MASTER BIOTECNOLOGÍA BACTERIANA
2014

Aplicaciones Bioinformáticas en Biodegradación

Florencio Pazos

Computational Systems Biology Group
Programa de Biología de Sistemas
Centro Nacional de Biotecnología (CNB-CSIC)

Florencio Pazos Cabaleiro
Computational Systems Biology Group (CNB-CSIC)
pazos@cnb.csic.es
http://csbg.cnb.csic.es

CNB, Madrid
Dic-2014

Aplicaciones Bioinformáticas en Biodegradación

- Introducción – Biodegradación/Bioremediación

- Bases de datos con información sobre biodegradación

- Estudio de la red de biodegradación desde un punto de vista sistémico

- Predicción de biodegradabilidad

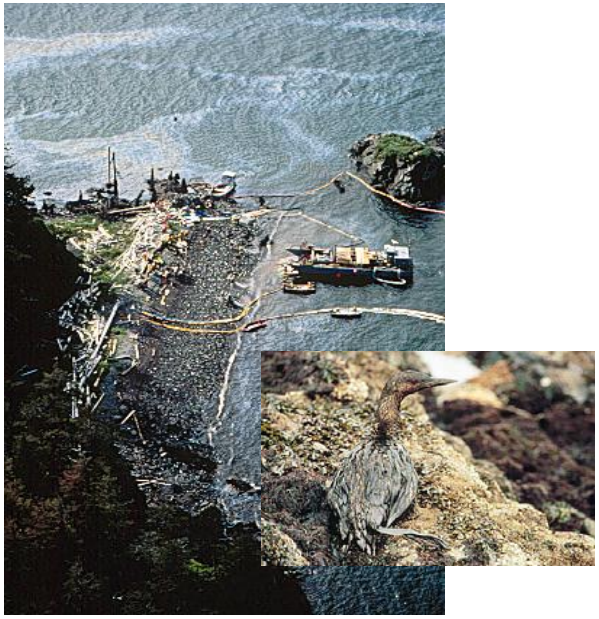
- Introducción – Biodegradación/Bioremediación

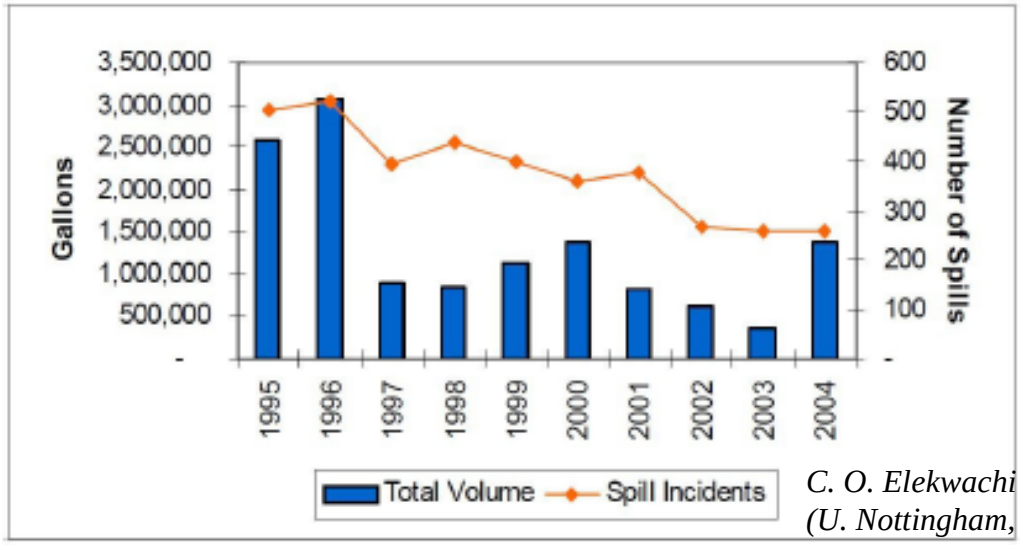
- Bases de datos con información sobre biodegradación

- Estudio de la red de biodegradación desde un punto de vista sistémico

- Predicción de biodegradabilidad

Introducción





C. O. Elekwachi PhD thesis (U. Nottingham, UK)

Figure 2: Graph showing volume and number of spills above 100 gallons in United States Coastal waters between 1995 and 2004. Prepared by US Congressional Research Service using data from the United States Coast Guards (USCG) oil spill compendium. Source: Ramseur (2010).

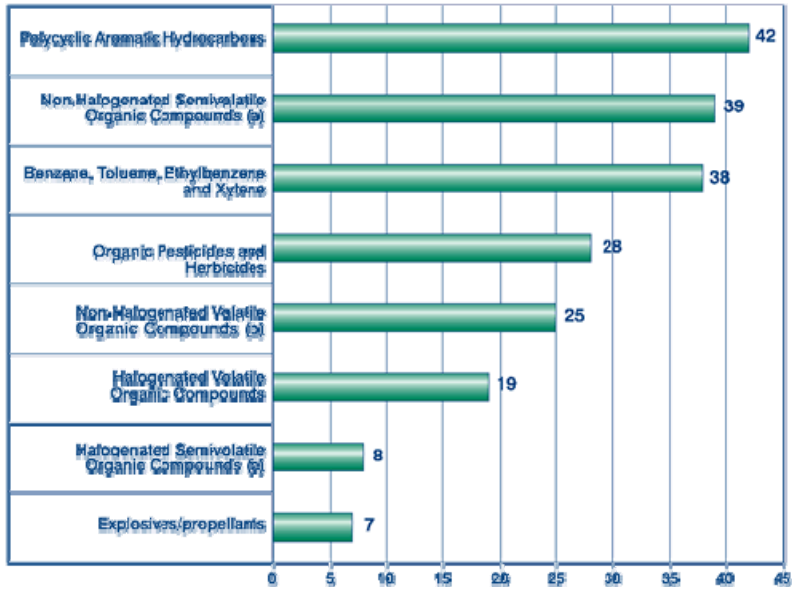


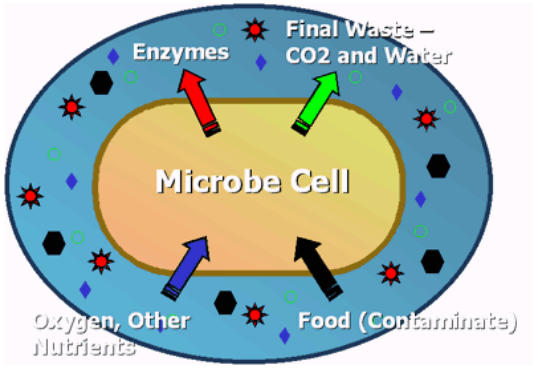
Table 3: Summary of Conventional Groundwater and Soil Remediation Methods*

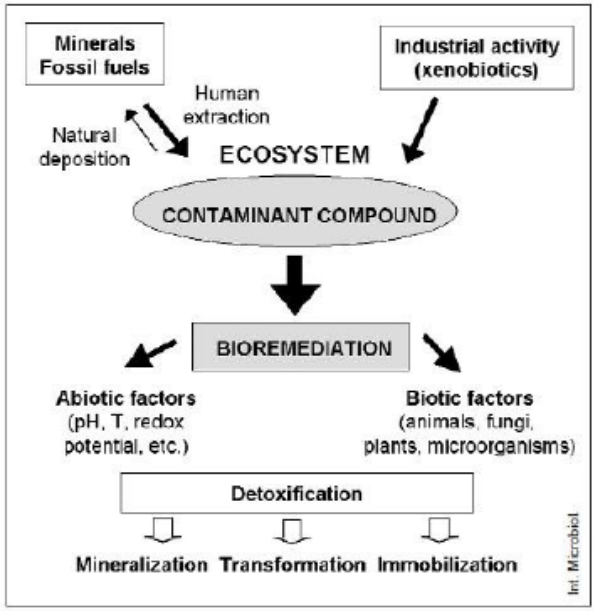
Methods	Type	Advantages	Disadvantages
	Conventional		
Excavation, transfer, and disposal to landfill	*Ex situ* soil	Effective at removal of shallow contaminants	Expensive, disruptive, relocates problem to another site (landfill)
Excavation and incineration, soil washing	*Ex situ* soil	Effective at removal of shallow contaminants	Expensive, disruptive
Bioremediation/ natural attenuation	*In situ* soil and groundwater	Works well when conditions are right	Costly in monitoring and time consuming
Soil vapour extraction	*In situ* soil, groundwater with extraction pumping	Works well when conditions are right	Costly to operate
Pump and treat	*Ex situ* groundwater treatment	Works well for free product removal and providing hydraulic control	Generally ineffective at remediation, costly in investment and very time consuming
	Less Conventional		
In situ jetting chemical delivery system	*In situ* injection using a lance to inject liquids for chemical oxidation, bioremediation, pH adjustment, and metals stabilization	Can be less costly than other methods and less time consuming, with minimal disruption; can be combined with ozone injection for oxidation projects	Site-specific design required related to site soil/water chemistry; bench tests and pilot studies recommended to verify site specific conditions
Passive systems; funnel and gate		Low cost for maintenance; truly passive groundwater treatment system	Very expensive for installation; works only with stable groundwater flow directions; treatment must be designed for site.
Trench collectors with ozone treatment wall	The water passively moves into the interceptor trench located perpendicular to the flow direction	Ozone works well with hydrocarbons and solvents and over time can oxidise volatile contaminants at oil and gas fields effectively	Works only with stable groundwater flow directions; ozone must be generated on site; high electrical usage of ozone generator; limited amount of ozone produced daily

C. O. Elekwachi PhD thesis (U. Nottingham, UK)

Biodegradación/Bioremediación

Biodegradation: Ability adquired by some microorganisms to process recalcitrant, often xenobiotic compounds that do not form part of their central metabolism (CM) by transforming them into compounds that can enter into CM. Biodegradation processes have enormous potential for environmental cleanup (bioremediation) and in biocatalysis (green chemistry).





