ELSEVIER

(Q) SAR study on the metabolic stability of steroidal androgens

Roberta Bursi,* Marcel E. de Gooyer,† Arijan Grootenhuis,† Peter L. Jacobs,‡ Jaap van der Louw,§ and Dirk Leysen§

*Departments of *Molecular Design & Informatics, †Pharmacology, ‡Analytical Chemistry for Research, and §Medicinal Chemistry Departments, N.V. Organon, Oss, The Netherlands*

Metabolic stability is a key issue in the development of orally active androgens for Partial Androgen Deficiency in Aging Males (PADAM) and male contraception. Rates of metabolism in human hepatocyte suspensions provide useful information on the stability of compounds that undergo a first pass metabolism. We have derived a structure–pharmacokinetic relationship for a data set of 32 in-house steroidal androgens by means of the decision-trees technique. Volume, shape, number of rotatable bonds, and surface turned out to be the most important descriptors for classification. Only 2 of the 32 compounds were misclassified. The most stable compounds were classified in three leaf nodes on different branches of the tree, suggesting that higher metabolic stability can be achieved for the same substrate by different steric modifications. Further, it is generally assumed that the first step in cytochrome P450s oxidation reactions takes place by hydrogen abstraction to form a radical intermediate. An electronic model for hydrogen abstraction in steroidal androgens was, therefore, developed by means of ab initio calculations. Activation energies of steroid radical systems calculated as energy differences between the reactants equilibrium geometry energies and their corresponding transition states energies could be used to predict relative rates of metabolism to guide the design and redesign process of metabolically more stable steroidal androgens. © *2001 by Elsevier Science Inc.*

Keywords: metabolic stability, male contraception, PADAM, decision trees, ab initio calculations, cytochrome P450, androgens

INTRODUCTION

In the last few years much attention has been paid to male targets with respect to contraception, Partial Androgen Deficiency in Aging Males (PADAM), and hormone-dependent diseases such as male hypogonadism and prostate cancer. Consequently, many pharmaceutical companies and universities are concentrating their efforts on the discovery and development of steroidal and nonsteroidal orally bioavailable, potent, and selective androgens.

The endogenous androgen testosterone (Figure 1), besides having low potency, possesses a half-life of only a few minutes in the human body, resulting into very low oral bioavailability.[1] In plasma, 98% of testosterone, biosynthesized in testis and adrenals, is protein bound (Color Plate 1). The remaining 2% is available for the activity as well as the inactivation of testosterone. The latter mainly occurs in the liver by action of the cytochrome P450s (CYP3A4) and the 17β-, 3α- and 3β-Hydroxy Steroid Dehydrogenase (HSD) enzyme families. It is generally assumed that the attack of these enzymes on a substrate is driven by an intermediate step involving radical formation.[2]

Further, some amount of the free fraction of testosterone peripherally circulates in organs such as prostate and brain, where the 5α-reductase and aromatase enzymes metabolize it into the active metabolites dihydrotestosterone (DHT) and estradiol (E_2), respectively. DHT and E_2 are further transported in blood where they are up to 99% protein bound, while the remaining free fraction is metabolized in the liver.

While potency and selectivity are no longer critical issues in the development of new androgens, metabolic stability still is. Human liver models are needed to come to a better understanding of the complex mechanisms of the first-pass metabolism of androgens. Over the years, many efforts have been devoted to the development of in vitro drug metabolism models that can be predictive for the corresponding in vivo models.[3–5] In the last few years the great advancement of computer techniques and methods in (Quantitative) Structure Activity Relationship ((Q)SAR) studies[6–9] have led to a fast development of in silico ADME (Absorption, Distribution, Metabolism and Excretion).[10–25]

We report the preliminary results of two studies performed on an in-house data set of steroidal androgens. In the first study, we investigated the capability of a classification technique such as decision trees[26–30] to classify the metabolism rates ($\tau_{1/2}$) determined in human hepatocytes of 32 steroidal

Color Plates for this article are on pages 607–608.

Corresponding author: R. Bursi, Molecular Design & Informatics, N.V. Organon, P.O. Box 20, 5340 BH Oss, The Netherlands. Tel.: +31 (0) 412 661468: Fax: +31 (0) 412 662539.

