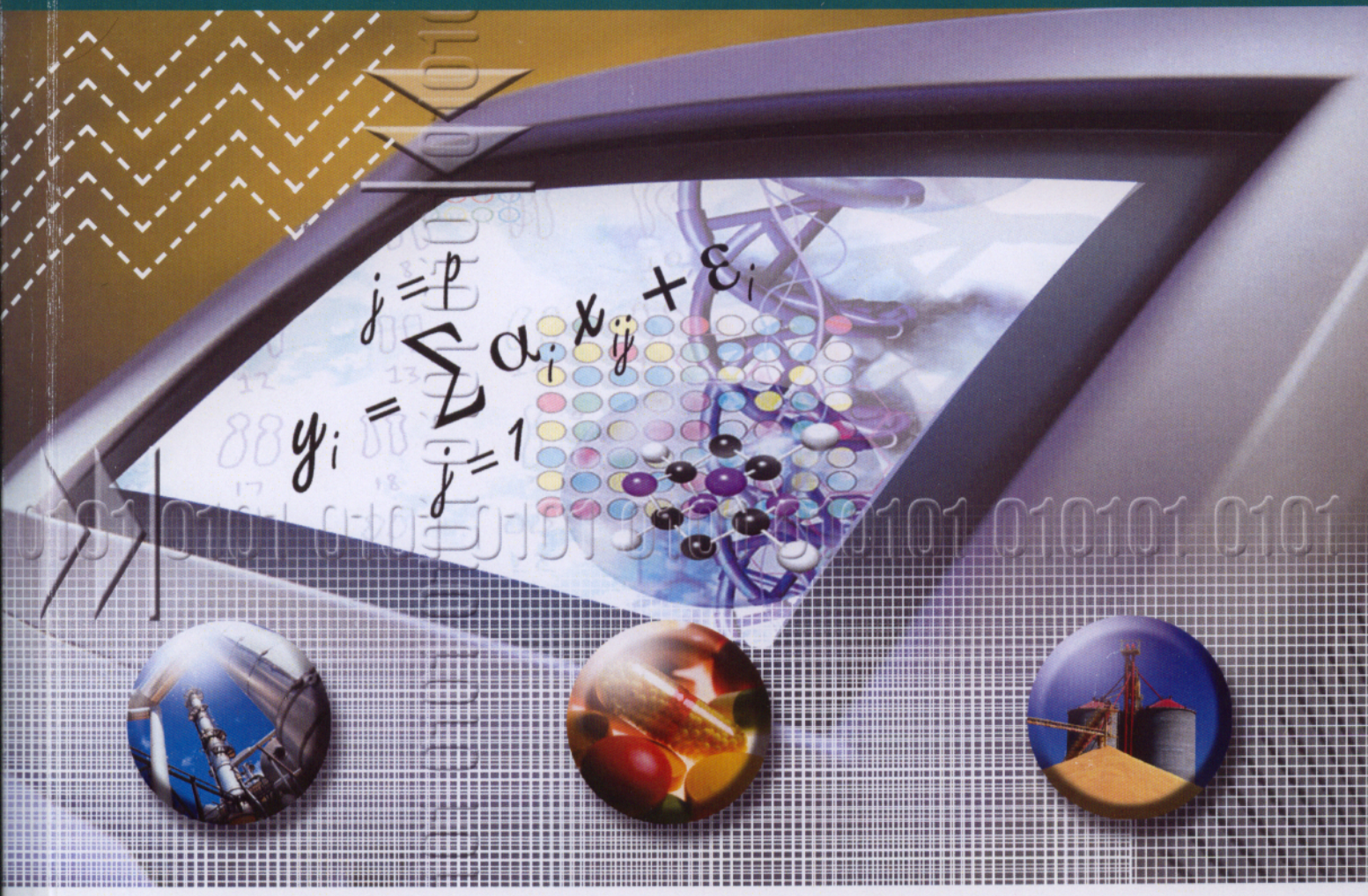
3-4 décembre 2003 - CNAM - PARIS



Chimiométrie 2003

PROGRAMME

Groupe Français de Chimiométrie

Avec le soutien de :

