

Physical and Molecular Properties of Agrochemicals: An Analysis of Screen Inputs, Hits, Leads, and Products

Eric D. Clarke* and John S. Delaney

Abstract: This work provides a comprehensive overview of agrochemical properties in terms of the way they change during progression from screen hit to product and in terms of their limits as expressed in modern commercial products. Most herbicides and fungicides readily meet the Lipinski 'rule of five' criteria for drug-like compounds with many meeting the more constrained limits reported for pharmaceutical leads.

Keywords: Agrochemicals · Bioavailability · Design guidelines · Property profiles · 'Rule of five'

1. Introduction

Identifying the particular balance of intrinsic potency and bioavailability required by a given chemical class to express a desired biological effect is a key part of the chemistry design and optimization process. Bioavailability in itself provides the challenge of understanding the balance of mobility and stability related properties required to allow expression of activity in lead generation screens and on progression from screen hit to lead series to product. In practice, chemists, organic, physical and computational, seek to define and exploit these balances in terms of physical and molecular properties. In 1997 Lipinski and co-workers at Pfizer published what is widely regarded the key paper defining physicochemical and structural properties profiles for optimal oral availability of drugs [1]. Their work unlocked results and analysis from other pharmaceutical companies, which further explored the concepts of drug-likeness and lead-likeness in terms of physical and molecular properties [2]. The Lipinski 'rule of five' placed upper limits for four properties; namely 500 for molecular weight, log P 5 for octanol/water parti-

tion coefficient, 5 for the number of hydrogen bond donors and 10 for the number of hydrogen bond acceptors. Subsequent publications, notably from AstraZeneca and GlaxoSmithKline, indicated that pharmaceutical leads are less structurally complex than marketed drugs and their property limits are more constrained [2].

Within the agrochemical industry Briggs was quick to follow up the Lipinski 'rule of five' with 'ground rules of three' outlined in a 1997 talk [3]; and in a recent poster Briggs and co-workers set limits to the physical properties of fungicides [4]. The most cited work on agrochemicals relating to the Lipinski 'rule of five' comes from Tice who focused on insecticides and foliar applied herbicides [5]. Our own initial contributions to understanding property profiles of agrochemicals, based on an analysis of compounds and physical properties taken from the Pesticide Manual [6] have been presented in talks and posters [7]. In many respects these reported property profiles for drugs and agrochemicals can be considered broadly similar, the most marked difference being that agrochemicals have a lower number of hydrogen bond donors. In 2002 we significantly expanded the scope of studies on agrochemicals to cover all types (insecticides, fungicides and herbicides), all stages of the progression stream (input, hit, lead, product) and a diverse range of physical and molecular properties. Preliminary reports of this work concluded that parameters relating to molecular size, atom types, hydrogen bonding, ionization state, lipophilicity and aqueous solubility can dif-

ferentiate agrochemicals by type and stage, but that differences in property profiles are subtle rather than dramatic [8]. This work has now been further developed and reviewed to give a comprehensive account of the physical and molecular properties of agrochemicals.

2. Properties

This section gives brief details of the molecular and physical properties used in this work. The structures of the fungicide azoxystrobin [9], the herbicide mesotrione [10] and the insecticide thiamethoxam [11] are given as examples of modern agrochemicals (Fig. 1).

2.1. Partition Coefficients

The octanol/water partition coefficients (log P oct) used are estimated values, designated ELOGP, defined as the mean value from three distinct prediction methods. These methods are CLOGP based on structural fragments, AlogP based on atom contributions and ACD/LogP based on a combination of atom and fragment contributions [12]. By definition, logP oct values relate to the un-ionized form of acids and bases.

2.2. Delta log P

Delta log P is defined as the difference between octanol/water and alkane/water partition coefficients ($\Delta\log P$) [13]. In practice this property has been directly predicted using the LFER equation for $\Delta\log P$

*Correspondence: Mr. E.D. Clarke
Syngenta
Jealott's Hill International Research Centre
Bracknell
Berkshire
RG42 6EY
United Kingdom
Tel.: + 44 1344 414067
Fax: + 44 1344 455629
E-Mail: eric.clarke@syngenta.com

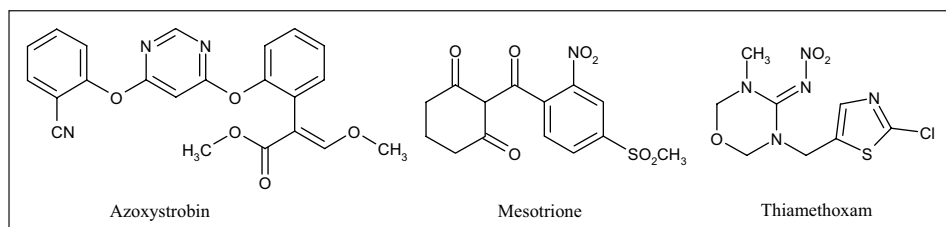


Fig. 1. Examples of modern agrochemical products

hexadecane ($\Delta \log P = \log P_{\text{octanol}} - \log P_{\text{hexadecane}}$) via PC based Absolv software [14].

