

# The Tropical Disease Initiative

*Comparative docking on protein structure models from ten tropical disease genomes.*



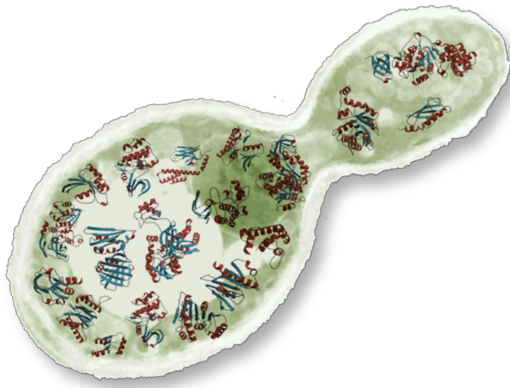
**Marc A. Marti-Renom**

<http://sgu.bioinfo.cipf.es>

Structural Genomics Unit  
Bioinformatics Department

Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain



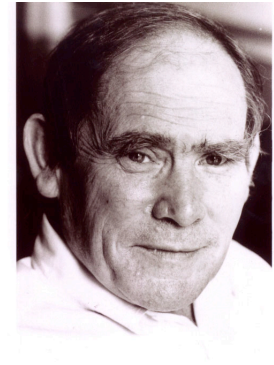


**Bioinformatics @ CIPF**  
**Introduction to CM**  
**Modeling genomes**  
**Modeling Genes**  
**Tropical Disease Initiative**

# Data in the post-genomic era

Progress in science depends on new techniques, new discoveries and new ideas, **probably in that order.**

*Sydney Brenner, 1980*



The introduction and popularization of high-throughput techniques has drastically changed the way in which biological problems **can** be addressed and hypotheses can be tested.

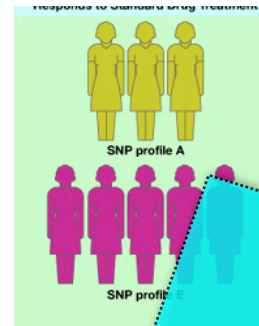
**But not necessarily the way in which we really address or test them...**

Genes in the DNA...



...code for proteins...

>protein kinase  
acctgttgatggcgacagggactgtatgctg  
atctatgctgatgcatgcatgctgactactgat

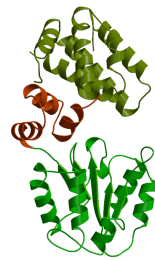


...produces the final phenotype

# From genotype to phenotype.

...whose structure accounts

...plus the environment...





...which can be different because of the variability. 10 million SNPs



...whose final effect configures the phenotype...

## Genes in the DNA...

Now: 22240 (NCBI build 35 12/04)  
50-70% display alternative splicing  
25%-60% unknown

Transfrags

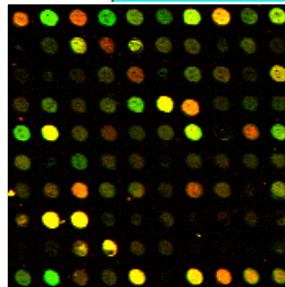
>protein kinase

```
acctgttgatggcgacagggactgtatgctgatctat  
gctgatgcatgcatgctgactactgatgtgggggcta  
ttgacttgatgtctatc....
```



...when expressed in the proper moment and place...

A typical tissue is expressing among 5000 and 10000 genes



## From genotype to phenotype.

(post-genomics scenario)

...code for proteins...

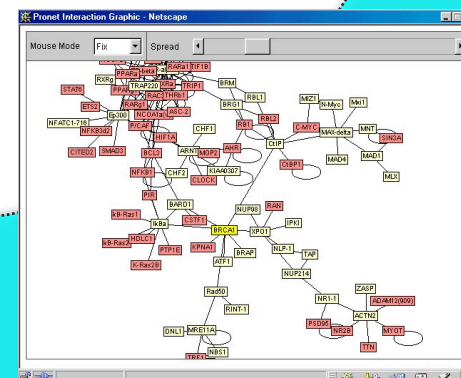
That undergo post-translational modifications, somatic recombination...

100K-500K proteins



...whose structures account for function...

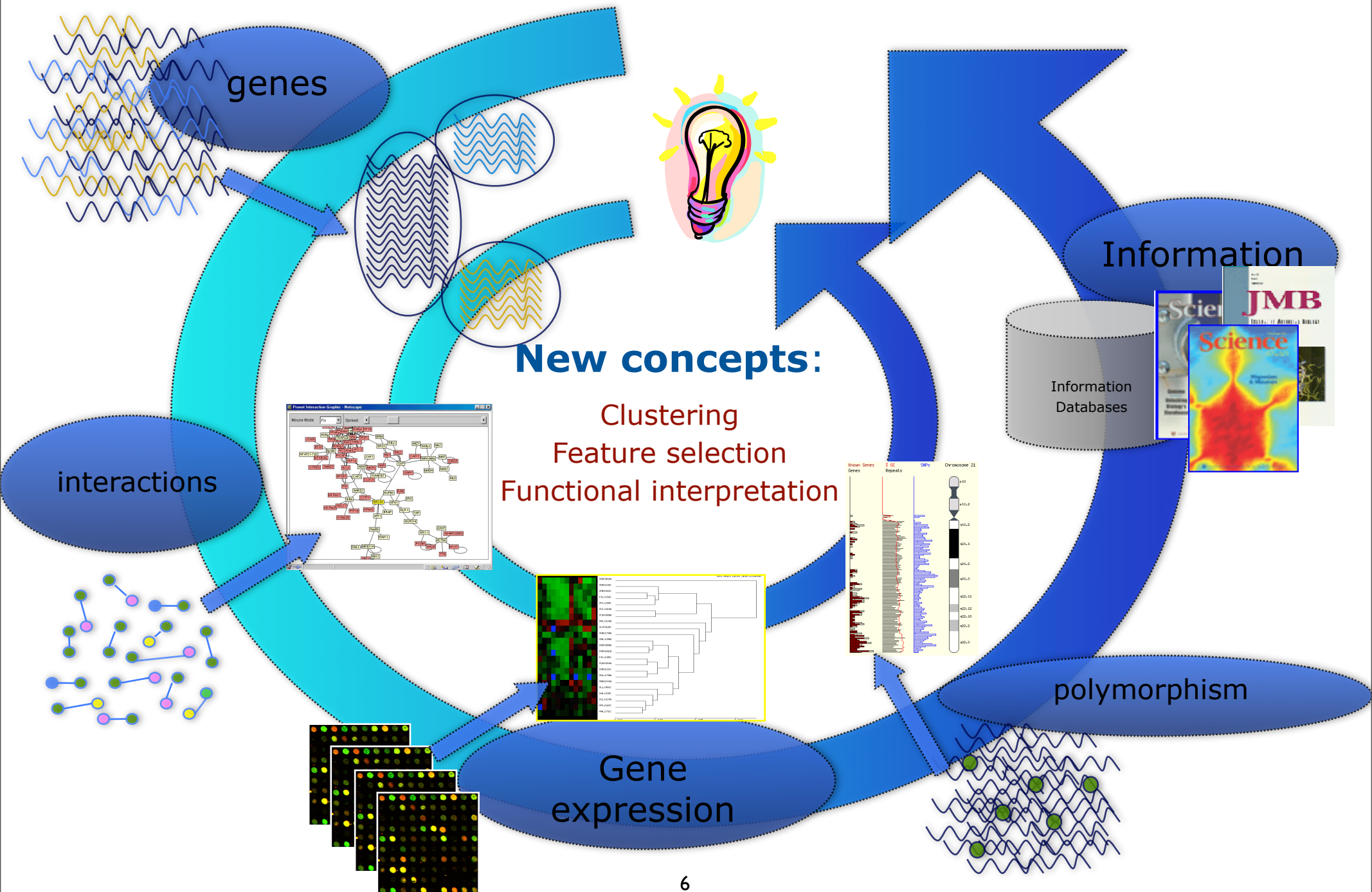
...conforming complex interaction networks...



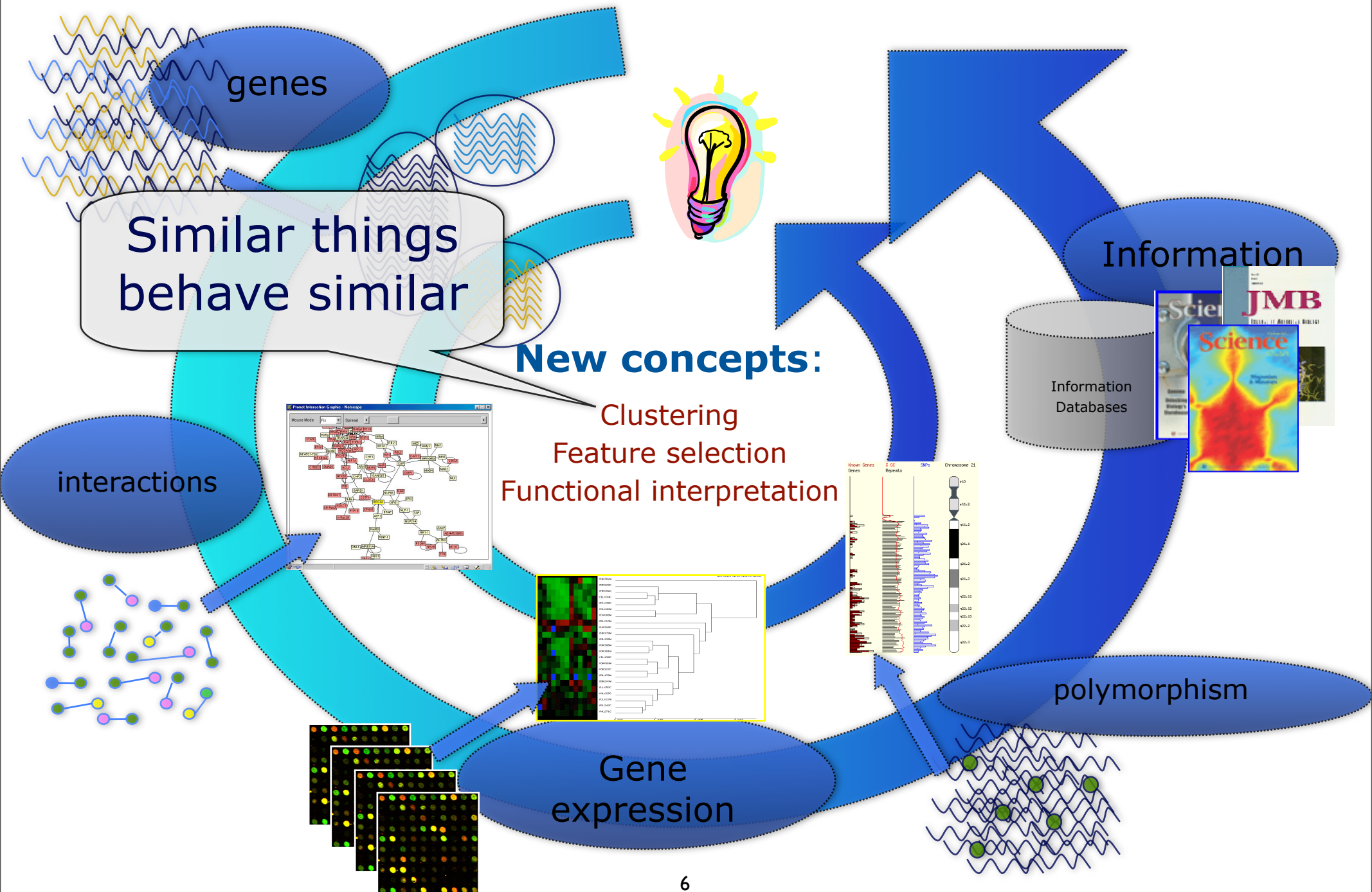
...in cooperation with other proteins...

Each protein has an average of 8 interactions

# Similar things behave similar...

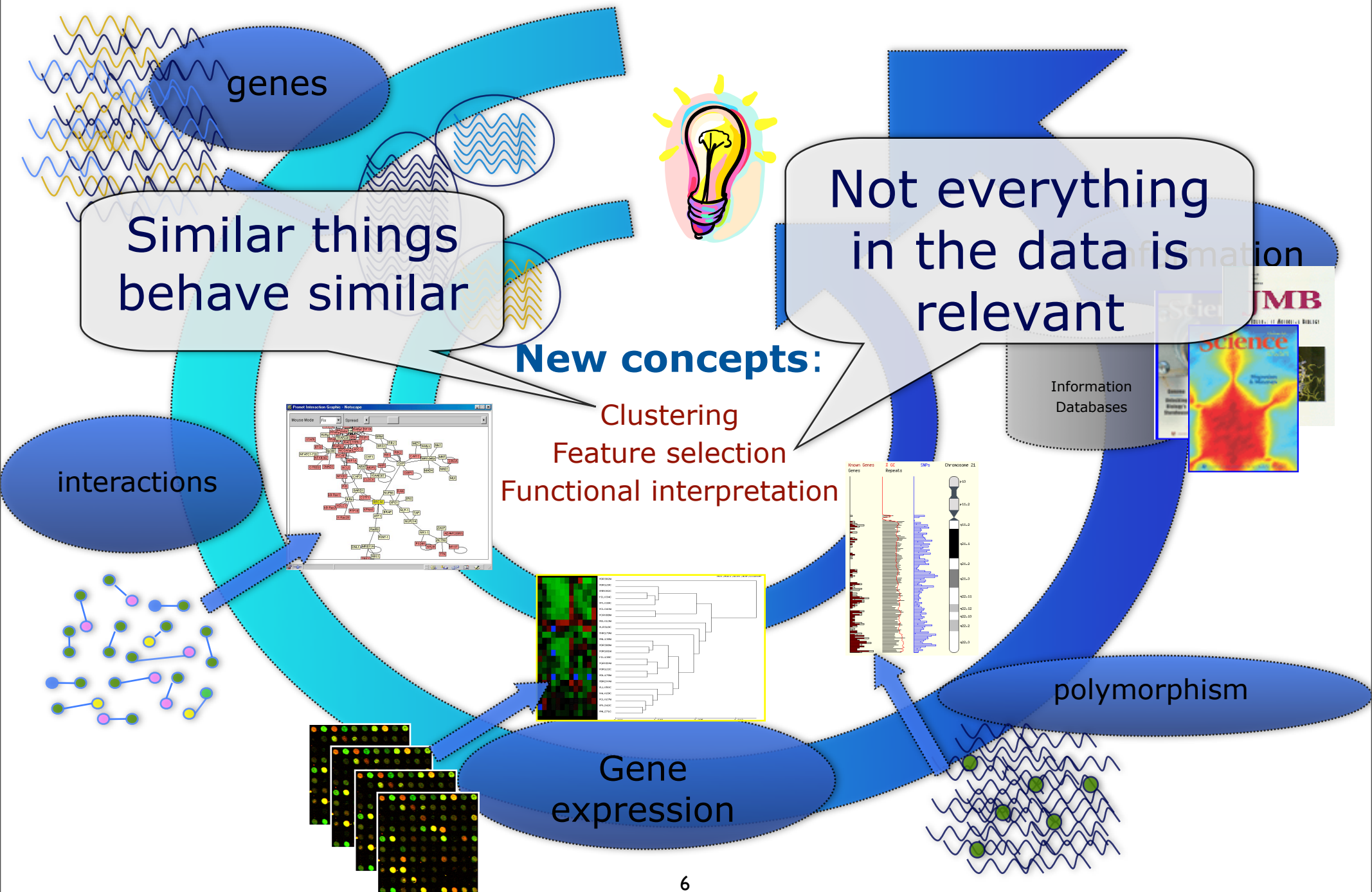


# Similar things behave similar...





# Similar things behave similar...



# Similar things behave similar...

genes

Similar things  
behave similar



Not everything  
in the data is  
relevant

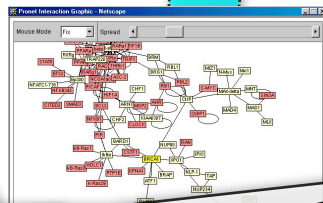
New concepts:

