



---

# 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
-

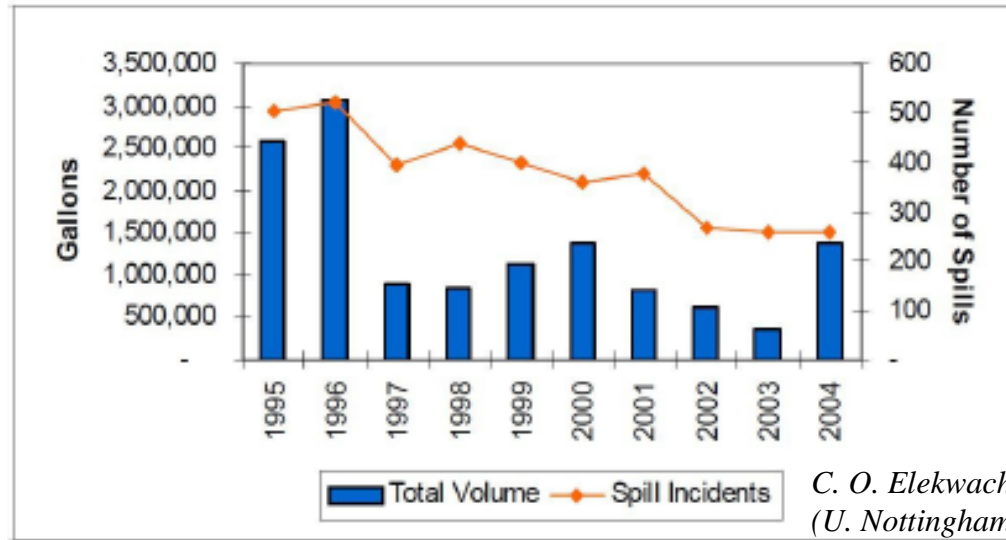
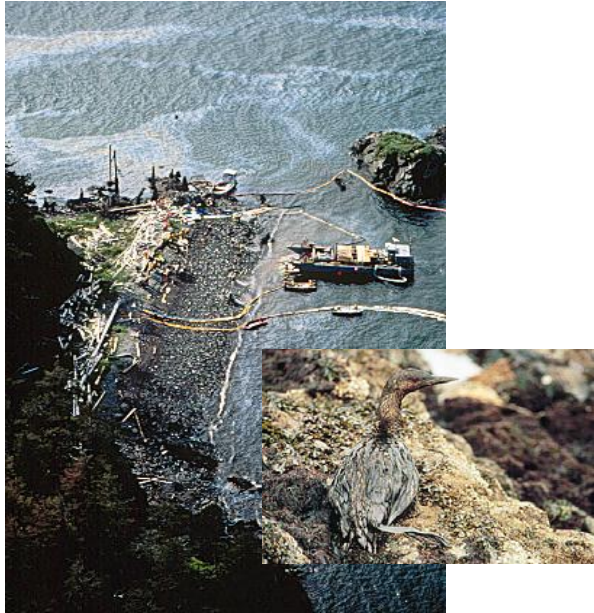
# Aplicaciones Bioinformáticas en Biodegradación

---

## - **Introducción – Biodegradación/Bioremediación**

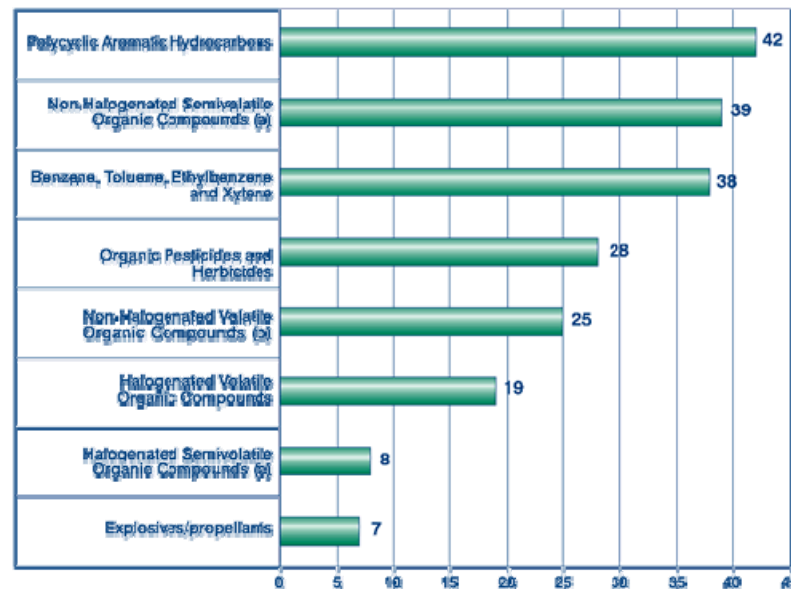
- Bases de datos con información sobre biodegradación
  - Estudio de la red de biodegradación desde un punto de vista sistémico
  - Predicción de biodegradabilidad
-

# Introducción



*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).

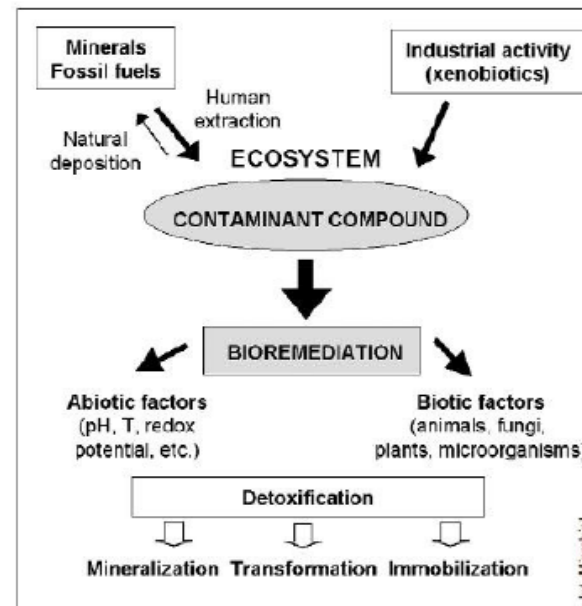
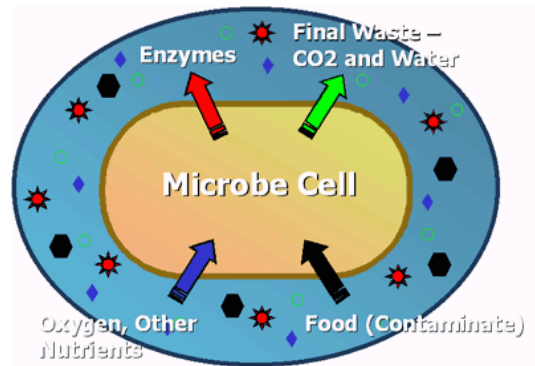


