Journal of Computer-Aided Molecular Design **17:** 849–859, 2003.

Prediction of *in vitro* metabolic stability of calcitriol analogs by QSAR

Berith F. Jensen[a], Morten D. Sørensen[b], Anne-Marie Kissmeyer[b], Fredrik Björkling[b], Kim Sonne[b], Søren B. Engelsen[a] & Lars Nørgaard[a,*]

[a]*Centre for Advanced Food Studies, Food Technology, Department of Dairy and Food Science, The Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark*

[b]*LEO Pharma A/S, Industriparken 55, DK-2750 Ballerup, Denmark*

Received 3 July 2003; accepted in revised form 21 November 2003

Key words: calcitriol analogs, metabolic stability, Partial Least Squares Regression (PLS), Quantitative Structure-Activity Relationships (QSAR), QikProp, BCUT_descriptors, variable selection

Summary

The metabolic stability of a drug is an important property for potential drug candidates. Measuring this property, however, can be costly and time-consuming. The use of quantitative structure-activity relationships (QSAR) to estimate the *in vitro* stability is an attractive alternative to experimental measurements. A data set of 130 calcitriol analogs with known values of *in vitro* metabolic stability was used to develop QSAR models. The analogs were encoded with molecular structure descriptors computed mainly with the commercial software QikProp and DiverseSolutions. Variable selection was carried out by five different variable selection techniques and Partial Least Squares Regression (PLS) models were generated from the 130 analogs. The models were used for prediction of the metabolic stability of 244 virtual calcitriol analogs. Twenty of the 244 analogs were selected and the *in vitro* metabolic stability was determined experimentally. The PLS models were able to predict the correct metabolic stability for 17 of the 20 selected analogs, corresponding to a prediction performance of 85%. The results clearly demonstrate the utility of QSAR models in predicting the *in vitro* metabolic stability of calcitriol analogs.

Introduction

Calcitriol (1α,25-dihydroxyvitamin D_3) is the physiologically active form of vitamin D and is essential for regulation of the extracellular concentrations of calcium and phosphorus. Moreover, calcitriol is found to have great impact on the regulation of cell growth (inhibition of proliferation and induction of differentiation) of various cell types, e.g. cancer cells, skin cells, and cells of the immune system [1–5]. In humans, calcitriol has shown a positive effect in the treatment of psoriasis and cancer [6, 7]. However, an overdose of calcitriol can lead to elevated levels of calcium and phosphorus in the blood, posing a risk of inducing soft tissue calcifications [8]. The pharmaceutical industry has shown an increasing interest in the search for analogs of calcitriol. The goal is to create analogs with strong effects on the cell growth regulation combined with a reduced activity on the calcium homeostasis [9]. The chemical structure of calcitriol is shown in Figure 1.

Poor absorption, distribution, metabolism and/or excretion (ADME) properties are some of the main reasons for terminating the development of chemical drug candidates [10]. This recognition has resulted in the development and application of a wide range of *in vitro* screening tests for classification of compounds with respect to ADME properties during the early drug discovery process. Today, the absorption and metabolic properties of new drug candidates are normally determined first in an *in vitro* assay and then *in vivo* [11]. The application of *in silico* technology offers considerable potential for reducing the number of experimental studies required for compound selection and for improving the success rate. *In silico* predictions have been investigated for several ADME

*To whom correspondence should be addressed. Fax: +45-35283245; E-mail: lan@kvl.dk

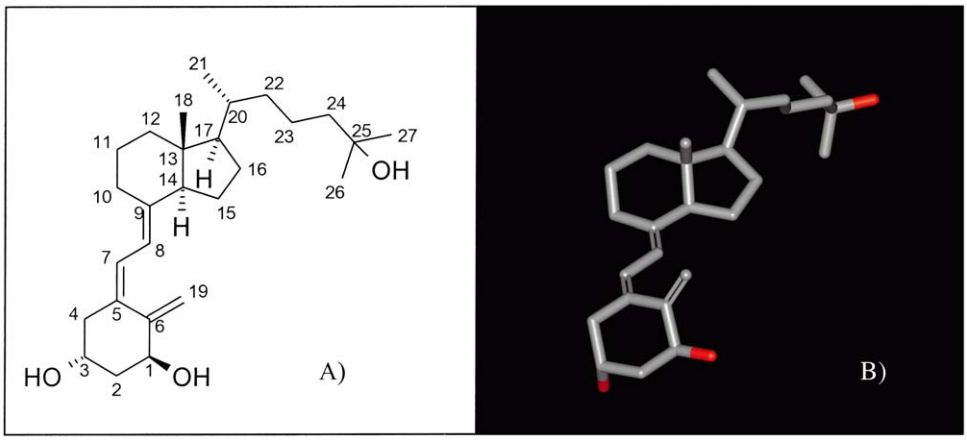

Figure 1. (A) The structure and labeling of calcitriol. (B) Heavy atom representation of the 3D structure of calcitriol.

