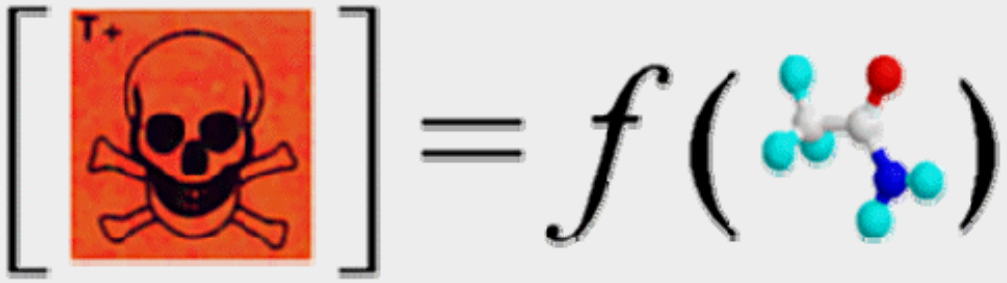

MARIO NEGRI
ISTITUTO DI RICERCHE FARMACOLOGICHE
Emilio Benfenati
Istituto Mario Negri
Laboratory of Environmental Chemistry and Toxicology
GRID approaches for Industrial Chemicals and Pharmaceutical Applications
UNICORE Summit 2008
Las Palmas de Gran Canaria, Spain - August 26, 2008
UNICORE SUMMIT

(Q)SAR

=

(quantitative) structure-activity relationship



in silico

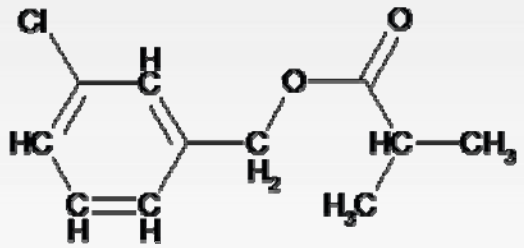
MARIO NEGRI

MARIO NEGRI

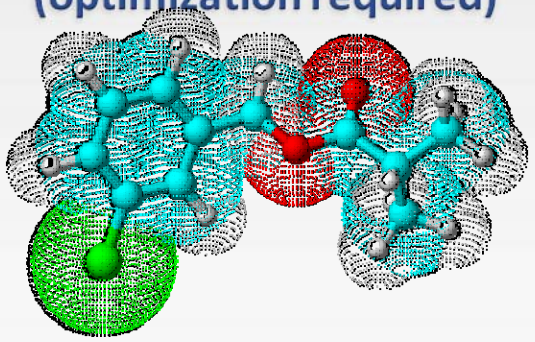
MARIO NEGRI

procedure to calculate DESCRIPTORS

2D descriptors
(no optimization required)



3D descriptors
(optimization required)



- Thousands of *2D/3D descriptors* can be calculated with different software starting from different input file format
- Thousands of *chemical fragments* can be used with other software

MOLECULAR DESCRIPTORS

Many *descriptors families* are used:

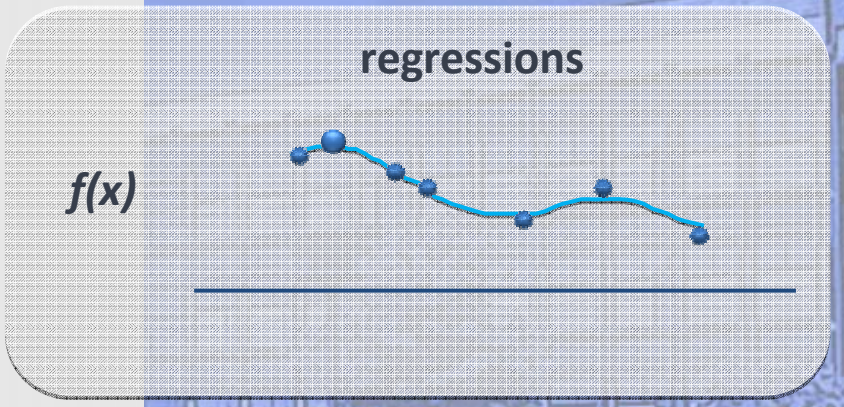
- **Constitutional / information descriptors**: molecular weight, number of chemical elements, number of H-bonds or double bonds, …
- **Physicochemical descriptors**: lipophilicity, polarizability, …
- **Topological descriptors**: atomic branching and ramification
- **Electronic, geometrical and quantum-chemical descriptors**
- **Fragmental / structural keys** defining *Booleans (bitmap) arrays*

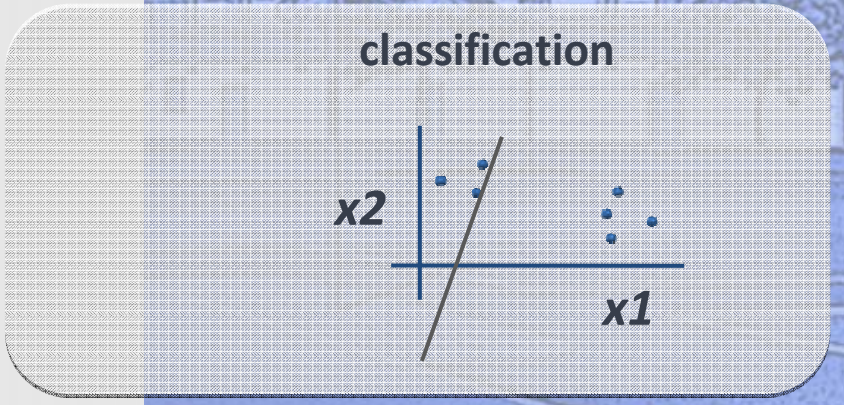
… … …

MARIO NEGRI

ALGORITHMS *Classifiers*

- Discriminant Analysis
- CART
- KNN
- Fuzzy logic
- Bayesian
- Self Organizing Map (SOM)
- Support Vector Machine (SVM)



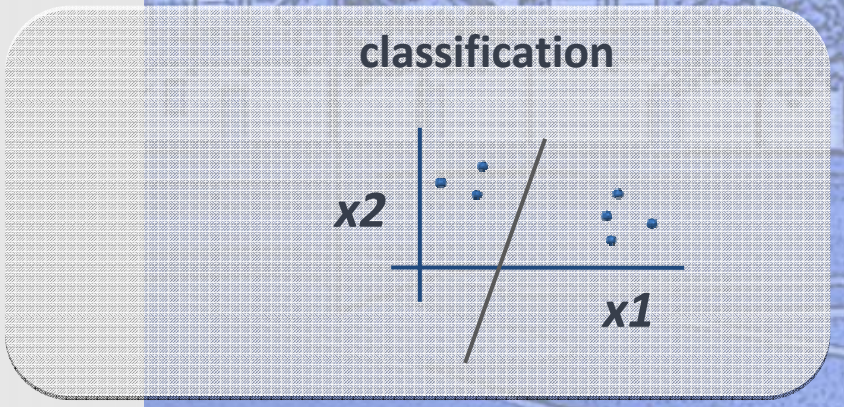


MARIO NEGRI

ALGORITHMS *Regressions*

- Multi Variate Analysis (MVA)
- Partial Least Squares (PLS)
- Neural Networks (NN)
- Other algorithms

  (PCA, GeneticAlgorithms)



IT tools for Chemicals/Property

2 main *experiences*

toxicity/environmental

pharma

Toxicity/environmental models

- Negative properties
- Human AND environmental properties
- Hundreds/thousands chemicals
- Focus on few chemicals
- Simpler tools
- Free tools
- No false negatives
- Public/confidential data
- Acceptability issue

Pharma specifities

- Beneficial properties
- Human properties
- Millions of candidates
- Focus on selected candidate drugs
- Complex tools
- Commercial tools
- No false positives
- Confidential data
- Commonly used

MARIO NEGRI

REACH

- REACH (registration, evaluation, authorization of chemicals) is the new EU legislation dealing with chemicals in Europe.
  The philosophy is: <u>NO DATA NO MARKET</u>

- Hazard profile and safe use should be granted for all substances marketed for more than 1 ton/year. Huge amount of data will be therefore needed since 2010 up to 2018

- About 30,000 chemicals to be evaluated
  Billions of Euros for testing

REACH & QSAR

- (Q)SAR is mentioned among other alternatives to animal testing as an acceptable method to fill data gap.

- According to *REACH regulation* (Annex XI) a (Q)SAR is valid if:

  - the model is recognized scientifically valid;
  - the substance is included in the applicability domain of the model;
  - results are adequate for classification and labelling and for risk assessment;
  - adequate documentation of the methods provided.

