Front page for deliverables

Project no.	003956
Project acronym	NOMIRACLE
Project title	Novel Methods for Integrated Risk Assessment of Cumulative Stressors in Europe
Instrument	IP
Thematic Priority	1.1.6.3, 'Global Change and Ecosystems' Topic VII.1.1.a ('Development of risk assessment methodologies')

Deliverable reference number and title:

D.2.3.9 Report on the progress of work on combining the reasoning approach of MEPPS and the probabilistic approach of CATABOL

Due date of deliverable: *April 30, 2007* *Actual submission date:* *May 14, 2007*

Start date of project: 1 November 2004 *Duration:* 5 years

Organisation name of lead contractor for this deliverable: Lhasa Limited

Revision [draft, 1, 2, …]: Final Report

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Authors and their organisation:

P N Judson, Lhasa Limited

Deliverable no: D.2.3.9	Nature: Report	Dissemination level: PU	Date of delivery: 14th May 07
Status: Final Report			Date of publishing: 5th June 2007
Reviewed by (name and period): Gabriela Espinosa, Sabcho Dimitrov			

Contents

Introduction

The main objective of the NoMiracle team at Lhasa Limited is to provide a demonstration application that advises on potential biodegradation pathways by reasoning about information from different software sources and with differing degrees of quantification. The demonstrator will draw on information from CATABOL and MEPPS. CATABOL uses a probabilistic approach to assess which are the more likely biodegradation pathways for a query structure. MEPPS uses a qualitative, reasoning-based approach. This report outlines progress to date and what remains to be done.

For brevity in this document, the "umbrella" application under development, which assesses information from CATABOL and MEPPS and produces reports to the user, is abbreviated to "NoMBP" (for "NoMiracle Biodegradation Predictor").

Summary of work completed

1 An overall design has been drawn up for NoMBP.

2 A command line interface has been developed between CATABOL and NoMBP, allowing NoMBP to give queries to CATABOL and to obtain the results.

3 A prototype converter is in use for generating Molfile descriptions of chemical structures from SMILES representations.

4 An interface has been developed between MEPPS and NoMBP.

5 Biodegradation reactions have been chosen for use as examples in the demonstrator, suitable for showing the advantages of reasoning about the output from different prediction packages.

Work completed

1 Overall design for the NoMBP package

The package is being developed using Open Source tools – Eclipse, Subversion, Ant, and Cobetura. The main components of NoMBP, and the links between them, are represented in the diagram in Figure 1.