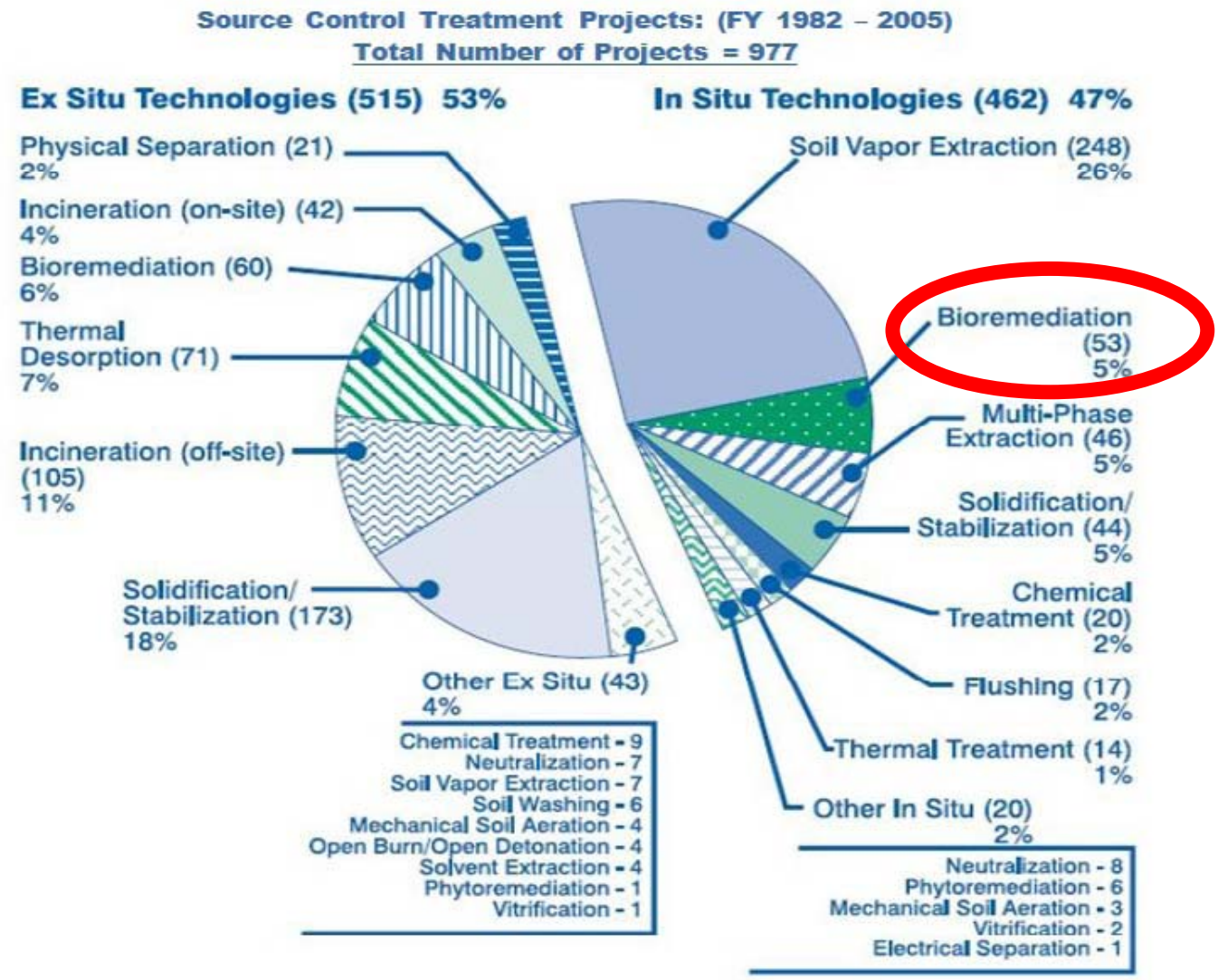
Parales, R.E., Bruce, N.C., Schmid, A. & Wackett, L.P. (2002). Biodegradation, biotransformation, and biocatalysis (b3). *Appl. Environ. Microbiol.*, **68,** 4699–4709.

Dua, M., Singh, A., Sethunathan, N. & Johri, A.K. (2002) Biotechnology and bioremediation: successes and limitations. *Appl. Microbiol. Biotechnol.*, **59**, 143–152.

Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M. & Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature*, **409**, 258–268.

Diaz, E. (2004). Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *International Microbiology* **7(3):** 173-180.

Bioremediación



C. O. Elekwachi PhD thesis (U. Nottingham, UK)

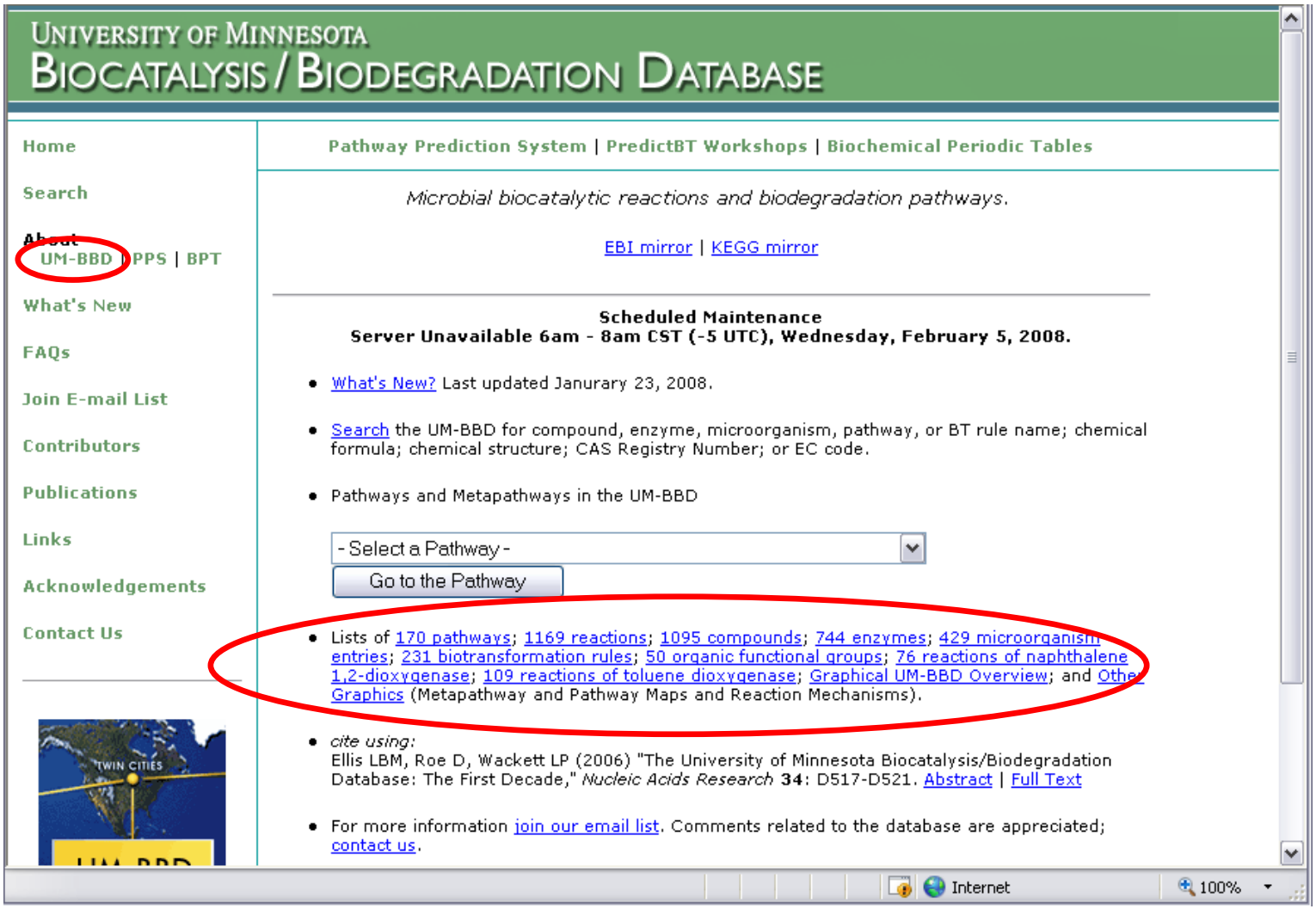
- Introducción – Biodegradación/Bioremediación

- **Bases de datos con información sobre biodegradación**

- Estudio de la red de biodegradación desde un punto de vista sistémico

- Predicción de biodegradabilidad

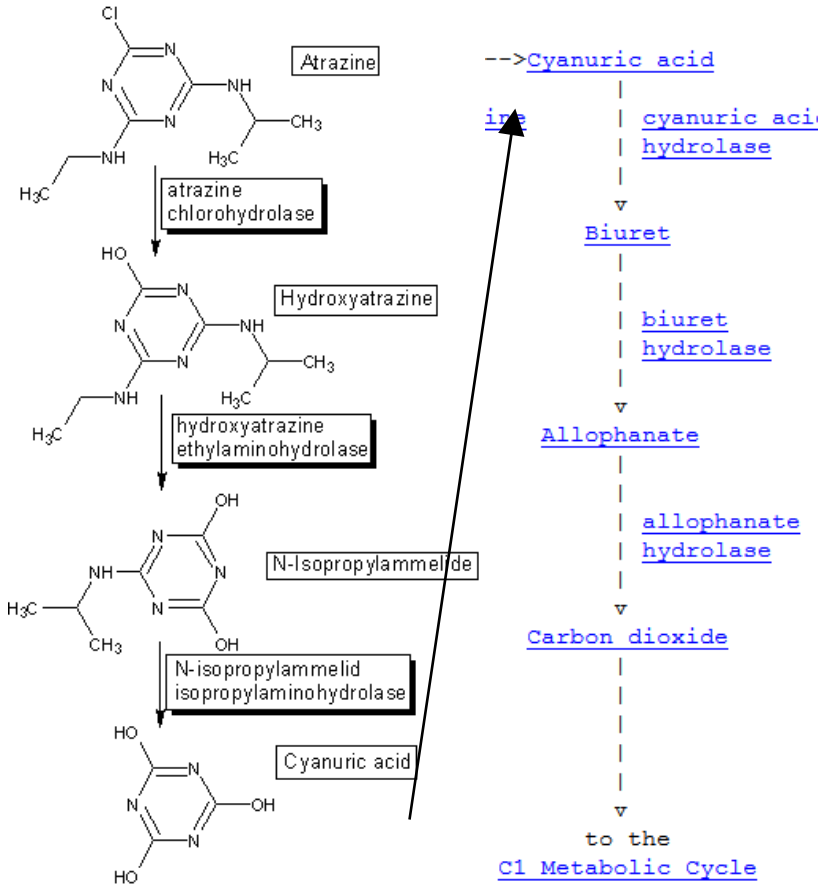
UM-BBD (Now EAWAG-BBD)

