Environmental Chemistry and North America's Great Lakes:

From phosphorus to QSAR

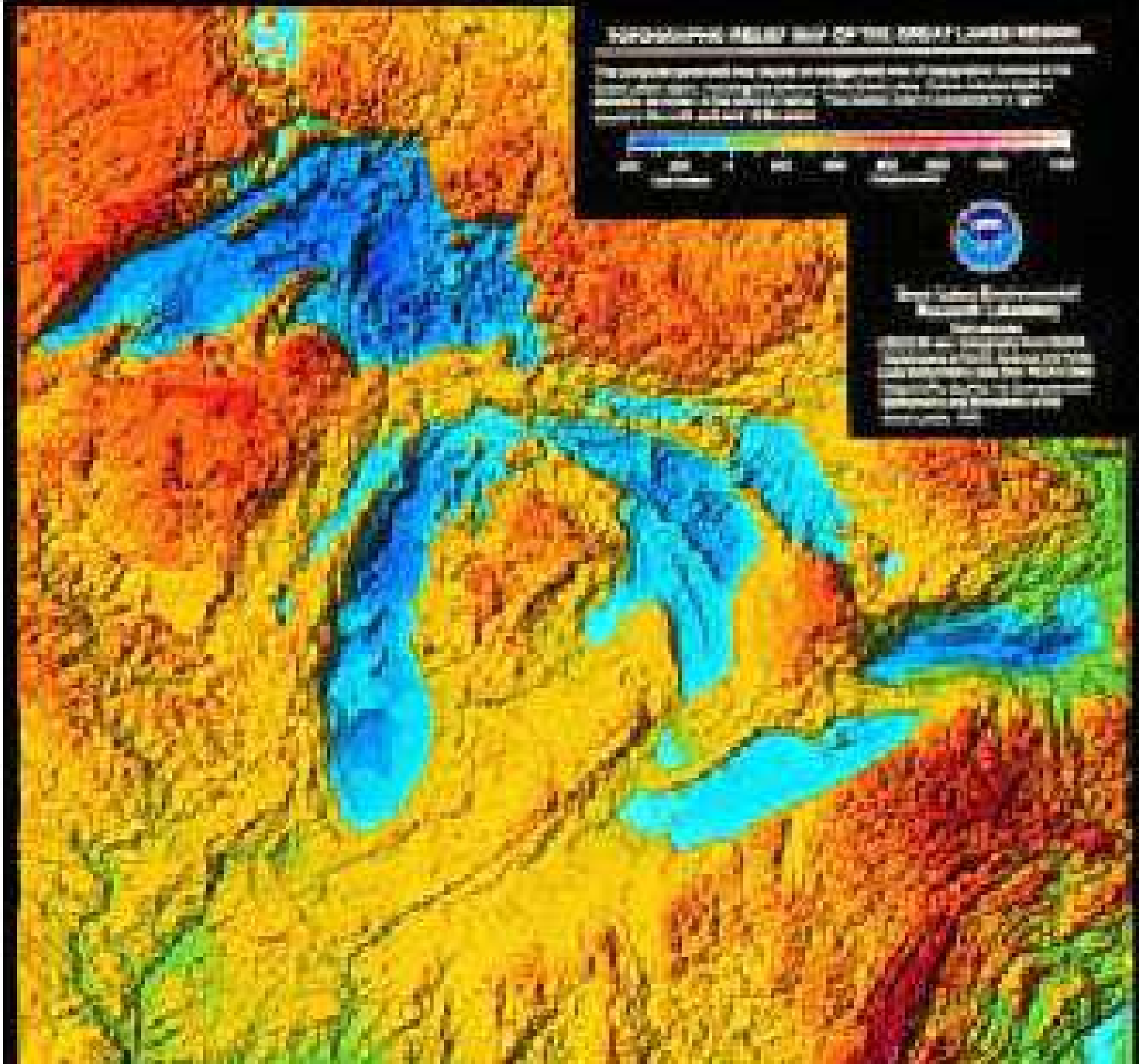
Klaus L.E. Kaiser

TerraBase Inc., Hamilton, Ontario
CANADA

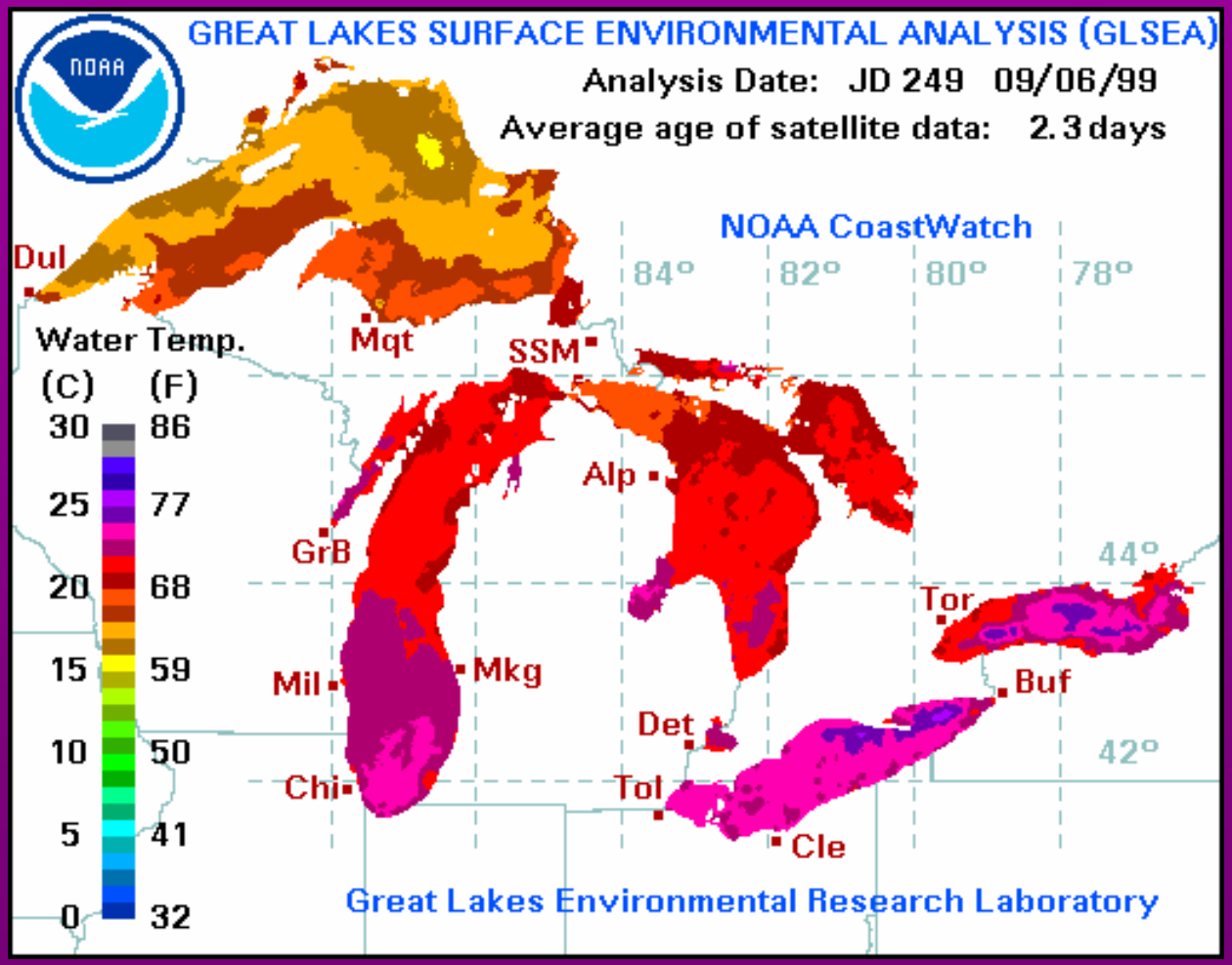
TerraBase Inc.

TerraBase Inc.

Great Lakes Water Depths



Summer Surface Temperatures



TerraBase Inc.

Great Lakes Surface Areas $(km)^2$

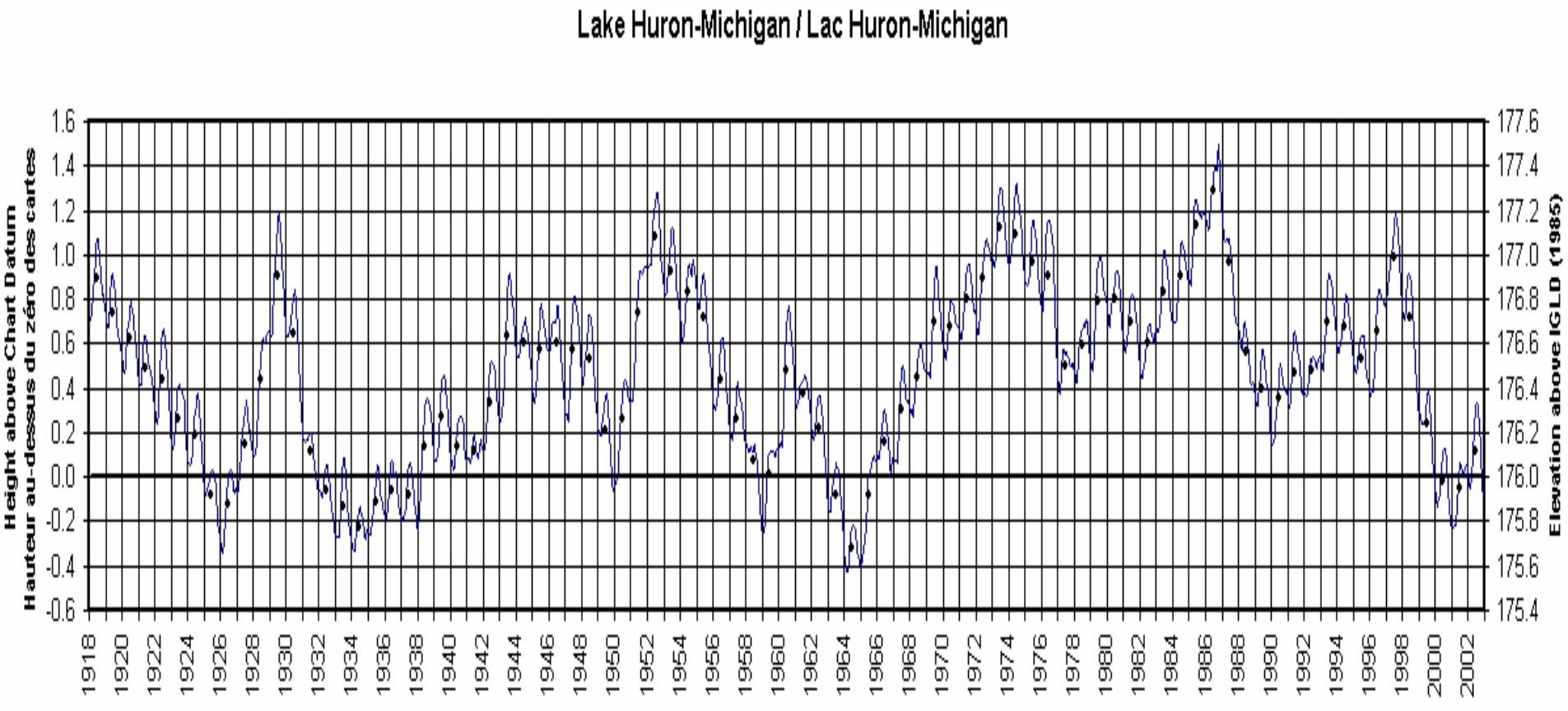
- **Superior** **85 000**
- **Michigan** **58 000**
- **Huron** **60 000**
- **St. Clair** **1 000**
- **Erie** **26 000**
- **Ontario** **19 000**