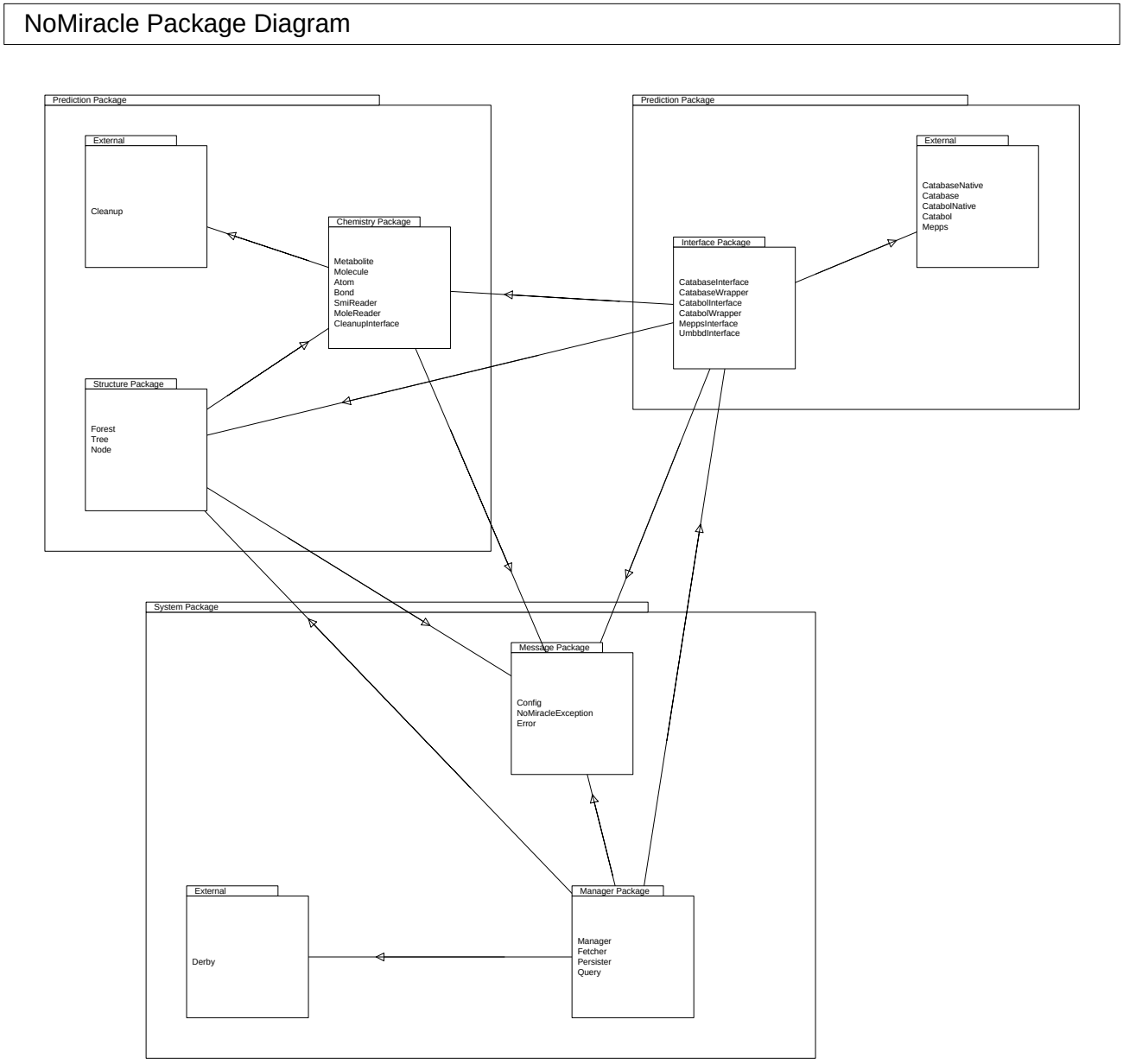

Figure 1. The overall design for the NoMBP package

The application is divided into three main code packages, and a number of sub packages. The three packages are "system", "chemistry", and "prediction". System handles the management

of the application and interfaces between the various internal and external components. Chemistry is responsible for the chemical aspects, converting between molecule formats and handling the tree data. Prediction is responsible for the interfaces to the external components and converting results from those components into a common format.

A query is received from the web interface and stored into the database (system.external.Derby). The entry is then acted on by the application. The system.manager.Manager detects the database entry, and calls a system.manager.Fetcher to handle the query. This Fetcher calls the prediction interfaces (eg prediction.interface.CatabolInterface), each of which is responsible for calling the relevant external component, possibly via an intermediary (eg CatabolNative) that handles computer language impedance issues. The interface returns an instance of chemistry.structure.Forest. Once all the predictions have been acquired the Fetcher calls the results interface (currently under development).

2 Command line interface between CATABOL/CATABASE and NoMBP

We have developed an interface so that NoMBP can pass the structure of a query molecule to CATABOL and CATABOL can return a predicted biodegradation pathway tree. NoMBP can display the tree, as illustrated in Figure 2.