2.3. Aqueous Solubility

The aqueous solubility of compounds was calculated using an in-house program called ESOL. This method was developed to produce solubility predictions to an accuracy comparable with Yalkowsky's general solubility equation but without the need for measured melting points [15]. The linear model uses four parameters ($\log P_{\text{oct}}$, molecular weight, number of rotatable bonds, proportion of heavy atoms defined as 'aromatic' by Daylight SMARTS [16]) to describe solubility in water ($\log S_w$ in ppm). The performance of ESOL ($R^2 = 0.51$) was comparable to the GSE ($R^2 = 0.48$) when judged against measured aqueous solubility values taken from the Pesticide Manual [6] for 469 agrochemical products; full details of the ESOL method will be described elsewhere.

2.4. Hydrogen Bonding

Hydrogen bonding capacity was assessed using two methods. Values obtained from Absolv [14] for the Abraham descriptors A (H-bond acidity) and B (H-bond basicity) gave a nuanced description of a molecule's overall hydrogen bonding capacity while counts of donors (any NH or OH group) and acceptors (any double bonded oxygen or aromatic nitrogen except NH – strong acceptors only) gave a simpler view of the same. It was found that H-bond donor count correlated quite well with Abraham's A ($R^2 = 0.78$) while H-bond acceptor count showed little correlation with Abraham's B ($R^2 = 0.20$). The poor correlation between B and acceptor count seemed to stem from the way the acceptor count ignored 'weak' acceptors. Abraham's B encompasses contributions from a far wider range of electronegative groups, making it more suitable for physical property prediction [17].

2.5. Aromatic and Heteroatom Proportion

The percent aromatic and percent non-carbon atoms for a compound were calculated using Daylight SMARTS sub-structures. An in-house Fortran program written

using the Daylight programming toolkit [16] ran a series of SMARTS sub-structural searches (patterns) against each molecule and reported the number of times each sub-structure occurred in the molecule. In each case the number of hits against the SMARTS pattern ("a" defining an atom in an aromatic system, "[!#6]" defining a non-carbon atom) was divided by the number of heavy (non-hydrogen) atoms and the result multiplied by 100 to give a percentage.

2.6. Formal Charge at pH 7

A count of formally charged centers was made using a program kindly provided by Peter Kenny (AstraZeneca, Alderley Park, UK) [18]. The program recognizes functionality predominantly charged at pH 7 (e.g. oxyanions, sulphonyls, aliphatic amines, guanidines) regardless of their registered protonation state using SMARTS patterns and the Daylight programming toolkit [16].

3. Analysis

This section gives detail of the datasets used, data formats and data visualization.

3.1. Datasets

Four sets of compounds were used in this work, representing the progression HTS input \rightarrow HTS hit \rightarrow lead series \rightarrow agrochemical product. These were a 9900 random subset of the Syngenta company compound collection (HTS inputs), a 6500 set of confirmed high throughput *in vivo* screen actives (HTS hits), 660 compounds which justified synthetic work (lead series), and the 1380 single organic products listed in the electronic Pesticide Manual [6]. The products were divided into pre- and post-1967 sets relative to date of Chemical Abstracts Service Registry Number (CAS RN). The median date of first use for products is about five years after allocation of CAS RN. It is the authors' experience-based view that the 750 post-1967 products are most representative for current agrochemical design. The hit, lead and product sets were further separated into the three main activity areas, fungicide, herbicide and insecticide. In total nine physical and

molecular properties were determined for compounds in each set [8].