E-mail address: r.buma@organon.oss.akzonobel.nl (R. Bursi)

655 Avenue of the Americas, New York, NY 10010

1093-3263/01/$–see front matter
PII S1093-3263(01)00089-4

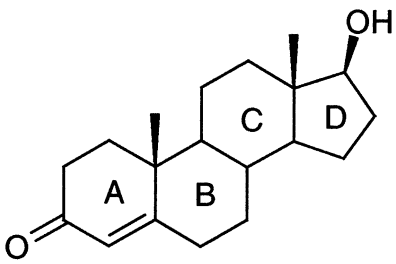

Figure 1. Testosterone. 2D representation. The pseudo-testosterone substrate consists of the C and D rings only.

androgens. In the second study, we derived an electronic model for hydrogen abstraction via radical formation in steroidal androgens by means of ab initio calculations.

METHODS

Decision Trees Analysis

The program CORINA[31] was used to convert the 2D structures of the data set of 32 androgens into 3D. Chemical structures are Organon proprietary information. 23 molecular properties, among which volume, molecular weight (MW), polar surface, number of rotatable bonds, and number of hetero-atoms were calculated by an in-house built program.[32]

The decision-tree program C5.0[33] was used to train the data set. C5.0 constructs decision trees in two phases. A large tree is first grown to fit the data closely and is then "pruned" by removing parts that are predicted to have a relatively high error rate. The pruning option affects the way error rates are estimated and, hence, the severity of pruning; values smaller than the default (25%) cause more of the initial tree to be pruned, while larger values result in less pruning. The Recursive Partitioning (RP) method implemented in C5.0 is confidential and unpublished. The splitting criterion algorithm, however, is similar to the Gain Ratio algorithm implemented in C4.5.[26]

In our study, default minimum leaf size and pruning confidence value were employed. Because of the limited amount of compounds in the training set, a simple distinction between *low* ($\tau_{1/2} \leq 30$ minutes) and *high* ($\tau_{1/2} > 30$ minutes) metabolically stable compounds was made. Classification into three classes *low* ($\tau_{1/2} \leq 30$ minutes), *medium* ($30 < \tau_{1/2} < 60$ minutes), and *high* ($60 \geq$ minutes) was also attempted resulting however in a less satisfactory data set classification, which therefore will not be discussed here. The decision trees error rates were estimated by the n-fold cross-validation procedure.

Metabolic Stability Data

Agarwal and Monder have already described the experimental protocol followed in this study.[34] Cryo-preserved human hepatocytes from healthy young men (25–45 years) were obtained from In Vitro Technologies (Baltimore, MD, USA). The hepatocytes were thawed at 37°C and immediately placed on ice. The cells were washed twice in cold medium: William's medium E, without phenol red, supplemented with glutamax I, gentamicin (50 μg/ml), insulin (1 μmol/l) and hydrocortisone (10 μmol/l). The hepatocytes were seeded at a density of 5×10^5 viable cells/well in a 12-well plate. The incubations were started by the addition of a test compound in medium. The final concentration of the test compound was 100 nmol/l. The incubations were performed at 37°C in an atmosphere of air/O_2/CO_2 (55/40/5) for different interval times (0, 30, 60, and 180 min). The incubations were terminated by pipetting the incubation mixture into one volume of acetone on ice. The acetone was evaporated under a stream of nitrogen, the volume was adjusted to 1.5 ml and centrifuged at 10,000g for 30 min at 4°C. The supernatants were collected for LC-MS/MS analysis.

The parent compound was determined using a Supelcosil LC-8-DB (33×4.0 mm) column and a gradient of methanol in 20 mmol/l ammonium acetate dissolved in formic acid (0.1 mol/l) at a flow rate of 1 ml/min. All mass measurements were performed at an API-3000 mass spectrometer (Perkin-Elmer Sciex, Toronto, Canada).

Half-life times, $\tau_{1/2}$, were determined by interpolation of the amount of parent compound present in the supernatants for each incubation time. These data are Organon proprietary information.

Electronic Model for P450 Oxidations: Hydrogen Abstraction

It is generally assumed that before functionalization, P450s interact with a given substrate via hydrogen abstraction, producing a short-lived radical intermediate substrate.[35] It has been discussed in the literature that an inverse relationship exists between the frequency of metabolism of the most occurring radicals and their corresponding relative stabilities.[36] Based on the Hammond postulate that for certain similar reactions exothermic reactions tend to have earlier transition states that are of lower energy than endothermic reactions; a linear relationship, the Brønsted relationship,[37] can be assumed between the reaction enthalpy ΔH_R and the activation energy $\Delta H^{\#}$ for a given reaction. It has also been shown that this linear relationship can be further improved by adding one extra term, which is the ionization potential of the product radical (IP_{rad}).[35] This extra term, which is the energy required for the extraction of the unpaired electron in the highest occupied molecular orbital of the product radical, accounts for some resonance effects. This means that to derive an operational model, i.e., a model with predictive power, we need to obtain the equilibrium energies of reactants, products, and their corresponding transition states for several hydrogen abstraction reaction profiles.