**Table 3: Summary of Conventional Groundwater and Soil Remediation Methods\***

Methods	Type	Advantages	Disadvantages
<b>Conventional</b>			
Excavation, transfer, and disposal to landfill	<i>Ex situ</i> soil	Effective at removal of shallow contaminants	Expensive, disruptive, relocates problem to another site (landfill)
Excavation and incineration, soil washing	<i>Ex situ</i> soil	Effective at removal of shallow contaminants	Expensive, disruptive
Bioremediation/ natural attenuation	<i>In situ</i> soil and groundwater	Works well when conditions are right	Costly in monitoring and time consuming
Soil vapour extraction	<i>In situ</i> soil, groundwater with extraction pumping	Works well when conditions are right	Costly to operate
Pump and treat	<i>Ex situ</i> groundwater treatment	Works well for free product removal and providing hydraulic control	Generally ineffective at remediation, costly in investment and very time consuming
<b>Less Conventional</b>			
In situ jetting chemical delivery system	<i>In situ</i> 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

# Biodegradación/Bioremediación

**Biodegradation:** Ability acquired 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).



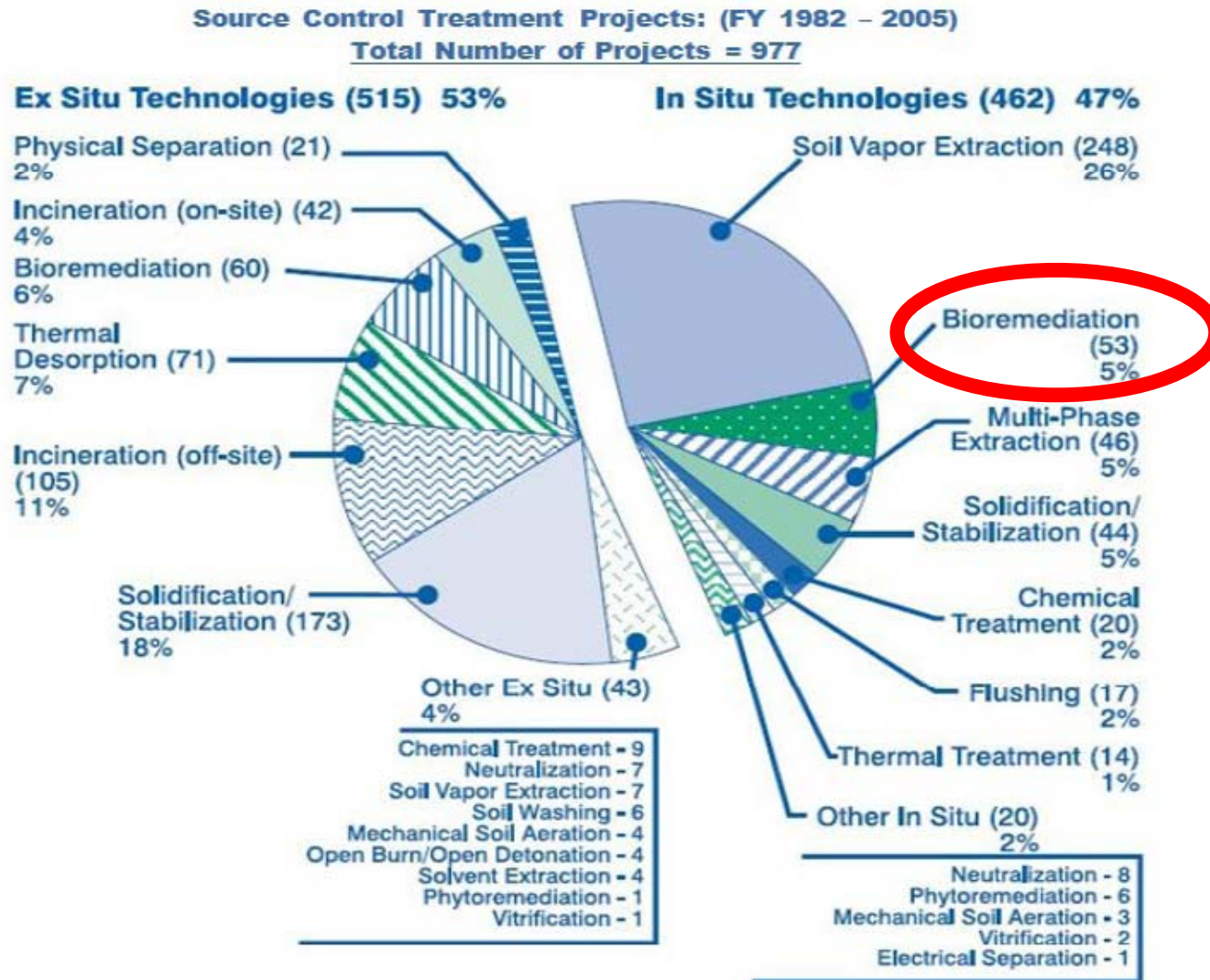
Parales, R.E., Bruce, N.C., Schmid, A. & Wackett, L.P. (2002). Biodegradation, biotransformation, and biocatalysis (b3). *Appl. Environ. Microbiol.*, **68**, 4699–4709.

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



# 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
-



# UM-BBD (Now EAWAG-BBD)

UNIVERSITY OF MINNESOTA  
BIOCATALYSIS / BIODEGRADATION DATABASE

Home | Pathway Prediction System | PredictBT Workshops | Biochemical Periodic Tables

Microbial biocatalytic reactions and biodegradation pathways.

[FBI mirror](#) | [KEGG mirror](#)

**Scheduled Maintenance**  
Server Unavailable 6am - 8am CST (-5 UTC), Wednesday, February 5, 2008.

- [What's New?](#) Last updated January 23, 2008.
- [Search](#) the UM-BBD for compound, enzyme, microorganism, pathway, or BT rule name; chemical formula; chemical structure; CAS Registry Number; or EC code.
- Pathways and Metapathways in the UM-BBD

- Select a Pathway -  
Go to the Pathway

- Lists of [170 pathways](#); [1169 reactions](#); [1095 compounds](#); [744 enzymes](#); [429 microorganism entries](#); [231 biotransformation rules](#); [50 organic functional groups](#); [76 reactions of naphthalene 1,2-dioxygenase](#); [109 reactions of toluene dioxygenase](#); [Graphical UM-BBD Overview](#); and [Other Graphics](#) (Metapathway and Pathway Maps and Reaction Mechanisms).
- *cite using:*  
Ellis LBM, Roe D, Wackett LP (2006) "The University of Minnesota Biocatalysis/Biodegradation Database: The First Decade," *Nucleic Acids Research* **34**: D517-D521. [Abstract](#) | [Full Text](#)
- For more information [join our email list](#). Comments related to the database are appreciated; [contact us](#).