Clustering

Feature selection

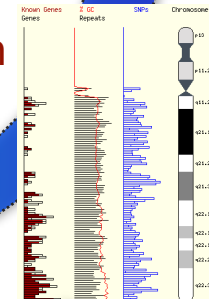
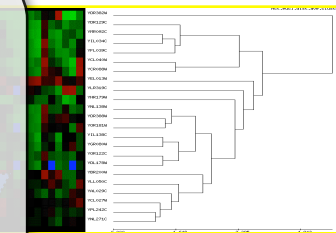
Functional interpretation

interactions

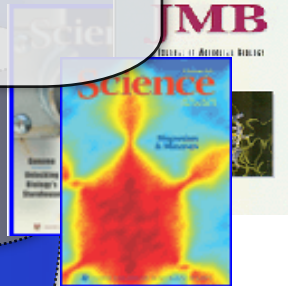


The data has to  
be transformed  
into biological  
information

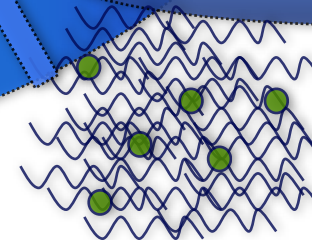
Gene  
expression



Information  
Databases



polymorphism



# Bioinformatics Department <http://bioinfo.cipf.es>

Bioinformatics at CIPF

<http://bioinfo.cipf.es/> Google

### The Department

People  
Functional Genomics Unit  
Pharmacogenomics & Comparative Genomics Unit  
Structural Genomics Unit

### Tools

DNA array data analysis  
SNPs data analysis  
Functional profiling  
Downloads

### Documents & Publications

Papers  
Communications  
Supplementary material

### Meetings & Courses

Meetings & workshops  
Courses  
On line courses  
Accommodation

 **CAMDA 2007, Dec 13-14**  
Seventh international conference for the Critical Assessment of Microarray Data Analysis

### Coming events...

  
**PRINCIPE FELIPE**  
CENTRO DE INVESTIGACION

 **INB** INSTITUTO NACIONAL DE BIOINFORMATICA

 **CEGEN**  
Centro Nacional de Genotipado

 **ciberer**  
Centro de Investigación Biomédica En Red de Enfermedades Raras

 **GECOBIO**  
VALENCIA

 Red Temática Nacional de Bioinformática

 **INDIGO**  
Integrated Highly Sensitive Fluorescence-based Biosensor for Diagnostic applications

### Bioinformatics

DNA array analysis SNPs analysis Functional Profiling

Today's research in biomedicine cannot be understood without the concurrence of bioinformatics. Our department aims to tackle biomedical problems from a system's biology perspective. Following this, the general objective we seek through the main lines of research is to relate the mutations (Pharmacogenomics and Comparative Genomics) to their effect at cellular and phenotypic level (Functional Genomics) trying to understand the mechanism of action (Structural Genomics).

### News

**May 2007** - Positions available: one biostatistician and one bioinformatician  
**April 2007** - Blast2GO, a universal Gene Ontology annotation, visualization and analysis suite for functional genomics research.  
**January 2007** - Phylemon, a suite of tools for molecular evolution, phylogenetics and phylogenomics.  
**September 2006** - Prophet, a tool for building a class predictor.  
**March 2006** - Computing journal award to the best R&D project.  
**February 2006** - New releases: GEPAS v3.0 and Babelomics v2.0. Also PupaSuite, interactive selection of optimal sets of SNPs for large-scale genotyping.  
**December 2005** - CIPF receives autelsi award: "The First European Cluster for Scientific Computing with Free software on GNU/Linux."

### More news...

Google    
☐ Search WWW ☐ Search bioinfo.cipf.es

**Postal address**  
Bioinformatics Department (CIPF)  
Av. Autopista del Saler, 16  
(Camino de las moreras)  
46013 Valencia, Spain  
Tel.: + 34 96 328 96 80  
Fax: + 34 96 328 97 01  
(see map)

Send comments to the [webmaster](#). Last updated: January 15, 2007

### Spotlight Tools

 **Access**  
Tool usage around the World.

 **Prophet**  
A tool for building a class predictor

 **PupaSuite**  
Interactive selection of optimal sets of SNPs for large-scale genotyping.

 **MARMITE**  
Functional profiling with PubMed words.

 **Fatigo+**  
Functional interpretation of large-scale experiments using GO, KEGG, Interpro, Transfac, CisRed...

 **CAAT**  
Draw, browse, analyze and validate interactively your hierarchical clustering results.

 **FatiScan**  
Detect blocks of functionally related genes (GO, KEGG) with significant coordinate (although modest) over- or under-expression.

 **SIDE**  
Interactive design of Small interfering RNA.

 **Blast2GO**  
A universal Gene Ontology annotation, visualization and analysis suite for functional genomics research.

### Packages

 **GEPAS**  
Gene Expression Pattern Analysis Suite (v3.0).

 **BABELOMICS**  
Suite of tools for Functional Profiling (v2.0).

 **Phylemon**  
Suite of tools for molecular evolution & phylogenetics (v1.0).

 **Data Analysis and Visualization in Genomics and Proteomics**  
Francisco Azuaje, Joaquín Dopazo (Editors)  
ISBN: 0-470-09439-7



# Bioinformatics Department <http://bioinfo.cipf.es>



**Functional Genomics**  
Dr. Joaquín Dopazo

## GEPAS



## BABELOMICS

<http://gepas.bioinfo.cipf.es>

Gene Expression Pattern Analysis Suite

<http://pupasuite.bioinfo.cipf.es>

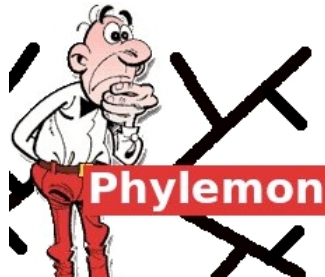
SNP Analysis Suite

<http://babelomics.bioinfo.cipf.es>

Functional Profiling Analysis Suite



**Comparative Genomics**  
**Pharmacogenomics**  
Dr. Hernán Dopazo



<http://phylemon.bioinfo.cipf.es>

Molecular Evolution Analysis Suite

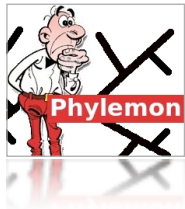


**Structural Genomics**  
Dr. Marc A. Marti-Renom

**DBAli**<sub>v2.0</sub>  
Home

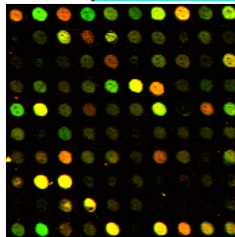
<http://www.dbali.org>

Structural Biology Analysis Suite



>protein kinase

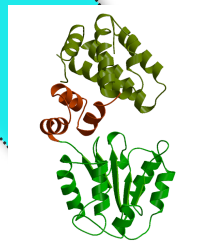
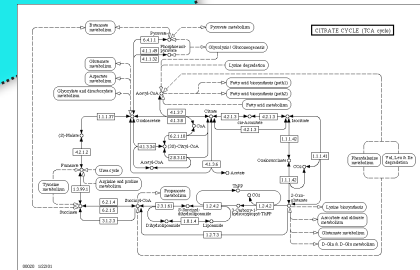
```
acctgttgatggcgacagggactgtatgt  
gatctatgctgatgcgatgctgactact
```



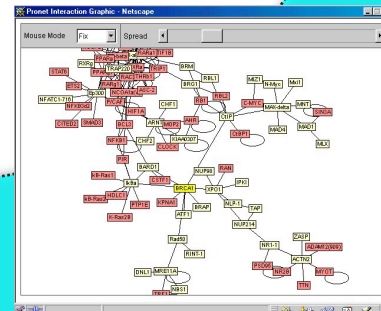
**From genotype to  
phenotype.**

(post-genomics scenario)

**GEPAS**



**DBAli**  
v2.0  
Home



**BABELOMICS**

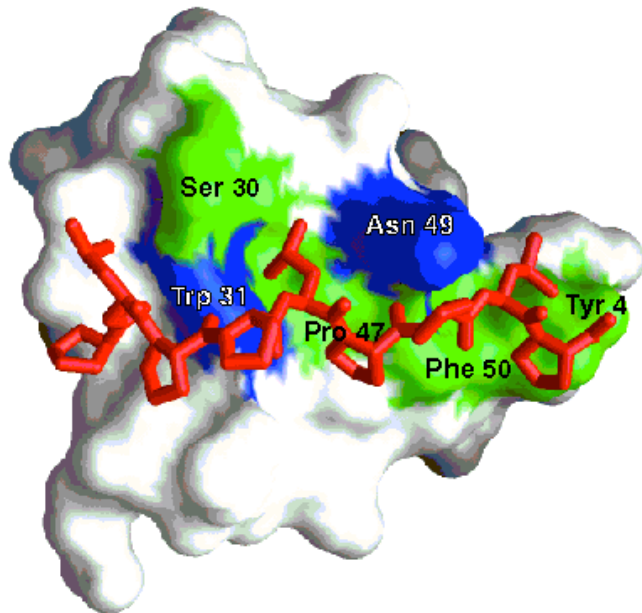
# Why is it useful to know the **structure** of a protein, not only its sequence?

- ◆ The biochemical function (activity) of a protein is defined by its interactions with other molecules.
- ◆ The biological function is in large part a consequence of these interactions.
- ◆ The 3D structure is more informative than sequence because interactions are determined by residues that are close in space but are frequently distant in sequence.

YDL117W  
(15-64)

10 20 30 40 50

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

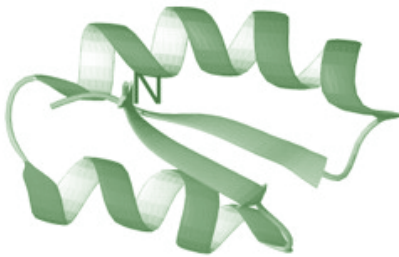


In addition, since evolution tends to conserve function and function depends more directly on structure than on sequence, **structure is more conserved in evolution than sequence.**

The net result is that **patterns in space are frequently more recognizable than patterns in sequence.**

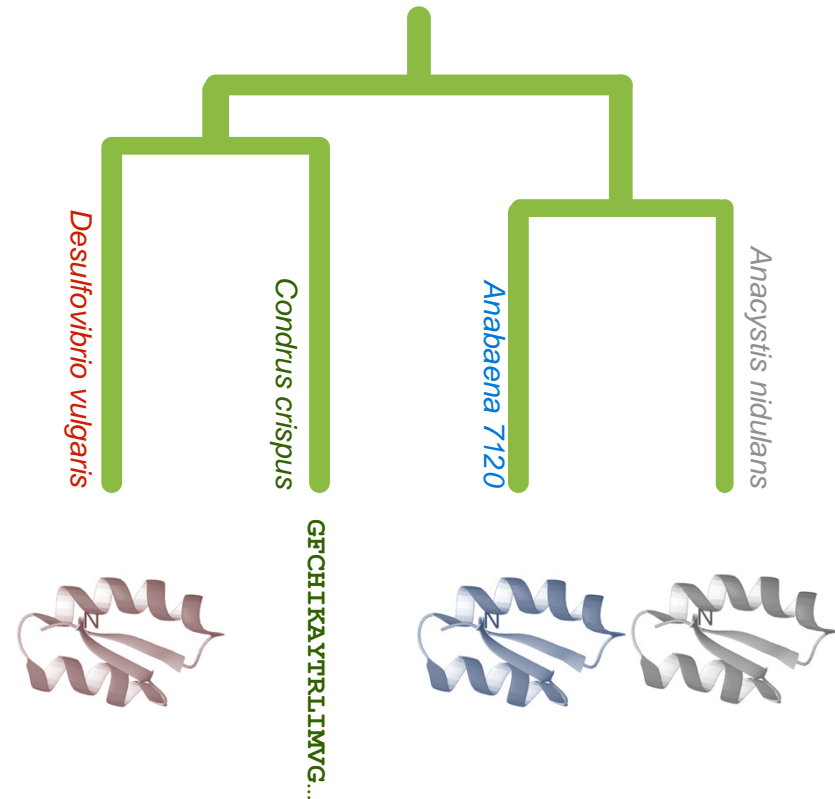
# Principles of protein structure

GFCHIKAYTRLIMVG...