Legislation, such as **REACH**,

aims to assess effects towards

human health and ***environment***

Targets:

- Man
- Ecosystem

MARIO NEGRI

SEVERAL METHODS

can be used to get data

- several **ANIMAL MODELS**
  preferably according to *OFFICIAL GUIDELIN*

and

- **ALTERNATIVE METHODS**

Thus, the **legislative approach**

is not **black** or **white**:

- there is a **grey scale**

- other legislations are *different*:

  e.g.

  pesticides require *animal models*

  cosmetics require *non-animal models*

REACH

is **richer** than other legislations

- more **complex**
- more **flexible**

the **evaluation of the methods,**
Including *alternative ones*, and *QSAR*,
is **specific to the law**

if the **scale** is not **zero** or **one**…

…how to weight

***in silico* contributions?**

we have two cases:

1. *in silico* model mimics an animal model
2. no animal model exists

1

in silico model mimicking animal model

we compare
the two methods

ANIMAL QSAR

animal model exists:

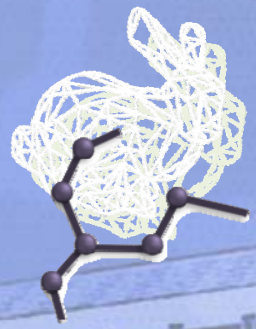
what do we **LOOSE**?

no model:

what do we **GAIN**?

2

in silico model without related model



we evaluate

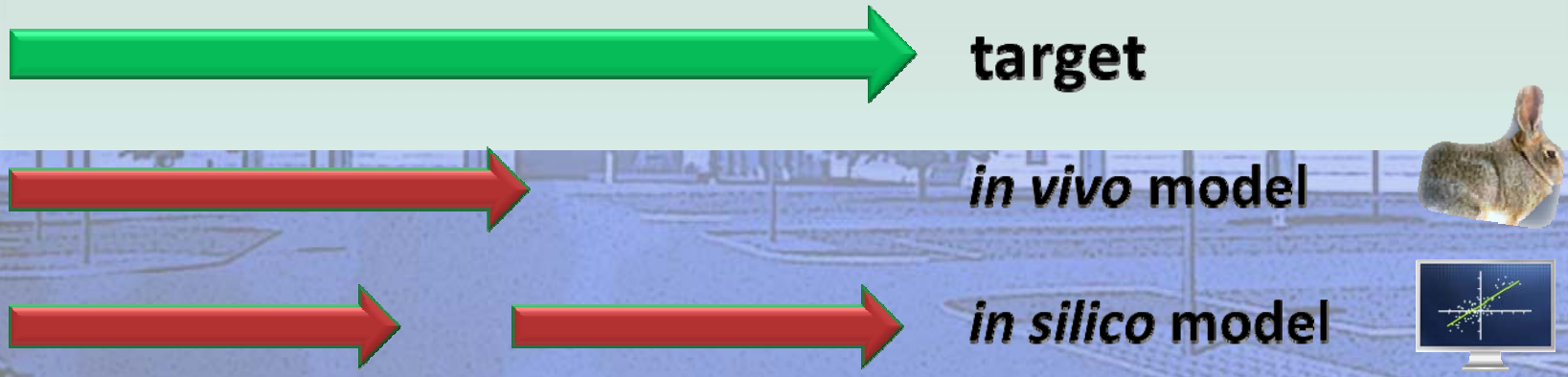
the *utility* for the TARGET

MARIO NEGRI

TARGET is environment (man)

question:

can the ***in silico*** approach be useful?



MARIO NEGRI

the ***in silico model*** should go in **parallel** to the **animal model** in a smooth way

the QSAR model

should adhere to the ideal situation as much as possible

Data -> Use

"Data" -> ***Model*** -> *Use*

what kind of DATA?

preferably from

official protocols

what kind of USE?

depends on the tonnage, etc:

C&L, Risk Assessment

MARIO NEGRI

QSAR model is a **statistical process**, which extracts **information** from ***several examples***

- **UNCERTAINTY** is necessary info
- **UNCERTAINTY** should be necessary in ALL cases, not only for QSAR

MARIO NEGRI

QSAR forces the debate into mathematical terms

we should aim to a more objective, mathematical description

given the *experimental uncertainty*,
we have to identify
the *in silico* uncertainty

considering the use we define
domains of acceptable use

OECD principles for QSAR validation

I. a defined endpoint

II. an unambiguous algorithm

III. a defined applicability domain

IV. appropriate measures of goodness-of-fit, robustness and predictivity

V. a mechanistic interpretation, if possible

Projects addressing the REACH legislation

CAESAR
QSAR models for REACH
→ models development

OSIRIS
ITS for REACH
→ ITS

CHEMOMENTUM
grids for knowledge-oriented applications
→ GRID approaches

SCARLET
SARs in *mutagenicity* and *carcinogenicity*
→ workshop

CHEMPREDICT
models based on simple chemical descriptors
→ descriptors development

CASCADE
chemicals as contaminants in the food chain
→ endocrine disruptors screening

Computer Assisted Evaluation of industrial chemical Substances According to Regulations

SIXTH FRAMEWORK PROGRAMME

http://www.caesar-project.eu/

The Consortium of the CAESAR project

1. Istituto di Ricerche Farmacologiche Mario Negri
   coordinator
2. Central Science Laboratory
3. BioChemicsConsulting SAS
4. Politecnico di Milano
5. KnowledgeMiner Software Frank Lemke
6. Liverpool John MooresUniversity
7. HelmholtzZentrumfürUmweltforschung - UFZ
8. KemijskiinštitutLjubljanaSlovenija
9. TNO - National Organisatievoor ToegepastNatuurwetenschappelijkOnderzoek