Home | Search | **About** | [UM-BBD](#) | [PPS](#) | [BPT](#) | What's New | FAQs | Join E-mail List | Contributors | Publications | Links | Acknowledgements | Contact Us

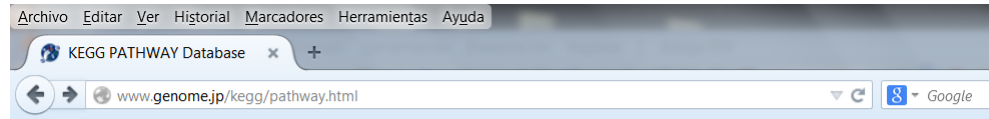
TWIN CITIES  
UM-BBD

Internet 100%

• <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)



## KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Select prefix

Enter keywords

map

Organism

Go

Help

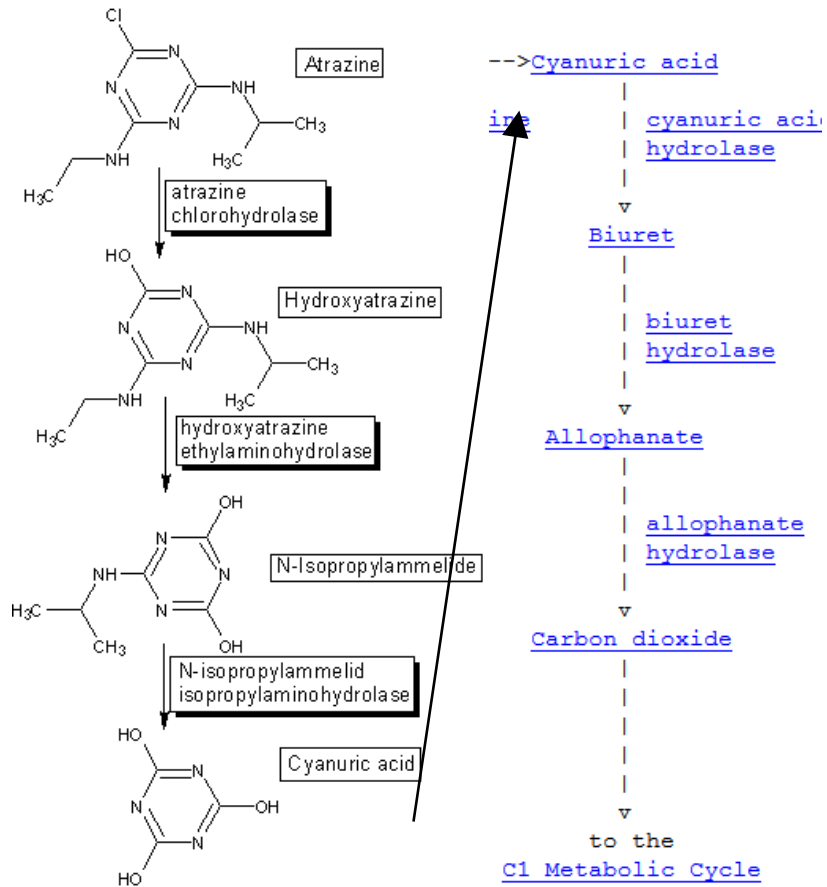
[ New maps | Update history ]

### Pathway Maps

1.11 Xenobiotics biodegradation and metabolism

- Benzoate degradation
- Aminobenzoate degradation
- Fluorobenzoate degradation
- Chloroalkane and chloroalkene degradation
- Chlorocyclohexane and chlorobenzene degradation
- Toluene degradation
- Xylene degradation
- Nitrotoluene degradation
- Ethylbenzene degradation
- Styrene degradation
- Atrazine degradation
- Caprolactam degradation
- DDT degradation
- Bisphenol degradation
- Dioxin degradation
- Naphthalene degradation
- Polycyclic aromatic hydrocarbon degradation
- Furfural degradation
- Steroid degradation
- Metabolism of xenobiotics by cytochrome P450
- Drug metabolism - cytochrome P450
- Drug metabolism - other enzymes

- Endocrine disrupting compounds
- Pesticides
- Cytochrome P450



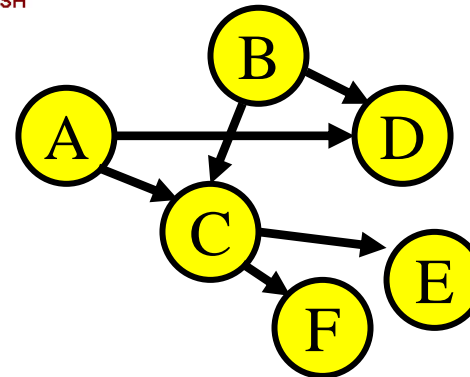
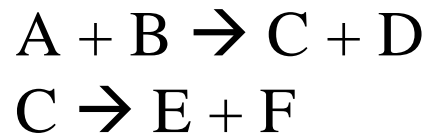
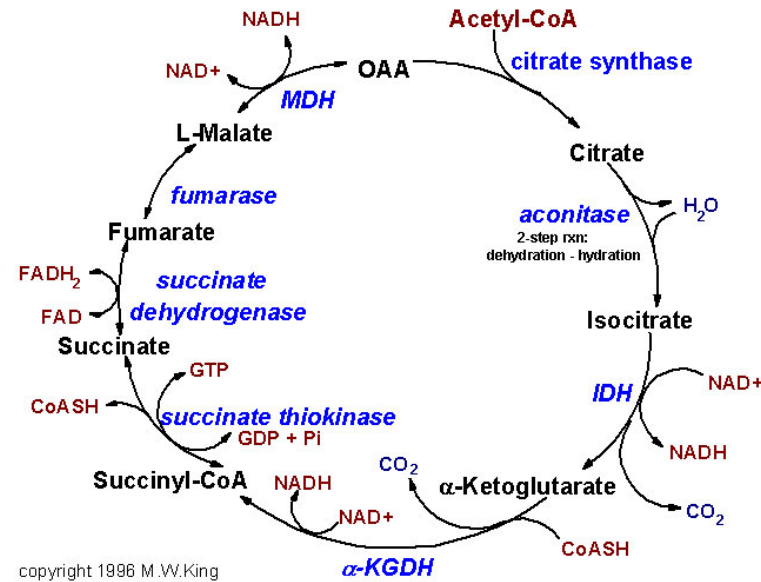
# Aplicaciones Bioinformáticas en Biodegradación

---

- 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
-

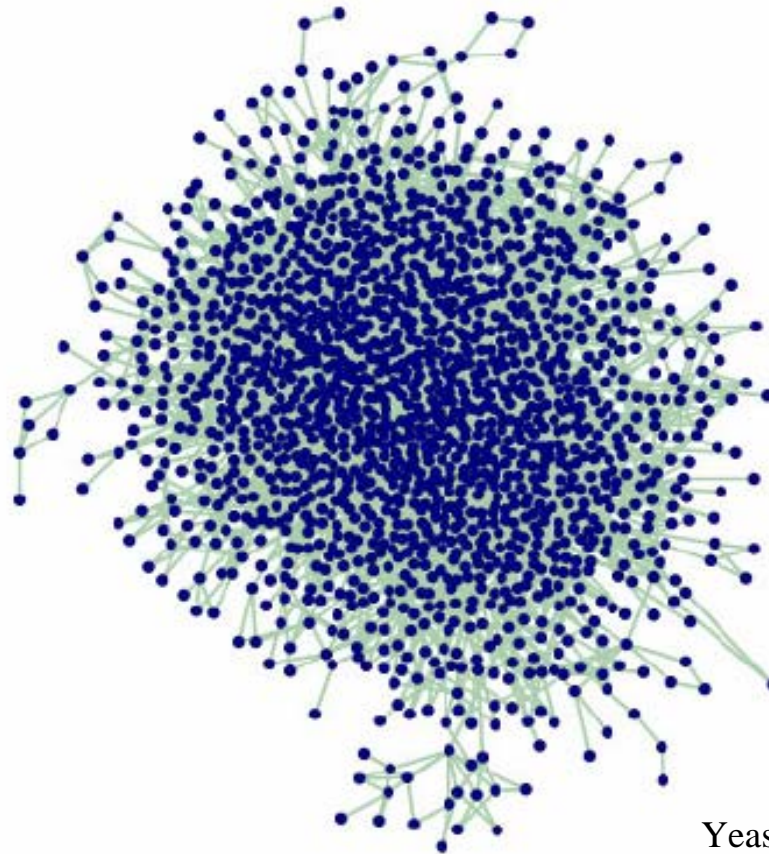
# Estudio sistémico de redes metabólicas (metabolismo central)