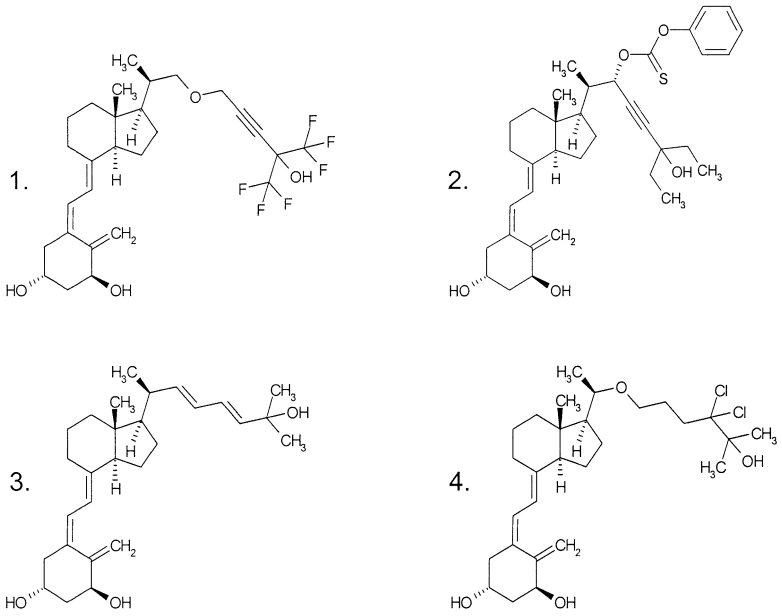

Figure 2. Four calcitriol analogs detected as outliers.

parameters, particularly absorption, distribution and metabolism [12]. There is a growing consensus that the *in silico* predictions are comparable with those made using *in vitro* tests, with the advantage of being able to significantly reduce experimental effects in the early screening phase of drug development; i.e., it is possible to screen virtual compounds [13]. The early assessment of ADME properties may help pharmaceutical scientists to select the best candidates for development as well as to reject candidates with a low probability of success.

In the present study six QSAR [14] models were developed to predict the metabolic stability of calcitriol analogs. The models were generated using Partial Least Squares Regression [15] and the descriptors selected using different variable selection methods. The models and selected descriptors are discussed and validated by a test set.

Experimental

In vitro assay

The set of 130 calcitriol analogs and their experimentally derived *in vitro* metabolic stability values were determined at LEO Pharma A/S, Denmark. The *in vitro* metabolic stability of the 130 analogs was de-

Table 1. Division of calcitriol analogs into three metabolic classes.

Metabolic stability	Class	$S9_m$[a]	Data set[b]	Training set[b]	Test set[b]
Unstable	1	<10%	56	38	18
Stable	2	10–35%	31	20	11
Very stable	3	>35%	43	29	14

[a]Non-metabolized calcitriol analogs.
[b]Division of calcitriol analogs.

termined by the following method: Each analog was incubated with a liver-metabolizing system containing the post-mitochondrial liver fraction (S9) from rats and a mixture consisting of 8 mM $MgCl_2$, 33 mM KCl, 5 mM glucose 6-phosphate and 4 mM NADP in 0.1 M phosphate buffer, pH 7.4. The final concentration of the test compound was 4 μM. Incubations were performed at 37 °C and after 1 hour the metabolism was stopped by addition of acetonitrile [16]. After centrifugation the concentration of the parent compound was determined by reversed phase HPLC with UV detection according to the method described by Kissmeyer et al. [17]. The part of the compound that had not been metabolized was calculated in percent ($S9_m$). Before the calibrations were performed, the analogs were divided into three metabolic classes as shown in Table 1. The critical values used for dividing the analogs into three metabolic groups were chosen to be $S9_m = 10\%$ and $S9_m = 35\%$. These values are empirical and were selected on the basis of the research and development program for calcitriol analogs at LEO Pharma A/S. The program includes treatment of skin diseases, cancer, immunological disorders, and bone mineralization. This means both topical and systemic treatment, which in addition to pharmacological requirements demands specific metabolic profiles. In general, calcitriol analogs for topical treatment should have a low metabolic stability (belong to Class 1) once they enter the systemic circulation, whereas analogs for systemic treatment should be rather stable (belong to Class 3) [18].

Descriptor generation

A total of 201 descriptors were generated for each of the 130 compounds. A set of 161 standard 3D BCUT_descriptors was computed with DiverseSolutions version 4.0.6 (Tripos, Inc., Missouri, USA). BCUT_descriptors are an extension of Burden's parameters [19] which are based on a combination of the atomic number for each atom and a description of the nominal bond-type for adjacent and non-adjacent atoms. The BCUT_descriptors incorporate various functions of intramolecular distances and one of the following atomic properties: atomic charge, polarizability, hydrogen bond donor- and acceptor-abilities, corresponding to the electrostatic, dispersion and H-bonding modes of intermolecular interactions [20]. A set of 31 physico-chemical descriptors was computed using QikProp version 1.6 (demo version kindly provided by Schrödinger, LLC, New York, USA). The last nine descriptors were generated by manual counting of topological features. The BCUT descriptors are listed in the supplementary material, while the QikProp descriptors and manual descriptors are shown in Tables 2 and 3. The 3D conformations used in the calculation of BCUT and QikProp descriptors were generated using the program MacroModel version 7.1 (Schrödinger, LLC). For all compounds, 100 conformations were calculated using the MacroModel Monte Carlo routine. The conformations were energy-minimized with the truncated Newton conjugate gradient method [21] using the MMFF94s force field [22, 23]. The solvent was set to water with a continuum solvation using a Generalized Born/Surface Area (GB/SA) [24] approach (the SLVNT command of MacroModel). The lowest energy conformation was used for descriptor calculation.

