# Molecular Descriptors Family on QSAR Modeling of Quinoline-based Compounds Biological Activities

Lorentz JÄNTSCHI, http://lori.academicdirect.org
Sorana BOLBOACĂ, http://sorana.academicdirect.ro

**The 10th Electronic Computational Chemistry Conference (ECCC10), April 1-30, 2005**
**held at http://eccc.monmouth.edu**

# Motivation

- Structure – activity relationships allow explaining of biological activities and can suggest the synthesizing of new active compounds.

# Idea

- The idea is to create a unitary approach, based on a minimal set of well-known truths, capable to generate an efficient model of activity behavior depending on molecular structure.

**The 10th Electronic Computational Chemistry Conference (ECCC10), April 1-30, 2005**
**held at http://eccc.monmouth.edu**

# Material

- A set of 15 Quinoline based compounds was taken into study.
- Quinolines, heterocyclic compounds, are an aromatic nitrogen compounds characterized by a double-ring structure contains a benzene and a pyridine, with antiseptic, antipyretic, and antiperiodic properties, widely used as a parent compound to make drugs.
- Mutagenicity and cytotoxicity of quinolines in Salmonella typhimurium was used as investigated activities.

# Method

- A new original set of molecular descriptors, called Molecular Descriptors Family (MDF) was used to make the Quantitative Structure - Activity Relationship study.
- The MDF is of pure structural nature and take into account both geometrical and topological model of molecules. The MDF use sets of atomic properties, distance metrics, interaction descriptors, overlapping descriptors methods, molecular fragmentation criterions, overall fragmental descriptors superposing methods, and linearization procedures in order to produce a number of 787968 MDF members with different calculation formulas.

# The MDF

It use sets of object oriented properties and methods as follows:

- Atomic property (p) can be one of: M (mass), Q(charge), C(cardinality), E(electronegativity), G(group electronegativity);
- Distance metric (d) can be one of: t(topological) and g(geometrical);
- Interaction descriptor (implies two participants) can be one of: D(d), d(1/d), O($p_1$), o(1/$p_1$), P($p_1p_2$), p(1/$p_1p_2$), Q($\sqrt{p_1p_2}$), q(1/$\sqrt{p_1p_2}$), J($p_1$d), j(1/$p_1$d), K($p_1p_2$d), k(1/$p_1p_2$d), L(d$\sqrt{p_1p_2}$), l(1/d$\sqrt{p_1p_2}$), V($p_1$/d), E($p_1/d^2$), W($p_1^2$/d), w($p_1p_2$/d), F($p_1^2/d^2$), f($p_1p_2/d^2$), S($p_1^2/d^3$), s($p_1p_2/d^3$), T($p_1^2/d^4$), t($p_1p_2/d^4$);

- The overlapping descriptors interaction can be one of: R and r (threat descriptors as scalars, compute resultant relative to a given atom j and respectively conventional origin), M and m (first calculate and then use the property center similarly to well-known mass center calculations), and D and d (threat descriptors as Cartesian vectors);
- Molecule fragmentation are made on pairs of atoms using one of: m (minimal fragments), M (maximal fragments), D (Szeged distance based fragments) and P (Cluj shortest paths based fragments);

- Molecular descriptor cumulates overall fragmental descriptors values by using one of:
  - Conditional group: m (smallest), M (highest), n (smallest absolute), N (highest absolute);
  - Averages group: S (sum), A(average of all valid values), a(S divided by number of all fragments), B (average first by atom and then my molecule), b (by bond);
  - Geometric group: P (multiplication), G(geometric mean, valid fragments), g (adjusted G), F (by atom and then by molecule), f (by bond);
  - Harmonic group: s (harmonic sum), H (harmonic mean, valid fragments), and similarly to above h, I, and i.