We simulated the oxidizing enzyme by a methoxy radical approaching the steroidal site of interaction. After hydrogen abstraction, methanol and a steroid radical are the resulting products.

Since we are simulating radical reaction profiles, the equilibrium geometries of reactants, products, and transition states are open-shell systems, i.e., doublets (in this specific case). The spin-unrestricted Hartree-Fock (UHF) theory[38] was therefore used to obtain these geometries by means of the GAUSSIAN 94 program.[39] When dealing with UHF type of calculations, one has to deal with spin contamination. This is due to the nature of the UHF wavefuntion, which is not a true eigenfunction of the total spin operator, unlike exact wavefuntions, which necessarily are. The exact expectation value of $<\hat{S}^2>$ for a doublet is 0.75. If the obtained equilibrium geometries correspond to an expectation value larger than 0.75–0.76, then higher spin states are mixed into the doublet. Since the amount of spin contamination is variable for each calculation, it is

impossible to perform a reliable comparison of calculated energies that are spin contaminated.

The degree of spin contamination generally decreases with increasing size of basis set. We first attempted to obtain equilibrium geometries via the semiempirical approach AM1,[40] which is computationally not very intensive. Spin contamination, however, was largely present. Gaussian-type basis functions were therefore used in our calculations. Only at the 3-21G-basis set level—after spin correction performed by GAUSSIAN—spin contamination was reduced to zero.

Equilibrium geometries of reactants and products could be obtained straightforwardly. On the average, they required 12 hr of CPU time on a SGI R10000 at the 3-21G basis set level. Force field calculations were performed at the end point of each optimization (reactants, products, and transition states). On the average, they required 5 hr of CPU time.

Optimization of transition states turned out to be much more difficult and computationally intensive than reactants and products equilibrium geometry calculations. Ionization potentials were automatically obtained at the end of the optimizations from Koopman's theorem.[38]

In this study we investigated seven substrates for which hydrogen abstraction was performed at the 17β-hydroxyl substituent position of a pseudo-steroidal skeleton (see Figure 1). Chemical structures are Organon proprietary information. Finally, multiple linear regression[41] was applied to the resulting ΔH_R, $\Delta H^{\#}$, and IP_{rad}.

RESULTS AND DISCUSSION

To maintain their competitiveness, pharmaceutical companies must aim at faster and more efficient drug development trajectories. Inappropriate ADME and toxicity (T) profiles are among the most frequent causes of failure of potential drugs in preclinical development. It is therefore not surprising that early ADMET, i.e., ADMET within Lead Optimization, is now receiving high priority in the final phase of the research trajectory.

The development of orally active steroidal and nonsteroidal androgens is strongly affected by ADMET. The known bioavailable androgens suffer from drawbacks such as the need for frequent administration in substantial doses or the occurrence of liver damage on the long term (methyl testosterone). Clearly, first-pass metabolism is a severe bottleneck for the development of new androgens.

We have attempted to tackle this problem with two different approaches and corresponding computational techniques. On the one hand, a classification method like decision trees was applied to an in-house data set of steroidal androgens to test whether a meaningful structure–pharmacokinetic model could be obtained. On the other hand, an ab initio electronic model for hydrogen abstraction on the 17β-hydroxyl substituent on a pseudo-steroidal substrate was derived to obtain ideas on steric and/or electronic effects on the transition state energies, i.e., on the relative rates of metabolism of those reactions.

Decision Trees Analyses

The decision tree obtained is shown in Color Plate 2. The tree starts from the lower right corner of the picture and develops in branches up to the upper left corner. Volumes, shape, number of bonds, and surface were the resulting classifiers of this data set. Compounds that possess a descriptor value larger than a given threshold value will be classified in a cluster on the left side of the variable descriptor (below the tree), while the rest of the compounds will be further classified by the next most important variable on the right side (above the tree). For instance, the molecular volume resulted as the most important discriminating descriptor. Five compounds possess a volume value larger than a given threshold that distinguishes these five compounds from the rest of the data set. These compounds are correctly classified as having high metabolic stability. Clearly, the decision tree suggests that the volume of the steroidal substrates must exceed a certain size (the cavity size of the enzyme) if high metabolic stability is desired. Among the rest of the compounds, the shape (described by the principal moments of inertia) becomes the second most important variable. Six compounds having a shape value larger than a certain threshold value were successfully classified separately from the rest as compounds with low metabolic stability. These six compounds belong to the same chemical class. Clearly, the structural/electronic feature(s) that characterize them do not improve their metabolic stability with respect to testosterone.

