying 2D gels



MASSP3

- MASSP3 (MultiAgent Secondary Structure Predictor with Post-Processing)
  - resorts to multiple experts for dealing with the problem of predicting secondary
  - performs an overall processing based on two main steps:
    - a “sequence-to-structure” prediction, by resorting to a population of hybrid genetic-neural experts
    - a “structure-to-structure” prediction, by resorting to a feedforward ANN

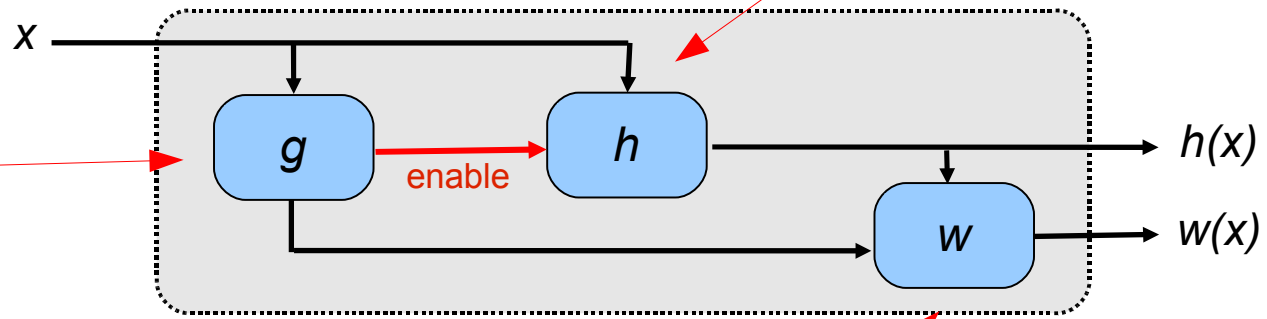


MASSP3

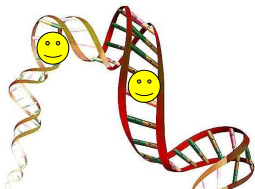
- Sequence to structure
  - the micro-architecture (based on **[Armano04]**)

- An embedded expert whose activation depends on g(x)

- A binary function that selects inputs according to some relevant features



- A weighting function used to perform output combination

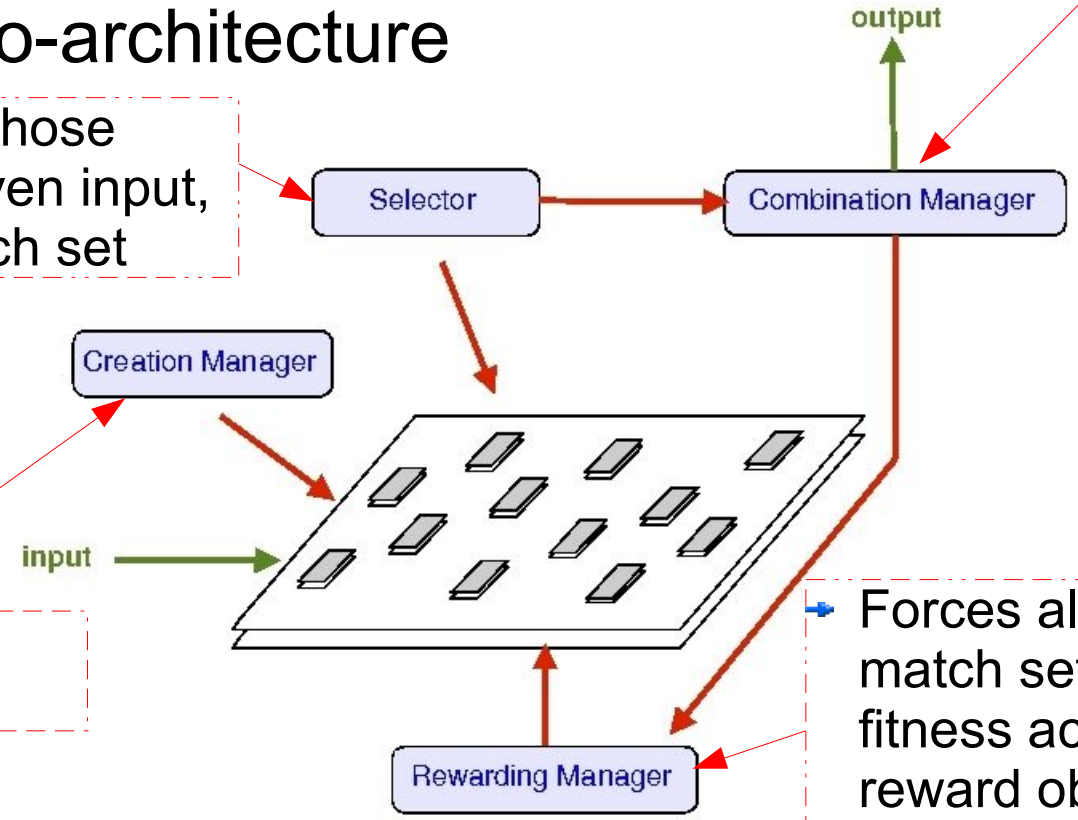


MASSP3

- Sequence to structure:
  - the macro-architecture

Combines the outputs of experts belonging to the match set, so that a suitable voting policy can be enforced on them

Collects all experts whose guards covers the given input, thus forming the match set

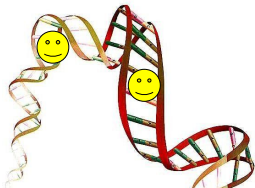


Creates experts, when needed

Forces all experts in the match set to update their fitness according to the reward obtained by the external environment

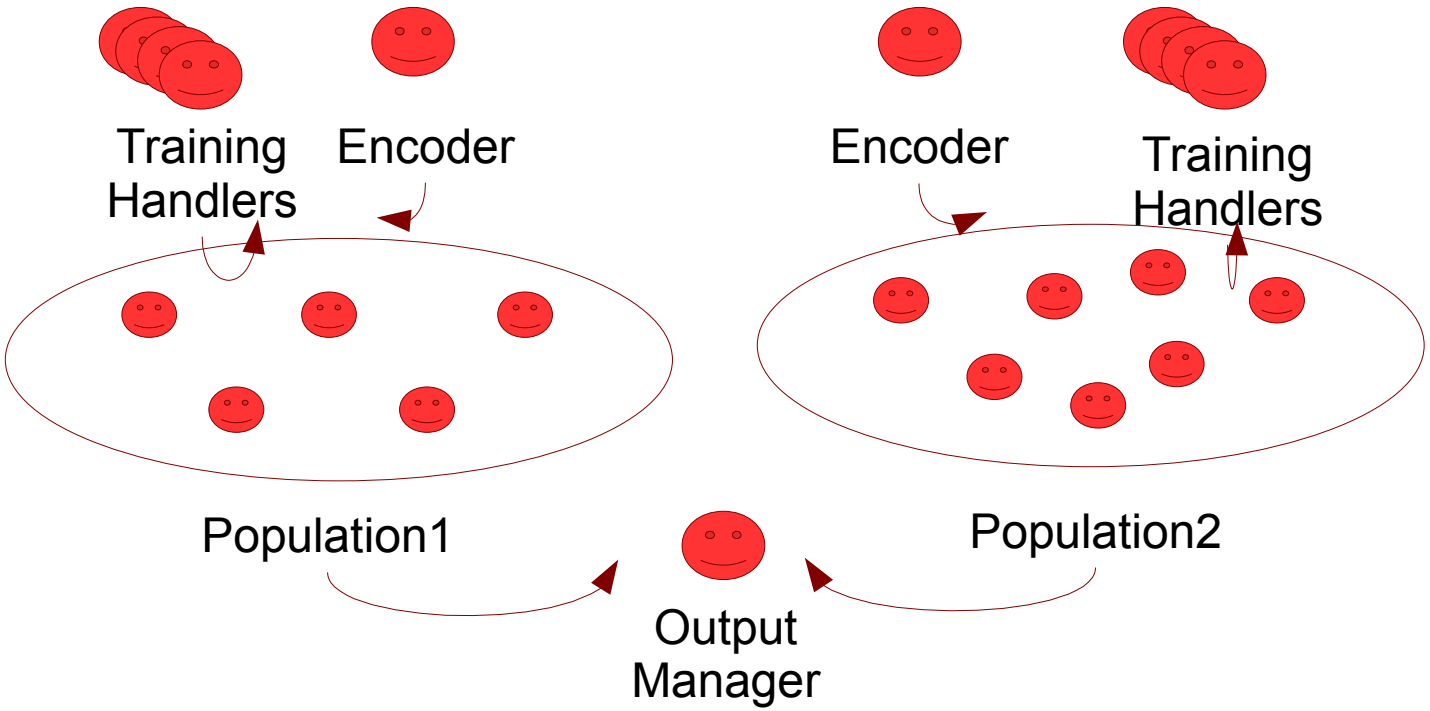
MASSP3

- Structure to structure:
  - to improve the prediction accuracy
    - a suitable post-processing is performed by a single MLP fed with the sequence-to-structure prediction expressed in terms of α, β, and c propensions
  - to augment the autocorrelation of the input signal
    - a suitable “low-pass” filtering is performed by resorting to a suitable gaussian shape

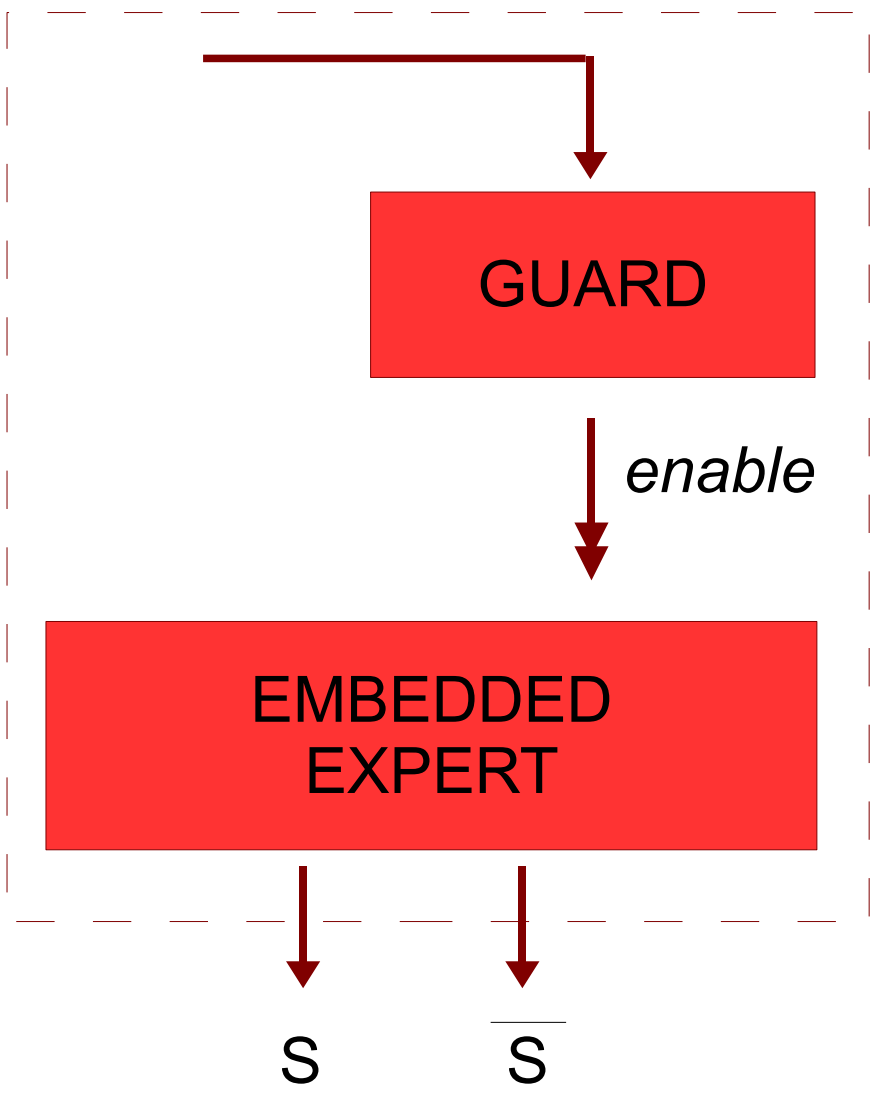


A Cooperative MultiAgent Learning Approach to SSP

- A MAS composed by a population of agents that cooperate and interact in accordance with generic requirements imposed by the adoption of evolutionary computation strategies