http://www.caesar-project.eu/

MARIO NEGRI

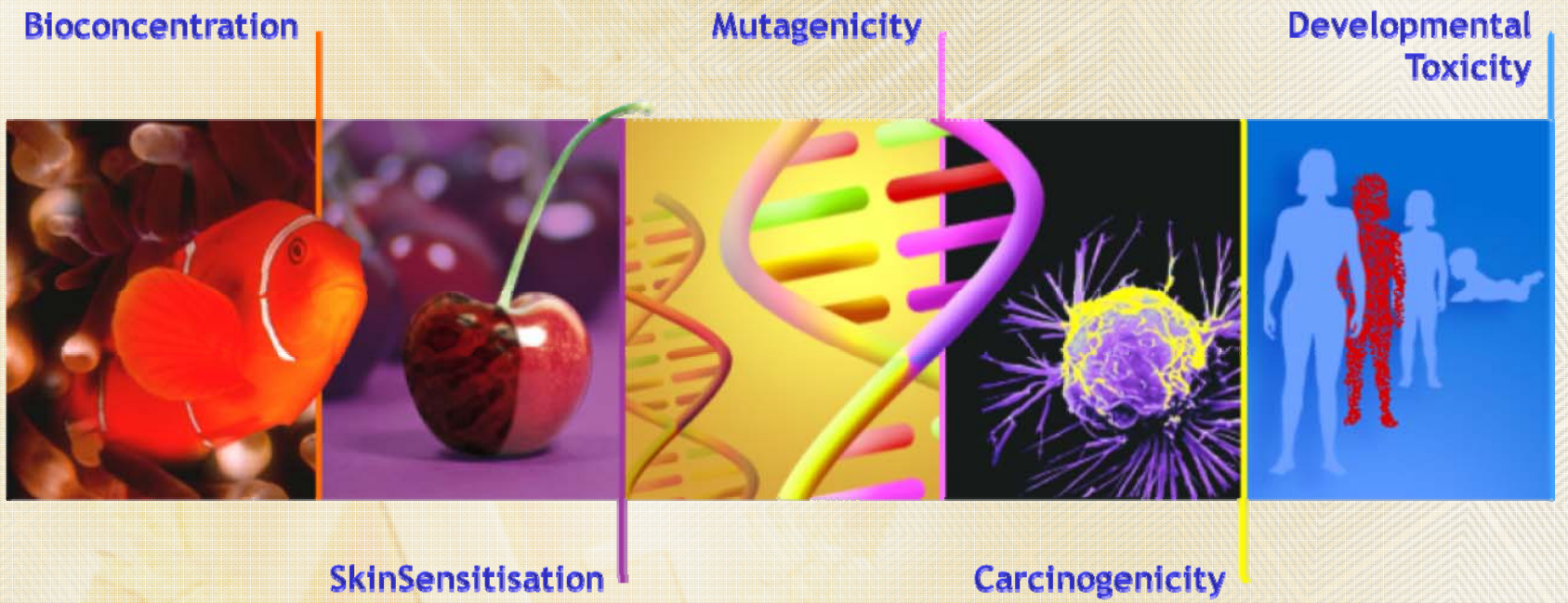
CAESAR the objective

- to develop **QSAR models** for **REACH**

CAESAR the method

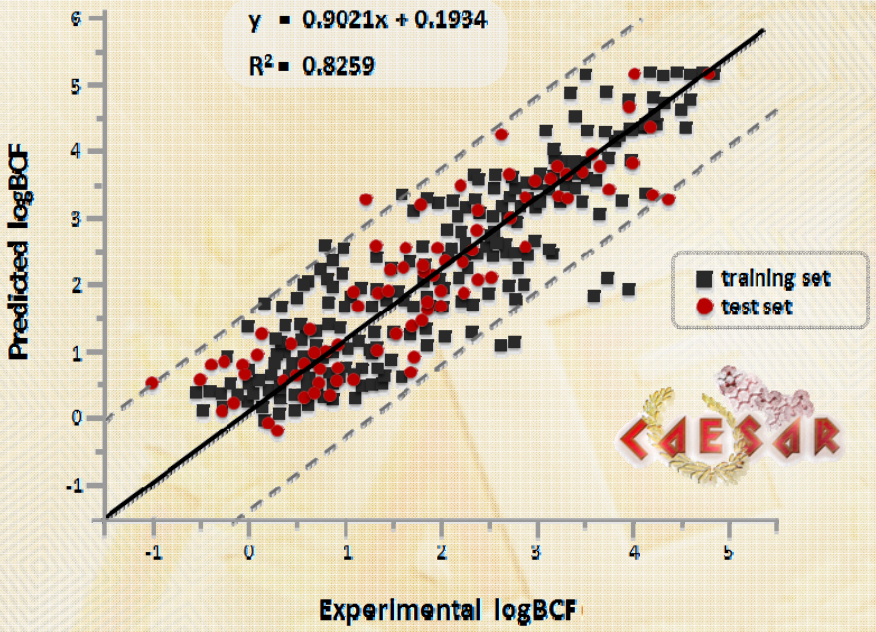
- **batteries of tools:**
  chemical descriptors, fragments, different algorithms
- **integrated tools**

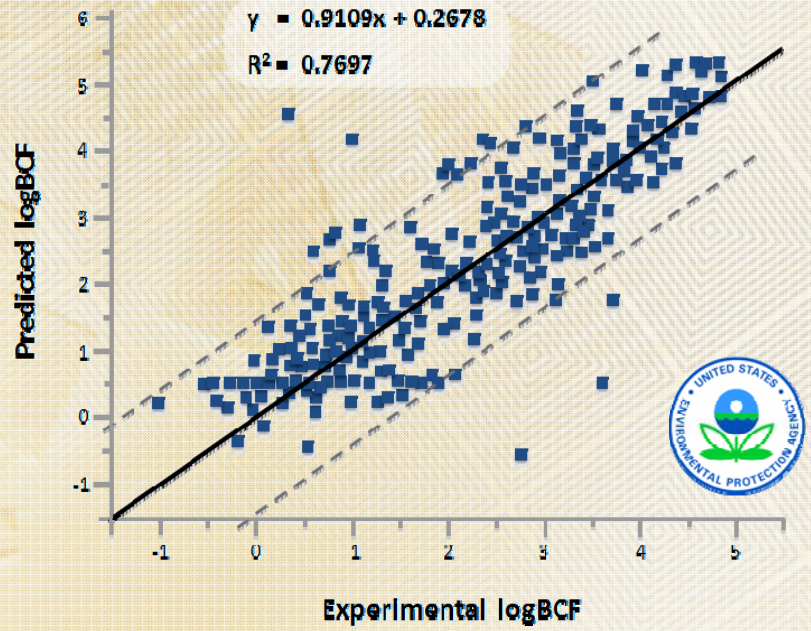
CAESAR the 5 selected end-points



CAESAR Results of modelling

CAESAR model for Bioconcentration Factor (BCF) vs EpiSuite® model (US EPA)



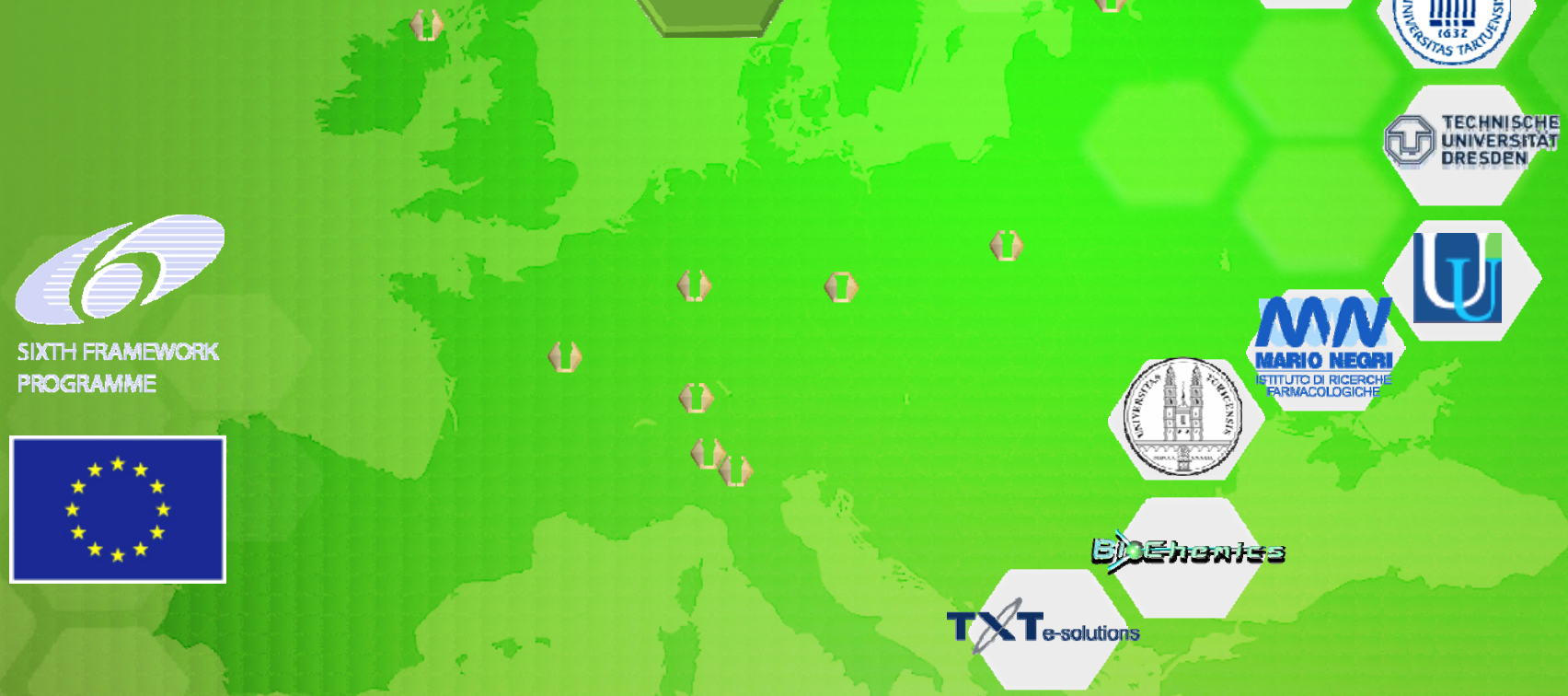


CAESAR Some open issues

- How to calculate tens of thousands of compounds?
- How to make user-friendly the tool?
- How to develop models more focussed on chemicals in a reproducible, robust way?
- There are *two approaches*: *global* and *local models*

