1 Introduction

One of the key insights of Machine Intelligence research has been that an appropriate machine representation is essential for success. This insight is the basis for the success of the machine learning technique of inductive logic programming (ILP). ILP uses logic programs (first order predicate logic) to represent learning problems. This contrasts with the use of attribute vectors (propositional logic) in standard machine learning, neural network, and statistical methods [22]. For certain problems the richer language of logic programs allows a more appropriate problem representations to be used, and this can result in greater predictive accuracy and/or more comprehensible results.

Within ILP arguably the most successful applications have been within the field of learning molecular structure activity relationships (SARs). Here the problem is to learn a predictive relationship between a molecular structure and its activity (usually a real number or class type). SAR problems are particularly suitable for ILP as the key to them is representing molecular structure, and this is a task well suited to ILP as it is relational - a molecule can naively be consider to be a set of atoms *related* together by bonds. Various relational representation of molecular structure have been studied within the ILP community [12–16, 18–20, 26, 30].

In this paper we present a new approach to representing molecular structure. This approach is a relational extension of our StruQT representation [1, 2]. The inspiration for StruQT comes from the atoms in molecules (AIM) theory of Bader [3, 4, 6, 24]. The AIM quantum chemistry approach is well suited for machine inference as it is topologically based and close to the naive atom-bond current ILP representation. Our new molecular representation is richer, and more chemically correct, than any previous molecular representation used with ILP (or to the best of our knowledge any field of AI). The approach opens the way to the general representation of fields for machine inference.

1.1 Representing molecular structure in ILP

The first example of applying ILP to problems associated with representing molecular structure was work on modelling the activity of trimethoprim analogues in dihydrofolate reductase [14]. The representation used in this work was close to that of standard propositional learners - where the data is represented as a table, with examples as rows and attributes as columns. This propositional like representation is only possible for SAR problems where all the molecules share a common template. The trimethoprim analogues share a template that can have molecular groups attached at ring positions 3,4, and 5. This enabled the molecular structures to represented in the form:

```
struc(d35,no2,nhcoch3,h).
```

which states that drug no. 35 has NO2 substituted at position 3, NHCOCH3 substituted at position 4, and no substitution at position 5 (hydrogen is default). The chemical properties of the substituted groups were represented by introducing predicates for each property, e.g.:

```
polar(br,polar3).
```

which states that Br (bromine) has a polarity of value 3. Arithmetic information was then required to compare values of properties, e.g.:

```
great_polar(polar4,polar2).
```

as the ILP algorithm Golem [12, 14] which was used for induction did not have built in arithmetic. This general form of representation was subsequently applied to triazines [13], and tacrine and steroid anologue problems [15].

The most natural representation for molecular structure is one based on atoms and bonds. This representation was combined with ILP in the SAR problem of predicting the mutagenicty of chemicals [16]. For example the following predicates are part of the representation of the molecule 3,4,3'-trinitrobipehyl in this strudy. Atoms are represented in the form:

```
atom(127,127_1,c,22,0191).
```

which states that that in compound 127 (3,4,3'-trinitrobipehyl), atom no. 1 is a carbon atom of type 22 (aromatic) and it has a partial charge of 0.191. Bonds are represented in the form:

```
bond(127,127_1,127_6,7).
```

which states that in compound 127 there is a bond between atoms no. 1 and 6 of type 7 (aromatic). This representation was designed to work with ILP programs such as Progol [23] which are designed to allow implicit background knowledge, such as of arithmetic. This representation is completely general for chemical compounds and no special attributes need to be invented. This general representation was applied to the prediction of chemical carcinogenicity [17, 27, 28].

The atom and bond representation could also be easily extended to the representation of 3-Dimensional molecular structure. This can be achieved by simply adding Cartesian coordinates to the atom predicate and coding information about 3-D geometery in the background knowledge - as first suggested in [9, 16, 27, 28]. For example the atom above could be represented in the form:

```
atom(127,127_1,c,22,0191,1.40,3.36,7.63).
```

where the X coordinate is 1.40, the Y coordinate is 3.36, and the Z coordinate is 7.63. There is no need with this representation to align each molecule to an external coordinate system, and each molecule can have its own local coordinate frame. This form of representation allows molecular pharmacophores (3D characterisations of drugs) to be learned. The representation can be further extended by allowing for different molecular conformations. These are distinct low energy shapes which a molecule can take (typically only one of these conformations has an "activity"). The multi-conformation problem can be elegantly represented and exploited in ILP by simply labelling the from which conformation a particular structure comes from, e.g.:

```
atom(127,3,127_1,c,22,0191,1.40,3.36,7.63).
```

would label the atom as coming from conformation 3. The problem then become that of finding at least one conformation with the desired pharmacophore (which can be easily represented in logic). This approach was first applied to inhibitors of angiotensin converting enzyme [10] and more recently compared with the standard 3D SAR method of CoMFA (see below) with favourable results for ILP [11].