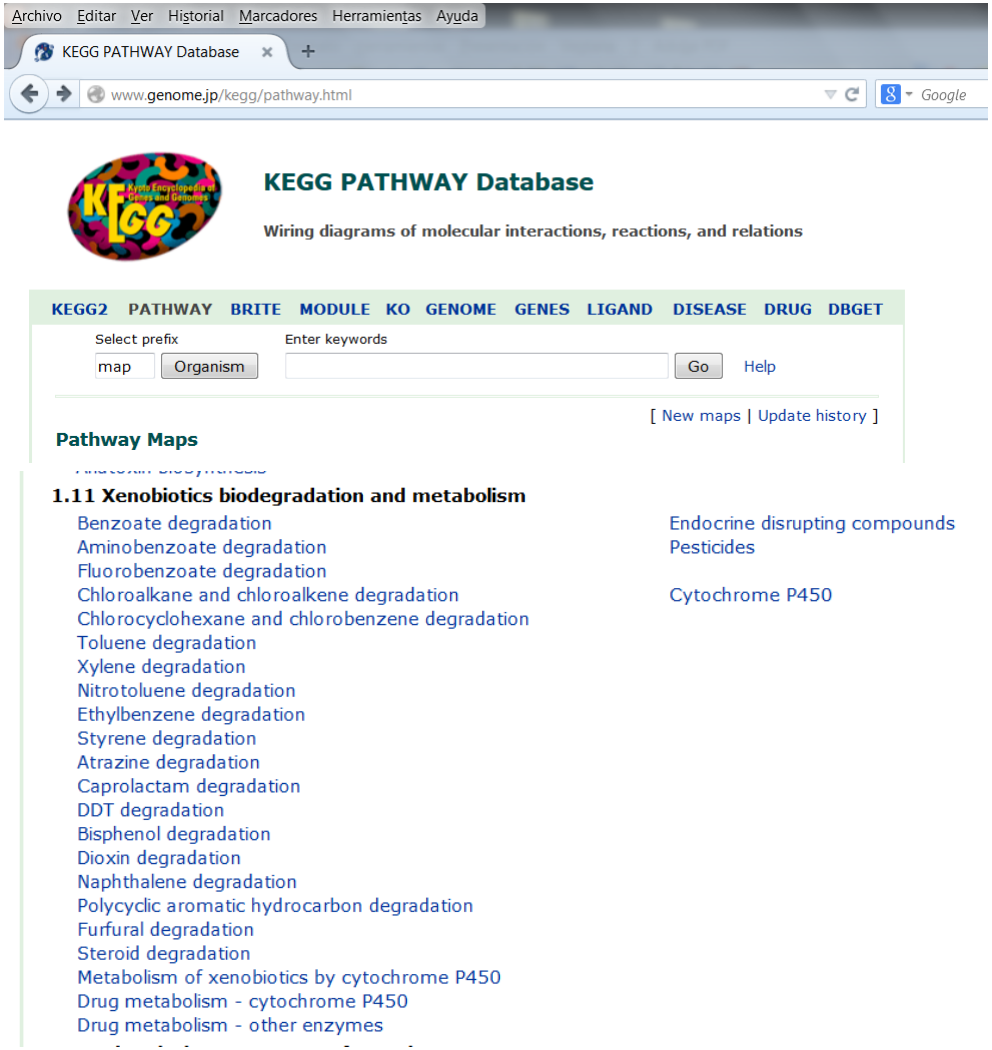


- **http://eawag-bbd.ethz.ch/**

- Ellis, L.B., Hou, B.K., Kang, W. and Wackett, L.P. (2003) The University of Minnesota Biocatalysis/Biodegradation Database: post-genomic data mining. *Nucleic Acids Res*, **31,** 262-265.

UM-BBD (Now EAWAG-BBD)





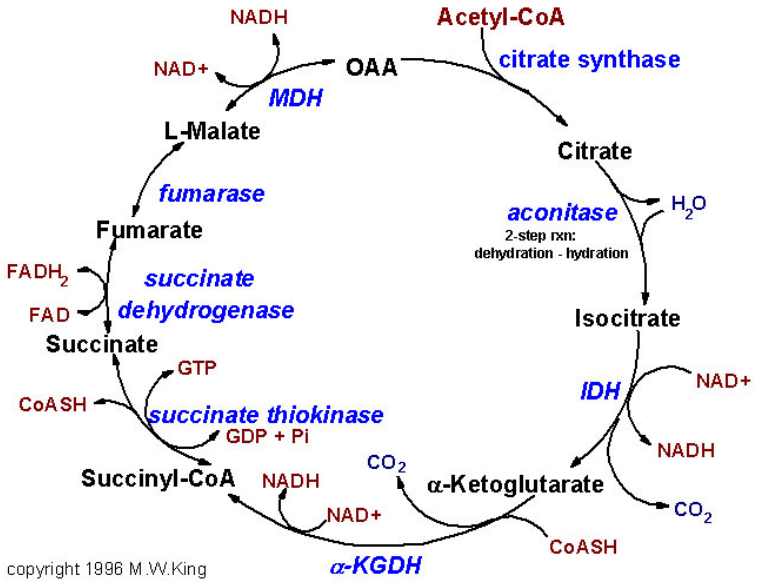
- Introducción – Biodegradación/Bioremediación

- Bases de datos con información sobre biodegradación

- **Estudio de la red de biodegradación desde un punto de vista sistémico**

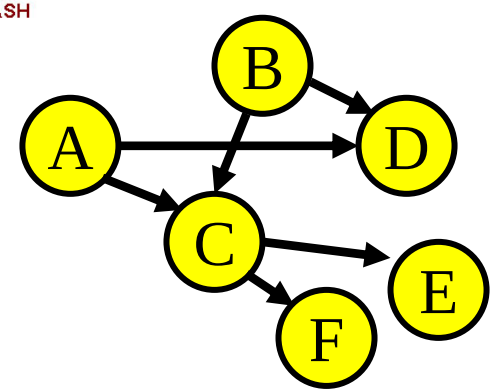
- Predicción de biodegradabilidad

Estudio sistémico de redes metabólicas (metabolismo central)
Representación del metabolismo como redes (grafos)



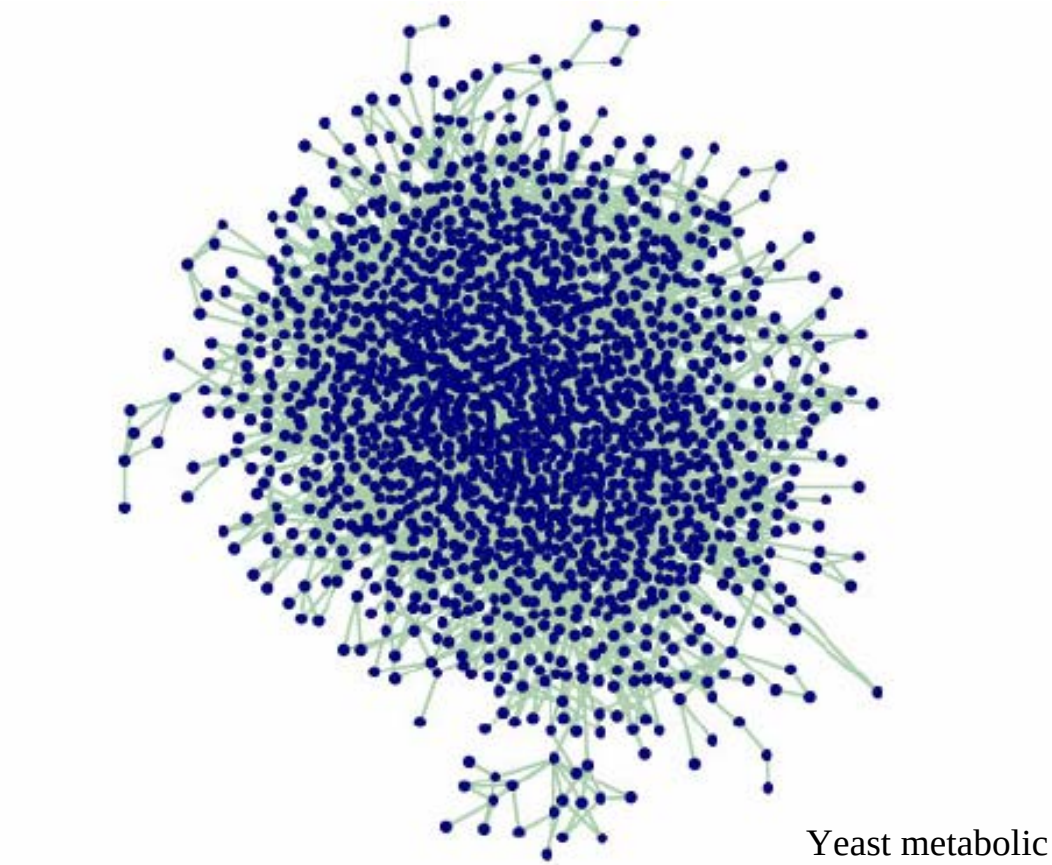
A + B → C + D

C → E + F



∀ A, B, C, ... ≠ H20, ions, cofactors, or Nr>X

Estudio sistémico de redes metabólicas (metabolismo central)



Yeast metabolic network

•Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, **5**, 101-113.