Water retention

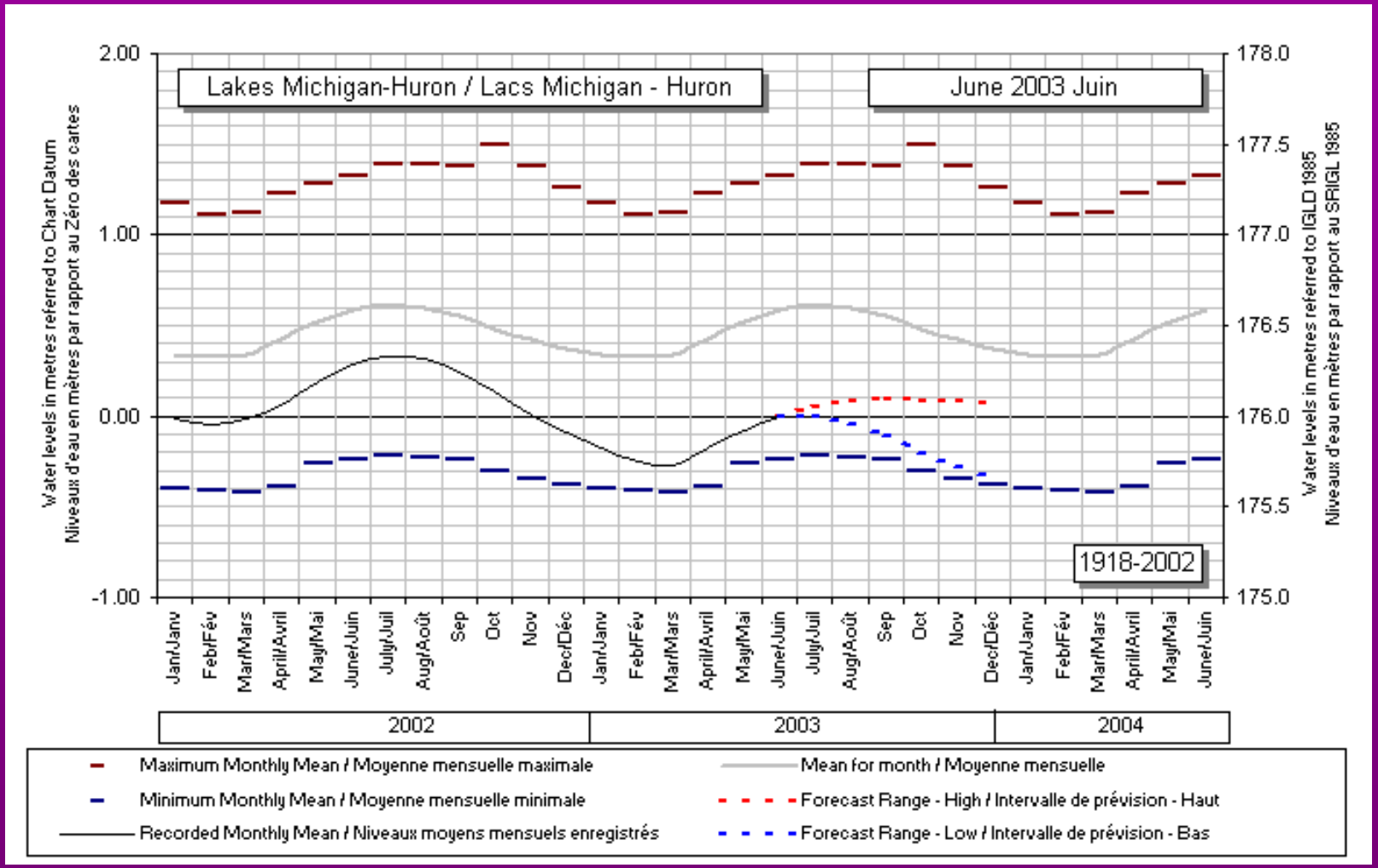
- L. Superior 191 years
- L. Michigan 99 years
- L. Huron 22 years
- Erie 3 years
- Ontario 6 years

Flow of the interconnecting rivers

- at ~2000 m^3 / sec
  and at ~100 000 sec / day
- volume ~200 000 000 m^3 / day
- or ~0.2 $(km)^3$ / day
- or ~70 $(km)^3$ / year



TerraBase Inc.



TerraBase Inc.

Lake Michigan / Lake Huron

Typical annual seasonal water level fluctuation:

0.6 m

Seasonal high water level: July

Seasonal low water level: January

TerraBase Inc.

Lake Michigan / Lake Huron

- Combined surface area: 120 000 $(km)^2$
- Assume zero in-flow, and
- 2000 m^3 / sec outflow for 6 months or 35 $(km)^2$ x 1000 m
- Resulting drop in water level: 35 / 120 = 0.35 m
- Question: What causes the winter seasonal drop of ~ 0.5 m ??

Lake Michigan / Lake Huron

- Question:
  What causes the large seasonal drop ??

- Answer:
  Evaporation

TerraBase Inc.

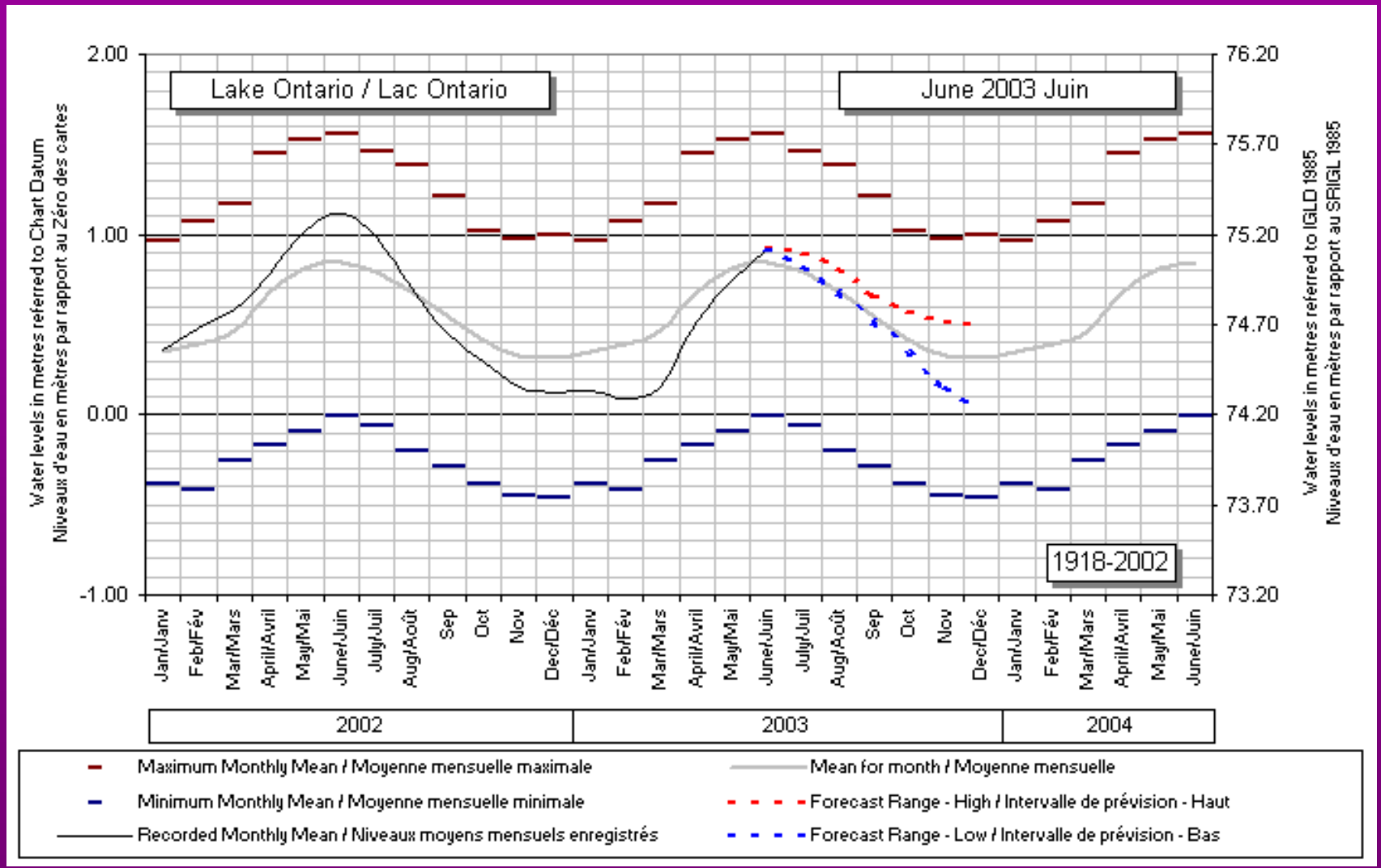
Lake Michigan / Lake Huron

Question: Why is the evaporation greater in winter than in summer ??

Answer: Vapor pressure gradient is greater

relative humidity in summer: 70- 90%
relative humidity in winter: 10-30%

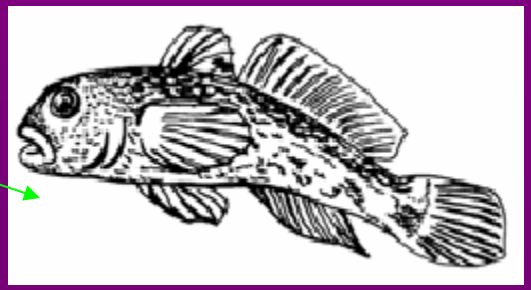
TerraBase Inc.



TerraBase Inc.

Invading species

- Sea lamprey (*Petromyzon*) ~ 1930
- Zebra mussel (*Dreissena*) ~ 1985
- Round gobi (*Neogobius*) ~ 1990
- Spiny water flea (*Bythotrephes*) ~ 1995
- Bighead carp (*Mylopharyngodon*) ~ 2003 (?)