# MDF formulas and names

- MDF values enter in QSAR modeling after a transformation (linearization procedure, one of: **I** (identity, no change), i (inverse), A (absolute), a (inverse of absolute), L (logarithm of absolute), l (logarithm). The mathematical formula of the calculation is like:

**$L_D(S_F(\{I_M(A_P,D_O,D_F(A_P,D_O),f)\ |\ f \text{ is from } F_C(Molecule)\}))$**

where $L_D$ is linearization descriptor, $S_F$ is overall superposing formula, $I_M$ is interaction model, $A_P$ is atomic property, $D_O$ is distance operator, $D_F$ is descriptor formula and $F_C$ is fragmentation criteria.

- As result, a number of 787968 MDF members are calculated:

  $2(D_O)*6(A_P)*24(D_F)*6(I_M)*4(F_C)*19(S_F)*6(L_D)$

- Labeling obey the construction of the descriptor. Thus, AiPdtQt are from:

$D_O = t, A_P = Q, D_F = t, I_M = d, F_C = P, S_F = i, L_D = A$

- Not all MDF members has real (computable) and not identical values. More, not all of them are distinct each from other. A procedure to clean the MDF set was implemented and applied.

# Data set

| Name of the compounds | Mutagenicity | Cytotoxicity |
|---|---|---|
| 8-methyl-quinoline | -0.71 | -3.39 |
| 8-aminoquinoline | -0.24 | N/A |
| 8-hydroxyquinoline | 0.79 | -2.30 |
| 8-chloroquinoline | 0.37 | -2.38 |
| 8-ethylquinoline | 0.40 | -2.73 |
| 8-cyanoquinoline | -0.46 | -3.59 |
| 8-ureidoquinoline | -1.93 | -3.71 |
| 8-fluoroquinoline | -0.55 | -3.67 |
| 8-sulfonamidoquinoline | -2.82 | -4.45 |
| 8-benzyloxyquinoline | 0.92 | -1.41 |
| quinoline | 0.09 | -3.12 |
| 8-methoxyquinoline | -1.50 | -3.75 |
| 8-ethoxyquinoline | -1.05 | -3.42 |
| 8-quinolinol acetate | -0.26 | -2.62 |
| N-8-quinolinyl acetamide | -1.09 | -3.64 |

# Results

- For the Quinolines set, only 319867 (all 15, Mutagenicity) and 319827 (14 of 15, Cytotoxicity) has real and not identical values and only 102608 (15) respectively 103411 (14) are distinct each from other. The selected members (102608 for Mutagenicity and 103411 for Cytotoxicity) enter into multiple linear regression analysis. Mono-varied and bi-varied models were applied. At the end of all pair's computations (for bi-varied model, 5264149528 pairs for Mutagenicity), the best QSAR models were selected and presented.