Chemomentum

Grid services based environment to enable innovative research

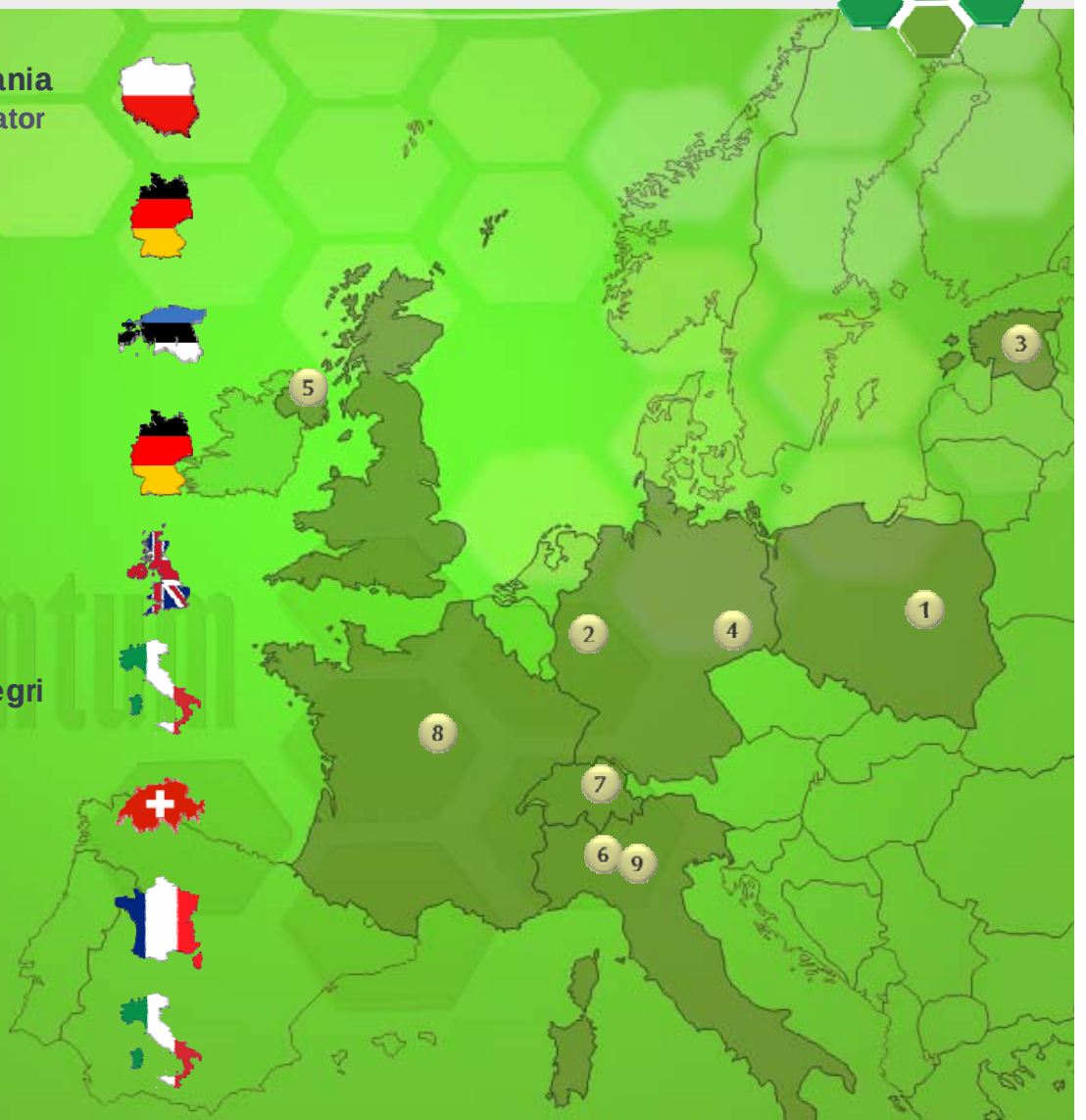


SIXTH FRAMEWORK PROGRAMME

http://www.chemomentum.org/

CHEMOMENTUM the Consortium

1. ICM - Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego-coordinator
2. ForschungszentrumJülichGmbH
3. TartuÜlikool - UniversityofTartu
4. TechnischeUniversitätDresden
5. Universityof Ulster
6. Istituto di Ricerche Farmacologiche Mario Negri
7. UniversitätZürich
8. BioChemicsConsulting SAS
9. TXT e-solutions S.p.A.