The representation of molecular structure in ILP has also been extended by inclusion of information about standard molecular groups. When a chemist looks at a molecule s/he generally reduces the molecule to standard groups such as alcohols, esters, aromatic rings, etc. These chemical groups can easily be defined in the learning background knowledge based on the atom/bond representation [29]. This has now been done for most of the standard groups that chemists would recognise, and structure is represented as the topological or geometric relationship between these chemical groups.

1.2 CoMFA

Probably the most popular current method of representing molecular structure in SARs is that of comparative molecular field analysis (CoMFA) technique [8, 21, 25, 31] where the basic idea is to represent the 3-D molecular field by a collection of sampled data points. Central to the method is the application of a 3-D grid which covers the structures of all the molecules in the data set. An attribute is assigned to each grid point and the the set of attributes collected into a vector to represented the 3-D shape of the field. Thus, the shape information of the field is indirectly coded as attribute index numbers. Each attribute contains a scalar which signifies the value of the field at the sampled XYZ grid point.

The key problem with the CoMFA approach is that it is difficult to find comparable XYZ sampling points between molecules. To achieve this the method requires a time consuming and error prone relative orientation of the molecules in the data set. A problem with this is that the general concept of molecular similarity may not be the most appropriate for the problem at hand. Indeed finding the correct similarity measure between molecules for a SAR problem is close to solving the SAR problem in the first place. The CoMFA method therefore fails both to represent molecular electronic structure efficiently and to be efficient for machine inference.

2 Molecular structure and quantum mechanics

The ideal machine representation of molecular structure should:

- Be firmly rooted in quantum mechanics - molecules are quantum mechanical objects not classical ones.
- Encode efficiently the electronic structure of the molecule.
- Be related to existing chemical knowledge and concepts.
- Be efficient for machine inference.

2.1 StruQT

The StruQT representation is based on the *topology* of the molecular 3D field of electron density. The topology is described by a set of connected *critical points* (CPs) which correspond to locations in 3D space where the gradient of the scalar field is zero. These CPs constitute an efficient compression of the field and can be considered as a general approach for representing nD continuous fields. Thus, nD fields are compared by inspecting their sets of critical points rather than their corresponding sampled points.

We have previously used StruQT in an attribute (propositional) based way, and have tested it with partial least squares regression and artificial neural networks [1, 2]. Unfortunately, in this form, StruQT has a correspondence problem similar to CoMFA: we need to find comparable critical points between molecules. Although there is less need for relative orientation using the StruQT approach. This correspondence problem can be solved by incorporating StruQT with ILP.

2.2 The quantum mechanical basis for StruQT

Quantum mechanics does not explicitly define what is meant by an atom. In the quantum topological atoms in molecules (AIM) theory, Richard Bader and co-workers [4, 5, 7] remides this by providing a rigorous quantum chemical definition for atoms. It is clearly desirable to retain the concept of the atom, as it has proved its worth over the centuries and is the key to our intuitive understanding of chemistry. AIM theory therefore provides a strong link between chemical intuition and the theory of quantum mechanics. This means that AIM theory is an excellent starting point for developing molecular representations for machine inference.

The electron density distribution $\rho(r)$ itself does not give much indication of interesting structures since most of the density is concentrated around the nuclei. However, by inspecting the gradient vector field $\nabla\rho(r)$ a different picture emerges. Here structures related to e.g. bonds, rings and cages can be found. In $\nabla\rho(r)$ we can trace the "field lines" which correspond to the path taken by particles following the gradient vector in each point, see Figure 1. These gradient paths originate at infinity and terminate at a critical point where $\nabla\rho(r) = 0$. We also see gradient paths linking different critical points. Critical points can be classified into different types by inspecting the diagonalized Hessian matrix associated with each CP.