Folding (physics)

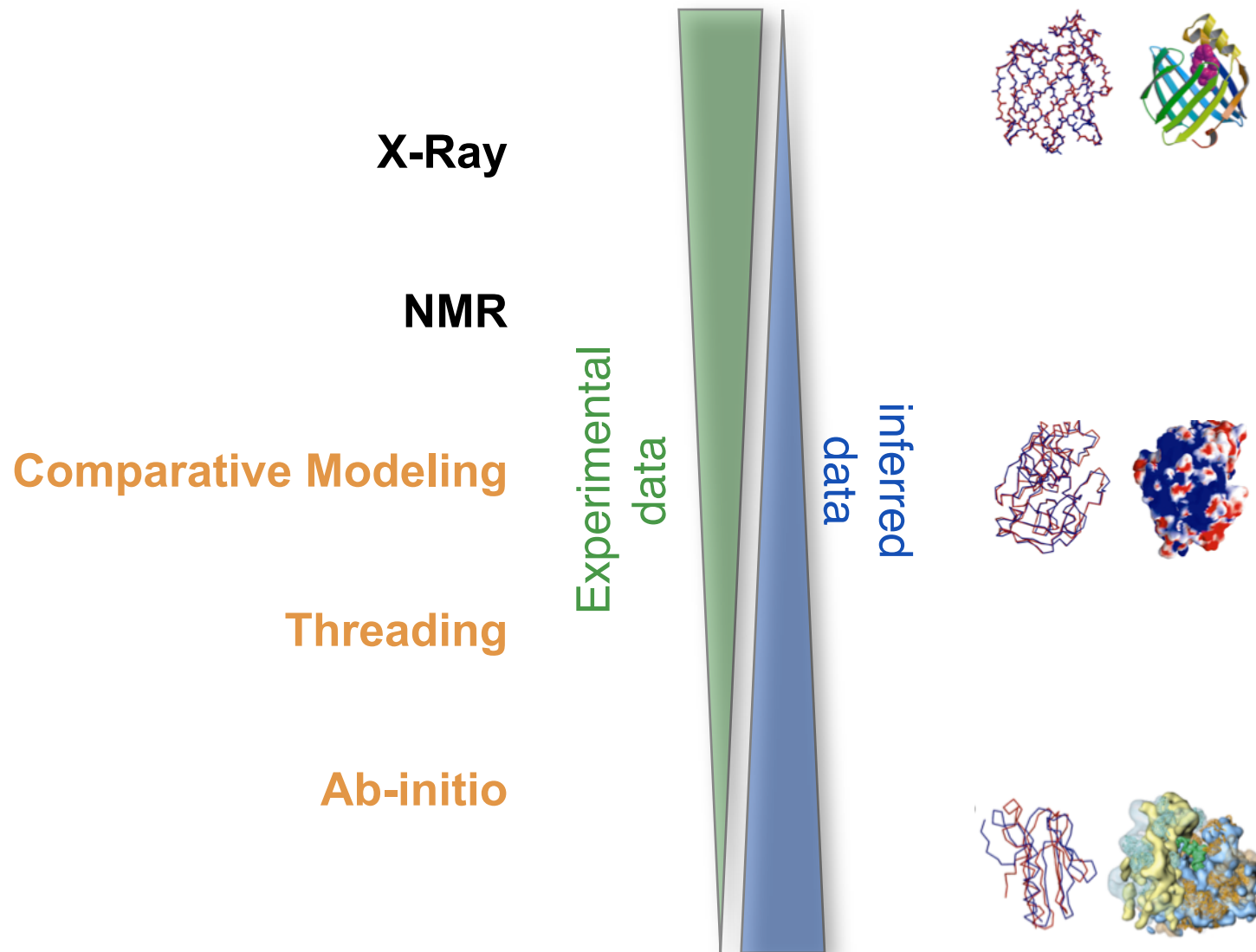
*Ab initio* prediction



Evolution (rules)

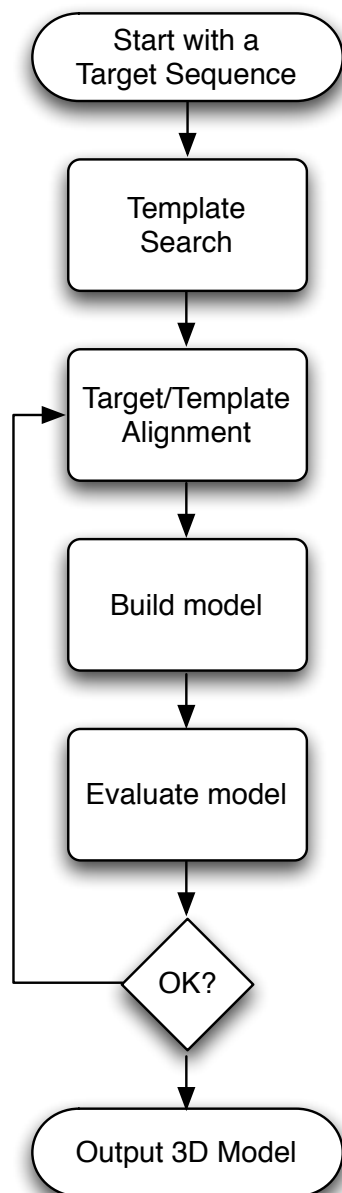
Threading  
Comparative Modeling

# protein prediction **vs** protein determination



# Comparative modeling by satisfaction of spatial restraints

## MODELLER



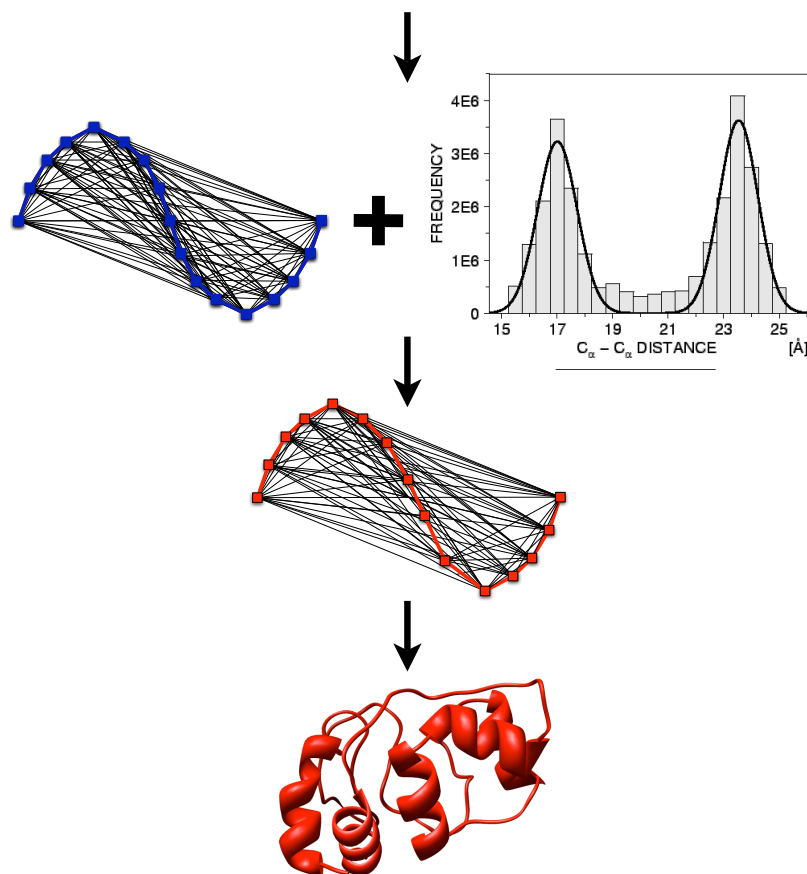
Given an alignment...

extract spatial features  
from the template(s)  
and statistics from  
known structures

apply these features  
as restraints on your  
target sequence

optimize to find the  
best solution for the  
restraints to produce  
your 3D model

MSVIPKR--GNCEQTSE  
ASILPKRLFGNCEQTSD

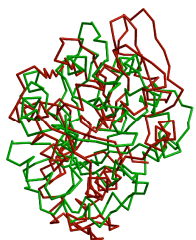


A. Šali & T. Blundell. *J. Mol. Biol.* 234, 779, 1993.  
J.P. Overington & A. Šali. *Prot. Sci.* 3, 1582, 1994.  
A. Fiser, R. Do & A. Šali, *Prot. Sci.*, 9, 1753, 2000.

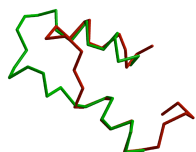


# Comparative modeling by satisfaction of spatial restraints

## Types of errors and their impact



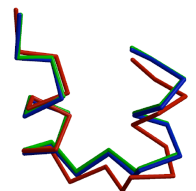
**Wrong fold**



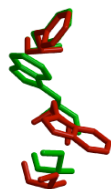
**Miss alignments**



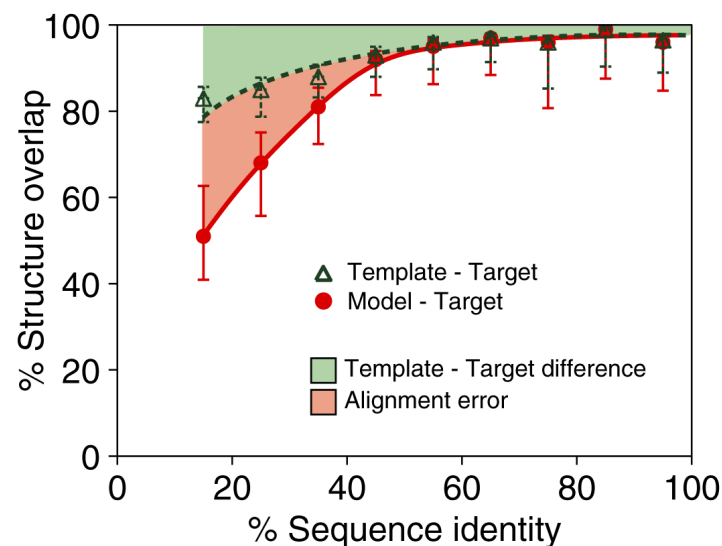
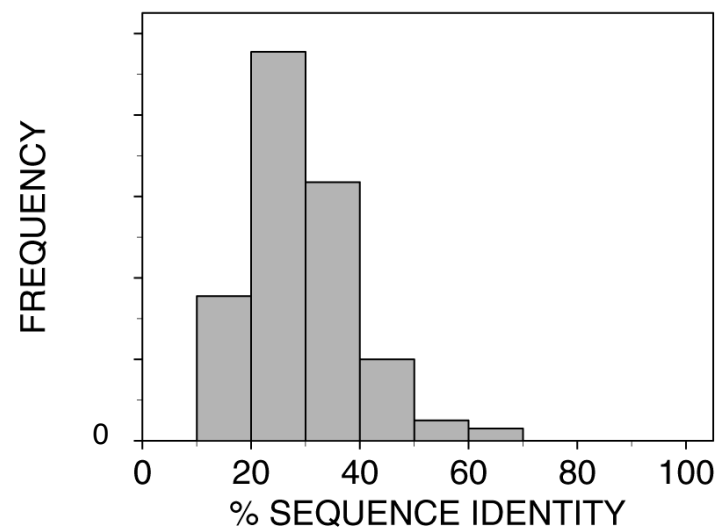
**Loop regions**