3.2. Data Format and Visualization

All data analysis and visualization work has been carried out using Microsoft Excel 2000. For each dataset as a whole and subdivided into activity areas, mean, standard deviation (SD), 90th and 10th percentile values were determined for the nine properties. Property variations for the different sets of compounds have also been displayed as 'radar' plots [8]. To enable this the mean values for a given dataset were normalized against the mean and standard deviation of the screen input set to give a scaled value equal to $(\text{Dataset Mean} - \text{Input Mean}) / (\text{Input SD})$. Property means identical to the screen input mean are plotted on the 0 ring of the radar plot. Higher/lower values are shown as displacements away from/toward the center of the plot. The units of displacement are the number of input standard deviations (SD) from the input mean.

4. Results

Figs 2 to 5 show the four radar plots which give the scaled values obtained for the nine representative parameters for all compounds in the hit, lead and product sets and on division into insecticide, fungicide and herbicide types. These plots highlight the shifts in property profile that can occur as compounds progress from screen hit, through lead series to agrochemical product. In Table 1 we give the screen inputs values used to normalize the plots and the means and standard deviations for hits, lead series, products and post-1967 products. The property ranges for post-1967 agrochemical products delimited by 10th & 90th percentiles shown in Table 2 serve to define preferred property profiles. All references to products for the rest of this paper will refer to post-1967 products unless otherwise stated.

5. Discussion

Agrochemicals can be required to express activity near their point of application, or following uptake and movement within host and target species. For many fungicides and insecticides the host species are plants and applications are made to foliage. Different balances of properties can be expected for compounds which act as deposits on leaf surfaces compared to those which need some degree of uptake in plants to function. Herbicides always require uptake and movement into plants following

application to express activity. Consequently, compared to other agrochemicals, the overall property profiles of herbicides are expected to be more distinct. These points are more extensively covered by Tice [5] but with this, albeit simplistic, view in mind we first address herbicides.

The property profiles for herbicides are of particular interest as they should provide a guide to systemic properties for fungicides and insecticides. The molecular weight, lipophilicity and aqueous solubility profiles appear to be strongly correlated and constrained, particularly in herbicide leads and products. There is a clear bias in products towards non-carbon atoms. It is likely that a lower carbon to heteroatom balance is required to achieve acceptable aqueous solubility and lipophilicity for uptake and movement in plants. Herbicides, in contrast to insecticides and fungicides, have a preference for negative charge. Acid and ester functions are quite common in herbicide products and phytotoxicity can often be an initial concern for insecticide and fungicide hits which are esters. Overall,

we find our property profiles for herbicide leads and products to be reasonably consistent with those reported by Tice for foliar applied herbicides [5].

Fungicides show some echoes of the herbicide story (*e.g.* products are more soluble than hits), but also seem to have their own requirements. It's noticeable that fungicidal products seem to lack H-bonding functionality and are quite small with a mean molecular weight of 300. The presence of positively charged compounds is most noticeable in fungicide hits but this preference decreases markedly in leads and products. Note that fungicides do not appear to tolerate negative charge, which may reflect problems with phytotoxicity. These observations generally agree with the published views of Briggs and co-workers [4].

The most noticeable changes affecting insecticides on progression are a rise in lipophilicity, a systematic stripping away of aromatic functionality and H-bond donors, and an increase in the proportion of non-carbon atoms in the molecule. These obser-

variations most likely reflect replacement of aromatic systems or their substitution with halogens in leads to compensate for the high metabolic capability of insects relative to fungi and plants. The relatively low solubility/high lipophilicity of products reflects the scarcity of systemic insecticides. Acids tend not to make insecticides, again phytotoxicity being a potential problem. The presence of positively charged compounds is most evident in leads but products tend to be neutral. This could reflect the relative ease of metabolism of basic amines or compromised uptake of the charged molecule. The trends seen in insecticide properties are similar to those reported by Tice [5].

The overall profile for agrochemical products is effectively shaped by the contributions from herbicides and fungicides (72% of total). For agrochemicals as a whole progression from hits to products is dominated by rising solubility, decreasing basicity and the removal of carbon, particularly in aromatic systems.