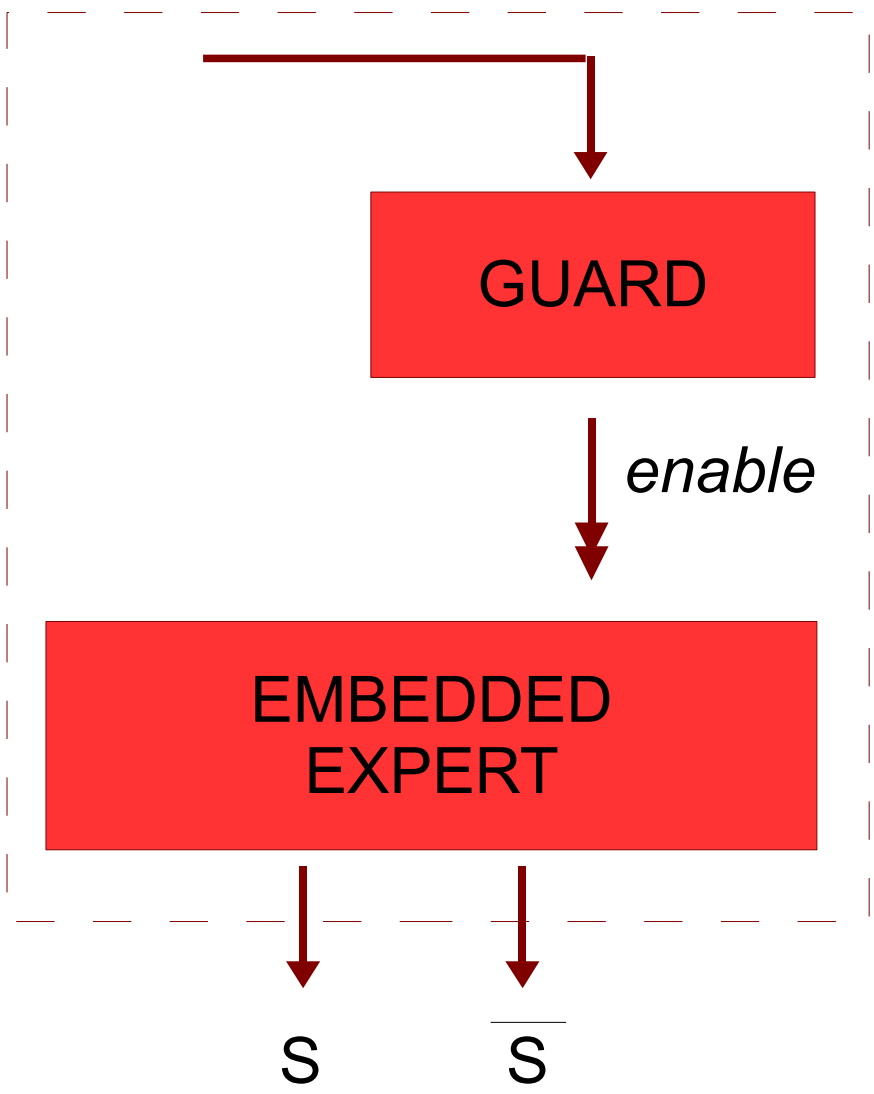


A Cooperative MultiAgent Learning Approach to SSP

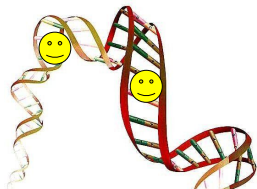


- The guard is devoted to split the input space into matching/non matching regions
- An approach based on activation masks: a mask is a pattern of amino acids that matches with the corresponding input

A Cooperative MultiAgent Learning Approach to SSP

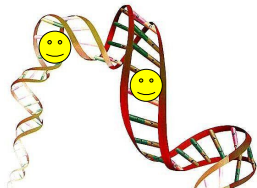


- The embedded expert plays a central role in the output combination
- Trained to recognize a specific secondary structure:
  - a label is assigned to each embedded expert
  - a feedforward ANN has been implemented



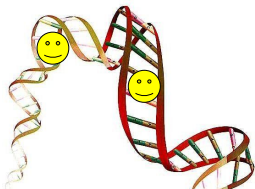
A Cooperative MultiAgent Learning Approach to SSP

- Output combination: all selected experts form the match set *M(x)*
  - *M(x)* is partitioned in three separate subsets $M_H(x)$, $M_E(x)$, and $M_C(x)$
  - several independent predictions made from different experts are expected
  - a weighted majority rule is adopted:
  - the final predicted label is the one that maximizes the corresponding *W* value

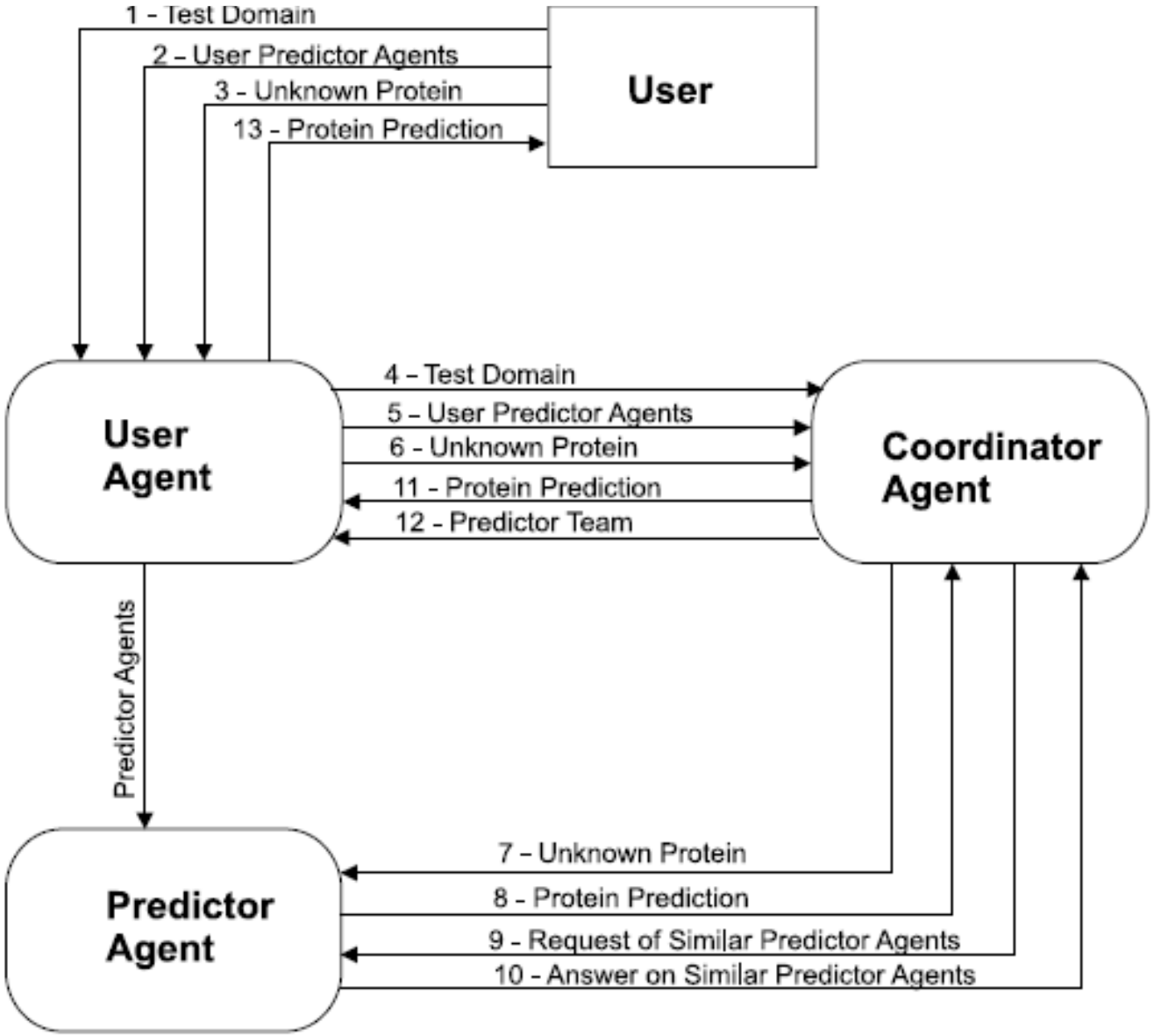


X-MACoP

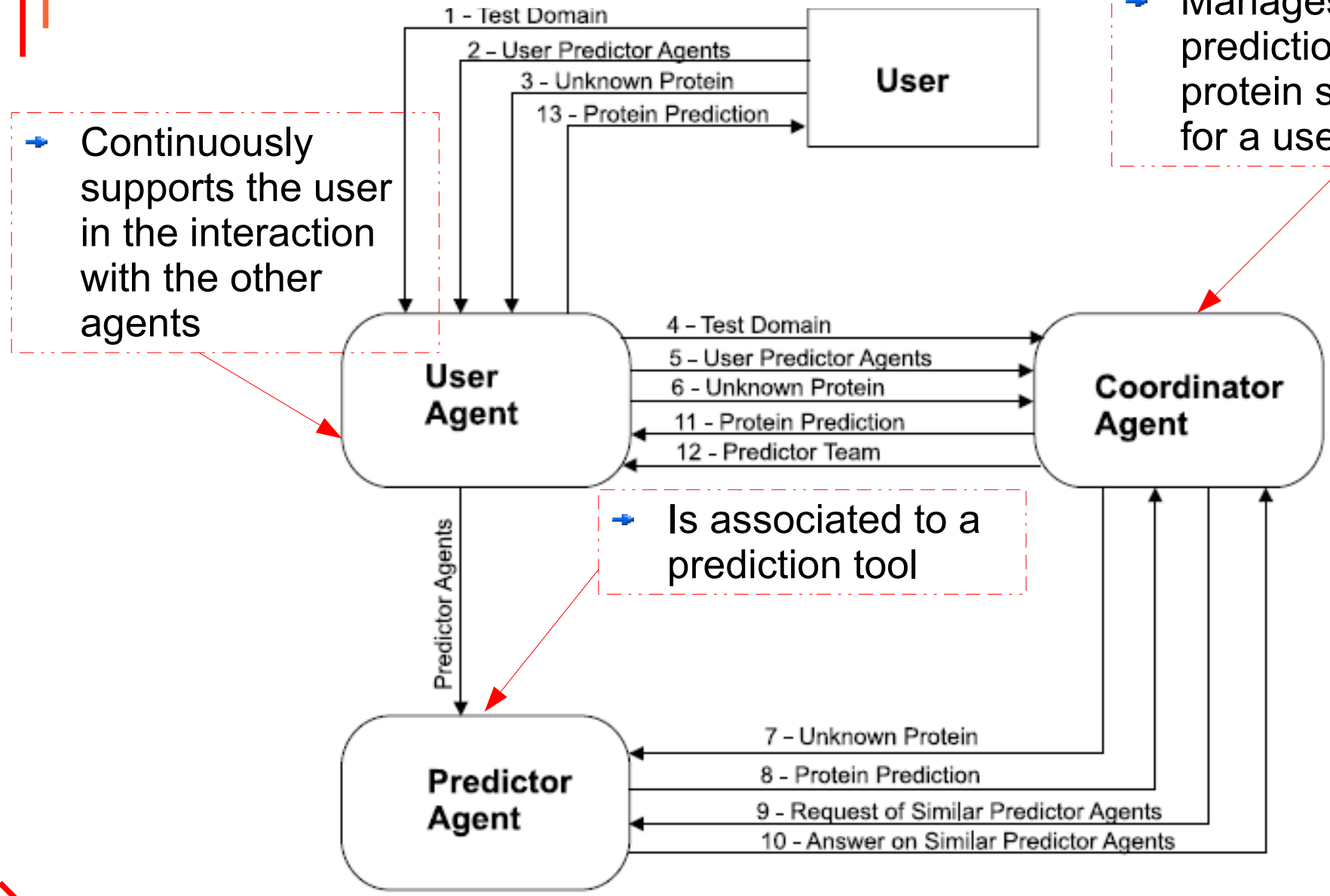
- X-MACoP (XML Multi-Agent system for the Collaborative Prediction of protein structures)
  - Consists of three kinds of agents
    - User Agent
    - Predictor Agent
    - Coordinator
  - Each agent is characterized by its ontology and its behavior.



X-MACoP

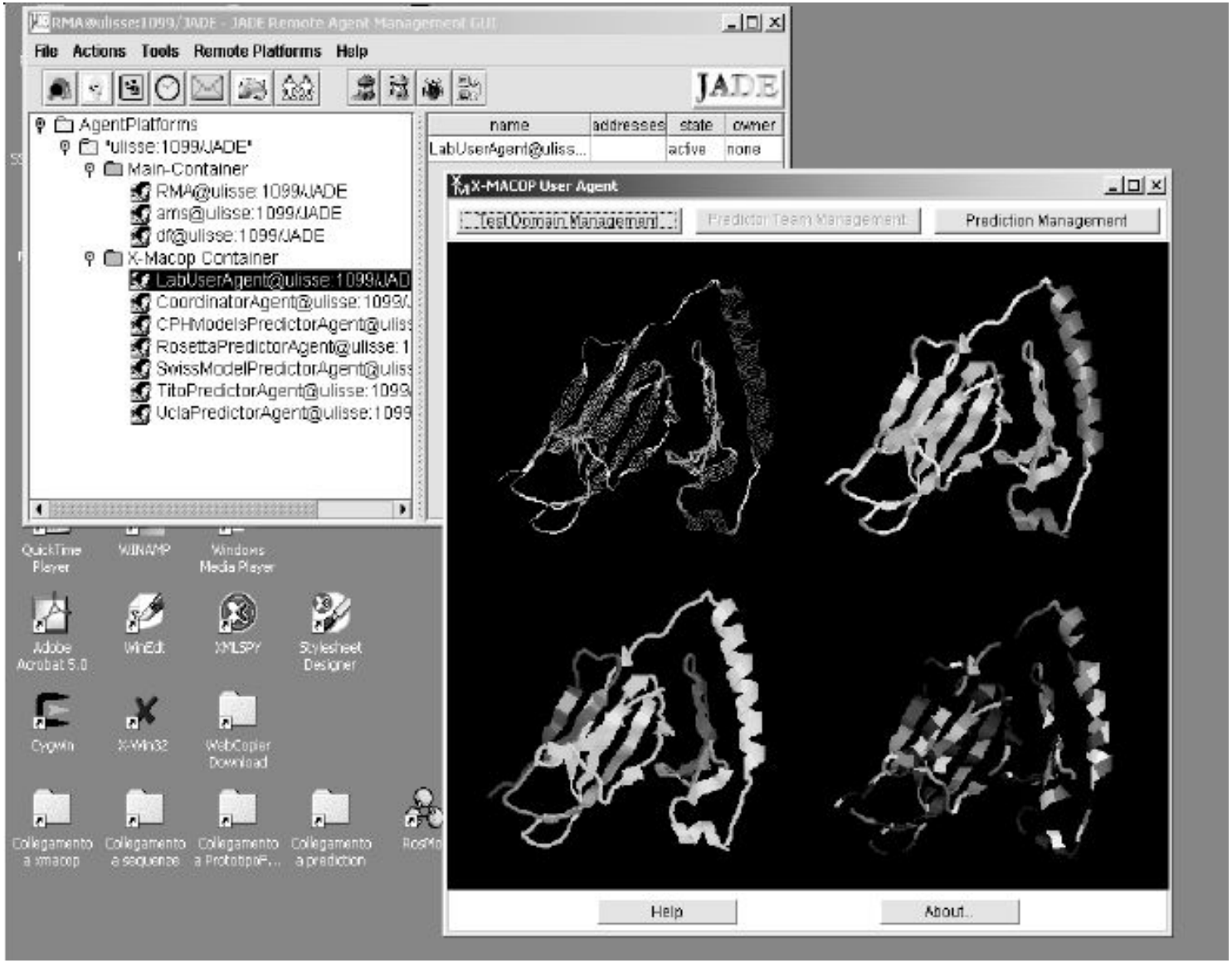

1 - Test Domain
2 – User Predictor Agents
3 - Unknown Protein
13 - Protein Prediction
User
User Agent
4 – Test Domain
5 – User Predictor Agents
6 - Unknown Protein
11 - Protein Prediction
12 - Predictor Team
Coordinator Agent
Predictor Agents
7 – Unknown Protein
8 - Protein Prediction
9 - Request of Similar Predictor Agents
10 - Answer on Similar Predictor Agents
Predictor Agent