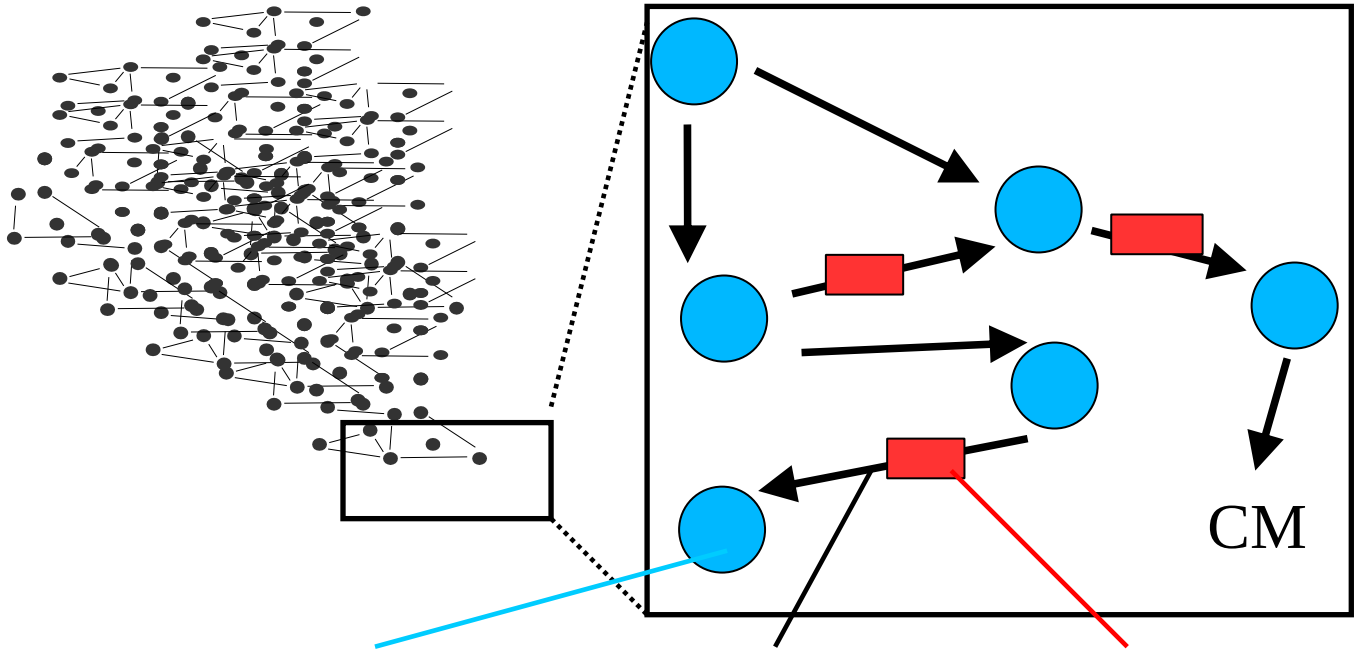
•Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000). The large scale organisation of metabolic networks. *Nature* **407,** 651-653

•Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, **5,** 101-113

•Ravasz, E., Somera, L., Mongru, D.A., Oltvai, Z.N. and Barabási, A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297,** 1551-1555.

Ensamblado de la red global de biodegradación

UMBBD
Enzyme
Chemfinder
SwissProt
….



Name
Formula
SMILES string
Properties: sol., mw., …
3D structure (PDB).
…

Enzymes/substrates/products
Links to databases
…

EC code
Reaction
Organisms
Sequences
Links to databases
…

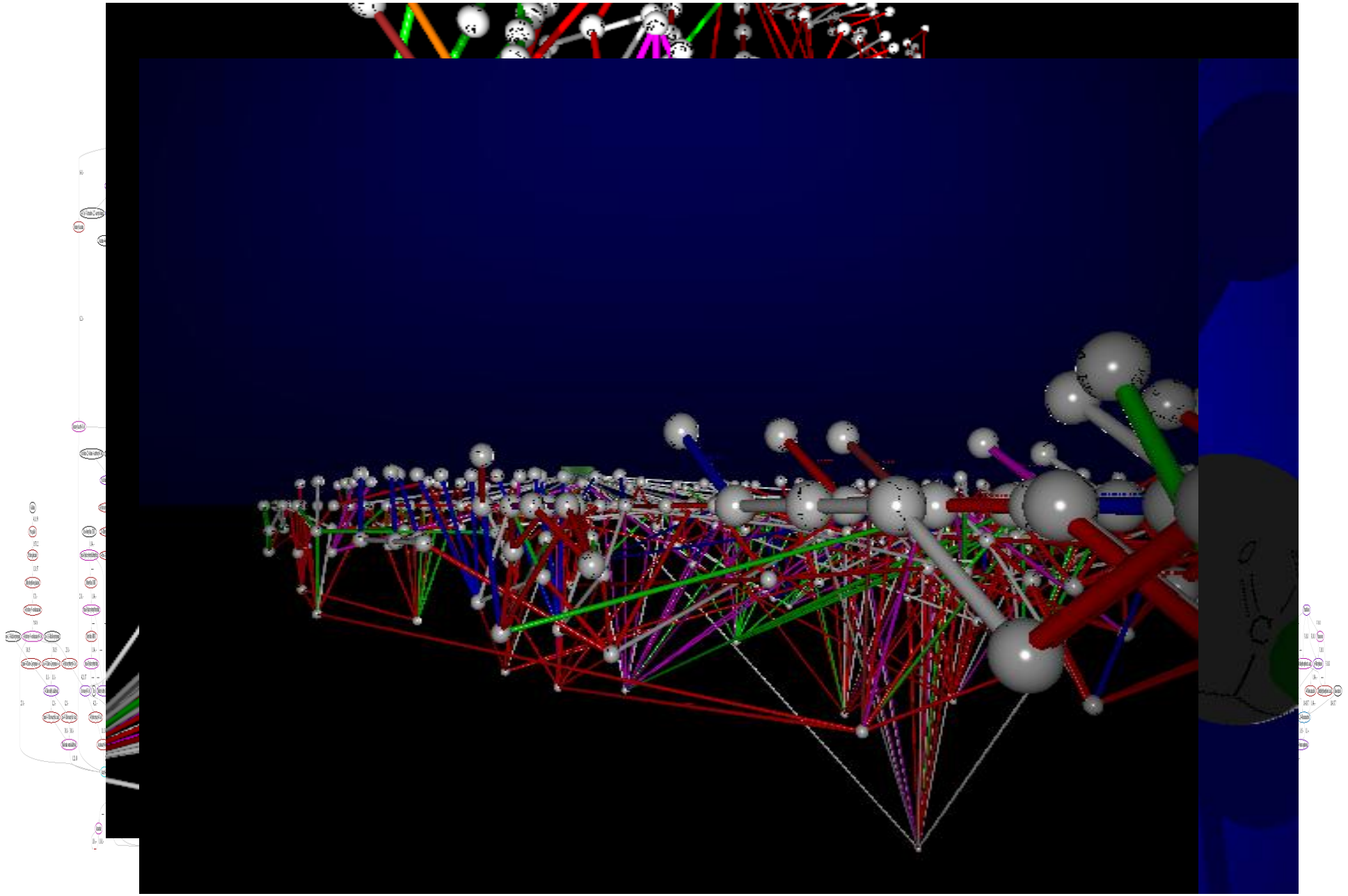
740 compounds
821 reactions
678 with associated enzymatic activity
308 specific enzymatic activity (4 EC numbers)

• Florencio Pazos, David Guijas, Alfonso Valencia & Victor De Lorenzo. (2005). MetaRouter: bioinformatics for bioremediation. *Nucleic Acids Research*. **33:**D588-D592.