Variable selection

In QSAR studies variable selection is an important element. Selecting a preferred set of molecular descriptors is crucial to obtaining a predictive and robust QSAR model. In this study, five variable selection techniques were used: Forward Stepwise Selection (FSS) [25, 26], Principal Variables (PV) [27, 28], Generating Optimal Linear PLS Estimations (GOLPE) [29, 30], Genetic Algorithm (GA) [31] and Jackknife [32]. The driving force behind each algorithm is the continuous reduction of the root mean square error values (RMSE values) from descriptor subset

Table 2. Descriptors calculated by QikProp.

Code	Description of the QikProp-descriptors
# stars	Drug-likeness
# rotor	The number of rotatable bonds
# rctvFG	The number of reactive functional groups
CNS	The predicted central nervous system activity on a −2 (inactive) to +2 (active) scale
MW	The molecular weight of the molecule
dipole	The dipole moment for the molecule
SASA	The total solvent-accessible surface area in Å^2
FOSA	The hydrophobic component of the SASA (saturated carbon and attached hydrogen)
FISA	The hydrophilic component of the SASA (SASA on N, O and H on heteroatom)
PISA	The pi (carbon and attached hydrogen) component of the SASA
WPSA	The weakly polar component of the SASA (halogens, P and S)
volume	The total solvent-accessible volume in Å^3
donorHB	The estimated number of hydrogen bonds that would be donated by solute to water molecules in an aqueous solution
accptHB	The estimated number of hydrogen bonds that would be accepted by solute from water molecules in an aqueous solution
dip^2/V	Kirkwood–Onsager dipole solvent index ((dipole moment)^2/volume)
ACxDN^.5/SA	Index of cohesive interactions in solids
glob	Globularity (4*pi*r^2/SASA)
QPpolrz	The predicted polarizability in Å^3
QPlogPC16	The predicted log of the hexadecane/gas partition coefficient
QPlogPoct	The predicted log of the octanol/gas partition coefficient
QPlogPw	The predicted log of the water/gas partition coefficient
QPlogPo/w	The predicted log of the octanol/water partition coefficient
QPlogS	The predicted aqueous solubility
BIPcaco	The predicted apparent Caco-2 cell permeability in nm/s. Boehringer-Ingelheim scale
AffyPCaco	The predicted apparent Caco-2 cell permeability in nm/s. Affymax scale
QPlogBB	The predicted log of the brain/blood partition coefficient
AffyPMDCK	The predicted apparent MDCK cell permeability in nm/s. Affymax scale
QPlogKp	The predicted log of the skin permeability
IP(eV)	The PM3[23] calculated ionization potential
EA(eV)	The PM3[23] calculated electron affinity
# metabol	The number of likely metabolic reactions that are listed in the propout file

to descriptor subset. Subsets of descriptors that give lower RMSE values are favored. RMSE in combination with the correlation coefficient (r) is used as a measure of model performance [27]. RMSE is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum (y_{pred} - y_{ref})^2}{N}}$$

where y_{pred} is the predicted value, y_{ref} is the laboratory/measured value, and N is the number of samples. RMSECV is RMSE calculated from the cross-validated samples [33], while r_{cv} are the correlation coefficients for the same situation. RMSEP is calculated in the same way from the independent test set.

Prior to variable selections a PLS regression model was established and outliers were detected on the basis of leverage vs. residual and U vs. T plots. All possible outliers were removed from the data set in order to reduce their influence on the variable selection.

Jackknife was performed with Unscrambler version 7.6 (CAMO A/S, Norway). FSS, PV, GOLPE and GA were performed with MATLAB version 6.1 (Math Works, Inc. Natick, MA) installed with PLS_Toolbox version 2.0.0b (Wise & Gallagher; Eigenvector Technologies, Manson, WA). FSS and PV were performed using in-house programs [27], while GOLPE was programmed in MATLAB language by the authors. The command 'CORDEXCH' (Statistics toolbox, Math Works, Inc., version 3) was used to produce

Table 3. Manually determined descriptors.

Code	Description of the manually determined descriptors
# double	The number of double bonds in the side chain[a]
# triple	The number of triple bonds in the side chain[a]
epi	If the stereochemistry at carbon 20 is epi or normal (1 = epi and 0 = normal)
# ether	The number of ether bonds in the side chain[a]
# thioether	The number of thioether bonds in the side chain[a]
# methyl	The number of methyl groups terminal in the side chain[a]
# ethyl	The number of ethyl groups terminal in the side chain[a]
# aromatic	The number of aromatic groups in the side chain[a]
blocked C24	If carbon 24 is blocked or not (1 = totally blocked, 0.5 = partly blocked, −0.5 = hydroxylated (blocked by a OH-group)

[a]The antenna side chain in calcitriol attached to carbon 17.