X-MACoP



- Manages the prediction of the protein structure for a user
- Continuously supports the user in the interaction with the other agents
- Is associated to a prediction tool

X-MACoP

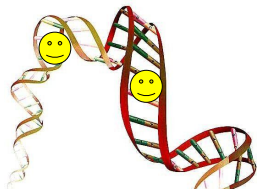

File Actions Tools Remote Platforms Help
JADE
AgentPlatforms
"ulisse:1099/JADE"
Main-Container
RMA@ulisse:1099/JADE
ams@ulisse:1099/JADE
df@ulisse:1099/JADE
X-Macop Container
LabUserAgent@ulisse:1099/JAD
CoordinatorAgent@ulisse:1099/
CPHModelsPredictorAgent@ulis
RosettaPredictorAgent@ulisse:1
SwissModelPredictorAgent@ulis
TitoPredictorAgent@ulisse:1099
UclaPredictorAgent@ulisse:1099
name
addresses
state
owner
LabUserAgent@uliss...
active
none
X-MACOP User Agent
TestDomain Management
Predictor Team Management
Prediction Management
Help
About...
QuickTime Player
WINAMP
Windows Media Player
Adobe Acrobat 5.0
WinEdt
XMLSPY
Stylesheet Designer
Cygwin
X-Win32
WebCopier Download
Collegamento a xmacop
Collegamento a sequence
Collegamento a PrototipoF...
Collegamento a prediction

X-MACoP

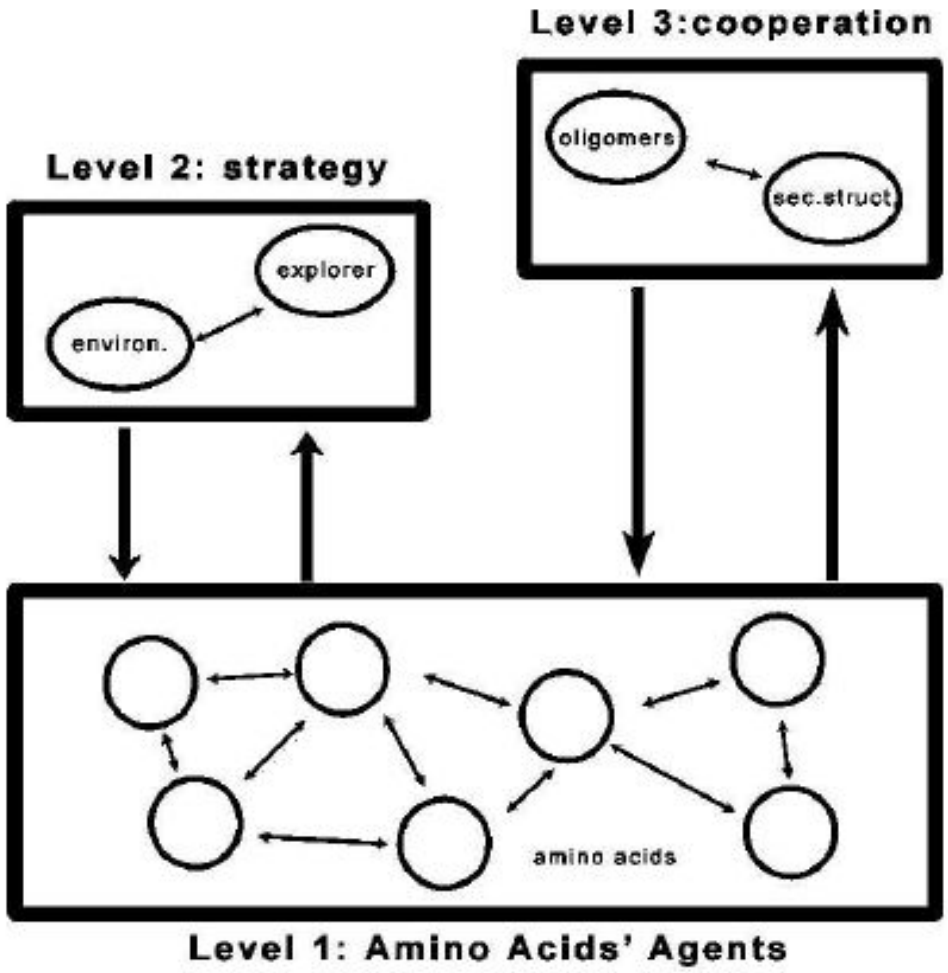

Build a Predictor Team PT for the current prediction task
Choose the predictors
ROSETTA
SWISS MODEL
T.I.T.O.
UCLA
CPH MODEL
weighing coeff. for Rosetta
weighing coeff. for Swiss
weighing coeff. for Tito
weighing coeff. for Ucla
weighing coeff. for Cph
0.8
1
0.5
0.2
0.3
choose the measures to be performed
THd
0.25
Predictor Local Precision
Predictor Global Precision
Team conjunctive local precision
Team conjunctive global precision
Team disjunctive local precision
Team disjunctive global precision
Predictors Affinity
Retrieve PDBs
Compute Predictions
3D Alignment
Compute measures
Back to Main Menu'

An Agent-Based Protein Structure Predictor

- Four level of agents
  - Level 0 deals with generation of an initial solution
  - Level 1 is focused on the stochastic search in the state space
  - Level 2 performs global strategic tasks
  - Level 3 deals with cooperation strategies
- There are agents of level 1,2 and 3 (level 0 agents are trivial)
- All agents interact and communicate in the Linda tuple space

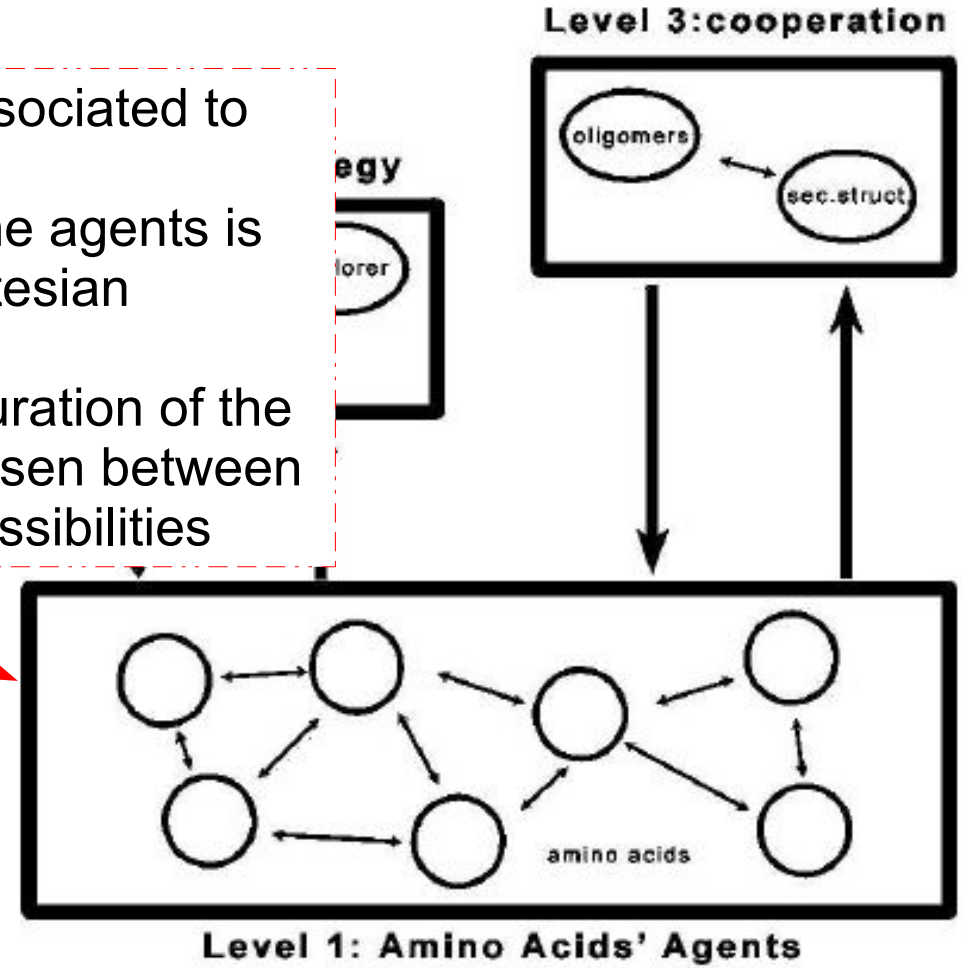


An Agent-Based Protein Structure Predictor



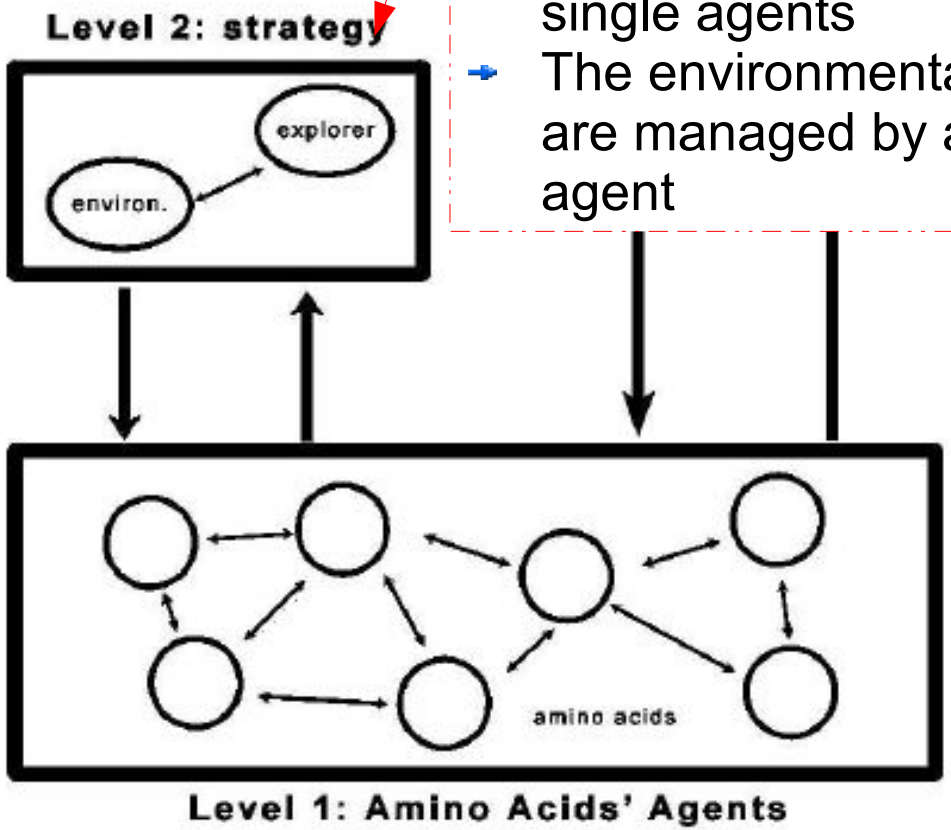
An Agent-Based Protein Structure Predictor

- Each agent is associated to an aminoacid
- The position of the agents is expressed in cartesian coordinates
- The initial configuration of the chain can be chosen between three different possibilities

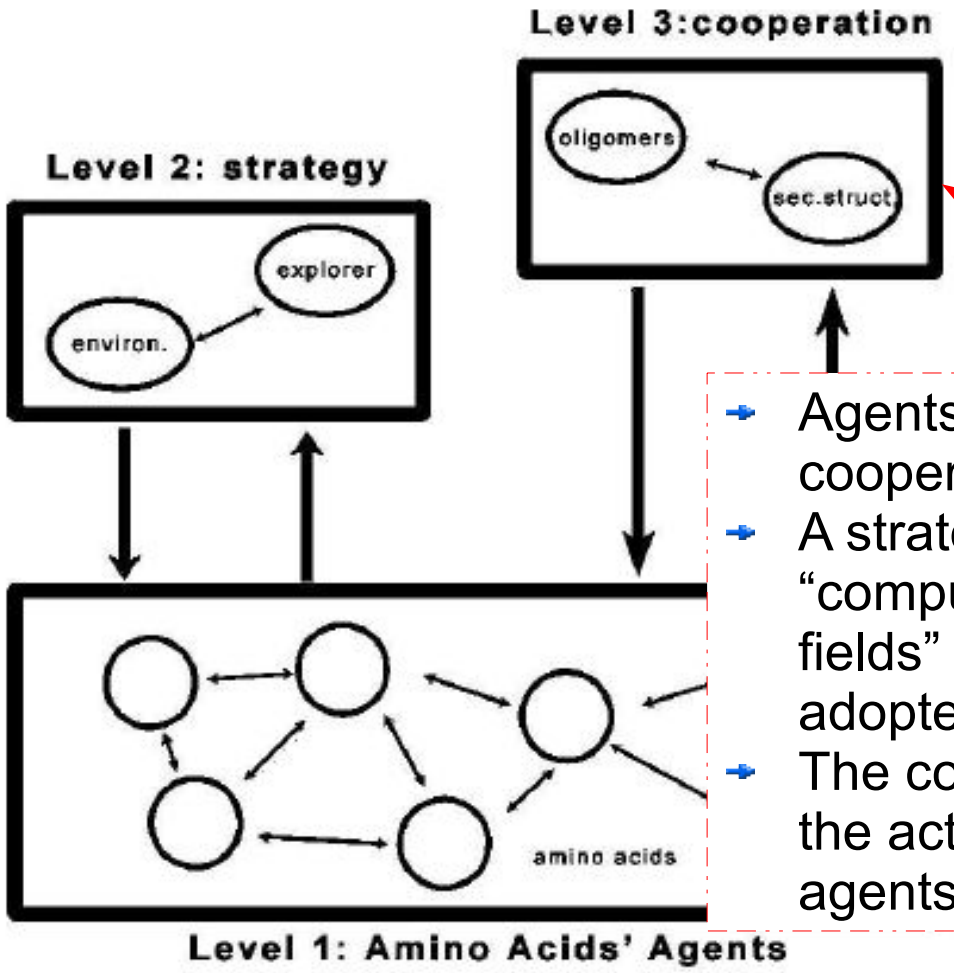


An Agent-Based Protein Structure Predictor

- Agents have a global knowledge of the current configuration
- Agents are aimed at controlling the activity of the single agents
- The environmental variables are managed by a specific agent