## Representación del metabolismo como redes (grafos)



$\forall A, B, C, \dots \neq \text{H}_2\text{O}, \text{ions, cofactors, or } \text{Nr} > \text{X}$

# Estudio sistémico de redes metabólicas (metabolismo central)

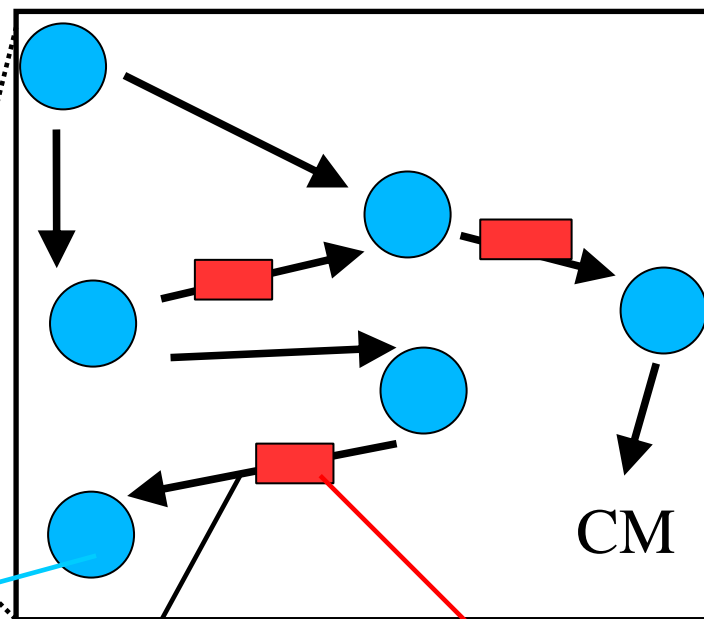
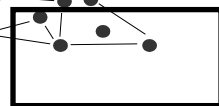
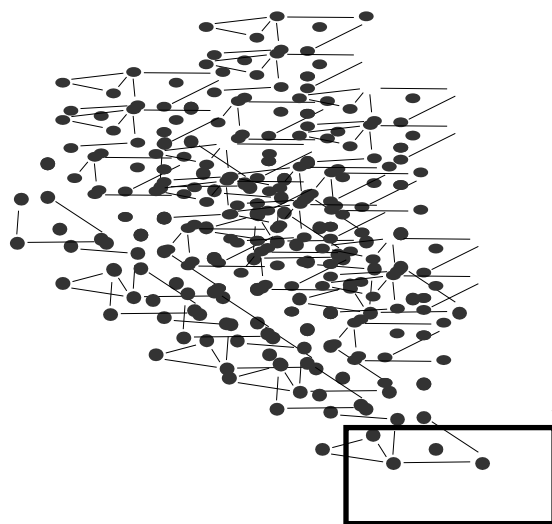


Yeast metabolic network

- 
- 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

**UMBBBD**  
 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

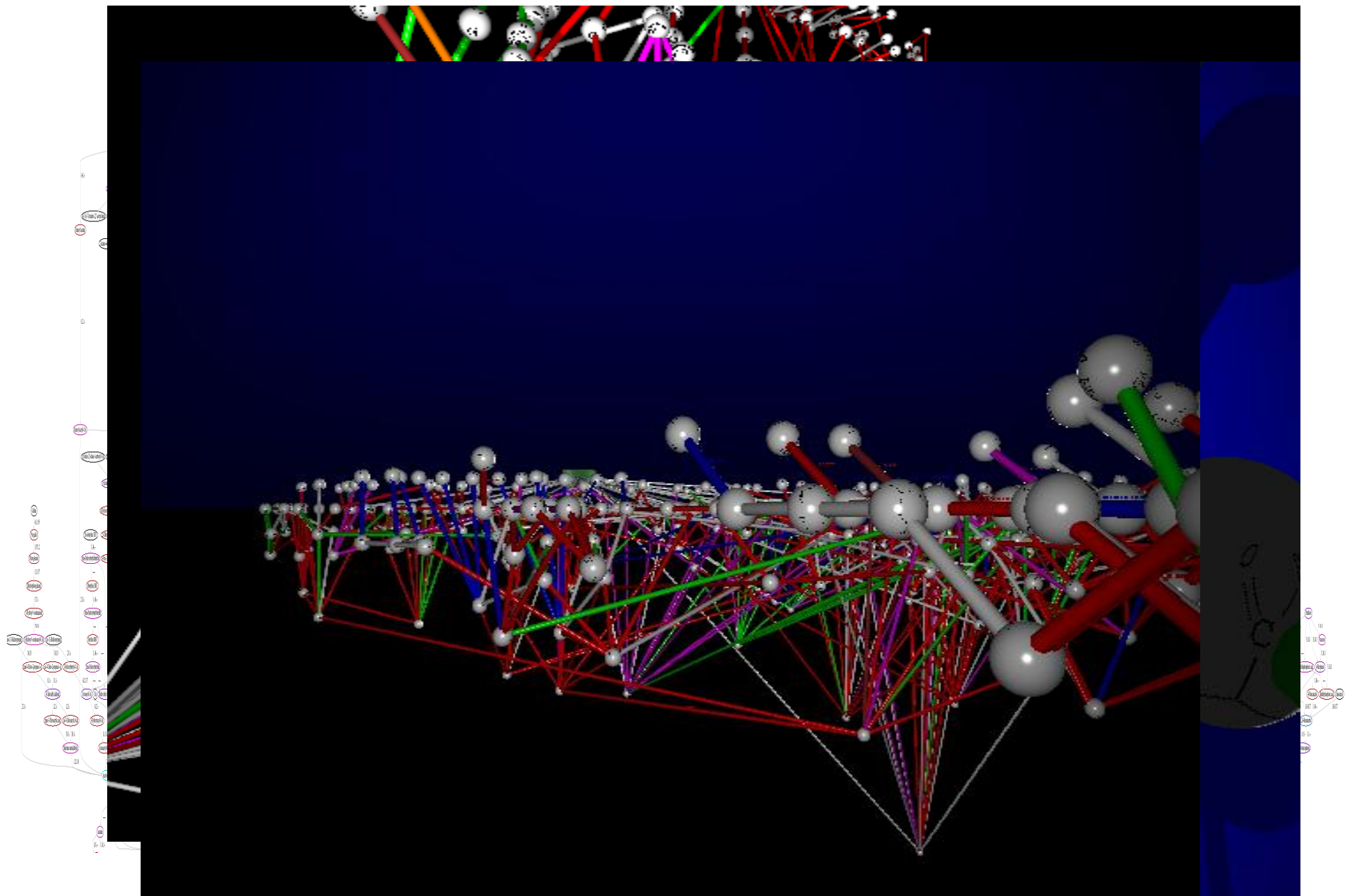
308 specific enzymatic activity (4 EC numbers)