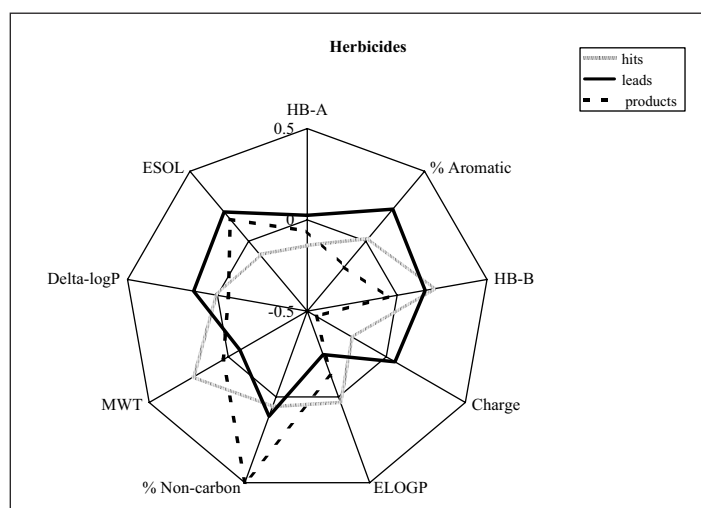


Fig. 2. Scaled values for herbicide hits, leads and products

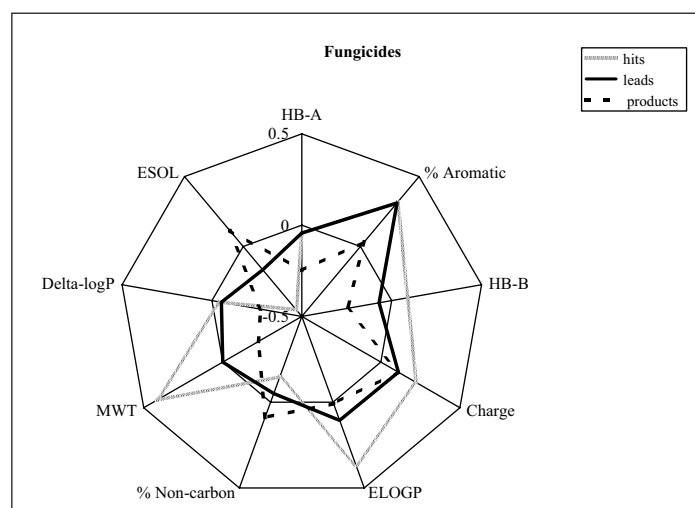


Fig. 3. Scaled values for fungicide hits, leads and products

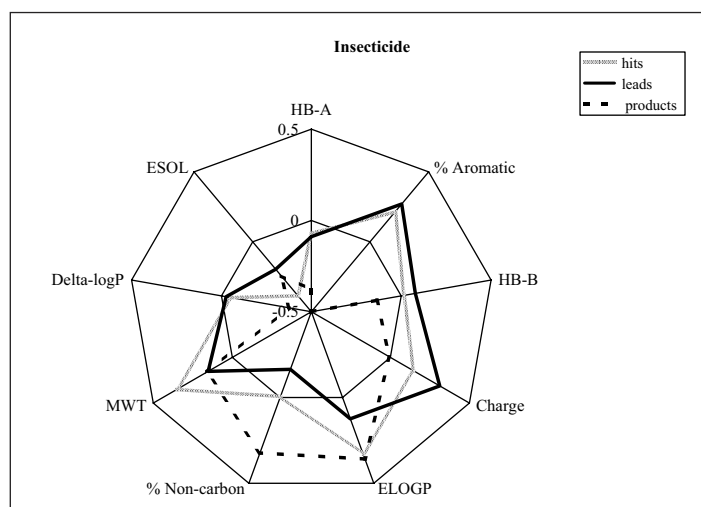


Fig. 4. Scaled values for insecticide hits, leads and products

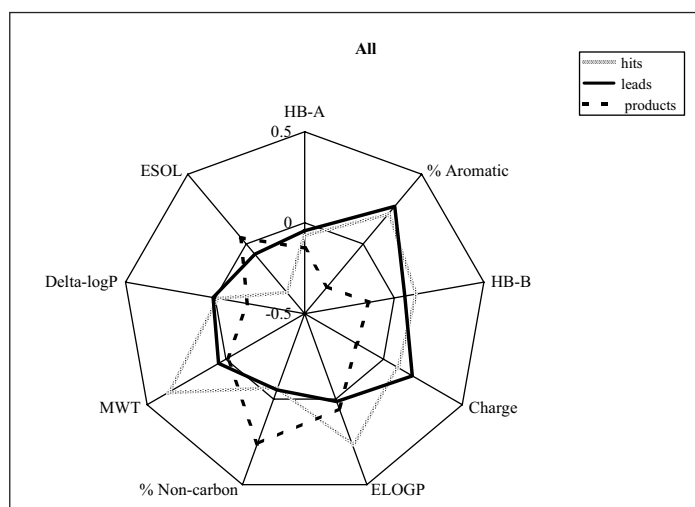


Fig. 5. Scaled values for agrochemical hits, leads and products

Table 1. Means and standard deviations for inputs, hits, leads, products and post-1967 products (mean/sd)