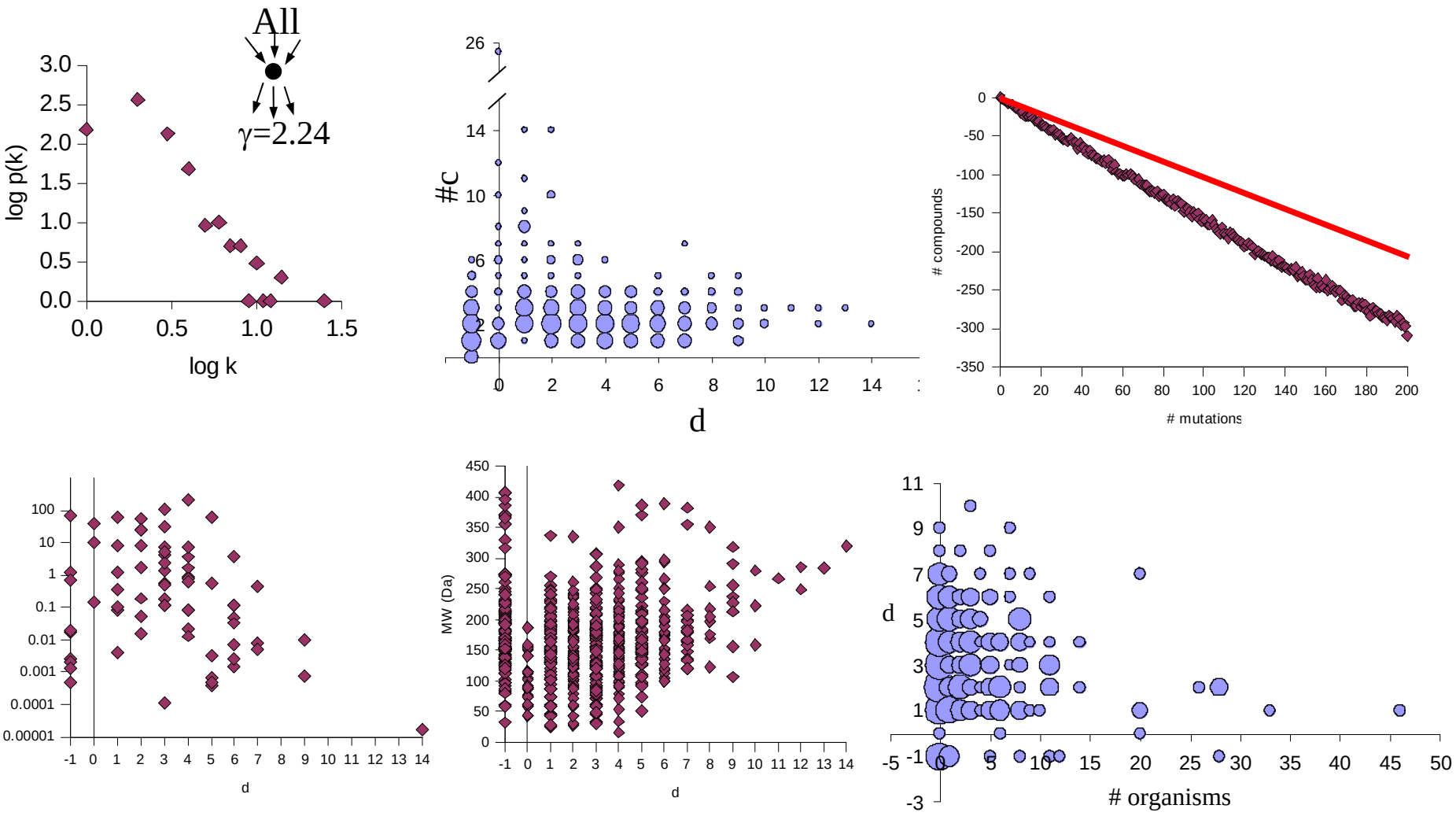
Red global de biodegradación

• Florencio Pazos, Victor De Lorenzo & Alfonso Valencia. (2003). The organization of the Microbial Biodegradation Network from a Systems-Biology perspective. *EMBO Rep.* **4(10):**994-999.

http://pdg.cnb.csic.es/biodeg_net

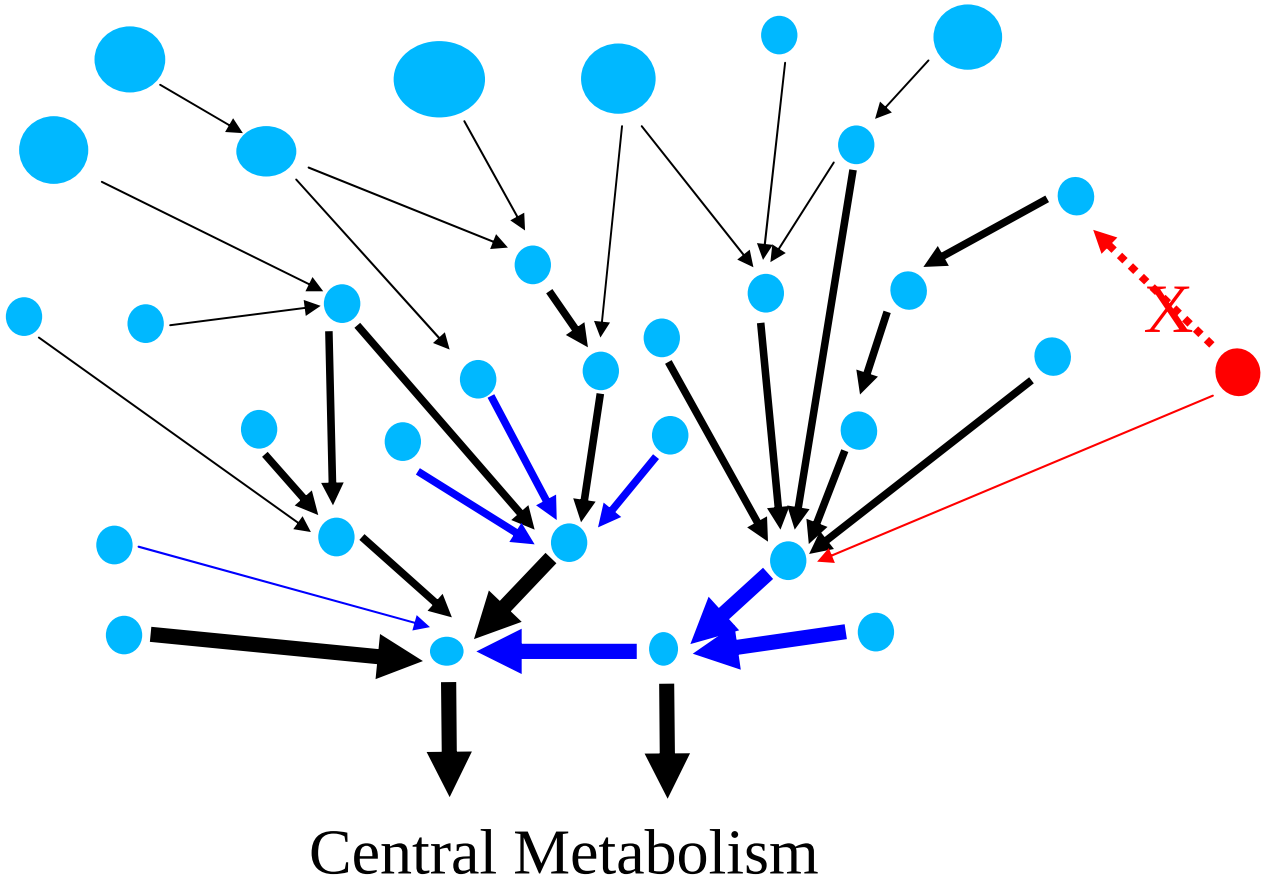


Propiedades de la red global de biodegradación



• Florencio Pazos, Victor De Lorenzo & Alfonso Valencia. (2003). The organization of the Microbial Biodegradation Network from a Systems-Biology perspective. *EMBO Rep.* **4(10):**994-999.

Propiedades de la red global de biodegradación



• Florencio Pazos, Victor De Lorenzo & Alfonso Valencia. (2003). The organization of the Microbial Biodegradation Network from a Systems-Biology perspective. *EMBO Rep.* **4(10):**994-999.

Aplicaciones Bioinformáticas en Biodegradación

- Introducción – Biodegradación/Bioremediación

- Bases de datos con información sobre biodegradación

- Estudio de la red de biodegradación desde un punto de vista sistémico

- **Predicción de biodegradabilidad**

Predicting “biodegradability”

Knowing whether a novel chemical compound is likely to be metabolised by microorganisms is crucial for assessing the environmental risks associated to its production, transportation, utilization and disposal

But... after 50 years of research on microbial biodegradation, detailed knowledge about biodegradative pathways is available for only a few hundreds species

New pesticides and pharmaceuticals are being produced at rates that cannot be matched by experimental attempts to determine the outcome when spilled or released into the environment. This makes essential to develop systems that can predict the fate of chemical compounds before releasing them into the environment.

Predicción de Biodegradabilidad

- Sistemas basados en conocimiento experto (i.e. *UM-PPS*)

- Sistemas basados en aprendizaje artificial (i.e. *BDPServer*)

Predicción de Biodegradabilidad - *UMBBD-PPS*

As of February 4, 2008 6:17:17 AM CST, the UM-BBD contains 291 biotransformation descriptions for 234 biotransformation rules.

Acetophenone derivative -----> Benzoyl acetate derivative (bt0202)

Acid chloride -----> Carboxylate (bt0026)

Acylsulfonate -----> Carboxylate + HSO3- (bt0227)

primary Alcohol -----> Aldehyde (bt0001)

secondary Alcohol -----> Ester (bt0002)

secondary Alcohol -----> Ketone (bt0002)

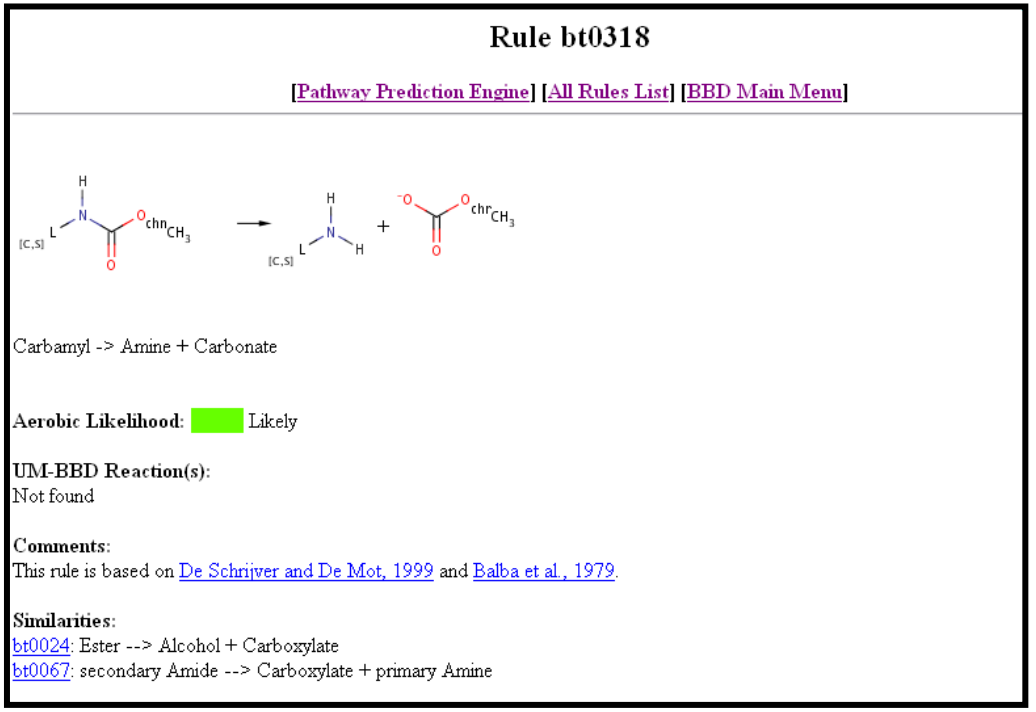
primary Alcohol with halogen on the same carbon -----> Aldehyde (bt0314)

Aldehyde -----> Carboxylate (bt0003)

.....