CNRS Centre National de la Recherche Scientifique

SFdS Société Française de Statistique

eurofins

Cnam

"IN SILICO" DESIGN OF NEW POTENTIAL METAL BINDERS

D. Fourches, V. P. Solov'ev and A. Varnek *

Laboratoire d'Infochimie, UMR 7551,
Université Louis Pasteur,
4, rue B. Pascal, 67000, Strasbourg
varnek@chimie.u-strasbg.fr

RÉSUMÉ. *Une suite d'outils informatiques (une base de données pour l'extraction liquide-liquide, un système expert et un générateur de bibliothèques combinatoires) a été employé pour le développement « in silico » de nouveaux extractants de l'uranyle. Les calculs effectués sur une famille de podands phosphorylés ont permis d'établir des modèles « structure - activité » qui ont été ensuite utilisés pour estimer les coefficients de distribution de l'uranyle (logD) par des composés virtuels générés à l'aide du module combinatoire. Huit nouveaux composés sélectionnés à partir de la bibliothèque combinatoire virtuelle, ont été synthétisés et testés expérimentalement. Les résultats de ces tests montrent que les calculs prédisent correctement les valeurs logD pour 7 composés sur 8.*

MOTS-CLÉS : *design « in silico », méthodes QSPR, podands phosphorilés, extraction liquide-liquide, uranyle.*

1. Introduction

Solvent extraction involves a cation-ligand complexation in one of the liquid phases or at the liquid/liquid interface, accompanied by transfer of the complexes into bulk organic phase. Development of new extraction systems with desirable properties generally proceeds in empirical manner because of complexity of studied processes. Indeed, thermodynamics parameters of extraction depend on many variables (the nature of metal(s), conterion(s), ligand(s), pH, organic solvent and background compounds) and, therefore, their theoretical modelling represents a very difficult task.

In fact, a rational development of new extraction systems with desired characteristics could be possibly based on the methods of chemoinformatics which suggest three main elements indispensable to *in silico* design of new extractants: (*i*) a comprehensive database, (*ii*) an expert system which builds quantitative structure-property relationships (QSPR) and (*iii*) a generator of combinatorial libraries. Figure 1 illustrates links between these modules: information collected in the database is treated empirically by the expert system which establishes relationships between structure of compounds and their extraction properties. Then, structure-property models are applied to screen virtual combinatorial library leading to potential actives which then can be experimentally tested, in turn providing the database with new information.