The elements of the matrix are given by:

$$H_{ij} = \frac{\partial^2\rho(r)}{\partial x_i \partial x_j}, \quad i \in [1, 2, 3] \tag{1}$$

and its associated eigenvalues $\lambda_1, \lambda_2, \lambda_3$ can be used to construct two measures for CP classification purposes:

- The signature, which is the algebraic sum of the *signs* of the three eigenvalues
- The rank, which is the number of non-zero eigenvalues

In this article we will only consider CPs which have rank equal to three. Lower rank CPs do exist and they are alway associated with some change in the system, this could be e.g. an internal rotation of a group or during a reaction where a bond is broken.

From the rank and signature (r, s) we have only four possible CPs with rank $= 3$:

- $(3, +3)$, a point with minimum electron density along all three dimensions. This is the *cage critical point* (CCP)
- $(3, +1)$, a point with minima along two dimensions and a maximum along a third dimension in the electron density. This is the *ring critical point* (RCP)

- $(3, -1)$, a point with maxima along two dimensions and one minimum along a third dimension in the electron density. This is the *bond critical point* (BCP)
- $(3, -3)$, a point with maximum electron density along all three dimensions. This is the *nuclear attractor* (NA) critical point. We may also have in special cases that such an attractor occurs when there is no nucleus, called the non-nuclear attractor critical point.

StruQT can in principle make use of all these CPs for describing the electronic structure. However, the most important of these points is the BCP. This is the only CP being used in the following SAR modelling. The BCP effectively describes the electronic structure of the molecule.

The set of critical points for each molecule are contained in an AIM descriptor matrix (ADM) [2] where each row contains a CP and each column contain a property for the critical point. These properties are mainly derived from the Hessian matrix associated with the critical point. Once the Hessian matrix has been computed a diagonalization will identify its eigenvalues $\lambda_1, \lambda_2, \lambda_3$. The most common properties used in StruQT for the BCP are :

- The XYZ coordinates of the BCP
- The Laplacian (computed as $\lambda_1 + \lambda_2 + \lambda_3$)
- The electron density, $\rho(r)$
- The ellipticity , ϵ (computed as $(\frac{\lambda_1}{\lambda_2}) - 1$)
- The two lengths involved in the interatomic interaction line (AIL)

Physical meaning can be attributed to the ellipticity and the Laplacian. The ellipticity measures the deviation of the electron density distribution from symmetrical around the bond. As a consequence of this the ellipticity can be used to measure the π character of a chemical bond. However, not all bonds with large π contributions have large ellipticity values. The typical counter example is etyne where we know there is a large overlap between π orbitals, however due to symmetry conditions, the ellipticity does not reflect this.

The Laplacian measures the degree of charge concentration of the bond. When the Laplacian is negative, we have charge concentration in the bond. This is typical for covalent and polar bonds. When the Laplacian is positive, we have charge depletion in the bond and is usually associated with ionic or hydrogen bonds.

3 The relational StruQT representation

A hierarchy of concepts are involved in AIM theory. At the top we have the scalar field type such as the electron density, the Laplacian, or the electrostatic potential. Given a (continuous) scalar field we can then compute critical points (after constructing the gradient vector field). It is here possible to make the approach even more general by stating that we start from the critical points from any vector field. A *gradient* vector field is just one example of a vector field. In fact, gradient vector fields have special properties that are not necessarily observed in vector fields in general. For instance, the curl $F(\mathbf{r})$, i.e $\nabla \times F(\mathbf{r})$ is zero for all gradient vector fields and the eigenvalues for critical points will always be real. This means that vorticies as found in aerodynamic or hydrodynamic flow cannot be found in gradient vector fields. To make our approach sufficiently general, we therefore state that the critical points used have been derived from a vector field specified type.

Given the field type, we need to encode the properties of the critical points. Irrespective of field type, the critical point is described by its Hessian matrix which contains information about the local curvature of the CP. As seen above we used the eigenvalues of the diagonalized Hessian matrix to classify into various types of CPs. The rank and signature approach will always be correct, however their interpretations in terms of bond, atoms, ring and cage structures may only be valid for the electron density field.

On the other hand, when we can make mappings into chemically meaningful concepts they are extremely valuable for the interpretation of the SAR models. We have therefore decided to include field specific predicates that will aid such chemical interpretations whenever possible. Here we show only for the electron density scalar field, but the same is also possible for CPs in the Laplacian scalar field where the CPs are closely linked to valence shell electron repulsion (VSERP) model which has been successful as an approximate molecular model. This is another example how AIM theory is able to provide a quantum based justification of intuitive models of molecular structure.