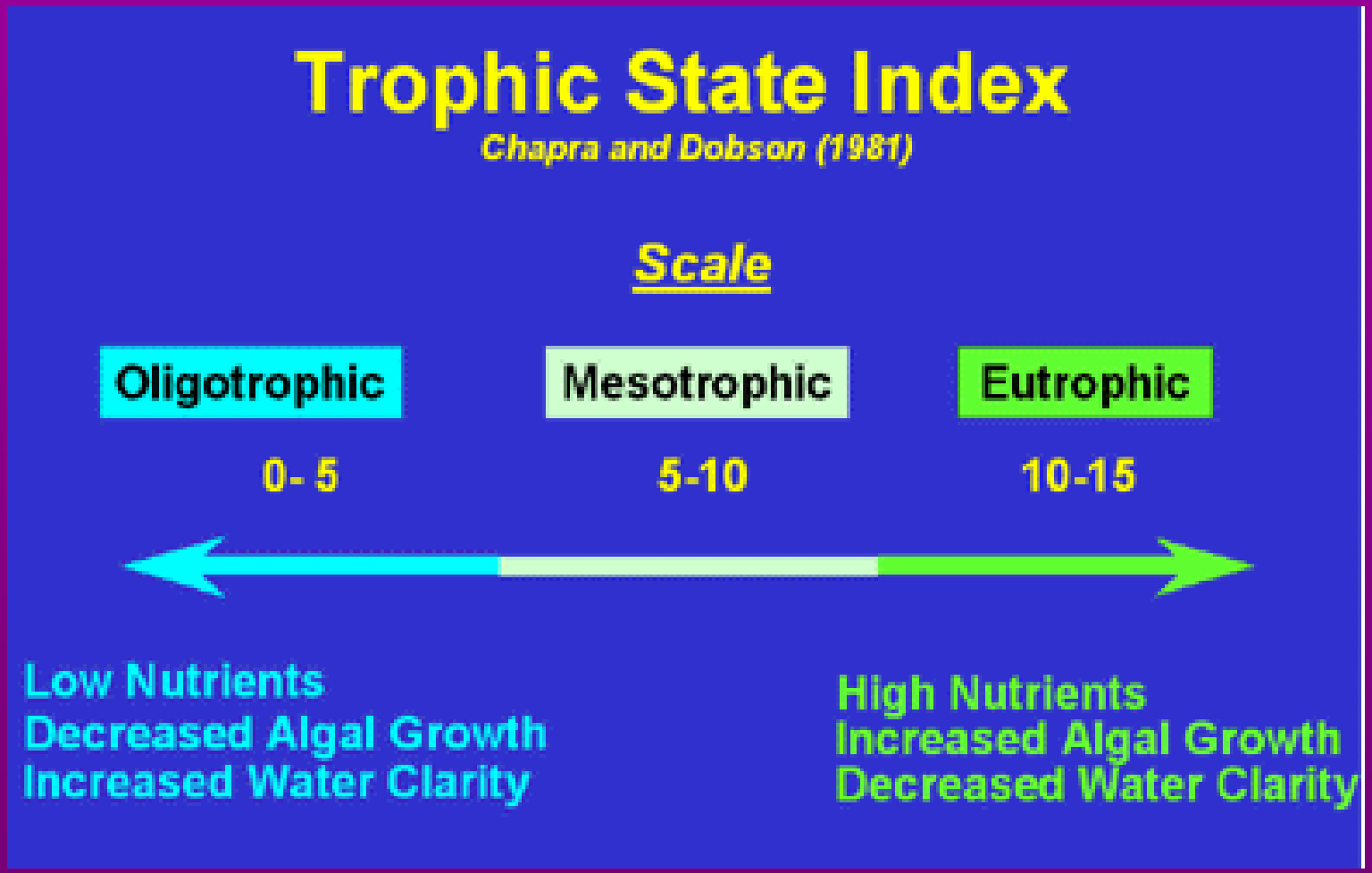
Lake Erie ~ 1970

- algae blooms (*Microcystis*)
- filamentous algae on shores (*Cladophora*)
- mayfly depletion (*Pontoporeia*)
- declining fish catches

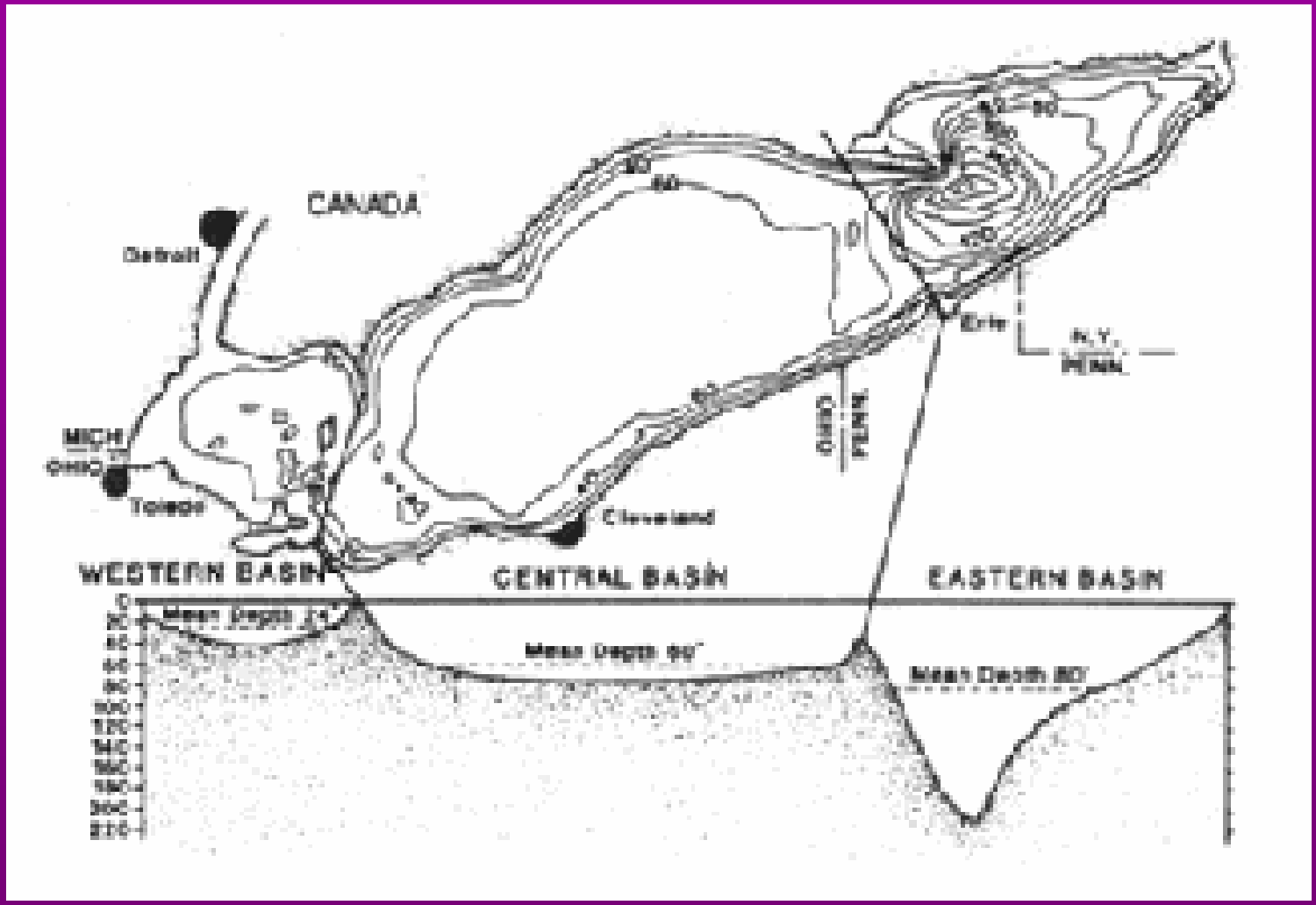
TerraBase Inc.

Trophic State Index

Chapra and Dobson (1981)



TerraBase Inc.



TerraBase Inc.

Lake Erie "dead", 1970

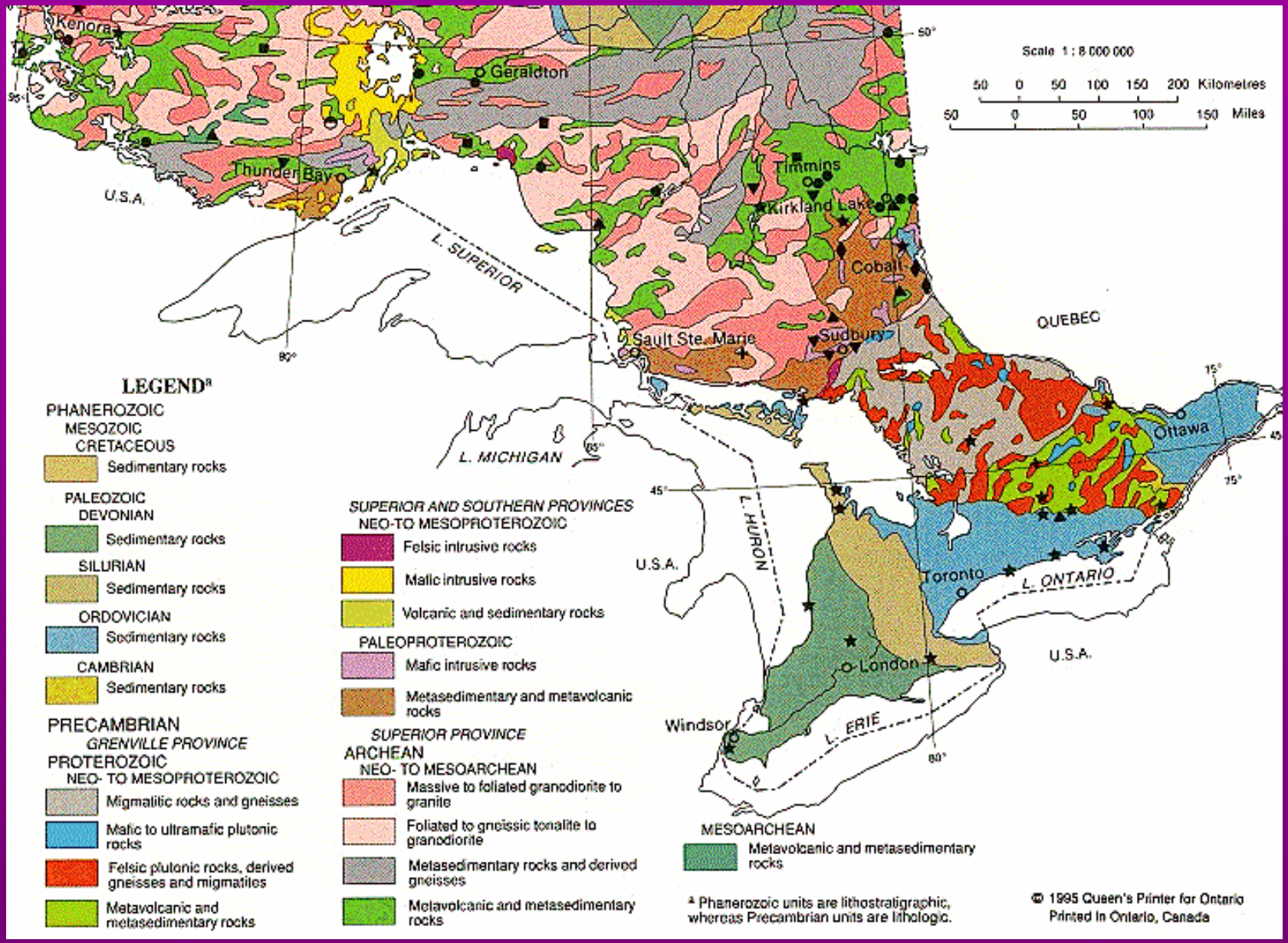
- Cause: Rapid eutrophication
- Solution: Reduction of phosphate loadings
- Effect: (~1990) healthy

TerraBase Inc.

TerraBase Inc.

Why are the lakes susceptible to invasive species & contaminants ?

- Answers:
- Very young (in geological terms)
- ~ 12 000 years since last ice age
- Relatively nutrient poor
- Many ecological niches un-occupied



TerraBase Inc.