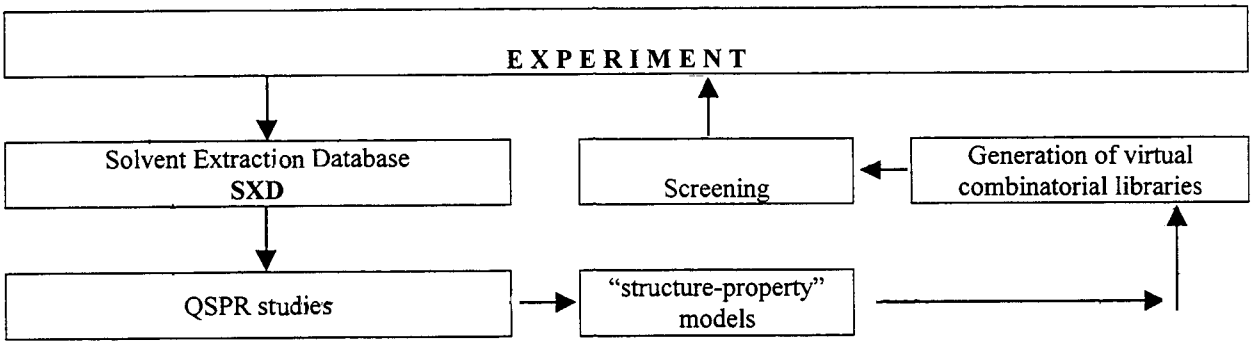
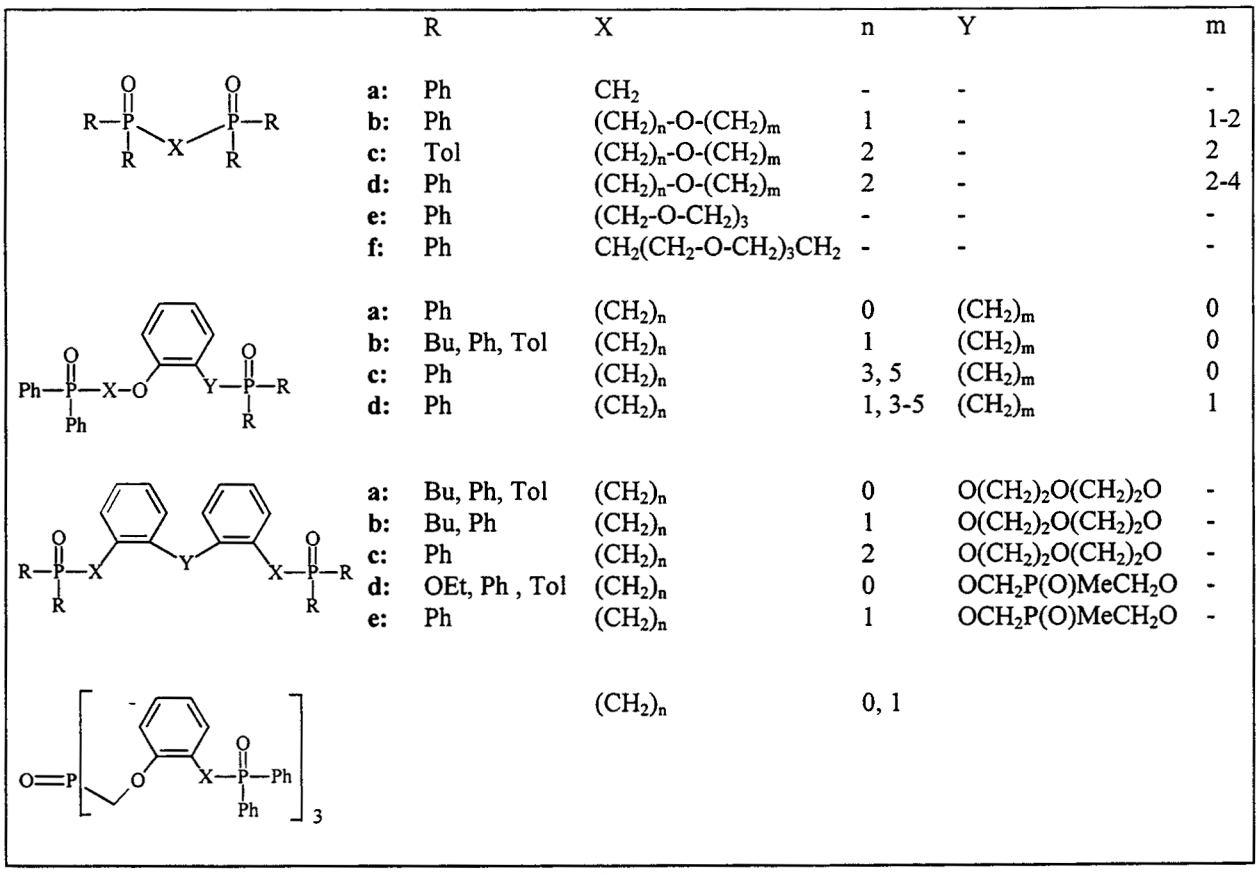


Figure 1. The strategy applied for "*in silico*" design of new ligands for solvent extraction.

The goal of this work is computer-aided design of new phosphoryl-containing podands which efficiently extract the uranyl cation from water to the organic solvent. Phosphoryl-containing podands (Figure 2) are acyclic molecules with polyether spacer(s) linking two (in monopodands), three (in dipodands and tripodands) or four (in tripodands) terminal phosphine oxide groups.