Table 4. Settings used in the GA function 'GENALG' in MATLAB.

Items	Settings
Population size	256
Window width	1
% Initial terms	10
Maximum generations	500
% At convergence	20
Mutation rate	0.005
Crossover	Double
Regression choice	PLS
# Of latent variables	10
Cross-validation	Random
Number of subsets	13
Number of iterations	5

the design matrix in GOLPE and a cross-validated NIPALS routine to generate the PLS regression models. GOLPE was performed with 500 combinations and three repetitions and the models were validated with segmented cross-validation with 9–10 analogs in each segment. In the PLS_Toolbox the GA interface (GENALG) requires the user to set a number of parameters. The settings used in this paper are listed in Table 4. In GA, three repetitions were performed and variables were selected for the subset, if they were represented in more than 75% of the models in two out of three runs.

Multivariate data analysis

The multivariate method used in this study was Partial Least Squares Regression (PLS). The purpose of the PLS regression is to build a linear model which enables prediction of a desired characteristic (**y**) from a matrix containing mutable descriptors (**X**). In matrix notification the linear model is $\mathbf{y} = \mathbf{Xb}$, where **b** contains the regression coefficients that are determined during the calibration step [34]. The Unscrambler version 7.6 (CAMO A/S, Norway) was used for the PLS modeling. All models were developed on the basis of the selected variables (**X**) and the response variable *in vitro* metabolic stability (**y**). **X** and **y** were autoscaled and the models were validated with segmented cross-validation with 9–10 analogs in each segment.

Before the calibrations were performed, the 130 analogs from the data set were sorted by the value of $S9_m$ and divided into a training set consisting of 87 analogs and a test set consisting of 43 analogs. The data set was divided, so every third analog was selected for the test set and the rest of the analogs went to the training set. For every subset of variables one PLS regression model was generated based on the 87 analogs from the training set. The models were cross-validated and then externally validated with the model independent test set for which the *in vitro* metabolic stability was known in advance. For each model the RMSECV (cross-validation) and RMSEP (independent test set) values were calculated and it was evaluated if the models were overfitting or if it was safe to re-estimate the models using all 130 analogs.

In case of no overfit the re-estimated models based on 130 analogs were used for prediction of a data set consisting of 244 analogs for which the *in vitro* metabolic stability was unknown. On the basis of the predictions made by the models, 20 of the 244 analogs were selected and the *in vitro* metabolic stability was subsequently determined. As previously men-

tioned, calcitriol analogs that belong to Class 1 and 3 are of greater value to the medicinal industry, as they can be used for topical and systemic treatments, respectively. For that reason almost all of the 20 selected analogs were predicted to Class 1 or 3. It was aimed to select analogs that were predicted to the same class in as many models as possible. In addition, the *in vitro* metabolic stability predicted by the models was compared to the measured *in vitro* metabolic stability.

Results and discussion

Result of the variable selections

Before the variable selections were performed, five analogs were detected as outliers. Four of the five outliers are shown in Figure 2[1].

The side chain of outlier 1, 2 and 5 (not shown) is more complex than the side chain of many of the other analogs in the data set. Outlier 3 looks very similar to the other analogs in the data set, but it is predicted to Class 3 by the models and determined to Class 1 in practice (*in vitro*). Moreover, outlier 4 is the only analog in the data set which has dichloro attached to one of the carbon atoms in the side chain. The outliers were excluded from the data set prior to the application of the variable selections methods. The variables selected from FSS, PV, GOLPE, GA and Jackknife are shown in Table 5.

Different variables were selected by the five selection methods. However, the descriptors #double, epi, #ether, QlogS, IP(eV) and #metabol were chosen in at least three of the five selection methods. The most consistently chosen descriptors were epi and #ether, which were selected by all the selection methods. A PLS regression model based on the six commonly chosen descriptors – #double, epi, #ether, QlogS, IP(eV) and #metabol – was generated and the loadings of the model were distributed as shown in Figure 3. The figure shows that of the six descriptors, #double, epi and #ether have loadings furthest away from zero in the first latent variable (LV1) and for that reason they are the most important for the modeling of $S9_m$. Therefore, a subset (Mix3) consisting of these three variables was included in the following analysis. The fact that these three descriptors have a high correlation with the *in vitro* metabolic stability of calcitriol analogs is in accordance with previous studies. 20-epi analogs are known to have a decreased metabolic stability compared to 20-normal analogs. This is partly caused by a decreased affinity to the vitamin D binding protein (DBP) which will give an increased access for degradation in the blood stream and liver [35, 36]. Furthermore, analogs with an ether linkage in the side chain are found to be less metabolically stable, while analogs with one or more double bonds in the side chain are found to have higher metabolic stability [37]. One or more double bonds in the side chain prevent hydroxylation at these sites, implying an increased stability. However, there is no obvious explanation for the decreased stability of ether containing analogs as compared to analogs with an aliphatic side chain.

[1] Due to confidentiality the last outlier cannot be shown.