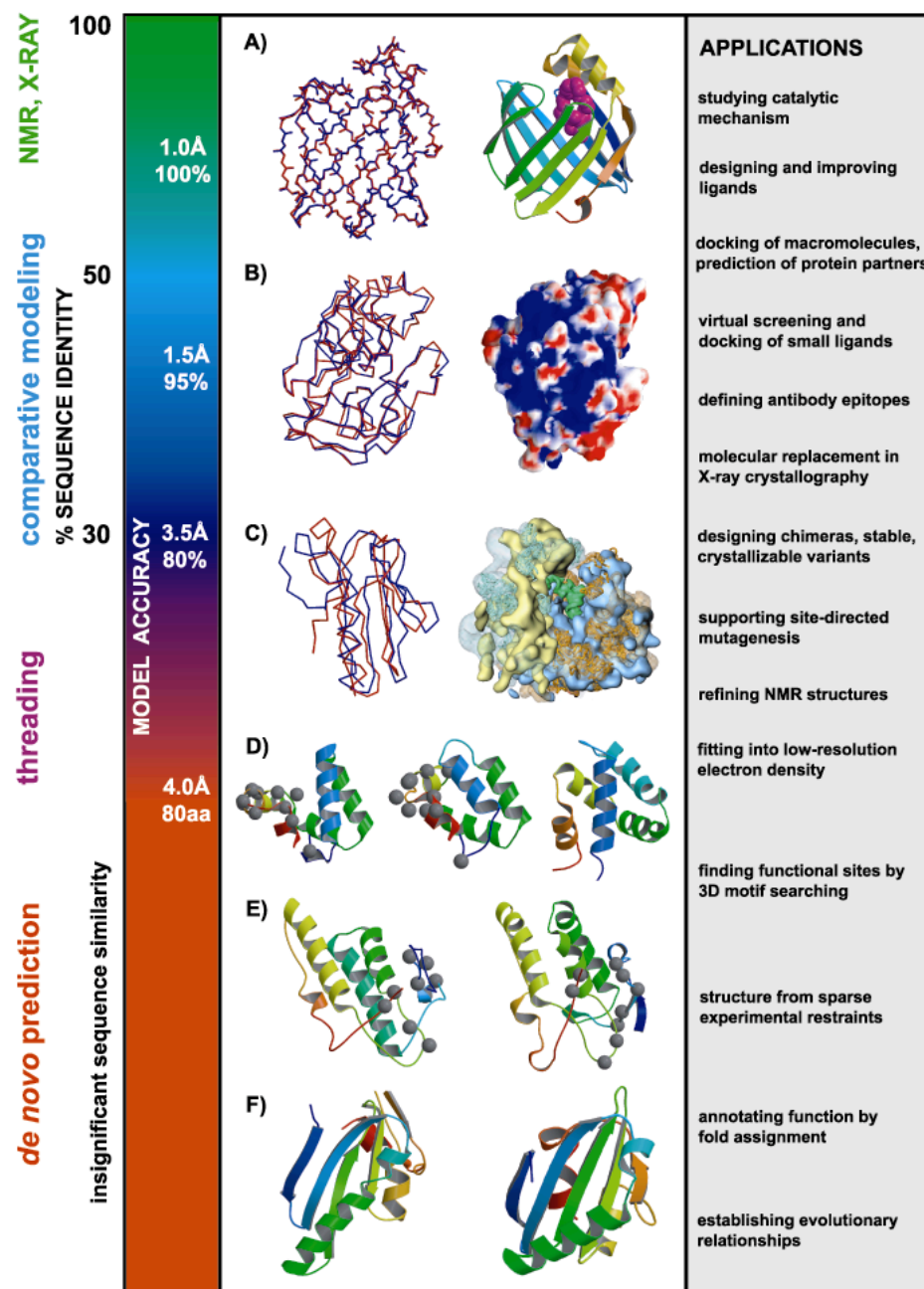
**Rigid body distortions**



**Side-chain packing**



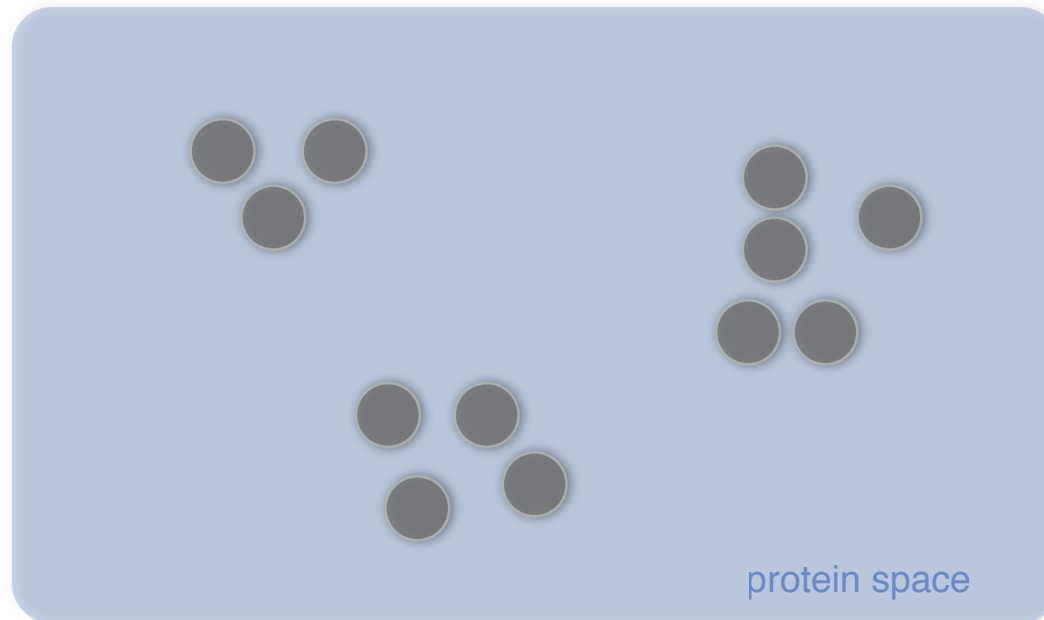
# Utility of protein structure models, despite errors



# Structural Genomics

Characterize most protein **sequences** based on related known **structures**

1. The number of “**families**” is much **smaller** than the number of proteins.
2. **Any one** of the members of a family is **fine**.



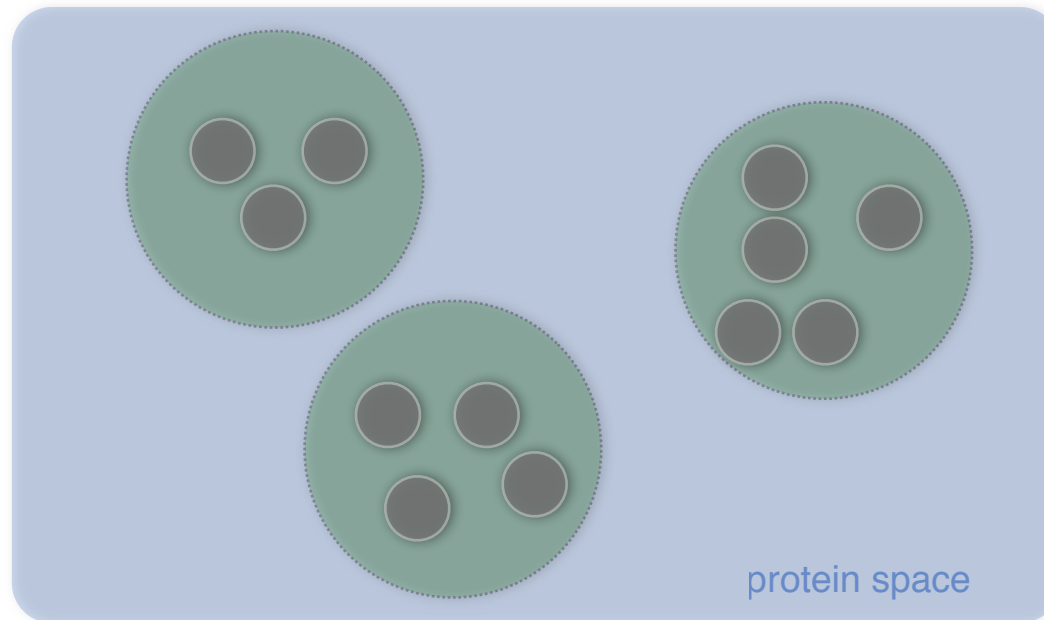
There are **~16,000** families (90%)  
@ 30% sequence identity cutoff

*Sali. Nat. Struct. Biol.* **5**, 1029, 1998.  
*Sali et al. Nat. Struct. Biol.*, **7**, 986, 2000.  
*Sali. Nat. Struct. Biol.* **7**, 484, 2001.  
*Baker & Sali. Science* **294**, 93, 2001.  
*Vitkup et al. Nat. Struct. Biol.* **8**, 559, 2001

# Structural Genomics

Characterize most protein **sequences** based on related known **structures**

1. The number of “**families**” is much **smaller** than the number of proteins.
2. **Any one** of the members of a family is **fine**.



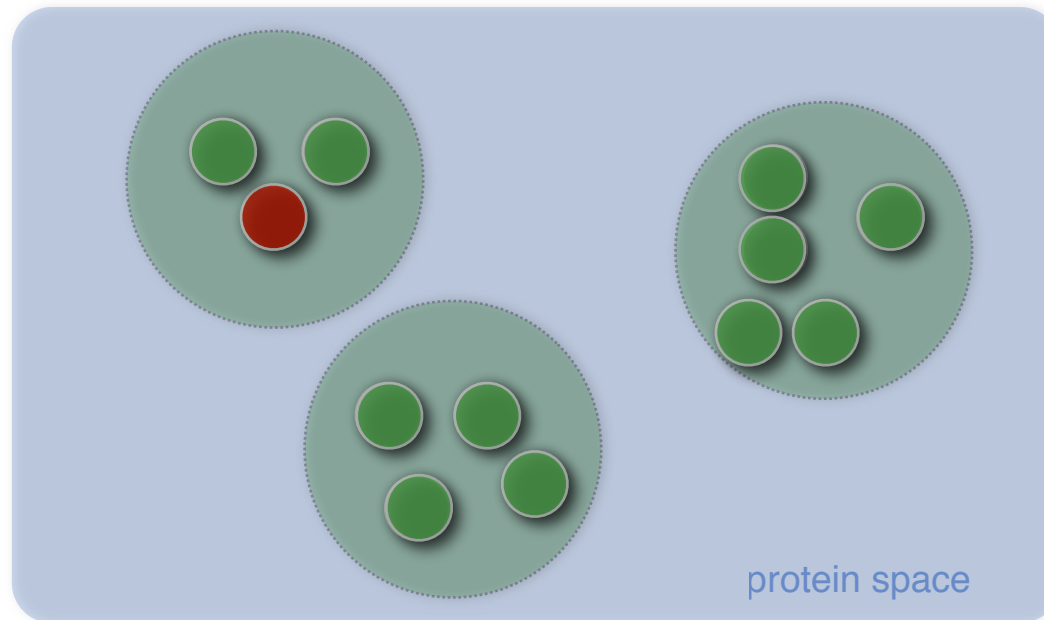
There are **~16,000** families (90%)  
@ 30% sequence identity cutoff

*Sali. Nat. Struct. Biol.* **5**, 1029, 1998.  
*Sali et al. Nat. Struct. Biol.*, **7**, 986, 2000.  
*Sali. Nat. Struct. Biol.* **7**, 484, 2001.  
*Baker & Sali. Science* **294**, 93, 2001.  
*Vitkup et al. Nat. Struct. Biol.* **8**, 559, 2001

# Structural Genomics

Characterize most protein **sequences** based on related known **structures**

1. The number of “**families**” is much **smaller** than the number of proteins.
2. **Any one** of the members of a family is **fine**.



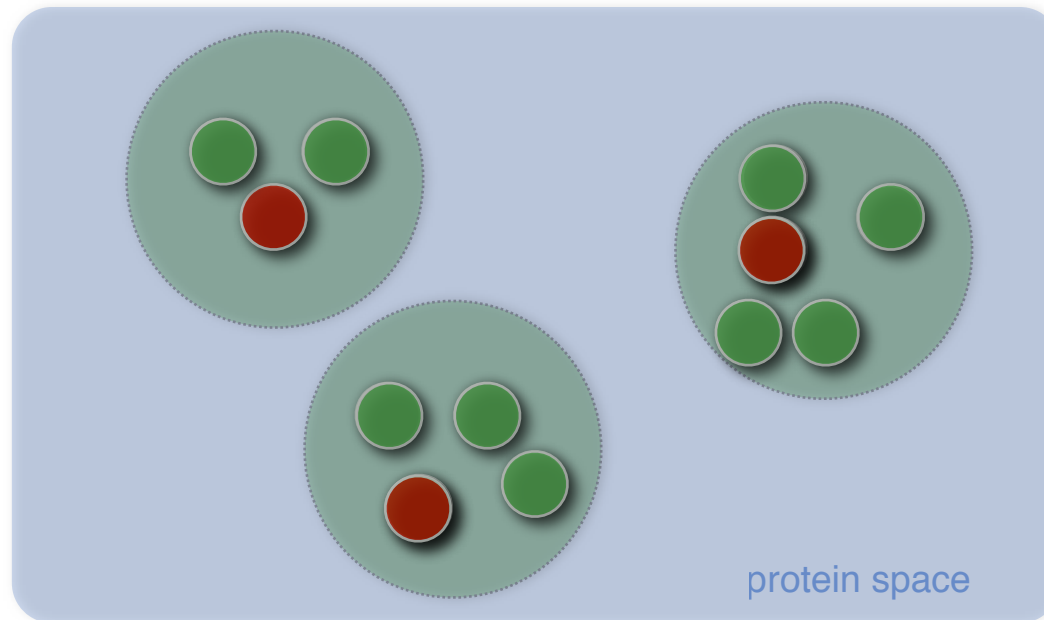
There are **~16,000** families (90%)  
@ 30% sequence identity cutoff

*Sali. Nat. Struct. Biol.* **5**, 1029, 1998.  
*Sali et al. Nat. Struct. Biol.*, **7**, 986, 2000.  
*Sali. Nat. Struct. Biol.* **7**, 484, 2001.  
*Baker & Sali. Science* **294**, 93, 2001.  
*Vitkup et al. Nat. Struct. Biol.* **8**, 559, 2001

# Structural Genomics

Characterize most protein **sequences** based on related known **structures**

1. The number of “**families**” is much **smaller** than the number of proteins.
2. **Any one** of the members of a family is **fine**.



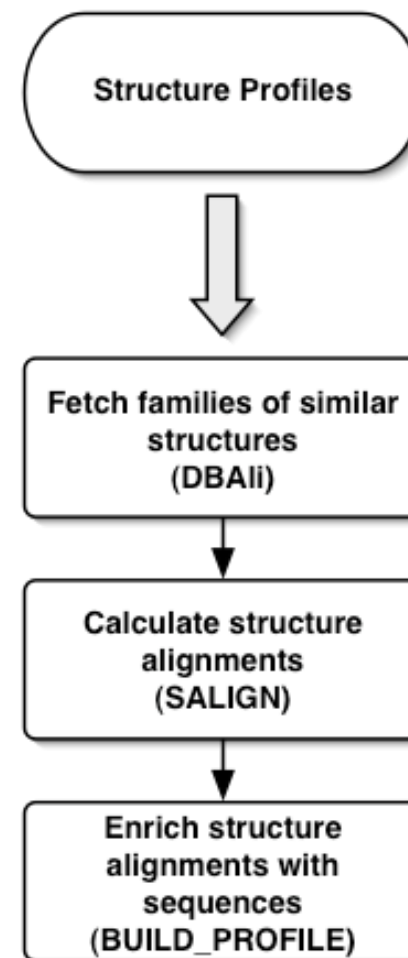
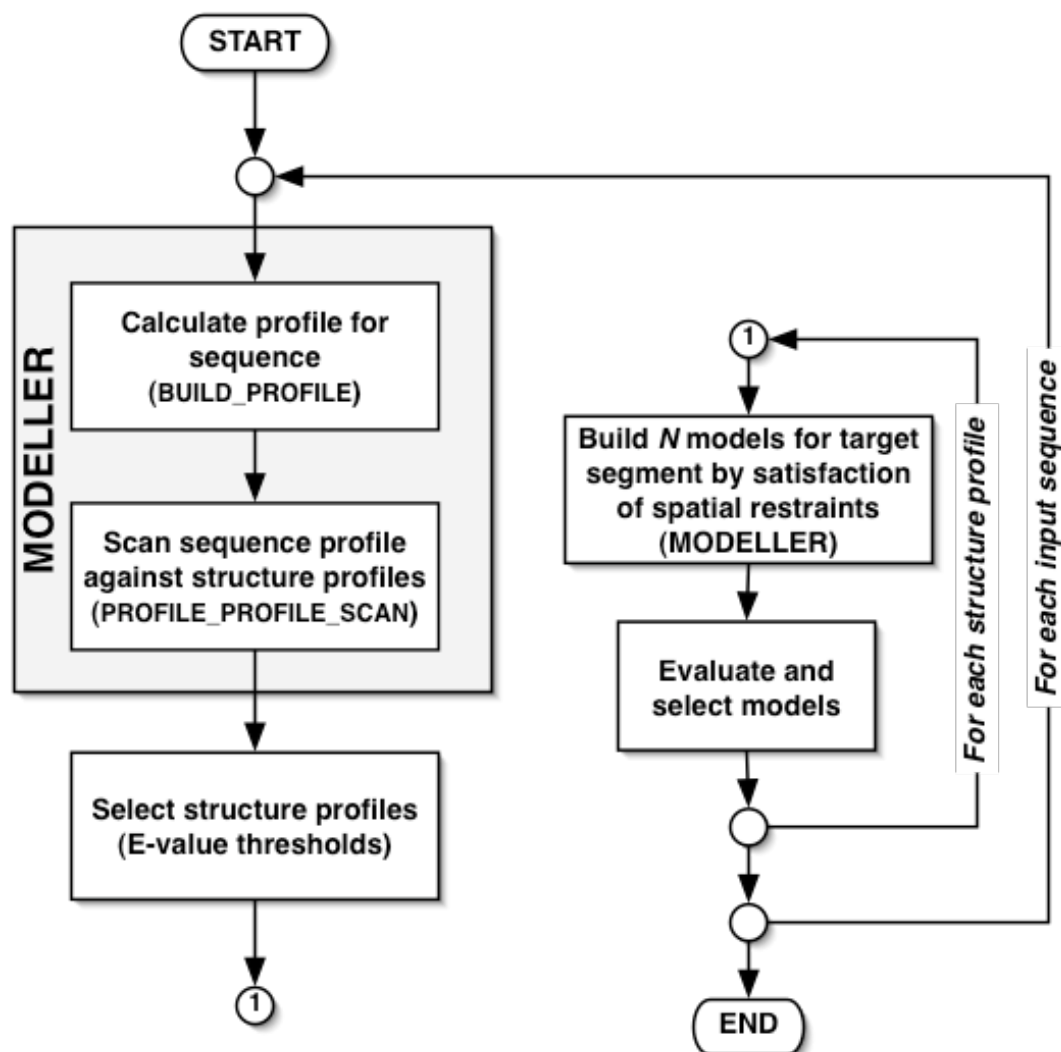
There are **~16,000** families (90%)  
@ 30% sequence identity cutoff

