BM Systems

PREDICTIVE INTEGRATIVE BIOLOGY

Methodology of Mapping in Integrative Molecular Biology.

the Differences & Complementarities
Between « Heuristic » and « Mathematical » approaches
Concepts & Examples

19th European Congress of Psychiatry – EPA 2011
Vienna Austria, March 12-15 2011

Author-speaker: Dr. François Iris: Founder & CSO BMSystems
To contact us: manuel.gea@bmsystems.net

The problem of Complex System Analysis and Biological Modelling.

If you dream to create the first operational bird model…

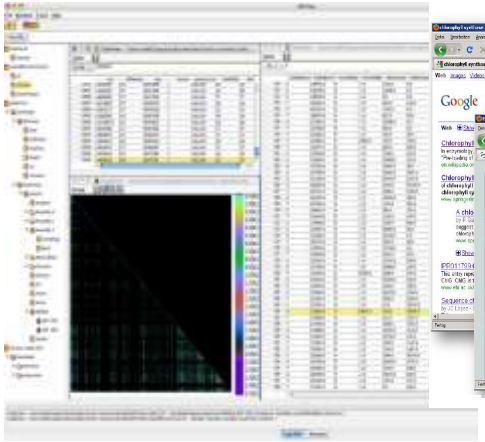
… a "basic" living Complex system that not only flies…

Be sure to use the appropriate modeling concepts & tools. If not…

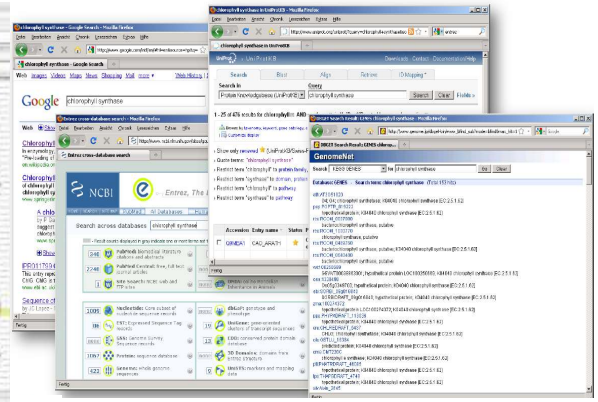
…you get a Complicated "Cartesian" system. It does fly, but…

The challenge is clearly not a question of technologies only

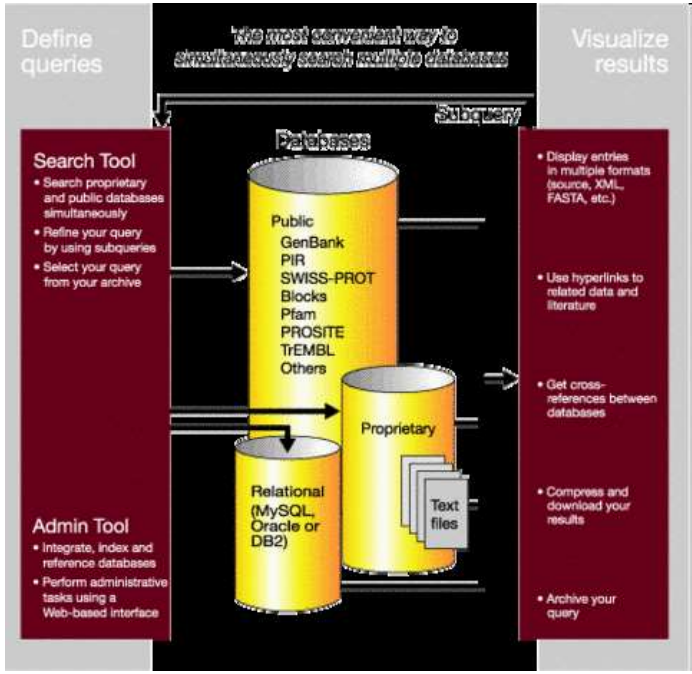
Analysing & integrating data.