# MDF QSARs for Mutagenicity of Quinolines

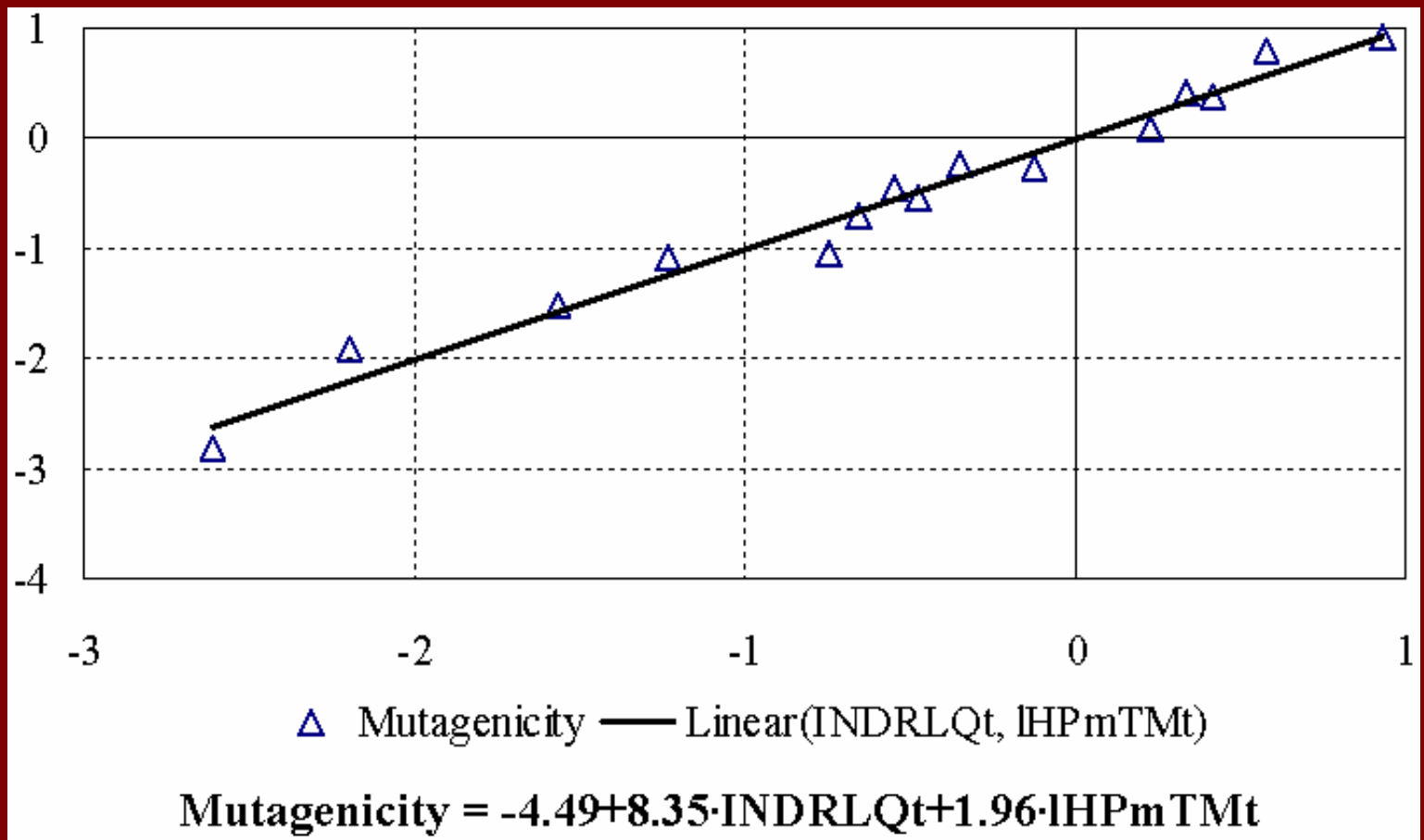
| No | QSAR | r, $r^2$, $r^2_{adj}$ | F, p% | $r^2_{cv\text{-}loo}$ | Remarks |
|---|---|---|---|---|---|
| 1 | 4.63<br>- 6.58·*iHPMdCg* | 0.8,<br>0.65,<br>0.62 | 23.8,<br>$3 \cdot 10^{-2}$ | 0.57 | Mono-varied model, n = 15 |
| 2 | -4.49<br>+8.35·*INDRLQt*<br>+1.96·*lHPmTMt* | 0.99,<br>0.98,<br>0.97 | 250,<br>$1.7 \cdot 10^{-8}$ | 0.96 | Bi-varied model, n = 15<br>$r^2$(*INDRLQt, lHPmTMt*) = 0.63<br>$r^2$(*INDRLQt*) = 0.123<br>$r^2$(*lHPmTMt*) = 0.03 |