*Sali. Nat. Struct. Biol.* **5**, 1029, 1998.  
*Sali et al. Nat. Struct. Biol.*, **7**, 986, 2000.  
*Sali. Nat. Struct. Biol.* **7**, 484, 2001.  
*Baker & Sali. Science* **294**, 93, 2001.  
*Vitkup et al. Nat. Struct. Biol.* **8**, 559, 2001



# MODPIPE2.0

## Large-Scale Protein Structure Modeling



Eswar et.al., (2003) Nucl.Acids.Res. 31(13)

# ModBase Statistics

Large-scale modeling of the TrEMBL-SWISSPROT databases

<http://www.salilab.org/modbase/>

Sequences (total)	2,186,210
Sequences (modeled)	1,340,687
Models	4,284,570

The screenshot shows the ModBase Search Page in a web browser. The browser's address bar displays the URL [http://modbase.compbio.ucsf.edu/modbase-cgi-new/search\\_form.cgi](http://modbase.compbio.ucsf.edu/modbase-cgi-new/search_form.cgi). The page features a navigation bar with links to Home, User Login, ModBase Search Page, ModWeb Modeling Server, Help, and Current Logins. The main heading is "Database of Comparative Protein Structure Models" with a subheading "Welcome to ModBase, a database of three-dimensional protein models calculated by comparative modeling." On the left, a sidebar lists various sections: General Information, Statistics, News, Project Pages, Documentation, Authors and Acknowledgements, Publications, Todo List, and Related Resources. The main content area contains a "ModBase search form" with a search button. Below the search form, it states "All available datasets are selected" and provides a link to "Select specific dataset(s)". There is also a "Search by properties" section with dropdown menus for "Property" (set to "Database Accession Number") and "Organism" (set to "ALL"), along with an "Advanced search" link. A "Note" at the bottom left states: "MODBASE contains theoretically calculated models, not experimentally determined structures. The models may contain significant errors." At the very bottom, there is a disclaimer and a request to cite the article in publications.



University of California  
San Francisco

Pieper et al. NAR 34, D291 (2006)

# What is the physiological ligand of Brain Lipid-Binding Protein?

Predicting features of a model that are not present in the template

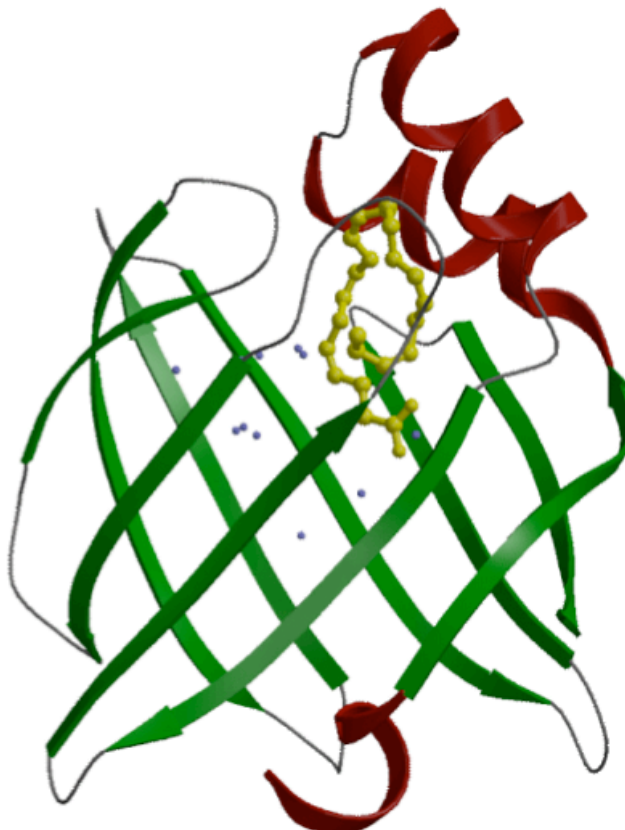
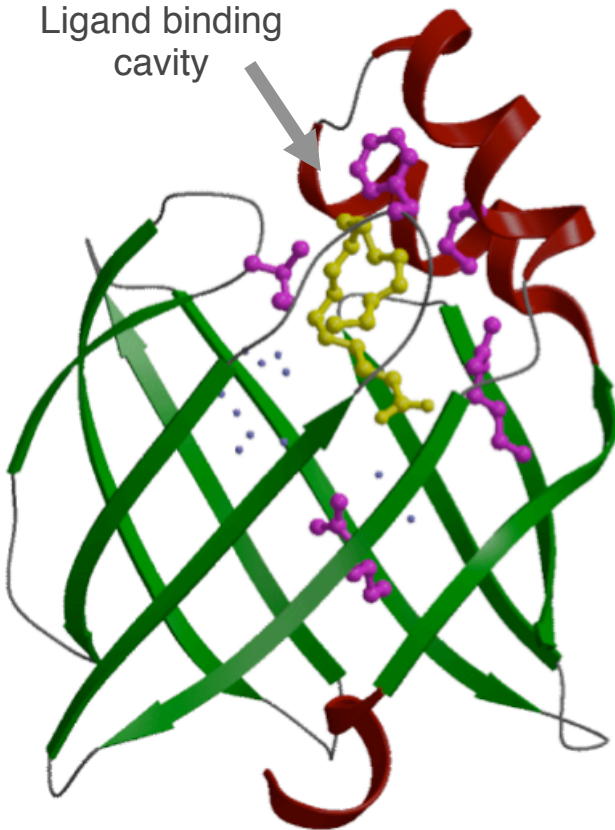
BLBP/oleic acid

BLBP/docosahexaenoic acid

Cavity is **not** filled

Cavity **is** filled

Ligand binding  
cavity



1. BLBP binds fatty acids.

2. Build a 3D model.

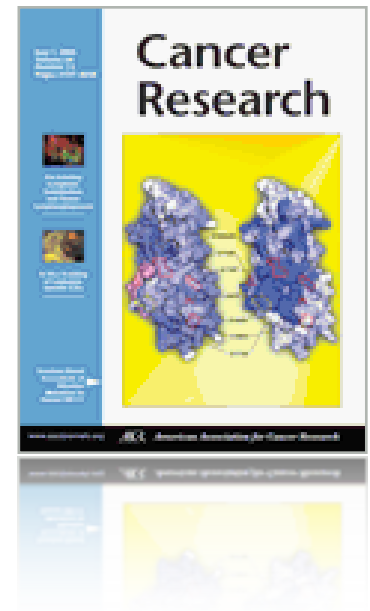
3. Find the fatty acid that fits most snugly into the ligand binding cavity.

# Structural analysis of missense mutations in human BRCA1 BRCT domains

Nebojsa Mirkovic, Marc A. Marti-Renom, Barbara L. Weber,  
Andrej Sali and Alvaro N.A. Monteiro

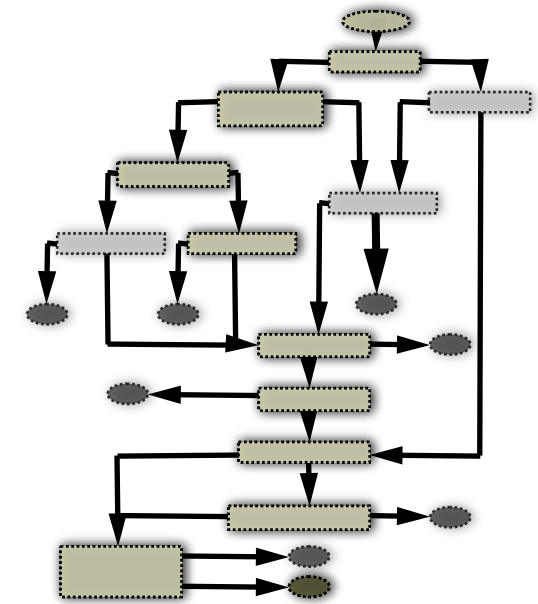
**Cancer Research (June 2004). 64:3790-97**

Cannot measure the functional impact of every  
possible SNP at all positions in each protein!  
Thus, prediction based on general principles of  
protein structure is needed.

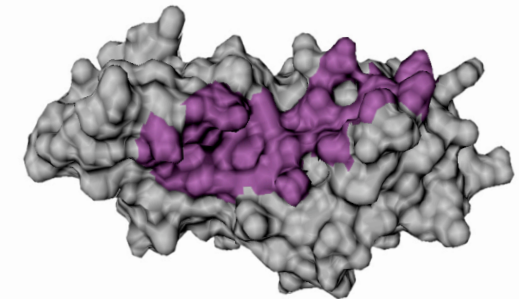
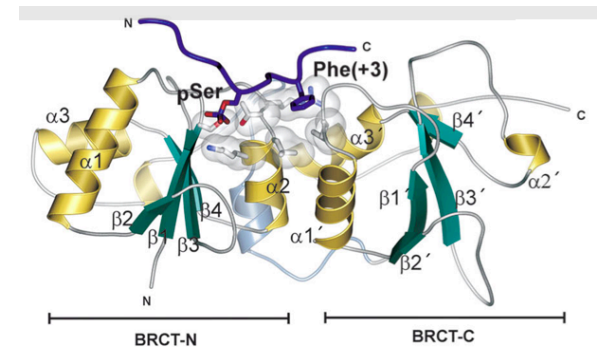
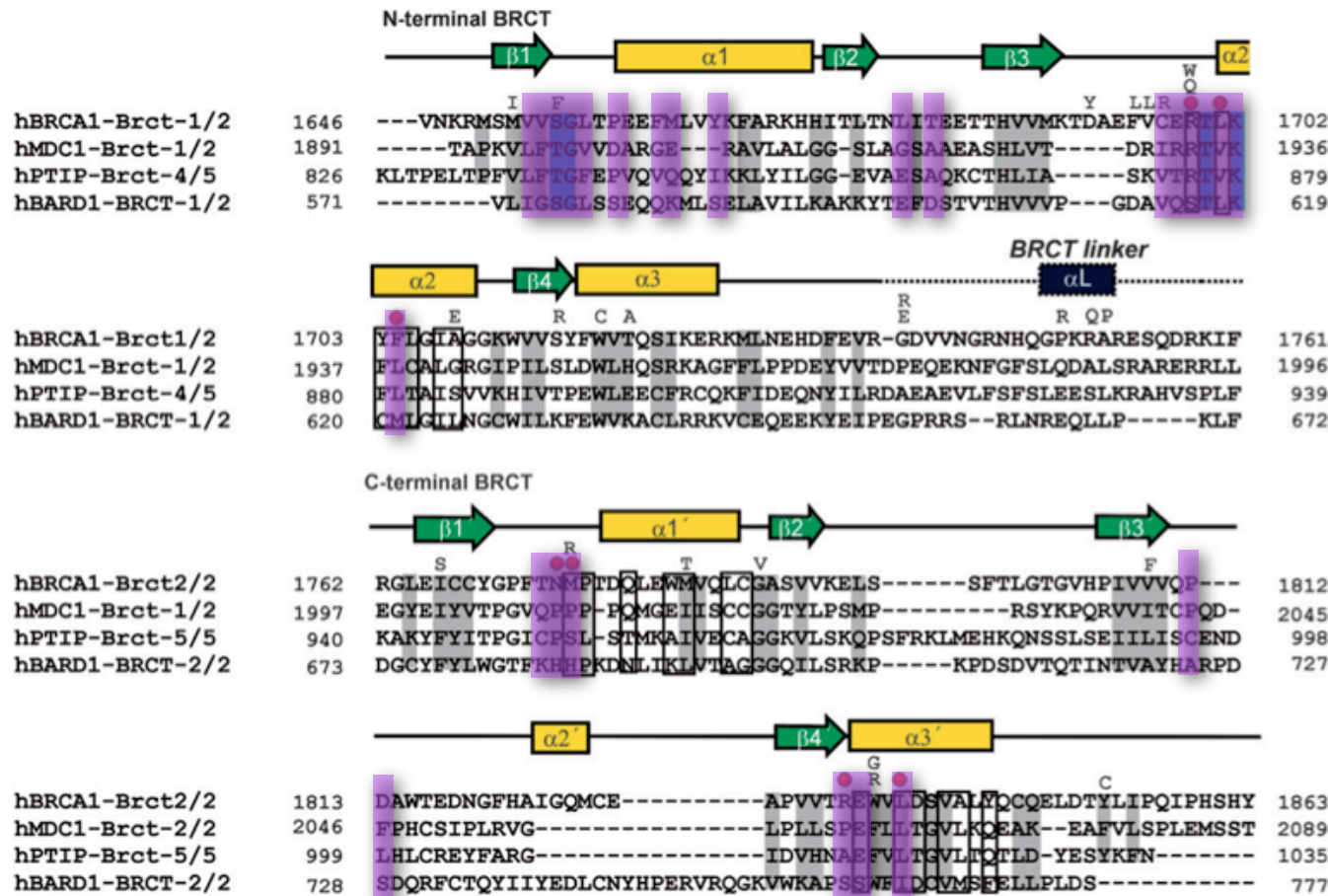