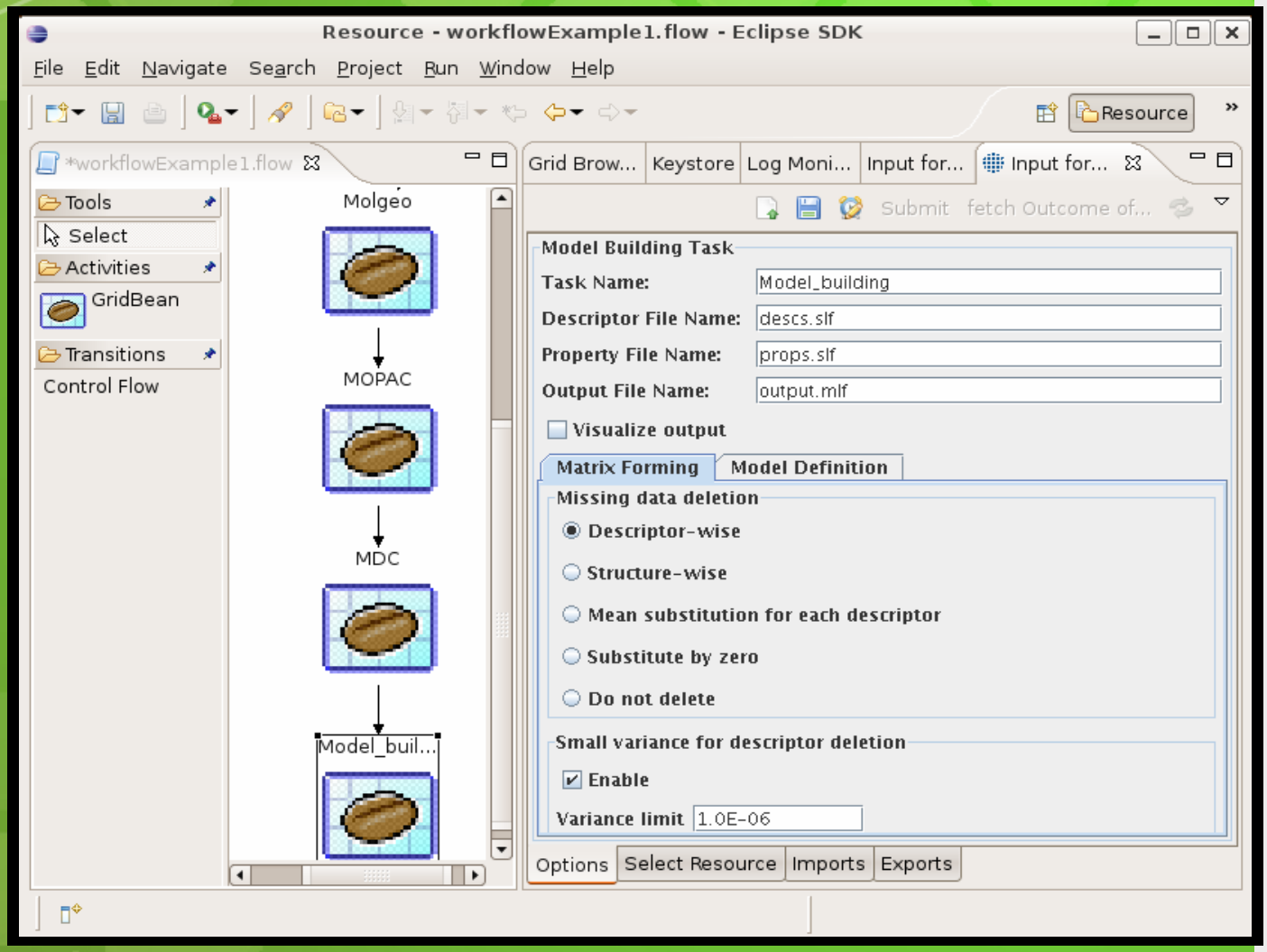


MARIO NEGRI

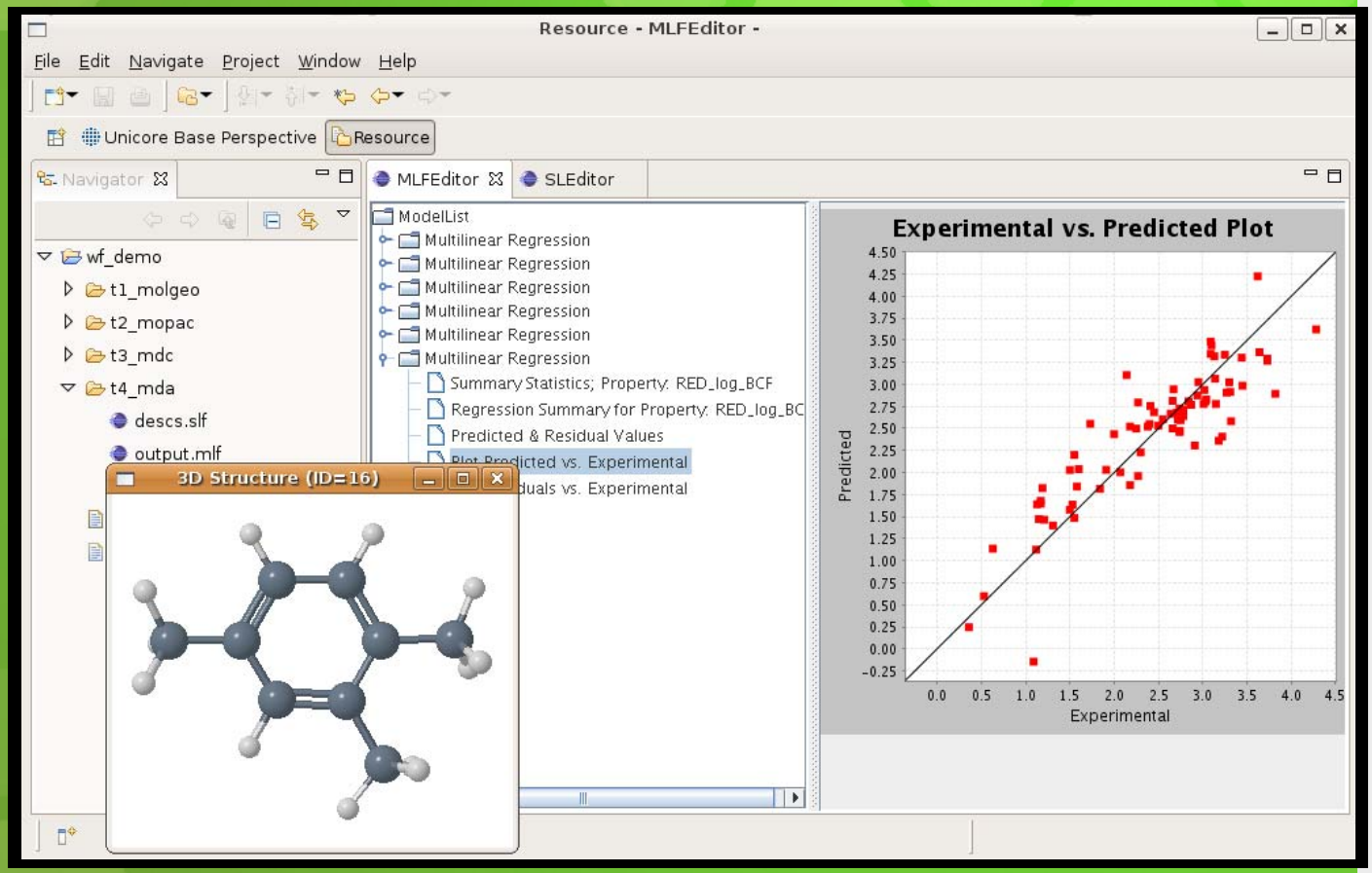
http://www.chemomentum.org/

CHEMOMENTUM Project Overview

- focused on the end users;
- user-friendly automatic approach for QSAR modelling;
- seamless approach;
- efficiently deal with data and knowledge:
  same environment, DB, SW, chem, properties, QMRF;
- Links with existing DB and SW:
  ECOTOX, CDK, CODESSApro, JOELib2, Marvin, Macromodel, MOLGEO, MOPAC, GAMESS, AFP, HSA(GA), DEMETRA, CAESAR, workflow tools, docking;
- Use cases:
  Drug discovery, toxicity prediction, environmental risk assessment, REACH.