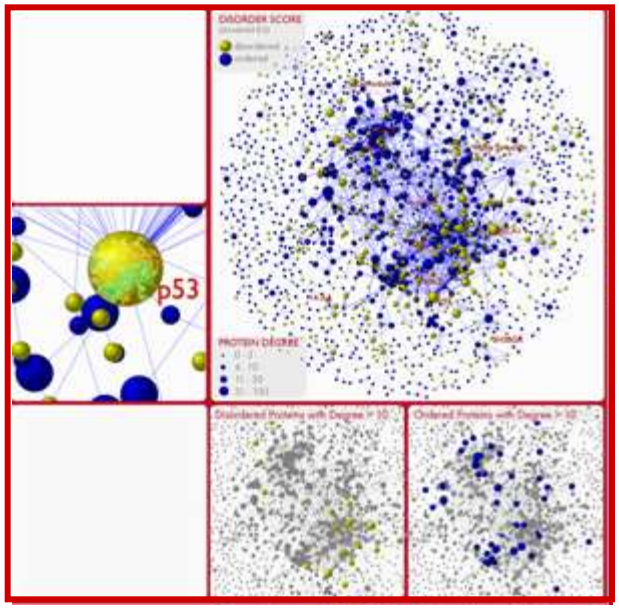
Data flatfile



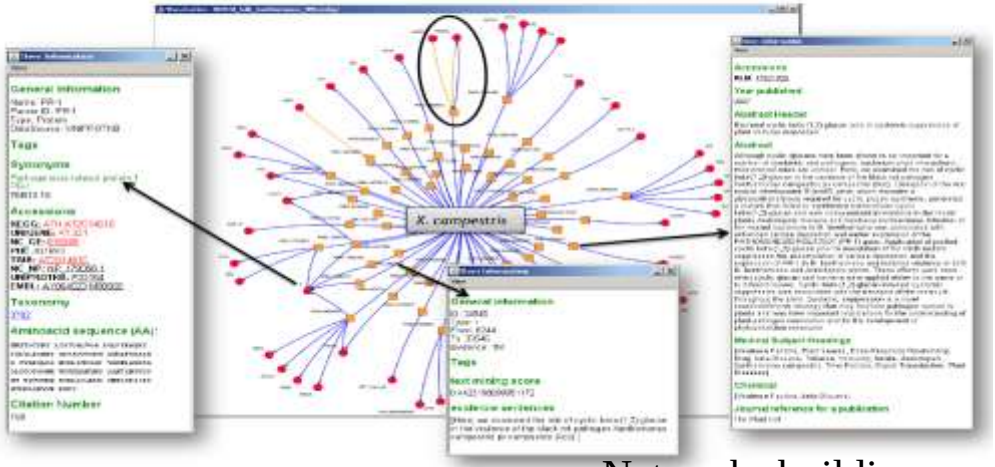
Literature sources & text mining



Analytical Web resources



Visualisation tools

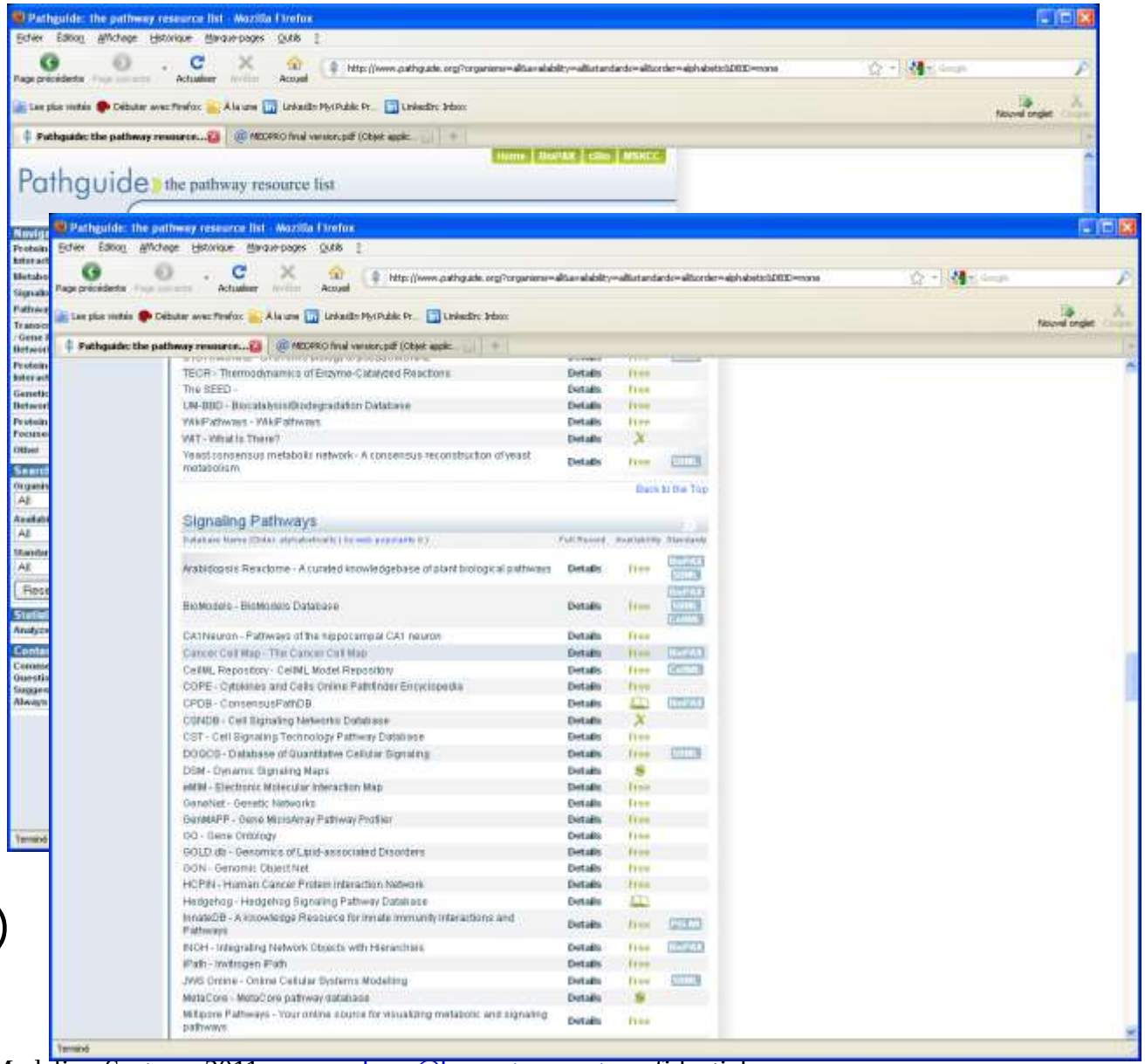


Networks building

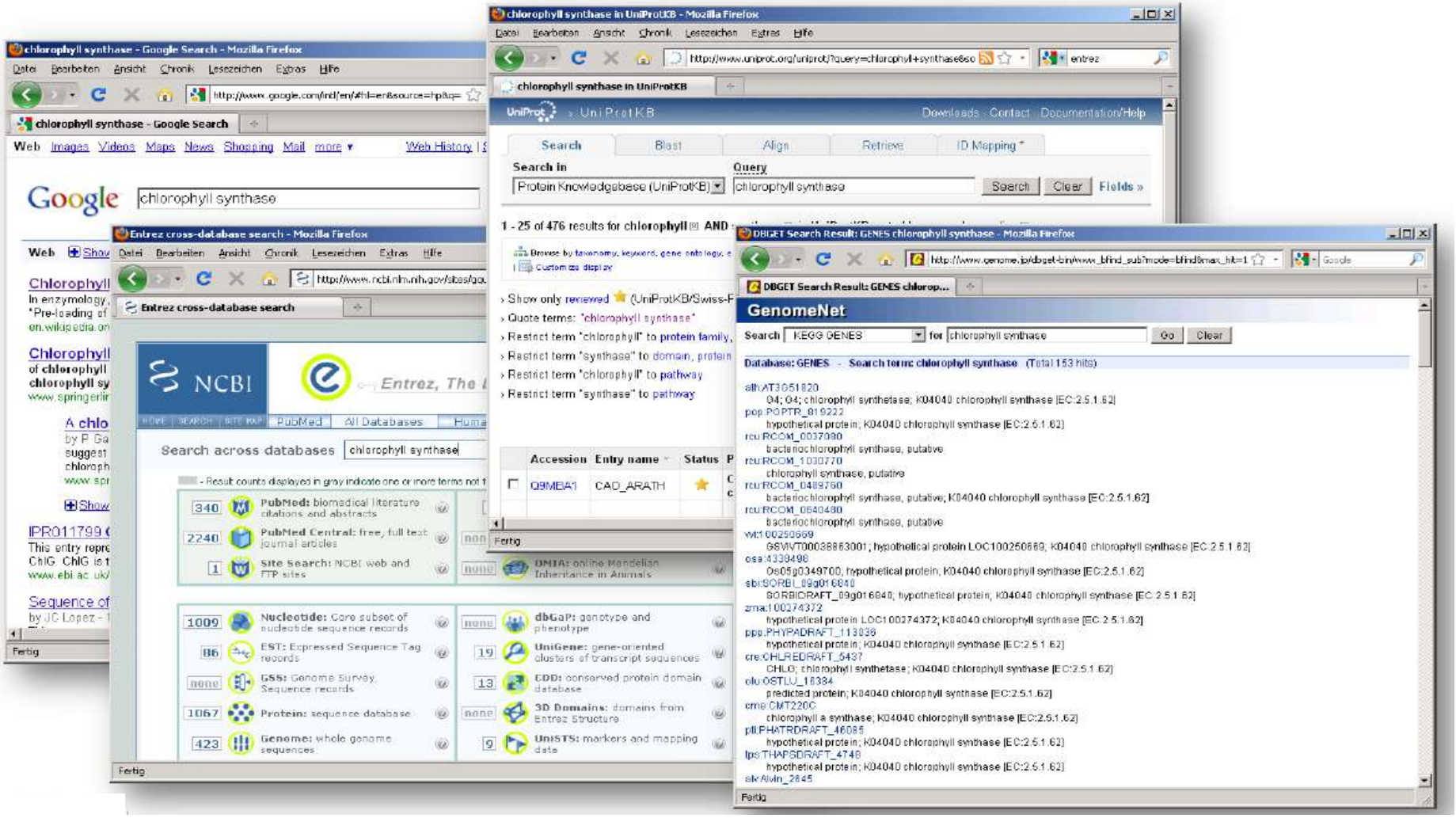
Currently, some **317** web resources (databases) provide access to

- Thousands of pathways and networks, documenting
- Millions of interactions between proteins, genes, small molecules..

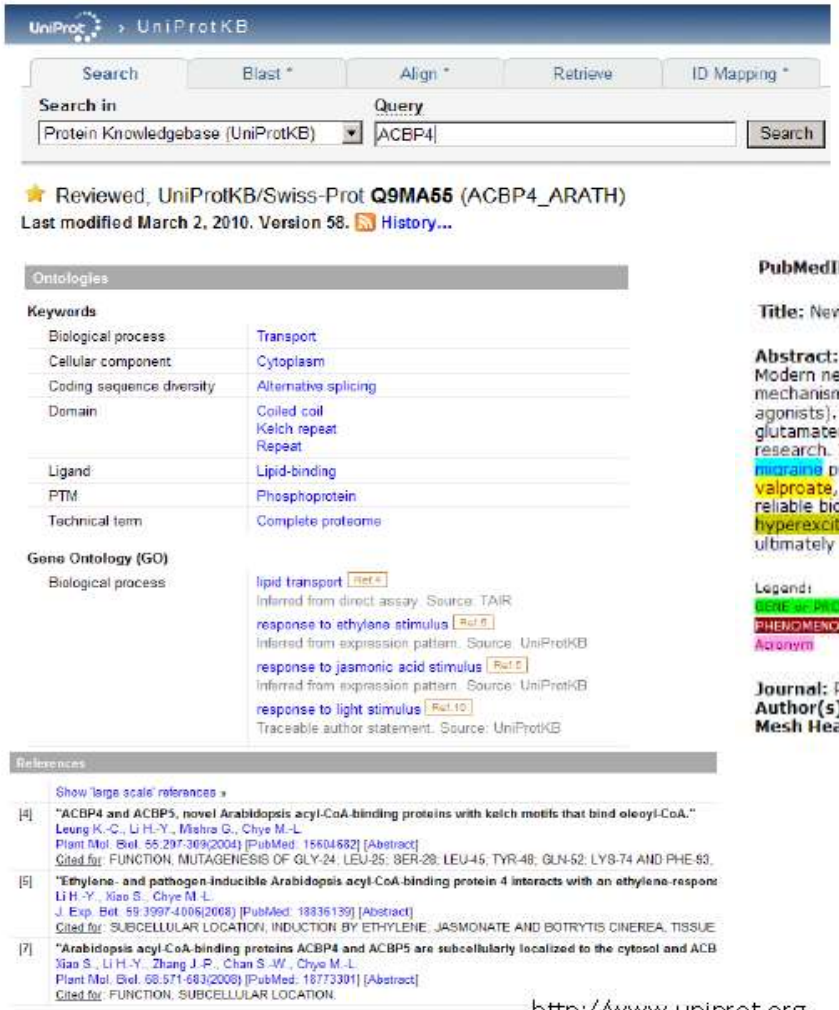
(http://pathguide.org/)

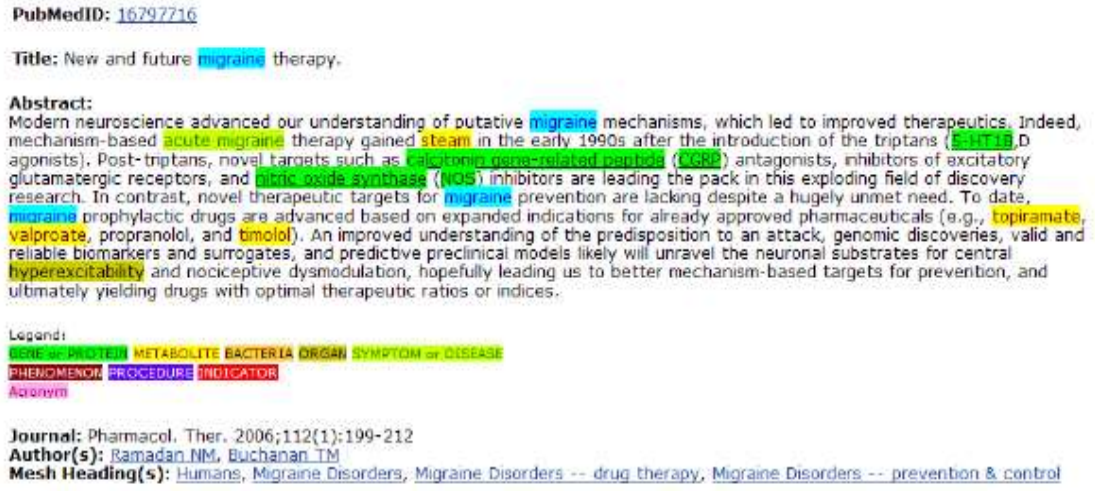


The challenge is to use as many of these biological databases as possible, CONCURRENTLY!



In practical terms, it means extracting as much information as possible from archived records.





http://www.uniprot.org

The Classical Process

The classical data mining approach pre-supposes that

- All reported information is *a priori* valid, and
- There are well defined rules for expressing relationships between components and they are practically always obeyed by authors.

Hence the systematic application of mining algorithms such as

- **Distance of co-occurrence :**

$$\log\left(\frac{N_{phr}(G).N_{phr}(G')}{N_{phr}(G \text{ et } G')^2}\right)$$

- **Functional relationships:**

$$N_{phr}(G_1, G_2, R)$$

to all 'omics data, from transcriptomics to proteomics & metabolomics

Complex System Analysis and Biological Modelling

It is first and foremost integrating masses of information

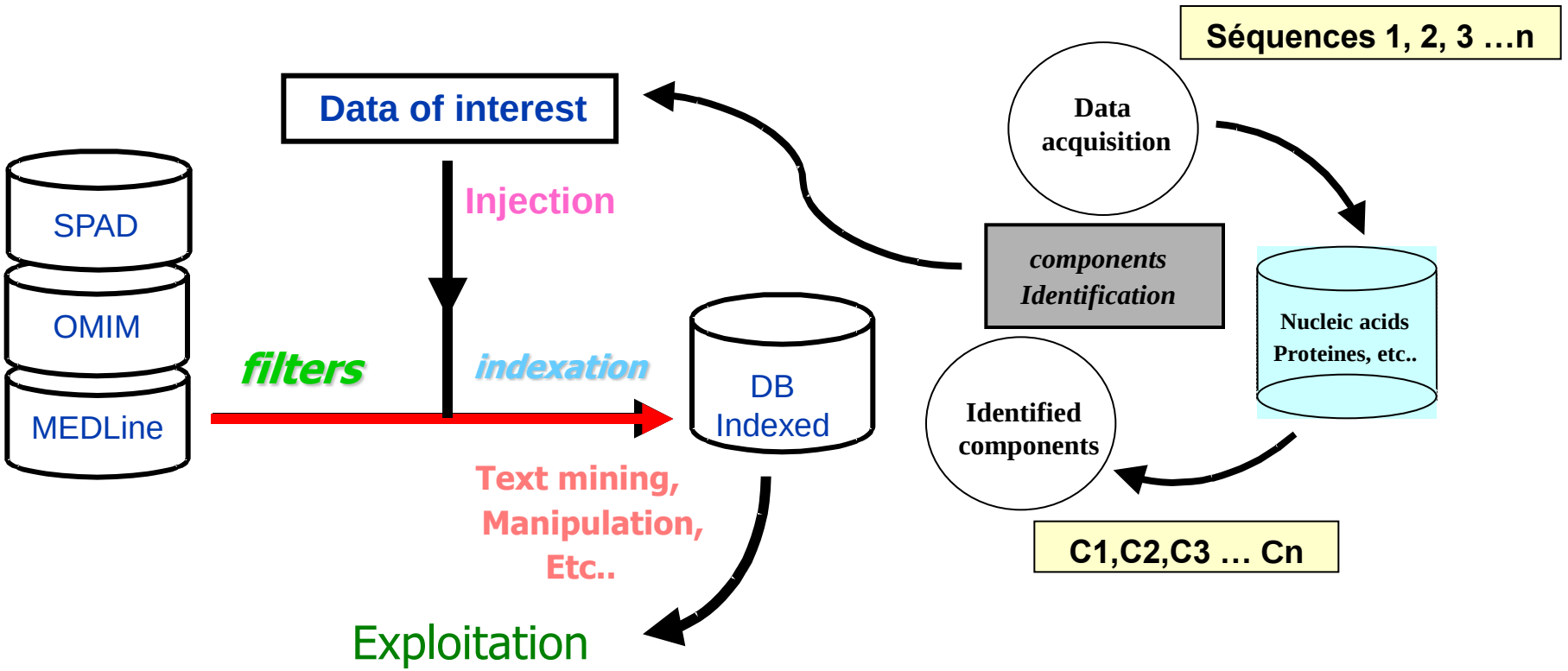
In the bio-medical realm, the relevant information touches on every biological domain, from anatomy to molecular biology and structural biophysics, spanning a multitude of functional contexts, corresponding to an enormous complexity.

The available information is necessarily **ALWAYS**

- **incomplete** to an unknown extent;
- **biased** to an unknown extent; and
- **erroneous** to an unknown extent.