The next branching of the tree occurs by the number of bonds. Subsequently, the surface and the volume once again emerge as important classifiers. The threshold values of the classifier volume unambiguously differ from each other in the initial and final phases of this classification.

Interestingly, three of the four classifiers are "steric" descriptors, that is, descriptors of size and shape. It was rewarding to see that testosterone (low) and methyl testosterone (high) were correctly classified in the upper branch of the tree split by the surface variable.

Overall, only two compounds of the entire data set were classified wrongly. They were clustered in the last two leaf nodes separated by the volume. However, their half-lives are very close to the arbitrary threshold of 30 min. In this case, the erroneous classification fell within the experimental error limits.

Further, it was encouraging to see that the three clusters with high metabolic stability contained compounds with very different steric and/or electronic modification(s) of the same steroidal skeleton. In our view, this result suggests that high metabolic stability can in fact be achieved for the same substrate in different ways and that as long as we can introduce enough structural and/or electronic diversity in our compounds, the chance of identifying a "winner" will be very high.

The decision tree's error rate was estimated by cross-validation (n-fold = 8) to be 50% with an estimated variability (Standard Error) of 6.7%. This result clearly shows that although qualitatively very useful, this model requires many more structures and corresponding experimental data to become really operational, i.e., reliably predictive. For this purpose, the metabolic stability of many more steroidal androgens using human hepatocytes is currently being measured in our laboratories.

Ab Initio Electronic Model for Hydrogen Abstraction

An example of the type of reaction profiles that were simulated is given in Color Plate 3. The equilibrium geometries of either the reactants (a pseudo-testosterone substrate, AH, in equilibrium with a methoxy radical, B*) or the products (a pseudo-

testosterone substrate radical, A*, and methanol, BH) were first obtained. Consistently, the same minima for both reactants and products had to be carefully considered, since more than one equilibrium geometry could be identified on the potential energy surface between AH and B* and A* and BH. With the equilibrium geometries of reactants and products, the corresponding transition states A—H—B could be identified.

As previously discussed in the Methods section, care had to be taken for spin contamination. In the first instance, our calculations at the AM1 semiempirical level were to a large extent spin contaminated ($<S^2> >> 0.75$). We switched therefore to the Gaussian-based basis functions and started with the smallest basis set STO-3G. As expected, at this basis-set level spin contamination was still present even after spin contamination correction (performed by the program Gaussian when the full electronic wavefunction is chosen). We therefore increased the basis set to 3-21G, leading for all calculations to a $<S^2>$ equal to 0.75–0.76 after spin correction. The results discussed hereafter were all obtained at the HF 3-21G basis-set level.

At this point we wish to indicate that we did not consider the entire steroidal skeleton in our calculations, but only a "pseudo" steroidal skeleton, i.e., only the C and D rings (see Figure 1). This choice was made to reduce the computational efforts and speed up the calculations. It might be argued that the A and B rings (with their steric and/or electronic modifications) can also influence the hydrogen abstraction mechanism at the 17 position. We agree that this cannot be excluded a priori. However, the main aim of this study was to look for a trend and obtain a qualitative comparison of modified testosterone-like substrates with respect to testosterone. Further, it is well known that at this level of theory (HF 3-21G), conjugation (3-21G) is poorly represented and electron correlation effects (HF) are completely neglected. The chance, therefore, that by inclusion of the A and B rings the resulting trend might change is very small.

We have simulated the hydrogen abstraction from the hydroxyl group at the 17β position of a pseudo-steroidal skeleton for 7 reaction profiles and subsequently monitored the effect(s) of skeleton steric and/or electronic modifications on the $\Delta H^{\#}$ and ΔH_R of these reactions. Clearly, modified substrates with a larger $\Delta H^{\#}$ than the reference substrate testosterone are expected to show higher metabolic stability than testosterone, while comparable or lower activation energies are expected to lead to lower metabolic stability.

Multiple linear regression analysis (see Color Plate 4) was performed on the resulting $\Delta H^{\#}$, ΔH_R and IP_{rad} leading to the following relationship: $\Delta H^{\#} = 29.42 + 0.86\ \Delta H_R - 5.26\ IP_{rad}$. A cross-validated r^2 coefficient of 0.64 and a final r^2 coefficient of 0.82 were obtained.