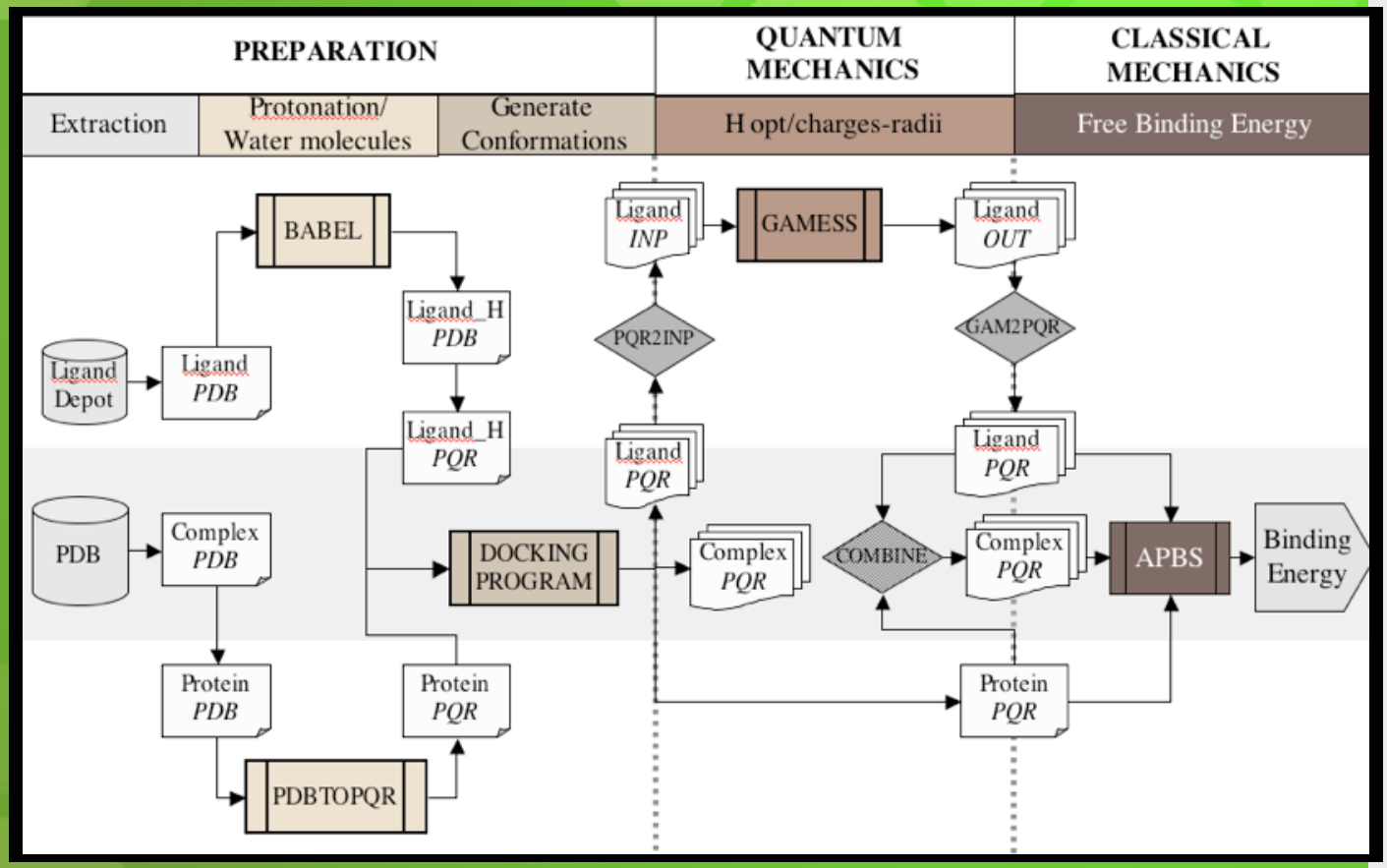


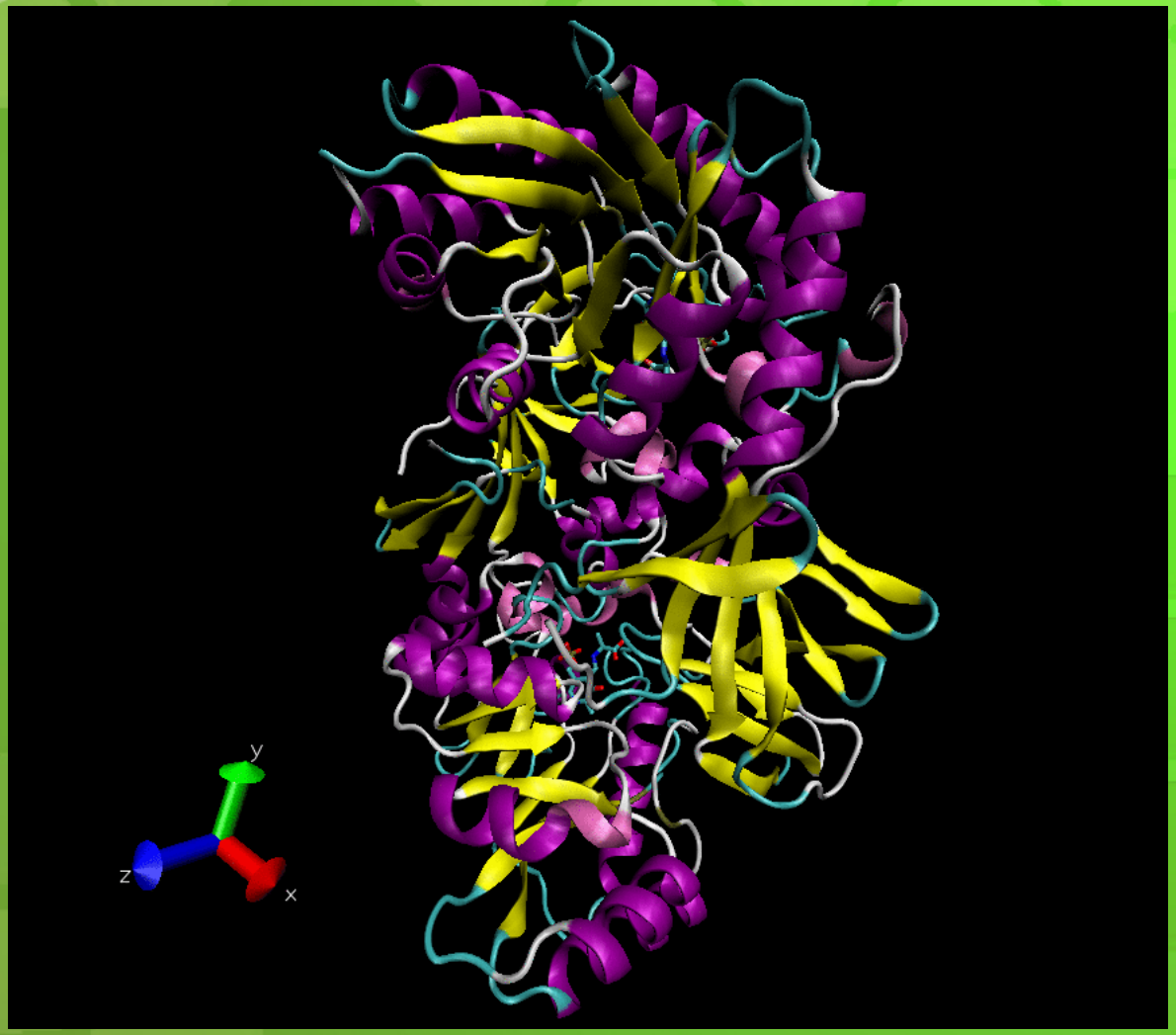
User chooses workflow



QSAR modelling results

Docking scheme





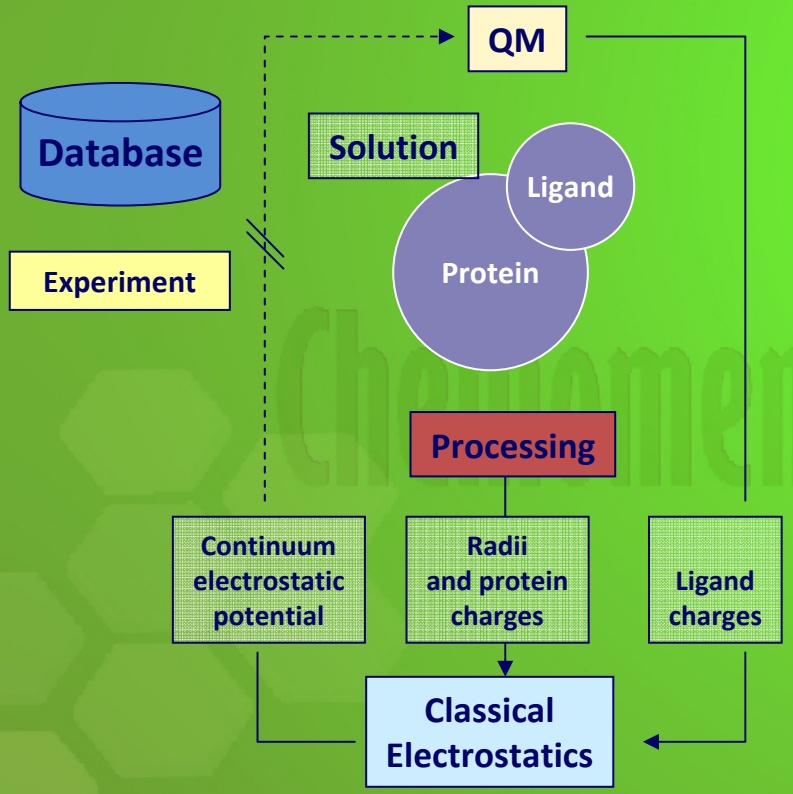
Docking workflow

each workflow step is a command line application

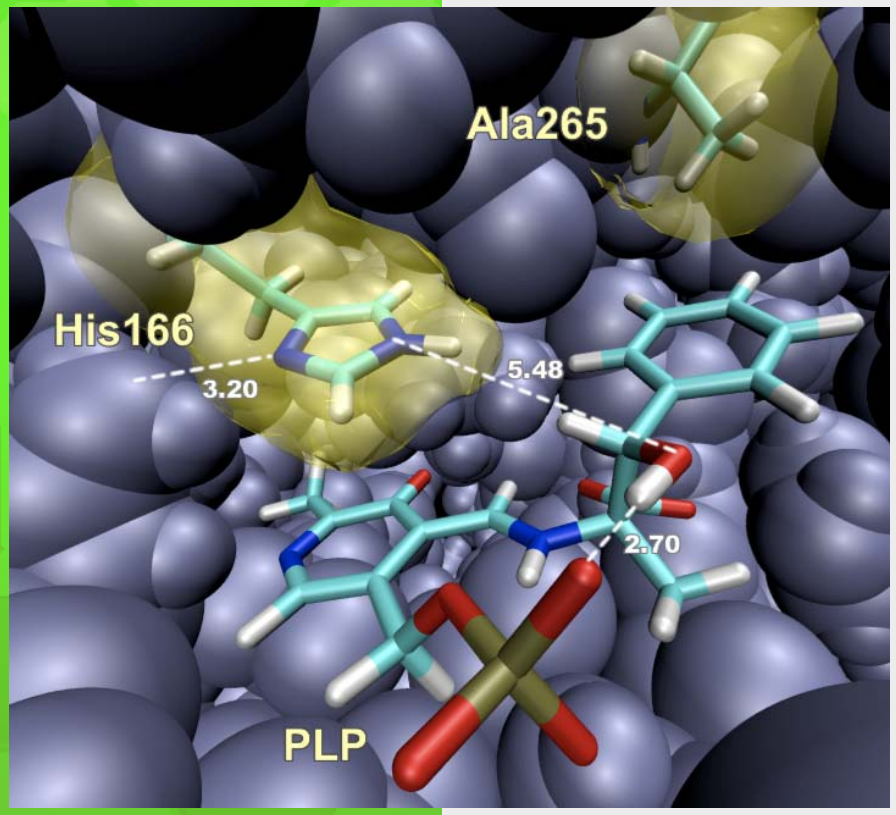
can be invoked with a command-line gridbean