All that goes by the name « ***Information*** » is not necessarily **Useful** and/or **Utilisable**

The Classical Systems Biology Process



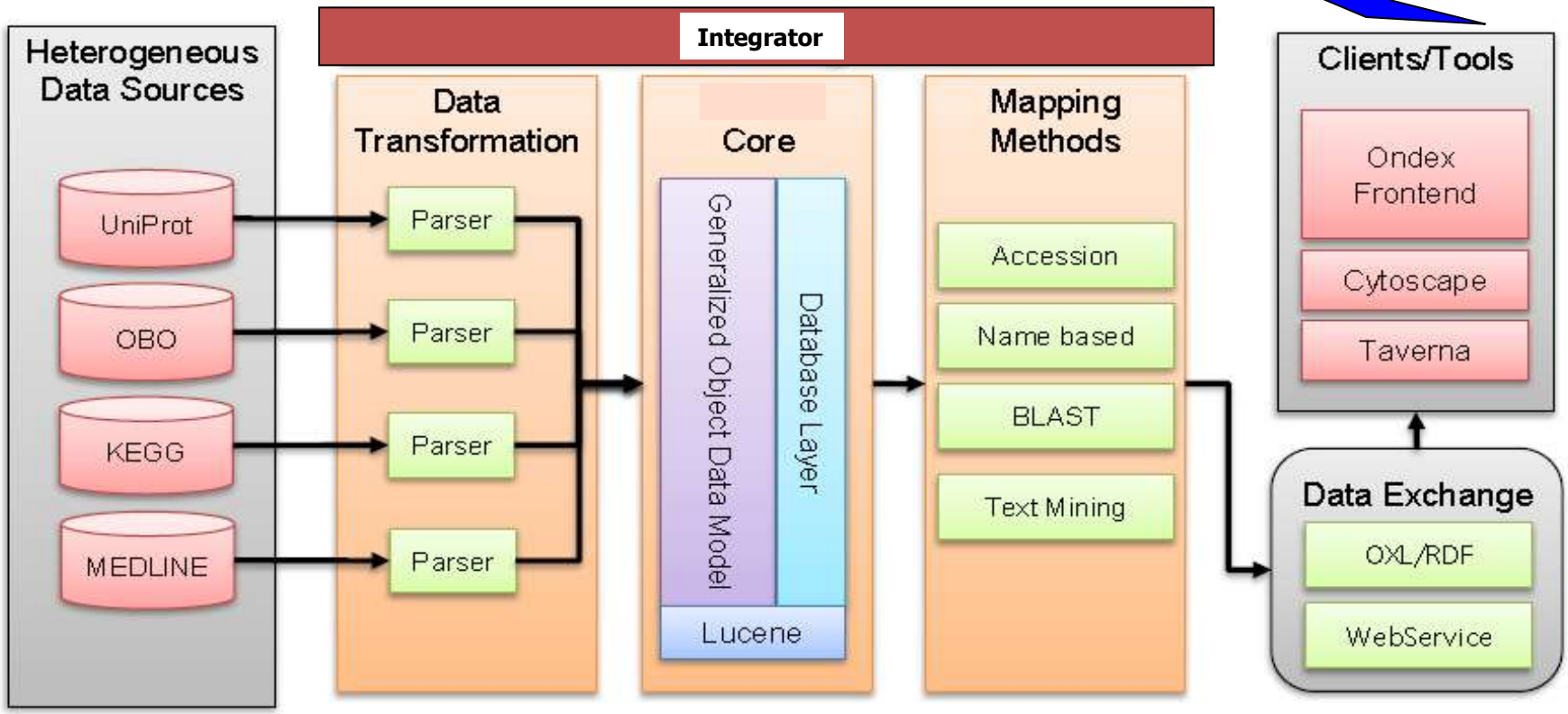
What constitutes a GOOD filter?

What constitutes a GOOD indexation strategy?

ALL information entered into the DB is ALWAYS *biased*, *incomplet*, *erroneous*, etc…

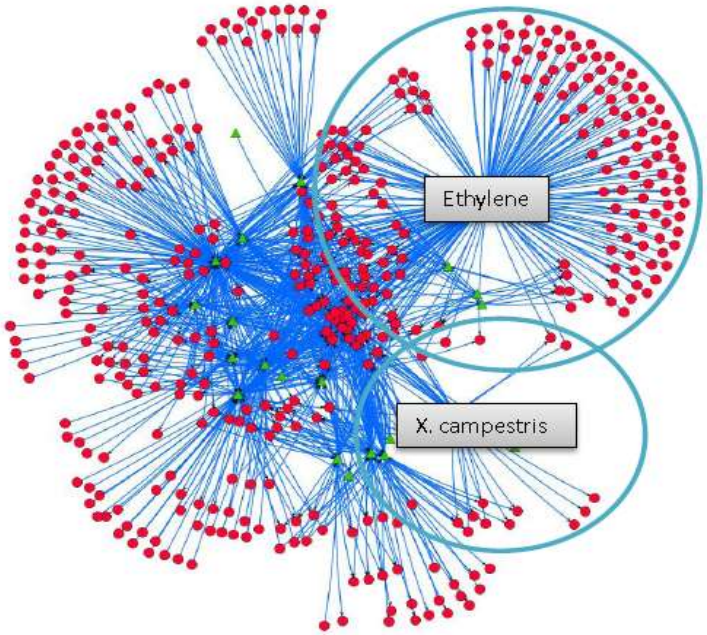
} Accumulation of inconsistencies

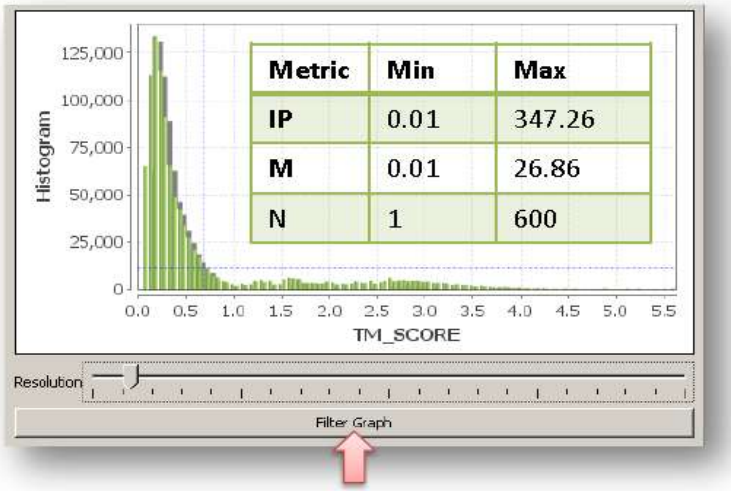
This, naturally, generates coherence issues affecting the entire analytical chain of events, irrespective of the visualisation and manipulation tools finally used!



So we not only end-up facing ever more intricate networks which require massive manual curation to become utilisable.

Protein-Stress Association Network

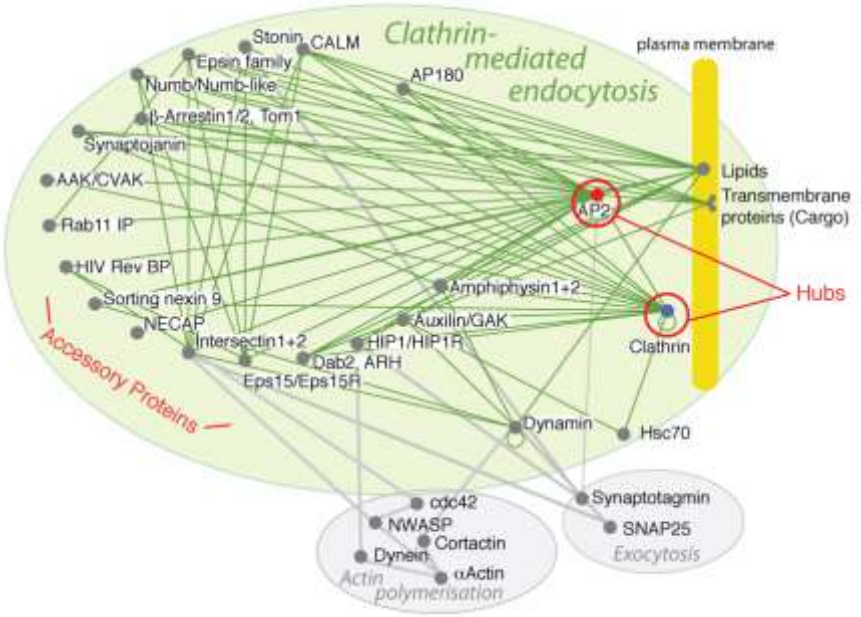




- 3145 proteins linked to 32 stresses by 10777 relations
- On average
  - each protein associated with 3.4 stresses
  - each stress associated with 337 proteins
- Filtering associations based on three scoring metrics IP, M and N
- Which metric and cut-off are most suited for filtering noise?

IP= Building blocks; M = distance of association; N= interactions

This also results in a highly misleading vision of protein interactions & networks.



How can a single hub protein bind so many different partners?

The problem is largely non-existent and resides in the construction and the representation of protein interaction networks.

Proteins derived from a single gene, even if different, are clustered in maps into a single node.

This leads to the impression that a single protein binds to a very large number of partners.

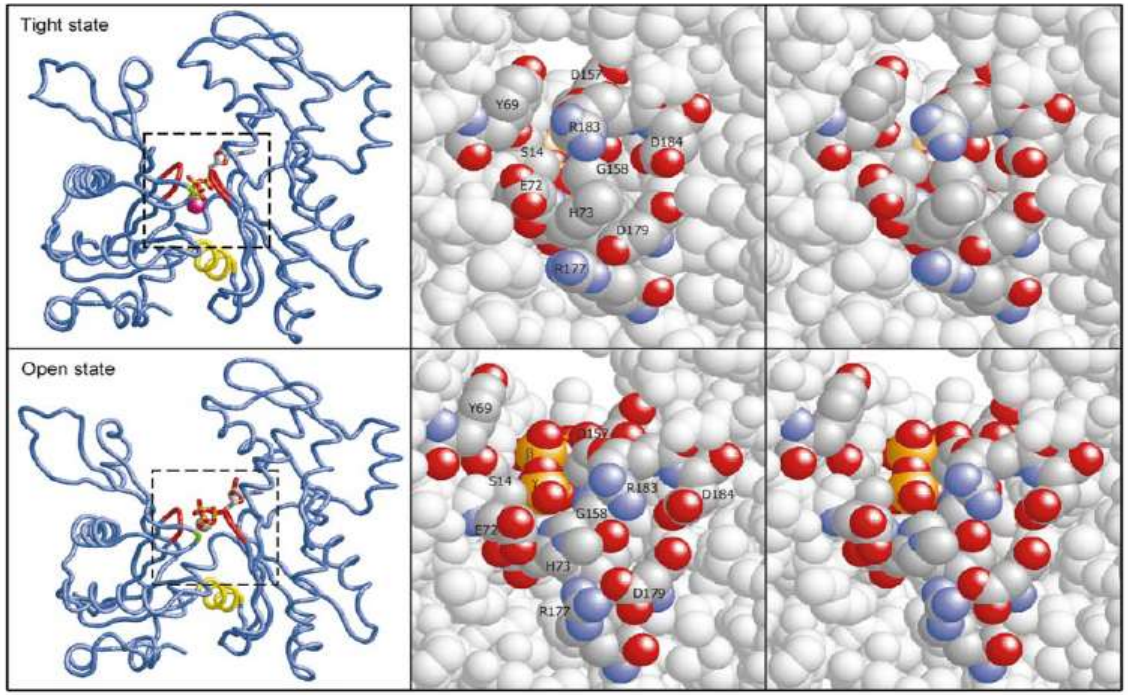
In reality, it does not.

Protein networks reflect confusions involving combinations of multiple distinct conformations of a same protein as well as functional plasticity addressing distinct proteins encoded by one gene.