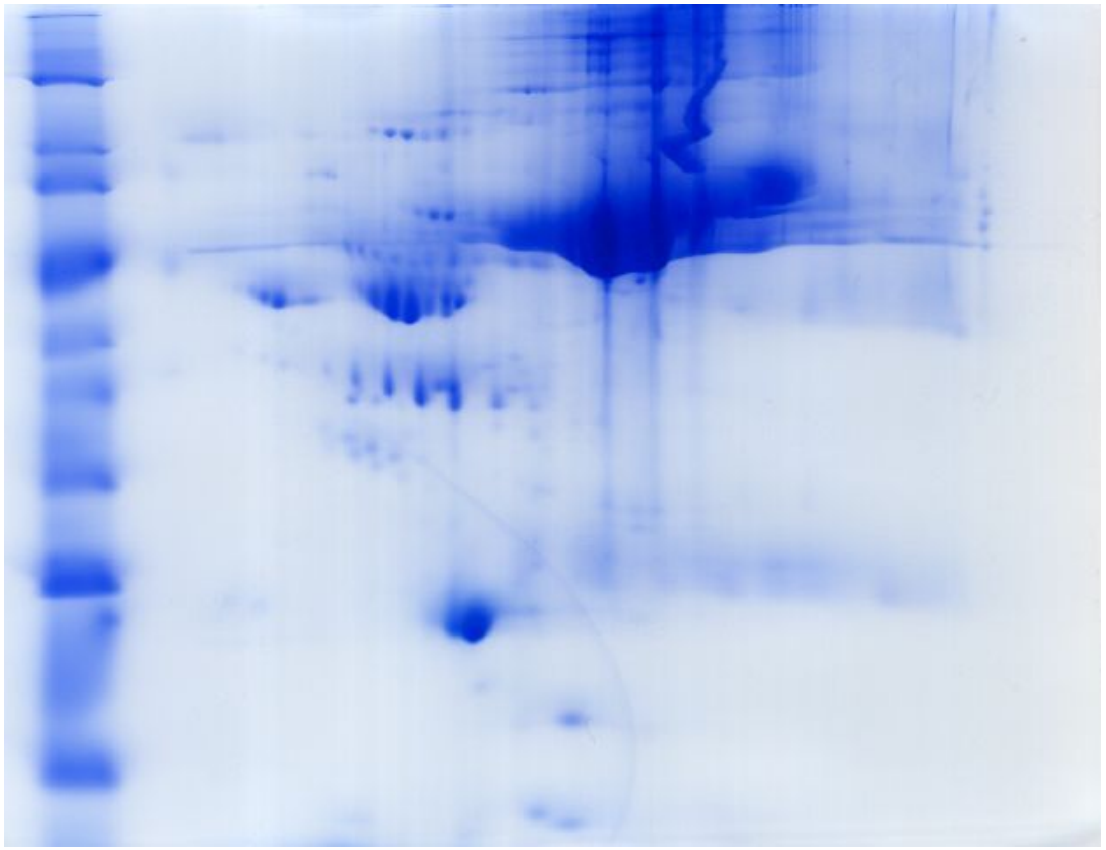
An Agent-Based Protein Structure Predictor



- Agents govern the cooperation
- A strategy similar to "computation fields" **[Leonardi02]** has been adopted
- The cooperation agent control the activity of all lower level agents

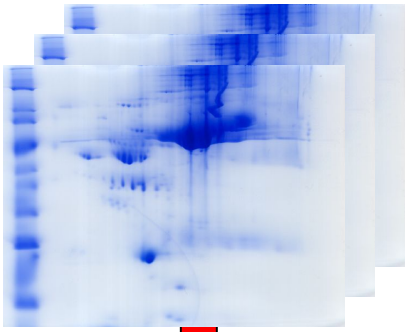
A MultiAgent System for Classifying 2D Gels

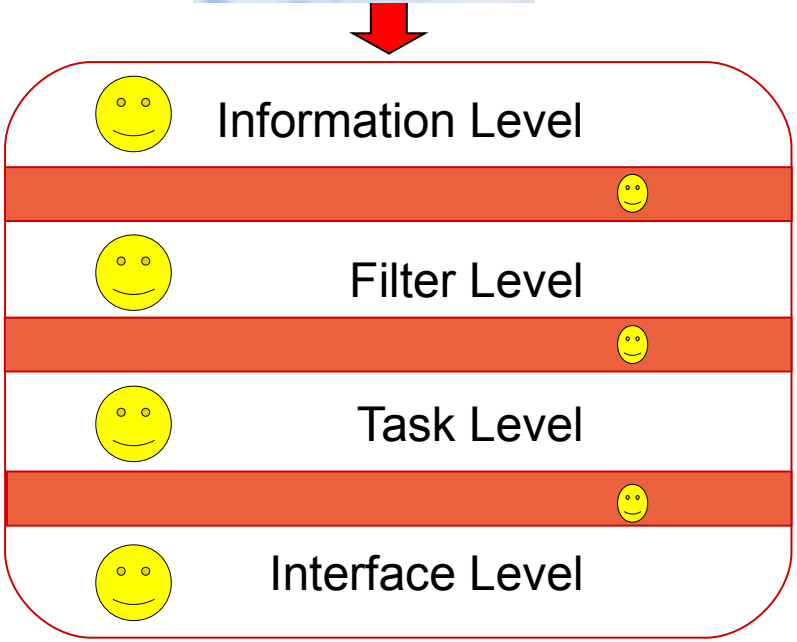
- We are currently customizing a generic multiagent architecture to classify 2D gel images

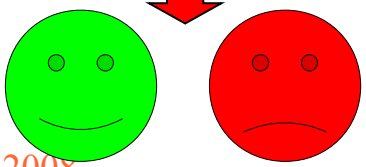


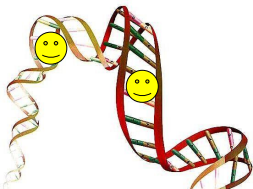
A MultiAgent System for Classifying 2D Gels

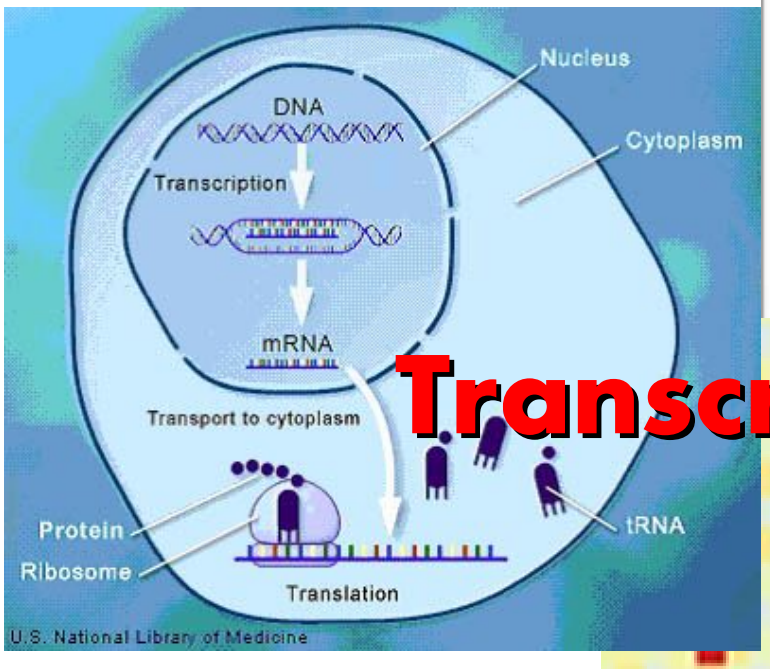
- The system will be able to support biologists in recognizing images corresponding to healthy / ill patients

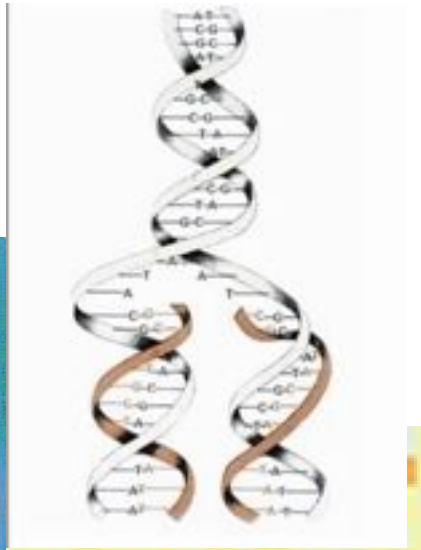




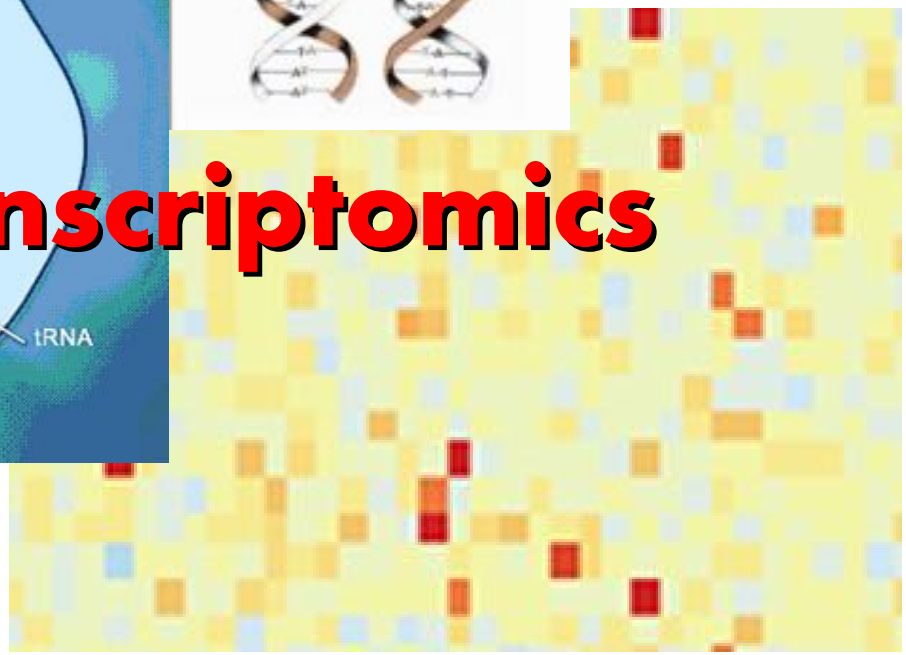






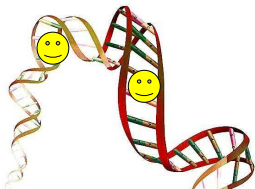


Transcriptomics



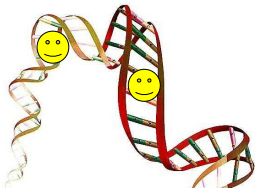
Transcriptomics

- Functional Genomics
  - Automatic Workflows Composition **[Orro06]**
- Gene Expression
  - A Multi-agent Approach to Gene Expression Analysis **[Lam06]**
  - Using Multi-Agent System for Gene Expression Classification **[Stiglic04]**



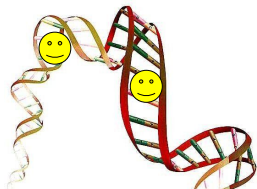
Automatic Workflows Composition in Homology Based Functional Genomics

- Most of the tasks in bioinformatics analysis of genomics sequence can not be carried out with a single standalone application
- An agent approach to support the composition, execution and management of bioinformatics workflows is presented