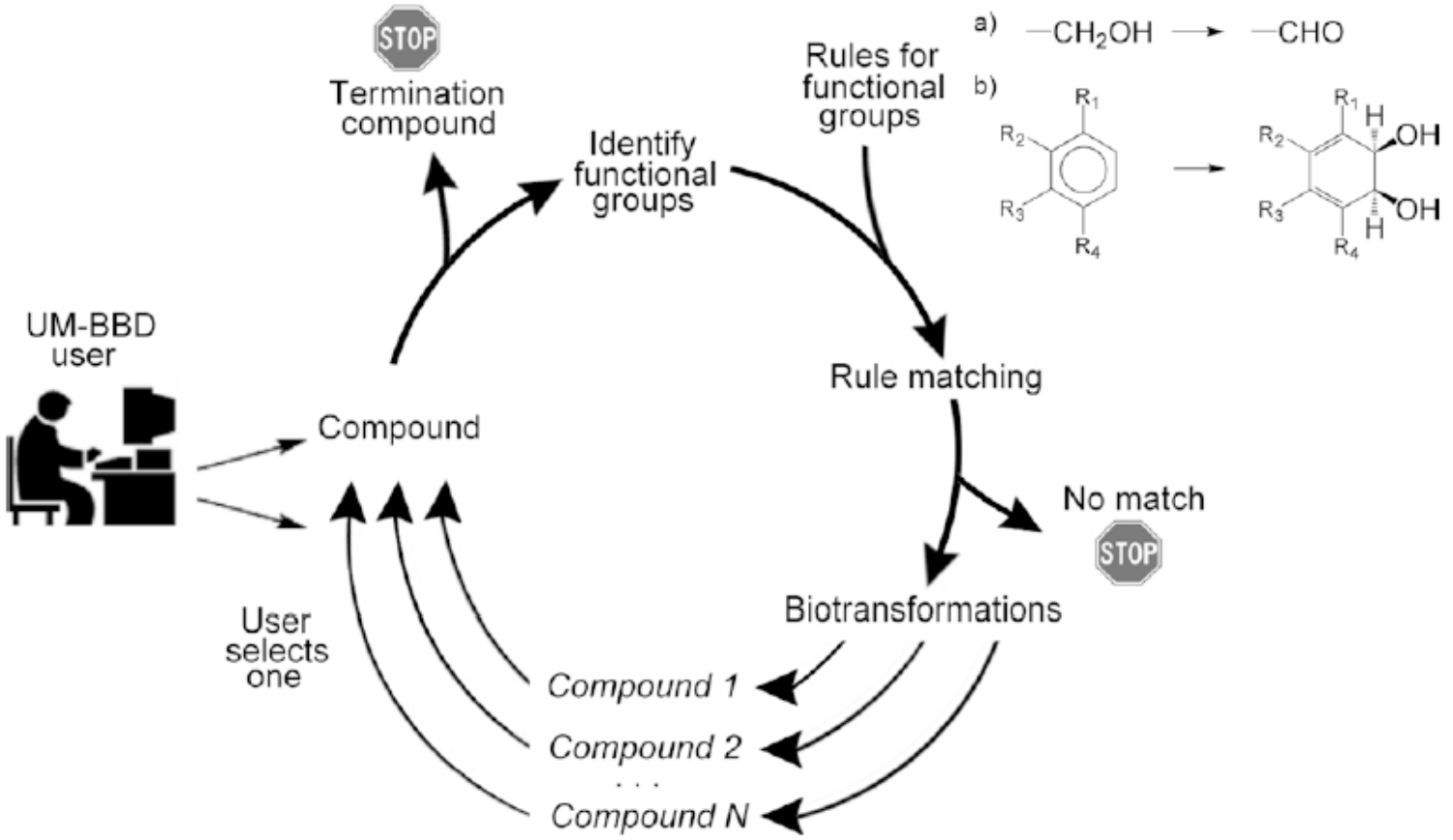
..

.



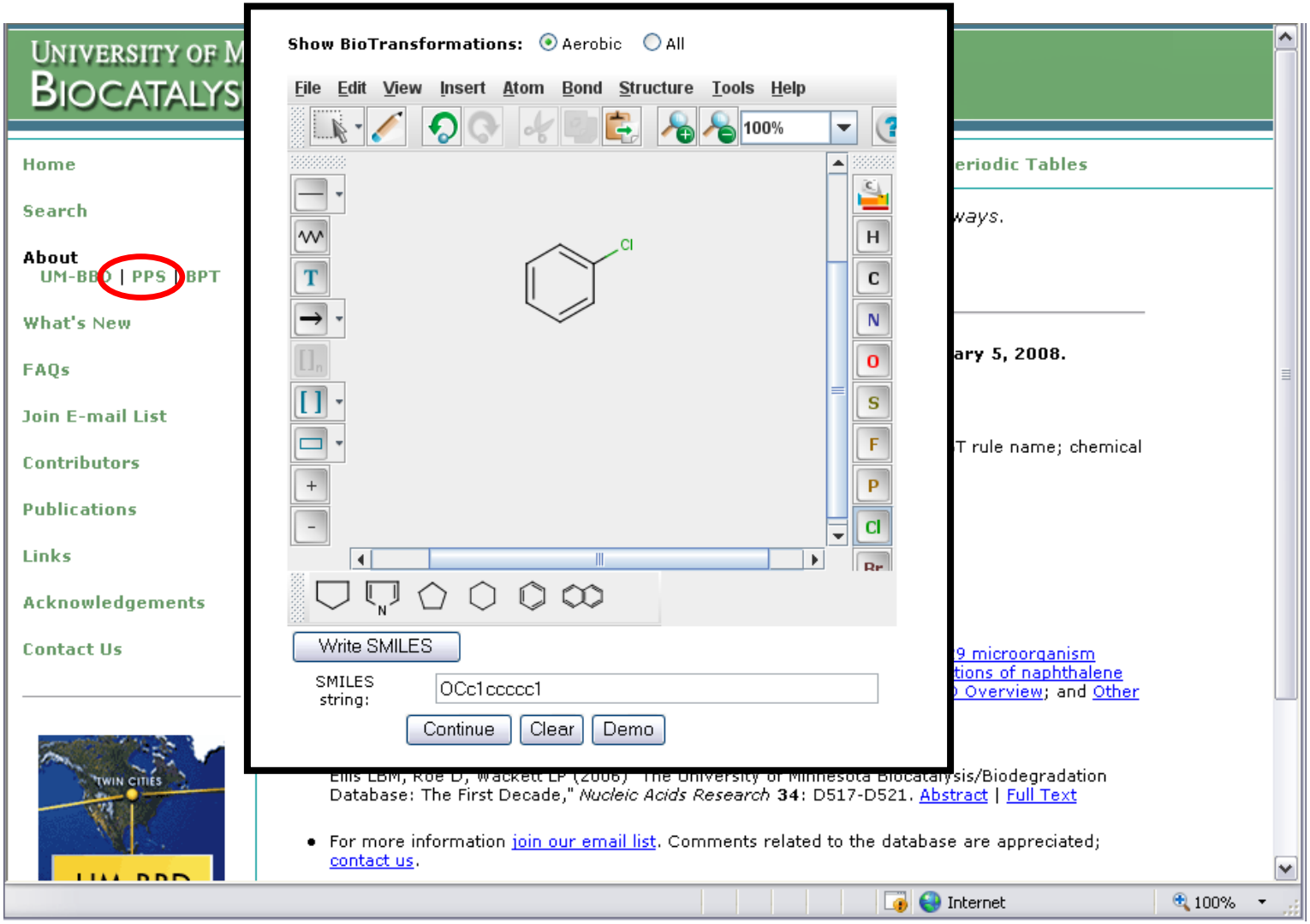
Hou, B.K., Ellis, L.B. and Wackett, L.P. (2004) Encoding microbial metabolic logic: predicting biodegradation. *J Ind Microbiol Biotechnol.*, **31**, 261-272.

Predicción de Biodegradabilidad - *UMBBD-PPS*



Hou, B.K., Ellis, L.B. and Wackett, L.P. (2004) Encoding microbial metabolic logic: predicting biodegradation. *J Ind Microbiol Biotechnol.*, **31**, 261-272.

Predicción de Biodegradabilidad - *UMBBD-PPS*



- **http://umbbd.ethz.ch/**

• Ellis, L.B., Hou, B.K., Kang, W. and Wackett, L.P. (2003) The University of Minnesota Biocatalysis/Biodegradation Database: post-genomic data mining. *Nucleic Acids Res*, **31,** 262-265.

Predicting biodegradability – *Machine learning*

Another approach: *machine learning*

Learn from examples of known recalcitrant/biodegradable compounds how to distinguish the two classes

Disadvantage: no information on the biodegradative pathway. Just final fate.