Results of the Multivariate Data Analysis

Table 6 shows the results of the multivariate regression analysis performed on 87 analogs and the subsequent validation with 43 analogs. The outliers shown in the table are different from the outliers shown in Figure 2. As seen in Table 6, the RMSEP values are slightly lower than or similar to the RMSECV values. This indicates that the models are not overfitting, and it is therefore safe to pool the training and independent test set and subsequently re-estimate the models with all 130 analogs with the same number of significant components as for the model based on 87 samples.

The re-estimated models were used for prediction of a data set consisting of 244 calcitriol analogs. The results are shown in Table 7. The models can be divided into two groups. Group A consists of the three models based on subsets selected by Jackknife, FS and Mix3. In these models 16–21% of the test analogs are predicted to Class 1, 27–35% to Class 2 and about 50% of the test analogs are predicted to Class 3. Group B consists of the three models based on subsets selected by PV, GOLPE and GA; here 2–4% of the test analogs are predicted to Class 1, 12–23% to Class 2 and 73–86% of the test analogs are predicted to Class 3. In both groups A and B the models assign the majority of compounds to the same classes. For example, 90% of the compounds assigned to Class 3 by the Mix3 model are also assigned to Class 3 by the Jackknife and the FS model. The results show that it may be an advantage to predict a new data set in several models instead of one particular model. The distribution of the test analogs in the models in Group A seems most reliable, as it seems very unlikely that only 2–4% of the test analogs should belong to Class 1, when calcitriol analogs in general are designed to have

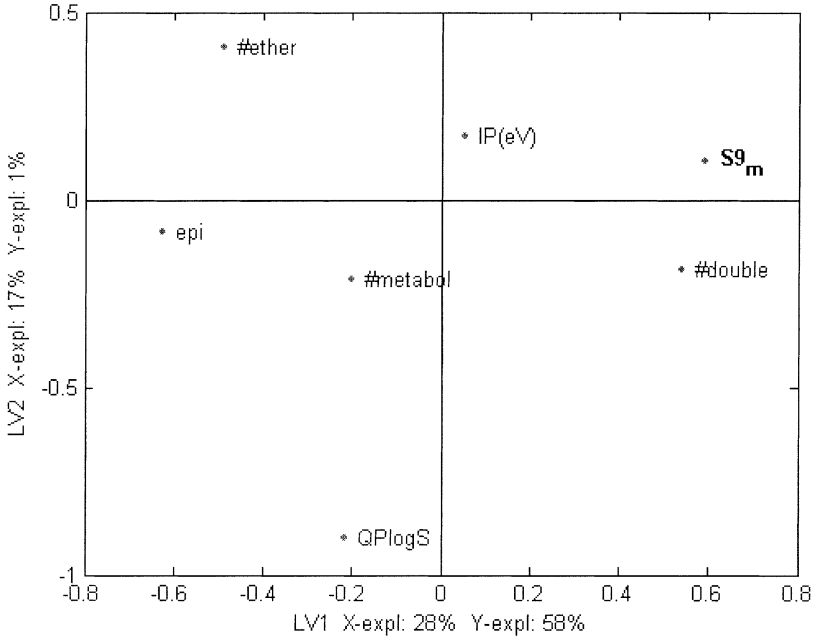

Figure 3. Loading plot from a PLS model based on 125 calcitriol analogs (5 analogs were removed as outliers) and the following descriptors: #double, epi, #ether, QlogS, IP(eV) and #metabol. Only the first latent variable (LV1) is important for the prediction of $S9_m$.

either very low or very high metabolic stability. On the basis of the predictions made by the models in Group A, 20 of the 244 analogs were selected for determination of *in vitro* metabolic stability. All except one of the 20 analogs were predicted to Class 1 or 3. The chemical structures of 10 representative analogs from the selected set are shown in Figure 4[2]. The analogs that are predicted as unstable (analogs B–J) all have at least one ether bond in the side chain and the stereochemistry at carbon 20 is epi. The analogs that are predicted as very stable (analogs K–T) all have normal stereochemistry at carbon 20. Furthermore, seven analogs (K–Q) have two or three double bonds in the side chain. This correlates very well with the theory mentioned in the previous section.

In Table 8 the in *vitro* metabolic stability predicted by the models is compared to the *in vitro* metabolic stability determined in practice. Of the 20 selected analogs from the data set the predicted stability of 14 analogs match the metabolic stability determined in practice. For three analogs (E, G and H) the difference between the predicted and determined stability is not larger than the standard deviation for the *in vitro* assay and the prediction may be assumed correct. For three analogs (B, M and T) there is no match between the predicted and determined metabolic stability: The models predict analog B to be unstable, because it has an ether bond and the stereochemistry at carbon 20 is epi. In practice, $S9_m$ is determined to 25%, which may be because carbon 25 is hydroxylated, protecting from primary oxidation at this position. Both analog M and T are by the models predicted to be very stable. The stereochemistry at carbon 20 is normal in both analogs. Analog M has three double bonds in the side chain, while analog T has none. However, in practice $S9_m$ is determined to be 5% and 16% for analog M and T, respectively. There is no obvious explanation for the different results obtained *in silico* versus *in vitro* for these compounds. The validation shows that the prediction power of the models in Group A is 85% (17*100/20).