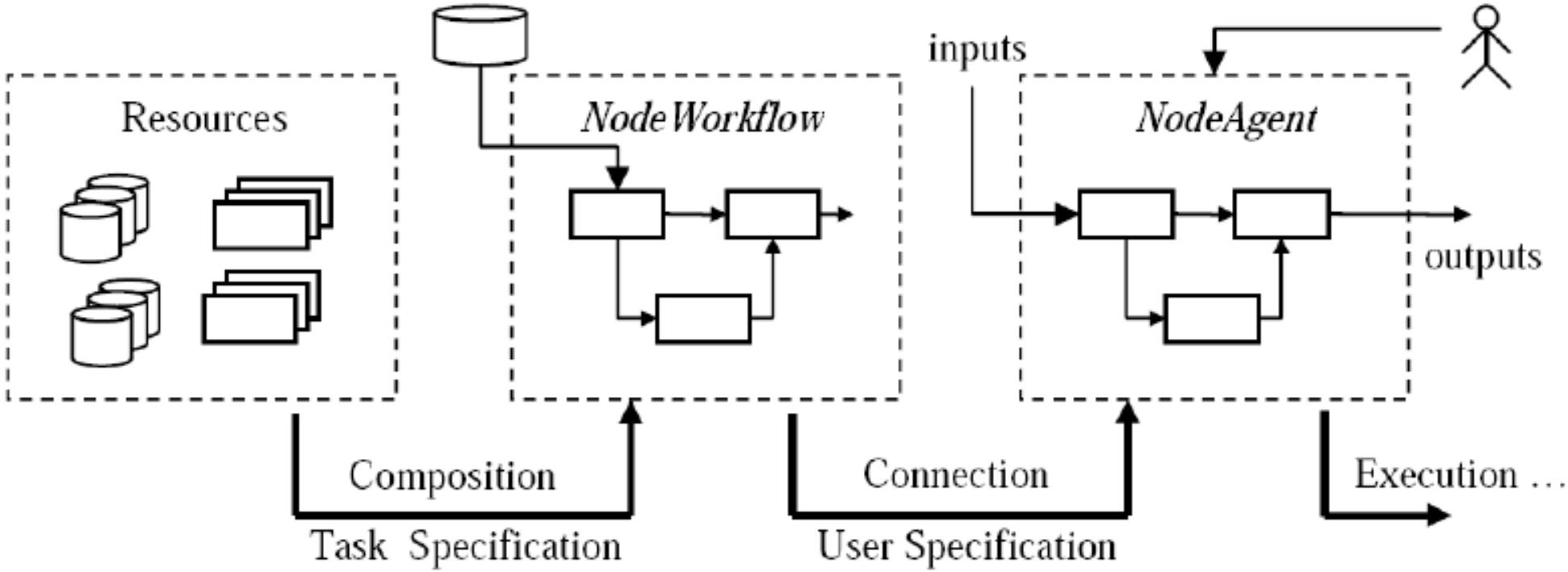


Automatic Workflows Composition in Homology Based Functional Genomics

- In this context an agent is a software entity that plays the role of a node in a workflow and exports the behavior of the underlying application with some additional features
  - it makes use of knowledge domain information to select a suitable connection of resources for the related task
  - it exposes only a set of high-level parameters that are more intuitive for the user
  - it is able to interact with the user at execution time to permit the monitoring of the overall process

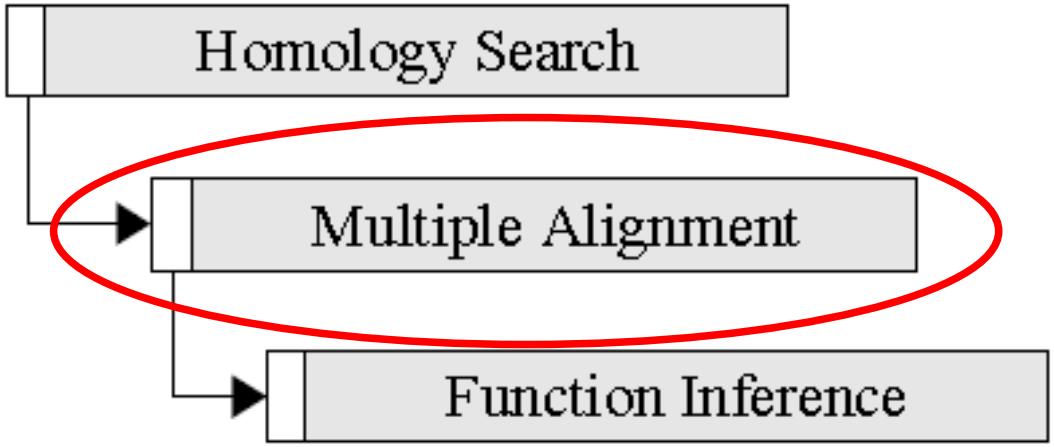


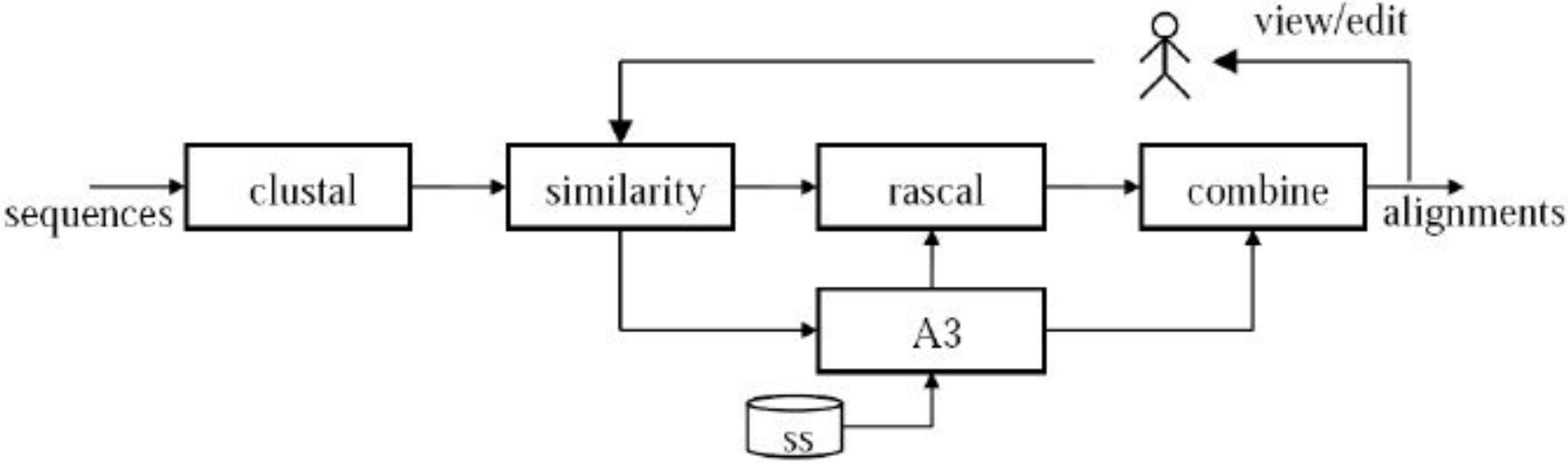
Automatic Workflows Composition in Homology Based Functional Genomics