# MDF QSARs for Cytotoxicity of Quinolines

| No | QSAR | r, $r^2$, $r^2_{adj}$ | F, p% | $r^2_{cv\text{-}loo}$ | Model remarks |
|---|---|---|---|---|---|
| 1 | -4.14<br>$+8.39 \cdot 10^{-3} \cdot aAmrKQt$ | 0.845,<br>0.715,<br>0.692 | 30,<br>$1.4 \cdot 10^{-2}$ | 0.573 | Mono-varied model, n = 14 |
| 2 | $-7.18 \cdot 10^{-1}$<br>$+2.25 \cdot 10^{-1} \cdot lsMrSQg$<br>$+9.87 \cdot 10^{-2} \cdot ASPrVQg$ | 0.98,<br>0.96,<br>0.95 | 125,<br>$2.7 \cdot 10^{-6}$ | 0.928 | Bi-varied model, n = 14<br>$r^2(lsMrSQg, ASPrVQg) = 0.38$<br>$r^2(lsMrSQg) = 0.06$<br>$r^2(ASPrVQg) = 0.004$ |
| 3 | -1.58<br>$+2.06 \cdot 10^{-1} \cdot lNMrSQg$<br>$+9.3 \cdot 10^{-2} \cdot ASPrVQg$ | 0.98,<br>0.96,<br>0.95 | 122,<br>$3.1 \cdot 10^{-6}$ | 0.934 | Bi-varied model, n = 14<br>$r^2(lNMrSQg, ASPrVQg) = 0.34$<br>$r^2(lNMrSQg) = 0.004$<br>Best cross-validation score |
| 4 | -1.6<br>$+3.37 \cdot 10^{-1} \cdot lNMrEQg$<br>$+9.47 \cdot 10^{-2} \cdot ASPrVQg$ | 0.98,<br>0.96,<br>0.95 | 119,<br>$3.5 \cdot 10^{-6}$ | 0.933 | Bi-varied model, n = 14<br>$r^2(lNMrEQg, ASPrVQg) = 0.35$<br>$r^2(lNMrEQg) = 5 \cdot 10^{-5}$ |

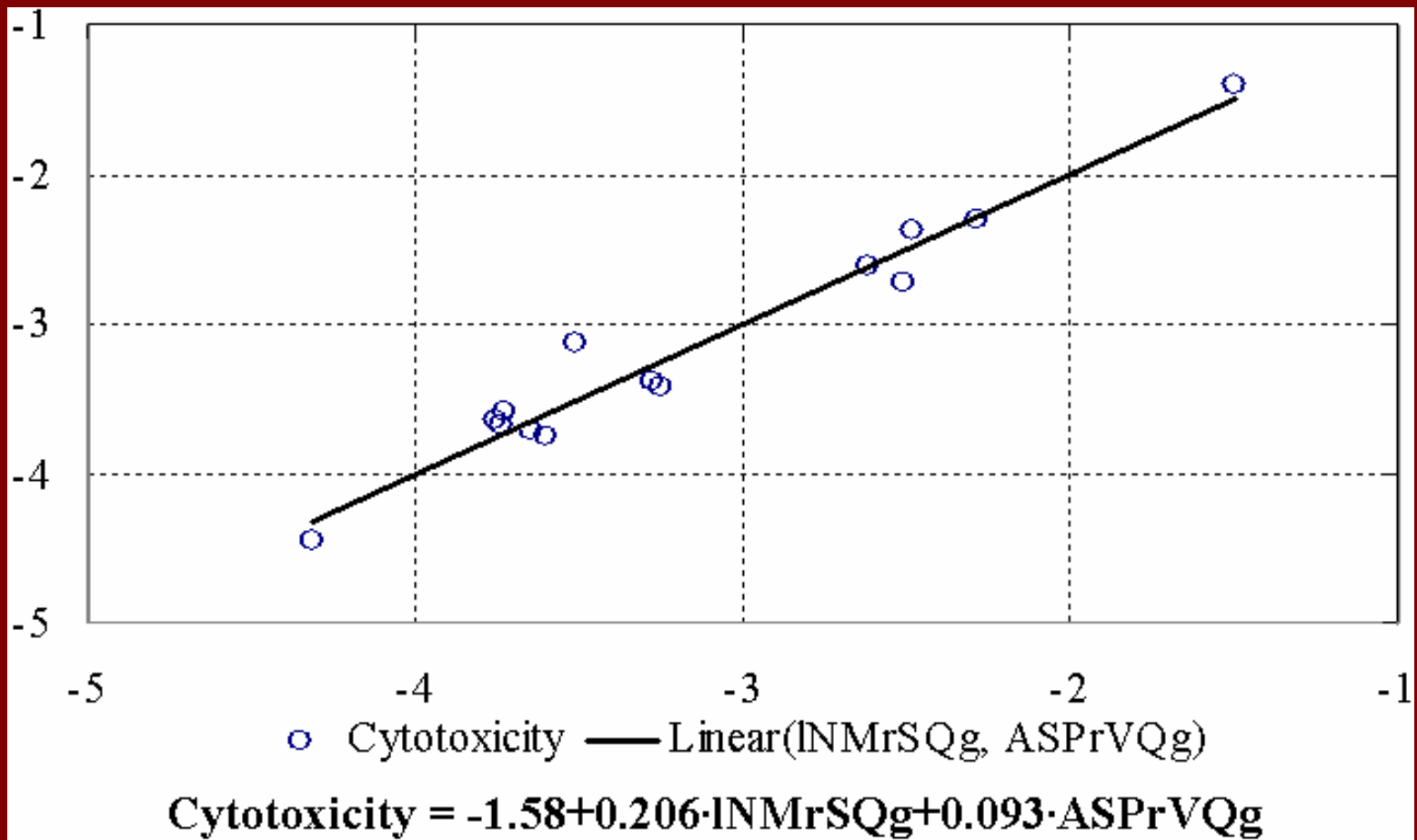
**The 10th Electronic Computational Chemistry Conference (ECCC10), April 1-30, 2005**
**held at http://eccc.monmouth.edu**