Quantitative determinations In ligand-protein active sites

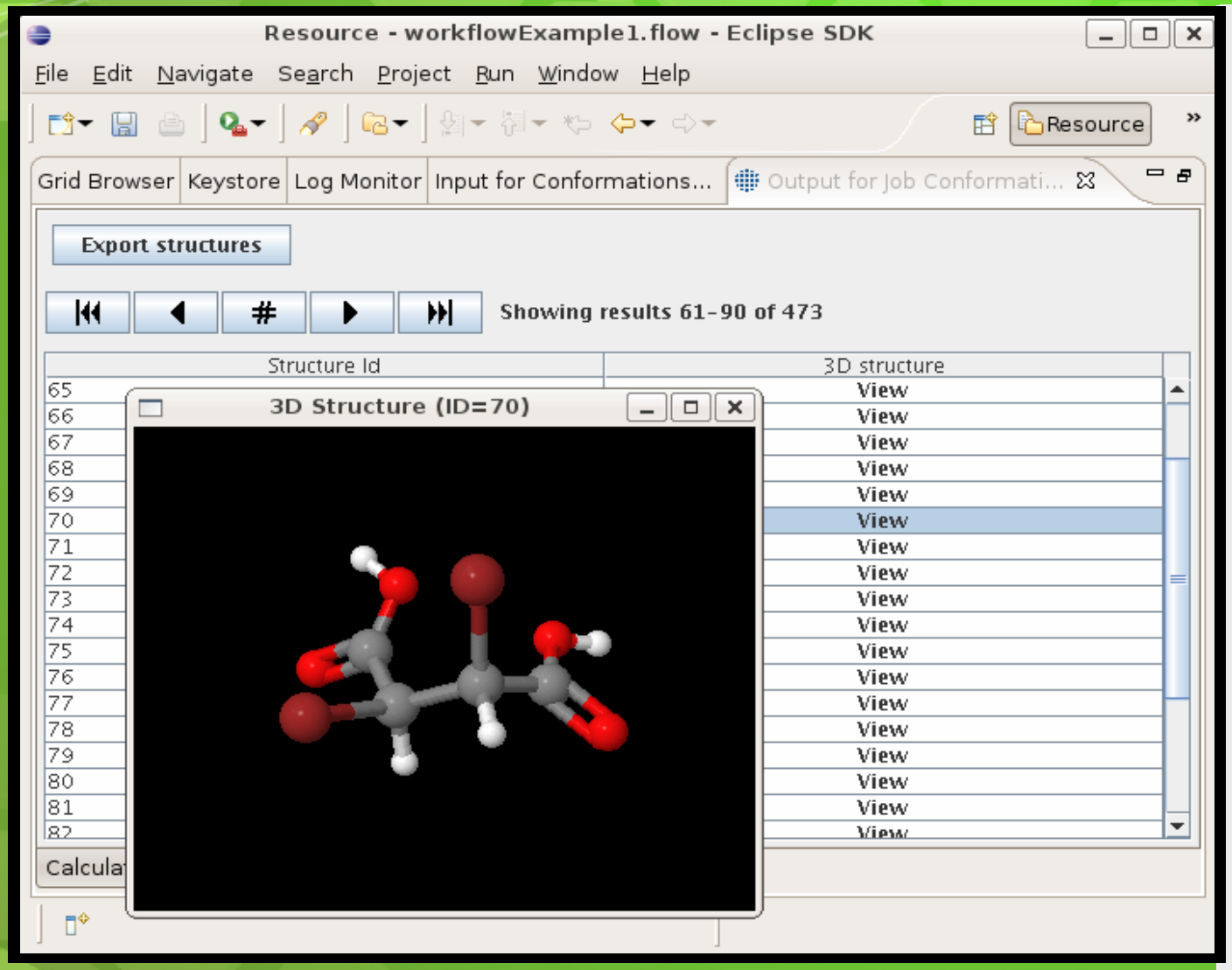
Macroscopic effects of sequence mutations on protein function



Hybrid QM-APBS-MD method
Grid Technologies
Middleware application driver



Seebeck, F.P.; Guainazzi, A.; Amoreira, C.; Baldridge, K.K.; Hilvert, D. "Stereoselectivity and expanded substrate scope of engineered PLP-dependent aldolase, *Angewante. Chemie. Int. Engl.*, 2006.



Powerful, Flexible visualisation

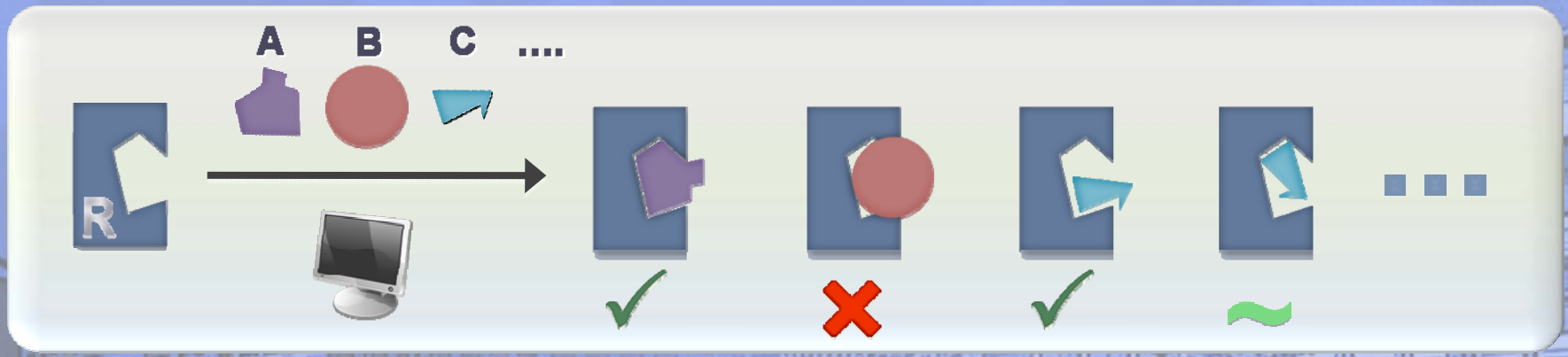
CHEMOMENTUM GRID Technology

- *CHEMOMENTUM Workbench* - Grid services based environment to enable automatic QSAR;
- based on the *UNICORE* Grid middleware;
- high computational power;
- high level of security, through UNICORE certificates;
- generic, flexible system for running workflow-centric, complex applications.
- web site: www.chemomentum.org

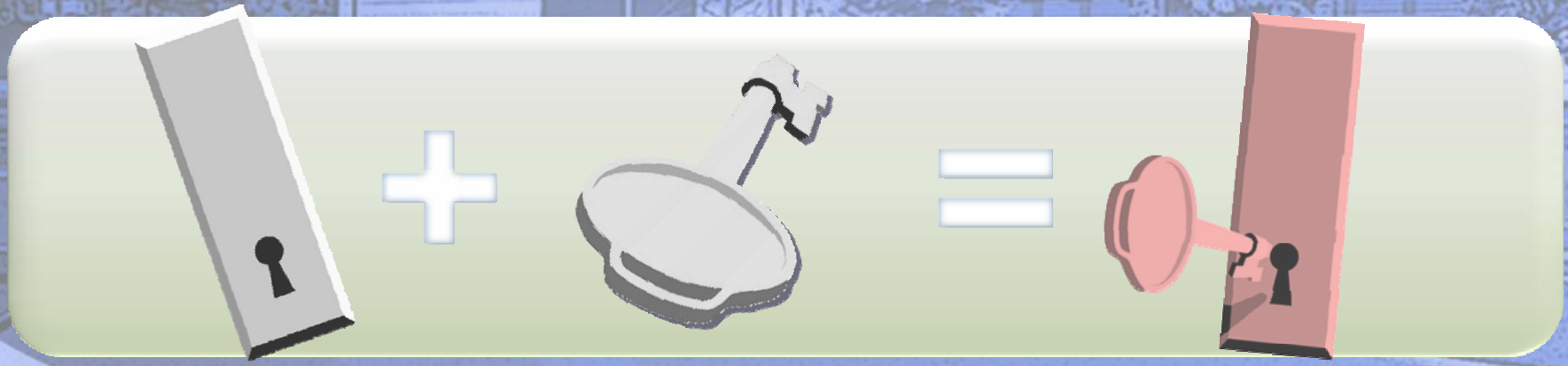
CHEMOMENTUM is useful for *REACH*

- Processing many chemicals
- User-friendly tool
- Report on the workflow, reproducible

Tools for Pharma: *Virtual Screening*



involves the computational analysis of chemical databases to identify compounds with the potential to bind to a given biological receptor



Virtual Screening: Approach used

X-ray crystallographic structure of the receptor is known

or

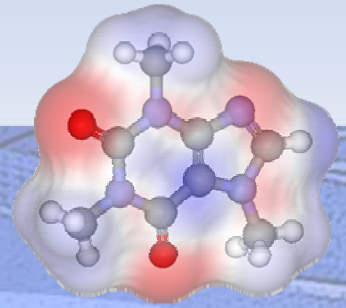
Closely related structures are available (Homology Model)

YES

NO

RECEPTOR FITTING

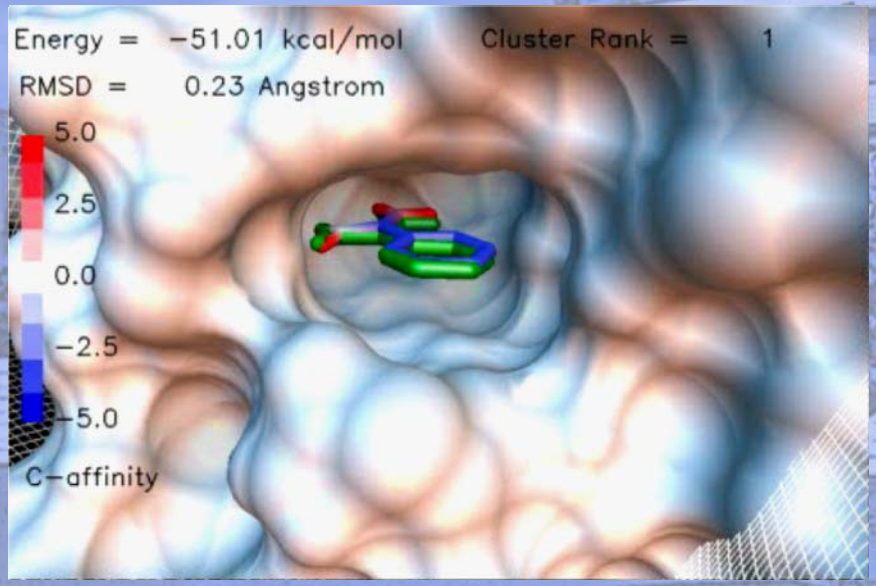
RECEPTOR MAPPING



MARIO NEGRI

Receptor fitting (docking)

computational methods for finding the best matching between two molecules: a receptor and a ligand



a scoring function is used to quantitatively rank the ligands according to binding affinity