Automatic Workflows Composition in Homology Based Functional Genomics

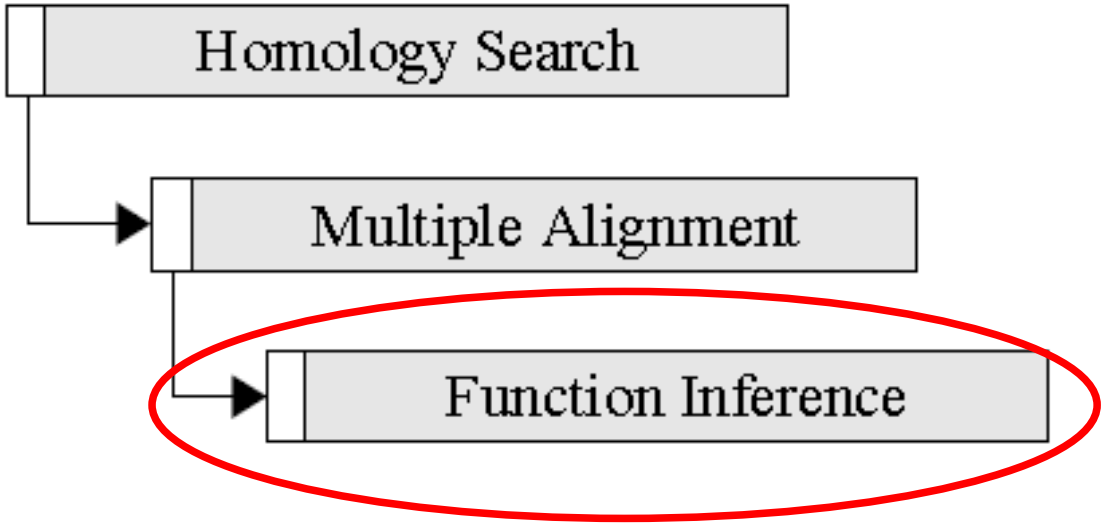
- Case study

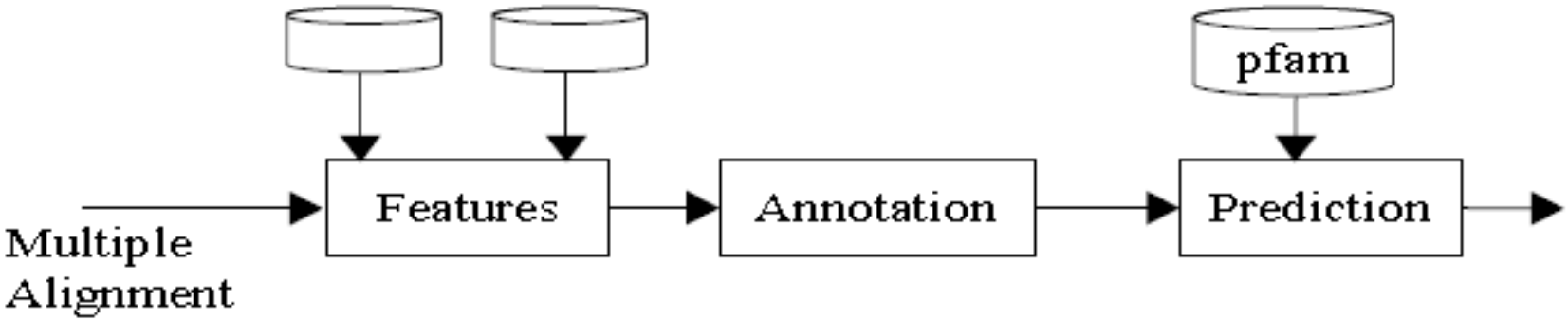




Automatic Workflows Composition in Homology Based Functional Genomics

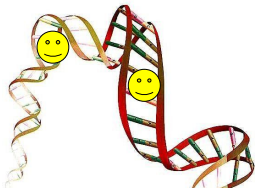
- Case study



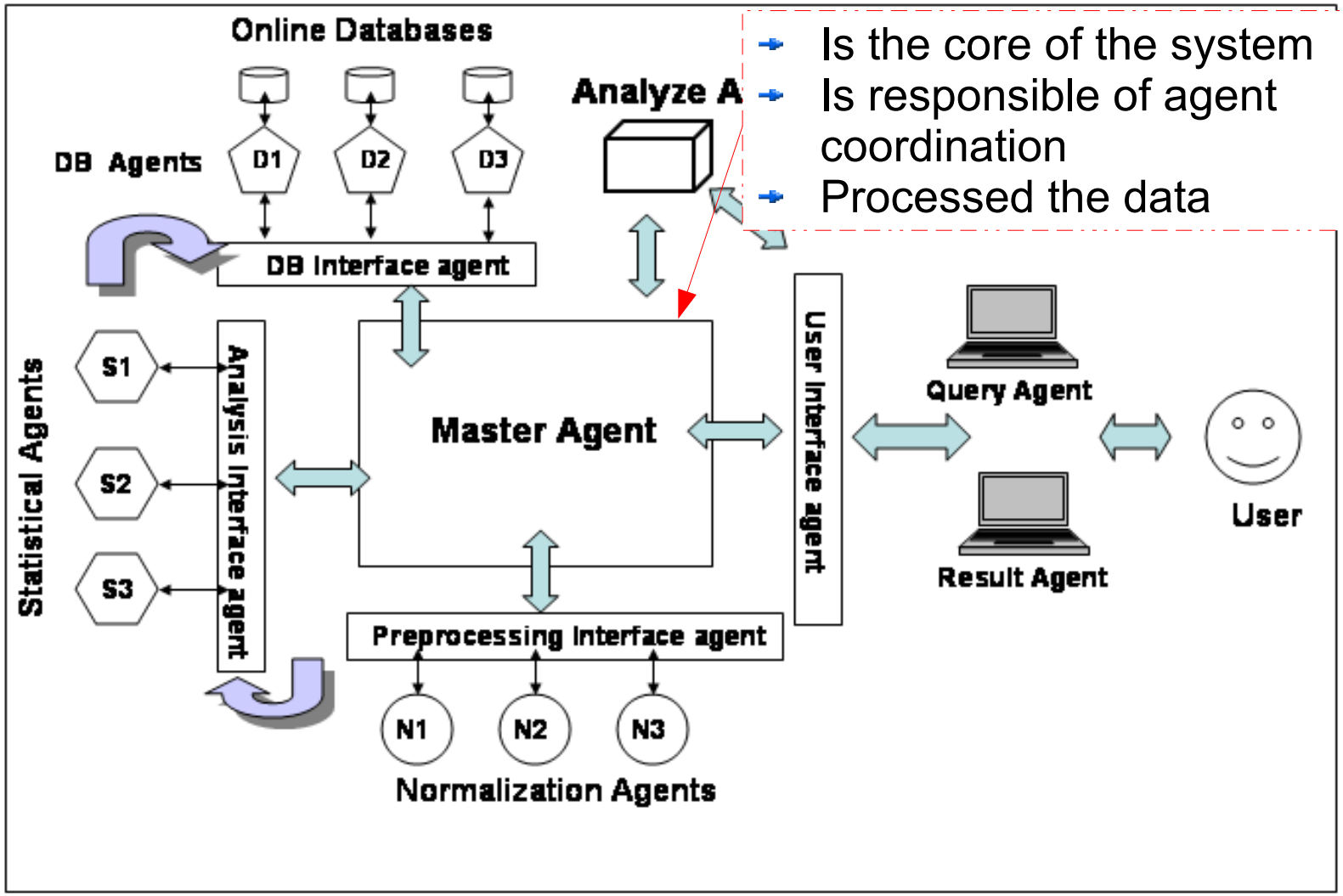


A Multi-agent Approach to Gene Expression Analysis

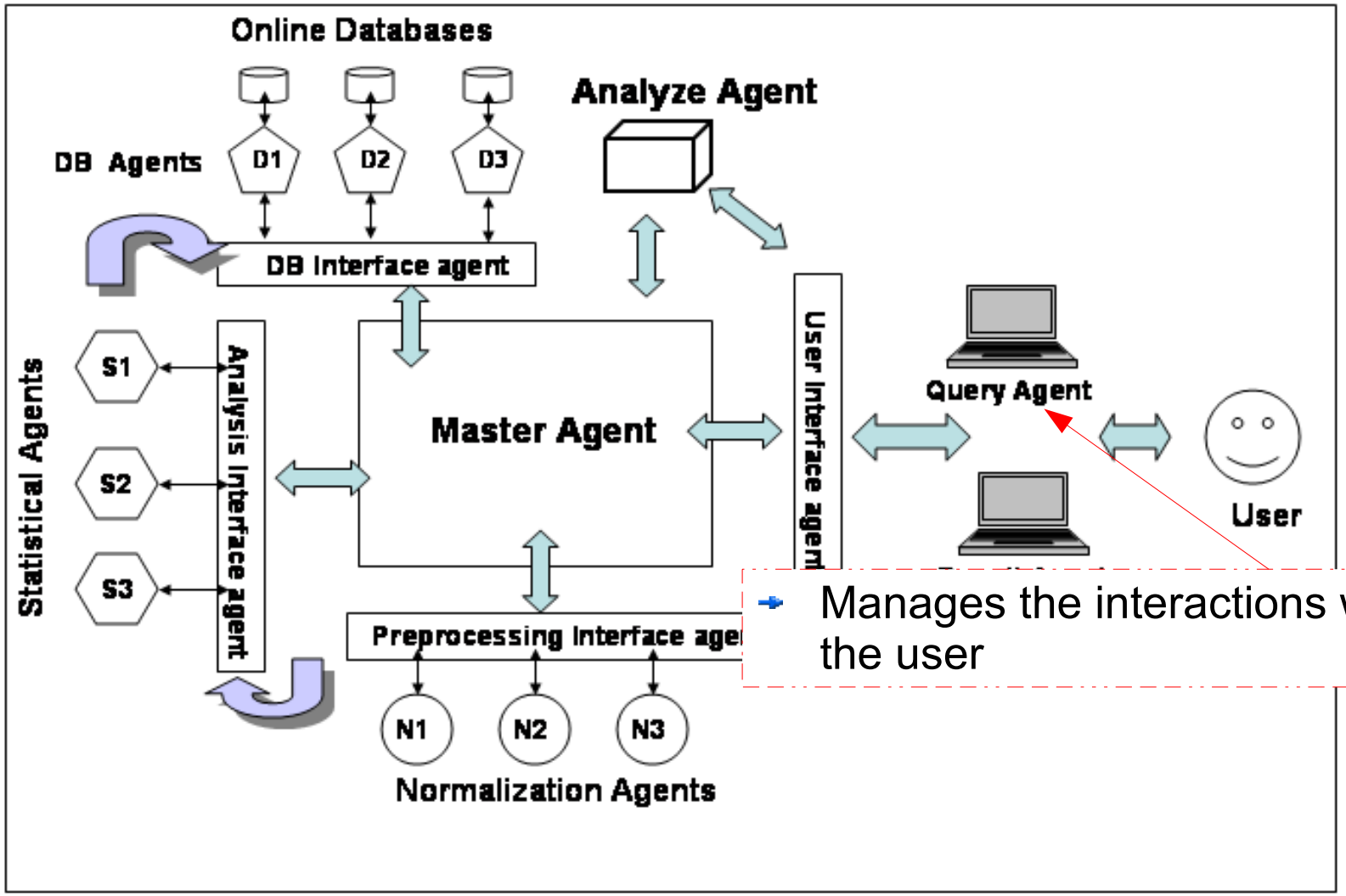
- Gene expression analysis utilizes many resources and the result can be difficult to interpret
- There is not a fixed way to do an analysis and most often researchers will try different strategies to analyze their data set
- The proposed approach is aimed at providing a system that satisfies this need by an adaptive multi-agent solution



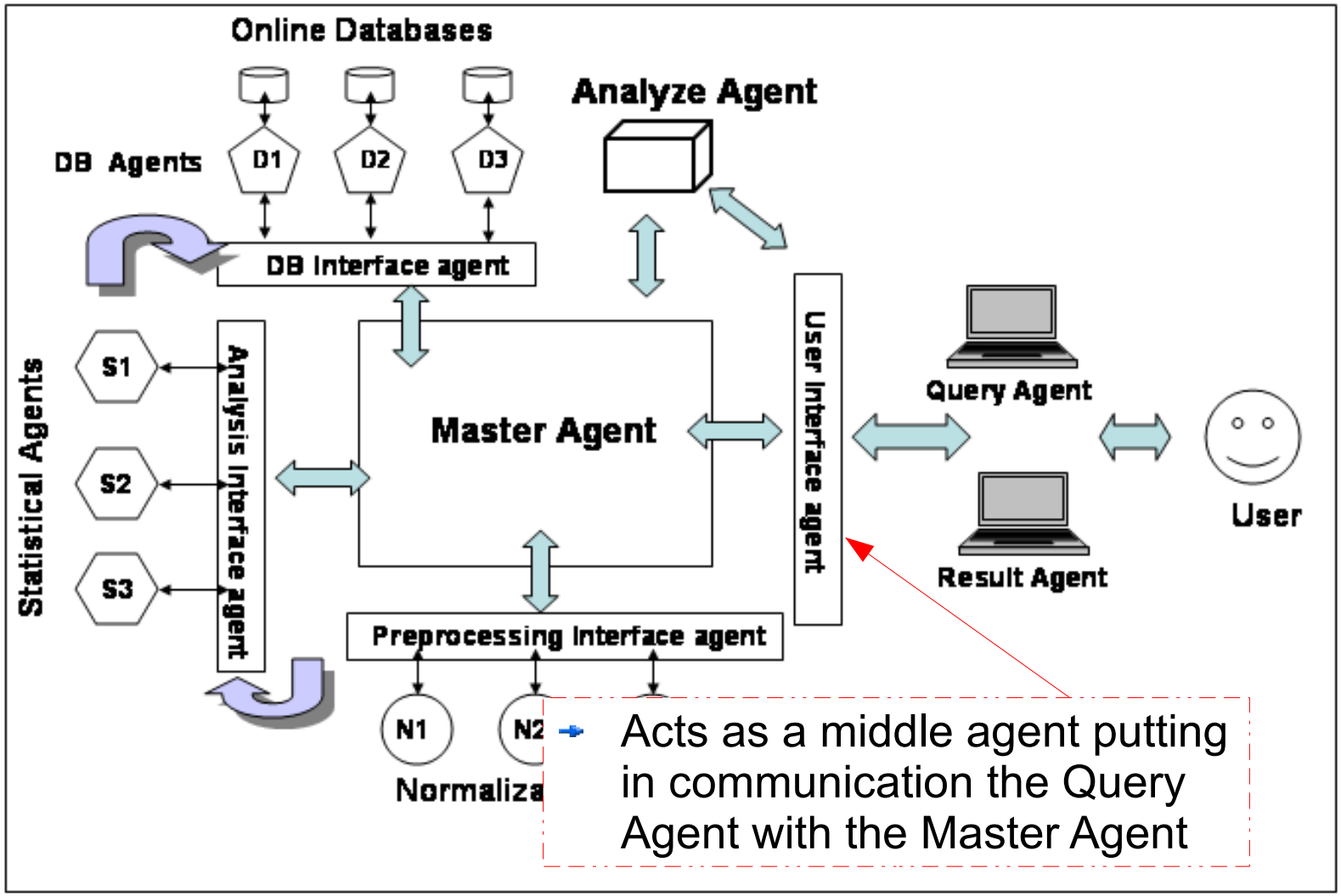
A Multi-agent Approach to Gene Expression Analysis



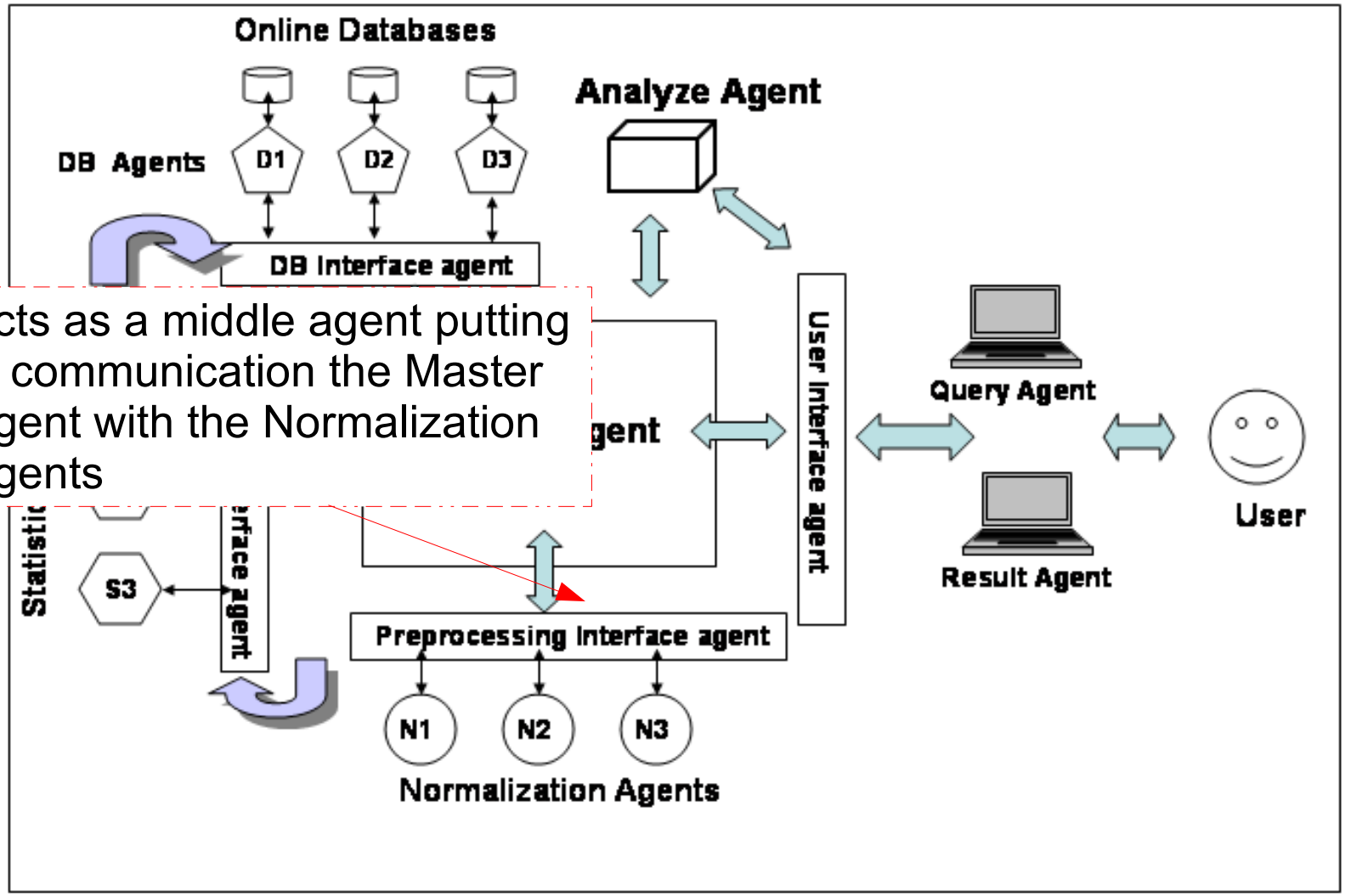
A Multi-agent Approach to Gene Expression Analysis