Some historic facts

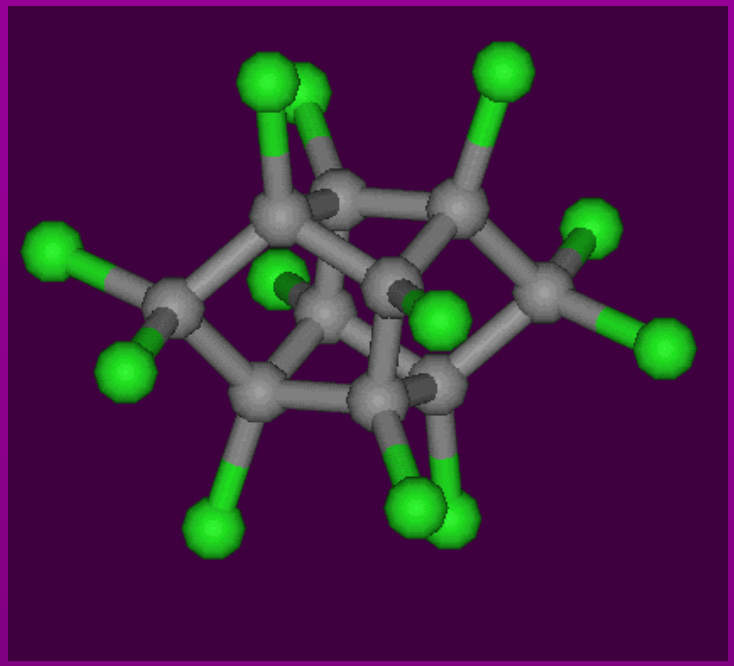
- Niagara River power generation: 1920s
- Electricity-using industry
- Diels-Alder diene synthesis: 1930s
- Chemical industry: electrolysis of NaCl:
  - chlorine generation
- chlorination: hexachlorocyclopentadiene
- Diels-Alder diene/chlorine products in 1940s: aldrin, heptachlor, mirex

Lake Ontario ~ 1970

- Contaminants in eel (*Anguilla*):
  - Polychlorobiphenyls: ~ 50 ppm
  - Mirex: ~10 ppm
  - Polychloronaphthalenes
  - Polychlorophenanthrenes
  - Polychlorostyrenes
  - Polychloroalkanes

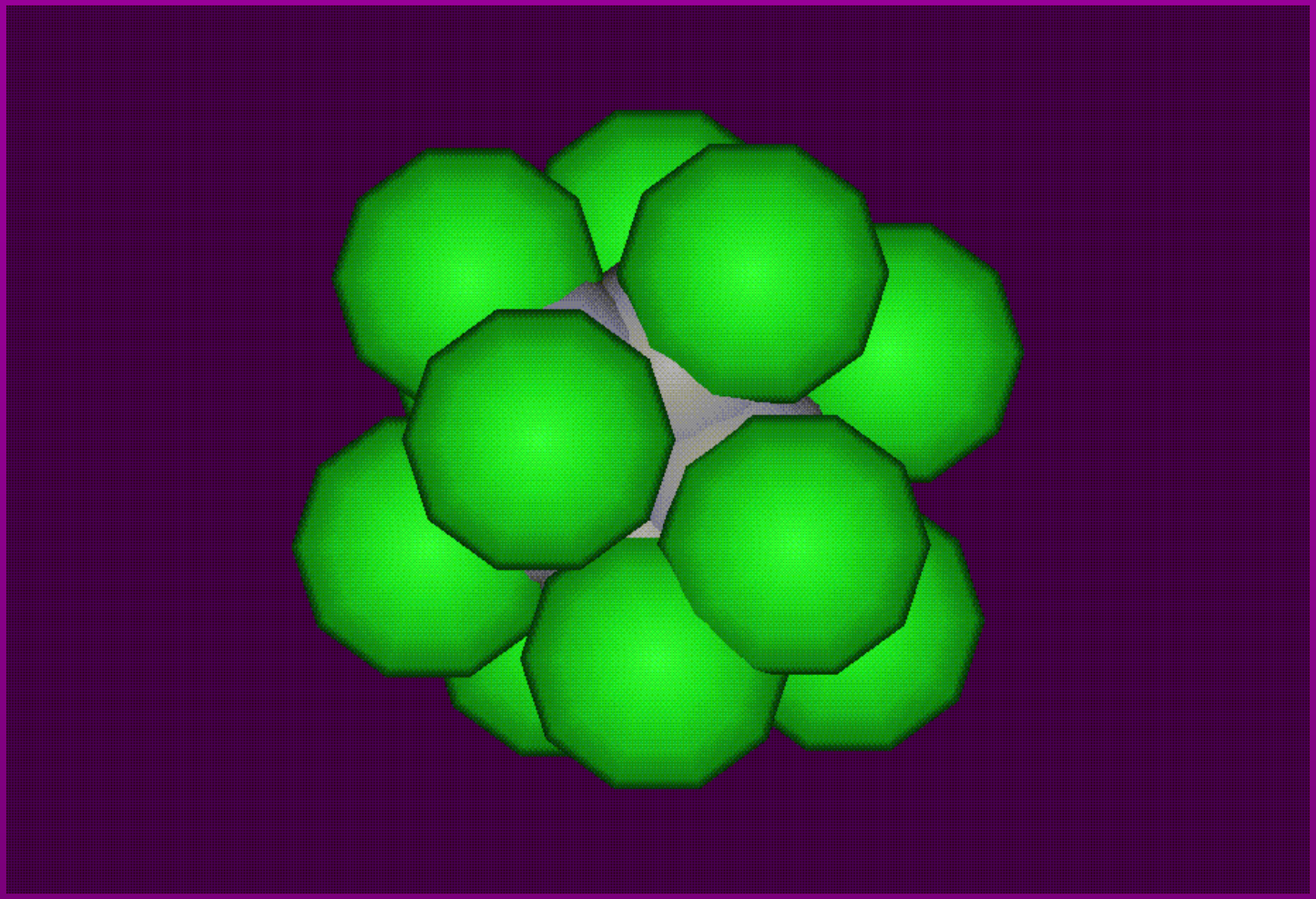
Mirex

- $C_{10}Cl_{12}$
- Mirex: insecticide, flame retardant
- 1973: Mirex in Lake Ontario fish
- 1976: ban in Canada, US



Mirex

Use:
flame retard.
insecticide



What are the effects of contaminants ?

We don't know

- Very few measured values of any sort
- 2 500 OECD-HPV chemicals: some values
- 100 000 chemicals in commerce: very few values
- Need for estimation of effects
- Quantitative Structure-Activity Relationships (QSAR)

QSAR

Important since ~ 1960 for development of pharmaceuticals, pesticides, etc.

1983: 1st Intl. Workshop on QSAR in Environmental Toxicology (Hamilton, Ont.)

1986: 2nd Intl. Workshop on QSAR in Environmental Toxicology (Hamilton, Ont.)

Since then: Workshops in 1988, 1990, 1992, 1994, 1996, 1998, 2000, 2002, 2004 (Liverpool)

TerraQSAR™

stand-alone

probabilistic neural network (PNN) based

toxicity/effect

computation programs

for PCs

TerraBase Inc.

TerraQSAR™ modules

TerraQSAR™ - FHM:

96-hr LC50 for fathead minnow

TerraQSAR™ - RMIV:

iv. LD50 for rat & mouse

TerraQSAR™ - E2-RBA:

estrogen receptor binding affinity (RBA) relative to 17*beta*-estradiol

Input

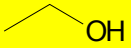
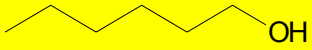
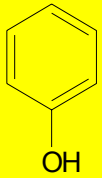
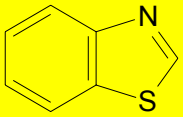
TerraQSAR modules use as input only the chemical's SMILES code (2-D or 3-D), which is an international code for the representation of chemical structures and amenable to computer analysis.

SMILES = acronym for
Simplified **M**olecular **L**ine **E**ntry **S**ystem

Output

The **TerraQSAR™** programs typically compute the **LC50, LD50, or EC50** in both mg/L or mg/kg b.w. (body weight) and in pT (log [L/mmole] or log ([kg b.w.]/mmole) units, as well as the molecular weight (MW) of substances entered, directly from the compound's structure.

SMILES string examples and structures (hydrogen atoms omitted)

ethanol	**CCO**	
n-hexanol	**CCCCCCO**	
phenol	**c1ccccc1O**	
benzothiazole	**c1cccc2ncsc12**	

Fragments

Fragments used in the **TerraQSAR** modules have been described in detail in several publications listed in the literature, especially in the works by Kaiser *et al.* An overview of basic fragment types considered is given in **Table 1**.

Table 1. Fragment types used in TerraQSAR

Fragment Types	Examples
Acidity fragment	C(=O)O, S(=O)(=O)O
Aliphatic ring fragment	C1CCCCC1, C1CCCC1
Aromatic ring fragment	c1ccccc1, c1ccccn1
Atom fragment	C, H, N, O
Bond fragment	C-C, C=C, C#C
Group fragment	C-O-H, C-O-C, O=C-O-C
Hydrophobicity fragment	C(C)(C)C, CCCC
Ionization fragment	[O-], [Na+]
Polarity fragment	O=N(=O)CC(O)
Reactivity fragment	C=CC=O
Stereo fragment	Cl[C@H](C)N, Cl[C@@H](C)N
Weight fragment	molecular weight

Computation

The computer evaluates the number and types of bonds and fragments present in the compound and computes the toxicity estimate on the basis of the same types of bonds and fragments present in the training data set.

Computation time varies with the complexity of the query structure and speed of the computer. Typically, for compounds without chiral centers, and molecular weights of <200, computation time on a 2 GHz machine takes <5 seconds.

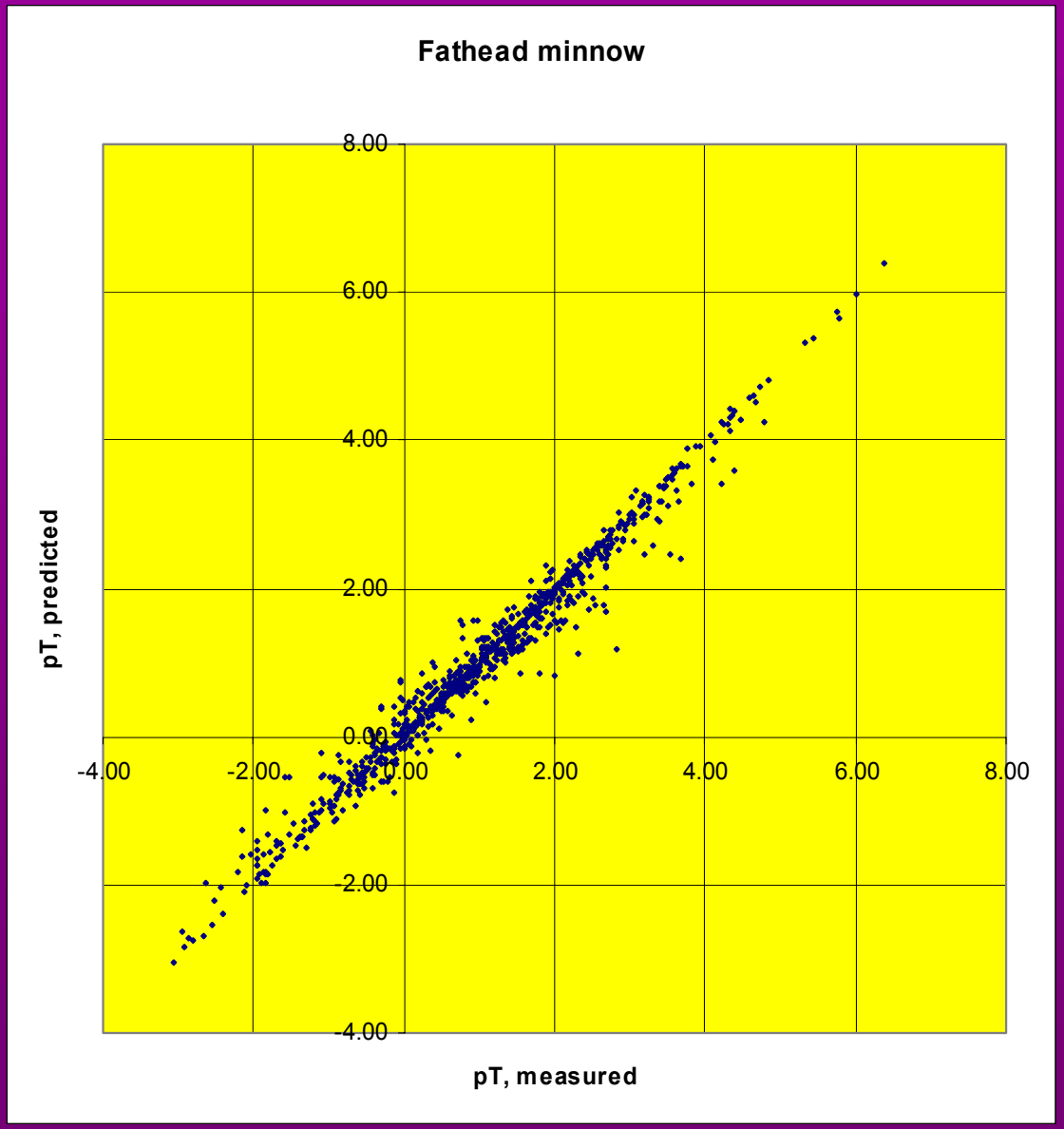
Results

Figure 1 shows the measured vs. predicted values for all 886 compounds fathead minnow values used in the development of the **TerraQSAR - FHM** estimation program, as obtained from the program.