One protein = Functional plasticity

Post-transcriptional modification of Actin

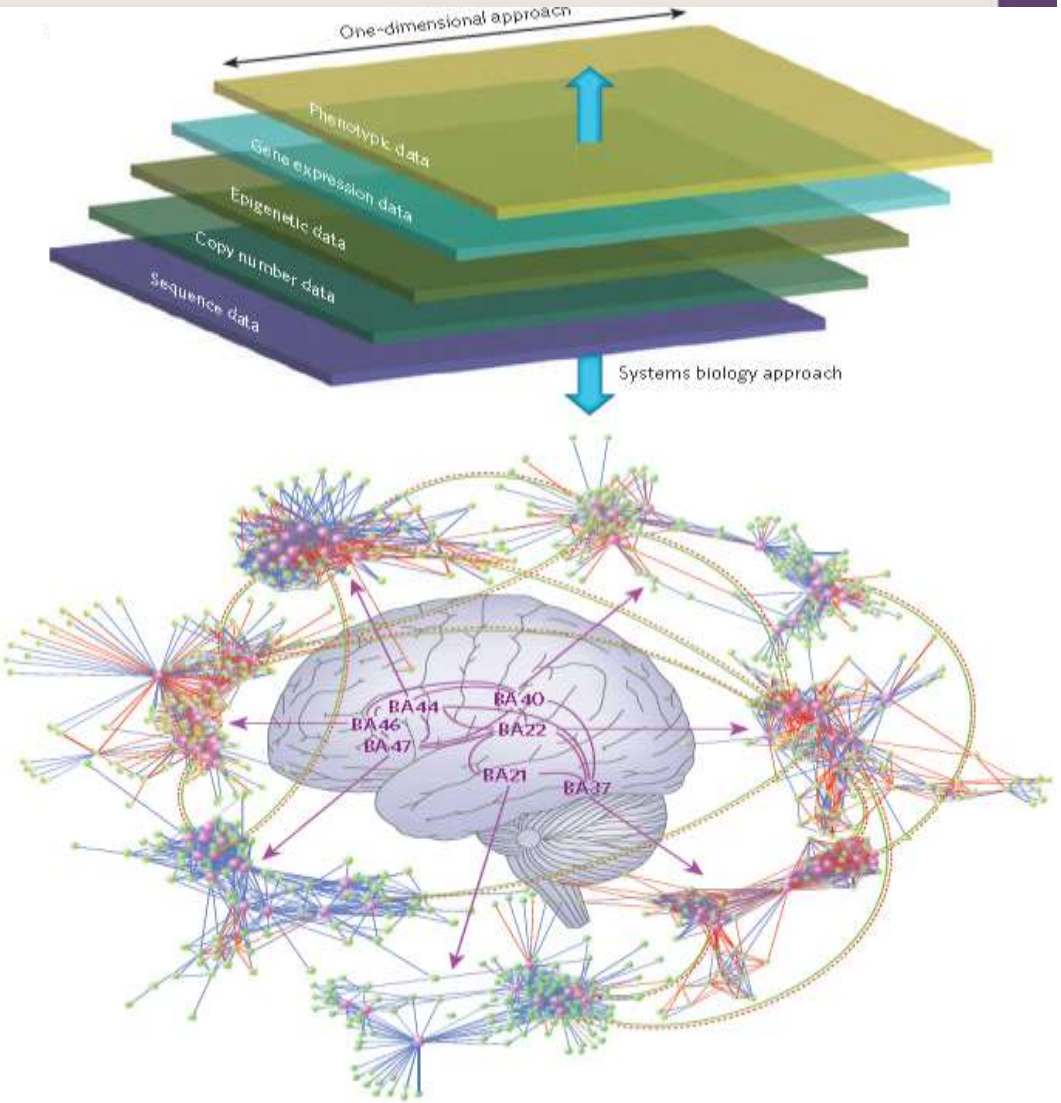
Actin His73 methylation = ATP exchange rate increased, ATP hydrolysis and phosphate release prior to and independent of filament formation while polymer formation delayed.



Cytoskeleton dynamics significantly altered.

DCD44 glycosylation patterns = drastically different effects on regulation of immune cells.

As a result, when it comes to really complex systems and poorly defined physiological processes, such as those associated with functions within the CNS and disorders thereof, Bayesian approaches lead to nearly intractable difficulties.

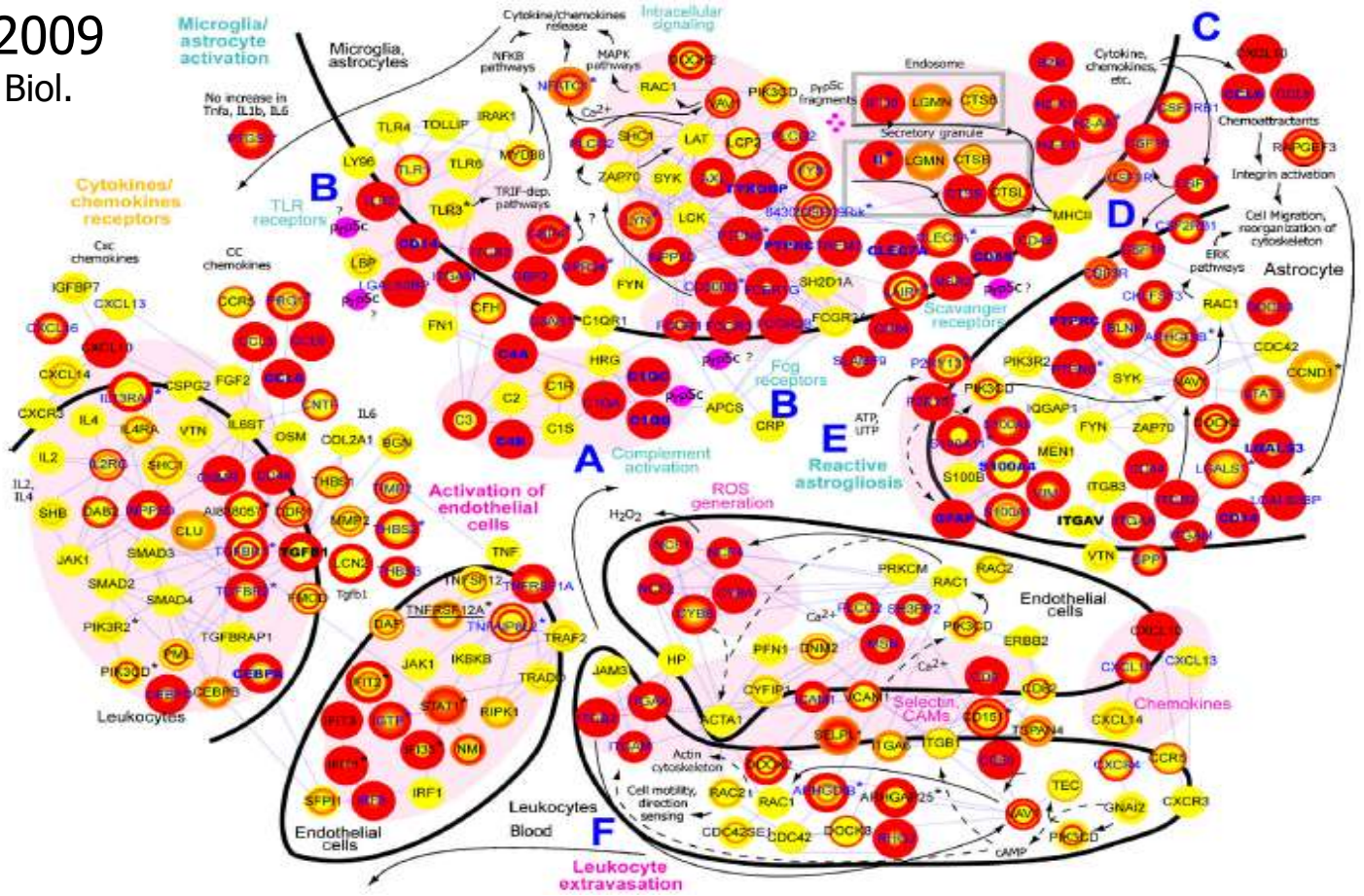


Mathematical models are remarkable validation/fine-tuning tools when applied to well defined processes.

They are inadequate discovery tools when applied to poorly understood multicellular processes.

An example of what must be avoided: the systems biology of Creutzfeld-Jakob Disease.

This Bayesian model was published in 2009 (Hwang D et al. Mol Syst Biol. **5**:252. PMID: 19308092)



Constructed from a list of 333 shared DEGs and protein–protein interactions information from novel targeted experiments and public databases, it describes the networks that are potentially involved in the activation of microglia and astrocytes.

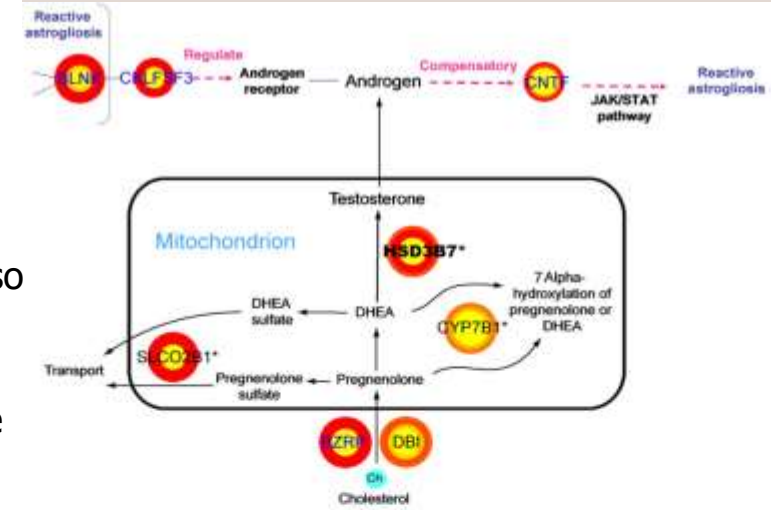
Although it certainly does lead to sub-networks that do make some sense, these cannot

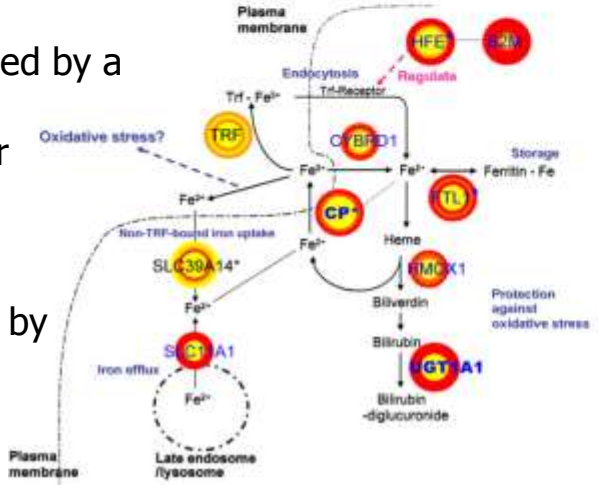
1) be distinguished from Alzheimer's Disease, also characterised by astrogliosis, nor

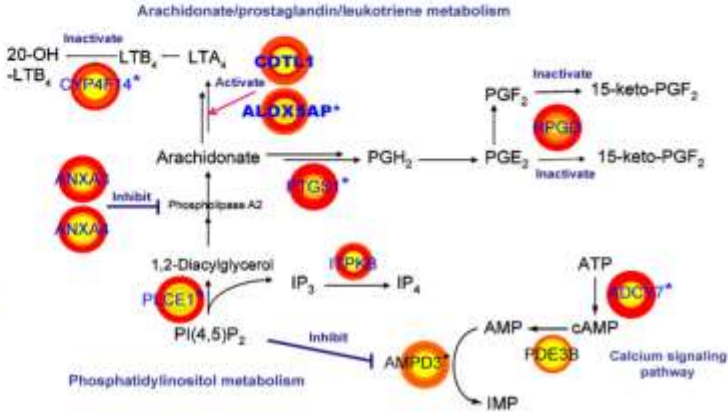
2) begin to explain why neurons are killed in the CNS and never in the spinal chord, nor

3) why the disease is characterised by a spongiosis absent in other neurodegenerative disorders, nor

4) why the disease progresses silently for many years followed by a sudden very rapid clinical progression.