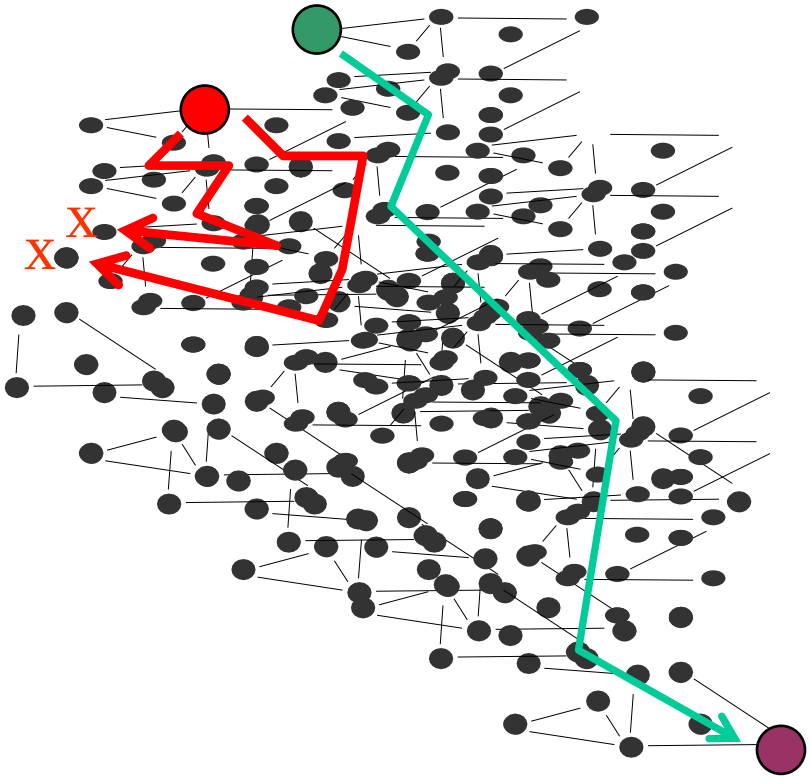
Advantage: fully automatic (non interactive) => can be applied to large collections of compounds

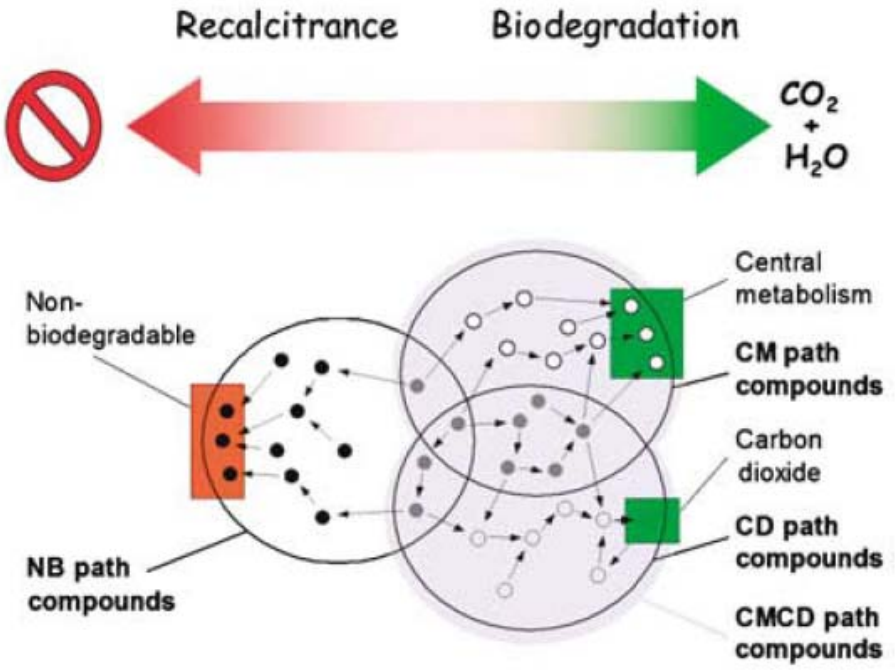
Requirements

- Examples of recalcitrant/biodegradable compounds (many)
- Description of the chemical structure which can be managed by the machine learning system
- ML method

Obtaining examples of biodegradable/recalcitrant compounds

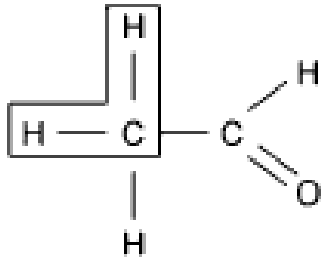
Assumption: biodegradable compounds are those for which it is possible to find a pathway to CM.
Otherwise they are recalcitrant.





Vectorial representation of molecules

Chemical structure



SMILES

CC=O

Atomic triads

C–C–H	4
C–C=O	1
H–C–H	3
H–C=O	1

MW

44 Da

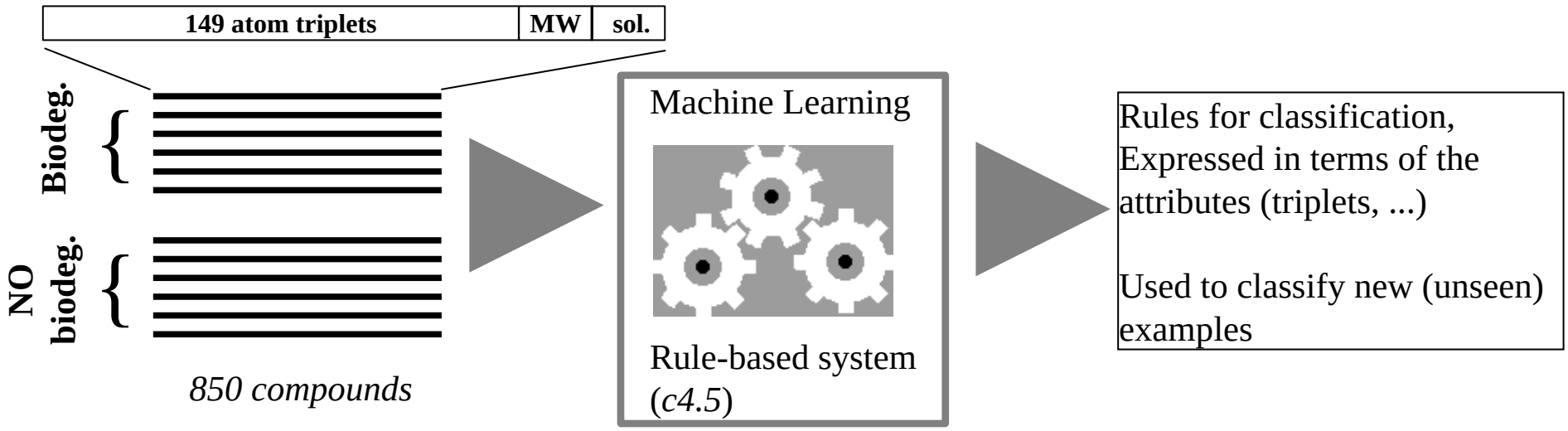
Solubility

0.1 g/100 ml

Compound vector = (0.1, 44 ,4 ,1 ,3 ,1 ,0 ,0 ,...)

Gómez, MJ, Pazos F, Guijarro FJ, de Lorenzo V, Valencia A. (2007). The environmental fate of organic pollutants through the global microbial metabolism. *Mol Syst Biol*. **3:**114.

Machine learning system



Advantages of rule-based systems compared with other ML, i.e. NN:

- Explicit rules in human-readable format
- Can handle missing values (solubility)

Quinlan, JR. (1993). *c4.5: Programs for Machine Learning.* San Mateo, CA. Morgan-Kaufmann

Results

Table I Example of propositional rule generated for the classification compounds in the scheme NB *or* No NB

Rule 55: IF	–C–C–C >19
	–O–C–C >1
	–O–C–C ⩽3

THEN, the compound belongs to the NB class (Confidence 90.6%)
Examples (14 cases)