# Red global de biodegradación



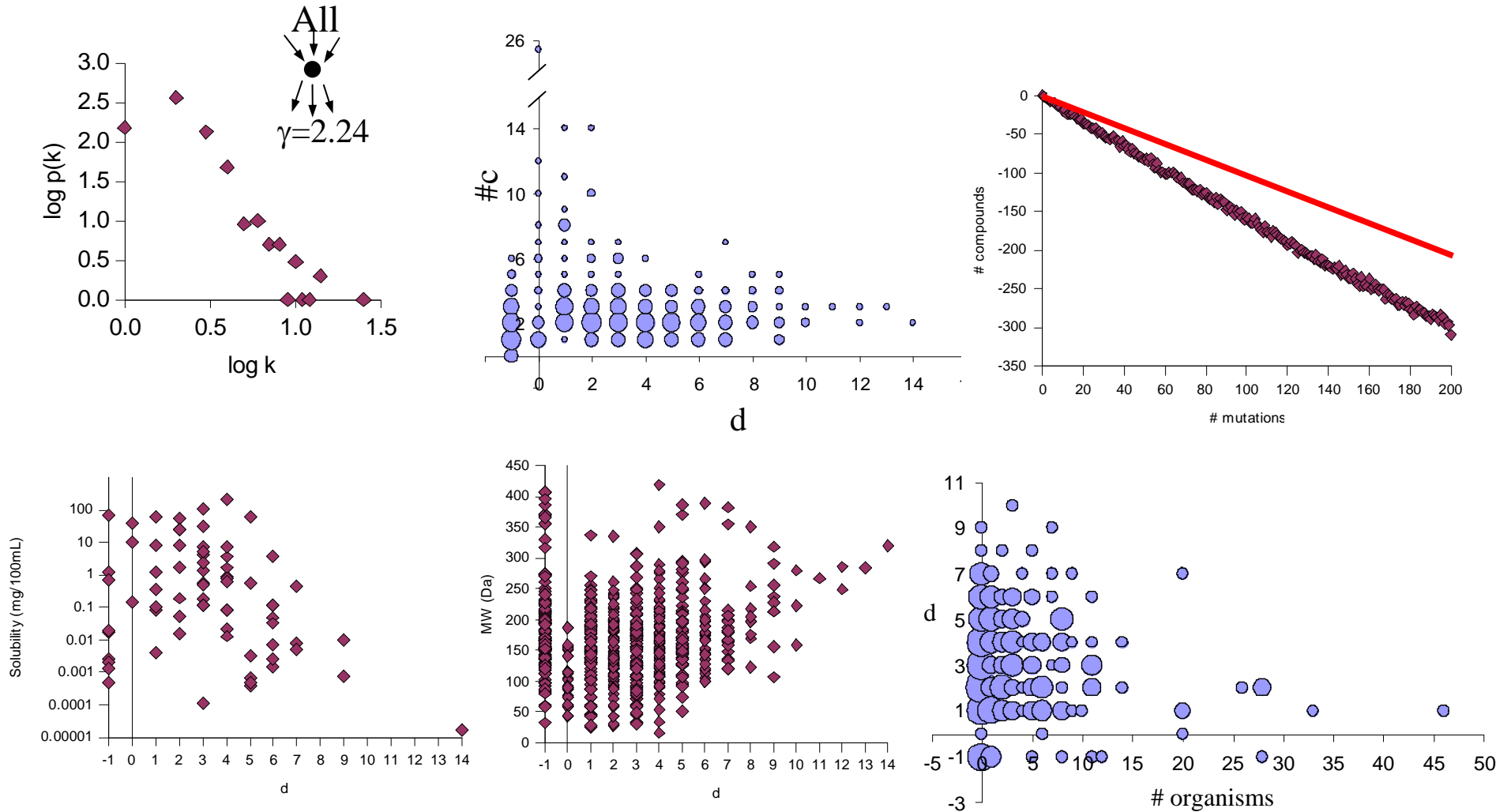
• Florencio Pazos, Victor De Lorenzo & Alfonso Valencia. (2003). The organization of the Microbial Biodegradation Network from a Systems-Biology perspective. *EMBO Rep.* **4(10)**:994-999.

[http://pdg.cnb.csic.es/biodeg\\_net](http://pdg.cnb.csic.es/biodeg_net)



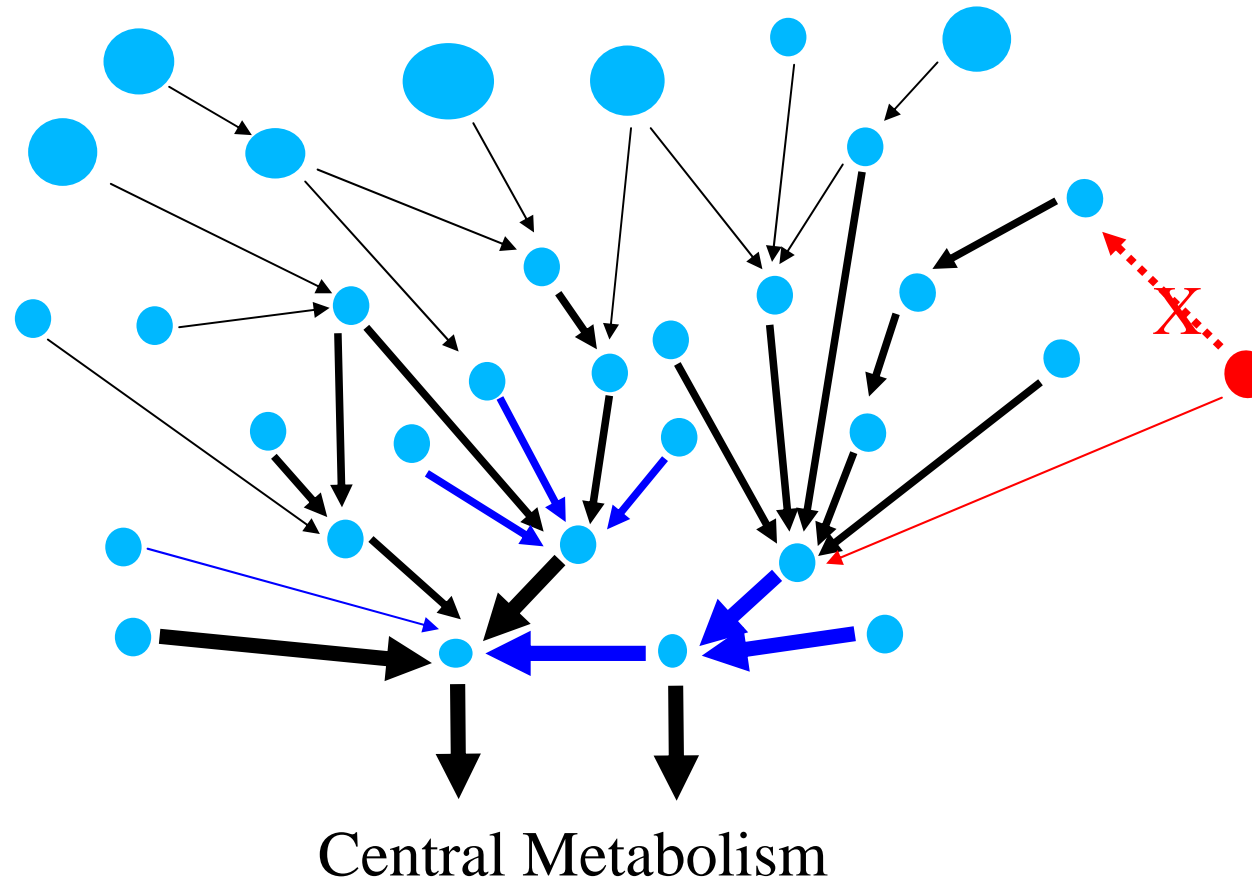


# Propiedades de la red global de biodegradación



• 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 + HSO<sub>3</sub><sup>-</sup> (bt0227)