# Missense mutations in BRCT domains by function

	cancer associated	not cancer associated	?				
no transcription activation	C1697R R1699W A1708E S1715R P1749R M1775R		M1652K L1657P E1660G H1686Q R1699Q K1702E Y1703HF1 704S	L1705PS1 715NS172 2FF1734L G1738EG 1743RA1 752PF176 1I	F1761S M1775E M1775K L1780P I1807S V1833E A1843T		
transcription activation		M1652I A1669S		V1665M D1692N G1706A D1733G M1775V P1806A			
?			M1652T V1653M L1664P T1685A T1685I M1689R D1692Y F1695L V1696L R1699L G1706E W1718C	W1718S T1720A W1730S F1734S E1735K V1736A G1738R D1739E D1739G D1739Y V1741G H1746N	R1751P R1751Q R1758G L1764P I1766S P1771L T1773S P1776S D1778N D1778G D1778H M1783T	C1787S G1788D G1788V G1803A V1804D V1808A V1809A V1809F V1810G Q1811R P1812S N1819S	A1823T V1833M W1837R W1837G S1841N A1843P T1852S P1856T P1859R



# Putative binding site on BRCA1



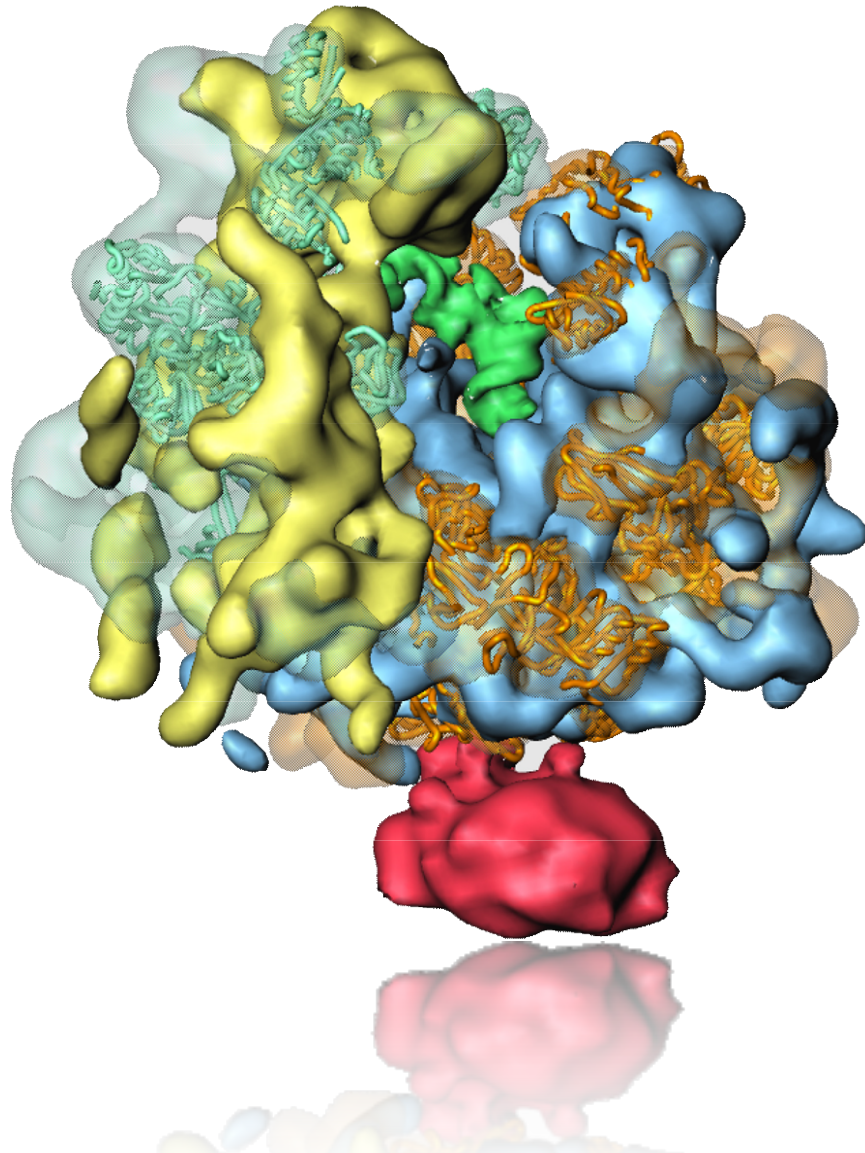
Putative binding site predicted in 2003  
and accepted for publication on March 2004.

Williams *et al.* 2004 Nature Structure Biology. June 2004 11:519

Mirkovic *et al.* 2004 Cancer Research. June 2004 64:3790



# *S. cerevisiae* ribosome



Fitting of comparative models into 15Å cryo-electron density map.

43 proteins could be modeled on 20-56% seq.id. to a known structure.

The modeled fraction of the proteins ranges from 34-99%.

# Tropical Disease Initiative (TDI)

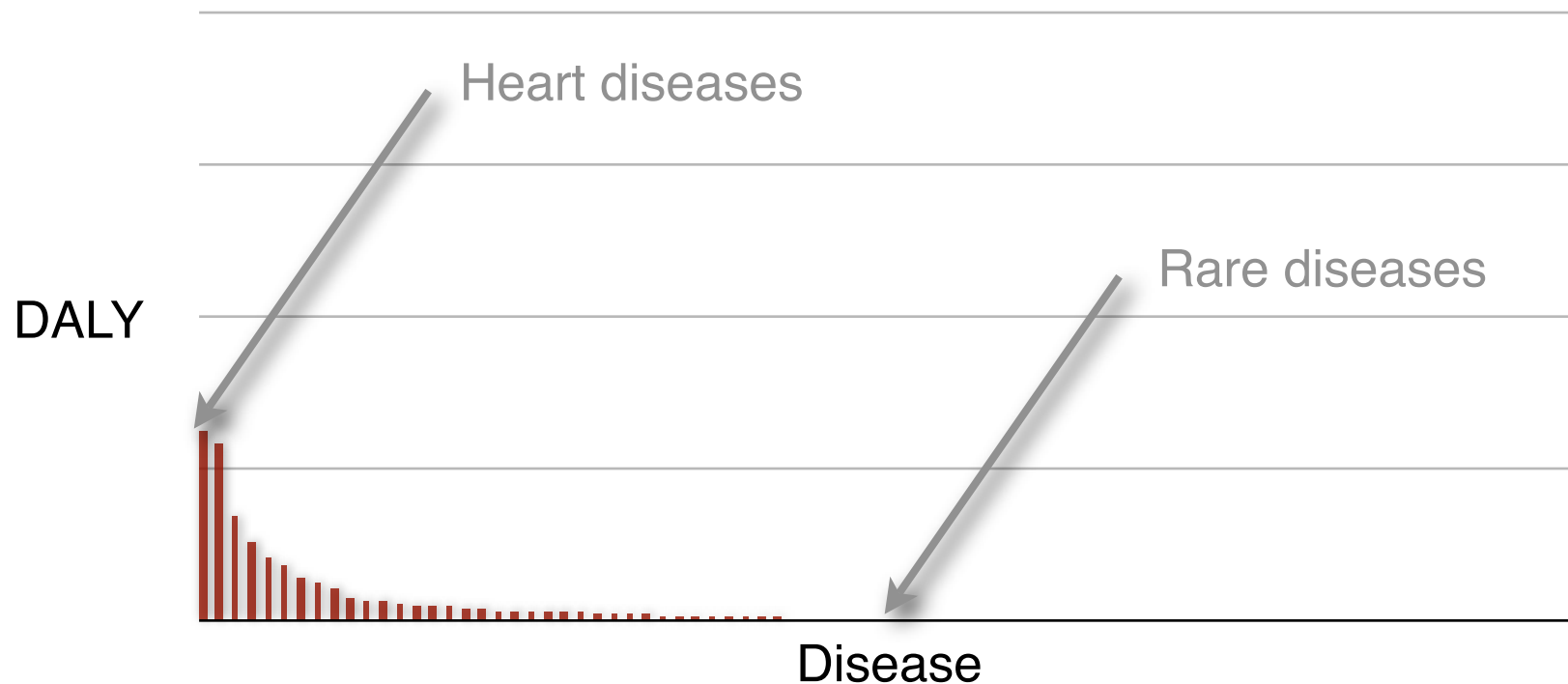
*Predicting binding sites in protein structure models.*



<http://www.tropicaldisease.org>

# Need is High in the Tail

- DALY Burden Per Disease in Developed Countries
- DALY Burden Per Disease in Developing Countries



Disease data taken from WHO, *World Health Report 2004*

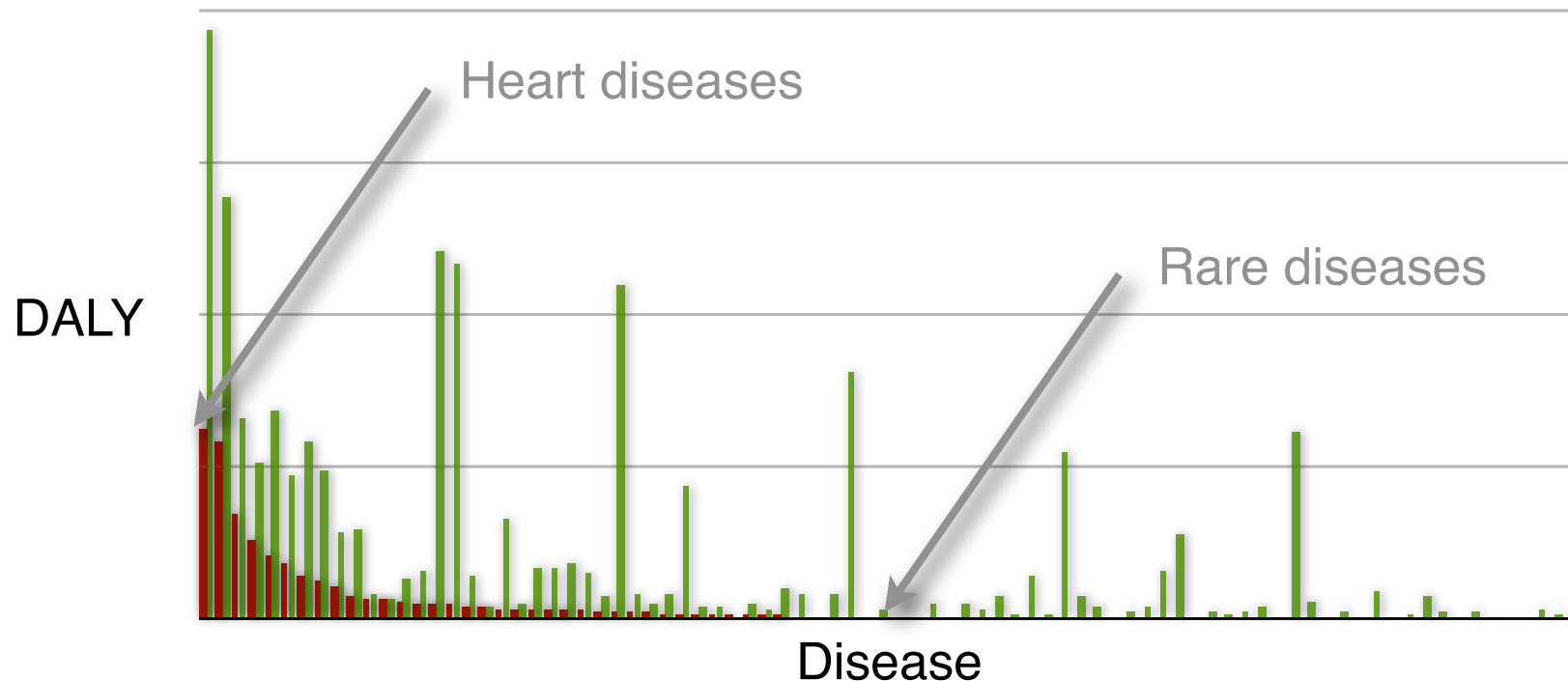
DALY - Disability adjusted life years

DALY is not a perfect measure of market size, but is certainly a good measure for importance.

*DALYs for a disease are the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the health condition. The DALY is a health gap measure that extends the concept of potential years of life lost due to premature death (PYLL) to include equivalent years of 'healthy' life lost in states of less than full health, broadly termed disability. One DALY represents the loss of one year of equivalent full health.*

# Need is High in the Tail

- DALY Burden Per Disease in Developed Countries
- DALY Burden Per Disease in Developing Countries



Disease data taken from WHO, *World Health Report 2004*

DALY - Disability adjusted life years

DALY is not a perfect measure of market size, but is certainly a good measure for importance.

*DALYs for a disease are the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the health condition. The DALY is a health gap measure that extends the concept of potential years of life lost due to premature death (PYLL) to include equivalent years of 'healthy' life lost in states of less than full health, broadly termed disability. One DALY represents the loss of one year of equivalent full health.*

# “Unprofitable” Diseases and Global DALY (in 1000’s)

<b>Malaria*</b>	<b>46,486</b>
Tetanus	7,074
<b>Lymphatic filariasis*</b>	<b>5,777</b>
Syphilis	4,200
Trachoma	2,329
<b>Leishmaniasis*</b>	<b>2,090</b>
Ascariasis	1,817
<b>Schistosomiasis*</b>	<b>1,702</b>
<b>Trypanosomiasis*</b>	<b>1,525</b>

Trichuriasis	1,006
Japanese encephalitis	709
<b>Chagas Disease*</b>	<b>667</b>
<b>Dengue*</b>	<b>616</b>
<b>Onchocerciasis*</b>	<b>484</b>
<b>Leprosy*</b>	<b>199</b>
Diphtheria	185
Poliomyelitis	151
Hookworm disease	59