52 out of the 152 attributes involved in rules

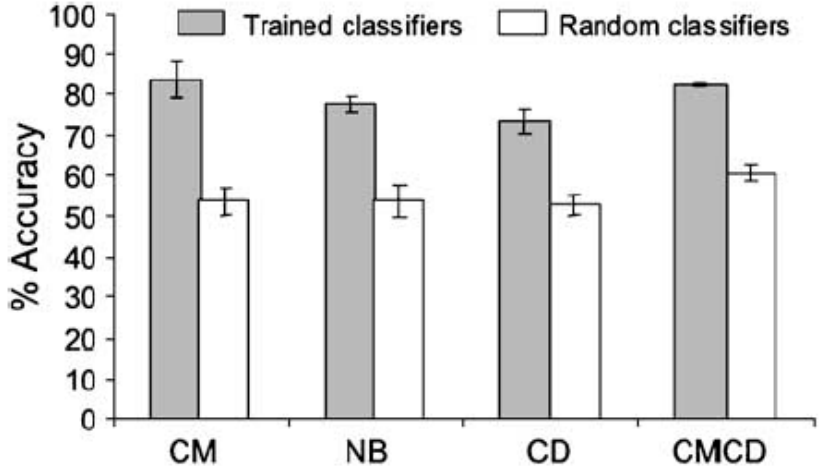


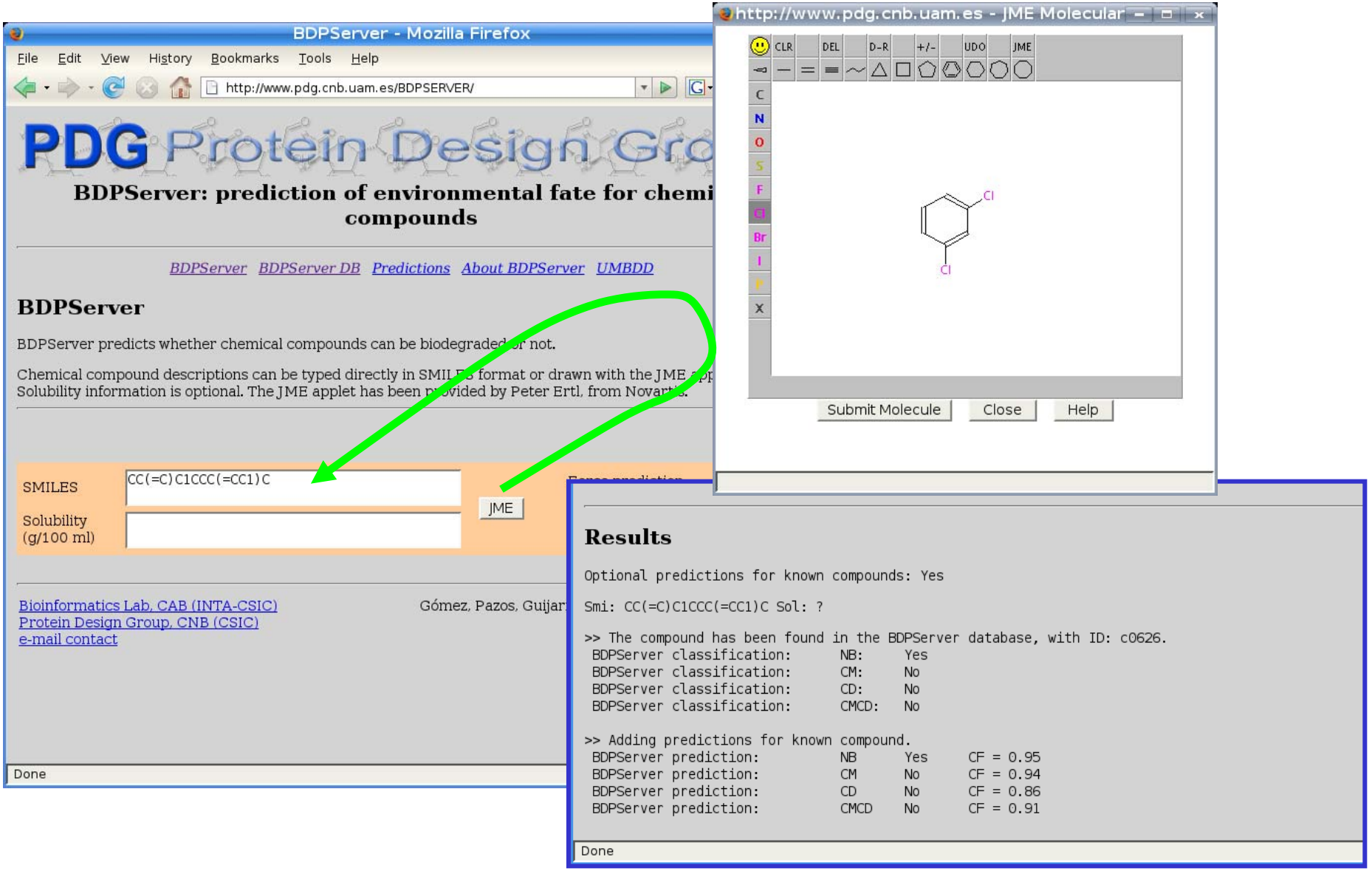
Table II Predictive performance in fivefold cross-validation experiments

Classification scheme[a]	CM or No CM	NB or No NB	CD or No CD	CMCD or No CMCD
Accuracy (%)	87±4	77±4	73±5	82±3
Significance respect to random—P(N)	1.1×10^{-39}	1.2×10^{-24}	2.4×10^{-16}	6.4×10^{-26}
Default class	CM	No NB	No CD	CMCD
Majority class	CM	No NB	No CD	CMCD
No. of cases	533	496	513	634
Sensitivity (%)	93±4	86±4	78±7	92±3
Specificity (%)	83±3	78±2	78±3	85±2
Minority class				
No. of cases	308	345	328	207
Sensitivity (%)	67±5	64±5	66±9	50±11
Specificity (%)	85±8	77±4	66±5	71±5

***E coli* CM (733 comps.): 90.31% CMCD; 81.44% *No*-NB**

Gómez, MJ, Pazos F, Guijarro FJ, de Lorenzo V, Valencia A. (2007). The environmental fate of organic pollutants through the global microbial metabolism. *Mol Syst Biol*. **3:**114.

Using the predictor - *BDPServer*



- Gómez, MJ, Pazos F, Guijarro FJ, de Lorenzo V, Valencia A. (2007). The environmental fate of organic pollutants through the global microbial metabolism. *Mol Syst Biol*. **3:**114.

- **http://pdg.cnb.csic.es/DBPSERVER/**

Updating the system…

[Computational Systems Biology Group @CNB-CSIC]

SVM Biodegradability predictor

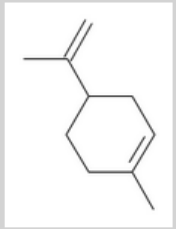
Welcome to SVM Biodegradability predictor. This is a tool for predicting biodegradability and/or toxicity, based on the different definitions provided for some of the most popular compound databases. It is easy to use, just introduce a compound in SMILES format or design it with the JME applet, select the databases that you want to obtain a prediction for and press GO button.

Input compound in SMILES format or use JME editor: CC(=C)C1CCC(=CC1)C

GO! Use sample

Available databases: ☑ UM-BBD ☑ PPDB ☑ NITE ☑ PPDB TOXICITY

CC(=C)C1CCC(=CC1)C



UM-BBD: No biodegradable (Reliability:87.50%, Score:-1.000)

PPDB: Non persistant (Reliability:84.55%, Score:0.962)

NITE: Ready biodegradable (Reliability:76.92%, Score:-0.018)

PPDB TOXICITY: Low toxicity (Reliability:81.44%, Score:-0.894)

Contact

For general questions/queries about the software, please contact Florencio Pazos (pazos@cnb.csic.es), for technical questions contact Juan A. Garcia Martin (juan.garcaamartan@bc.edu).

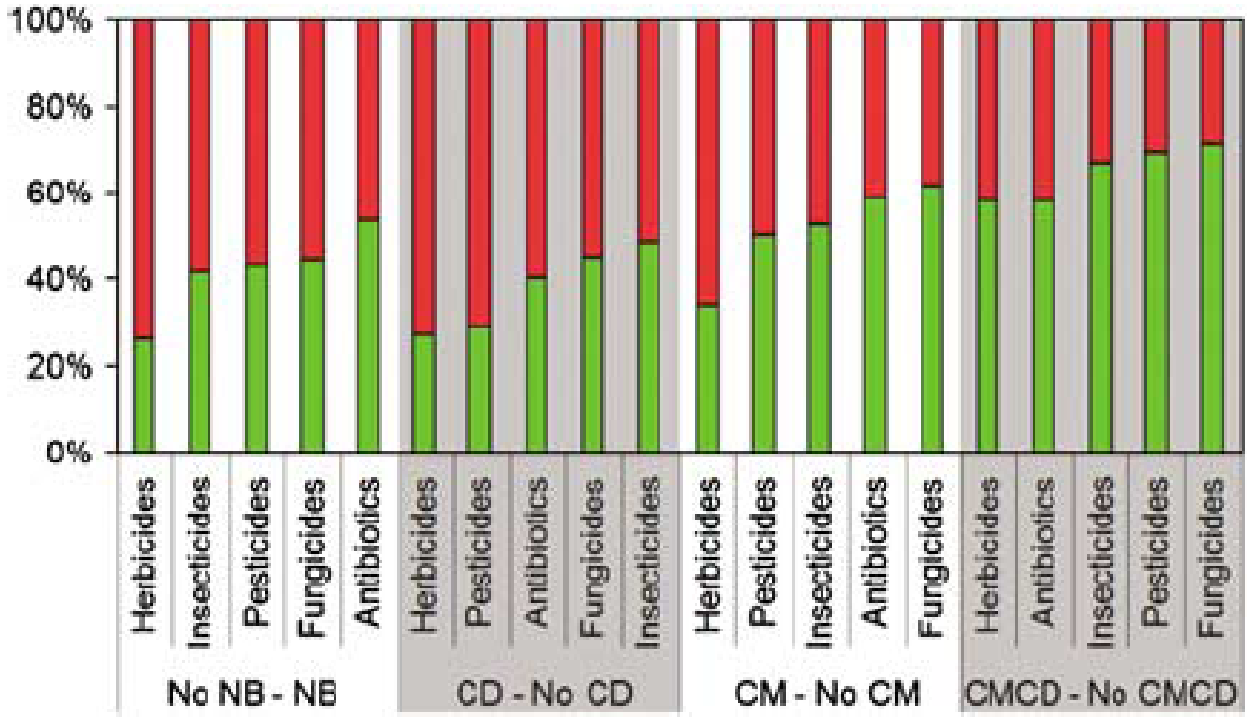
[Computational Systems Biology Group @CNB-CSIC]

Juan A. Garcia-Martin, M. Chavarria, V. de Lorenzo, F. Pazos. *In preparation*

Blind predictions

- Chemical compounds regulated by the ECB (HPVCs, LPVCs & *Annex-I*): 70% CMCD

- PubChem (USA). 3600 comps. aprox.



Gómez, MJ, Pazos F, Guijarro FJ, de Lorenzo V, Valencia A. (2007). The environmental fate of organic pollutants through the global microbial metabolism. *Mol Syst Biol*. **3:**114.

Agradecimientos

- Alfonso Valencia (CNIO)
- Victor De Lorenzo (CNB)
- Javier Guijarro (CNB)
- Manuel J. Gomez-Rodriguez (CAB-INTA - CNIC)
- David Guijas (Alma Bioinformatics)
- L. Wacket, L. Ellis (UMBBD)
- Juan A. García Martín (Boston University)

Florencio Pazos Cabaleiro
Computational Systems Biology Group (CNB-CSIC)
pazos@cnb.csic.es
http://csbg.cnb.csic.es