[2]Due to confidentiality the remaining analogs in the selected set cannot be shown.

Conclusions

A data set of 130 calcitriol analogs with known values of *in vitro* metabolic stability was used to develop QSAR models. The analogs were represented by 9 manually determined descriptors and 192 molecular structure descriptors computed with the commercial software QikProp and DiverseSolutions using a very pragmatic approach to represent conformational

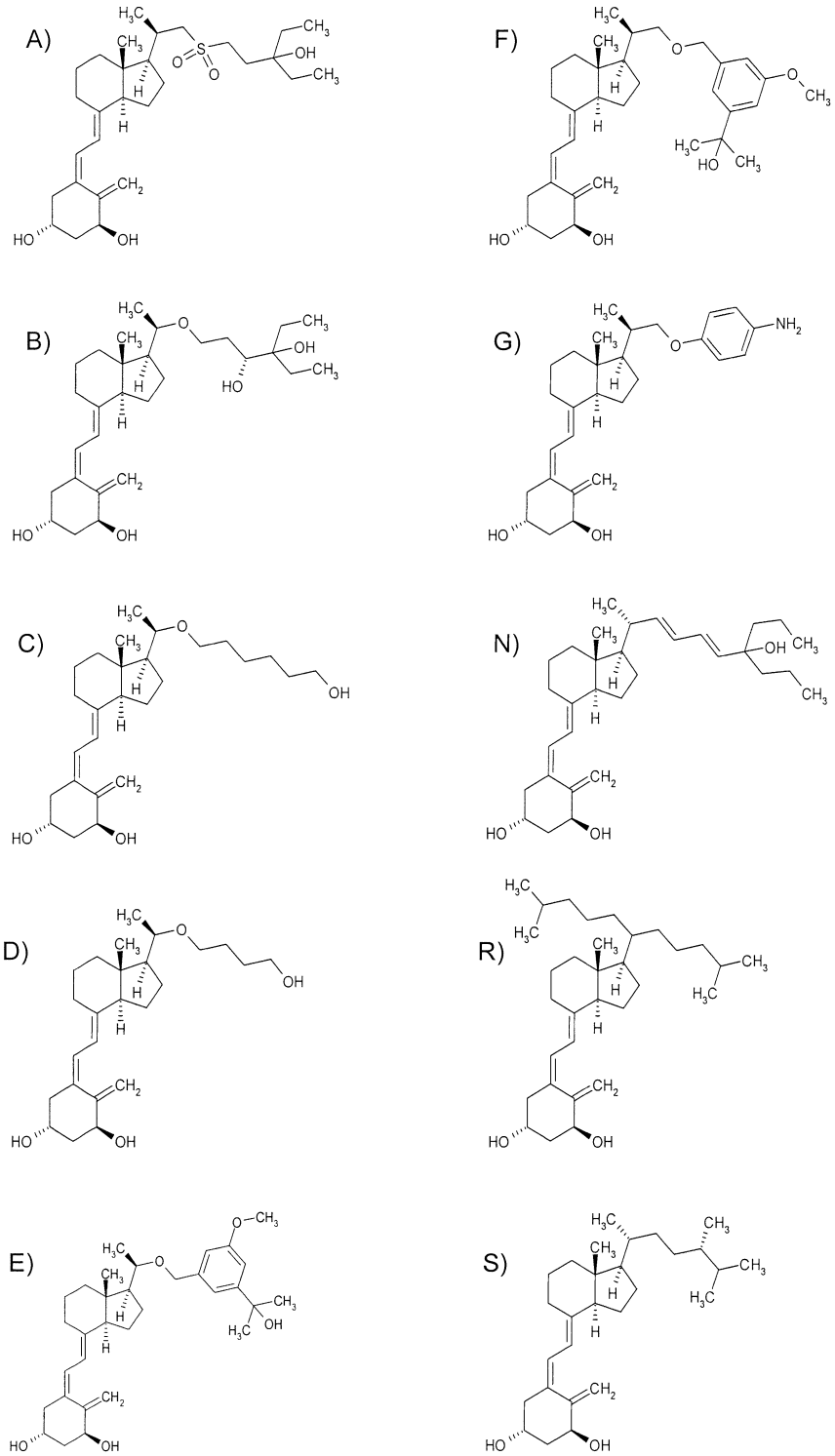

Figure 4. Chemical structures of 10 representative analogs from the selected set. Analog A is predicted to be stable (Class 2), analogs B–G to be unstable (Class1) and analogs N, R and S are predicted to be very stable (Class 3).

Table 5. Selected subsets of variables from five different variable selection methods. Variables written in bold letters are selected by at least three of five selection techniques.