Property	Inputs	Hits	Leads	Products	Post-1967 Products
MWt	323/107	363/105	328/89	297/99	322/96
ELOGP	3.1/2.1	3.7/1.9	3.2/1.8	3.0/2.1	3.2/2.2
ESOL (log ppm)	1.5/1.7	0.9/1.5	1.4/1.4	1.9/1.7	1.5/1.7
$\Delta\log P$	2.9/2.1	2.8/1.8	2.9/1.9	2.3/2.2	2.5/2.3
Charge @ pH 7	0.0/0.6	0.1/0.6	0.1/0.6	-0.1/0.5	-0.1/0.5
H-Bond Acidity	0.3/0.4	0.3/0.4	0.3/0.4	0.3/0.4	0.3/0.4
H-Bond Basicity	1.4/0.7	1.4/0.6	1.4/0.5	1.1/0.7	1.3/0.6
% Aromatic atom	44/22	49/19	50/21	35/24	37/22
% Non-carbon atom	29/11	28/11	28/10	34/14	32/12

Table 2. 10th and 90th percentile limits for post-1967 products and sub-divisions into herbicides, fungicides and insecticides

Property	Agrochemicals	Herbicides	Fungicides	Insecticides
MWt	210 to 430	230 to 430	210 to 380	210 to 500
ELOGP	0.7 to 5.7	0.7 to 4.9	1.4 to 4.8	0.9 to 6.6
ESOL (log ppm)	-0.4 to 3.6	0 to 3.5	0.4 to 3.0	-1.3 to 3.4
$\Delta\log P$	0.4 to 5.3	0.7 to 5.9	0.7 to 3.8	0.4 to 4.6
Charge @ pH 7	-1 to 0	-1 to 0	0 to 0	0 to 0
H-Bond Acidity	0 to 0.9	0 to 0.9	0 to 0.6	0 to 0.7
H-Bond Basicity	0.6 to 2.0	0.7 to 2.1	0.7 to 1.6	0.7 to 1.8
% Aromatic atom	0 to 63	23 to 60	15 to 74	0 to 55
% Non-carbon atom	17 to 48	20 to 50	20 to 43	15 to 50

6. Conclusions

We continue our efforts to influence organic chemists through these observations and guides to physical and molecular property profiles found in the majority of agrochemicals as they progress from hits on high throughput *in vivo* screens through to lead series and finally to products. It needs to be emphasized that the profiles for hits and leads mainly reflect compounds arising from conventional and combinatorial chemistry approaches to lead generation. They may not necessarily be appropriate for compounds arising from structure based design or natural product approaches, particularly if an active transport process is thought to operate. What is most evident from this work is that herbicides and fungicides generally meet the Lipinski 'rule of five' criteria for drug-like compounds. Indeed, many compounds synthesized and selected as inputs to agrochemical high throughput *in vivo* screens fall within the more constrained property limits reported for lead-like compounds of interest to the pharmaceutical industry.

Whilst instructive and useful, we are well aware that such rules, guides and limits can be rather blunt instruments used to cut out what we believe to be screen inputs and hits with compromised bioavailability. The role of metabolism in defining the mobility-stability balance for agrochemicals is often unknown and more needs to be done in this area if we are to improve our overall knowledge of bioavailability. Finally, effort put in to gain bioavailability will count for nothing if intrinsic potency is lacking.

Acknowledgements

We thank Mr David Adams and the Physical Chemistry Measurements Team (Syngenta), Dr Kei Enomoto (Syngenta/UCL PhD student 1999-2003) and especially Miss Berengere Esparcieux (industrial placement student 2002-2003) for generating quality controlled SMILES representations of current and superseded compounds listed in the e-Pesticide Manual. In addition we thank Dr Clive Tomlin for providing electronic structures for current compounds in the Pesticide Manual. Miss Elizabeth Draper (University of Sheffield, MSc project 2002) and Mr Graham Mullier (Syngenta) are acknowledged for their key contributions to the development of ELOGP.