The data cover approximately ten orders of magnitude, ranging from pT= -3 to pT= 7, where pT is the negative logarithm of the millimolar LC50 concentration, pT = log (L/mmol).

Figure 1

FHM



TerraBase Inc.

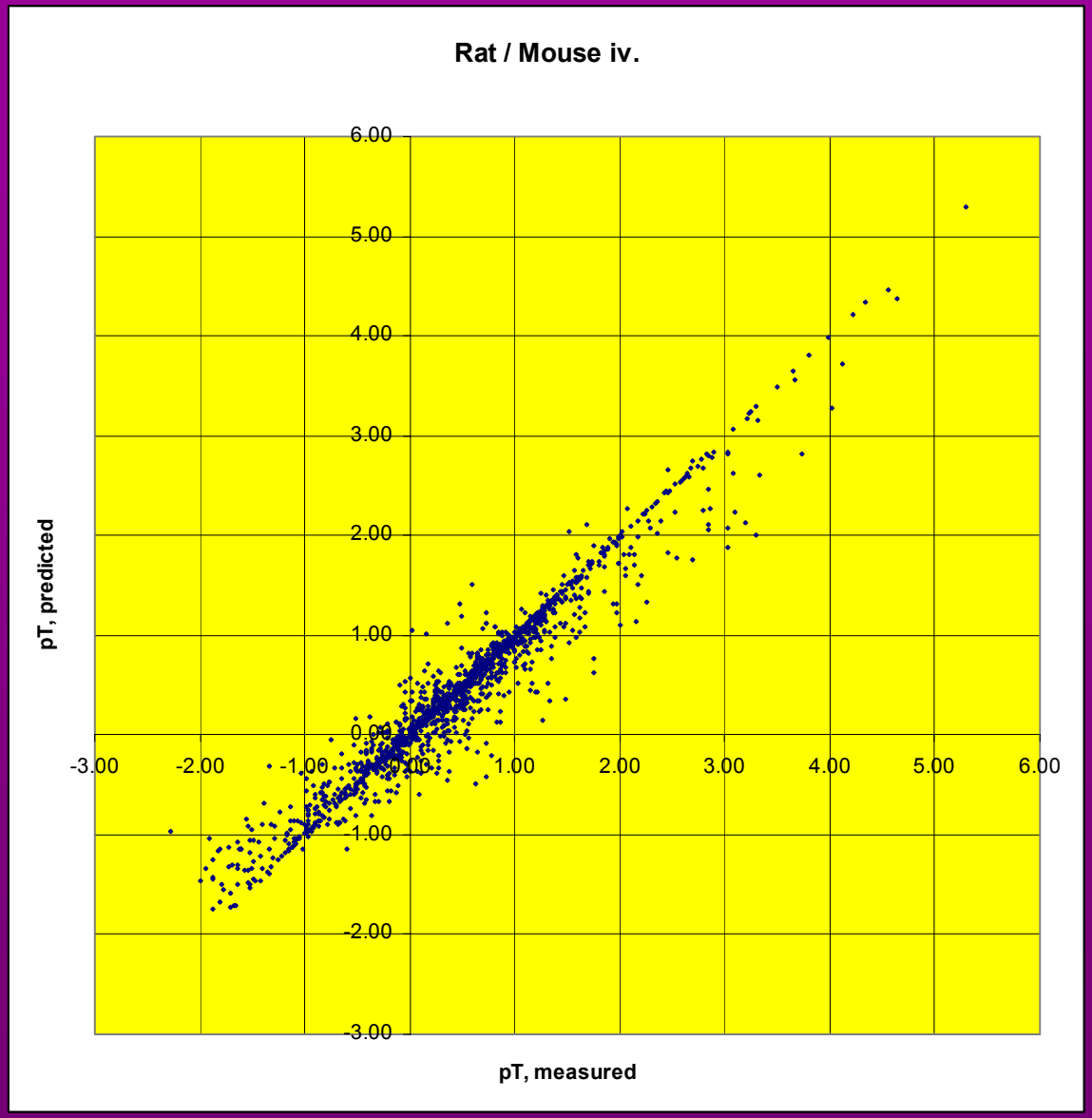
Table 3. Statistcal indicators for TerraQSAR - FHM

Program Module	Indicator	Variable
FHM	Number of compounds	886
	Range (log units)	10.5
	Correl. coeff. (r^2)	0.979
	Slope	1.064
	Intercept	-0.061
	RMS* error (leave-out 33%)	0.192
	RMS error (full training set)	0.063

* RMS: root mean square error

Figure 2

RMIV



TerraBase Inc.

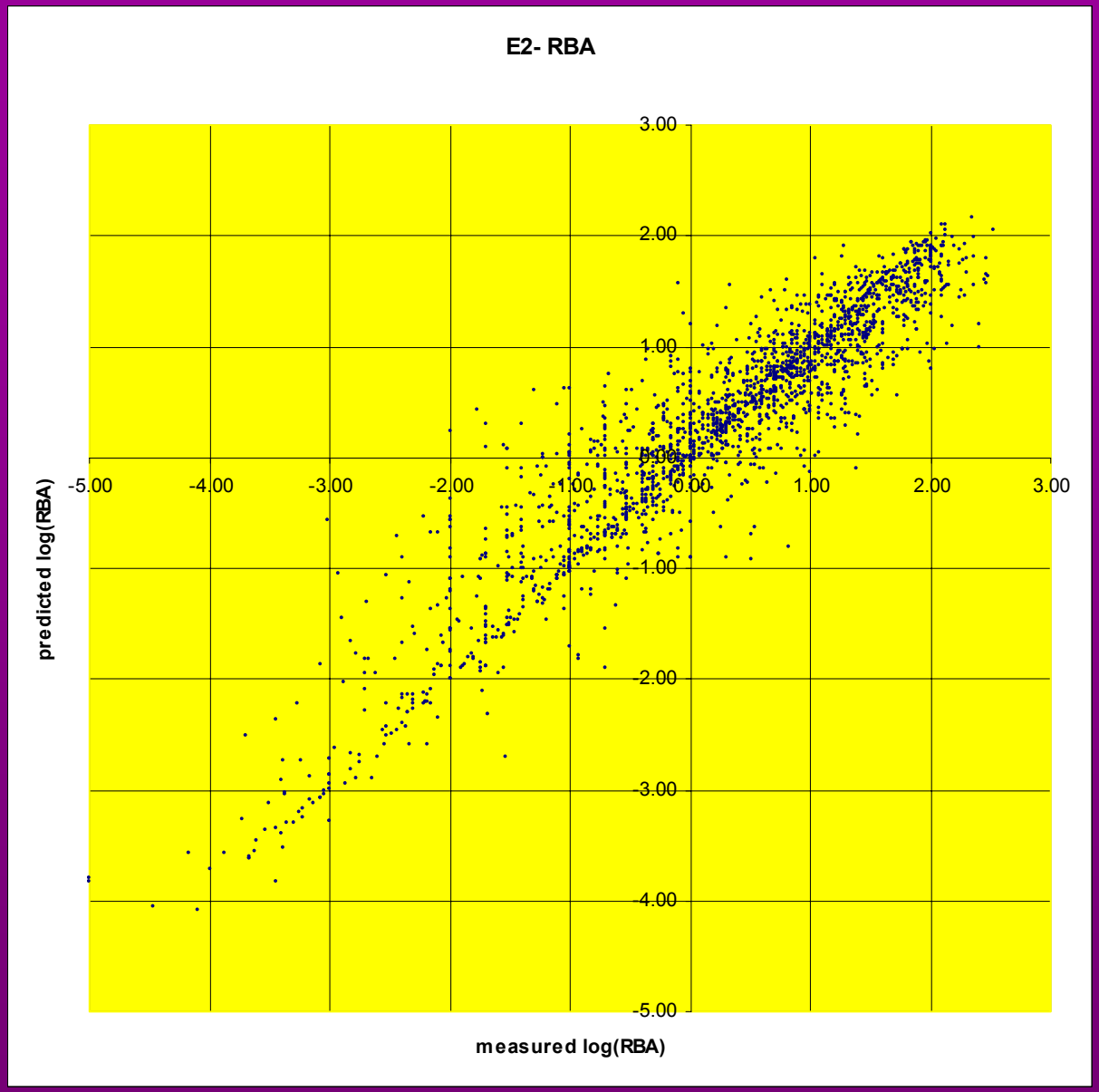
Table 4. Statistical indicators for TerraQSAR - RMIV

Program Module	Indicator	Variable
RMIV	Number of compounds	1497
	Range (log units)	7.5
	Correl. coeff. (r^2)	0.969
	Slope	1.057
	Intercept	0.005
	RMS* error (leave-out 33%)	0.199
	RMS error (full training set)	0.076

* RMS: root mean square error

Figure 3

E2-RBA



TerraBase Inc.