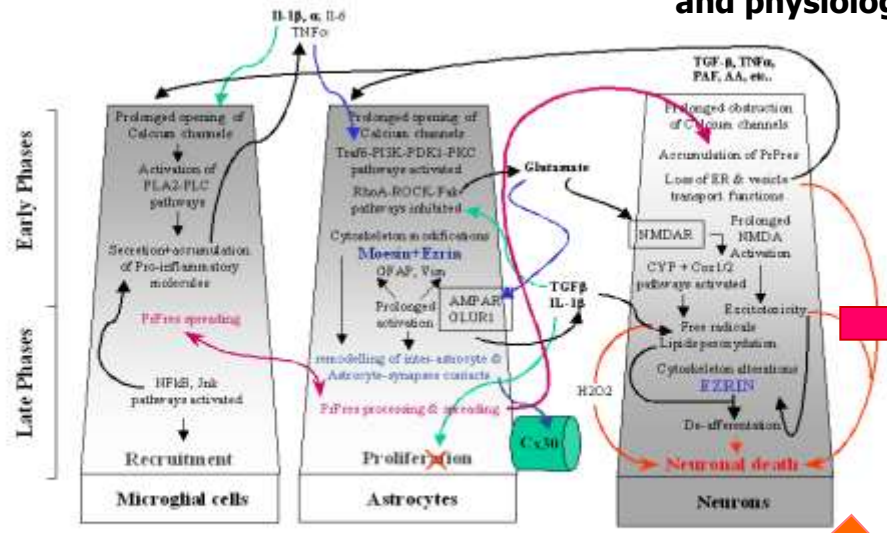


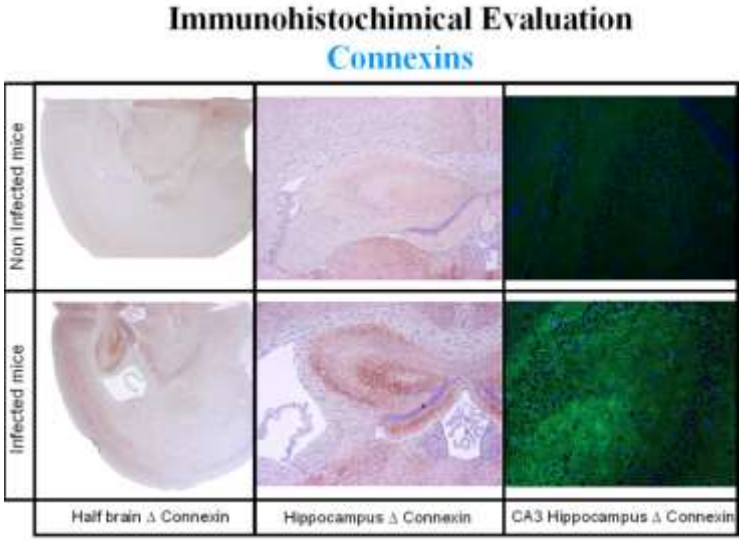
Mathematical models are remarkable validation/fine-tuning tools when applied to well defined processes.

They are inadequate discovery tools when applied to poorly understood multicellular processes.

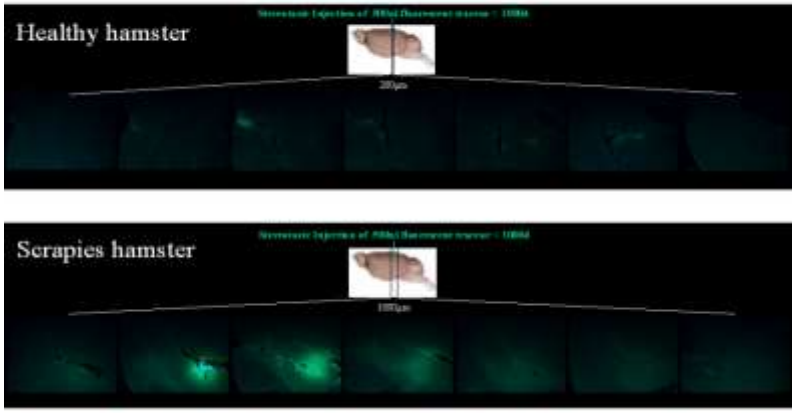
As compared to the heuristic CJD model finalised in early 2008.

Which predicts and explains the pathological mechanisms at both molecular and physiological levels.

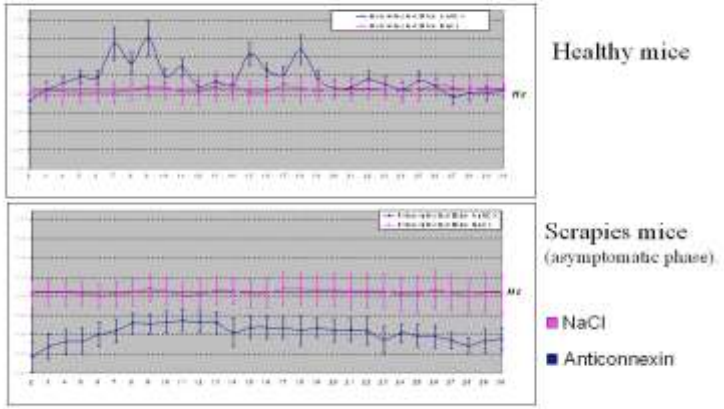




Functional modifications of glial connexions



Functionnal EEG alterations linked with modifications of glial connections.



Source:
BMSys & CEA

The differences between « heuristic » and « mathematical » approaches.

Heuristics:

A problems solving approach evaluating each step in a process, searching for satisfactory solutions rather than for optimal solutions, using all available *qualitative* information instead of *quantitative* information.

Thus,

Heuristic modelling starts from accumulated information to produce a model capable of describing the mechanisms that generated the observed outcome / data and predict their modifications associated with a different outcome**;**

It plays the role of an architect.

While

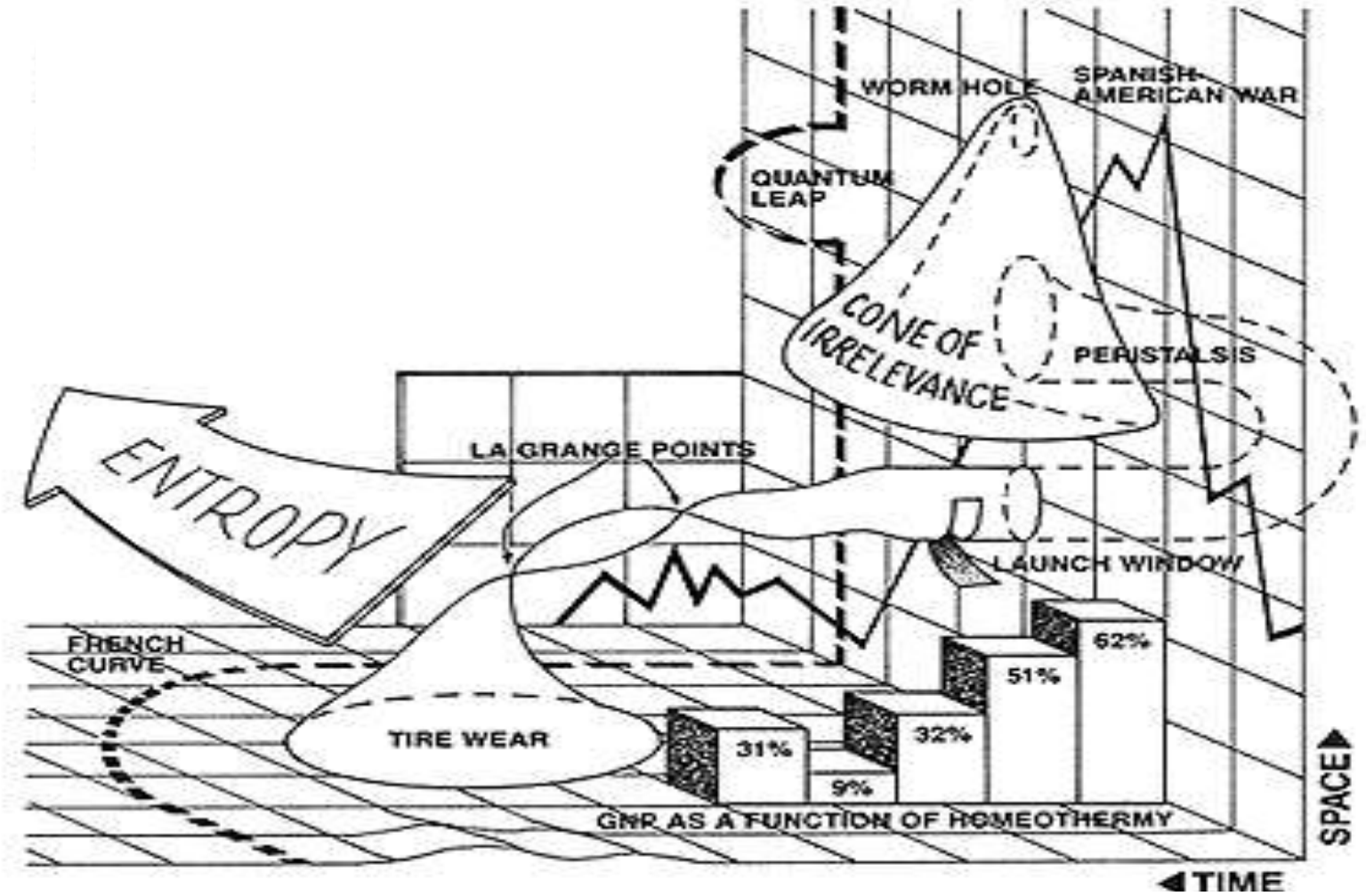
mathematical (Bayesian) modelling starts from quantitative data to produce models capable of reiterating this data and predict the outcome of a different experimental paradigm.

It plays the role of an engineer.

Hence

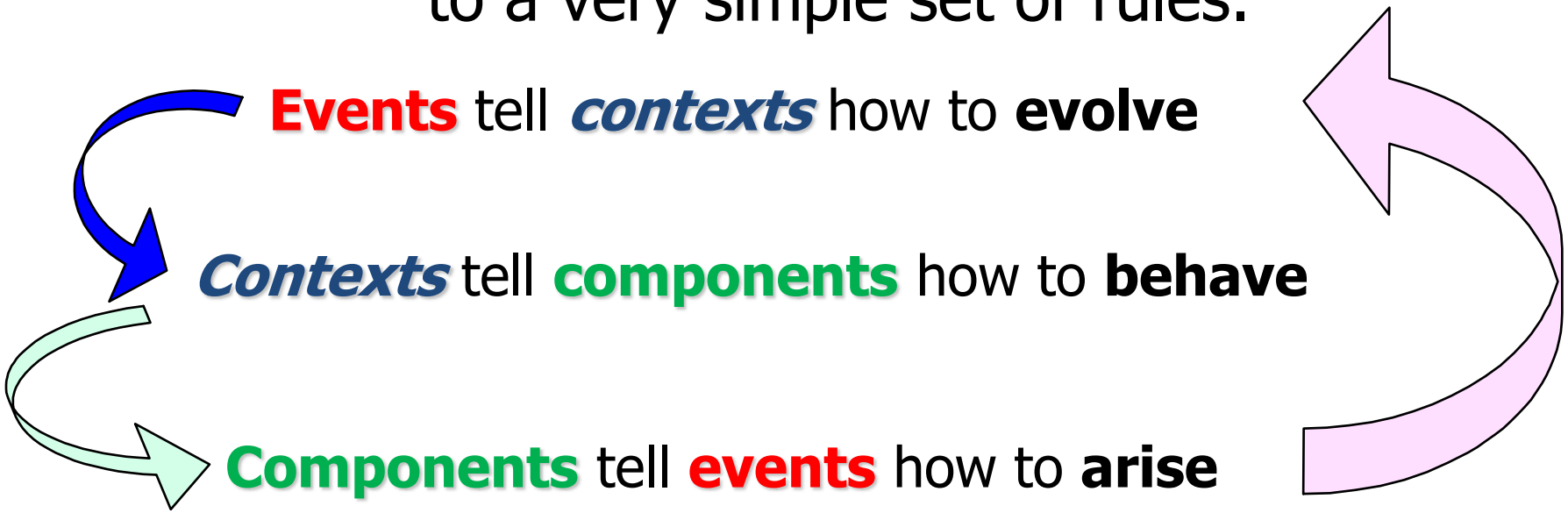
Far from being incompatible, these two approaches are complementary.

The heuristic analytical process must follow a « relativistic » approach.



Within this framework, *Non-linearity*, *Irrelevance*, *Wear*, *Relative weights* & *Contexts* are key concepts.

And the « relativistic » approach leads to a very simple set of rules.