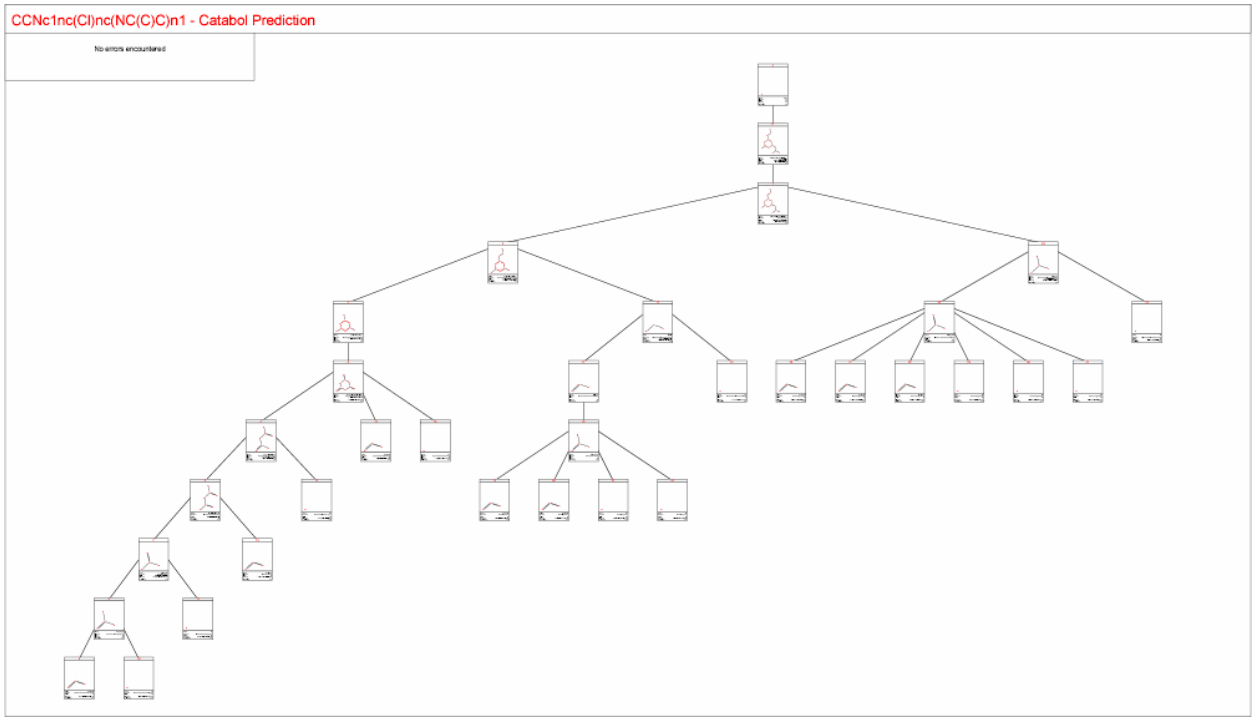

Figure 2. A biodegradation tree from CATABOL

Similarly, a tree can be requested from CATABASE and displayed, as illustrated in Figure 3. CATABASE returns information from a database about the observed biotransformation of known substances, whereas CATABOL can make predictions for substances for which there are no observations, either because the studies have not been done or reported, or because the substances have not been synthesised.

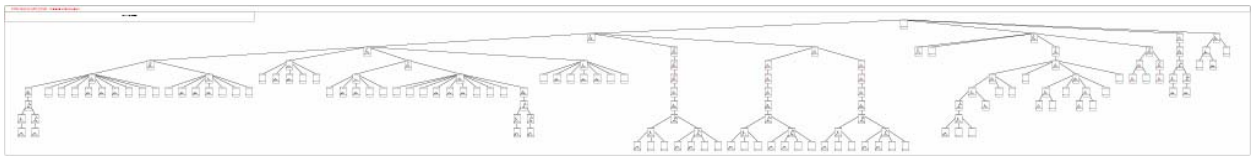

Figure 3. A biodegradation tree from CATABASE

The user can click on a node in a tree display in order to see the node in more detail, as illustrated in Figure 4.

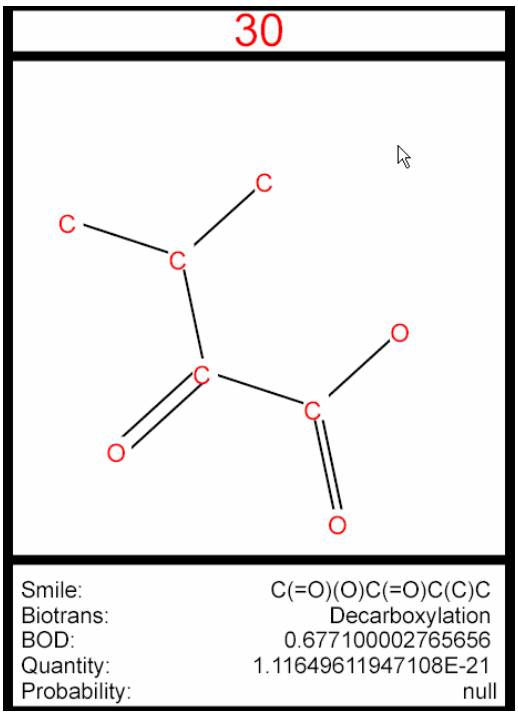

Figure 4. A node expanded from a biodegradation tree display

When the demonstrator is completed, NoMBP will compare these trees with trees generated by MEPPS and will apply reasoning to make recommendations about areas of similarity and difference in the trees. For example, NoMBP might recommend that, for a particular type of query structure, there is high confidence in predictions for certain biotransformations when both applications are in agreement, whereas in others confidence might be high if the prediction comes from one application but low if it comes from the other; there might be rules about how to interpret contradiction between CATABOL and MEPPS.