# Plot of best MDF QSAR for Mutagenicity of Quinolines




**Mutagenicity = -4.49+8.35·INDRLQt+1.96·lHPmTMt**

# Plot of best MDF QSAR for Cytotoxicity of Quinolines



# Remarks

1. As are expected, a pair of MDF members provides significantly better explains of the activity (mutagenicity and cytotoxicity) compared with single member. More, comparing with previous reported results (*):
   - Mutagenicity, n = 13 - two Quinolines omitted, $r^2$ = 0.87 vs. 0.98 (our result with MDF members)
   - Cytotoxicity, n = 13 - one Quinoline omitted, $r^2$ = 0.8 vs. 0.96 (our result with MDF members)

   the MDF produces better explanation of structure-activity relationship.

(*) Smith J.C., Hansch C., Morton J.M., QSAR treatment of multiple toxicities: the mutagenicity and cytotoxicity of quinolines, Mutation Research, 1997, 379, p. 167-175.

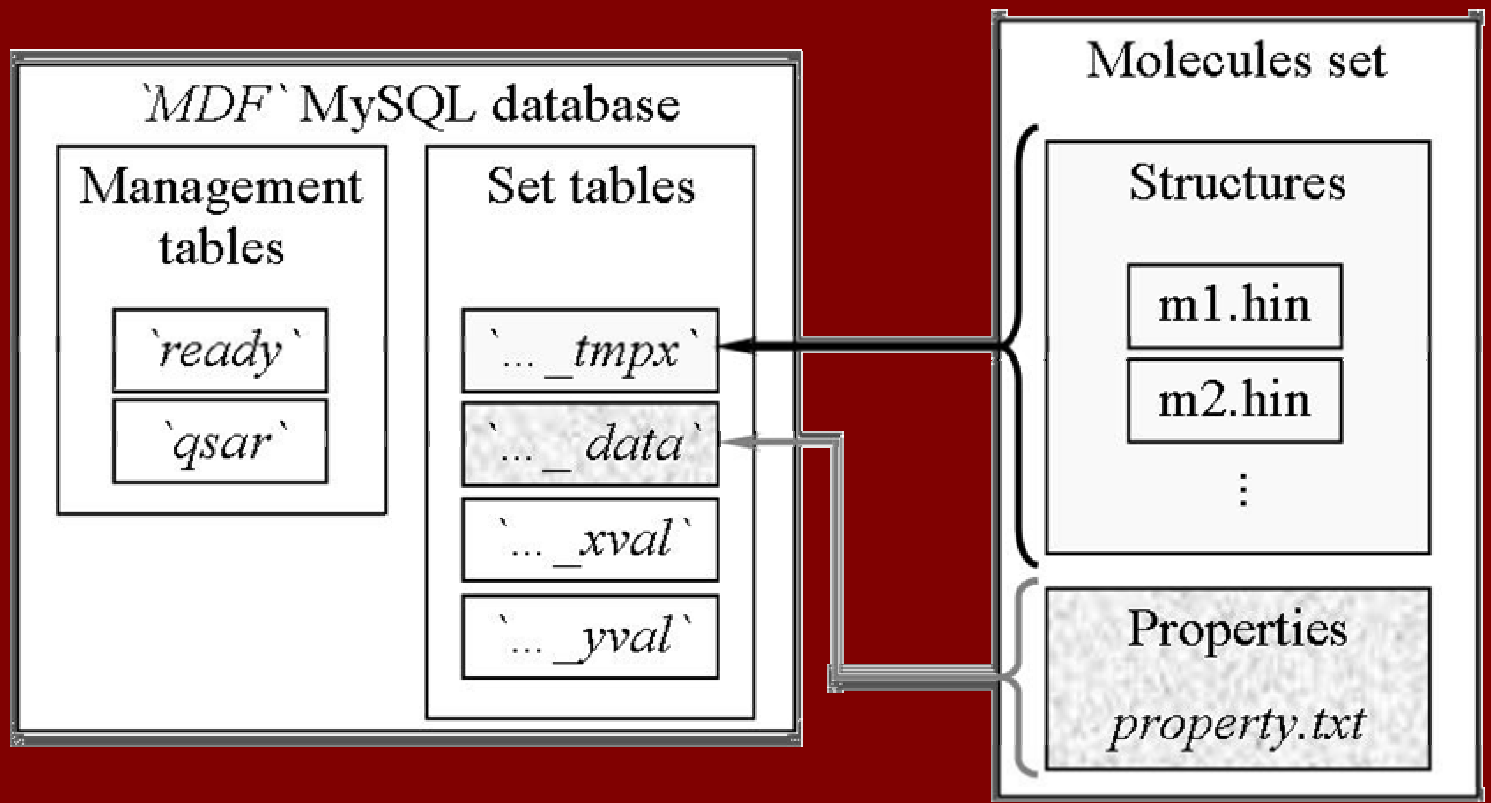
2. The absence of the best descriptor from the mono-varied model from the pair(s) of best bi-varied model, the almost null correlation between every single descriptor from bi-varied model and activity, and the presence of statistically significant link between descriptors from the best bi-varied pairs demonstrates that it is no link between using of orthogonal descriptors (Principal and/or Dominant Component Analysis) and QSAR modeling.