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

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

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

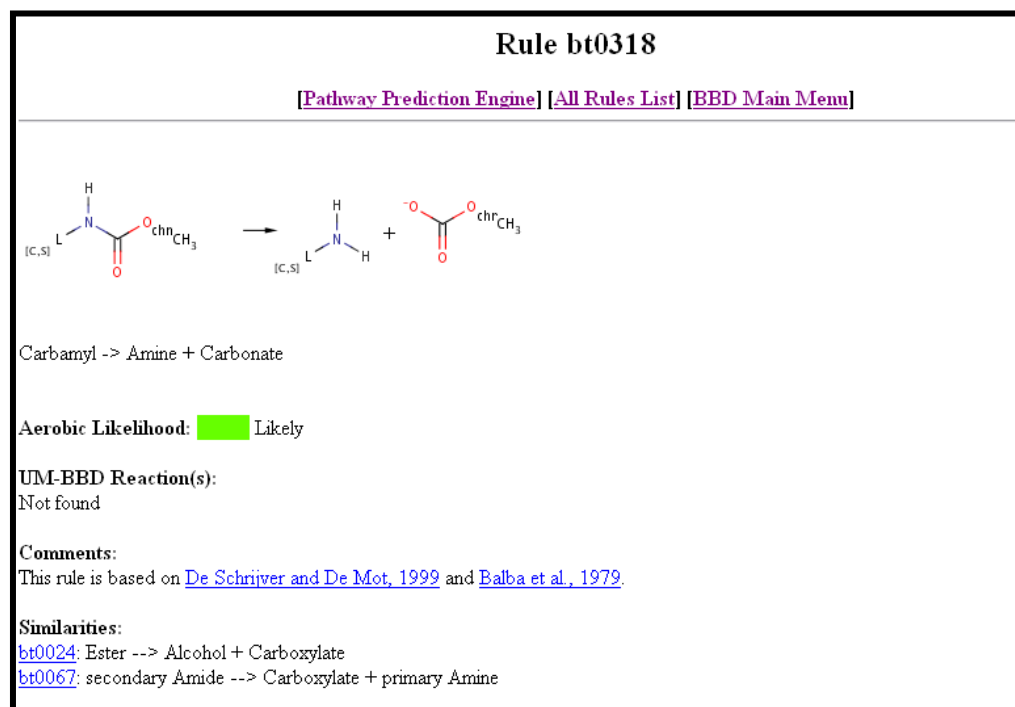
primary Alcohol with halogen on the same carbon -----> Aldehyde (bt0314)

Aldehyde -----> Carboxylate (bt0003)

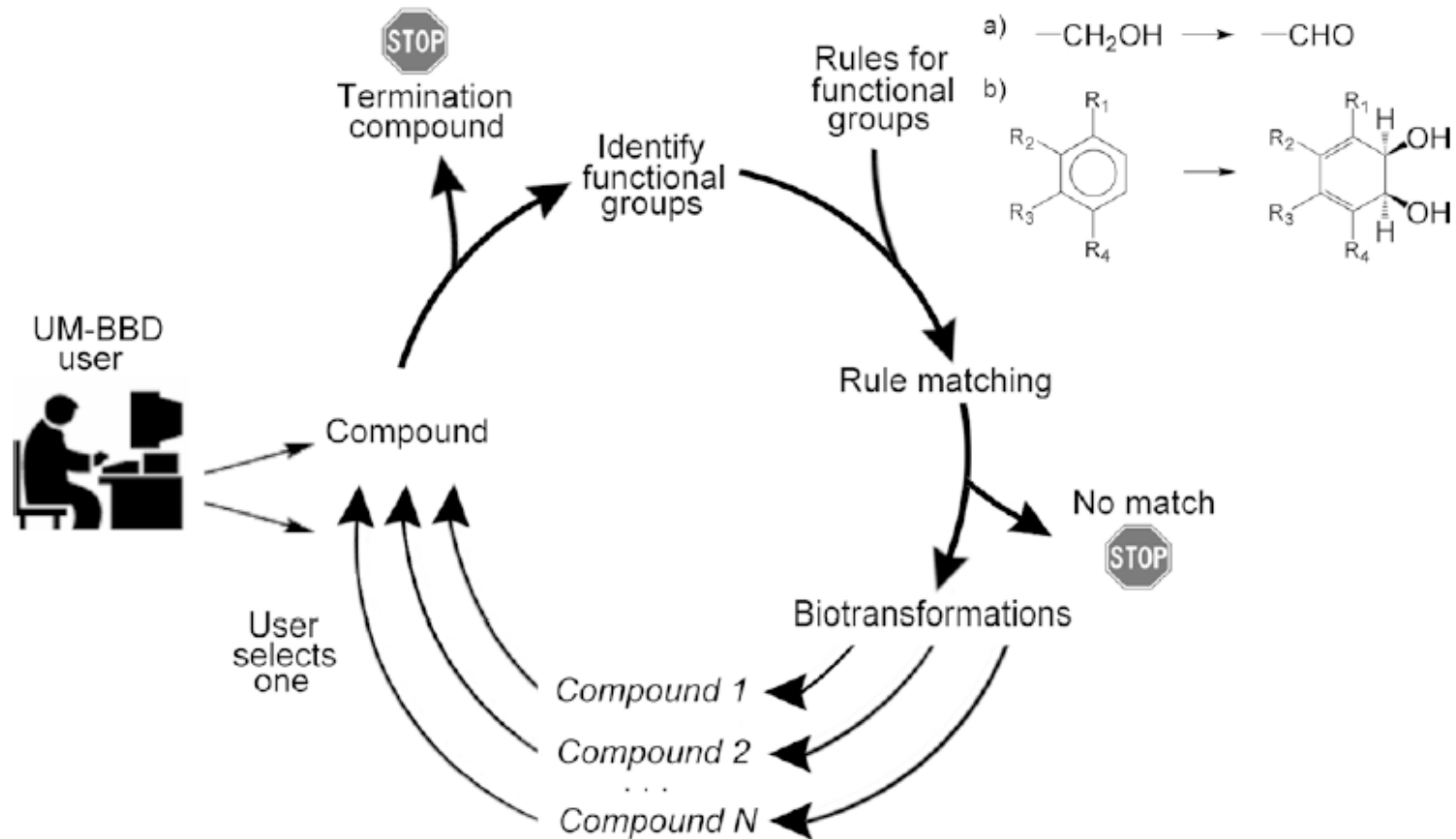
.....