3.0.1 Predicates

First we list the most general predicates that will be valid for all types of fields and critical points:

- `cp(Id,Field,[x,y,z])` where `cp` is the predicate for a critical with unique identity value `Id`, coordinates `[x,y,z]` (a list) and the Hessian. Here it is assumed that the Hessian is represented by the full matrix, however in most cases only the eigenvalues of this matrix are needed.
- `mol_conf(Id,Mol_Id, Conf_Id)` which identifies the molecular conformation. `Mol_Id` is the unique identifier for the molecule and `Conf_Id` is the identifier for the conformation.
- `signature(Id,Signature)` which contains the information related to the signature of the Hessian matrix. This is the sum of the *signs* of the eigenvalues. As seen from above the signature value `Signature` can have the following values: $[-3, -1, 1, 3]$.
- `rank(Id,Rank)` which contain the information about the rank (i.e. number of non-zero eigenvalues) of the Hessian.
- `con_path(Id_1,Id_2)` is a predicate that indicates which critical points are connected. They connected by gradient paths that either originate from infinity or a critical point.
- `hessian(Id,L_1,L_2,L_3)` contains the eigenvalues of the Hessian matrix for critical point `Id`.

In many applications we make use of `Field = electron density` which means we can attach various physical interpretations to the critical points. The same is also true for the `Field = Laplacian of the electron density` but the interpretations of the physical role of each CP is different. When `Field = electron density` we can make use of the following predicates:

- `nuc_attr(Id)` which states that the CP in question is a nuclear attractor

- `bcp(Id)` which states that the CP in question is a bond critical point
- `rcp(Id)` which states that the CP in question is a ring critical point
- `ccp(Id)` which states that the CP in question is a cage critical point
- `electron_density(Id,Density)` which contains the electron density for CP `Id`
- `ellipticity(Id,Ellipticity)` which contains the ellipticity value for CP `Id`
- `laplacian(Id,Laplacian)` which contains the laplacian value for CP `Id`
- `ail(Id,Id_1,Id_2)` contains the identifiers for the three CPs involved in the definition of an atomic interaction line (AIL). `Id_1` and `Id_2` are connected to `Id`.
- `ail_dist1(Id,Dist_23)` contains the total distance from `Id_1` to `Id_2` plus the distance from `Id_3` to `Id_3` in the atomic interaction line (AIL).
- `ail_ratio(Id,Ratio)` contains the ratio of the minimum distance divided by the maximum distance in the atomic interaction line (AIL).

4 Results

In this section we have applied the new representation to a set of polychlorinated dibenzofuran molecules. Here is an example involving the first two BCPs for the first molecule in the data set (2-chloro dibenzofuran):

```
bcp(1)
cp(1,'eldens',[ 7.927,-4.395, 0.000])
signature(1,-1.000)
hessian(1,-0.725,-0.712, 0.480)
ail(1, 1, 2)
ail_dist1(1, 102.081)
ail_ratio(1,  0.012)
electron_density(1, 0.276)
laplacian(1,-0.957)
ellipticity(1, 0.019)
bcp(2)
cp(2,'eldens',[ 5.677,-4.240, 0.000])
signature(2,-1.000)
hessian(2,-0.641,-0.524, 0.334)
ail(2, 2, 4)
ail_dist1(2, 102.643)
ail_ratio(2,  0.026)
electron_density(2, 0.309)
laplacian(2,-0.830)
ellipticity(2, 0.224)
```

Here are the predicates for the first two RCPs in the same molecule:

```
rcp(24)
cp(24,'eldens',[ 4.705,-2.135, 0.000])
signature(24, 1.000)
```

```
hessian(24,-0.015, 0.078, 0.094)
electron_density(24, 0.020)
laplacian(24, 0.157)
rcp(25)
cp(25,'eldens',[ 0.647,-1.885, 0.000])
signature(25, 1.000)
hessian(25,-0.048, 0.165, 0.209)
electron_density(25, 0.044)
laplacian(25, 0.327)
rcp(26)
cp(26,'eldens',[-2.644, 0.518, 0.000])
signature(26, 1.000)
hessian(26,-0.015, 0.077, 0.092)
electron_density(26, 0.020)
laplacian(26, 0.154)
```

Here are the nuclear attractors (nuclei) positions:

```
nuc_attr(27)
cp(27,'eldens',[ 4.533,-2.558, 0.000])
nuc_attr(28)
cp(28,'eldens',[ 3.625,-1.934, 0.000])
```

The program MORPHY which was used to extract the information of the CPs of each molecule in the data set does not provide the eigenvalues of the Hessian for the nucleic attractors, but they could in principle be obtained.