3. In bi-varied QSARs, when is more than one model that obtains the same values of correlation coefficient, the cross-validation score allows to choose the best model.

4. Even if using of MDF in QSAR modeling is time consuming, it has doubtless advantages, such as better QSAR and a much closer structure activity explanation.
5. Best bi-varied QSAR model of Mutagenicity uses the INDRLQt and lHPmTMt members, and is statistically significant ($p = 1.7 \cdot 10^{-8}$ %).
6. All three best of bi-varied QSARs of Cytotoxicity use members which consider the geometrical shape (g) as well as the atomic property represented by the partial change (Q) and are statistically significant ($p \approx 3 \cdot 10^{-6}$ %).
7. The obtained QSAR models allow making of important remarks on structural nature of Mutagenicity and Cytotoxicity activities.

# Conclusions

- Mutagenicity of Quinolines is almost of molecular topology nature and is strongly dependent on both atomic mass and partial atomic charge (99% for bi-varied MDF QSAR model with INDRL**Qt** and lHPmT**Mt** members).
- Cytotoxicity of Quinolines is strongly dependent on the partial change atomic property and its behavior is almost of molecular geometry nature (98% for bi-varied MDF QSAR model with lNMrS**Qg** and ASPrV**Qg** members).

# The End.


`MDF` MySQL database
Management tables
`ready`
`qsar`
Set tables
`..._tmpx`
`..._data`
`..._xval`
`..._yval`
Molecules set
Structures
m1.hin
m2.hin
Properties
property.txt