Table 5. Statistical indicators for TerraQSAR – E2-RBA

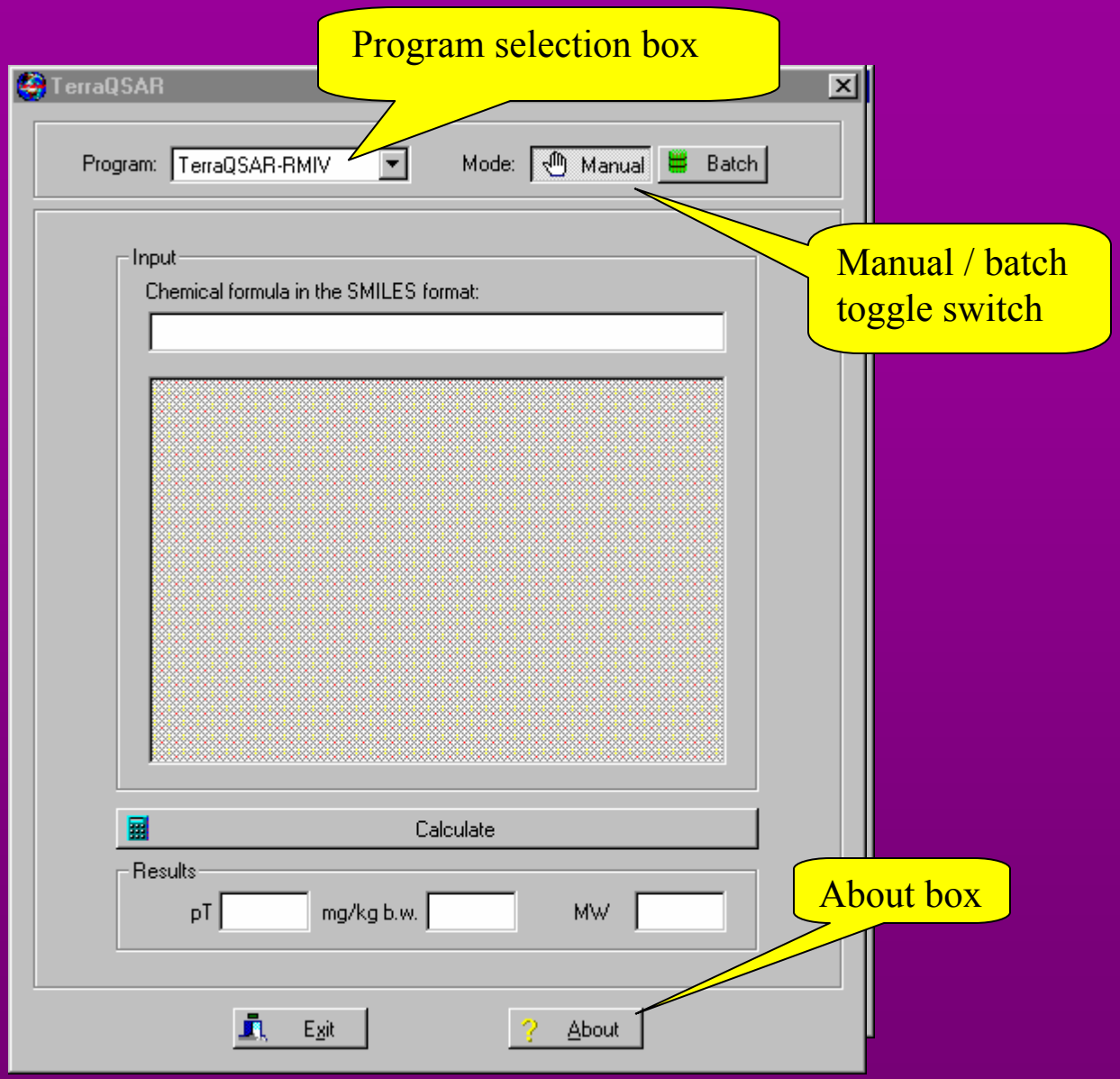
Program Module	Indicator	Variable
E2-RBA	Number of compounds	> 2000
	Range (log units)	7.5
	Correl. coeff. (r^2)	0.935
	Slope	1.078
	Intercept	-0.064
	RMS* error (leave-out 33%)	0.238
	RMS error (full training set)	0.121

* RMS: root mean square error

Program Interface

The program interface of the **TerraQSAR** toxicity prediction modules is shown in the next slide.

It is simple,
intuitive, and
highly functional.



TerraBase Inc.

Example 1:

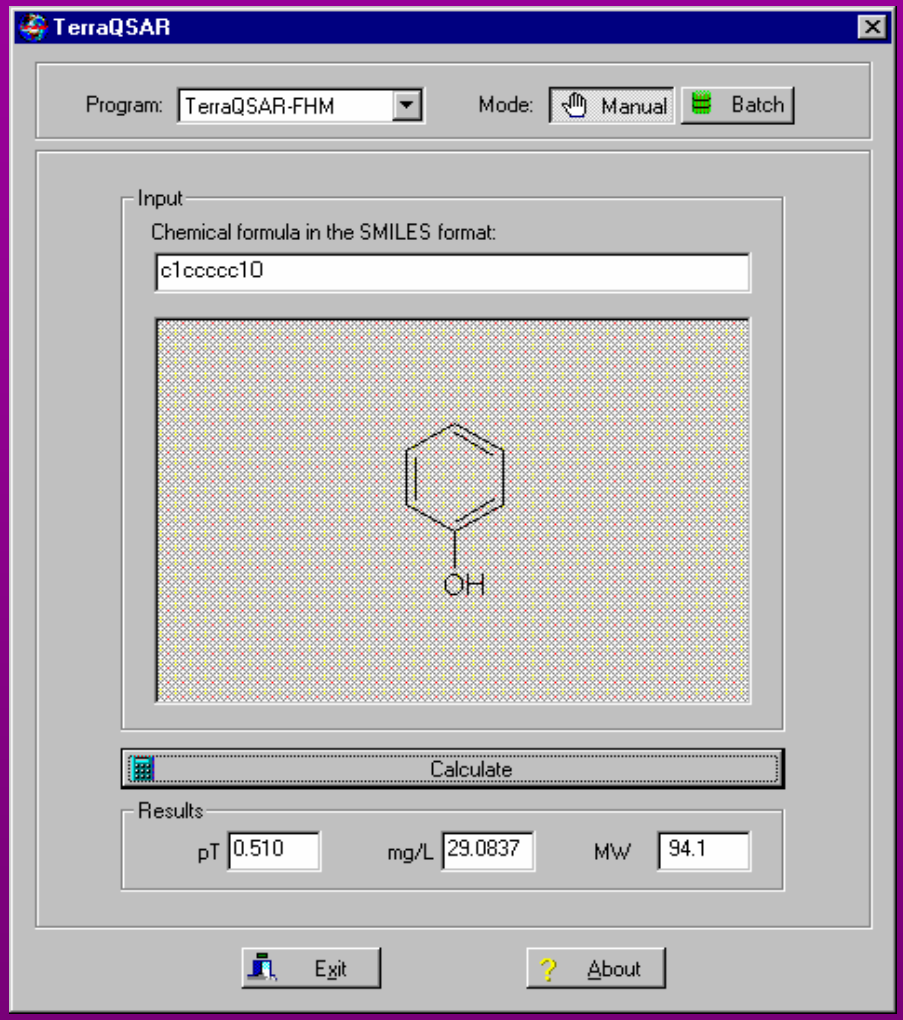
Phenol has the SMILES string "c1ccccc1O". Copying this string into memory, for example from this text (making sure the quotation marks are omitted), and pasting it into the Input field will result in the appearance of the chemical structure of phenol in the shaded, rectangular field below, as shown in the next slide.

Once the user has ascertained that the structure of the compound is that of the desired chemical, a simple click of the Calculate bar below the structure field will result in the three fields below the bar to be filled with the predicted values for the compound, as shown in the next slide.

Field 1 (pT) is the - log LC50 [mmol/L];

Field 2 (mg/L) is the LC50 in mg/L (FHM module);

Field 3 (MW) shows the molecular weight of the compound.



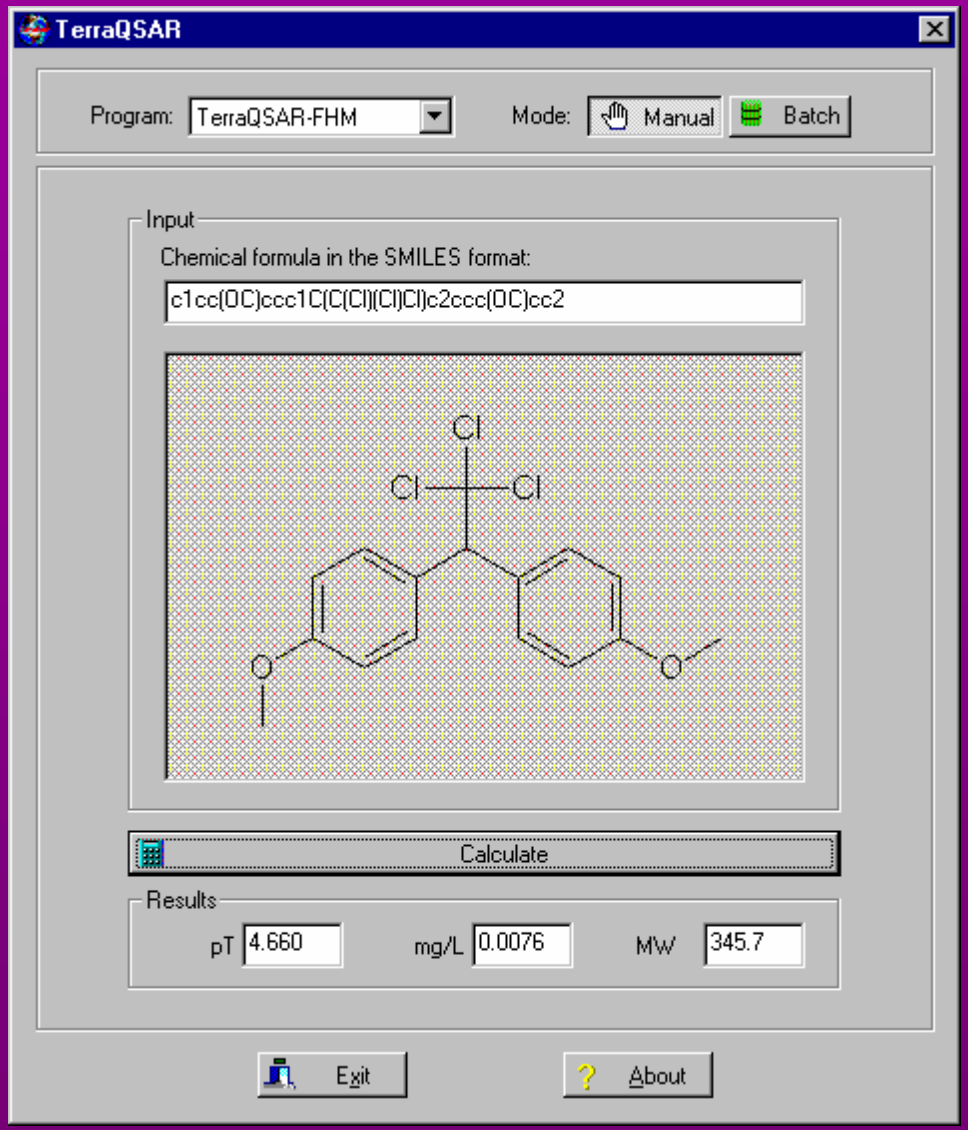
TerraBase Inc.