Received: September 18, 2003

- [1] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Advanced Drug Delivery Reviews* **1997**, *23*, 3.
- [2] a) S.J. Teague, A.M. Davis, P.D. Leeson, T.I. Oprea, *Angew. Chem. Int. Ed.* **1999**, *38*, 3743; b) T.I. Oprea, A.M. Davis, S.J. Teague, P.D. Leeson, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308; c) M.M. Hann, A.R. Leach, G. Harper, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856; d) T.I. Oprea, *J. Comp. Aided Molec. Design.* **2002**, *16*, 325; e) M.C. Wenlock, R.P. Austin, P. Barton, A.M. Davies, P.D. Leeson, *J. Med. Chem.* **2003**, *46*, 1250.
- [3] G.G. Briggs, 'Predicting uptake and movement of agrochemicals from physical properties' talk given at the SCI Meeting on uptake of agrochemicals and pharmaceuticals, London, UK, December, **1997**.
- [4] G.G. Briggs, P. Desbordes, P. Genix, 'Are there limits to the physical properties of fungicides?' poster presentation, 10th IUPAC Int. Congress on the Chemistry of Crop Protection, Basel, Switzerland, August **2002**. (philippe.desbordes@bayer-cropscience.com).
- [5] a) C.M. Tice, *Pest. Manag. Sci.* **2001**, *57*, 3; b) C. M. Tice, *Pest. Manag. Sci.* **2002**, *58*, 219.
- [6] a) C.D.S Tomlin (Ed.), 'The Pesticide Manual', 12th Edition, **2000**, Brit. Crop Protect. Council; Farnham, Surrey, UK; b) The e-Pesticide Manual, **2002-2003**, version 2.2.
- [7] a) E.D. Clarke, J.S. Delaney, UK QSAR & Chemoinformatics Meeting, Syngenta, UK, April **2001**, talk (<http://www.ukqsar.co.uk>); b) E.D. Clarke, J.S. Delaney, 10th IUPAC Int. Congress on the Chemistry of Crop Protection, Basel, Switzerland, August **2002**, poster & talk.
- [8] E.D. Clarke, J.S. Delaney, in 'Euro QSAR 2002: Designing Drugs and Crop Protectants' (Proceedings of the 14th European Symposium on Quantitative Structure-Activity Relationships, 8-13 September **2002**, Bournemouth, UK), Eds. M. Ford, J. Dearden, D. Livingstone, H. van de Waterbeemd, Blackwell Publishing, UK, **2003** (in press).
- [9] D.W. Bartlett, J.M. Clough, J.R. Godwin, A.A. Hall, M. Hamer, B. Parr-Dobrzanski, *Pest. Manag. Sci.* **2002**, *58*, 649.
- [10] G. Mitchell, D.W. Bartlett, T.E.M. Fraser, T.R. Hawkes, D.C. Holt, J.K. Townson, R.A. Wichert, *Pest. Manag. Sci.* **2001**, *57*, 120.
- [11] P. Maienfisch, H. Huerlimann, A. Rindlisbacher, L. Gsell, H. Dettwiler, J. Haettenschwiler, E. Sieger, M. Walti, *Pest. Manag. Sci.* **2001**, *57*, 165.
- [12] For current review of fragment-based methods see P. Japertas, R. Didziapetris, A. Petrauskas, *Quant. Struct.-Act. Relat.* **2002**, *21*, 23; also see a) R. Mannhold, A. Petrauskas, *QSAR Comb. Sci.* **2003**, *22*, 466; b) A. J. Leo, D. Hoekman, *Perspectives in Drug Discovery and Design.* **2000**, *18*, 19. AlogP predictions were made using Diamond Discovery (version 1.5) software obtained from Accelrys Inc., San Diego, California, USA and ACD logP values obtained using ACD Phys. Chem. Batch (version 4.76) software obtained from Advanced Chemistry Development Inc., Toronto, Canada.
- [13] P. Seiler, *Eur. J. Med. Chem.* **1974**, *9*, 473.
- [14] Absolv (v.1.4) software obtained from Sirius Analytical Instruments, Forest Row, East Sussex, UK. See J.A. Platts, D. Butina, M.H. Abraham, A. Hersey, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 835.
- [15] N. Jain, S.H. Yalkowsky, *J. Pharm. Sci.* **2001**, *90*, 234.
- [16] Daylight Toolkits and ClogP (v. 4.71) software obtained from Chemical Information Systems, Santa Fe, New Mexico.
- [17] a) M.H. Abraham, J.A. Platts, *J. Org. Chem.* **2001**, *66*, 3484; b) M.H. Abraham, K. Enomoto, E.D. Clarke, G. Sexton, *J. Org. Chem.* **2002**, *67*, 4782.
- [18] P. Kenny, Handling heterocyclic tautomerism. Daylight User Group Meeting, Santa Fe, New Mexico, February **1999**, talk.