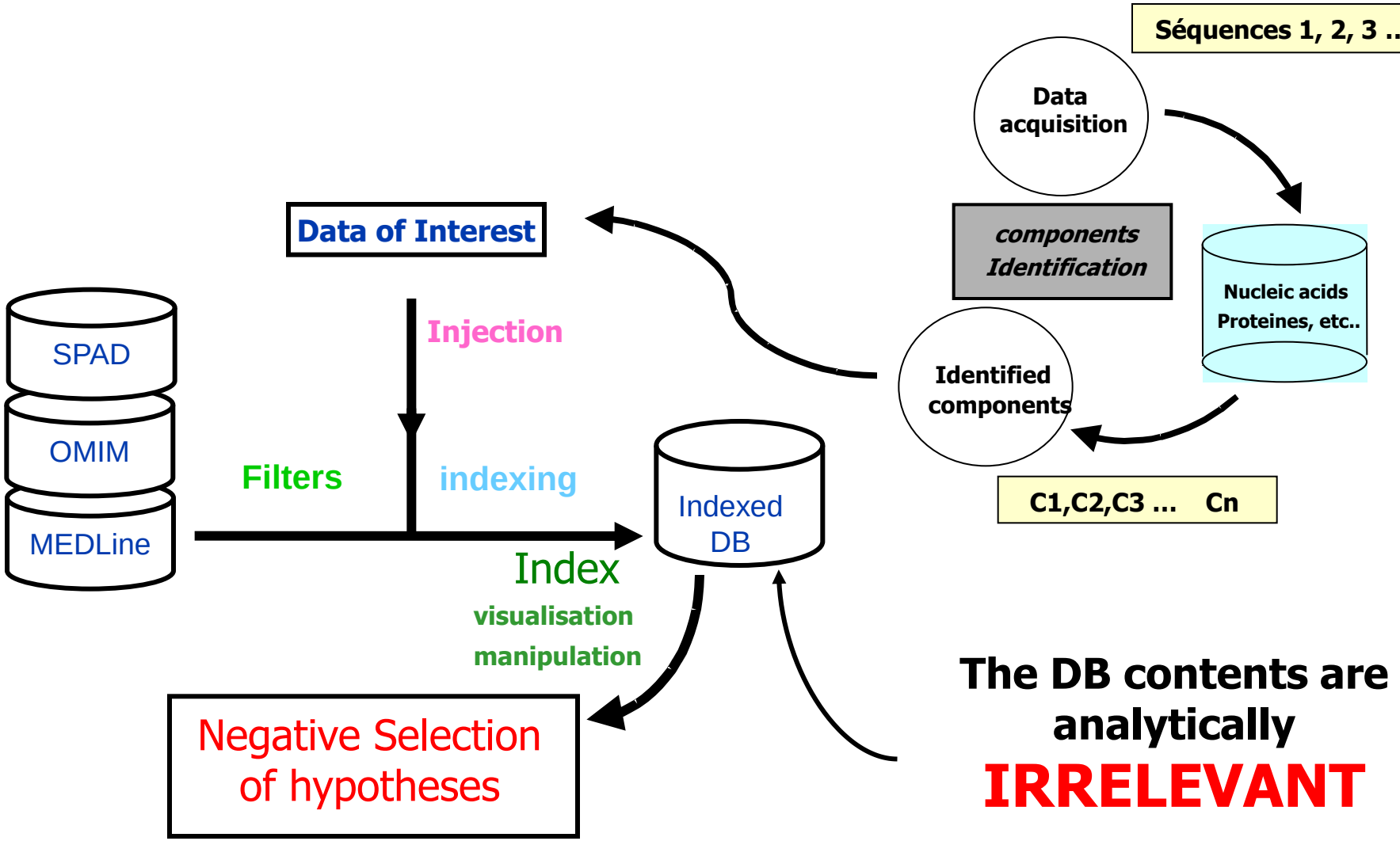


Analyses in terms of biological components and functions are now IRRELEVANT.

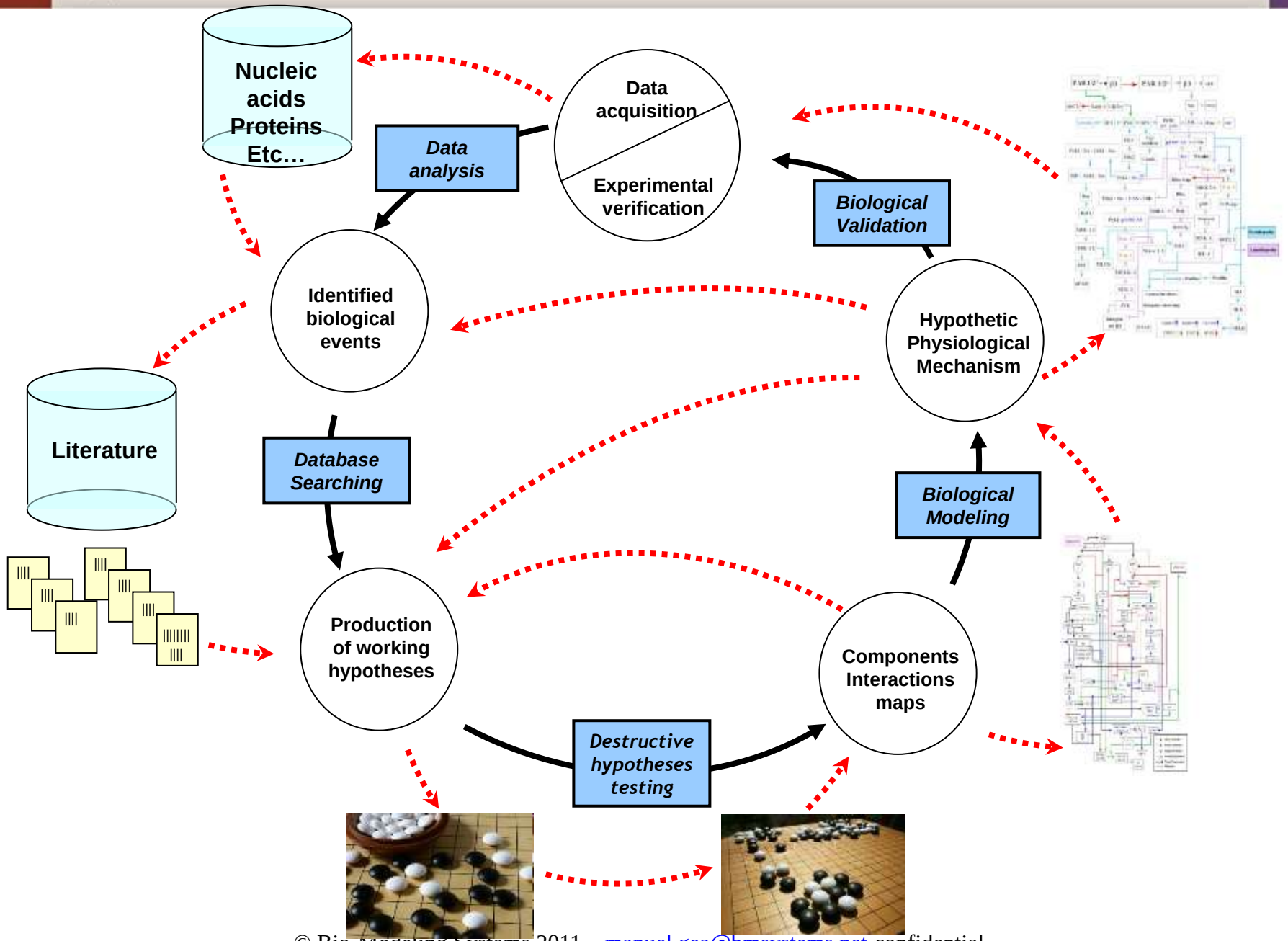
EVENT-DRIVEN analytical approaches become necessary.

This, in turn, imposes analytical procedures based upon the negative selection of working hypotheses.

The *CADI™* Integration & Modelling Process.

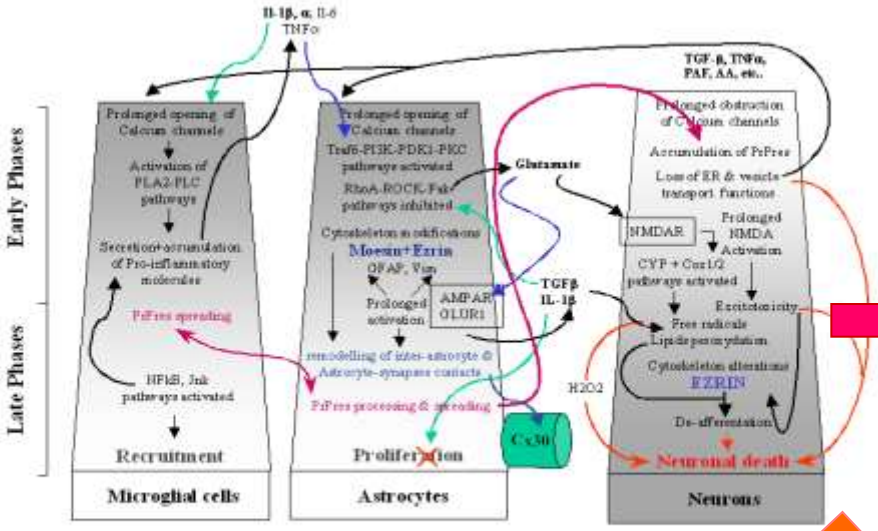


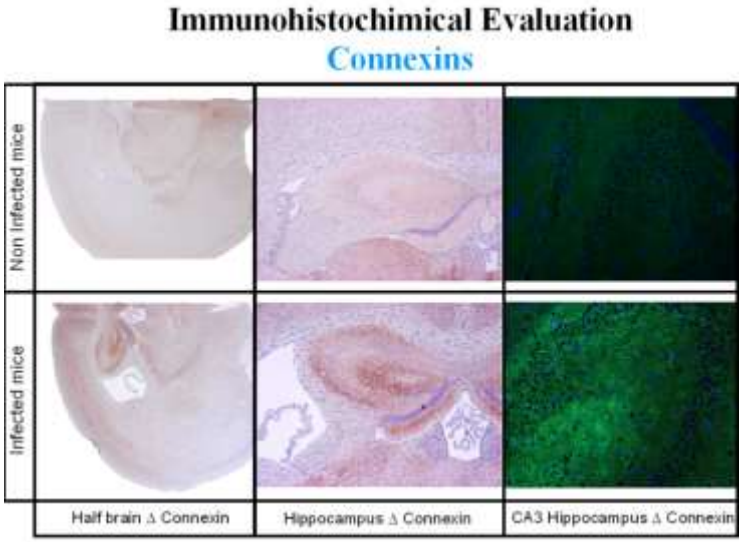
To arrive at a model that can be biologically challenged!



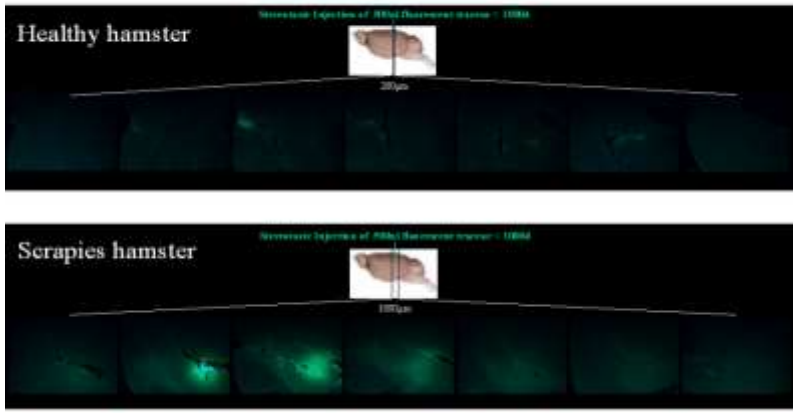
The heuristic model of CJD finalised in early 2008.

It predicts and explains the pathological mechanisms at both molecular and physiological levels.