Selection method	Selected variables
Jackknife	**#double, #ether, epi, QPlogS,** glob, #rotor, **#metabol**, BCUT_53[a], BCUT_54[a], BCUT_129[d], BCUT_106[c]
PV	**epi**, accptHB, **#double**, BCUT_137[d], #methyl, glob, **#ether**, BCUT_69[a], CNS, BCUT_107[c], PISA, **#metabol**, BCUT_46[a], BCUT_52[a], **IP(eV)**
FS	**epi, #ether**, **QPlogS**, BCUT_136[d], BCUT_75[b], **IP(eV)**, PISA, QPlogPo/w, CNS, BCUT_66[a], BCUT_98[c], BCUT_68[a], BCUT_157[d], AffyPCaco, BCUT_148[d]
GOLPE	**#double**, **#ether**, **epi**, #ethyl, SASA, **QlogS**, BIBCaco, glob, #rotor, QPlogKp, **IP(eV)**, **#metabol**, BCUT_46[a], BCUT_69[a], BCUT_148[d]
GA	#triple, **epi**, **#ether**, **#rctvFG**, QPpolrz, **QPlogS**, BIBcaco, QPlogKp, **IP(eV)**, **#metabol**, BCUT_26[a], BCUT_34[a], BCUT_37[a], BCUT_58[a], BCUT_66[a], BCUT_69[a], BCUT_85[b], BCUT_98[c], BCUT_125[d], BCUT_144[d]

[a]Encodes for atomic charges.
[b]Encodes for H-bond acceptor abilities.
[c]Encodes for H-bond donor abilities.
[d]Encodes for polarizabilities.

Table 6. Cross-validated PLS models based on 87 calcitriol analogs externally validated with a test set consisting of 43 analogs.

Selection method	# Outliers	# LV	r_{cv}	RMSECV in % point	RMSEP in % point
Jackknife	6	2	0.72	20	17
PV	3	2	0.72	20	18
FS	1	2	0.66	21	21
GOLPE	4	2	0.74	19	17
GA	4	7	0.75	19	16
Mix	11	1	0.75	19	19

diversity. Variable selection amongst the molecular descriptors was carried out by five different variable selection techniques. Generating a QSAR model to predict the *in vitro* metabolic stability of calcitriol analogs, the most important descriptors were found to be the number of double and ether bonds in the antenna side chain attached to carbon 17 and the stereochemistry at carbon 20.

Before the calibrations were performed, the 130 analogs were divided into a training set containing 87 analogs (2/3) and a test set containing 43 analogs (1/3). For every subset of variables one cross-validated PLS model was generated based on data from the training set. The models were validated by the model independent test set for which the *in vitro* metabolic stability was known in advance. The RMSEP values were slightly lower than or similar to the RMSECV values, indicating that the models were not overfitting, and the models were subsequently re-estimated with all 130 analogs.

The re-estimated models were used for prediction of a data set consisting of 244 calcitriol analogs for which the *in vitro* metabolic stability was unknown. The models could be divided into two groups according to their prediction of the data set. In both groups the models assigned the majority of compounds to the same classes. The results indicate that it may be an advantage to predict a new data set in several models instead of one particular model. Based on the predictions made by the models, 20 analogs were selected and the *in vitro* metabolic stability was determined. The PLS models were able to predict the correct metabolic stability for 17 of the 20 selected analogs, corresponding to a prediction performance

Table 7. PLS models for prediction of *in vitro* metabolic stability ($S9_m$) based on a data set compiled from data on 244 calcitriol analogs.

Selection method	% Predicted to Class 1	% Predicted to Class 2	% Predicted to Class 3
Jackknife	17	35	48
PV	2	12	86
FS	16	32	52
GOLPE	3	23	73
GA	4	16	80
Mix3	21	27	52

Table 8. $S9_m$ predicted by the PLS models compared with experimentally determined $S9_m$ (standard deviation $= \pm 5\%$).

Test analog	$S9_m$ predicted by the PLS models[a] (%)	$S9_m$ determined (%)	Test analog	$S9_m$ predicted by the PLS models[a] (%)	$S9_m$ determined (%)
A	10%–35%	21	K	> 35	80
B	< 10	25	L	> 35	85
C	< 10	0	M	> 35	5
D	< 10	0	N	> 35	60
E	< 10	10	O	> 35	56
F	< 10	2	P	> 35	47
G	< 10	10	Q	> 35	49
H	< 10	13	R	> 35	69
I	< 10	0	S	> 35	68
J	< 10	0	T	> 35	16

[a]The models used for prediction are Jackknife, FS and Mix3.

of 85%. The final models can compete with previously published 3D-QSAR models [38, 39] and the results indicate that QSAR models indeed are useful in predicting the *in vitro* metabolic stability of calcitriol analogs.

Supplementary Material

Table S1: 161 BCUT_descriptors calculated with DiverseSolutions. Table S2: 105 representative structures of the 130 calcitriol analogs in the calibration set. Table S3: Data set used for PLS modeling. Table S4: Data set used for PLS prediction. Supplementary material can be obtained from the authors.

Acknowledgements

This work was partly financed by grants from the Danish Ministry of Science, Technology and Innovation (Centre contract: Centre for New Sensor Systems for the Measurement of Food Quality). Gilda Kischinovsky is acknowledged for assistance with the manuscript.