A Multi-agent Approach to Gene Expression Analysis

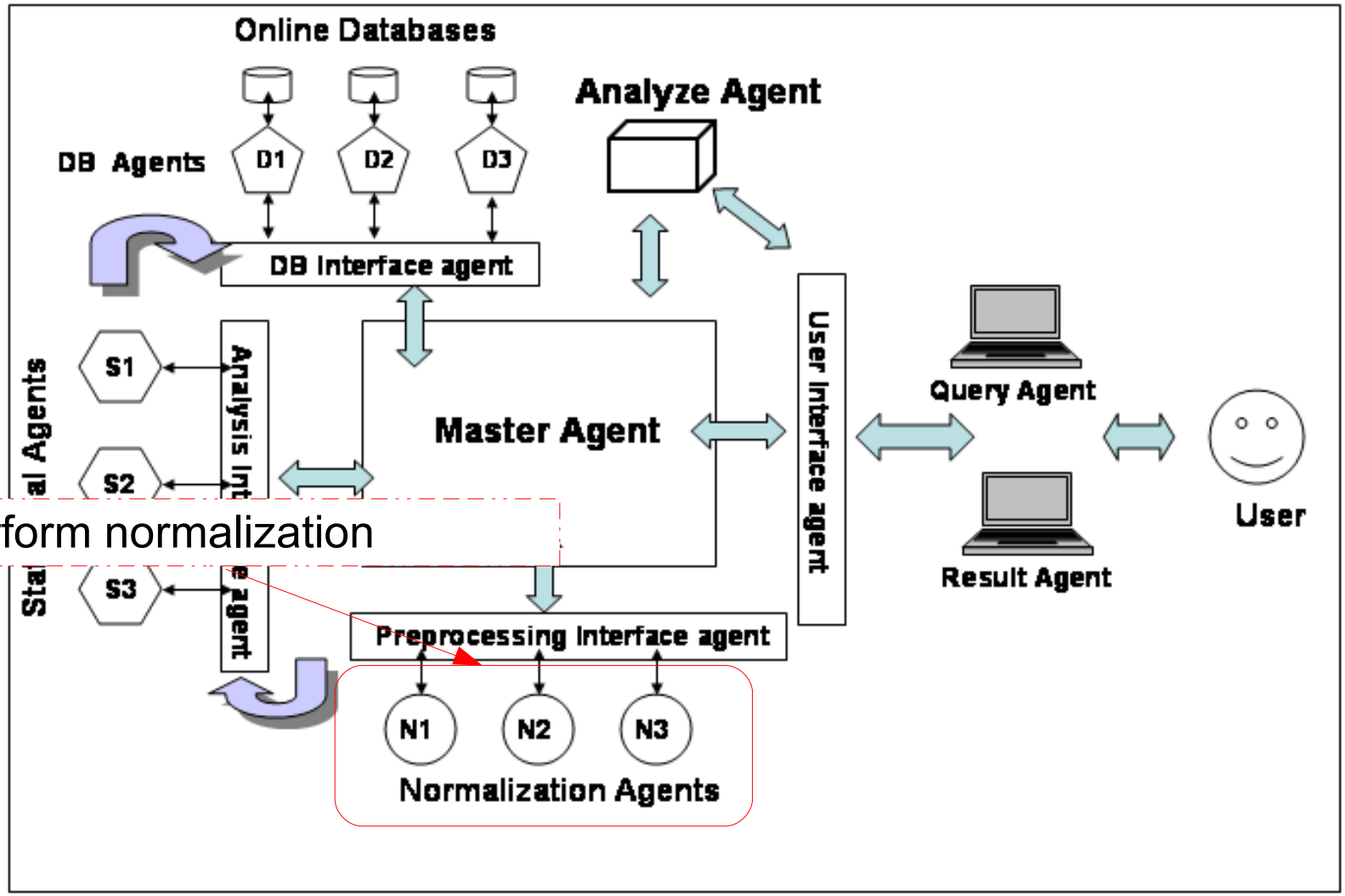


A Multi-agent Approach to Gene Expression Analysis



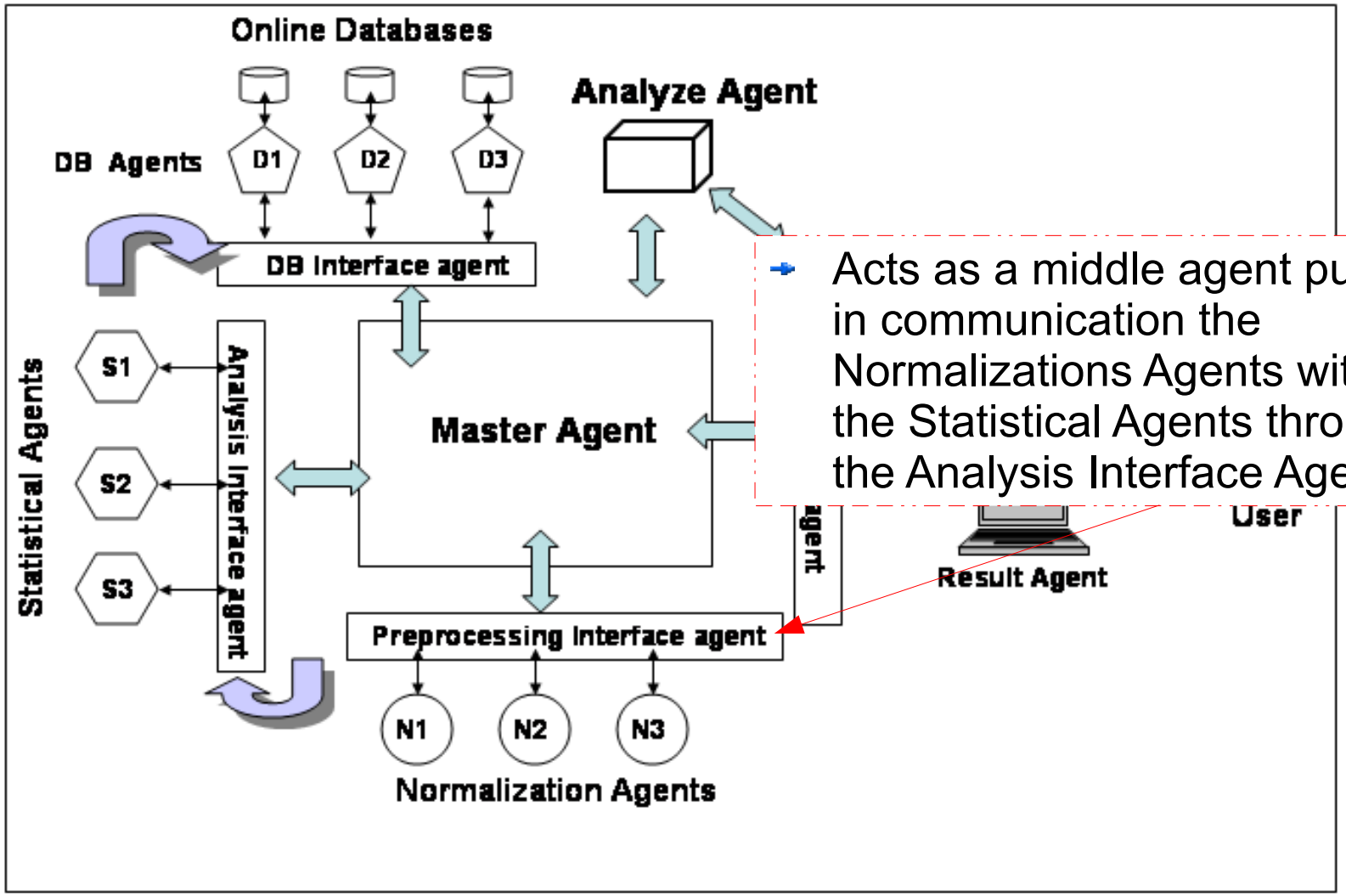
- Acts as a middle agent putting in communication the Master Agent with the Normalization Agents

A Multi-agent Approach to Gene Expression Analysis



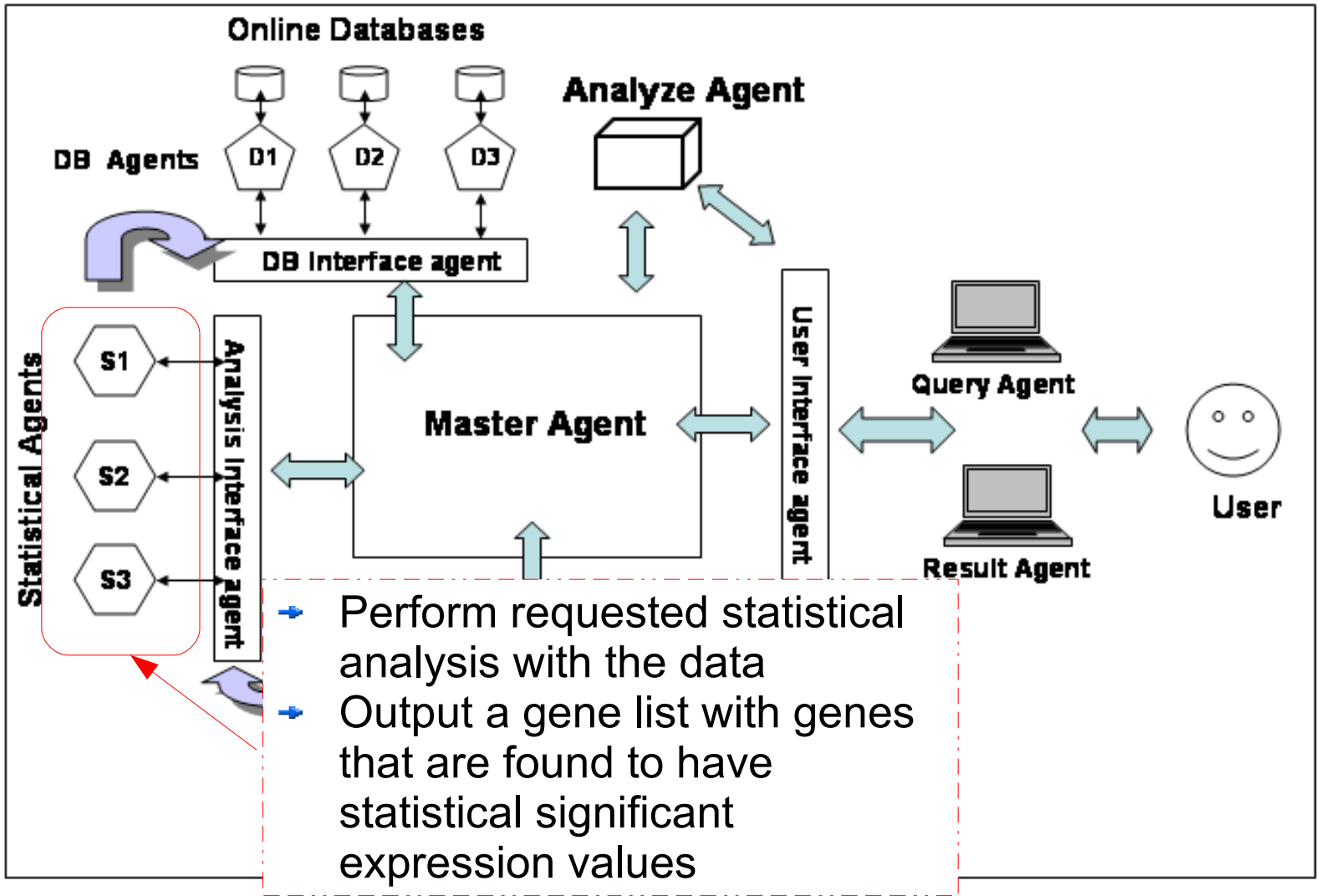
- Perform normalization

A Multi-agent Approach to Gene Expression Analysis

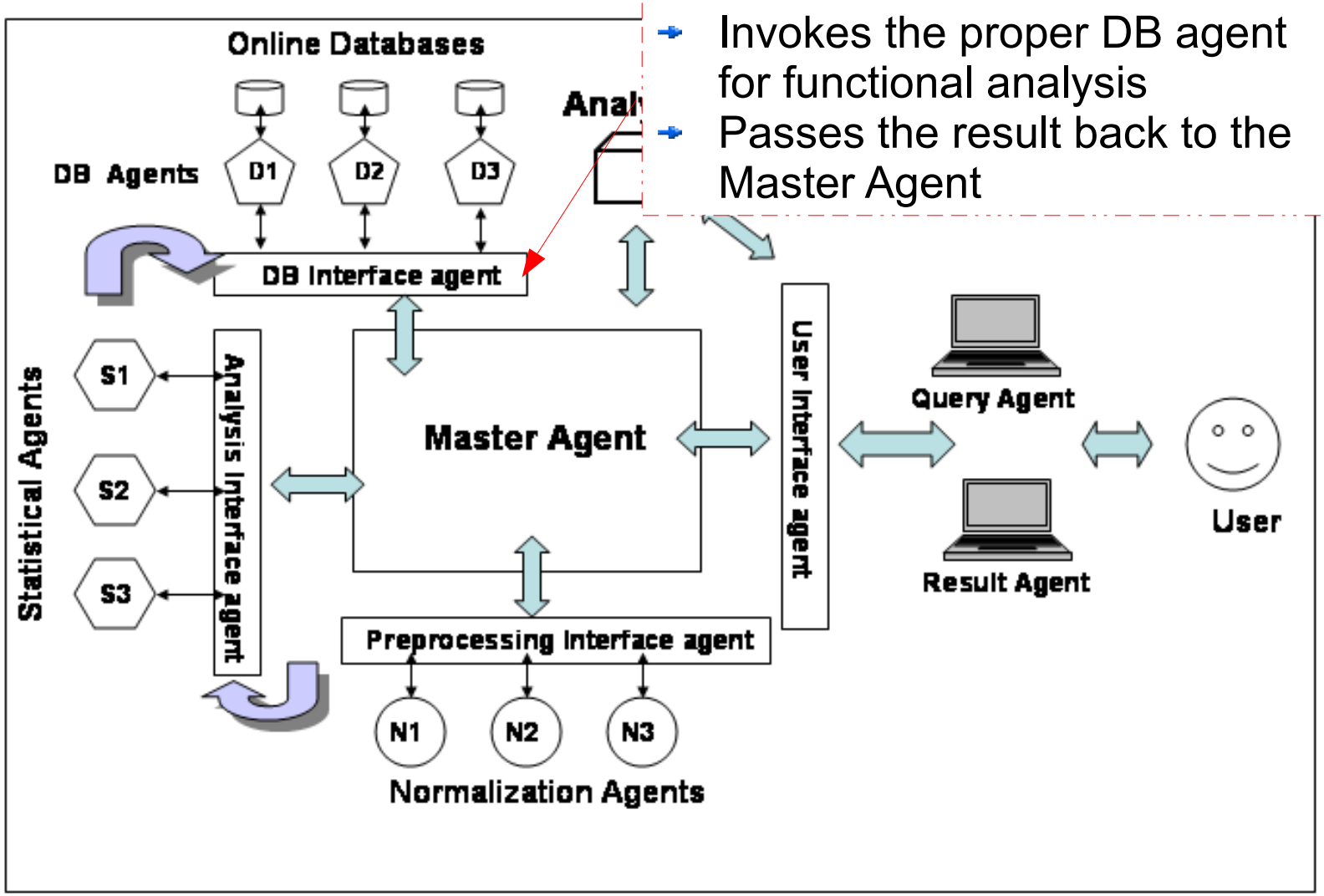


- Acts as a middle agent putting in communication the Normalizations Agents with the Statistical Agents through the Analysis Interface Agent