3 Prototype converter from Molfile to SMILES format

Communication between CATABOL or CATABASE and other applications depends on the SMILES representation of chemical structures. Examples of SMILES code appear in the header in Figure 2 and in the box at the bottom of the display in Figure 4. Many other applications, including MEPPS and NoMBP, communicate via Molfiles. Molfiles describe

chemical structures as connection tables in a format originally published by MDL Information Systems Inc. and now widely used. We have provided a converter between the two types of representation so that NoMBP can communicate with CATABOL and CATABASE.

4 Interface between MEPPS and NoMBP

We have developed an interface so that NoMBP can pass the structure of a query molecule to MEPPS and MEPPS can return a predicted biodegradation pathway tree. NoMBP can display the tree, as illustrated in Figure 5.

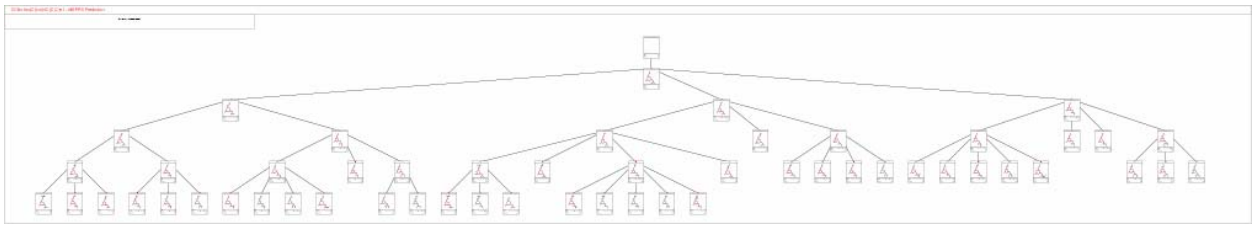

Figure 5. A biodegradation tree from MEPPS

As mentioned earlier in this report, when the demonstrator is completed, NoMBP will compare trees generated by MEPPS and CATABOL and will apply reasoning to make recommendations about areas of similarity and difference in the trees.

5 Choice of biodegradation reactions

NoMBP will be a prototype system containing a limited knowledge base of information about the relationships between different biodegradation reactions and pathways and how they are predicted by CATABOL and MEPPS. It is necessary to choose just one or two key areas of chemistry to be covered by reasoning in the demonstrator. The choice needs to be in an area for which sufficient information has been published for us to be able to confirm that the recommendations coming out of NoMBP are meaningful. In addition, it should relate to a topic that is, or has been, of practical relevance in environmental science.

Our preferred candidates are dehalogenation and oxidative ring hydroxylation and competition between them, which are relevant to the biodegradation of chemicals such as atrazine. It is expected that these reactions will provide the scope and interest needed for a meaningful demonstration as well as being important to the environmental and health objectives of NoMiracle. In the event that they turn out to be disappointing as examples, alternatives will be considered.

Work in progress

The following is a summary of work currently in progress.

1 Development of the knowledge base and reasoning rules for NoMBP. The knowledge-base researcher in the project team visited collaborators at the University of Minnesota early in May in connection with work on MEPPS, and the collaborators visited us in Leeds in mid-May. We expect spin-off benefit from these meetings for NoMiracle.

2 Incorporation of the Lhasa reasoning engine into NoMBP. This will be the core of the system for making judgements about the inter-relationships between predictions, whether they be corroboratory or contradictory.

3 Modifications to data collation to allow use of more complex and varied data.

4 Completion of the NoMBP user interface.

5 Further technical development the NoMBP software (e.g. on persistence of data).

6 Ongoing creation of design documentation as new functionality is added.

7 Development of improved software tools for entering knowledge into the knowledge-base.

Work scheduled for the next six months

1 Addition of a module for the comparison of complete molecular structures.

2 Finalisation of the user interface and output functionality, including completion of design documentation.

3 Completion of functional testing.

4 Provision of hardware to allow web access to the demonstrator by NoMiracle partners and other appropriate individuals.

P N Judson
5th June 2007