..

.



# Predicción de Biodegradabilidad - *UMBBB-PPS*



# Predicción de Biodegradabilidad - UMBBD-PPS

UNIVERSITY OF MINNESOTA  
BIOCATALYSIS

Home  
Search  
About  
UM-BBD | PPS | BPT  
What's New  
FAQs  
Join E-mail List  
Contributors  
Publications  
Links  
Acknowledgements  
Contact Us

Show BioTransformations:  Aerobic  All

File Edit View Insert Atom Bond Structure Tools Help

Write SMILES

SMILES string:

Continue Clear Demo

Periodic Tables

ways.

ary 5, 2008.

T rule name; chemical

9 microorganism tions of naphthalene Overview; and Other

Internet 100%

• <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.

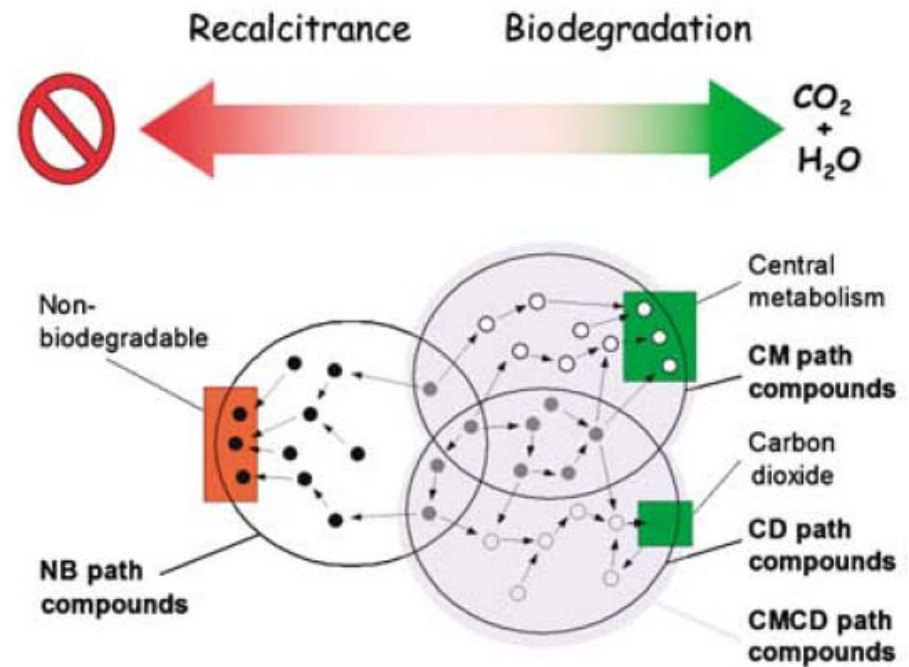
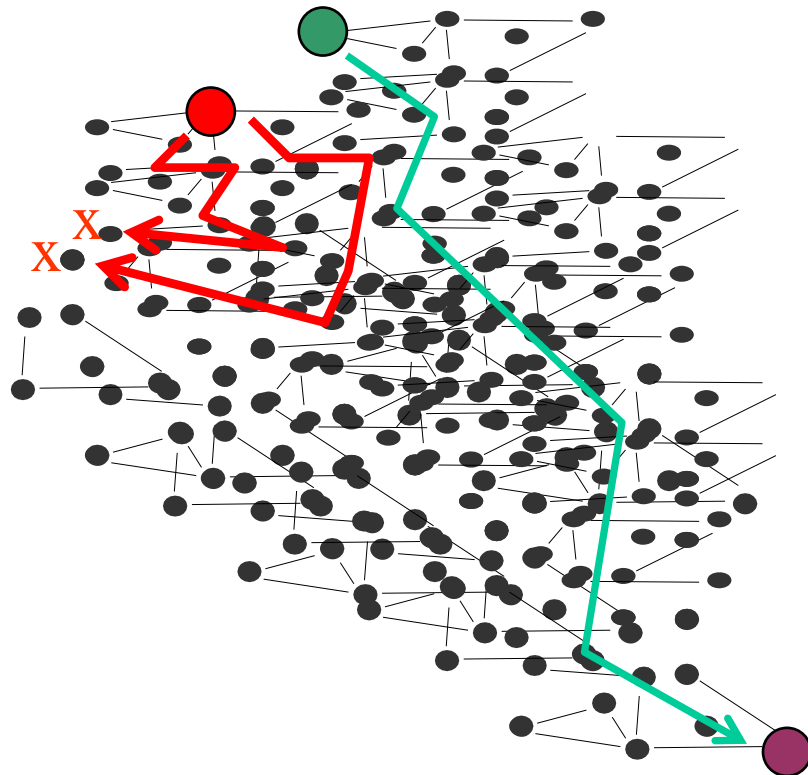
Advantage: fully automatic (non interactive) => can be applied to large collections of compounds

## **Requirements**

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

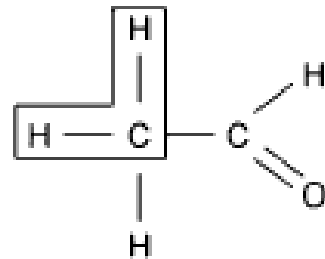
# Obtaining examples of biodegradable/recalcitrant compounds