Examples 2 - 4:

The following examples of more complex molecules demonstrate the versatility of the TerraQSAR - FHM program:

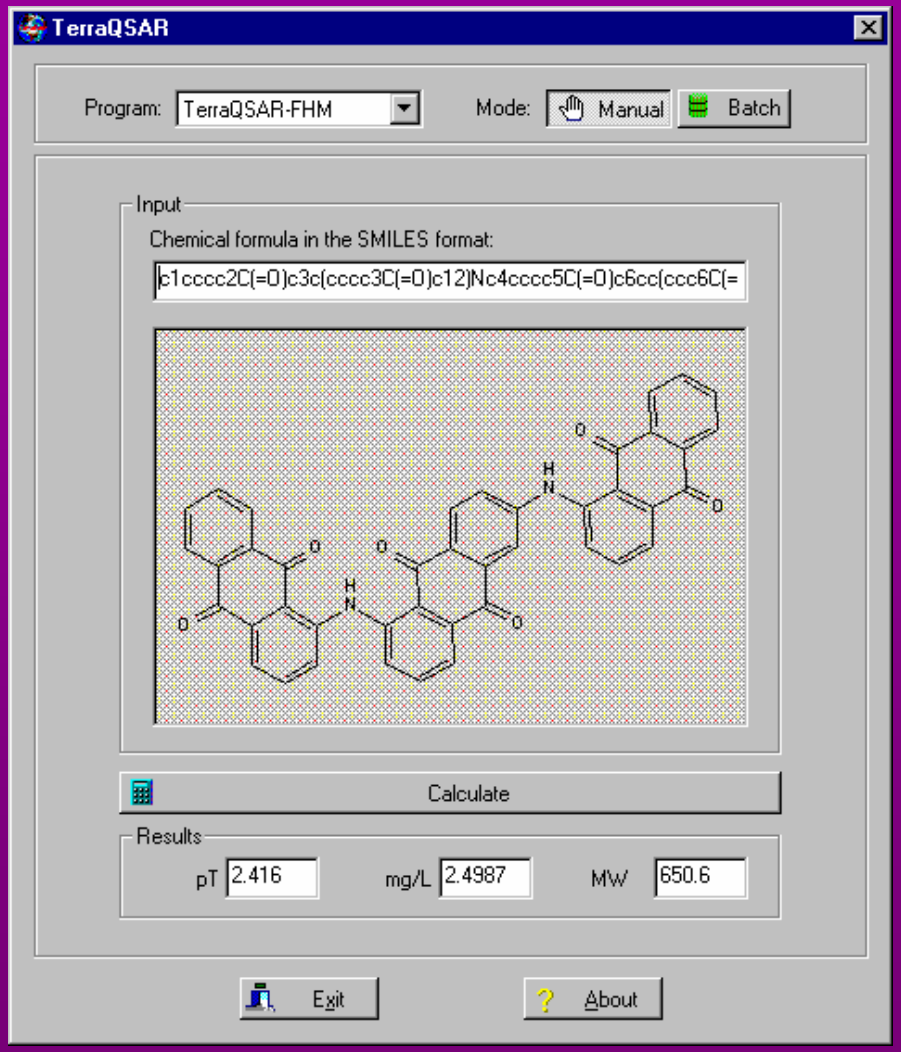
- Methoxychlor (pesticide)
- anthraquinone derivative (dye)
- Erythromycin (antibiotic)

FHM: Methoxychlor (insecticide)



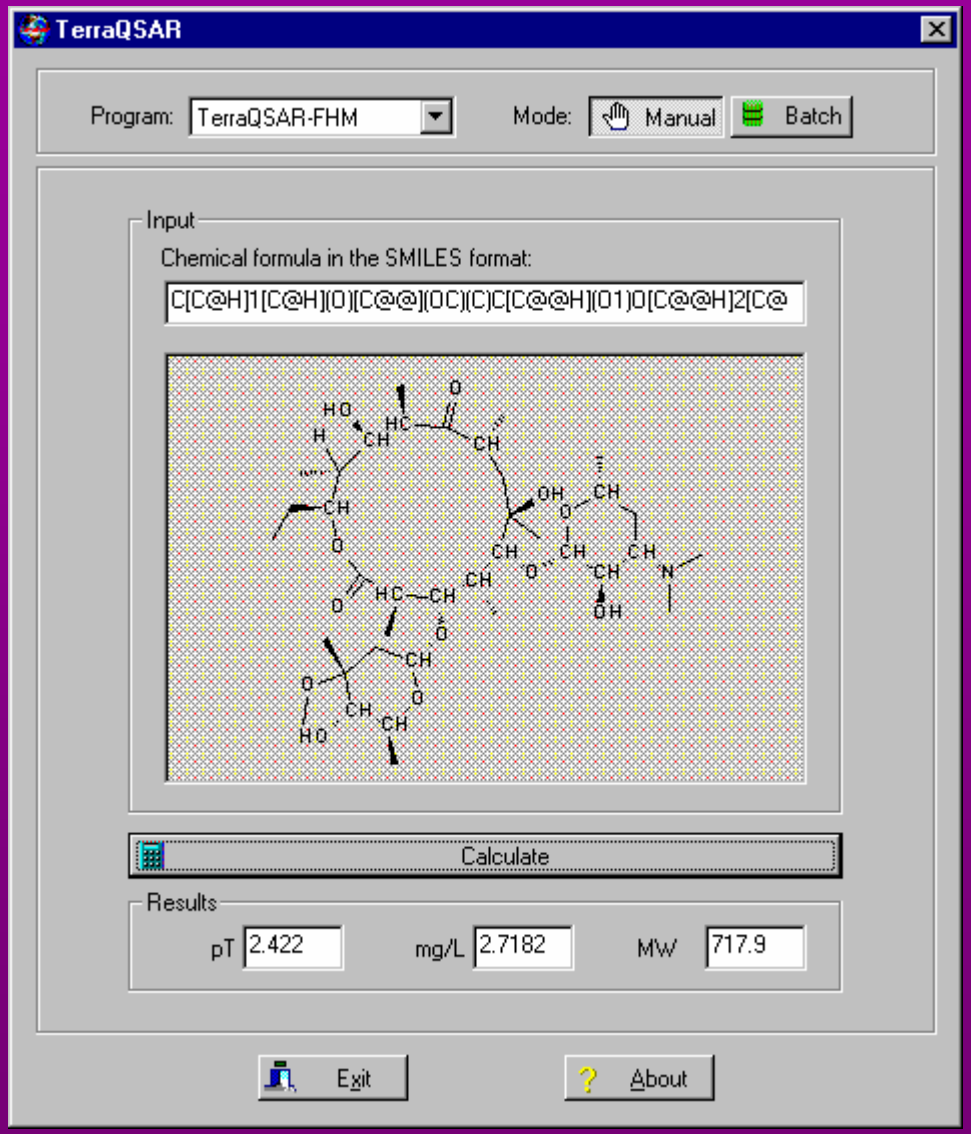
TerraBase Inc.

FHM: anthraquinone dye (dye / pigment)



TerraBase Inc.

FHM: Erythromycin (antibiotic)



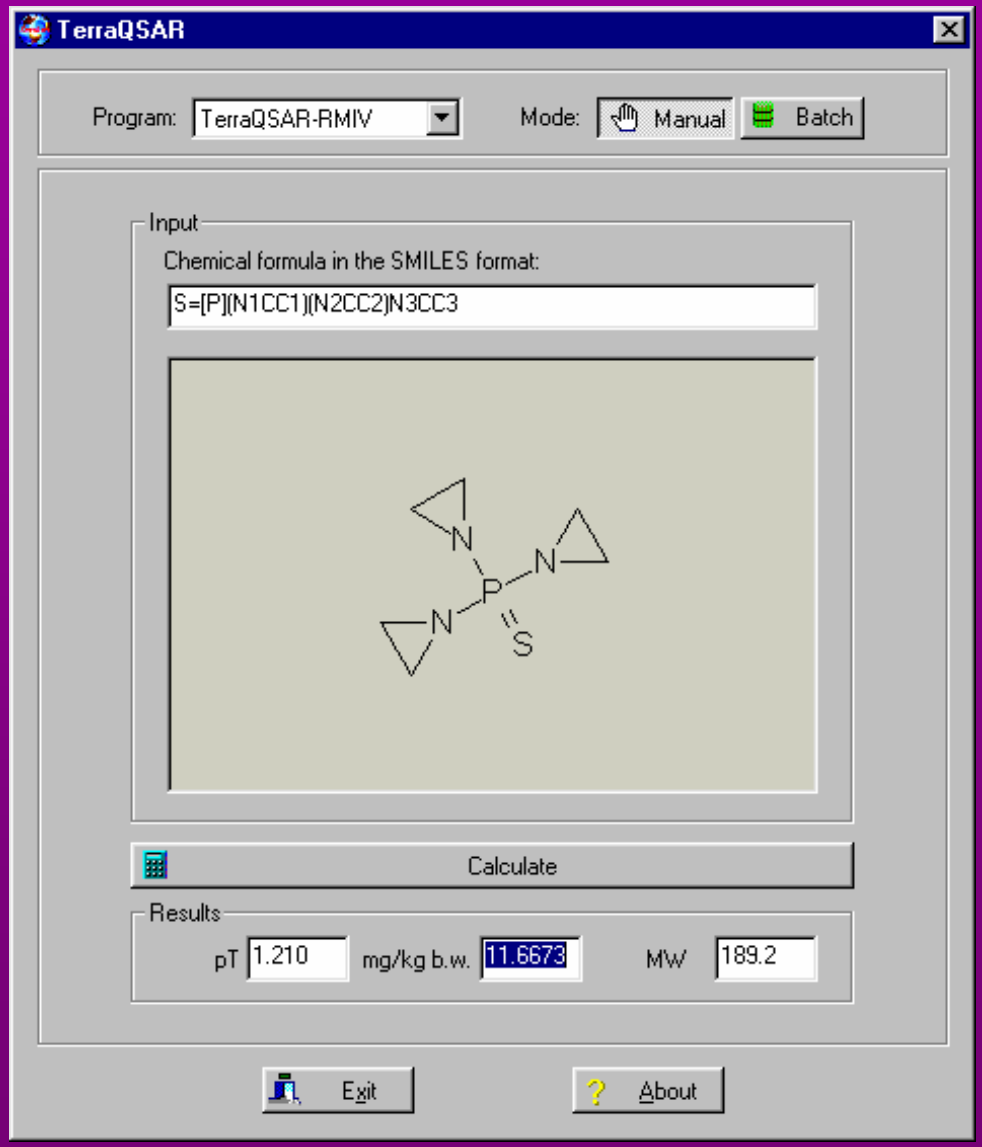
TerraBase Inc.

Examples 5-8:

The following examples of more complex molecules demonstrate the versatility of the TerraQSAR - FHM program:

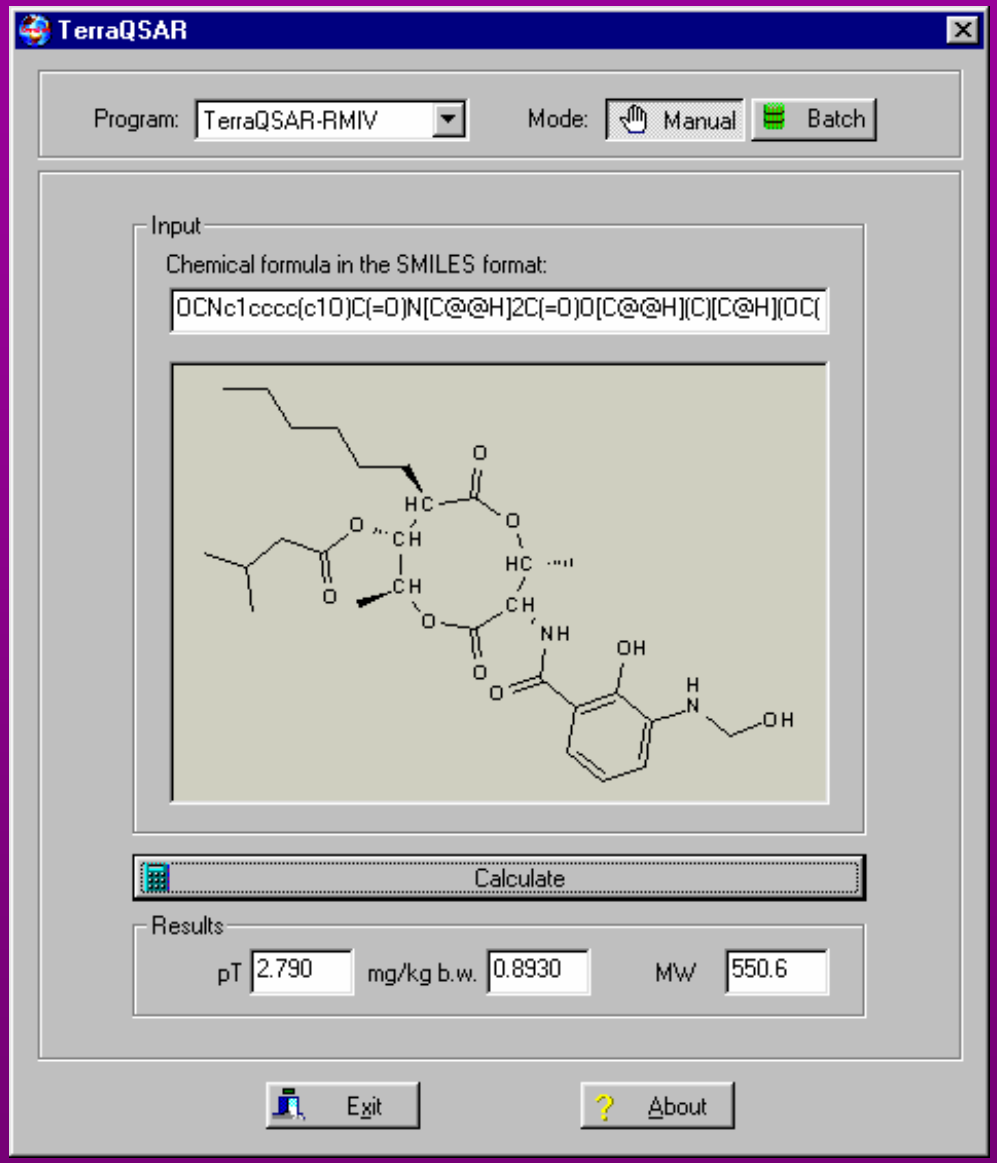
- Thiotepa (chemosterilant)
- Antimycin A (antibiotic)
- Alkuronium dichloride (neuromuscular agent)
- Tetrodotoxin (puffer fish toxin)