References

1. Abe, E., Miyaura, C. and Sakagami, H., Proc. Natl. Acad. Sci. USA, 78 (1981) 4990.
2. Kuribayashi, T., Tanaka, K., Abe, E. and Suda, T., Endocrinology, 113 (1983) 1992.
3. Bhalla, A.K., Clemens, T., Amento, E., Holick, M.F. and Krane, S.M., J. Clin. Endocr. Metab., 57 (1983) 1308.
4. Stumpf, W.E., Clark, S.A., Sar, M. and DeLuca, H.F., Cell Tissue Res., 238 (1984) 489.
5. Provvedini, D.M., Tsoukas, C.D., Deftos, L.J. and Manolagas, S.C., J. Immunol., 136 (1986) 2734.
6. Reichrath, J. and Holick, M.F., In Holick, M.F. (Ed.), Vitamin D: Physiology, Molecular Biology, and Clinical Applications, Vol. 1. Humana Press Inc., Totowa, NJ, 1999, pp. 357–374.
7. Holick, M.F., In Holick, M.F. (Ed.), Vitamin D: Physiology, Molecular Biology, and Clinical Applications, Vol. 1. Humana Press Inc., Totowa, NJ, 1999, pp. 207–216.
8. Holick, M.F. In Holick, M.F. (Ed.), Vitamin D: Physiology, Molecular Biology, and Clinical Applications, Vol. 1. Humana Press Inc., Totowa, NJ, 1999, pp. 1–16.

9. van den Bemd, G.J.C.M. and Chang, G.T.G., Curr. Drug Targets, 3 (2002) 85.
10. Venkatesh, S. and Lipper, R.A., J. Pharm. Sci., 89 (2000) 145.
11. Thompson, T.N., Med. Res. Rev., 21 (2001) 412.
12. Ekins, S., Waller, C.L., Swann, P.W., Cruciani, G., Wrighton, S.A. and Wikel, J.H., J. Pharmacol. Toxicol., 44 (2000) 251.
13. Boobis, A., Gundert-Remy, U., Kremers, P., Macheras, P. and Pelkonen, O., Eur. J. Pharm. Sci., 17 (2002) 183.
14. Hansch, C. and Fujita, T., J. Am. Chem. Soc., 86 (1964) 1616.
15. Wold, S., Martens, H. and Wold, H., In Kågström, B. and Ruhe, A. (Eds.), Lecture Notes in Mathematics Vol. 973. Springer-Verlag, Berlin, 1983, pp. 286–293.
16. Kissmeyer, A.-M. and Mortensen, J.T., Xenobiotica, 30 (2000) 815.
17. Kissmeyer, A.-M., Mathiasen, I.S., Latini, S. and Binderup, L., Endocrine, 3 (1995) 263.
18. Kissmeyer, A.-M., Preclinical pharmacokinetics and metabolism of EB 1089 and other vitamin D analogs. Leo Pharmaceutical Products Ltd, Denmark, 2000.
19. Burden, F.R., J. Chem. Inf. Comput. Sci., 29 (1989) 225.
20. Pearlman, R.S. and Smith, K.M., Perspect. Drug Discov. Des., 9/10/11 (1998) 339.
21. Ponder, J.W. and Richards, F.M., J. Comput. Chem., 8 (1987) 1016.
22. Halgren, T.A., J. Comput. Chem., 20 (1999) 720.
23. Halgren, T.A., J. Comput. Chem., 20 (1999) 730.
24. Qiu, D., Shenkin, P.S., Hollinger, F.P. and Still W.C., J. Phys. Chem. A, 101 (1997) 3005.
25. Stewart, J.J.P., J. Comput. Chem., 10 (1989) 209.
26. Sutter, J.M. and Kalivas, J.H., Microchem. J., 47 (1993) 60.
27. Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L. and Engelsen, S.B., Appl. Spectrosc., 54 (2000) 413.
28. Höskuldsson, A., Chemometr. Intell. Lab., 23 (1994) 1.
29. Baroni, M., Clementi, S., Cruciani, G., Costantino, G. and Riganelli, D., J. Chemometr., 6 (1992) 347.
30. Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R. and Clementi, S., Quant. Struct.-Act. Relat., 12 (1993) 9.
31. Holland J.H. (Ed.) Adaption in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI, 1975.
32. Martens, H. and Martens, M., Food Qual. Prefer., 11 (2000) 5.
33. Wold, S., Technometrics, 20 (1978) 397.
34. Haaland, D.M. and Thomas, E.V., Anal. Chem., 60 (1988) 1193.
35. Calverly, M.J., Binderup, E. and Binderup, L., In Norman, A.W., Bouillon, R. and Thomasset, M. (Eds.), Proceedings on the Eighth Workshop on Vitamin D, Paris, France, 5–10 July 1991, Vol. 1. Walter de Gruyter, New York, 1991, pp. 163–166.
36. Dilworth, F.J., Calverley, M.J., Kakin, H.L.J. and Jones, G., Biochem. Pharmacol., 47 (1994) 987.
37. Jones, G., In Holick, M.F. (Ed.), Vitamin D: Physiology, Molecular Biology, and Clinical Applications, Vol. 1. Humana Press Inc., Totowa, NJ, 1999, pp. 57–84.
38. Wessel, M.D., Jurs, P.C., Tolan, J.W. and Muskal, S.M., J. Chem. Inf. Comput. Sci., 38 (1998) 726.
39. Yoshida, F. and Topliss, J.G., J. Med. Chem., 43 (2000) 2575.