A Multi-agent Approach to Gene Expression Analysis

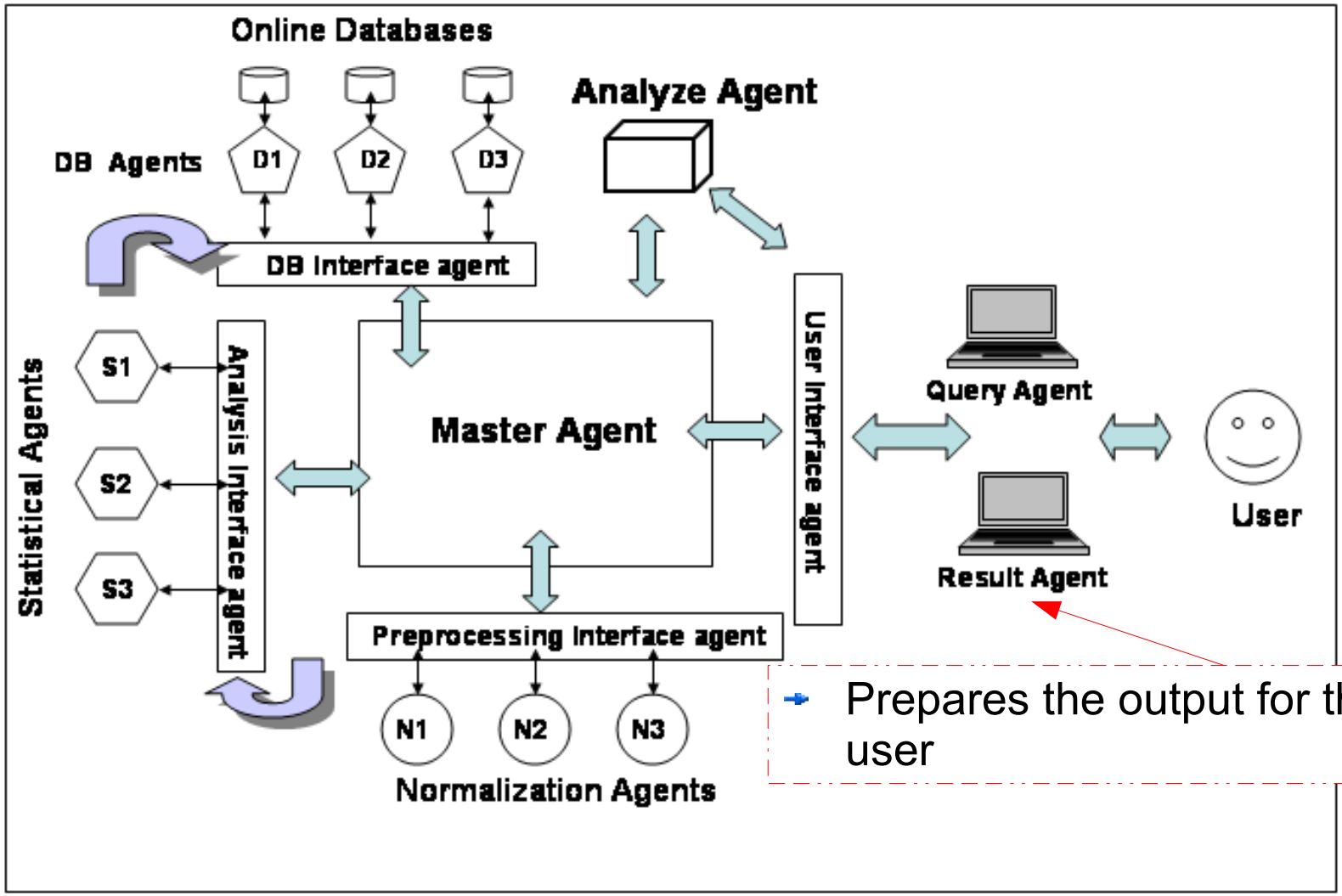


A Multi-agent Approach to Gene Expression Analysis



- Invokes the proper DB agent for functional analysis
- Passes the result back to the Master Agent

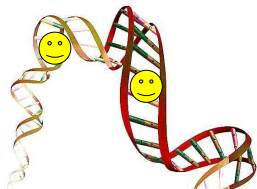
A Multi-agent Approach to Gene Expression Analysis



- Prepares the output for the user

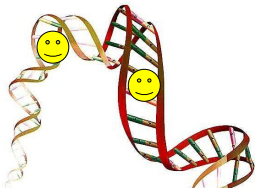
Using Multi-Agent System for Gene Expression Classification

- The MAS is aimed at finding significant classification genes
- Every possible pair of genes defines a point in the search space
- Agents achieve their goals in two steps
  - Exploring the search space using Monte Carlo method
  - Looking for the best solution by following vertical and horizontal lines from the best classifying position

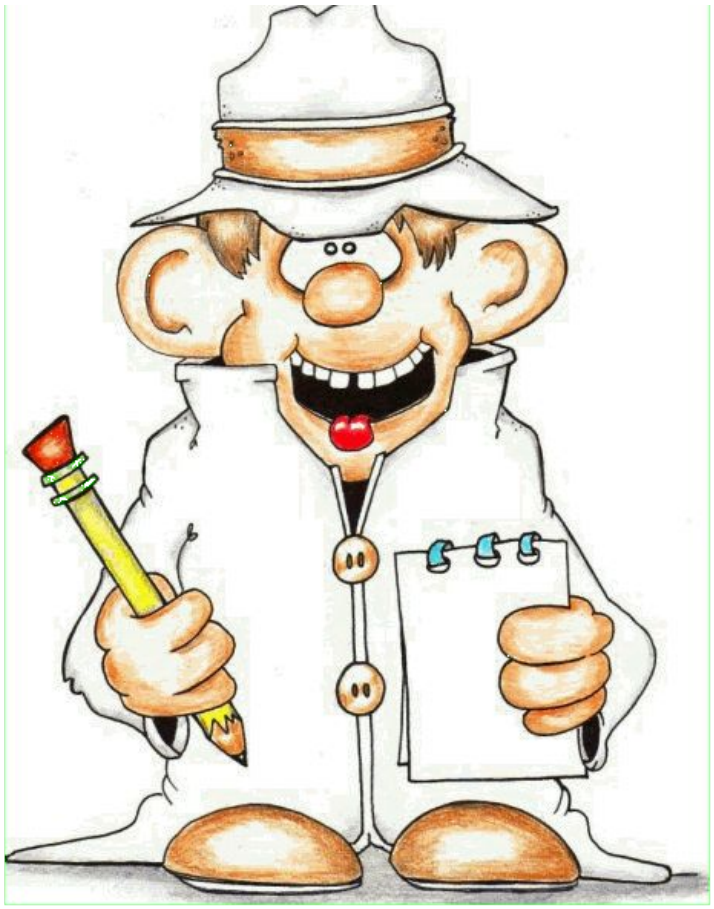


Using Multi-Agent System for Gene Expression Classification

- Fitness level of an agent is expressed by the ability of finding the best combination of genes for classification
- The classification task is performed by adopting the k-NN technique

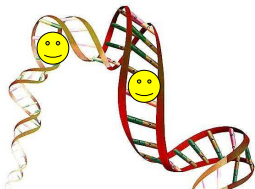


Conclusions



Conclusions

- Agent technology is a polymorphic concept
  - It denotes a variety of different approaches and perspectives
- In bioinformatics the agent technology can be adopted to conceptualize, analyze, design and implement complex software systems

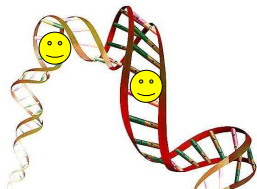


References

- **[Addis07]**
  A. Addis, G. Armano, F. Mascia, and E. Vargiu.
  Protein Secondary Structure Prediction through a Cooperative MultiAgent Learning Approach.
  System and Information Sciences Notes, Vol. 2 N° 1, September 2007, H. Tianfield (ed.), pp. 122—125, 2007.

- **[Altschul90]**
  S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman
  Basic local alignment search tool.
  Journal of Molecular Biology, Vol. 215 N°3, pp. 403–410, 1990.

- **[Armano04]**
  G. Armano
  NXCS Experts for Financial Time Series Forecasting.
  Applications of Learning Classifier Systems, pp. 68—91, 2004.

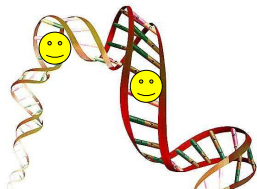


References

- **[Armano06]**
  G. Armano, A. Orro, and E. Vargiu.
  MASSP3: A System for Predicting Protein Secondary Structure.
  EURASIP Journal on Appl. Signal Processing, Special Issue on Advanced Signal Processing Techniques for Bioinformatics, X.W. Chen, S. Kim, V. Pavlovic, D. Casasent (eds.), Vol. 2006, pp. 1—9, 2006.

- **[Armano07]**
  G. Armano, A. Manconi, and E. Vargiu.
  A MultiAgent System for Retrieving Bioinformatics Publications from Web Sources.
  IEEE Transactions on Nanobioscience, Special Session on GRID, Web Services, Software Agents and Ontology Applications for Life Science, Vol. 6 N° 2, June 2007, pp. 104—109, 2007.

- **[Bortolussi07]**
  L. Bortolussi, A. Dovier, and F. Fogolari.
  Agent-based protein structure prediction.
  Multiagent Grid Syst. 3, 2 (Apr. 2007), 183-197, 2007.

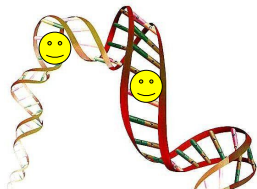


References

- **[Bryson01]**
  K. Bryson, M. Luck, M. Joy, and D.T. Jones.
  Agent Interaction for Bioinformatics Data Management.
  Applied Artificial Intelligence, 15(10):917-947, 2001.

- **[Corradini05]**
  F. Corradini, E. Merelli, and M. Vita.
  A multi-agent system for modelling the oxidation of carbohydrate cellular process.
  Proceeding of the *1st International Workshop on Modelling Complex Systems*, pp. 1265-1273, 2005.

- **[Corradini05b]**
  F. Corradini and E. Merelli.
  Hermes: agent-based middleware for mobile computing.
  Tutorial Book of *5th International School on Formal Methods for the Design of Computer, Communication and Software Systems: Mobile Computing*, 2005.



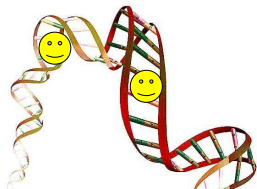
References

- **[Decker02]**
  K.S. Decker, S. Khan, C. Schmidt, G. Situ, R. Makkena, and D. Michaud.
  BioMAS: A Multi-Agent System for Genomic Annotation.
  International Journal Cooperative Information Systems, 11(3), pp. 265-292, 2002.

- **[Decker97]**
  K.S. Decker and K. Sycara.
  Intelligent Adaptive Information Agents.
  Journal of Intelligent Information Systems, 9(3), pp. 239-260, 1997.

- **[Garro04]**
  A. Garro, G. Terracina, and D. Ursino.
  A multi-agent system for supporting the prediction of protein structures.
  Integr. Comput.-Aided Eng. 11, 3 (Aug. 2004), 259-280, 2004.

- **[Graham00]**
  J. Graham and K.S. Decker.
  Towards a Distributed, Environment-Centered Agent Framework.
  Intelligent Agents VI, pp. 290-304, 2000.

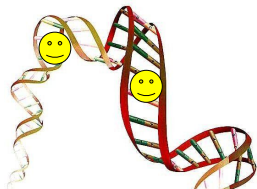


References

- **[Lam06]**
  H.C. Lam, M. Vazquez, B. Juneja, S.C. Fahrenkrug, and D. Boley.
  Gene Expression Analysis in Multi-Agent Environment.
  International Transactions on Systems Science and Applications, Special Issue on Agents in Bioinformatics and Health Informatics, G. Armano and R. Unland (eds.), Vol. 1 N° 1, pp. 35—42, 2006.

- **[Leonardi02]**
  L. Leonardi, M. Mamei, and F. Zambonelli
  A physically grounded approach to coordinate movements in a team.
  In 1st International Workshop on Mobile Teamwork (at ICDCS). IEEE CS, pp. 373—378, 2002.

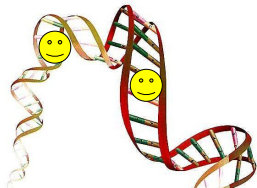
- **[Orro06]**
  A. Orro and L. Milanesi
  An Agent System for Automatic Workflows Composition in Homology Based Functional Genomics.
  International Transactions on Systems Science and Applications, Special Issue on Agents in Bioinformatics and Health Informatics, G. Armano and R. Unland (eds.), Vol. 1 N° 1, pp. 59-66, 2006.



References

- **[Stevens03]**
  R. Stevens, A. Robinson, C. Goble
  myGrid: Personalized Bioinformatics on the Information Grid.
  Bioinformatics, 19(Suppl 1) pp. I302-I304, 2003.

- **[Stiglic04]**
  G. Stiglic and P. Kokol.
  Using multiagent system for gene expression classification.
  Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE, Vol. 2, pp. 2952-2955, 2004.



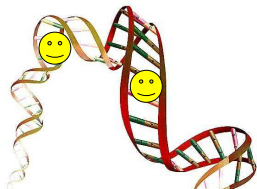
Further Readings

- **[Armano07b]**
  G. Armano.
  Why software agents can be effective in biomedical science.
  Multiagent Grid Syst. 3, 2 (Apr. 2007), 167-172, 2007.

- **[Baxevanis06]**
  A.D. Baxevanis and B. F. F. Ouellette.
  Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins.
  Wiley Higher Education, Third Edition, 2006.

- **[Merelli05]**
  E. Merelli, G. Armano, N. Cannata, F. Corradini, M. d'Inverno, A. Doms, P. Lord, A. Martin, L. Milanesi, S. Möller, M. Schroeder, and M.Luck.
  Agents in bioinformatics, computational and systems biology.
  Briefings in Bioinformatics, 2005.

- **[Valle06]**
  G. Valle, M. Helmer Citterich, M. Attimonelli, and G. Pesole.
  Introduzione alla bioinformatica.
  Zanichelli, Fourth Edition, 2006.



That's all folks!

Thanks for your attention!

Contact Information:

vargiu@diee.unica.it

http://iasc.diee.unica.it

Acknowledaments

ITALBIONET