**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

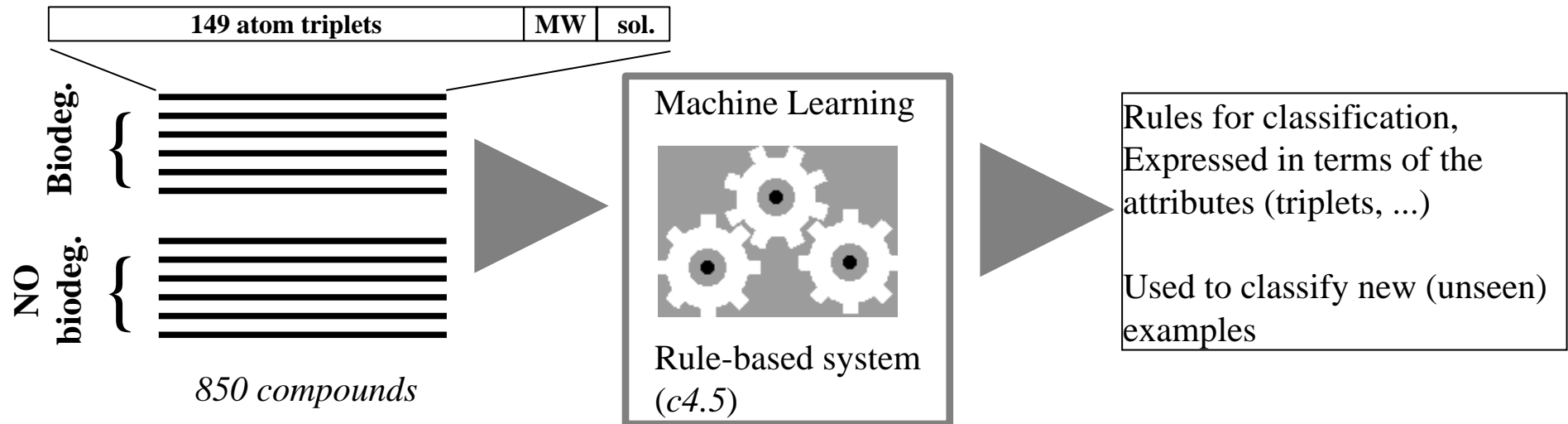
44 Da

Solubility

0.1 g/100 ml

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

# Machine learning system



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

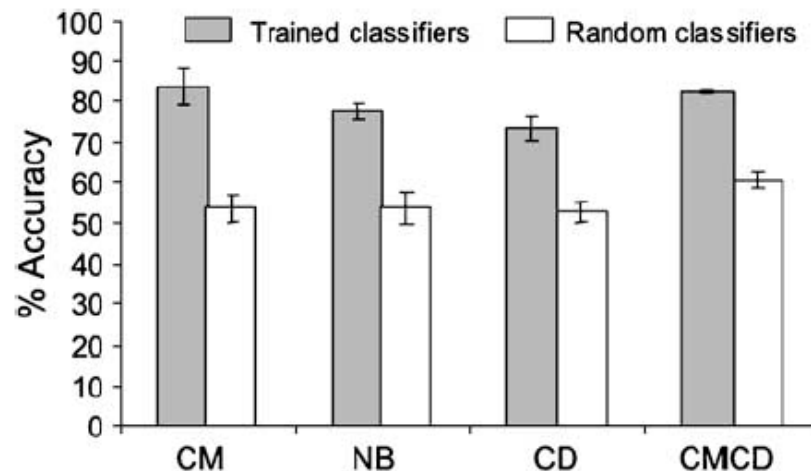
- Explicit rules in human-readable format
- Can handle missing values (solubility)

## Results

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

Rule 55: IF  $-C-C-C > 19$   
 $-O-C-C > 1$   
 $-O-C-C \leq 3$   
 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 <sup>a</sup>	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 <sup>-39</sup>	1.2 × 10 <sup>-24</sup>	2.4 × 10 <sup>-16</sup>	6.4 × 10 <sup>-26</sup>
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**

# Using the predictor - *BDP*Server

**BDP Server - Mozilla Firefox**  
http://www.pdg.cnb.uam.es/BDPSEVER/

## PDG Protein Design Group

### BDP Server: prediction of environmental fate for chemical compounds

[BDP Server](#) [BDP Server DB](#) [Predictions](#) [About BDP Server](#) [UMBDD](#)

#### BDP Server

BDP Server predicts whether chemical compounds can be biodegraded or not.  
Chemical compound descriptions can be typed directly in SMILES format or drawn with the JME applet. Solubility information is optional. The JME applet has been provided by Peter Ertl, from Novartis.

SMILES:

Solubility (g/100 ml):

[Bioinformatics Lab, CAB \(INTA-CSIC\)](#) [Protein Design Group, CNB \(CSIC\)](#) [e-mail contact](#)  
Gómez, Pazos, Guijarro

Done

#### JME Molecular

http://www.pdg.cnb.uam.es - JME Molecular

Submit Molecule Close Help

#### Results

Optional predictions for known compounds: Yes  
Smi: CC(=C)C1CCC(=CC1)C Sol: ?

>> The compound has been found in the BDP Server database, with ID: c0626.

BDP Server classification:	NB:	Yes
BDP Server classification:	CM:	No
BDP Server classification:	CD:	No
BDP Server classification:	CMCD:	No

>> Adding predictions for known compound.

BDP Server prediction:	NB:	Yes	CF = 0.95
BDP Server prediction:	CM:	No	CF = 0.94
BDP Server prediction:	CD:	No	CF = 0.86
BDP Server prediction:	CMCD:	No	CF = 0.91

Done

- 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

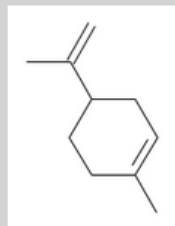
Welcome to SVM Biodegradability predictor. This is a tool for predicting biodegradability and/or toxicity, based on the different definitions provided for some of the most popular compound databases. It is easy to use, just introduce a compound in [SMILES](#) format or design it with the JME applet,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

Input compound in [SMILES](#) format or use [JME](#) editor:

[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)

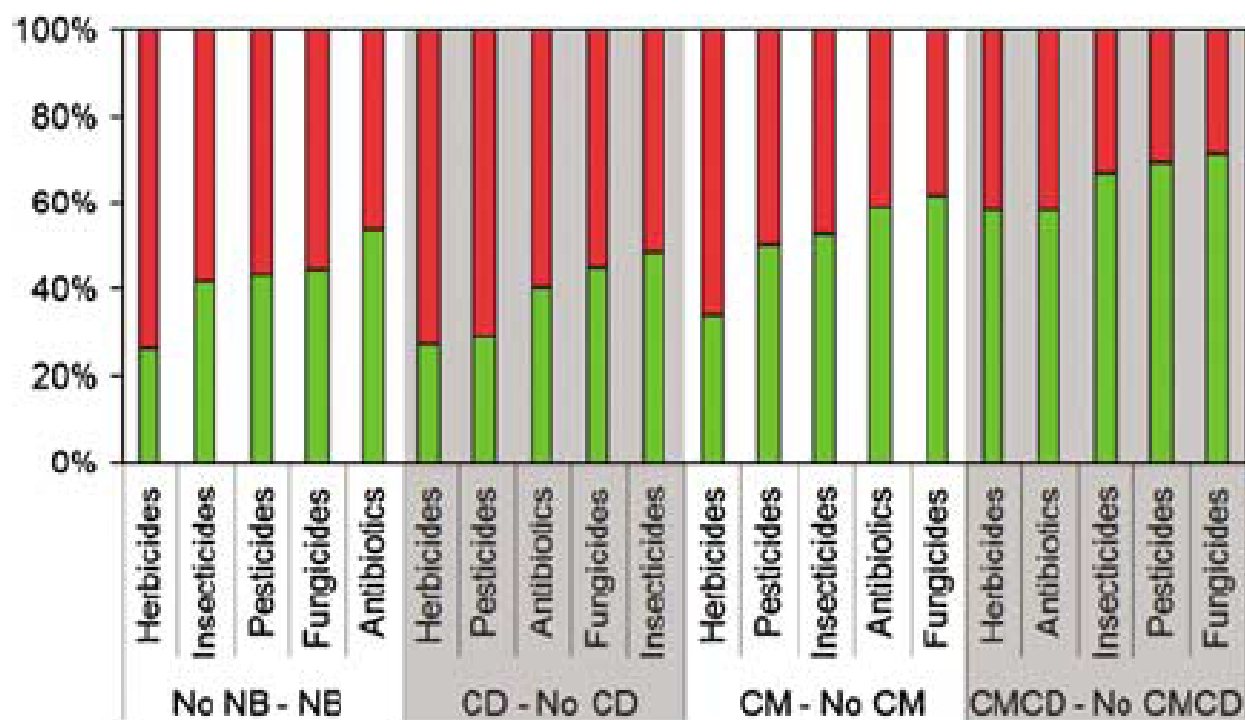
### Contact

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

[Computational Systems Biology Group @CNB-CSIC]

## 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.



# Aplicaciones Bioinformáticas en Biodegradación

---

## 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>*