A given study involves four main steps: (*i*) selection of the data from comprehensive Solvent eXtraction Database, (*ii*) QSPR modeling of the logarithm of the distribution coefficient (log*D*) of UO_2^{2+} which quantitatively measures extraction efficiency of ligands, (*iii*) generation and screening of a virtual combinatorial library, and, (*iv*) synthesis of theoretically predicted compounds and their experimental extraction studies.

Two different QSPR approaches were used in this work: a Hansch–type approach which uses as descriptors some physico-chemical parameters calculated either by quantum mechanical methods or by some empirical techniques, and the Free-Wilson–type approach which uses molecular fragments as variables in a multiple regression analysis. The first type of calculations was performed with the CODESSA-PRO program [KAR 99], whereas the calculations with fragment descriptors (atom/bond sequences or augmented atoms) were performed using Substructural Molecular Fragments method implemented into the TRAIL program [SOL 00, VAR 01]. All these descriptors are derived solely from molecular structure and do not require experimental data or expensive theoretical calculations to be obtained.



	R	X	n	Y	m
a:	Ph	CH_2	-	-	-
b:	Ph	$(CH_2)_n$-O-$(CH_2)_m$	1	-	1-2
c:	Tol	$(CH_2)_n$-O-$(CH_2)_m$	2	-	2
d:	Ph	$(CH_2)_n$-O-$(CH_2)_m$	2	-	2-4
e:	Ph	$(CH_2$-O-$CH_2)_3$	-	-	-
f:	Ph	$CH_2(CH_2$-O-$CH_2)_3CH_2$	-	-	-
a:	Ph	$(CH_2)_n$	0	$(CH_2)_m$	0
b:	Bu, Ph, Tol	$(CH_2)_n$	1	$(CH_2)_m$	0
c:	Ph	$(CH_2)_n$	3, 5	$(CH_2)_m$	0
d:	Ph	$(CH_2)_n$	1, 3-5	$(CH_2)_m$	1
a:	Bu, Ph, Tol	$(CH_2)_n$	0	$O(CH_2)_2O(CH_2)_2O$	-
b:	Bu, Ph	$(CH_2)_n$	1	$O(CH_2)_2O(CH_2)_2O$	-
c:	Ph	$(CH_2)_n$	2	$O(CH_2)_2O(CH_2)_2O$	-
d:	OEt, Ph , Tol	$(CH_2)_n$	0	$OCH_2P(O)MeCH_2O$	-
e:	Ph	$(CH_2)_n$	1	$OCH_2P(O)MeCH_2O$	-
		$(CH_2)_n$	0, 1		

Figure 2. Phosphoryl-containing podands studied in this work.

2. Results

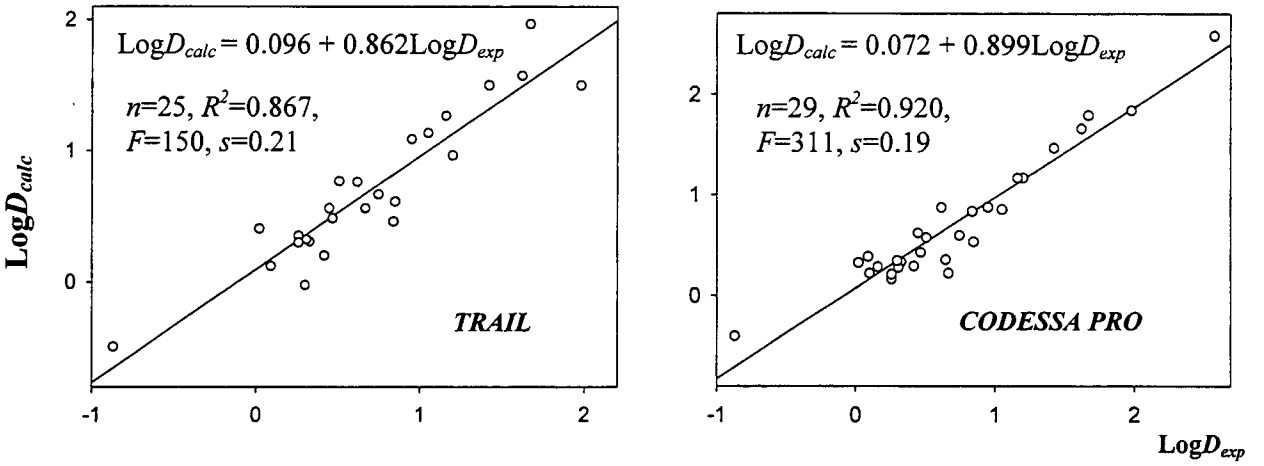
2.1 Building the structure – property models.