Specific opportunities for GRID

- Millions of candidate compounds
- Higher complexity of tools
- Different uses: internal development/registration
- Different confidentiality levels

CONS of *in silico* models

- *No experimental value*
- *Need of a validation process*
- *Need of a better definition of the RELIABILITY*
- *Improve acceptability*
- *Less training and education available*
- *Limited models available*

MARIO NEGRI

PROS of *in silico* models - I

- *Cheap (sometimes FREE)*
- *FAST (sometimes < 1 second; batch process)*

- *No synthesis is required*
- *No ANIMAL USE*

PROS of *in silico* models - II

Worldwide access

It's SAFER

- NO LAB
- NO SOLVENTS
- NO LAB POLLUTION
- NO POSSIBLE CONTAMINATION

It can produce new knowledge

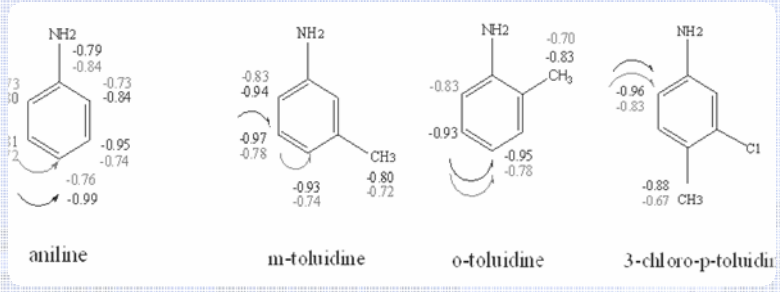
- IT PREDICST EFFECT
- IT EXPLAINS EFFECT

PROS of *in silico* models - III

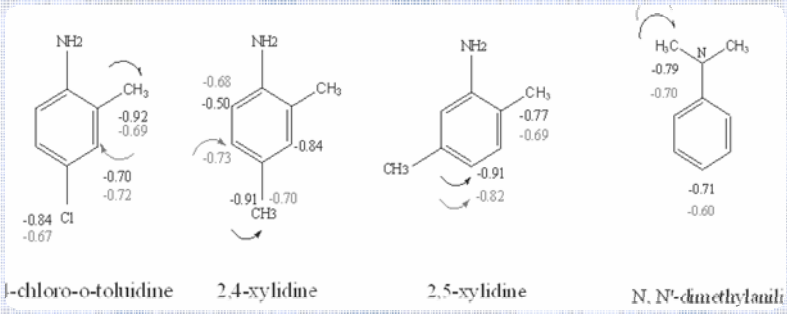
- *It can be PROACTIVE*
- *It can describe the observed phenomenon in MATHEMATICAL TERMS*
- *It can address MULTI TARGETS & DIFFERENCES*
- *It can address HUMAN ENDPOINTS*
- *It can EVOLVE and BE HIGHLY IMPROVED depending on the NEW CHEMICALS*

PROS of *in silico* models - IV

It can distinguish between different ANIMALS



for instance…

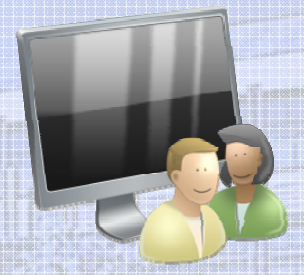


…RAT and MOUSE

It can incorporate results from OTHER METHODS, in an INTELLIGENT STRATEGY, addressing MORE COMPLEX TARGETS

Conclusions: possibilities for GRID - I

- ***Interaction human/computer***:
  extraction of knowledge
  from computer methods
  and human-based rules specified



- ***Interaction/platform***:
  Researchers
  Regulators
  Industry

Conclusions: possibilities for GRID - II

Starting point must be the identification of application/regulation and specific needs;

The definition of the use is a fundamental support to assessment; Models for screening (lower accuracy) and substitutive models (higher accuracy / sensitivity / specificity);

Attention to the usability and utility of the model

Future perspectives for GRID

MUCH HIGHER COMPLEXITY
FURTHER CHALLENGE

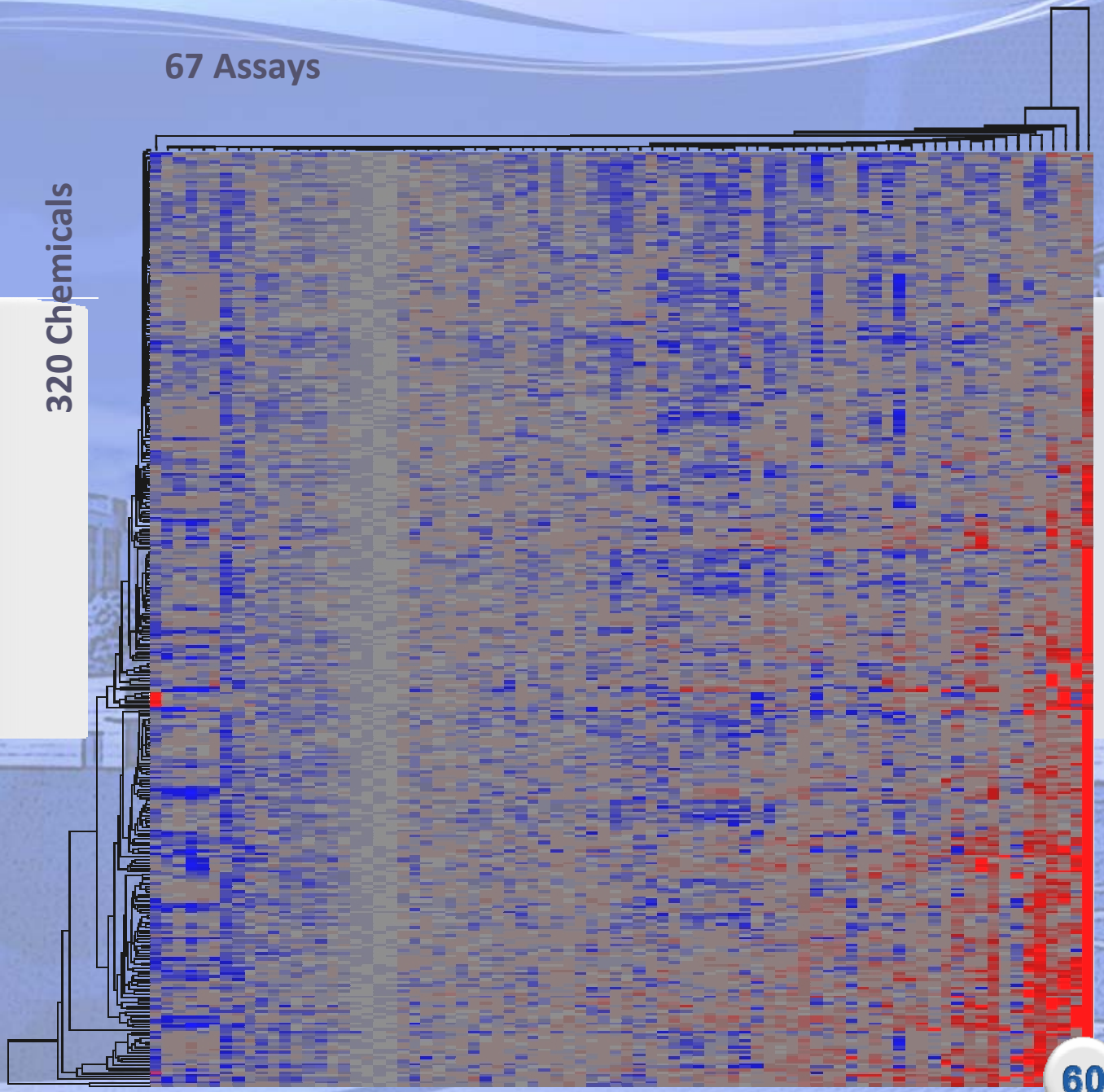
Omics

Proactive approach for synthesis

Combined features

Many endpoints

Many chemicals




MARIO NEGRI
ISTITUTO DI RICERCHE FARMACOLOGICHE
GRAZIE!
UNICORE Summit 2008
Las Palmas de Gran Canaria, Spain - August 26, 2008
UNICORE SUMMIT