Clearly, this model needs to be expanded and improved with many more reaction profiles to ensure statistically sounder and more robust predictions. These preliminary results, however, provide already useful information (confirmed by the experimental feedback) for the design of new androgens with potentially improved metabolic stability. These compounds have been synthesized and are currently being tested in human hepatocyte assays.

CONCLUSIONS

Androgens with favorable pharmacokinetics after oral administration and without undesired side-effects on the prostate and liver are highly desired for male contraception and the treatment of PADAM.

One major obstacle to the achievement of such an objective is the hepatic first-pass metabolism.

In this study we have reported the preliminary results of two in silico models for the metabolism of steroidal androgens: a structure–pharmacokinetic classification of human hepatocytes half-lives with decision trees and a hydrogen abstraction electronic model by means of ab initio calculations. Both methods produced good results. While the decision trees model can be used as rough filter for steric properties of steroidal androgens, the ab initio model provides insight into the electronic mechanisms most likely involved in the intermediate steps of the metabolic oxidation reactions.

Our work shows that, alongside with in vitro and/or in vivo models, in silico drug metabolism models can contribute significantly to the design of metabolically stable androgens.

REFERENCES

1 Rommerts, F.F.G. An overview of biosynthesis, transport, metabolism and nongenomic actions. In: *Testosterone Action Deficiency Substitution*, Nieschlag, E., and Behre, H.M., Eds., Springer-Verlag, Berlin Heidelberg, 1990, pp. 1–31

2 Rogozkin, V.A. *Metabolism of Anabolic Androgenic Steroids.* CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida, 1991

3 Houston, J.B. Utility of in vitro drug metabolism data in predicting in vivo metabolic clearence. *Biochem. Pharmacol.* 1994, **47**, 1469–1479

4 Bertz, R.J., and Granneman, G.R. Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin. Pharmacokinet.* 1997, **32**, 210–258

5 Iwatsubo, T., Hirota, N., Ooie, T., Suzuki, H., Shimada, N., Chiba, K., Ishizaki, T., Green, C.E., Tyson, C.A., and Sugiyama, Y. Prediction of in vivo drug metabolism in the human liver from in vitro metabolism data. *Pharmacol. Therap.* 1997, **73**, 147–171

6 Hansch, C., and Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology.* American Chemical Society: Washington DC, 1995, Vol. 1

7 Kubinyi, H. *3D QSAR in Drug Design: Theory, Methods and Applications.* ESCOM, Leiden, 1993

8 Kubinyi, H., Folkers, G., and Martin, Y.C. *3D QSAR in Drug Design: Ligand–Protein Interactions and Molecular Similarity.* Kluwer Academic Publishers, Dordrecht, 1998

9 Kubinyi, H., Folkers, G., and Martin, Y.C. *3D QSAR in Drug Design: Recent Advances.* Kluwer Academic Publishers, Dordrecht, 1998