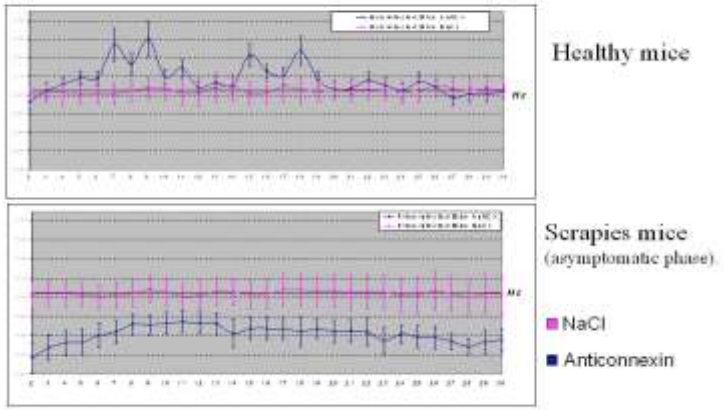




Functional modifications of glial connexions



Functionnal EEG alterations linked with modifications of glial connections.

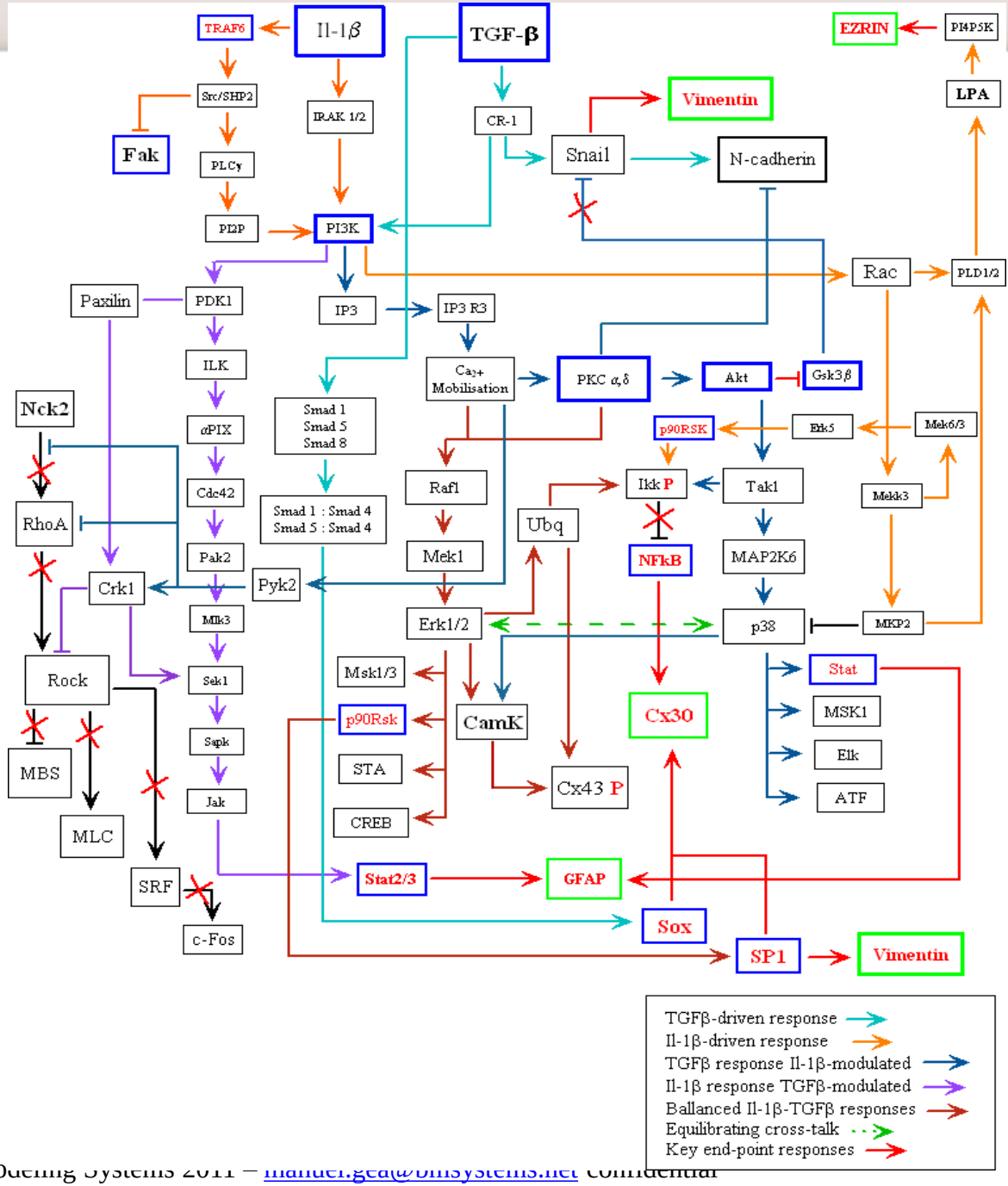


Source:
BMSys & CEA

The main driving mechanism.

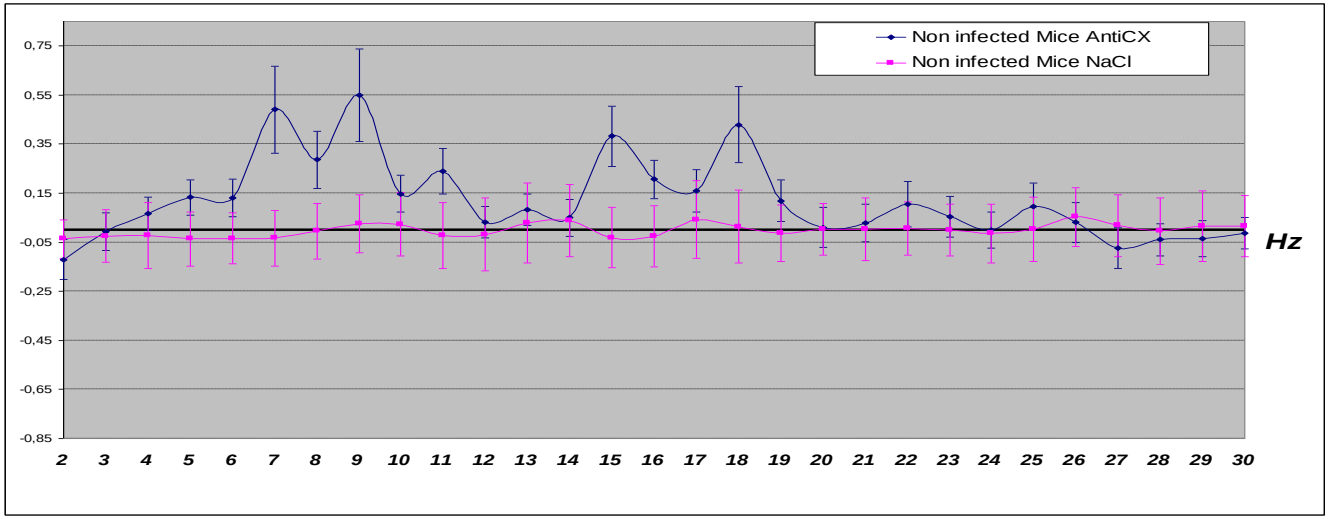
Il-1b-mediated signalling in hippocampus astrocytes

The pathways through which chronic neuronal stress signalling and concurrent glial pro-inflammatory responses lead to reactive astrocyte activation (GFAP + Vim) associated with cytoskeleton reorganisation (ezrin). This leads to a major switch in Cx targeting & distribution, resulting in the formation of a syncythium with massively altered diffusive properties and neurotrophic functions.



Practical consequences.

One of the roles of connexins is to dampen neuronal synchronisation.



Healthy mice

- NaCl
- Anticonnexin

In healthy animals, pharmacological blockade of Cx activity results in quantitative EEG patterns closely resembling an epileptic crisis
(frequency range-specific hyper-synchronisations).

This CEA/BMSystems collaborative research in CNS (psychiatric and neurological disorders) led to the co-owned patent WO201029131 (Use of anti-connexin agents for modulating the therapeutic effect of psychotropic drugs). Sept, 2008.

The net results.

CJD is not a neurological disease *stricto sensus*.
It is a disease that primarily affects astrocytes structures and functions which, over time, lethally affects healthy glial & neuronal cells through « bystander effects », leading to widespread CNS disorganisation (spongiosis) and functional failure.

But this model also provides an understanding of key mechanisms associated with psychiatric & neurology disorders.

An entirely new approach for their effective treatment was designed, tested <u>in vivo</u> and validated.

Patent covering novel therapeutics for psychiatry & neurodegenerative disorder (CEA/BMSystems).

This CEA/BMSystems collaborative research in CNS (psychiatric and neurological disorders) led to the co-owned patent WO201029131 with a worldwide exclusive license to Theranexus CEA's spin-off currently in Phase II.

Neither of which have much to do with CJD *per se*…

This work received a Bio-IT World « Best Practices » award from the Cambridge HealthTech Institute (USA).

AND

Was selected as 1 of the 3 pan-European « state of the art examples of systems biology approaches of benefit to medicine » by the European Commission's DG Research, Directorate of Health (June 2010).

Example 2:

Understanding the co-evolutionary interplays between bacteria and bacteriophages leads to the discovery of the means whereby ***undefined*** multi-resistant bacterial pathogens can be efficiently controlled.

The questions (French Defence)

- How to rapidly (less than 30 min) and efficiently ***detect*** the presence of any given LIVE bacterial pathogen?

- How to rapidly and efficiently ***destroy*** any ***unknown*** bacterial pathogen or emerging strain ***without using***

A) Antibiotics: too many resistant strains, and very rapid resistance acquisition.

B) Vaccines: much too slow to act, and small strain variations often lead to inefficacy.

In other words, what is required is a "detector-killer".

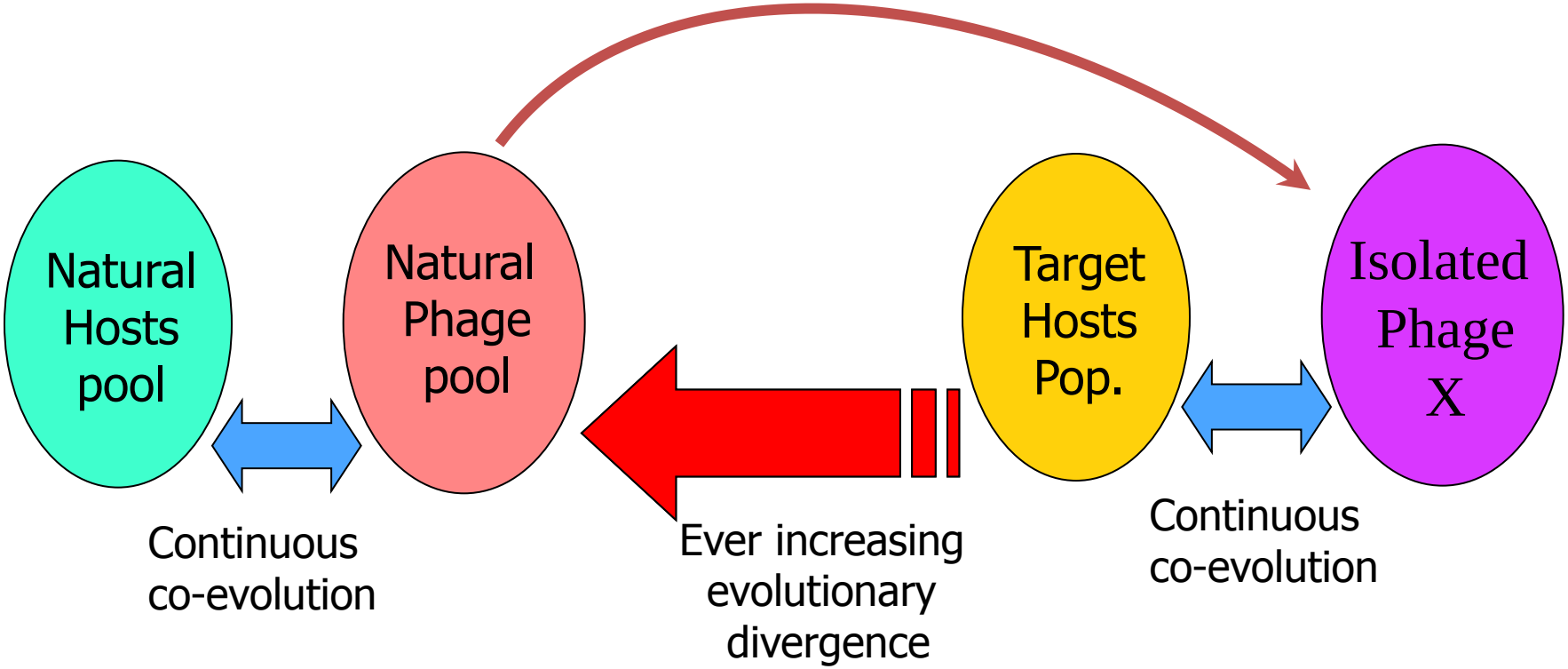
The first apparent answer

Bacteriophages, the natural predators of bacteria, could present the best potential to act as detectors-killers.