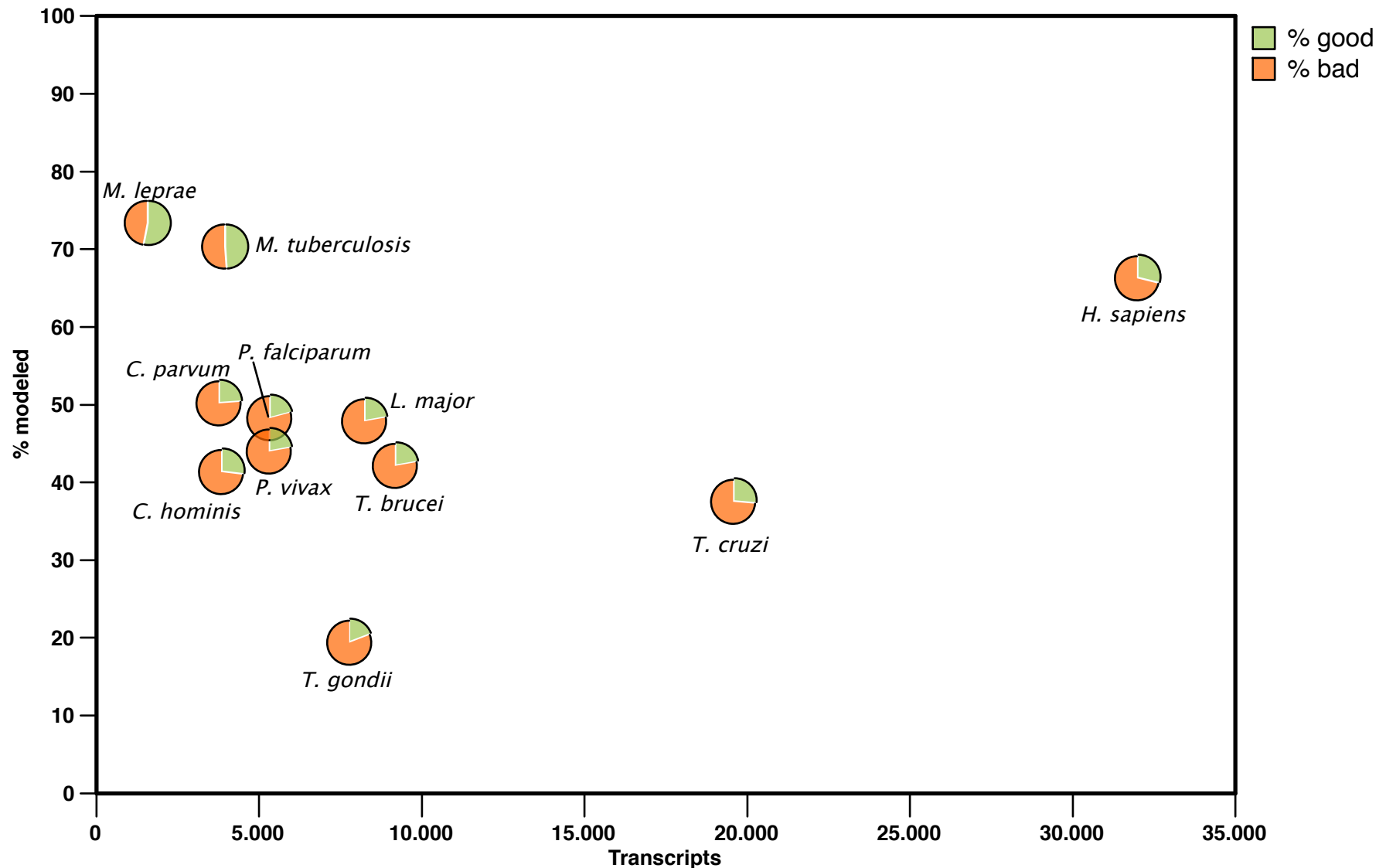
Disease data taken from WHO, *World Health Report 2004*

DALY - Disability adjusted life year in 1000’s.

\* Officially listed in the WHO Tropical Disease Research [disease portfolio](#).

# Modeling Genomes

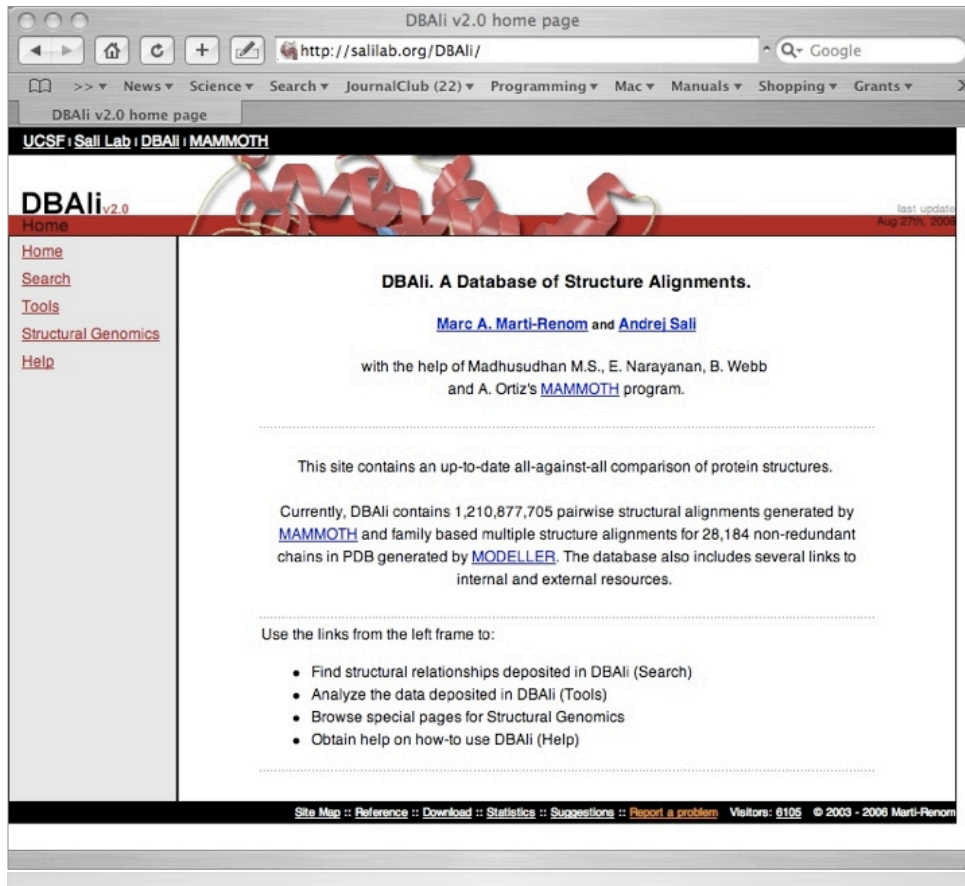
*data from models generated by ModPipe (Eswar, Pieper & Sali)*



*A good model has MPQS of 1.1 or higher*

# DBAli<sub>v2.0</sub> database

<http://www.dbali.org>



- ✓ Fully-automatic
- ✓ Data is kept up-to-date with PDB releases
- ✓ Tools for “on the fly” classification of families.
- ✓ Easy to navigate
- ✓ Provides tools for structure analysis

Does not provide a stable classification similar to that of CATH or SCOP

Pairwise structure alignments	
Last update:	July 5th, 2007
Number of chains:	93,307
Number of structure-structure comparisons:*	1,617,719,157
Multiple structure alignments	
Last update:	March 22nd, 2007
Number of representative chains:	31,848
Number of families:	11,900

Uses MAMMOTH for similarity detection

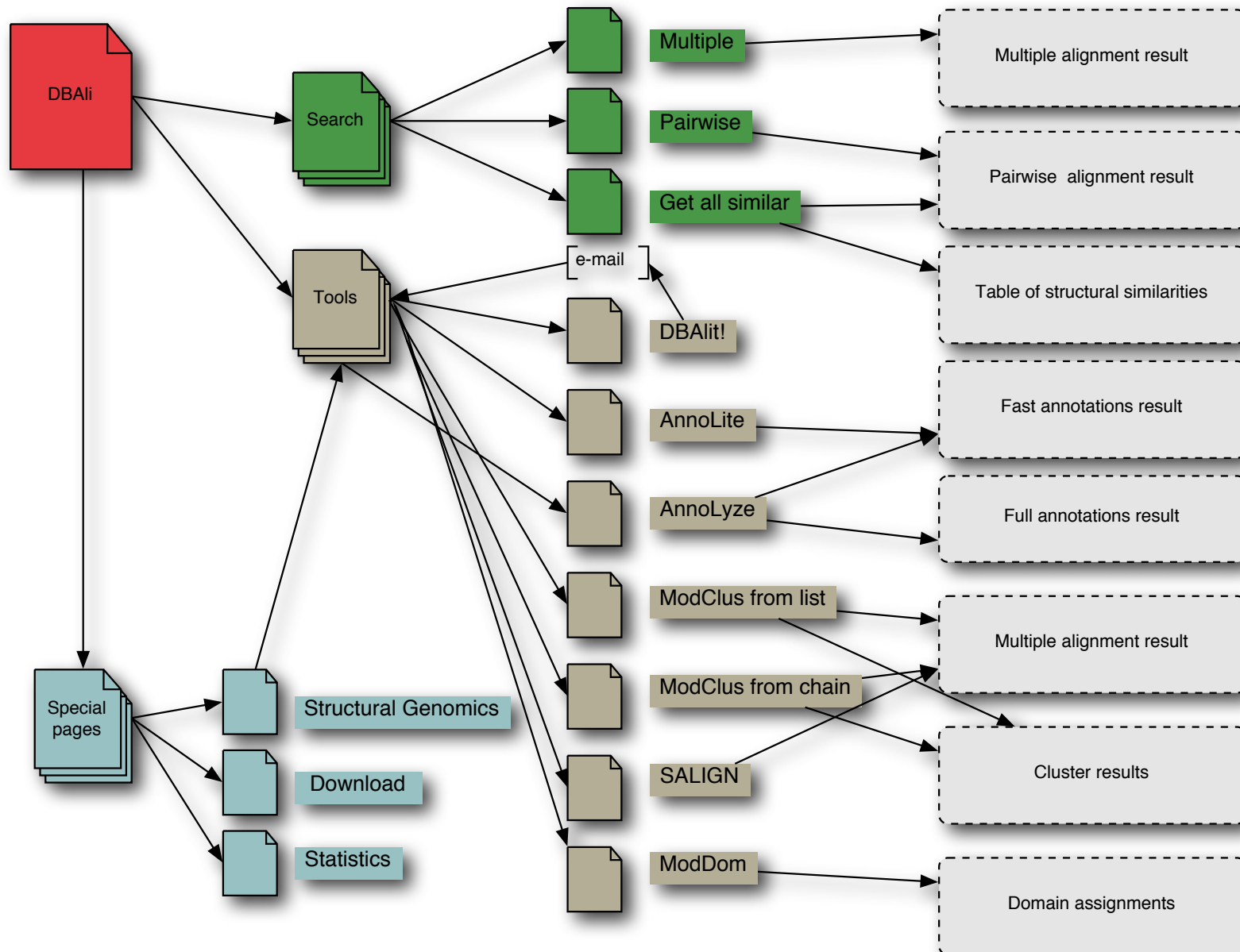
- ✓ VERY FAST!!!
- ✓ Good scoring system with significance

Ortiz AR, (2002) *Protein Sci.* 11 pp2606

Marti-Renom et al. 2001. *Bioinformatics.* 17, 746

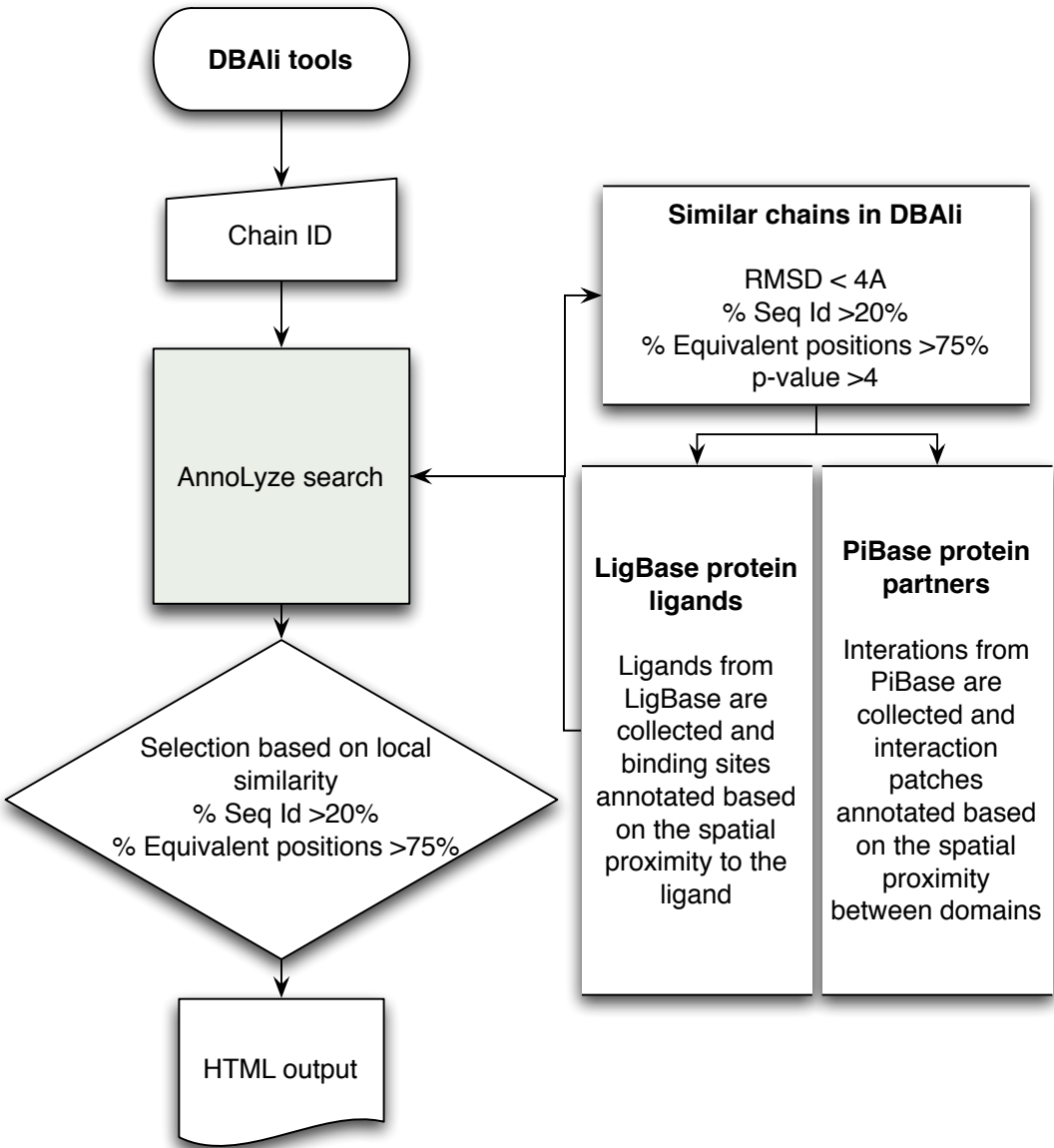
# DBAli<sub>v2.0</sub> database

<http://www.dbali.org>



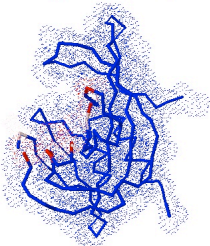


# Method



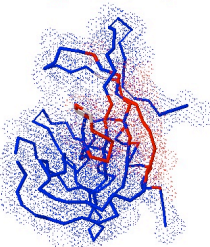
Inherited ligands: 4

Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
<a href="#">MO2</a>	59.03	<a href="#">0.185</a>	48 49 52 62 63 66 67 113 116
<a href="#">CRY</a>	20.00	<a href="#">0.111</a>	23 29 31 37 44 48 49 83 85 94 96 103 121
<a href="#">BOG</a>	20.00	<a href="#">0.111</a>	19 20 21 48 49 51 96 98 136
<a href="#">ACY</a>	15.87	<a href="#">0.163</a>	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135