The parent data set containing logD values for 32 podands studied at the same conditions was selected from the Solvent eXtraction Database [VAR 01]. The QSPR models were obtained for 29 molecules from the training set, which then are used for prediction calculations for the 3 compounds from the test set (Figure 3). Calculations with TRAIL involved 147 computational models using 49 different types of fragments and 3 linear and non-linear fitting equations. Several best models involving atom/bond sequences and "augmented" atoms were selected from those studies. The best models built by CODESSA-PRO involved both physico-chemical descriptors and fragment descriptors generated by TRAIL. Figure 3 shows that the log*D* values calculated for the test set are very close to the experimental ones.

2.2 Generation and screening the combinatorial virtual library and experimental blind test.

To generate virtual combinatorial libraries of potential extractants we used an algorithm based on the *Markush* structures which allows user to attach certain number of fragments to a molecular "core". The generation module allows user to select interactively molecular "bricks" and to launch computer "synthesis" of their possible combinations. Thus, a set of 2200 virtual podands was prepared and saved in the SD format to be used as an input by TRAIL for the assessment of log*D* values. Then, we have selected 8 podands which span the range of log*D* variation

for experimentally studied molecules (Figure 4) and estimated their extraction ability using the best models of CODESSA-PRO. Then, those podands were synthesized [BAU] and experimentally studied [TUR] as uranyl extractants using the same protocol as in previous studies of the compounds from the parent set [TUR 98]. Comparison of experimental and predicted log*D* values (Figure 5) shows that our QSPR models reasonably estimate log*D* for 7 from 8 compounds from the "blind test" set.



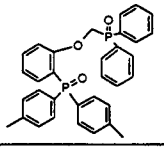
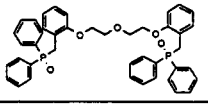
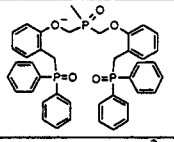
no.	compound	LogD exp	LogD predicted TRAIL	LogD predicted CODESSA
1		1.20	0.70	0.93
2		-0.20	-0.29	0.04
3		1.72	1.70	1.51
	R^2		*0.931*	*0.981*
	s		*0.37*	*0.14*

Figure 3. Extraction of uranyl nitrate by phosphoryl-containing podands in dichloroethane. The graphs show the correlation between experimental and calculated with TRAIL (left) and with CODESSA-PRO (right) distribution coefficients (log*D*) at the training stage. The **Table** reports the experimental (exp) and "predicted" (calc) log*D* values for the molecules from the test set. *R* and *s* are correlation coefficient and standard deviation for the correlation between experimental and predicted log*D* values.