- Many are very host-specific,
- They only replicate in LIVE bacterial cells,
- Many kill the cells in which they replicate,
- As the phage progeny population increases that of the target diminishes (in a « closed » environment, few targets, if any, should escape), and
- They are extremely numerous and varied (they probably represent the most numerous « life forms » on the planet).

BUT the matter is NOT as simple as it first appears!

Co-evolution versus unidirectional predatory pressure.



+

Ever decreasing chances to find a new efficient lytic phage.

Bacteria have existed for nearly 4 BILLION Years. They have so far resisted to EVERYTHING.
And it is certainly **NOT** for lack of phages!

The model-derived solution.

- ***What, in essence, is the problem?***

The bacterial targets will try anything to escape predation and we have no idea what will be the successful strategy. Furthermore, this strategy is likely to vary between locations (populations) for a same target.

- ***What do we need to achieve?***

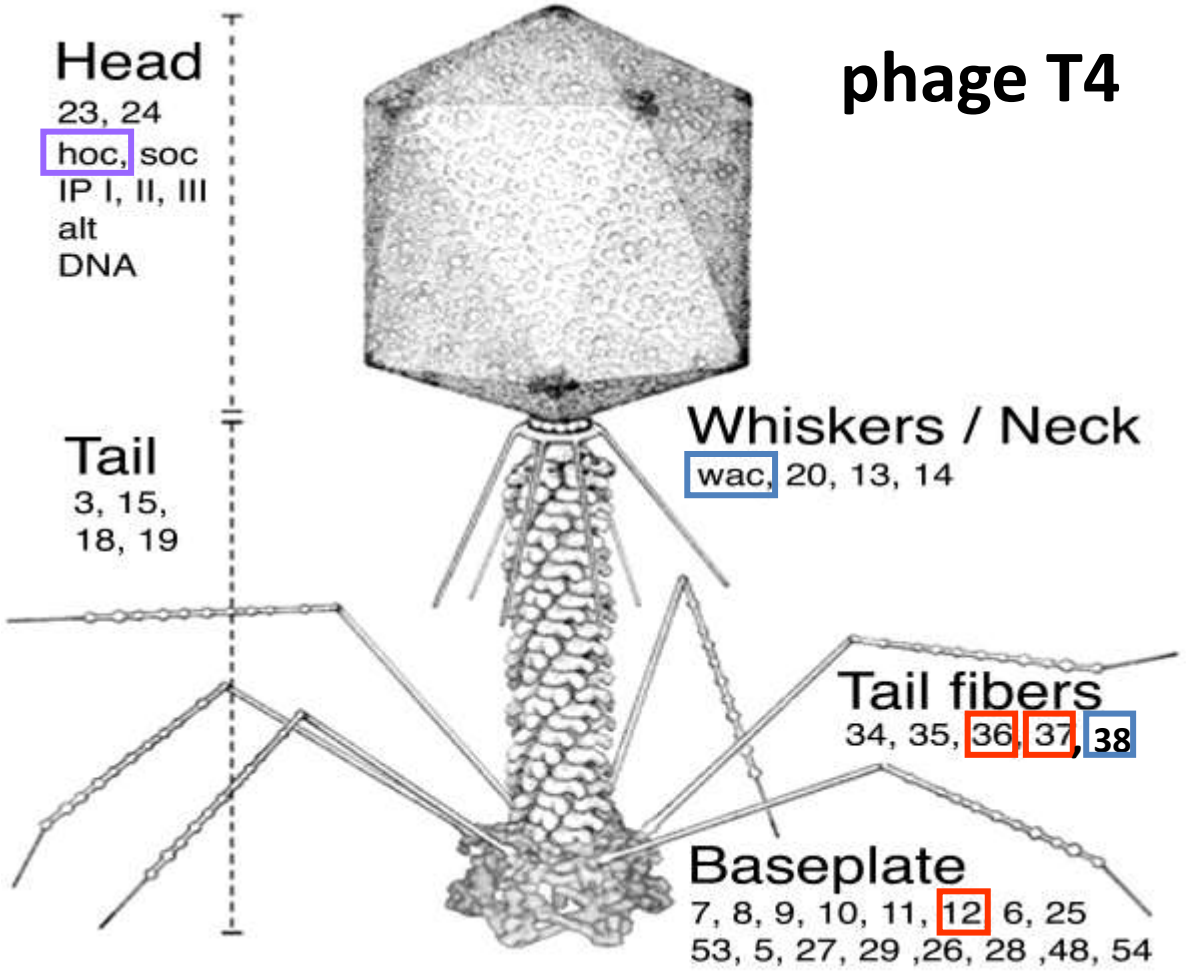
We must be capable of always preceding the targets escape strategies, no matter what they could be.

- ***The best-fit solution*** (model-derived):
  we MUST adopt a stochastic approach.

It becomes necessary to
- abandon all idea of « natural phage pools » and,

- stochastically engineer phage banks in order to produce particles capable of targeting anything and everything while maintaining their capacity to replicate in the face of targets evasion attempts.

The problems:



How to modify any of these proteins in N different regions, at X different sites, in Z different manners, all this simultaneously and then recombine **the multitude of variants** generated in a **population** of **obligate lytic** phages?

The technological answers.

Three proprietary technologies (invented at BMSystems) allowing the production of stochastically engineered phage banks.

TAPE™ Ⓟ (WO 2008/093009):

A technology allowing to rapidly & simultaneously introduce defined densities of random mutations in any number of selected regions within a gene while conserving intact any number of defined coding domains in this same gene.

Applicable to any known coding sequence.

RipH™ Ⓟ (WO 2009/090081):

A technology allowing to reversibly inactivate the genome of an obligate lytic phage within its host and carry out high efficiency homologous recombinations targeting multiple genes simultaneously without adversely affecting the host bacteria and the replicative capacity of the phage.

Ab-ACCUS™ Ⓟ (WO 2008/093010):

A recombination technology allowing the rapid & efficient production of lytic phage banks in which every individual differs from all others for any number of selected genes or other sequences.

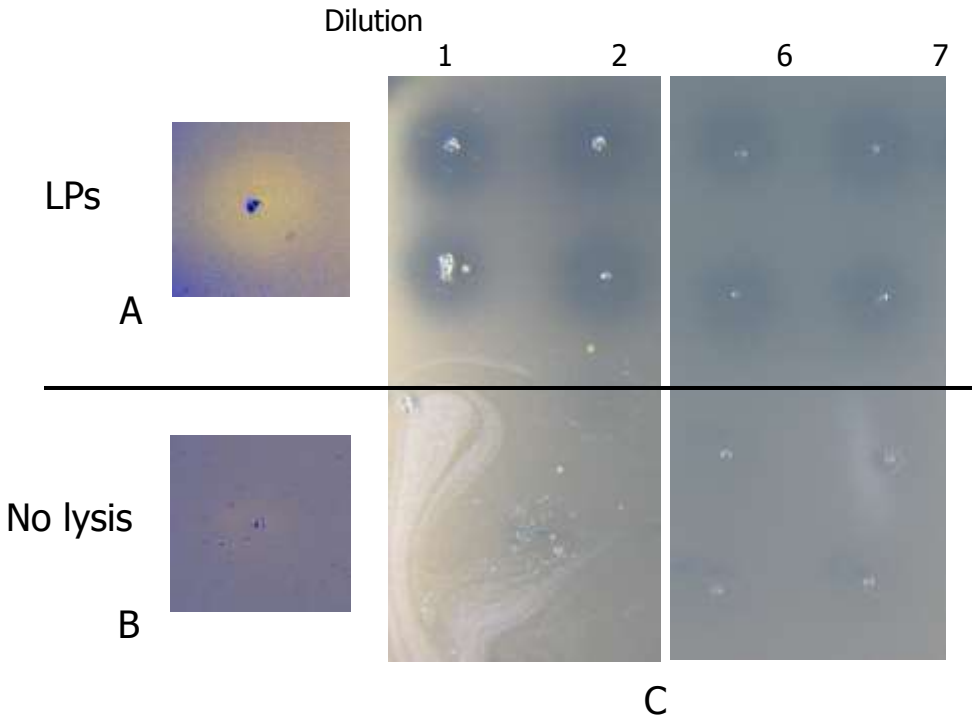
Applicable to any phage and to any known sequence.

The results.

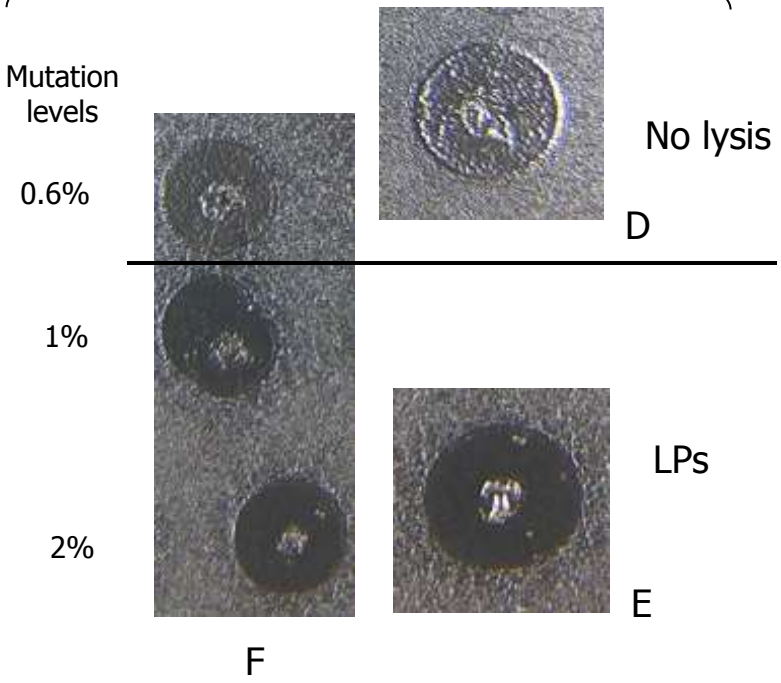
While T4 is specific to a narrow range of E. coli strains,

An engineered T4 bank contains variants capable of detecting and killing gram$^-$ bacteria far removed from E. coli.

Yersinia sp.



Pseudomonas sp.



Pouillot F, Blois H & Iris F (2010); Biosecur Bioterror. **8** (2): 155-169.

Practical consequences
(besides high-profile publications)

- Three technological patents with broad applications,

- Creation and financing of a Bio-Pharma company (Pherecydes Pharma) specialised in biodefense & biosecurity (Spin-off of BM-Systems),

- Research program with French civil and defence health-services,
- Discussions with US food industry firms,
- Contract with anti-bodies producer & discussions with enzymes producers.

All this in less than 3 years.

While the CJD-TheraNexus project took less than
4 years to complete.

From our point of view, under the street light is definitely not the right place to search!

Nevertheless, it MUST be remembered that

Models are AIDS to thought

NOT a replacement for it.

Thank you to

Bio-informatics

The "Engineering" Teams at BM-Systems,
directed by Paul-Henri Lampe, Pablo Santamaria & Manuel Gea

Biologists

The "Integrative Biology" Team at BM-Systems,
directed by Dr Francois Iris.

and, most of all

Academic collaborators

The CEA Prionics group, European Centre of Excellence,
directed by Prof. Jean-Philippe Deslys & Dr. Franck Mouthon.

..and to you for your attention.

Questions