10 Van de Waterbeemd, H., and Kansay, M. Hydrogen-bonding capacity and brain penetration. *Chimia* 1992, **46**, 299–303

11 Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 1997, **23**, 3–25

12 Palm, K., Stenberg, P., Luthman, K., and Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* 1997, **14**, 568–571
13 Norinder, U., Österberg, T., and Artursson, P. Theoretical calculation and prediction of Caco-2 cell permeability using MolSurf parametrization and PLS statistics. *Pharm. Res.* 1997, **14**, 1786–1791
14 Raevsky, O.A., and Schaper, K.-J. Quantitative estimation of hydrogen bond contribution to permeability and absorption processes of some chemicals and drugs. *Eur. J. Med. Chem.* 1998, **33**, 799–807
15 Wessel, M.D., Jurs, P.C., Tolan, J.W., and Muskal, S.M. Prediction of human intestinal absorption of drug compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* 1998, **38**, 726–735
16 Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* 1999, **88**, 807–814
17 Kelder, J., Grootenhuis, P.D.J., Bayada, D.M., Delbressine, L.P.C., and Ploemen, J.-P. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* 1999, **16**, 1514–1519
18 Alifrangis, L.H., Christensen, I.T., Berglund, A., Sandberg, M., Hovgaard, L., and Frokjaer, S. Structure–property model for membrane partitioning of oligopeptides. *J. Med. Chem.* 2000, **41**, 103–113
19 Ertl, P., Rohde, B., and Selzer, P. Fast calculation of molecular polar surface area as a sum of fragments-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* 2000, **43**, 3714–3717
20 Levin, V.A. Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability. *J. Med. Chem.* 1980, **23**, 682–684
21 Young, R.C., Mitchell, R.C., Brown, T.H., Ganellin, C.R., Griffith, R., Jones, M., Rana, K.K., Saunders, D., Smith, I.R., Sore, N.E., and Wilks, T.J. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H_2 receptor histamine antagonists. *J. Med. Chem.* 1988, **31**, 656–671
22 Abraham, M.H., Chadha, H, S., and Mitchell, R.C. Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci.* 1994, **83**, 1257–1268
23 Lombardo, F., Blake, J.F., and Curatolo, W.J. Computation of brain–blood partitioning of organic solutes via free-energy calculations. *J. Med. Chem.* 1996, **39**, 4750–4755
24 Feher, M., Sourial, E., and Schmidt, J.M. A simple model for the prediction of blood–brain partitioning. *Int. J. Pharm.* 2000, **201**, 239–247
25 Crivori, P., Cruciani, G., Carrupt, P.-A., and Testa, B. Predicting blood–brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* 2000, **43**, 2204–2216
26 Quinlan, J.R. *C4.5: Programs for Machine Learning.* Morgan Kaufmann: San Mateo, CA, 1993
27 Hawkins, D.M., Young, S.S., and Rusinko, A., III. Analysis of a large structure–activity data set using recursive partitioning. *Quant. Struct. Act. Relat.* 1997, **16**, 296–302
28 Cho, S.J., Shen, C.F., and Hermsmeier, M.A. Binary formal inference-based recursive modeling using multiple atom and physicochemical property class pair and torsion descriptors as decision criteria. *J. Chem. Inf. Comput. Sci.* 2000, **40**, 668–680
29 Izrailev, S., and Agrafiotis, D. A novel method for building regression tree models for QSAR based on artificial ant colony systems. *J. Chem. Inf. Comput. Sci.* 2000, in press
30 Miller, D.W. Results of a new classification algorithm combining k nearest neighbors and recursive partitioning. *J. Chem. Inf. Comput. Sci.* 2000, in press
31 Sadowski, J., Wagener, M., and Gasteiger, J. CORINA: Automatic generation of high-quality 3D-molecular models for application in QSAR. In: *10th Europian Symposium on Structure–Activity Relationships: QSAR and Molecular Modelling.* Sanz, F., Ed. Prous Science Publishers, 1994
32 Bayada, D.M., Hamersma, H., and van Geerestein, V.J. Molecular diversity and representativity in chemical databases. *J. Chem. Inf. Comput. Sci.* 1999, **39**, 1–10
33 C5.0, release 1.08a; RuleQuest Research Pty Ltd., St Ives NSW, Australia (http://www.rulequest.com)
34 Agarwal, A.K., and Monder, C. In vitro metabolism of 7α-methyl-19-nortestosterone by rat liver, prostate, and epididymis. *Endocrinology* 1988, **123**, 2187–2193
35 Korzekwa, K.R., Jones, J.P., and Gillette, J.R. Theoretical studies on cytochrome P-450 mediated hydroxilation: A predictive model for hydrogen atom abstraction. *J. Am. Chem. Soc.* 1990, **112**, 7042–7046
36 Korzekwa, K.R., Grogan, J., DeVito, S., and Jones, J.P. Electronic models for cytochrome P450 oxidations. *Adv. Exp. Med. Biol.* 1996, **387**, 361–369
37 Kresge, A.J. The Brønsted relation – recent developments. *Chem. Soc. Rev.* 1973, **2**, 475–503
38 Hehre, W.J., Radom, L., v.R. Schleyer, P., and Pople, J. A. In: *Ab initio molecular orbital theory.* John Wiley & Sons, Inc., New York, 1986
39 Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T.A., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Cioslowski, J., Stefanov, B.B., Nanayakkara, A., Challacombe, M., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Repogle, E.S., Gomperts, R., Martin, R.L., Fox, D. J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C., and Pople, J.A. GAUSSIAN 94, Gaussian, Inc., Pittsburgh PA, 1995
40 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 1985, **107**, 3902–3909
41 Wold, S., Albano, C., Dunn, W.J., Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., and Sjostrom, M. In: *Chemometrics: Mathematics and Statistics in Chemistry.* Kowalsi, B., Ed., Dordrecht, 1984