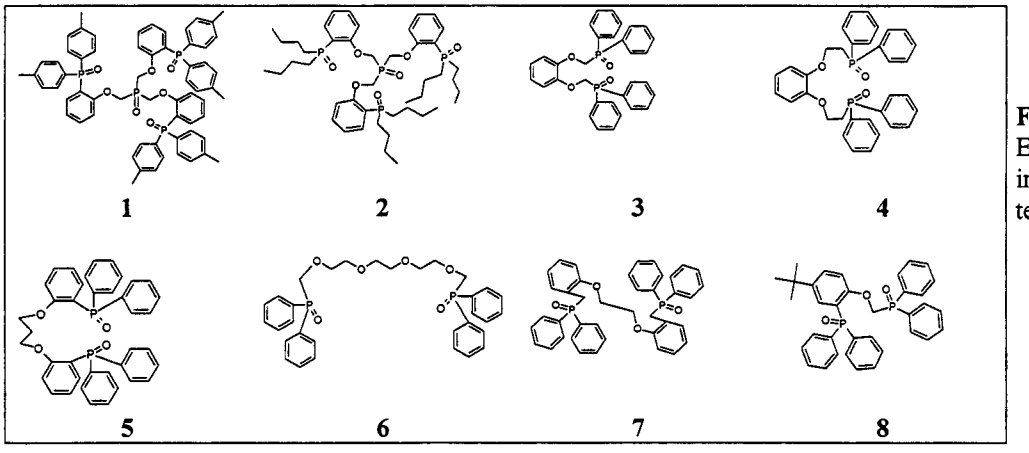


Figure 4. Eight compounds included in the "blind test" set.

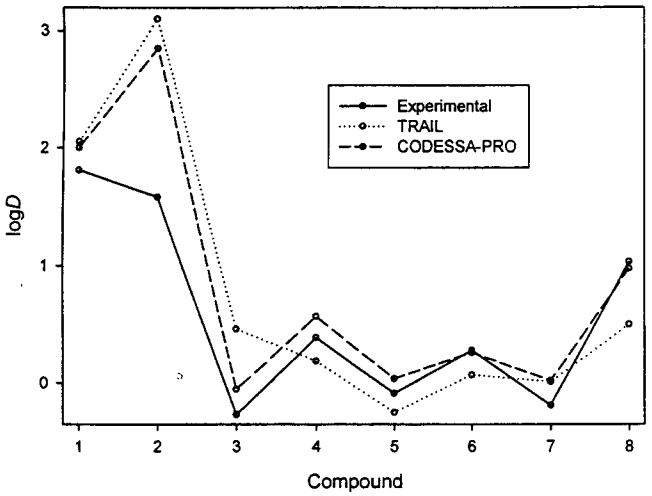


Figure 5 . Experimental and predicted log*D* values for the 8 compounds from the "blind test" set.

3. Conclusion

A number of tools including a comprehensive database, an expert system, and a combinatorial module has been developed and applied for "*in silico*" design of several phosphoryl-containing podands. A structure - property models for the distribution coefficient (log*D*) of uranyl extracted by podands were applied to assess extraction ability of virtual molecules from generated combinatorial library. Eight of these hypothetical compounds were then synthesized and tested experimentally. Comparison of calculated and new experimental results show that the log*D* values for 7 of 8 compounds were successfully predicted. This opens a perspective for computational design of new efficient extractant molecules.

4. Bibliographie

[KAR 99] KARELSON M., MARAN U., WANG Y., KATRITZKY A. R. "QSPR and QSAR models derived using large molecular descriptor spaces. A review of CODESSA applications", *Collection of Czechoslovak Chemical Communications,* vol. 64, n° 10, 1999, p. 1551-1571.

[SOL 00] SOLOV'EV V. P., VARNEK A., WIPFF G. "Modeling of ion complexation and extraction using substructural molecular fragments", *J. Chem. Inf. Comput. Sci.*, vol. 40, n° 3, 2000, p. 847-858.

[VAR 01] VARNEK A., WIPFF G., SOLOV'EV V. P. "Towards an information system on solvent extraction", *Solvent Extr. Ion Exch.,* vol. 19, n° 5, 2001, p. 791-837.

[BAU] Synthesis of new podands has been performed by Dr V. E. Baulin from the Institute of Physiologically Active Substances of Russian Academy of Sciences, Chernogolovka.

[TUR] Extraction experiments have been performed by Dr A. N. Turanov from the Institute of Solid State Physics of Russian Academy of Sciences, Chernogolovka.

[TUR 98] TURANOV A. N., KARANDASHEV V. K., BAULIN V. E. "Extraction of uran and thorium with neitral phosphoryl containing podands from nitric acid solutions", *Radiokhimiya (Rus.),* vol. 40, n° 1, 1998, p. 36-43.