Inherited partners:1

Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
<a href="#">d.113.1.1</a>	23.68	<a href="#">0.948</a>	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145



# Sensitivity .vs. Precision

	Optimal cut-off	Sensitivity (%) Recall or TPR	Precision (%)
Ligands	30%	71.9	13.7

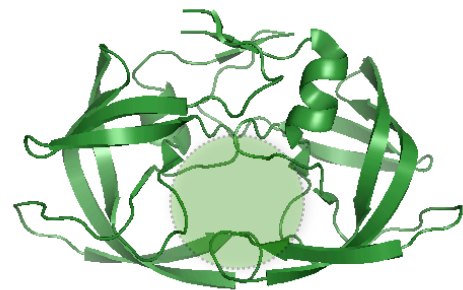
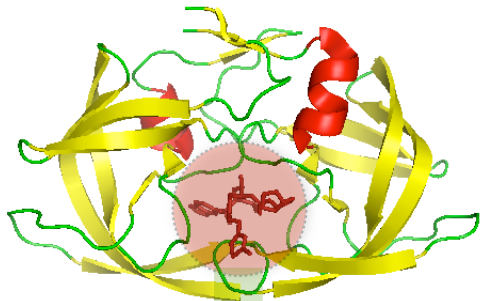
$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP}$$

**~90-95% of residues correctly predicted**

# Comparative docking

## 1. Expansion

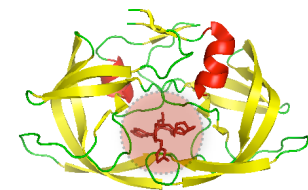
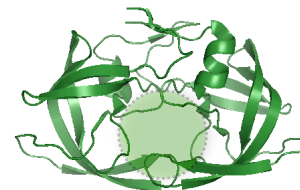
co-crystalized protein/ligand



crystalized protein

## 2. Inheritance

model



template

# Summary table

models with inherited ligands

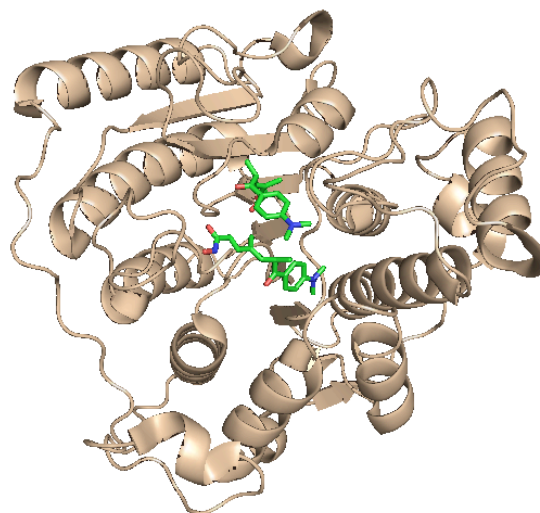
from 16,284 good models, 295 inherited a ligand/substance with at least one compound already approved by FDA and ready to be used from ZINC

	Transcripts	Good	Ligands	Lipinski	Lipinski+ZINC	FDA+ZINC
<i>C. hominis</i>	3,886	886	183	131	28	12 (10)
<i>C. parvum</i>	3,806	949	219	145	30	12 (10)
<i>L. major</i>	8,274	1,845	488	334	84	44 (34)
<i>M. leprae</i>	1,605	1,321	286	189	39	29 (25)
<i>M. tuberculosis</i>	3,991	2,887	404	285	71	44 (37)
<i>P. falciparum</i>	5,363	1,057	271	191	48	20 (16)
<i>P. vivax</i>	5,342	1,042	267	177	37	18 (15)
<i>T. brucei</i>	921	1,795	440	309	94	46 (36)
<i>T. cruzi</i>	19,607	3,915	730	493	127	62 (52)
<i>T. gondii</i>	7,793	587	174	124	28	8 (7)
<b>TOTAL</b>	<b>60,588</b>	<b>16,284</b>	<b>3,462</b>	<b>2,378</b>	<b>586</b>	<b>295 (242)</b>

# Example of inheritance (expansion)

*LmjF2 1.0680 from L. major “Histone deacetylase 2” (model 1)*

*Template 1t64A a human HDAC8 protein.*



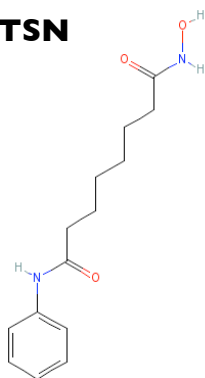
	Origen	Formula	Name	Cov.	Seq. Id. (%)
<b>ZN</b>	X-ray	Zn <sup>2+</sup>	Zinc ion	--	--
<b>NA</b>	X-ray	Na <sup>+</sup>	Sodium ion	--	--
<b>CA</b>	X-ray	Ca <sup>2+</sup>	Calcium ion	--	--
<b>TSN</b>	X-ray	C <sub>17</sub> H <sub>22</sub> N <sub>2</sub> O <sub>3</sub>	Trichostatin A	--	--
<b>SHH</b>	Expanded	C <sub>14</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub>	Octadenioic acid hudroxyamide phenylamide	100.00	83.8

# Example of inheritance (inheritance)

*LmjF21.0680 from L. major "Histone deacetylase 2" (model 1)*

	Formula	Name	Cov.	Seq. Id. (%)	Residues
<b>TSN</b>	C <sub>17</sub> H <sub>22</sub> N <sub>2</sub> O <sub>3</sub>	Trichostatin A	100.00	90.9	90 131 132 140 141 167 169 256 263 293 295
<b>SHH</b>	C <sub>14</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub>	Octadenioic acid hydroxyamide phenylamide	100.00	90.9	

**TSN**



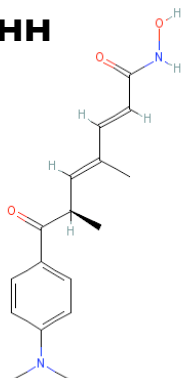
## suberoylanilide hydroxamic acid

### Pharmacological Action:

[Anti-Inflammatory Agents, Non-Steroidal](#)  
[Antineoplastic Agents](#)  
[Enzyme Inhibitors](#)  
[Anticarcinogenic Agents](#)

Inhibits histone deacetylase 1 and 3

**SHH**



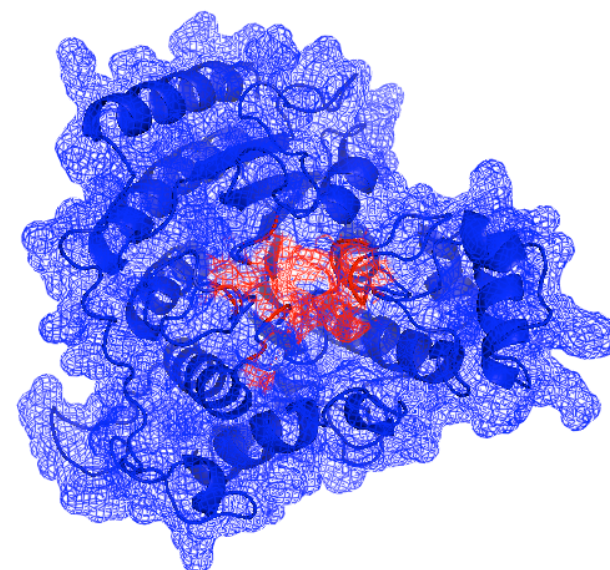
## trichostatin A

### Pharmacological Action:

[Antibiotics, Antifungal](#)  
[Enzyme Inhibitors](#)  
[Protein Synthesis Inhibitors](#)

chelates zinc ion in the active site of histone deacetylases, resulting in preventing histone unpacking so DNA is less available for transcription

	LmjF21.0680.1.pdb
<b>Template</b>	1t64A
<b>Seq. Id (%)</b>	<b>38.00</b>
<b>MPQS</b>	1.47



# Example of inheritance (CDD-Roos-literature)

*LmjF2 1.0680 from L. major “Histone deacetylase 2” (model 1)*

*Proc. Natl. Acad. Sci. USA*  
Vol. 93, pp. 13143–13147, November 1996  
Medical Sciences

## **Apicidin: A novel antiprotozoal agent that inhibits parasite histone deacetylase**

(cyclic tetrapeptide / Apicomplexa / antiparasitic / malaria / coccidiosis)

SANDRA J. DARKIN-RATTRAY<sup>\*†</sup>, ANNE M. GURNETT<sup>\*</sup>, ROBERT W. MYERS<sup>\*</sup>, PAULA M. DULSKI<sup>\*</sup>,  
TAMI M. CRUMLEY<sup>\*</sup>, JOHN J. ALLOCCO<sup>\*</sup>, CHRISTINE CANNOVA<sup>\*</sup>, PETER T. MEINKE<sup>‡</sup>, STEVEN L. COLLETTI<sup>‡</sup>,  
MARIA A. BEDNAREK<sup>‡</sup>, SHEO B. SINGH<sup>§</sup>, MICHAEL A. GOETZ<sup>§</sup>, ANNE W. DOMBROWSKI<sup>§</sup>,  
JON D. POLISHOOK<sup>§</sup>, AND DENNIS M. SCHMATZ<sup>\*</sup>

Departments of <sup>\*</sup>Parasite Biochemistry and Cell Biology, <sup>‡</sup>Medicinal Chemistry, and <sup>§</sup>Natural Products Drug Discovery, Merck Research Laboratories, P.O. Box 2000, Rahway, NJ 07065

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Apr. 2004, p. 1435–1436  
0066-4804/04/\$08.00+0 DOI: 10.1128/AAC.48.4.1435–1436.2004  
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

Vol. 48, No. 4

## **Antimalarial and Antileishmanial Activities of Aroyl-Pyrrolyl-Hydroxyamides, a New Class of Histone Deacetylase Inhibitors**

# Models database

<http://bioinfo.cipf.es/sgu/services/TDIModels/>

The TDIModels server

http://bioinfo.cipf.es/sgu/services/TDIModels/index.p Google

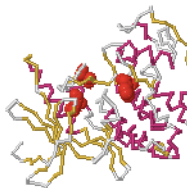
The TDIModels server

**TDIModels** ⓘ

Results for **O96526** [O96526 Cdc2-related kinase (Cell division related protein)]  
Number of models: 2

[SGU-HOME]  
DBAli  
Eva-CM  
SeqProfCod  
TDIModels

JMOL



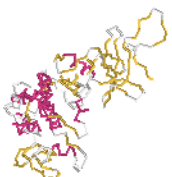
Jmol

This model has 1 predicted ligands.

	Lipinski	ZINC	FDA Coverage	Seq. Id.
<input checked="" type="checkbox"/> NO3	•	•	•	100.00 100.00

SEQUENCE IDENTITY: 58.00  
MODPIPE QUALITY SCORE: 1.73  
TEMPLATE PDB: 1gz8  
TEMPLATE CHAIN: A  
TARGET LENGTH: 311  
TARGET BEGIN: 20  
TARGET END: 309  
[Download PDB file](#)

JMOL



Jmol

This model has 2 predicted ligands.

	Lipinski	ZINC	FDA Coverage	Seq. Id.
<input checked="" type="checkbox"/> NO3	•	•	•	100.00 100.00
<input checked="" type="checkbox"/> KCX	•	•	•	100.00 93.75

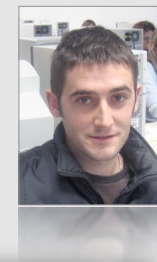
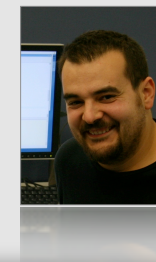
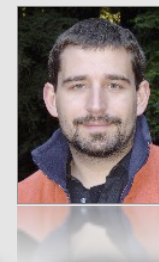
SEQUENCE IDENTITY: 29.00  
MODPIPE QUALITY SCORE: 1.13  
TEMPLATE PDB: 2cn5  
TEMPLATE CHAIN: A  
TARGET LENGTH: 311  
TARGET BEGIN: 1  
TARGET END: 311  
[Download PDB file](#)

[<- new search](#)

HELP:



# Acknowledgments



## **Structural Genomics Unit (CIPF)**

Marc A. Marti-Renom

Emidio Capriotti

Peio Ziarsolo Areitioaurtena

## **Comparative Genomics Unit (CIPF)**

Hernán Dopazo

Leo Arbiza

Francisco García

## **Functional Genomics Unit (CIPF)**

Joaquín Dopazo

Fátima Al-Shahrour

José Carbonell

Ignacio Medina

David Montaner

Joaquín Tárraga

Ana Conesa

Toni Gabaldón

Eva Alloza

Lucía Conde

Stefan Goetz

Jaime Huerta Cepas

Marina Marcet

Pablo Minguez

Jordi Burguet Castell

## **FUNDING**

Prince Felipe Research Center

Marie Curie Reintegration Grant

STREP EU Grant

Generalitat Valenciana

## **Tropical Disease Initiative**

Stephen Maurer (UC Berkeley)

Arti Rai (Duke U)

Andrej Sali (UCSF)

Ginger Taylor (TSL)

Barri Bunin (CDD)

## **STRUCTURAL GENOMICS**

Stephen Burley (SGX)

John Kuriyan (UCB)

NY-SGXRC

## **MAMMOTH**

Angel R. Ortiz

## **BIOLOGY**

Jeff Friedman (RU)

James Hudsped (RU)

Partho Ghosh (UCSD)

Alvaro Monteiro (Cornell U)

Stephen Krilis (St. George H)

## **FUNCTIONAL ANNOTATION**

Fatima Al-Shahrour

Joaquín Dopazo

## **COMPARATIVE MODELING**

Andrej Sali

M. S. Madhusudhan

Narayanan Eswar

Min-Yi Shen

Ursula Pieper

Bino John

Maya Topf

## **FUNCTIONAL ANNOTATION**

Andrea Rossi

Fred Davis



<http://bioinfo.cipf.es>  
<http://sgu.bioinfo.cipf.es>