RMIV: Thiotepa (chemosterilant)



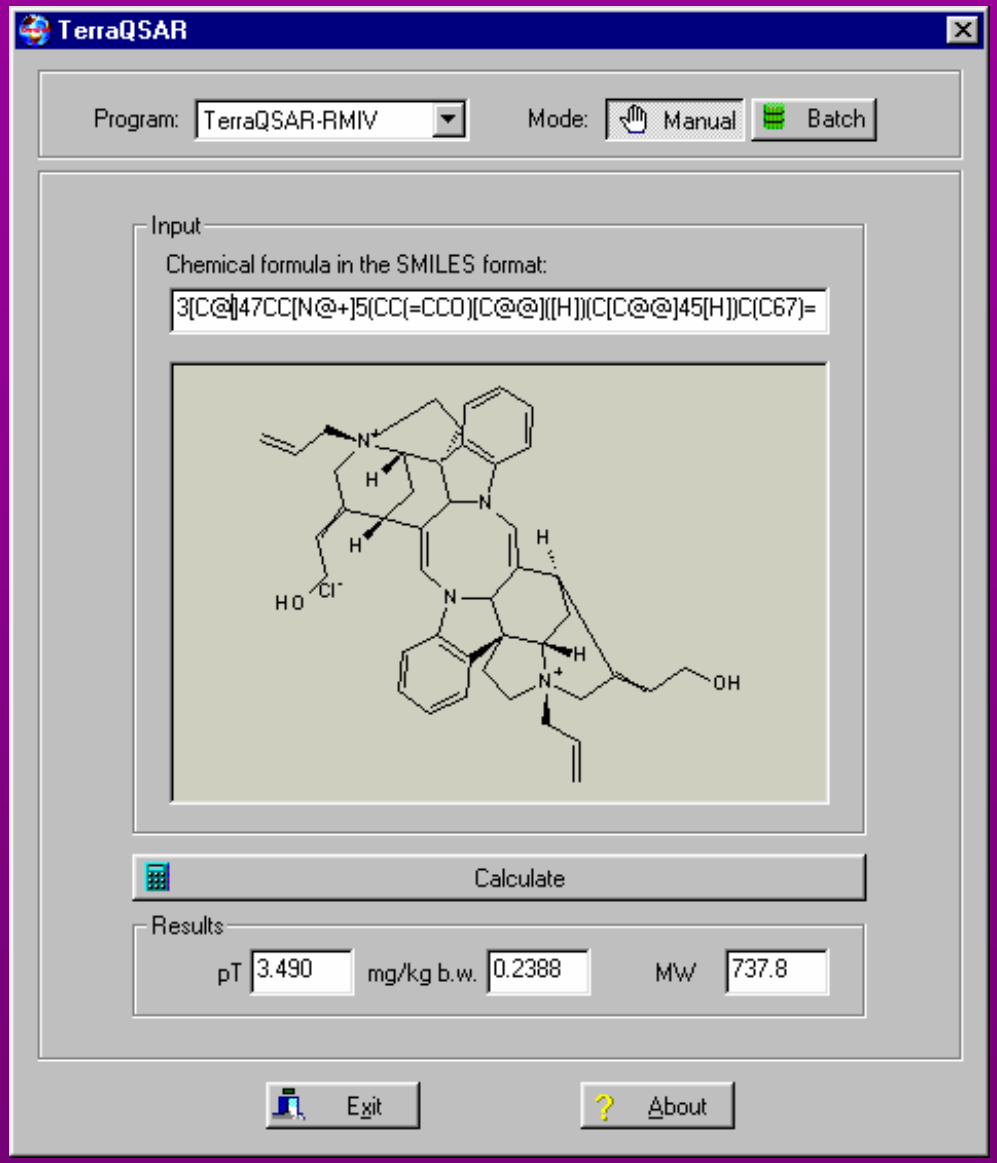
TerraBase Inc.

RMIV: Antimycin A (antibiotic)



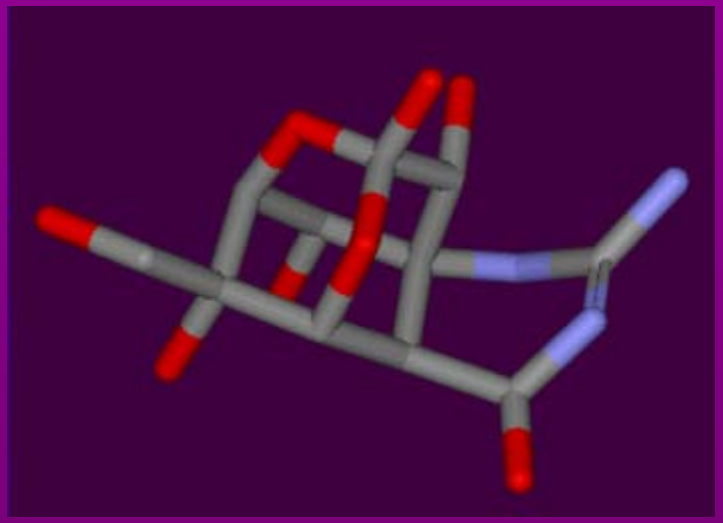
TerraBase Inc.

RMIV:
Alcuronium dichloride (neuromuscular blocking agent)



TerraBase Inc.

RMIV:
Tetrodotoxin
(puffer fish toxin)



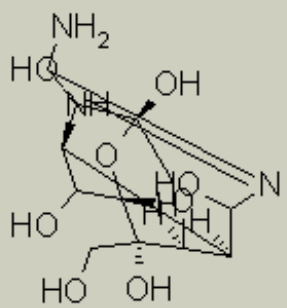
TerraQSAR

Program: TerraQSAR-RMIV Mode: Manual Batch

Input

Chemical formula in the SMILES format:

[C@@]23([H])[C@@]([C@]4([H])[C@]1([H])C(O)N=C(N)N[C@@]1(



Calculate

Results

pT 4.640 mg/kg b.w. 0.0073 MW 319.3

Exit About

TerraBase Inc.

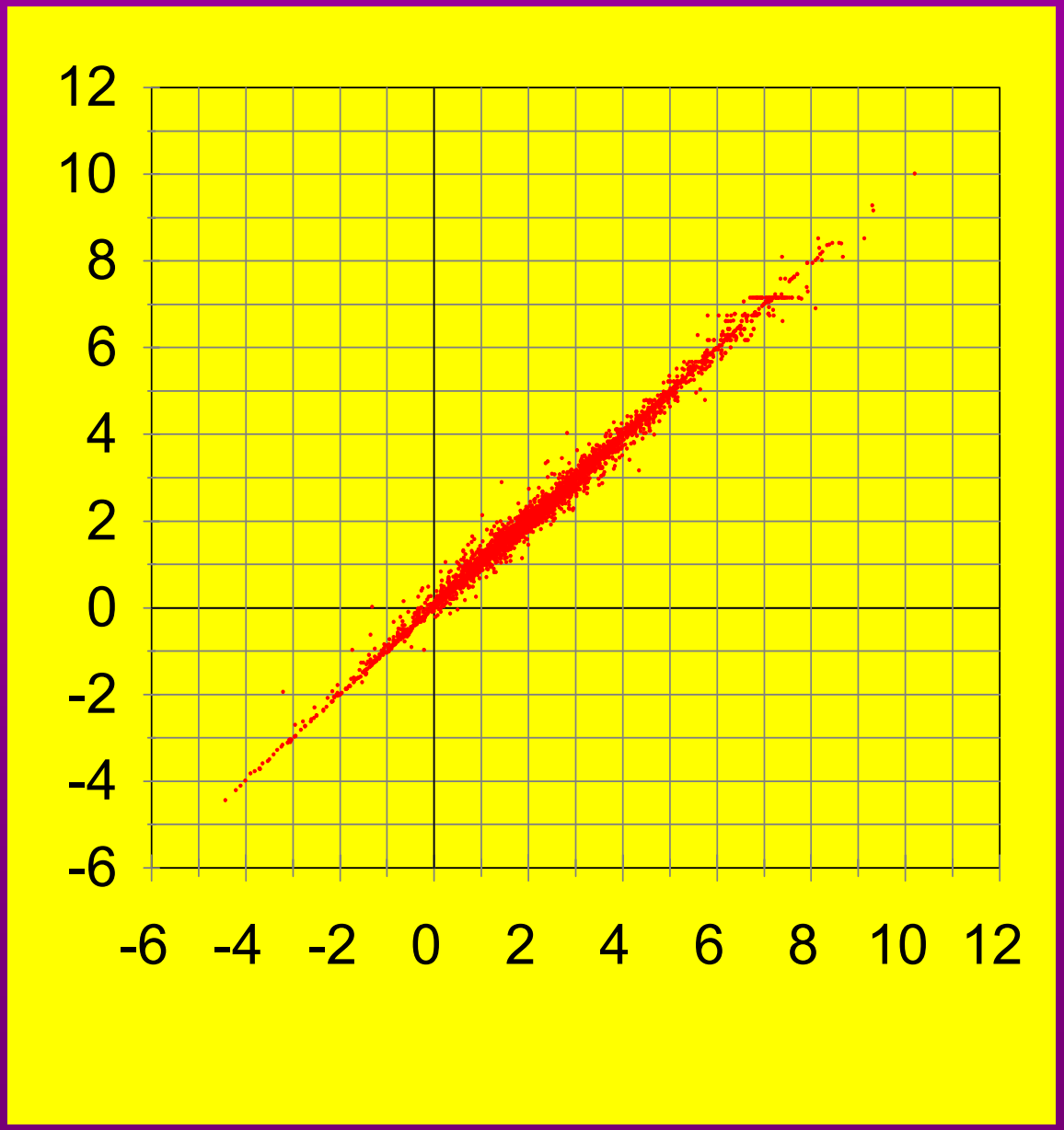
logP / log K_{ow}

(octanol / water partition coeff.)

4000 + data

14 orders of magnitude

(in preparation)



TerraBase Inc.

Thank you for your attention

Internet: http://www.terrabase.ca

TerraBase Inc.