5 Discussion

We have presented a new relational way of representing molecular structure (StruQT) for use in machine inference. The representation has the advantages: of being rooted in quantum mechanics, efficiently representing molecular electronic structure, being related to common chemical knowledge and concepts, and in being deigned to be efficient for machine inference (especially induction). This representation builds on previous work in representing molecular structure with ILP.

The relational StruQT method has clear theoretical advantages over existing SAR methods:

- Compared to standard 3D-SAR methods such as CoMFA there is no need to explicitly align the compounds.
- Compared to the propositional StruQT method there is no need to explicitly align the compounds.
- Compared to standard 3D-ILP methods the chemical description of the molecules is closer to reality.

There are two ways such advantages can be demonstrated: by ILP StruQT producing more easily comprehended results, and/or by producing better accuracy (or preferably error costs). We have not yet demonstrated either of these adavantages in practice. Finding a public domain suitable to do this is not trivial. We used the problem of predicting the wavelength of the lowest UV transition for a series of anthocyanidins to initially demonstrate the adavntage of propositional StruQT over standard chemometric programs. This dataset is unsuitable for comparison with ILP StruQT as the molecules all come from the same series (the problem is essentially propositional). The most commonly used ILP SAR database is that of "mutagenesis" [17]. This dataset is (currently) unsuitable for comparison with ILP StruQT because some of the molecules in this dataset are too large to apply the quantum chemistry programs required to calculate the electron densities. The most suitable available datasets are those of the thermolysin and glycogen phosphorylase inhibitors [11]. We have previously used these datasets to demonstrate that standard (pharmacophore) based 3D-ILP methods can outperform CoMFA with more comprehensible and more accurate predictions [11]. In these datasets the SAR problem is relational and the molecules small enough to apply quantum chemistry programs. We therefore plan to carry-out a comparative test of relational StruQT on these problems - we already have the results for standard standard 3D-ILP and ILP and will supplement them with propositional and ILP StruQT. Experience suggests that for many problems simpler representations will suffice, and in such cases there is no adavantage to be gained by using ILP StruQT. However, for a subset of problems ILP StruQT will be the representation of choice.

The relational StruQT representation would also appear suitable for general machine inference problems where the detailed quantum mechanical nature of the object needs to be represented, e.g. understanding high-temperature superconductivity, quantum control, the design of quantum computers, etc. In fact, the general representation we propose would be suitable for representing any problem with fields, which would make the representation potentially applicable to many problems in aerodynamics, hydrodynamics, geophysics, astronomy, etc.

References

[1] B.K. Alsberg, N. Marchand-Geneste, and R.D. King. Modeling quantitative structure-property relationships in calculated reaction pathways using a new 3d quantum topological representation. *Analytica Chimica Acta*, page (in press), 2001.

[2] B.K. Alsberg, N. Marchand-Geneste, and R.D. King. A new 3D molecular structure representation using quantum topology with application to structure-property relationships. *Chemometrics and Intelligent Laboratory Systems*, 54(2):75–91, 2001.

[3] R. F. W. Bader. Atoms in molecules in external fields. *Journal of Chemical Physics*, 91(11):6989–7001, 1989.

[4] R. F. W. Bader. 1997 polanyi award lecture - why are there atoms in chemistry? *Canadian Journal of Chemistry*, 76(7):973–988, 1998.

[5] R. F. W. Bader, M. T. Carroll, J. R. Cheeseman, and C. Chang. Properties of atoms in molecules - atomic volumes. *Journal of the American Chemical Society*, 109(26):7968–7979, 1987.

[6] R. F. W. Bader and P. J. Macdougall. Toward a theory of chemical-reactivity based on the charge-density. *Journal of the American Chemical Society*, 107(24):6788–6795, 1985.

[7] Richard F. W. Bader. *Atoms in molecules, a quantum theory.* International Series of Monographs on Chemistry, 22. Oxford University Press, Oxford, 1994.

[8] R. D. Cramer, D. E. Patterson, and J. D. Bunce. Comparative molecular field analysis (CoMFA). 1. effect of shape on binding of steroids to carrier proteins. *Journal of American Chemical Society*, 110:5959–5967, 1988.

[9] L. Dehaspe, H. Toivonen, and R. D. King. Finding frequent substructures in chemical compounds. In P. R. Agrawal, Stolorez, and G. Piatetsky, editors, *The Fourth International Conference on Knowledge Discovery and Data Mining*, pages 30–36. AAAI Press, Menlo Park.

[10] P. Finn, S. Muggleton, D. Page, and A. Srinivasan. Pharmacophore discovery using the inductive logic programming system progol. *Machine Learning*, 30:241–271, 1998.

[11] N. T. Geneste, B. Alsberg, and R.D. King. A new approach to pharmacophore mapping and qsar analysis using inductive logic programming. application to thermolysin inhibitors and glycogen phosphorylase. *Medicinal Chemistry*, page (accepted subject to revision), 2001.

[12] J. D. Hirst, R. D. King, and M. J. E. Sternberg. Quantitative structure-activity-relationships by neural networks and inductive logic programming .1. the inhibition of dihydrofolate- reductase by pyrimidines. *Journal of Computer-aided Molecular Design*, 8(4):405–420, 1994.

[13] J. D. Hirst, R. D. King, and M. J. E. Sternberg. Quantitative structure-activity-relationships by neural networks and inductive logic programming .2. the inhibition of dihydrofolate- reductase by triazines. *Journal of Computer-aided Molecular Design*, 8(4):421–432, 1994.

[14] R. D. King, S. Muggleton, R. A. Lewis, and M. J. E. Sternberg. Drug design by machine learning - the use of inductive logic programming to model the structure-activity-relationships of trimethoprim analogs binding to dihydrofolate-reductase. *Proceedings of the National Academy of Sciences of the United States of America*, 89(23):11322–11326, 1992.

[15] R. D. King, S. Muggleton, A. Srinivasan, C. Feng, R. A. Lewis, and M.J.E. Sternberg. Drug design using inductive logic programming. In *Proceedings of the 26th Hawaii International Conference on System Sciences*, pages 646–655. IEEE Computer Society Press, Los Alamitos.

[16] R. D. King, S. H. Muggleton, A. Srinivasan, and M. J. E. Sternberg. Structure-activity relationships derived by machine learning: the use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. *Proceedings of the National Academy of Sciences of the United States of America*, 93(1):438–442, 1996.

[17] R. D. King and A. Srinivasan. Prediction of rodent carcinogenicity bioassays from molecular structure using inductive logic programming. *Environmental Health Perspectives*, 104(S5):1031–1040, 1996.

[18] R. D. King and A. Srinivasan. The discovery of indicator variables for qsar using inductive logic programming. *Journal of Computer-aided Molecular Design*, 11(6):571–580, 1997.

[19] R. D. King, M. J. E. Sternberg, and A. Srinivasan. Relating chemical activity to structure - an examination of ilp successes. *New Generation Computing*, 13(3-4):411–433, 1995.

[20] De Raedt L. and Kramer S. The level-wise version space algorithm and its application to molecular fragment finding. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, page (in press). Morgan Kaufmann, San Francisco.

[21] C. Marot, P. Chavatte, and D. Lesieur. Comparative molecular field analysis of selective cyclooxygenase-2 (cox-2) inhibitors. *Quantitative Structure-activity Relationships*, 19(2):127–134, 2000.

[22] T. M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[23] S. Muggleton. Inverse entailment and progol. *New Generation Computing*, 13:245–286, 1995.

[24] O. Popelier. *Atoms in Molecules*. Prentice Hall, London, 2000.

[25] G. Roma, N. Cinone, M. Dibraccio, G. Grossi, G. Leoncini, M. G. Signorello, and A. carott. Synthesis, antiplatelet activity and comparative molecular field analysis of substituted 2-amino4h-pyrido[1,2-a]pyrimidin-4-ones, their congeners and isosteric analogues. *Bioorganic & Medicinal Chemistry*, 8(4):751–768, 2000.

[26] A. Srinivasan and R. D. King. Feature construction with inductive logic programming: a study of quantitative predictions of biological activity by structural attributes. *Lecture Notes in Artificial Intelligence*, 1314:89–104, 1997.

[27] A. Srinivasan, R. D. King, and D. Bristol. An assessment of submissions made to the predictive toxicology challenge. In T. Dean, editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 270–275. Morgan Kaufmann, San Francisco CA.

[28] A. Srinivasan, R.D. King, S.H. Muggleton, and M.J.E. Sternberg. The predictive toxicology evaluation challenge. In *Fifteenth International Joint Conference on Artificial Intelligence*. Morgan Kaufmann, San Francisco, CA.

[29] A. Srinivasan, S. H. Muggleton, M. J. E. Sternberg, and R. D. King. Theories for mutagenicity: a study in first-order and feature-based induction. *Artificial Intelligence*, 85(1-2):277–299, 1996.

[30] M. J. E. Sternberg, R. D. King, R. A. Lewis, and S. Muggleton. Application of machine learning to structural molecular-biology. *Philosophical Transactions of the Royal Society of London Series b- Biological Sciences*, 344(1310):365–371, 1994.

[31] S. E. Yoo, S. K. Kim, S. H. Lee, K. Y. Yi, and D. W. Lee. A comparative molecular field analysis and molecular modelling studies on pyridylimidazole type of angiotensin ii antagonists. *Bioorganic & Medicinal Chemistry*, 7(12):2971–2976, 1999.

6 Figure captions

- Figure 1 This figure shows the gradient vector field for a flat molecule. Hhere this molecule is 2-chloro-dibenzofuran. Three different critical points are observed: nucelar attractor (black circles), the bond critical point (black squares) and the ring critical point (black triangle). We also see the gradient field lines which either originate at infinity and terminate at a critical point, or could originate at a critical point and terminate at another. The electon density is increasing towards the nuclear attractor, but decreasing towards a ring critical point.

7 Figures

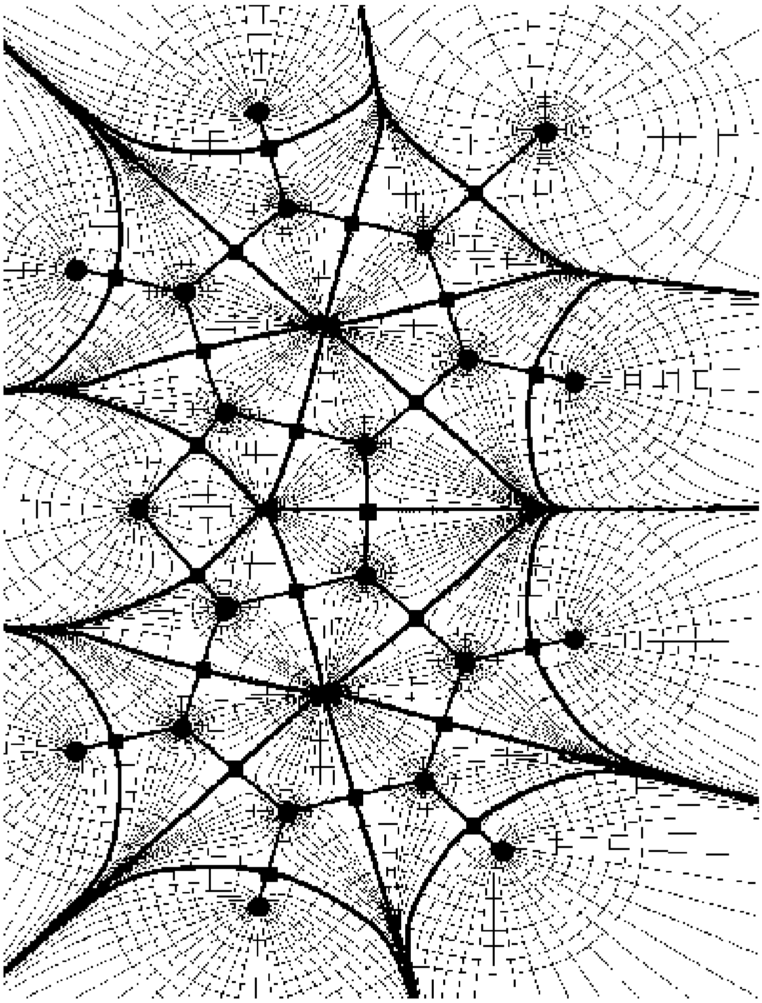

Figure 1: