Journal of Computer-Aided Molecular Design, **13:** 271–296, 1999.
KLUWER/ESCOM

Evaluation of a novel molecular vibration-based descriptor (EVA) for QSAR studies: 2. Model validation using a benchmark steroid dataset

David B. Turner[a,*], Peter Willett[a], Allan M. Ferguson[b] & Trevor W. Heritage[b]
[a]*Krebs Institute for Biomolecular Research and Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN, U.K.;* [b]*Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63144, U.S.A.*

Received 22 January 1998; Accepted 7 September 1998

Key words: 3D-QSAR, alignment-free QSAR, AM1, CoMFA, IR-vibration, PLS, PM3

Abstract

The EVA molecular descriptor derived from calculated molecular vibrational frequencies is validated for use in QSAR studies. EVA provides a conformationally sensitive but, unlike 3D-QSAR methods such as CoMFA, superposition-free descriptor that has been shown to perform well with a wide range of datasets and biological endpoints. A detailed study is made using a benchmark steroid dataset with a training/test set division of structures. Intensive statistical validation tests are undertaken including various forms of crossvalidation and repeated random permutation testing. Latent variable score plots show that the distribution of structures in reduced dimensional space can be rationalized in terms of activity classes and that EVA is sensitive to structural inconsistencies. Together, the findings indicate that EVA is a statistically robust means of detecting structure-activity correlations with performance entirely comparable to that of analogous CoMFAs. The EVA descriptor is shown to be conformationally sensitive and as such can be considered to be a 3D descriptor but with the advantage over CoMFA that structural superposition is not required. EVA has the property that in certain situations the conformational sensitivity can be altered through the appropriate choice of the EVA σ parameter.

Introduction

EVA is a molecular descriptor that is based upon fundamental infra-red (IR) and Raman range vibrational frequencies [1,2]. The descriptor has the advantage over 3D-QSAR methods such as CoMFA [3] inasmuch as it is invariant to rotation and translation of the structures concerned. Extensive studies [2] have indicated that EVA can successfully be used to develop robust QSAR models for a range of different structural classes, exhibiting various degrees of conformational freedom and with a variety of biological endpoints. These studies also found that EVA can perform well with heterogeneous sets of structures. In most cases the EVA models were found to be entirely statistically comparable to those obtained using CoMFA but without the difficulties associated with structural superposition. The EVA descriptor has also been found to be an effective component of a similarity measure for the similarity searching of a structural database [4].

The main purposes of this study are two-fold. Initially attention is focused upon validation of EVA QSAR based upon a much rigorous evaluation of a variety of EVA models than has previously been performed. A much-used benchmark steroid dataset [3] has been chosen for this work. In particular, we concentrate upon validation of the models obtained following the procedures outlined by Wold and Eriksson [5] wherein the structures are divided into a training set (to develop a model) and an entirely separate test set upon which to evaluate the model predictivity. Various forms of training set crossvalidation are used to determine the model dimensionality that should be used to predict the activities of the test set compounds. Intensive random permutation of the training set data

*To whom correspondence should be addressed.

is applied to assess the probability that observed correlations could be due to chance effects. A series of analogous CoMFA analyses is performed so as to provide a direct means of comparison of the EVA results with those of this popular technique. This is important because direct comparison of the results obtained here with those of most previously published QSAR studies with this dataset is not entirely valid since in most cases there are coding errors for various structures [6]. The effectiveness (judged by EVA QSAR statistical scores) of different levels of geometry optimisation (molecular mechanics, semiempirical and ab initio) is investigated together with consideration of how one might set about interpreting an EVA QSAR model in structural terms.

The second major aim of this study is to consider the conformational sensitivity of EVA. Previous studies with EVA either have involved compounds for which an alignment is not relevant [1] or the studies have started with conformations previously aligned for the purposes of CoMFA [2]. The development of a CoMFA model requires the selection of conformations and the overlay of these conformations within a 3D-grid. These tasks are of course inter-related and typically are driven by an underlying pharmacophore hypothesis; for example, a particular conformational selection may prevent the overlay of certain pharmacophoric features. This process is much simplified if, as is the case with the steroid dataset used here, there is a common rigid skeleton that can be used to provide the basis of an overlay. Whilst any true 3D descriptor must by definition be sensitive to conformation CoMFA can also be sensitive to small changes in the relative overlay of compounds. This is primarily due to the steepness of the distance-dependent potential functions at the van der Waals surface [7] but is further exacerbated when grid resolution-related sampling errors occur [8]. These and related issues provided the initial incentive for the development of the alignment-free EVA descriptor [1]. The term alignment-free is used here to indicate that EVA is not subject to the practical problems of overlay within a grid that add a further element of complexity and ambiguity to CoMFA. However, as demonstrated herein EVA is sensitive to 3D structure and it is shown that there is a relationship between the conformational-sensitivity of the technique and the value of σ, the main EVA parameter.

Methods

Notation

The following is an alphabetic list of abbreviations related to PLS analyses: CV – crossvalidation; F – Fischer test for significance; G – number of CV groups; LOO – leave-one-out (CV, where G = M); LV_{opt} – optimum number of latent variables (LVs); M – number of training set compounds; pr^2 – predictive-r^2 scores (external test set); PRESS – Predictive Residual Sum of Squares; q^2 – crossvalidated-r^2 (internal validation); r^2 – fitted-r^2; SE and SE_{CV} – Standard Error and CV Standard Error.

Software and hardware

All the work described herein was carried out using a Silicon Graphics Origin 200 R10000. The molecular modeling software used was Sybyl 6.3 [9] and the software required to perform the EVA standardisation process was custom written in 'C'.

Determination of the EVA descriptor

The derivation of the EVA descriptor has previously been described in some detail [1,2] and only a brief description of the technique will be given here. The descriptor is derived from IR- and Raman-range molecular vibrational frequencies typically obtained through the application of a normal coordinate analysis (NCA) to an appropriately energy minimized structure. For a compound with N atoms there are 3N−6 (or 3N−5 for a linear structure such as acetylene) normal modes of vibration. The frequency set for a given structure is projected onto a linear bounded frequency scale (BFS) covering a range from 1 to 4000 cm^{-1}. The use of this range means that all fundamental vibrational frequencies are included in the analysis – should a frequency exceed 4000 cm^{-1} then all frequencies from all molecules can be scaled according to the scale factors discussed below. Next a Gaussian kernel of fixed standard deviation (σ) is placed over each and every frequency value. The BFS is then sampled at fixed increments of L cm^{-1} and the value of the resulting EVA descriptor, EVA_x, at each sample point, x, is the sum of the amplitudes of the overlaid kernels at that point:

$$EVAx = \sum_{i=1}^{3N-6} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-f_i)^2/2\sigma^2} \quad (1)$$

where f_i is the i^{th} normal mode frequency of the compound concerned. This procedure is repeated for each dataset compound and results in a descriptor set consisting of 4000/L variables. Typically a descriptor set has been derived using a σ of 10 cm^{-1} and an L of 5 cm^{-1} giving 800 descriptor variables [1]. For a standard QSAR dataset the number of variables is thus very much larger than M and Partial least squares to Latent Structures (PLS) [10] is hence used to provide a robust regression analysis. It is important to note that the purpose of the EVA smoothing procedure is not to simulate an experimental IR spectrum (transition dipole data is discarded) but rather it is to apply a smearing function such that vibrations at slightly different frequencies in different compounds can be compared to one another. As such the results obtained with EVA QSAR can be dependent upon the chosen kernel width (σ) [2] since this parameter determines whether or not proximal kernels overlap. The general approach to selecting an EVA QSAR model is to derive a range of models using various σ values and, on the basis of crossvalidation results, select an optimal model to use for the prediction of previously unseen compounds. The effectiveness of such an approach is demonstrated in the results described below. Finally, the use of a fixed Gaussian standard deviation (kernel height, width and shape) means that each frequency (i.e., each part of the spectrum) is equally weighted prior to regression analysis.

Dataset

The selected dataset consists of 31 steroids for which binding affinities (expressed as log [K], where K is the association constant) to corticosterone-binding globulin (CBG) are available (Table 1). For each training set steroid confidence limits are available for the experimental relative binding affinities (RBA) from which the association constants are determined [11]. These limits enable the estimation of the precision to which log [K] can legitimately be predicted, giving a lower bound to the SE of ∼0.08. It is thus legitimate to make quite precise predictions of these activity values.

In the first, validation, part of this study, the structures utilised were provided by Wagener et al. [6]. This data set was also used in the original CoMFA publication of Cramer et al. [3], the structures from which are supplied with the Sybyl software [9], and feature in a study by Good et al. [12], the structures from which are supplied with Oxford Molecular software [13]. The former will henceforth be referred to as the Cramer dataset and the latter the OML dataset. Initially, much of the work described below was performed using the OML dataset. However, due to the discovery of various coding errors in this (and the Cramer) dataset the work was repeated using the Wagener dataset – these authors indicate those structures for which errors (primarily stereochemical) were found in both the Cramer and OML datasets [6].

Table 1. Steroid CBG-binding affinities

	Compound	CBG affinity
Training set		log [K]
M1	Aldosterone	6.279
L2	Androstanediol	5.000
L3	Androstenediol	5.000
L4	Androstenedione	5.763
L5	Androsterone	5.613
H6	Corticosterone	7.881
H7	Cortisol	7.881
M8	Cortisone	6.892
L9	Dehydroepiandrosterone	5.000
H10	Deoxycorticosterone	7.653
H11	Deoxycortisol	7.881
M12	Dihydrotestosterone	5.919
L13	Estradiol	5.000
L14	Estriol	5.000
L15	Estrone	5.000
L16	Etiocholanolone	5.255
L17	Pregnenolone	5.255
L18	17-Hydroxypregnenolone	5.000
H19	Progesterone	7.380
H20	17-Hydroxyprogesterone	7.740
M21	Testosterone	6.724
Test set		
H22	Prednisolone	7.512
H23	Cortisol 21-acetate	7.553
M24	4-Pregnene-3,11,20-trione	6.779
H25	Epicorticosterone	7.200
M26	19-Nortestosterone	6.144
M27	16α,17-Dihydroxy-4-pregnene-3,20-dione	6.247
H28	17-Methyl-4-pregnene-3,20-dione	7.120
M29	19-Norprogesterone	6.817
H30	11β,17,21-Trihydroxy-2α-methyl-4-pregnene-3,20-dione	7.688
M31	11β,17,21-Trihydroxy-2α-methyl-9α-fluoro-4-pregnene-3,20-dione	5.797

Structure numbers and activity group classification prefixes are those used by Good et al. [12]: H – high activity; M – medium; L – low.

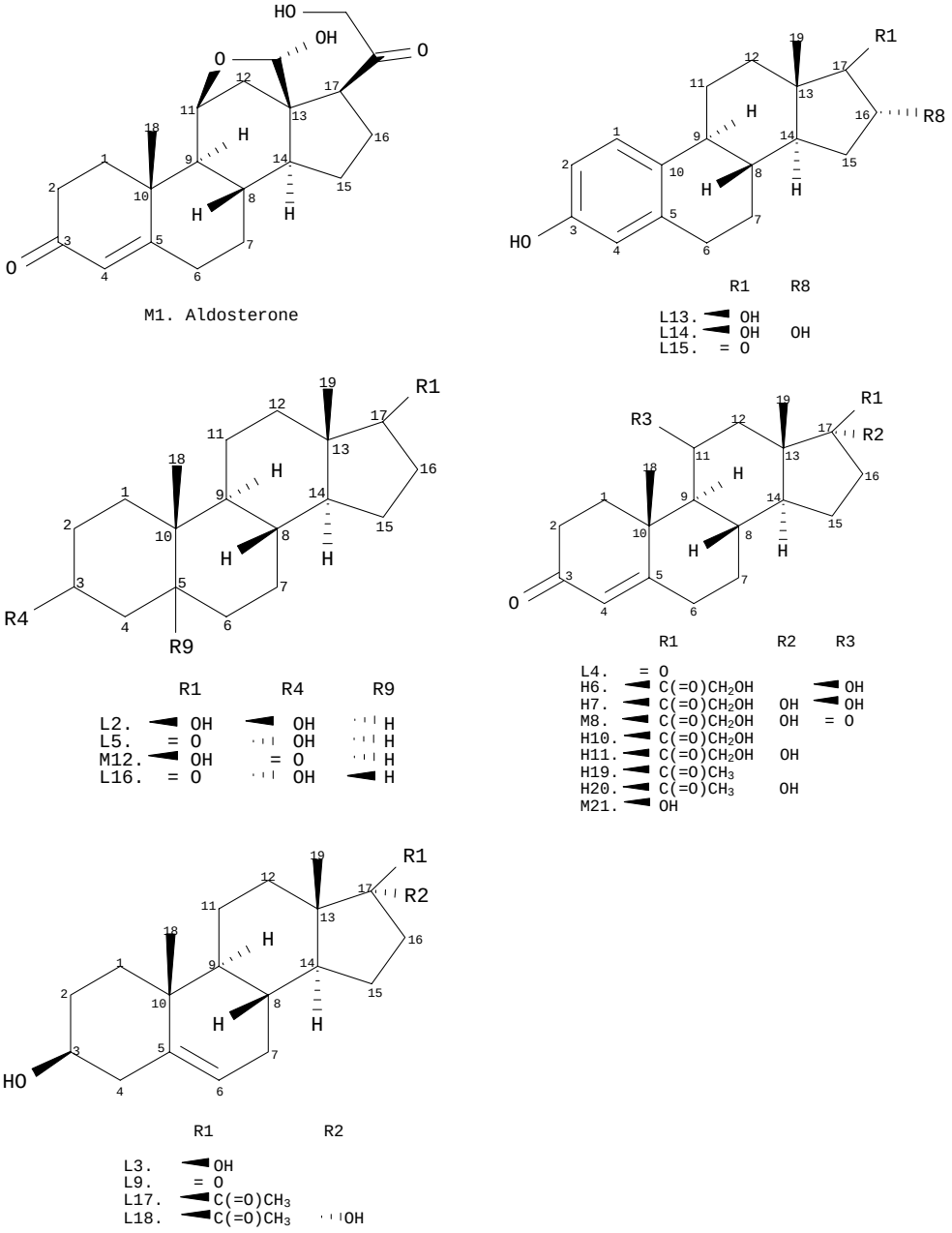
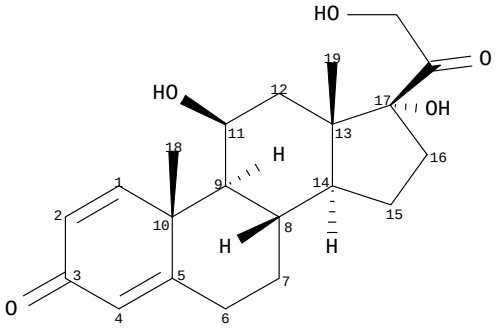


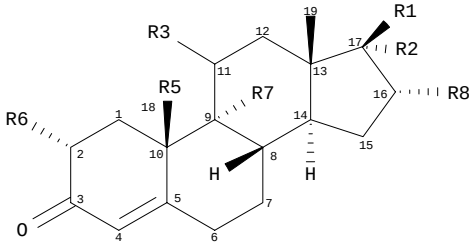
Figure 1. The 21 training set steroids.

These steroids have become a benchmark dataset largely because they are conformationally restricted and have an obvious choice of alignment that provides satisfactory results with CoMFA; this alignment is based on the simple least-squares fit of the 3, 5, 6, 13, 14, and 17 carbon atoms (Figure 1) with deoxycortisol as the template. This straightforward alignment-rule means that alternative conformations can be readily overlaid for CoMFA, a useful feature when many different conformations are to be considered as is the case in the later parts of this study.

The dataset is divided into a training set of 21 structures and a test set of 10 structures – Figures 1 and 2 respectively. This division of structures is that used in the original CoMFA publication [3] and subsequently by many other authors [6,12,14–21]. The



H22. Prednisolone



	R1	R2	R3	R5	R6	R7	R8
H23.	C(=O)COCOCH	OH	OH	CH_3		H	
M24.	C(=O)CH_3		= O	CH_3		H	
H25.	C(=O)CH_2OH		OH	CH_3		H	
M26.	OH			H		H	
M27.	C(=O)CH_3	OH		CH_3		H	OH
H28.	C(=O)CH_3	CH_3		CH_3		H	CH_3
M29.	C(=O)CH_3			H		H	
H30.	C(=O)CH_2OH	OH	OH	CH_3	CH_3	H	
M31.	C(=O)CH_2OH	OH	OH		CH_3	F	

Figure 2. The 10 test set steroids.

test set consists of 10 structures selected from a larger number of structures for which activity data became available after the original 21-structure CoMFA model was developed by Cramer et al. The only selection criterion was the availability of related structures in the Cambridge Structural Database [22]. In other words, a statistical experimental design method [23,24] has not been applied to these 31 structures and the data set division of Cramer et al. has been retained solely to permit some degree of comparison of our results with those of the numerous QSAR studies referenced above.

The test set (Figure 2) has a particularly obvious structural outlier (M31) which is the only compound with a position-9 substituent and with a highly electronegative fluorine substituent. The expectation prior to analysis is that activity-prediction for this structure may be poor since QSAR models can only be interpolative. This, of course, depends on whether structural variation at this point influences activity but it seems that it does since the activity of M31 is about 100 times (two log units) lower than that of H30, to which it is identical apart from the fluorine substituent. Therefore, predictions were made both including and excluding M31 from consideration; previous authors have used the same approach [3,12,14]. There are a number of other test set compounds with structural features not explicit in the training set the most important of which are indicated in the Results section. However, an important feature of both EVA and CoMFA descriptors is that, in principle, they can be used to make predictions for compounds whose structural features are not explicit in a training set [2] – this property is to some extent reflected in the results described below.

Calculation of normal mode frequencies (EVA)

Unless otherwise indicated EVA descriptors are based on vibrational frequencies derived using the MOPAC 6.0 [25] AM1 Hamiltonian. The effect of using frequencies derived using other levels of theory is discussed in the Results section. For the MOPAC optimisations all the structures were first geometry optimized (MOPAC keyword ALL_BONDS_AND_ANGLES) with convergence criteria of GNORM = 0.01 and SCFCRT = 10^{-12}; all other MOPAC settings were left at their default values. MOPAC geometries and back-calculated point charges were retrieved. This was followed by MOPAC FORCE calculations to determine the force constants from which molecular vibrational frequencies are calculated using a NCA. None of the 31 structures had imaginary ('negative') normal mode frequencies, indicating that the optimized geometries were at or very close to a stationary point [2].

EVA descriptors

The EVA descriptors were generated from normal mode frequencies and imported into a Sybyl molecular spreadsheet for PLS analysis. Initially, a wide range of models was developed based upon each of the EVA σ terms in the set {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 16, 18, 20, 22, 24, 26, 28, 30, 40, 50}. This procedure enables the 'best' model to be selected on the basis of training set CV results which can then be used for test set prediction. A number of other models were used to make test set predictions so as to evaluate the effect of alternative choices of σ to that indicated to be optimal by crossvalidation. These evaluations

can be undertaken rapidly when using the highly efficient SAMPLS [26] formulation of PLS. In all cases sampling increments, L, were kept sufficiently small relative to σ such that there was no omission of information encoded in the descriptor. In our previous study [2] the maximum values of L, which are σ-specific, were determined. These values are referred to as L_{crit} values and, provided $L < L_{crit}$, there is only an insignificant sampling error associated with extracting the EVA descriptor.

CoMFA analyses

CoMFA analyses were undertaken using standard Sybyl parameters viz.: a 1 Å grid-spacing in all directions, with the grid extending 4 Å beyond the union molecular volume; a carbon sp^3 probe with a unit positive charge; steric and electrostatic cut-off values were both ±30 kcal/mol, with electrostatic values lying within the steric cut-off region excluded from the analyses. MOPAC 6.0 AM1 calculated charges were utilised. Analyses were done using steric and electrostatic fields, both separately and combined, and were performed for unscaled, pre-autoscaled and pre-blockscaled (Sybyl COMFA_STD scaling option) data where appropriate.

CoMFA PLS modeling results can be dependent upon the location of the structures relative to the 3D-grid [8]; i.e., with a 2 Å grid rotation/translation of all structures as a rigid body has been shown to result in quite different statistical scores (q^2, r^2). Such sampling errors are a consequence of an insufficiently high grid-resolution and, therefore, the use of a 1 Å grid-spacing is recommended [8]. Such descriptor-space sampling errors do not arise with EVA provided that the L_{crit} values previously tabulated [2] are not exceeded. In order to assess the reliability of a 1 Å CoMFA model the aggregated structures were systematically reoriented within the grid and PLS modeling applied; detail of the orientations used is given in the results section. The resulting distribution of q^2, r^2, and pr^2 scores were then examined for consistency. In addition, Cross-validated R^2-Guided Region Selection (CVR^2-GRS) [8] has been presented as providing more consistent PLS results. This domain-based variable selection procedure involves: dividing the conventional CoMFA 3D-grid into 125 non-overlapping boxes of equal size; performing crossvalidated PLS on each sub-region separately using a 1 Å resolution; combining those sub-regions that individually provide a q^2 greater than a user-defined threshold (typically, 0.3) to give a re-combined region (1 Å resolution) that is used in a final PLS analysis.

It is important to note that there are two sets of conformations to which CoMFA can be applied. The first set consists of those supplied by Wagener et al. [6] – the 3D structure generator Corina [27] was applied to 2D connection table information following the stereochemistry indicated in Figures 1 and 2. The orientation of the C17 side chain (where present) was manually adjusted to maximize interaction of the C20 carbonyl group with a CBG-binding site hydrogen donor group. This procedure provides a suitable overlay for CoMFA. However, in order to perform a NCA (for EVA descriptor generation) the structures must first be geometry optimised. This means that the geometries used for EVA modeling are likely to (but may not) be somewhat different to the pre-optimised geometries. Therefore, in order to assess the presence and effect of these geometry alterations, CoMFA was also applied to the MOPAC 6.0 optimized geometries, aligned as before to deoxycortisol using a rigid-body least-squares fit of the six steran ring atoms.

PLS modeling (EVA and CoMFA)

As an initial procedure, LOO-crossvalidated PLS analysis was performed for each EVA and CoMFA descriptor set. In all cases the Sybyl minimum variance cut-off was set to zero. The LOO procedure, rather than leave-x-out, is the most widely used form of CV, having the advantage that the analyses are reproducible in the absence of CV group membership information. It is acknowledged, however, that LOO is not always the best approach to CV [28] and that, depending upon the grouping of compounds within the dataset, there may be more appropriate values for the number of CV groups (G). This is particularly the case with large datasets where the q^2 of LOO cross-validation approaches r^2 at the limit [5,29]. Therefore, additional CV runs were performed with G set to the recommended value of 7 [28]. Since CV group membership is assigned at random it is necessary to repeat CV runs many times to avoid effects resulting from the chance selection of compounds on any given run. Therefore, CV was repeated 200 times and LV_{opt} chosen from the mean of the SE_{CV} so obtained. There are, however, no fixed rules for determining G and further tests were thus performed where $G = 2$ and 3 (200 runs each) and $G = 4$, 5, 6 and 8 (50 runs each).

There are also no fixed rules for determining LV_{opt}. The models reported here are selected according to the

first Sybyl PLS SE_{CV}-minimum subject to the constraint that the maximum number of LVs should not exceed M/4 [30,31]. Thus a maximum of five LVs are extracted in these analyses. LVs were added to the models in order of their extraction by PLS. In addition, it is not acceptable to make predictions of the bioactivity of structures to greater precision than the error in the original measurements (a lower bound for the SE has been given above).

Test set predictions

The test set predictions are compared through predictive-r^2 (pr^2) scores together with examination of the predictive residuals themselves. Pr^2 and q^2 scores are defined in an analogous manner; separate terms have been used to distinguish clearly between the training and test set scores. The formula for the pr^2 is:

$$pr^2 = \frac{SD - PRESS}{SD} \quad (2)$$

where SD is the sum of the squared deviations between the affinities of molecules in the test set and the mean affinity of the training set molecules. If each test set compound were predicted to have the affinity of the mean value of the training set then the pr^2 would be zero.

Random permutation testing

Random permutation tests are applied to selected EVA and CoMFA training set models in order to establish confidence levels for the 'real' QSAR models and are considered to be one of the most powerful methods of validating a QSAR model [5,32]. The bioactivity data is first thoroughly scrambled by repeatedly swapping the activity values for randomly selected pairs of compounds – this method retains the original characteristics (mean, standard deviation etc.) of the bioactivity data. PLS is then applied in the form of both LOO crossvalidation and fitted model generation up to a maximum of 5 LVs – this provides q^2 and r^2 values that can be compared to those of the 'real' QSAR models. This process of activity data scrambling and model derivation was repeated 1,000 times per descriptor set thus providing a distribution of q^2 and r^2 scores using which the significance of the 'real' QSAR model can be assessed.

Results and discussion

Characteristics of the EVA descriptor

Figure 3 provides plots of the EVA variable values for deoxycortisol (one of three structures with the highest CBG-binding activity) and estradiol (one of the inactive structures) over the spectral range 1 to 4000 cm^{-1}; also provided is a plot of the univariate standard deviation (STDEV) of the descriptor over all 21 training set structures. These plots indicate that there is a great deal of vibrational information (many peaks) in the fingerprint region (1–1500 cm^{-1}) and considerably less information in the functional group region (1500–4000 cm^{-1}); this is a typical feature of an experimental IR spectrum whether there are many different functional groups present or very few as is the case here. The relative importance of the different parts of the spectrum for correlating with bioactivity is discussed below. The EVA descriptor values and the STDEV over the whole training set are largest at frequencies centered around 1400 and 3100 cm^{-1} (the latter is the hydrogen-stretching region); this is discussed further below in relation to autoscaling of the EVA descriptor.

Overview of the EVA training set models: various σ terms

A summary of the LOO CV results obtained using a wide range of EVA σ terms is given by plots of the cumulative fraction of variance of CBG-binding activity explained by successive PLS LVs (Figure 4). This figure clearly indicates that the most internally predictive and parsimonious models are to be found at low EVA σ values, with a peak around 3 cm^{-1}, where LV1 explains 77% (i.e., $q^2 = 0.77$) and LV2 explains only a further 3% of the CBG-binding activity; the use of additional LVs with this σ results in a deterioration in q^2. As the EVA σ is increased from 3 cm^{-1} there is a gradual and ultimately substantial loss of activity explained by LV1 which is recovered to some extent, but not completely, with additional LVs. This is particularly the case with LV2, for which the additional variance explained increases from 0.03 (where $\sigma = 3$ cm^{-1}) to 0.42 (where $\sigma = 28$ cm^{-1}), while the LV1 q^2 drops from 0.77 to 0.25 respectively. The 4 cm^{-1} model was selected as the optimal EVA model since it is insignificantly different from that at 3 cm^{-1} while the larger σ permits the use of a larger L [2] and thus a lower descriptor dimensionality; this is a useful means of reducing computational effort (particularly

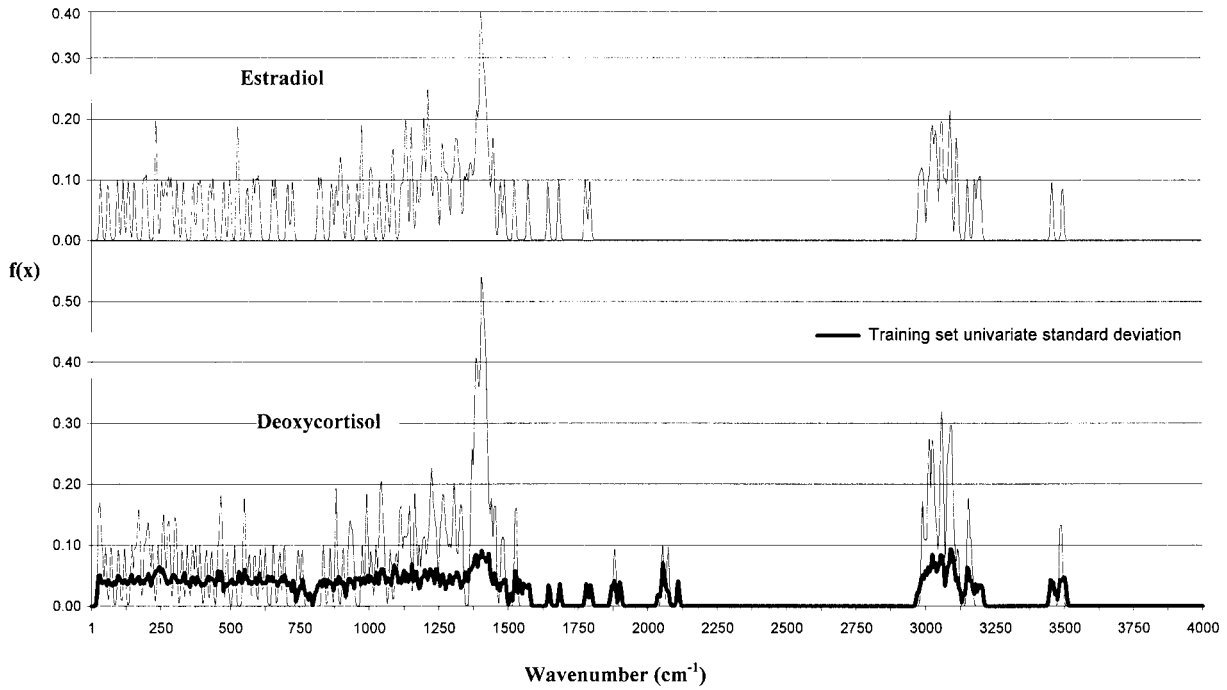


Figure 3. Characteristics of the EVA descriptor. EVA pseudo-spectra for estradiol (inactive for CBG-binding) and deoxycortisol (highest CBG-binding activity); a Gaussian σ of 4 cm^{-1} has been used. The univariate standard deviation for the 21 training set structures is given by the heavy line.

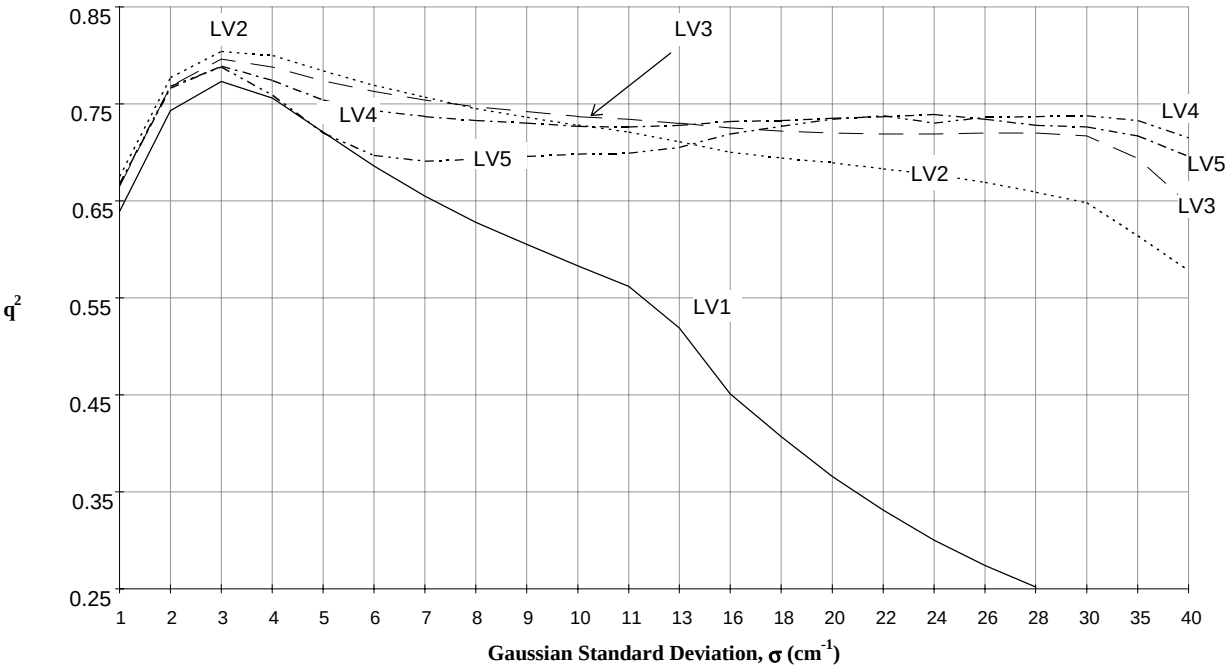


Figure 4. Cumulative fraction of CBG-binding activity variance (q^2) explained by successive PLS LVs for EVA descriptor sets derived using a wide range of σ values.

Table 2. PLS training and test scores in detail for selected EVA analyses[a]

EVA σ	Scaling	LV	LOO crossvalidation				Fitted model				Predict[b]
			q^2	SE_{CV}	F	p^c_{RAND}	r^2	SE	F	p^c_{RAND}	pr^2
4 cm^{-1}	None[d]	1	0.76	0.60	9.3	–	0.91	0.37	184.8	–	0.64 (0.68)
		2	**0.80**	**0.55**	**12.0**	**0.0010**	**0.96**	**0.24**	**224.1**	**0.0029**	**0.69 (0.74)**
		3	0.79	0.59	11.1	–	0.98	0.17	305.4	–	0.68 (0.73)
10 cm^{-1}	None[d]	1	0.58	0.78	4.2	–	0.73	0.63	50.4	–	0.50 (0.57)
		2	**0.73**	**0.65**	**8.0**	**0.0019**	**0.88**	**0.42**	**68.5**	**0.0007**	**0.59 (0.65)**
		3	0.74	0.65	8.4	–	0.93	0.33	77.0	–	0.61 (0.67)
20 cm^{-1}	None	1	0.37	0.96	1.7	–	0.54	0.82	22.0	–	0.41 (0.52)
		2	0.69	0.69	6.6	–	0.83	0.52	42.5	–	0.33 (0.48)
		3	**0.72**	**0.67**	**7.7**	**0.0019**	**0.92**	**0.37**	**62.5**	**0.0004**	**0.36 (0.50)**
		4	0.74	0.68	8.3	–	0.94	0.31	66.6	–	0.17 (0.38)
	Auto	1	0.54	0.81	3.6	–	0.77	0.58	63.0	–	0.34 (0.60)
		2	**0.61**	**0.77**	**4.7**	–	**0.88**	**0.42**	**68.3**	–	**0.20 (0.40)**
		3	0.62	0.78	4.9	–	0.94	0.32	82.5	–	0.14 (0.31)
40 cm^{-1}	None	1	0.18	1.10	0.6	–	0.34	0.98	10.0	–	0.42 (0.57)
		2	0.58	0.80	4.1	–	0.73	0.64	24.5	–	<0 (0.32)
		3	0.65	0.76	5.5	–	0.85	0.49	32.8	–	0.09 (0.36)
		4	**0.70**	**0.72**	**6.9**	**0.0067**	**0.91**	**0.39**	**40.9**	**0.0015**	**0.16 (0.56)**
		5	0.72	0.72	7.5	–	0.94	0.33	46.8	–	0.01 (0.51)
	Auto	1	0.44	0.90	2.3	–	0.61	0.75	29.5	–	0.34 (0.69)
		2	0.55	0.83	3.7	–	0.79	0.57	33.0	–	0.03 (0.36)
		3	**0.58**	**0.82**	**4.2**	–	**0.90**	**0.40**	**50.5**	–	**<0 (0.10)**
		4	0.58	0.86	4.1	–	0.92	0.38	44.5	–	<0 (0.13)

[a] The selected (first SE_{CV}-minimum-based) PLS model is indicated in bold type.
[b] Test set predictive-r^2 scores for 10 or 9 (bracketed) compounds. See main text for further details.
[c] Probability that the observed r^2 or q^2 scores could arise by chance estimated from 1000 random permutation tests; see main text for further details.
[d] Autoscaling resulted in models for all of which $q^2 < 0$.

important when applying random permutation tests). As discussed below, in general, an EVA model based upon a low σ is preferred because the narrow kernel width should result in less intra-structural kernel overlap (frequency information mixing) and, therefore, perhaps facilitate model interpretation. However, in order to test the statistical validity of models based upon larger σ terms, EVA descriptors derived using kernel σ's of 10, 20 and 40 cm^{-1} were also selected for further study.

Detail of the selected EVA models

Table 2 provides a summary of the LOO crossvalidated PLS results for each of the four selected EVA descriptor sets. Models based upon a range of LVs have been included to make explicit the trends in the models derived, while the preferred model dimensionality (SE_{CV}-minimum) is in bold type. For both the EVA 4 and 10 cm^{-1} descriptors two-LV models are optimal in terms of LOO CV scores. For the EVA 4 cm^{-1} descriptors this choice of model dimensionality was reinforced by leave-x-out CV results (Table 3), where values of G in the set {2, 3, 4, 5, 6, 7, 8} were used; in only very few cases did $LV_{opt} \neq 2$. With the EVA 20 and 40 cm^{-1} descriptor sets the optimal models (from LOO CV only) are with three and four LVs respectively. The effect of pre-autoscaling the EVA descriptors (using the Sybyl PLS AUTO_SCALING option) relative to no scaling is detrimental in all cases.

Turning to the test set it is clear that the EVA model suggested by CV to be the most predictive ($\sigma = 3/4$ cm^{-1}) is in fact the most predictive with these compounds. This means that at least for this dataset it is effective to use training set CV and parsimonious model selection to guide the choice of the EVA σ parameter. The pr^2 scores with the EVA 4 cm^{-1} model are 0.69 where M31 is included and 0.74 where M31 is

Table 3. Summary of crossvalidation results where G < N

Descriptors	G	Runs	LV^a_{opt}			SE_{CV} for LV_{opt}			q^2 for LV_{opt}		
			1	2	3	Max	Min	Mean	Max	Min	Mean
EVA 4 cm^{-1}	8	50	–	48	2	0.64	0.52	0.56	0.82	0.73	0.79
	7	200	–	200	–	0.70	0.49	0.57	0.84	0.69	0.79
	6	50	–	50	–	0.69	0.51	0.57	0.83	0.69	0.79
	5	50	–	50	–	0.65	0.53	0.59	0.82	0.73	0.78
	4	50	2	47	1	0.76	0.48	0.60	0.85	0.63	0.76
	3	200	11	177	12	1.00	0.47	0.62	0.86	0.35	0.74
	2	200	33	152	15	1.70[b]	0.45	0.70	0.87	−0.89[b]	0.67
CoMFA[c]	8	50	–	50	–	0.58	0.44	0.49	0.87	0.78	0.85
	7	200	–	197	3	0.59	0.43	0.49	0.88	0.77	0.84
	6	50	–	46	4	0.69	0.43	0.49	0.88	0.69	0.84
	5	50	–	48	2	0.72	0.42	0.49	0.89	0.66	0.84
	4	50	–	48	2	0.64	0.42	0.50	0.89	0.74	0.84
	3	200	4	168	25[d]	0.85	0.39	0.51	0.90	0.53	0.83
	2	200	16	143	29[e]	1.59[b]	0.37	0.57	0.91	−0.66[b]	0.78

Each analysis was repeated at least 50 times providing a range of models from which to select the optimal number of LVs (LV_{opt}).

[a] Number of cases where LV_{opt} is 1, 2 or 3.

[b] The next largest SE_{CV} scores are 1.01 (EVA) and 0.94 (CoMFA) while the next smallest q^2 scores are 0.33 (EVA) and 0.42 (CoMFA) – these models both arose where one of the two CV groups included all of the high activity structures.

[c] Combined steric and electrostatic fields – 2 Å grid resolution and no pre-scaling.

[d] Two of the four-LV and one of the five-LV models were optimal.

[e] Twelve of the four-LV models were optimal.

excluded; in the latter case there is therefore a reasonable match between q^2 and pr^2. This match is closer still if the mean q^2 scores from leave-x-out crossvalidation (Table 3) are compared to the pr^2 scores; this is particularly so where G is 2, 3 or 4 suggesting that LOO CV here provides a slightly optimistic estimate of the external predictivity of the EVA 4 cm^{-1} model.

An LV score plot for the EVA 4 cm^{-1} fitted training set model (Figure 5) illustrates that the distribution of compounds in LV space reflects the high-medium-low activity classification (Table 1). It should be noted that M8 is the most active of the medium activity structures and L4 is the most active of the low activity structures, so the clustering of these two compounds with compounds of a higher activity class is not inconsistent. In the aforementioned OML dataset testosterone (M21) has a topological error in the sense that the position-3 carbonyl has been substituted with a hydroxy group; this is significant because a position-3 hydrogen bond acceptor is a partial requirement for CBG-binding activity (Figure 6). This error is clearly reflected on an analogous EVA LV score plot (not shown) in which M21 clusters on its own some distance from any other compound.

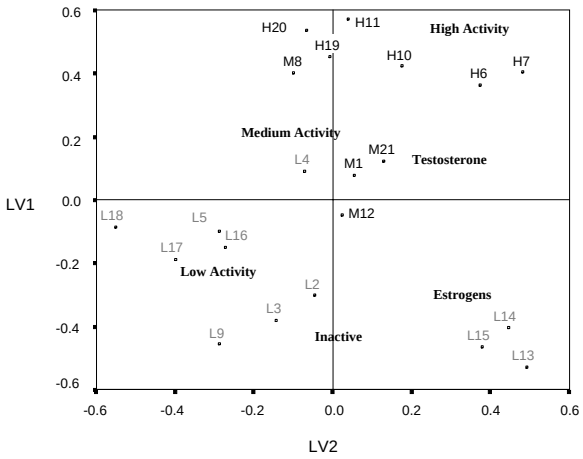


Figure 5. EVA 4 cm^{-1} fitted model LV score plot.

This sensitivity of EVA to structural inconsistency provides, albeit in a limited way, a test of the validity of the EVA method. More thorough and extensive validation is provided by random permutation tests. These were applied to the training set data for each of the selected EVA descriptor sets (Table 2) and indicate that

Table 4. PLS results for the CoMFA combined-field analyses

Grid interval/ Pre-scaling	Geometries	LV	LOO crossvalidation				Fitted model				Predict
			q^2	SE_{CV}	F	p^a_{RAND}	r^2	SE	F	p^a_{RAND}	pr^2
2 Å / None[b]	Wagener	1	0.78	0.57	10.3	–	0.84	0.49	97.8	–	0.50 (0.90)
		2	**0.85**	**0.47**	**17.6**	0.0001	**0.93**	**0.32**	**126.0**	0.00003	**0.35 (0.84)**
		3	0.82	0.54	13.7	–	0.96	0.26	131.5	–	0.34 (0.81)
1 Å / None[b]	Wagener	1	0.77	0.58	18.6	–	0.82	0.51	87.6	–	0.54 (0.87)
		2	**0.87**	**0.45**	**36.4**	0.0001	**0.93**	**0.32**	**126.4**	0.00002	**0.45 (0.84)**
		3	0.83	0.52	28.4	–	0.95	0.28	110.2	–	0.41 (0.81)
1 Å / None[b]	AM1-	1	–	0.75	0.61	11.8	0.81	0.52	83.4	–	0.53 (0.85)
	optimised	**2**	–	**0.83**	**0.52**	**18.8**	**0.93**	**0.34**	**111.1**	–	**0.43 (0.81)**
		3	–	0.82	0.54	17.9	0.95	0.29	105.2	–	0.36 (0.76)

See the footnotes to Table 2 and the main text for further information.
[a] Probability that the observed r^2 or q^2 scores could arise by chance estimated from 1000 random permutation tests; see main text for further details.
[b] Blockscaling gives identical results to no scaling. Training set autoscaled models are equally good but test set predictions are very poor ($pr^2 < 0$).

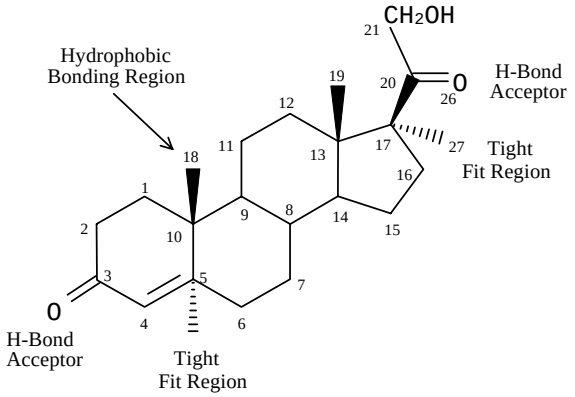


Figure 6. CBG binding site information. Deduced features required for binding to human corticosteroid-binding globulin (CBG); after Mickelson et al. [33]. The tight fit regions adjacent to carbons 5 and 17 are α with respect to the steroid plane, while the hydrophobic 'bonding' region is in the β-position.

the probability, p, that the observed models could have occurred by chance is in all cases extremely low, with a maximum p value of 0.0029. The findings support the assertion that statistically robust QSAR models can be developed for EVA descriptor sets derived using a wide range of different σ terms.

CoMFA

CoMFA analyses were performed using both the alignments originally provided by Wagener et al. and the AM1 optimised geometries from which EVA descriptors were calculated (Table 4). In the former case, and using a 1 Å grid resolution, the PLS scores are high with a q^2 of 0.87 (two LVs) and an r^2 of 0.93 (use of a 0.5 Å grid gave identical scores). As with EVA this choice of model dimensionality was reinforced by leave-x-out CV (Table 3) where in the vast majority of cases LV_{opt} is two. Random permutation tests (Table 4) support the validity of these models since the estimate of chance correlation, p, is ≤ 0.0001. This model has a test set pr^2 of 0.84 (M31 omitted) and 0.45 (M31 included), indicating that M31 is an outlier for the CoMFA model. If a 2 Å resolution is used (Table 4) the PLS scores are almost identical although there is a significant reduction (0.1 units) in the pr^2 score where M31 is included; this difference hints at the grid-resolution issues noted above. Indeed, systematic reorientation of the all compounds as a rigid aggregate relative to the bounding 3D-grid provides a wide-range of q^2 and pr^2 scores at a 2 Å grid resolution (Table 5). The pr^2 scores are particularly variable with a range of 0.22 to 0.60 (M31 included) or 0.67 to 0.91 (M31 excluded). Furthermore, there is virtually zero correlation between q^2 and either of the pr^2 scores so selecting an overall grid location based on q^2 does not guarantee the quality of the pr^2 scores. If the same set of analyses are repeated at a 1 Å resolution then much more consistent results are obtained (Table 5) – the pr^2 scores again cover the greatest range of values (∼0.05 units) and there is almost zero correlation between q^2 and either of the pr^2 scores. These results confirm that a 2 Å grid-resolution is too coarse to provide wholly reliable results while a 1 Å interval appears to be sufficient to provide consistent, reliable QSARs.

Table 5. Summary of CoMFA model reorientation tests for robustness[a]

Grid resolution	LOO crossvalidated q^2			Fitted model r^2			Test set pr^2		
	Range	Mean	StDev	Range	Mean	StDev	Range	Mean	StDev
2 Å	0.76 – 0.90	0.85	0.022	0.88 – 0.96	0.93	0.010	(0.22 – 0.60	0.44	0.046)[b]
							(0.63 – 0.91	0.83	0.037)[c]
1 Å	0.85 – 0.88	0.86	0.005	0.93 – 0.94	0.93	0.002	(0.42 – 0.48	0.45	0.010)[b]
							(0.81 – 0.86	0.84	0.008)[c]

[a] CoMFA PLS scores obtained when all structures are reoriented as an aggregate rigid body within the bounding 3D-grid. Starting with the original orientation, the aggregate was reoriented 1300 times through randomly selected angles in the x, y and z planes. The aggregate also was reoriented through 359° in 1° intervals in each plane separately (an extra 1078 PLS analyses); the overall range, mean and distribution for these analyses was identical (to 2 d.p.) to the randomised results.
[b] Test set including M31.
[c] Test set excluding M31.

The application of CVR^2-GRS (results not given) does not significantly alter PLS scores (q^2, r^2, pr^2 differ by up to 0.02 units) until the q^2 cut-off threshold for region exclusion is raised to 0.8 (leaving only three sub-regions in the analysis) whence the pr^2 scores drop dramatically to 0.04 (or 0.48 excluding M31). Thus, while CVR^2-GRS does not enhance PLS scores it certainly provides simplified (reduced-variable) models which may aid in interpretation. However; model reliability seems to be a consequence of the use of a 1 Å grid and CVR^2-GRS does not provide enhanced training or test set PLS scores.

The application of CoMFA to the MOPAC optimised structures used for EVA descriptor calculation provides q^2 and pr^2 scores (Table 4) that are slightly, but not significantly, lower than those obtained with the original manually adjusted structures. These results are based upon a grid resolution of 1 Å which, judging by the evaluations described above, should provide reliable models in terms of grid sampling-related errors. Whilst it is difficult to summarise the differences between the pre- and post-MOPAC optimised compounds some pertinent information is given in Table 6. The relative internal strain energies, whilst meaningless in absolute terms, indicate substantial relaxation of the structures after MOPAC optimisation. The pair-wise RMS fit values (based on the 3, 5, 6, 13, 14 and 17 superposition atoms only) also confirm that the differences between the compounds are very small. There is some variation in the orientation of the C17 side chain (where present) as judged by the C13-C17-C20-O26 torsion angle (this angle was manually adjusted to ∼60° in the original structures [6]) indicating that the C20-O26 moieties are not quite as well overlaid after MOPAC optimisation. However, if this torsion is again manually adjusted (either to 60° as before or ∼90° as suggested by the MOPAC minima) and CoMFA repeated there is no significant change in the resultant PLS training or test set scores suggesting that exact overlay of these groups is not important for CoMFA.

Comparison of EVA and CoMFA results

As noted above the 1 Å CoMFA results with the original and AM1-optimised structures are not significantly different (Table 4). The main intention behind assessing the conformational differences was to provide an alternative view-point of the relative performance of EVA and CoMFA on identical conformations. Clearly, the CoMFA results are in both cases a little better than those with EVA having q^2 scores of either 0.83 or 0.87 compared to 0.80 for the EVA 4 cm^{-1} model (Table 2). More significant differences between the EVA and CoMFA models are apparent when test set predictions are considered (final columns of Tables 2 and 4). Where M31 is included CoMFA predicts very poorly overall with pr^2 scores around 0.45 or 0.43 (Table 4) whereas the optimal EVA model has a pr^2 score of 0.69 (Table 2). When M31 (the fluorine-containing structure) is excluded from consideration the EVA prediction is slightly improved (to 0.74) while the CoMFA predictions are dramatically improved to 0.84 (non-MOPAC optimised geometries) or 0.81 (AM1 geometries). CoMFA over-estimates the activity of M31 by 1.8 log units or more (depending on the model considered) while EVA over-predicts by only 0.67 units (Table 7). The quality of the EVA prediction for M31 is in stark contrast to those of other QSAR techniques including CoMFA and, for example, COMPASS [14] and MS-WHIM [20], where the published residuals for M31 are +1.98 and +1.70 respectively.

Table 6. Comparison of the steroid structures pre- and post-AM1 geometry optimisation

Compound	RMS distance[a]	Torsion angle[b]		Strain energy[c]	
		Pre-AM1	AM1	Pre-AM1	AM1
M1	0.086	−60.7	−53.4	216.8	42.1
L2	0.112	–		164.6	39.1
L3	0.083	–		156.2	34.1
L4	0.121	–		165.6	34.0
L5	0.083	–		181.1	44.9
H6	0.109	59.3	88.6	345.0	46.5
H7	0.103	59.3	91.8	353.3	45.3
M8	0.083	59.3	91.5	136.1	36.0
L9	0.080	–		172.6	35.4
H10	0.078	59.3	87.6	158.5	42.3
H11	0.086	59.3	92.3	166.7	50.6
M12	0.131	–		157.5	37.2
L13	0.091	–		139.4	43.7
L14	0.090	–		141.0	44.9
L15	0.081	–		151.4	45.0
L16	0.109	–		196.4	41.5
L17	0.096	59.3	89.8	155.6	29.2
L18	0.096	59.3	89.5	166.1	38.4
H19	0.120	59.3	89.4	148.1	28.7
H20	0.082	59.3	96.5	158.5	41.1
M21	0.132	–		147.5	32.3
H22	0.119	59.3	90.4	361.4	54.6
H23	0.097	59.3	87.4	344.5	41.2
M24	0.104	59.3	90.7	117.5	16.0
H25	0.072	59.3	93.2	171.1	38.8
M26	0.090	–		92.9	27.1
M27	0.096	59.3	92.8	158.9	38.3
H28	0.101	59.3	88.3	145.7	27.3
M29	0.081	59.3	90.6	93.5	23.1
H30	0.104	59.3	90.6	357.5	50.1
M31	0.105	59.3	89.1	365.4	72.3

See Table 1 for further information on structures.

[a] RMS distance based on the 3, 5, 6, 13, 14 and 17 atoms for each pair of pre- and post-MOPAC optimised structures.

[b] Torsion angle based on C13-C17-C20-O26 atoms (Figure 6) where present.

[c] The relative strain energies (kcal/mol) are calculated using the Tripos 5.2 force-field released with Sybyl 6.3 [9]; electrostatic terms were included.

This result suggests that the EVA descriptor has a different information content to the steric and electrostatic CoMFA fields and the other noted descriptors. Alternatively, in the case of CoMFA and COMPASS, the alignment rules utilised may be incorrect.

The PLS residuals (Table 7) indicate that CoMFA has over-predicted in every case and that, at least in absolute terms, the EVA and CoMFA predictions in many cases have a similar predictive residual. The obvious exceptions are the aforementioned M31 and H22 – the latter has the poorest EVA and fourth poorest CoMFA prediction. Prednisolone (H22) is topologically almost identical to cortisol (H7) but has an extra double bond in the steran A-ring which is not present in any training set structure and which results in a slightly lower binding activity than H7. Both CoMFA and EVA poorly predict M27, which is identical to H20 except that it has a position-16 hydroxy group, the presence of which substantially reduces activity (1.5 log units difference). This substituent is explicit

Table 7. External test set predictions

Molecule	Experimental log [K]	EVA 4 cm^{-1}		CoMFA			
		Prediction	Residual	2 Å Resolution		1 Å Resolution	
				Prediction	Residual	Prediction	Residual
H22	7.51	6.47	−1.04	7.84	+0.33	7.69	+0.17
H23	7.55	7.81	+0.26	7.59	+0.04	7.73	+0.18
M24	6.78	6.91	+0.13	6.97	+0.19	7.02	+0.24
H25	7.20	6.92	−0.28	7.63	+0.43	7.67	+0.47
M26	6.14	5.87	−0.28	6.35	+0.21	6.39	+0.25
M27	6.25	7.17	+0.92	7.20	+0.95	7.21	+0.96
H28	7.12	7.03	−0.09	7.20	+0.08	7.18	+0.06
M29	6.82	6.38	−0.44	7.12	+0.30	7.13	+0.32
H30	7.69	7.61	−0.08	7.95	+0.26	7.71	+0.02
M31	5.80	6.47	+0.67	7.94	+2.14	7.71	+1.91

Predictions made using the EVA 4 cm^{-1} and CoMFA combined-field analyses using the respective 'optimal' two-component models.

in the training set in L14 (estriol) alone, which is one of three inactive estrogens (L13–L15). So, whilst it appears from the test set that this C16 hydroxy group modifies binding affinity, there is no obvious reason why training set PLS should weight variance due to the presence of this group to be explanatory for CBG-binding activity; hence, the poor M27 predictions might be expected.

Quality of normal mode frequencies

Plots such as those in Figure 3 emphasize the errors in calculation of normal mode frequencies associated with the MOPAC AM1 Hamiltonian and other semi-empirical methods. For example, carbonyl stretches are expected from experiment to appear at around 1700 cm^{-1} whilst here such stretches are represented by the peaks at 2040 to 2080 cm^{-1} (deoxycortisol). Provided, however, that such errors are consistent, which is the case with this steroid dataset where the structures are very similar and contain a very limited range of functional groups, then this should not be a problem. However, where more heterogeneous sets of structures are concerned the calculated frequencies may be more erratic [34] which is expected to have detrimental effects on a QSAR study. However, reasonable EVA QSARs previously have been obtained with heterogeneous datasets [2], although these certainly were not as good as those obtained herein with the steroids but then neither were the respective CoMFA results [2].

One simple method of treating such errors is to apply scale factors so as to match the calculated frequencies more closely to experimental values. Scott and Radom [34] and references therein provide details of such schemes – the recommended scale factor for AM1-based frequencies is 0.9532. Application of such scaling should aid one familiar with experimental IR spectra in the interpretation of an EVA QSAR model whilst at the same time have little or no effect on the QSAR statistical results. Intuitively the only difference may be to alter the EVA σ at which the optimal model is obtained. This assumption was in fact tested using the scale factor noted above and a range of descriptor sets derived using various σ values. The best models were again obtained where $\sigma = 3/4\ cm^{-1}$ with identical LV_{opt} and PLS scores (2 d.p.) to those originally obtained (Table 2). Only where both heterogeneous datasets are used and localised re-scaling of the frequencies is carried out is the application of scale factors expected to make a difference to the statistical results of a QSAR study.

In order to provide some indication of the effect of using different levels of theory to calculate normal mode frequencies the steroid set was treated using the following four methods: (1) MM3(94) [35] molecular mechanics; (2) MOPAC 6.0 PM3 [25] and (3) AMSOL 6.1.1 PM3 [36] semiempirical mechanics; and (4) a minimal basis set ab initio method – RHF/STO-3G. For the two semiempirical methods the original Wagener et al. [6] CoMFA conformations were used as starting points for geometry optimisation. However, in order to minimise computational effort (the steroid structures are quite large systems with between 20 and 29 heavy atoms) PM3 geometries were used as starting points for the RHF-level calculations. The keywords

Table 8. PLS modelling results based on various sources of normal mode frequencies

Source		Best σ (cm^{-1})	LV_{opt}	LOO CV		Fitted model		Test set predict	
				q^2	SE_{CV}	r^2	SE	pr^2	M31 residual
MOPAC 6.0	AM1	3 or 4	2	0.80	0.55	0.96	0.24	0.69 (0.74)	+0.67
	PM3	4	2	0.84	0.50	0.97	0.23	0.56 (0.75)	+1.36
AMSOL 6.1.1	PM3	4	2	0.81	0.53	0.97	0.22	0.68 (0.84)	+1.24
MM3(94)		3	2	0.83	0.52	0.98	0.18	0.74 (0.76)	+0.54
HF/STO-3G		3	2	0.81	0.54	0.98	0.18	0.64 (0.80)	+1.20
CoMFA (1 Å)		–	2	0.87	0.45	0.93	0.32	0.45 (0.84)	+1.91

See Table 2 and the main text for further information. The optimal models were chosen on the basis of the SE_{CV} minima which in all cases corresponded to the q^2 maxima.

for the MOPAC and AMSOL PM3 optimisations were either identical or essentially the same as those used for the AM1 runs. The structures for the RHF/STO-3G runs were geometry optimised using, first, SPARTAN [37] PM3 and then SPARTAN RHF/STO-3G with the default convergence criteria in both cases. Further geometry optimisation using the default convergence criteria and a normal coordinate analysis were then performed using GAUSSIAN 92 [38]. Since some of the frequency values exceeded 4000 cm^{-1} all of the resulting RHF/STO-3G frequencies were scaled using a factor of 0.90.

The EVA PLS results for this study (Table 8) indicate that with this dataset there are generally no great differences in either the q^2 scores (range 0.80 to 0.84) or the optimal σ at which these models are found (3/4 cm^{-1}). The pr^2 scores show greater variation, ranging from 0.74 to 0.84 (excluding M31) and from 0.56 to 0.74 (all 10 compounds). The AMSOL PM3 and RHF/STO-3G methods provide the best pr^2 scores where M31 is excluded (0.84 and 0.80 respectively) and these scores closely match the estimate of predictivity given by the respective q^2 scores (both 0.81). In all cases the predictions made for M31 by these models are considerably better than that with CoMFA, although the aforementioned AMSOL and RHF/STO-3G level models give much poorer predictions of M31 (+1.2 residual) than either the MOPAC- or MM3-originated models. However, it could be argued that MM3(94) gives the best result since the pr^2 with all ten test set compounds is best of all (0.74 compared to 0.69 with AM1). Overall, it appears that with this set of structures there is little or no benefit to be had from using higher levels of theory to determine the normal mode frequencies.

Visualisation and interpretation of EVA PLS results

In this section brief consideration is given to how an EVA QSAR model might be visualised and interpreted. It is possible to display the results of a PLS analysis in a number of ways [28]. With CoMFA 3D iso-contour plots can be used to visualize those regions of space indicated by PLS to be most highly positively or negatively correlated with bioactivity (Y). With EVA, however, such straightforward 3D visualisation is not possible. However, 2D plots of, for example, the magnitudes of the regression coefficients (B) or more sophisticated scores such as the VIP (Variable Influence on Projection) measure [31] can be used to indicate the relative importance of the x variables for modeling Y. The VIP is given by:

$$VIP_k = \sqrt{\sum_a w_{ak}^2 \cdot SSY_a \cdot \left[K/(SSY_{tot,expl} \cdot A)\right]} \quad (3)$$

where w_{ak} is the weight of the k^{th} explanatory variable in the a^{th} LV dimension, A is the total number of model LVs, K is the total number of explanatory variables, SSY_a is the amount of sum of squares of Y explained in the a^{th} dimension, and $SSY_{tot,expl}$ is the total sum of squares of Y explained in all LV_{opt} dimensions.

Plots of the B and VIP values across the spectrum for the optimal EVA 4 cm^{-1} model (Figure 7) provide an interpretation mechanism for an EVA QSAR that is analogous in some senses to the interpretation of an experimental IR spectrum. It should be reiterated that peak heights represent the relative importance of EVA variables for fitting Y and are not related to vibrational intensity. The main feature of these plots is that, for both the B and VIP measures, variables from all regions of the spectrum for which vibrational fre-

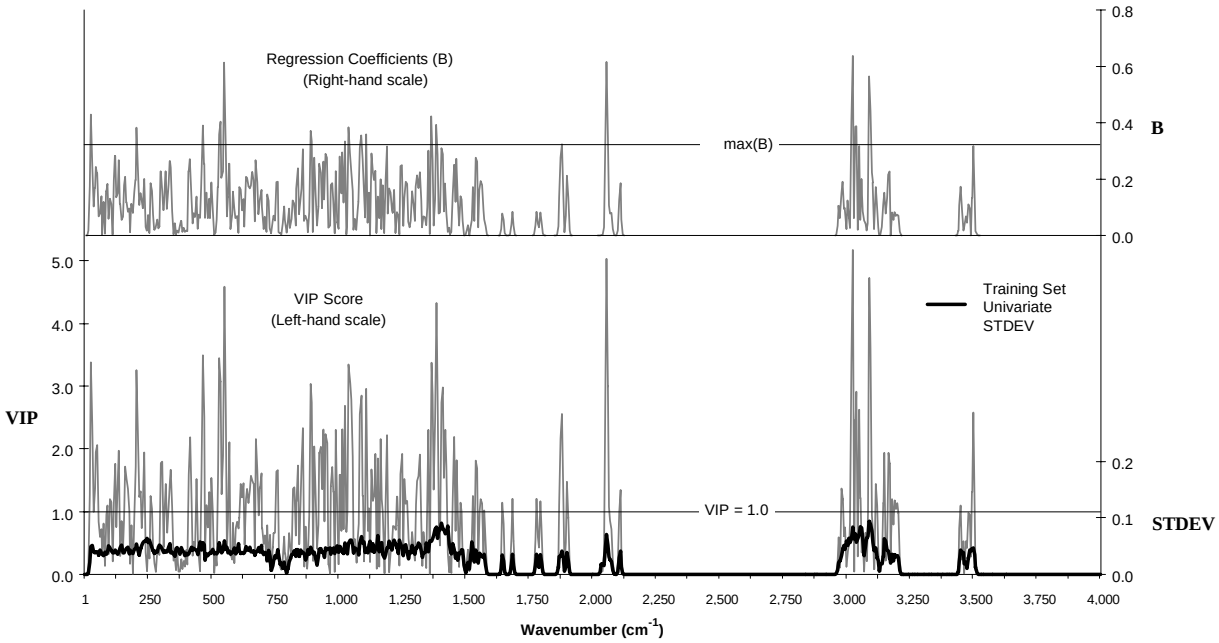


Figure 7. Regression coefficient (B) and VIP score plots: EVA 4 cm^{-1} two-LV fitted model. STDEV is the univariate standard deviation of the EVA descriptor (training set structures only).

quencies are present (essentially those where STDEV > 0) are significantly weighted by PLS.

The absolute magnitude of the B and VIP scores can be used to select the most important variables in an analysis [31]. Thus, for example, the carbonyl stretches (here at $\approx$2050 cm^{-1}) are heavily weighted (Figure 7) which is reasonable given that they are considered to be important determinants of binding strength (Figure 6). However, the heavy weighting of many other skeletal and hydrogen-stretching region frequencies is less readily explicable in terms of underlying structural features. The relative importance of various spectral sub-regions can be demonstrated by taking variables from such regions alone and performing training set PLS analysis and test set prediction. The PLS results (Table 9) indicate that the skeletal region itself can provide training and test set scores very close to those obtained with the full spectrum. The hydrogen-stretching region alone provides slightly poorer CV results ($q^2 = 0.71$) but relatively poor test set predictions (roughly 0.2 units lower than those of the skeletal region). The carbonyl-stretching region again provides a reasonable q^2 (0.70) but test predictions are very poor ($pr^2 \approx -0.05$ both including and excluding M31). So it appears that although the carbonyl stretching frequencies provide significant information for CV (where the low or inactive compounds tend not to have carbonyl groups) the weightings are not useful for predicting test set compounds (all of which have the requisite carbonyl groups); this is a reflection of the lack of training/test set design.

Detailed interpretation of an EVA analysis requires that: (1) the variables most highly correlated with activity be identified; (2) these variables are decomposed to their contributory vibrational frequencies; and (3) the indicated normal mode vibrations are examined (perhaps animated) to enable interpretation. This process is simplified if the number of variables to consider is minimised and where only one frequency per variable is indicated. Variables can be excluded according to – or VIP thresholds and a number of rules-of-thumb have been suggested [31]. A variable (x_k) can be considered important if its coefficient (b_k) is $>$ 1/2 max(B) or where $VIP_k > \sim 1.0$ – unimportant variables have $VIP_k < \sim 0.8$. It is clear (Figure 7) that each of the fingerprint, functional group and hydrogen-stretching regions provide important variables. The VIP method indicates a large number of variables to be important – 215 have VIP $\geq$ 0.8 while 183 have VIP $\geq$ 1.0 – whilst the B threshold indicates far fewer variables – 23 have b $\geq$ 1/2 max(B). The latter are a subset of those 33 variables for which VIP $\geq$ 2.5 so these measures essentially indicate the same set of variables to be the most important. In order

Table 9. PLS statistics: Importance of main spectral regions

Selection criterion	Wavenumber range (cm^{-1})[a]	LV_{opt}	LOO CV			Fitted model			Predict pr^2
			q^2	SE_{CV}	F	r^2	SE	F	
All regions	1 – 4000	2	0.80	0.55	12.0	0.96	0.24	224.1	0.69 (0.74)
Skeletal only	1 – 2000	2	0.77	0.59	10.3	0.96	0.23	245.5	0.71 (0.76)
Carbonyl stretch only	2000 – 2150	1	0.70	0.66	6.9	0.76	0.59	59.7	<0 (<0)
Hydrogen-stretch only	2875 – 3625	3	0.71	0.68	7.5	0.92	0.35	67.4	0.39 (0.57)
No carbonyl stretch	1 – 2000 2875 – 3625	2	0.79	0.56	11.5	0.96	0.24	231.0	0.69 (0.74)
No skeletal	2000 – 3625	2	0.71	0.67	7.30	0.89	0.41	71.4	0.43 (0.50)
No hydrogen-stretch	1 – 2150	2	0.79	0.57	11.1	0.96	0.23	245.8	0.71 (0.77)

EVA 4 cm^{-1} descriptors. See Table 2 for further information.
[a] Wavenumber range covered by the stated criterion (N.B., the wavenumbers have *not* been scaled to more closely match experiment).

Table 10. PLS statistics: Influence of variable selection

Selection criterion	NVARS[a]	LV_{opt}	LOO CV			Fitted model			Predict pr^2
			q^2	SE_{CV}	F	r^2	SE	F	
All variables	800[b]	2	0.80	0.55	9.3	0.96	0.24	224.1	0.69 (0.74)
VIP $\geq$ 0.8	215	2	0.86	0.46	18.8	0.96	0.24	227.3	0.67 (0.72)
VIP $\geq$ 1.0	183	2	0.87	0.45	19.2	0.96	0.25	212.9	0.67 (0.71)
VIP $\geq$ 2.5	33	2	0.84	0.48	15.7	0.89	0.41	147.2	0.62 (0.63)
VIP $\geq$ 3.0	17	2	0.85	0.47	17.3	0.92	0.35	105.2	0.54 (0.54)
$B \geq 1/2B_{max}$	23	2	0.88	0.43	21.9	0.94	0.31	134.8	0.69 (0.72)

EVA 4 cm^{-1} descriptors. See Table 2 for further information.
[a] Number of variables selected by the stated criterion.
[b] Of these, 375 have zero variance.

to determine the extent to which the variable subsets selected by these criteria summarise the information content of the full all-variable model PLS modeling was applied (Table 10). It is apparent that for the five cases tested the effect of variable selection is to enhance q^2 scores by up to 0.08 units. On the other hand the corresponding pr^2 scores tend to gradually worsen as the VIP threshold is raised. However, the use of the 1/2 max(B) threshold provides both the greatest enhancement in q^2 scores (0.88) and almost identical pr^2 scores to the all-variable model but using only 23 of the original 345 with non-zero variance.

Overall the variable-selection results suggest that these, or perhaps more sophisticated methods such as a method analogous to CVR^2-GRS [8], GOLPE [39], IVS-PLS [40] and others [41], can provide a simplified EVA PLS model that contains essentially the same information as a full-variable model. This must be considered a useful feature when it comes visualising the underlying normal mode vibrations whilst attempting to interpret an EVA QSAR. However, interpretation is further complicated since most of these variables come from frequency-dense spectral regions and as such encode information from two or more normal mode vibrations (intra-structural kernel overlap).

The extent of intra-structural kernel overlap can be assessed (Figure 3) by looking at regions of the spectrum for which the EVA descriptor values exceed the maximum possible height for a single isolated kernel – this value is 0.10 (obtained by putting $\sigma = 4$ and $x = f_i$ in Equation 1). Note that peaks with apparent maxima less than this threshold occur because actual (original kernel) maxima and sampling points need not coincide. Clearly for deoxycortisol (and all the steroids herein) there is substantial overlap in a number of spectral regions, with a maximum close to 1400 cm^{-1}, a number of lesser peaks to either side of this, and further maxima in the range 3000–3200 cm^{-1}. Comparison of these *high density* regions with the B and VIP plots (Figures 7 and 8) shows that many (but not all) of the most important variables

come from these regions. This suggests that these variables may be favoured at least in part simply because of their relatively high variance. A priori one possible solution to this problem is to autoscale the variance of all variables to unity which should reduce the importance of such high density regions. However, it is clear that autoscaling has a significant detrimental effect on both training and test set PLS scores (Table 2) and, for reasons which are at present unclear, autoscaling is certainly not a useful procedure with this dataset.

Conformational sensitivity of EVA

In this section consideration is given to the conformational sensitivity of EVA QSAR. CoMFA can be extremely sensitive to small conformational differences and the relative superposition of structures; this sensitivity, associated problems and possible solutions have been discussed extensively in the literature [7,8,17,21,42,43]. The EVA descriptor is, of course, completely invariant to rotation and translation of the structures concerned. This is only the case where modeling and NCA are done (as is usually the case) without external modifying factors which would be present should, for example, analysis be done in a receptor binding cavity.

In the work already described the conformations used as the starting points for MOPAC geometry optimisation (a necessary step prior to NCA) were those aligned for the purposes of CoMFA. This means that some effort had been made to match the conformations of the structures concerned, a necessary step in CoMFA, for which both conformation selection and superposition of those conformations is required. These two factors are of course inter-related and can be complicated by the issues associated with grid-resolution quantified above. With EVA, of course, there are no superposition problems and the main questions dealt with here are: 'does it matter what conformations are used in an EVA analysis?' and, if so, 'how sensitive is EVA to conformation?' Intuitively, a conformational alteration (that remains after geometry optimisation) will result in a change to the force constants and thence the normal mode frequencies and displacements obtained. This should then result in a change to the EVA descriptor and, if the differences are sufficient, significant alteration of the PLS weights and resultant QSAR scores. It is also intuitive that large differences in conformation are liable (but not certain) to result in larger differences in the derived vibrational frequencies than small conformational differences. Therefore, in order to focus EVA analysis upon structural differences and the relationship of these to changes in bioactivity it makes sense to use conformations that have been matched in a CoMFA-like manner – precise superposition (a frequently time-consuming and/or ambiguous step in CoMFA) is, however, not necessary. Furthermore, it is expected that there is a relationship between EVA σ and conformational sensitivity since σ determines the extent to which there is overlap of vibrations in different compounds with similar wave number. A large σ is expected to permit the EVA model to be less sensitive to small differences in wave number that are related to conformational differences; this feature is demonstrated below.

The conformational sensitivity of EVA can be demonstrated simply by taking an alternative conformation of one or more steroids and noting predictive differences for these structures relative to that of the original training or test set compound. There are a number of ways to generate alternative conformations and two methods have been chosen here. For the first method conformations were generated entirely de novo from 2D SMILES strings whilst the second method involved tweaking flexible bonds in an original conformation to produce alternative conformations.

For method one all 31 steroid conformations were regenerated from SMILES 2D representations and 3D structures obtained using Concord [44] rather than Corina which was used to generate the original conformations [6]. Unlike in Wagener et al., the orientation of the C17 side-chains was not adjusted (Figure 6) – this was done for CoMFA in order to best overlay the local carbonyl moieties. However, the specified orientation (α-/β-designations) of attachments (Figures 1 and 2) relative to the steran ring-plane was retained. These compounds were submitted to MOPAC 6.0 AM1 for further geometry optimisation and NCA using the keywords described previously. The most obvious difference between these (*non-matched*) conformations and the original (*matched*) structures relates to the orientation of the C20 carbonyl group which, where present, shows a great deal of variation as judged by the C13-C17-C20-O26 torsion angle (Table 11). An EVA analysis using these non-matched compounds again provides an optimal model where $\sigma = 3/4\,\mathrm{cm}^{-1}$; there is a slightly improved q^2 of 0.83 but pr^2 scores that are 0.10 units poorer than the original matched analysis (Table 2); viz., 0.65 (M31 excluded) and 0.62 (M31 included). If the C13-C17-C20-O26 torsion is adjusted

Table 11. Comparison of the steroid structures derived from Corina and Concord after AM1 geometry optimisation

Compound	RMS distance[a]	Torsion angle[b]		Strain energy[c]	
		Concord	Original	Concord	Original
M1	0.098	108.2	−53.4	41.4	42.1
L2	0.016		–	37.8	39.1
L3	0.034		–	33.3	34.1
L4	0.008		–	33.8	34.0
L5	0.032		–	41.0	44.9
H6	0.091	90.2	88.6	35.4	46.5
H7	0.110	−82.8	91.8	45.3	45.3
M8	0.132	−101.4	91.5	38.5	36.0
L9	0.012		–	35.4	35.4
H10	0.131	90.9	87.6	37.2	42.3
H11	0.090	−83.9	92.4	47.5	50.6
M12	0.009		–	36.3	37.2
L13	0.024		–	28.7	43.7
L14	0.024		–	29.0	44.9
L15	0.012		–	30.2	45.0
L16	0.008		–	41.6	41.5
L17	0.007	89.7	89.8	29.1	29.2
L18	0.016	−81.4	89.5	41.8	38.4
H19	0.035	89.5	89.4	27.6	28.7
H20	0.094	−81.4	96.5	40.1	41.1
M21	0.010		–	31.8	32.3
H22	0.083	−70.3	90.4	45.1	54.6
H23	0.145	−22.7	87.4	33.9	41.2
M24	0.089	90.3	90.7	13.4	16.0
H25	0.076	90.6	93.2	34.7	38.8
M26	0.032		–	26.0	27.1
M27	0.088	−39.1	92.8	41.3	38.3
H28	0.062	88.4	88.3	25.4	27.3
M29	0.052	89.6	90.6	21.9	23.1
H30	0.099	79.9	90.6	47.5	50.1
M31	0.065	81.8	89.1	59.8	72.3

See Table 1 for further information on structures.

[a] RMS distance based on the 3, 5, 6, 13, 14 and 17 atoms (Figure 6) between each pair of matched and non-matched MOPAC optimised structures.

[b] Torsion angle (after MOPAC optimisation) based on C13-C17-C20-O26 atoms (Figure 6) where present.

[c] The relative strain energies (kcal/mol) are calculated using the Tripos 5.2 force-field released with Sybyl 6.3 [9]; electrostatic terms were included.

to 90° and the structures reoptimised using MOPAC as before all the torsion angles remain close to 90°. EVA PLS analysis again provides an optimal model where $\sigma = 3/4$ cm^{-1}, with a q^2 of 0.83, while test set pr^2 scores are enhanced by 0.04 units to 0.69 (M31 excluded) and 0.66 (M31 included). These results suggest that EVA modeling is enhanced where effort is made to match the conformations of the structures concerned.

The activities of the newly generated non-matched steroid conformations were then predicted using the training set models developed with the original matched conformations of Wagener et al. (Table 2); this was done for a range of Gaussian σ (Figure 9). If only an optimal EVA model is considered for now (i.e., where $\sigma = 3/4$ cm^{-1}) then it is clear that there is a difference in the PLS scores with the training set compounds (fitted-$r^2 = 0.96$) and the equivalent 21

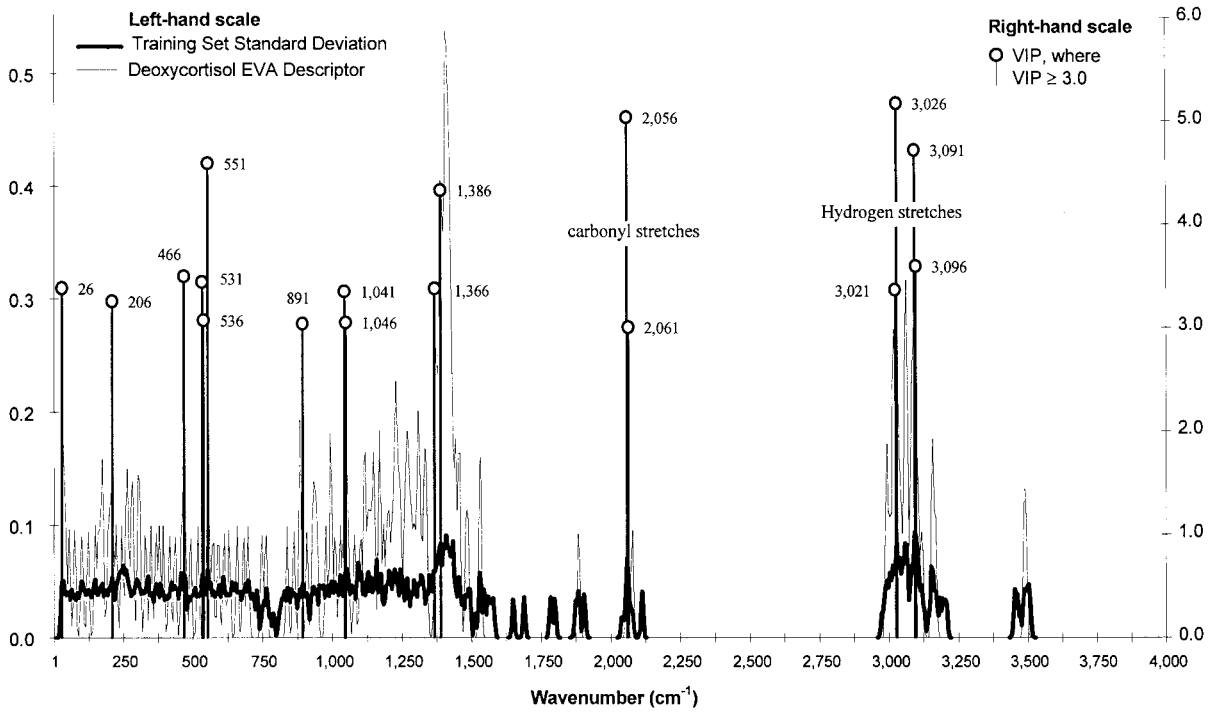


Figure 8. VIP score plot where VIP ≥ 3.0. This plot indicates the 17 variables for which VIP ≥ 3.0 in the EVA 4 cm^{-1} two-LV fitted model. The values beside the markers (○) are the wavenumbers (cm^{-1}) at which the variables were sampled. Also indicated is the training set univariate standard deviation.

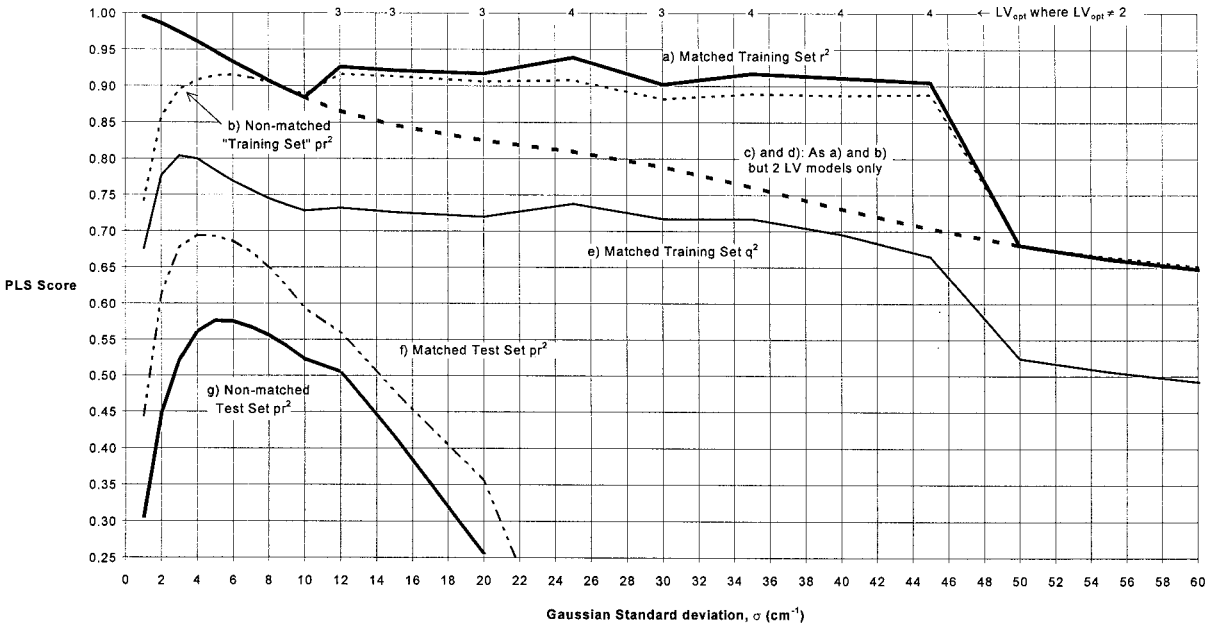


Figure 9. Matched and non-matched compound fitted and predictive PLS scores over a range of different Gaussian σ values: models derived using the original matched-compound training set; see main text for further details.

non-matched conformations ($pr^2 = 0.91$). The pr^2 score for the ten non-matched test set compounds is significantly poorer ($pr^2 = 0.56$) than that with the matched test set ($pr^2 = 0.69$). The corresponding CoMFA (1 Å grid resolution) pr^2 scores are 0.71 for the 21 non-matched compounds and, for the test set, 0.52 and 0.56 with and without M31 respectively; the non-matched compounds were overlaid as before using an RMS fit to the training set deoxycortisol conformation. Thus it is clearly demonstrated that EVA is sensitive to conformational differences and that CoMFA, in this instance, exhibits greater sensitivity to these differences than EVA; results with alternative σ are discussed below.

In a second approach to assessing conformational sensitivity a systematic conformational search was made in which the C17-C20 bond (Figure 6) of the original matched conformation of deoxycortisol was rotated about its axis through 360° in 15° increments. The resulting 24 conformations were submitted to MOPAC 6.0 AM1 for optimisation/NCA as before; six redundant conformers were identified and discarded. Predictions were then made for each of the remaining 18 conformers using the original Corina training set models, again for a range of EVA σ. If only the optimal 3/4 cm^{-1} EVA training set model is considered (Figure 10) then there is a strong correlation ($r^2 = 0.83$) between the C13-C17-C20-O26 torsion angle relative to the original matched compound torsion angle (ω) and the log [K] predictions (Table 12). With a couple of exceptions there appears to be little change in the prediction where $\omega < \sim\pm 50°$ (conformers 6–13) while where $\omega < -50°$ the prediction quality deteriorates significantly. The fact that there is not a smooth relationship between ω and the predictive quality reflects the fact that there is not a straightforward linear relationship between ω and the underlying normal mode frequency values. However, ω has been used as a rough guide to the extent of conformational difference. If the EVA descriptors for each of the 18 unique conformers are compared then it is apparent that there are substantial differences (variance) in all parts of the spectrum (Figure 11a); this is the case even when the two most similar conformations are compared (Figure 11b). In fact the only parts of the spectrum where there is training set univariate variance but zero variance for the 18 deoxycortisol conformers relates to vibrations unique to the three estrogen compounds (L13-L15) and the dienone compound (H22). Thus the localised conformational changes appear to have a modifying influence on all vibrational normal modes. Table 12 lists the carbonyl group stretching frequencies for both the C20 and the C3 carbonyl groups (Figure 6). In general these would be expected to be relatively invariant to conformation and this is clearly the case for the C3 carbonyl stretch which ranges from 2055.3 to 2056.4 cm^{-1} across the 18 deoxycortisol conformers. On the other hand the C20 carbonyl stretch varies from 2037.3 to 2076.9 cm^{-1} indicating that the stretch is quite sensitive to the local environment. Furthermore, these frequency values are correlated both with ω ($r^2 = 0.69$) and the corresponding EVA 4 cm^{-1} model predictions ($r^2 = 0.67$); that these correlations are not higher reflects the fact that regions of the molecule other than just the C17-C20 locale have conformational differences subsequent to MOPAC optimisation even with these relatively rigid steroids. Note that whilst it is possible to instruct MOPAC to geometry optimise specific bonds and angles the result of doing this here was to generate a number of structures with imaginary frequencies. Therefore, full geometry minimisation (ALL_BONDS_ANGLES) was performed for each of the initial 24 conformations; this explains the extent of the variation in the univariate EVA descriptor values of the 18 conformers across the spectrum (Figure 11a).

Relationship of EVA σ to conformational sensitivity

The relationship of EVA σ to conformational sensitivity can be assessed by again considering predictions of the activities of the newly generated non-matched steroid conformations using the models developed with the original matched structures. It is apparent that predictive differences are greatest where σ is very small (Figure 9, lines a and b) and that as σ is raised these differences gradually reduce until convergence at $\sim 8\ cm^{-1}$. Thereafter, there are some differences in the matched/non-matched r^2 and pr^2 scores (lines a and b) but these seem to be a function of the value of LV_{opt} (selected from the first SE_{CV}-minimum); values of LV_{opt} other than two are indicated on Figure 9. Thus, for example, if predictions are made using two LVs only (lines c and d) then it is clear that once σ exceeds a threshold value ($\sim 8\ cm^{-1}$) the predictive differences are very small (lines c and d are overlaid). In other words the two sets of conformations are indistinguishable where $\sigma > 8\ cm^{-1}$ and two LVs are employed (this is also the case if LV1 only is considered – not shown). The addition of a third or fourth LV (where indicated by CV and thus to some extent dependent upon the rule used to select LV_{opt}) immediately separates

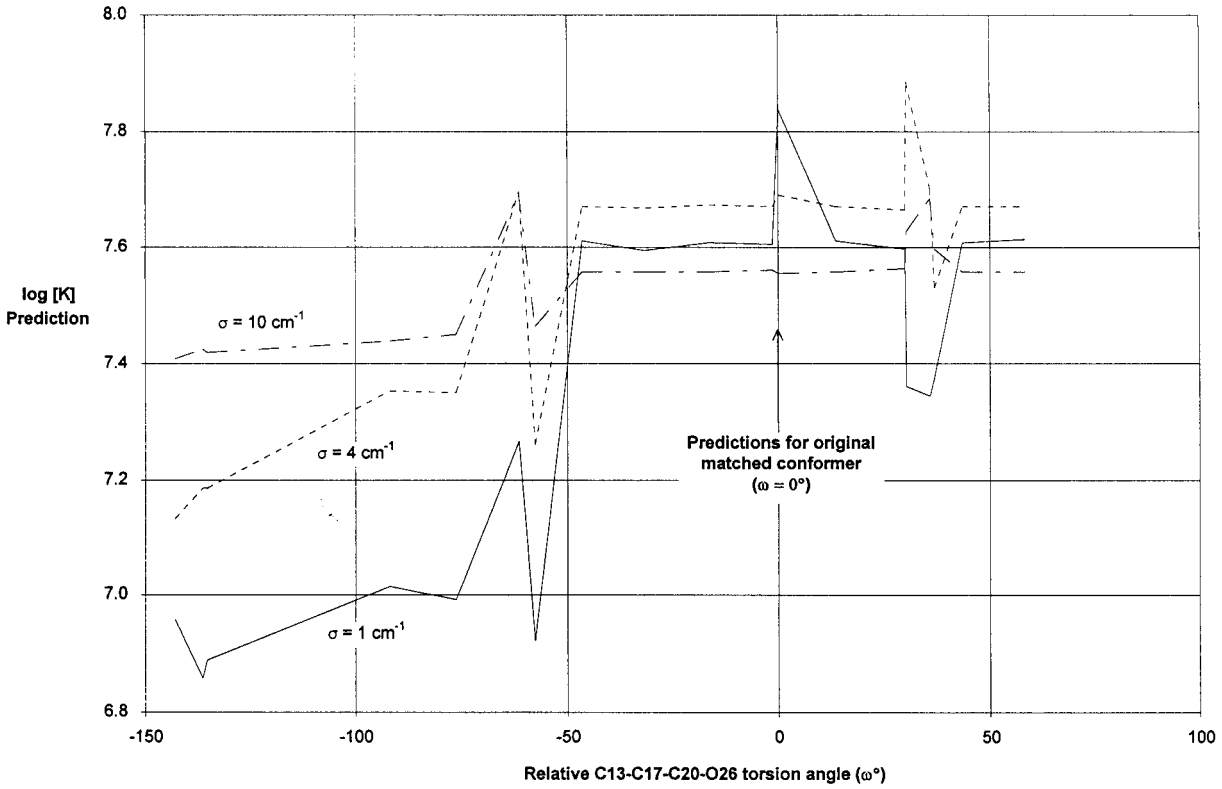


Figure 10. Prediction of deoxycortisol conformer log [K]: example of how a larger EVA Gaussian σ enables the predictive differences between conformers to be reduced. Sensitivity to conformational differences is greatest where $\sigma = 1$ cm^{-1} and the least where $\sigma = 10$ cm^{-1}. See main text for further details.

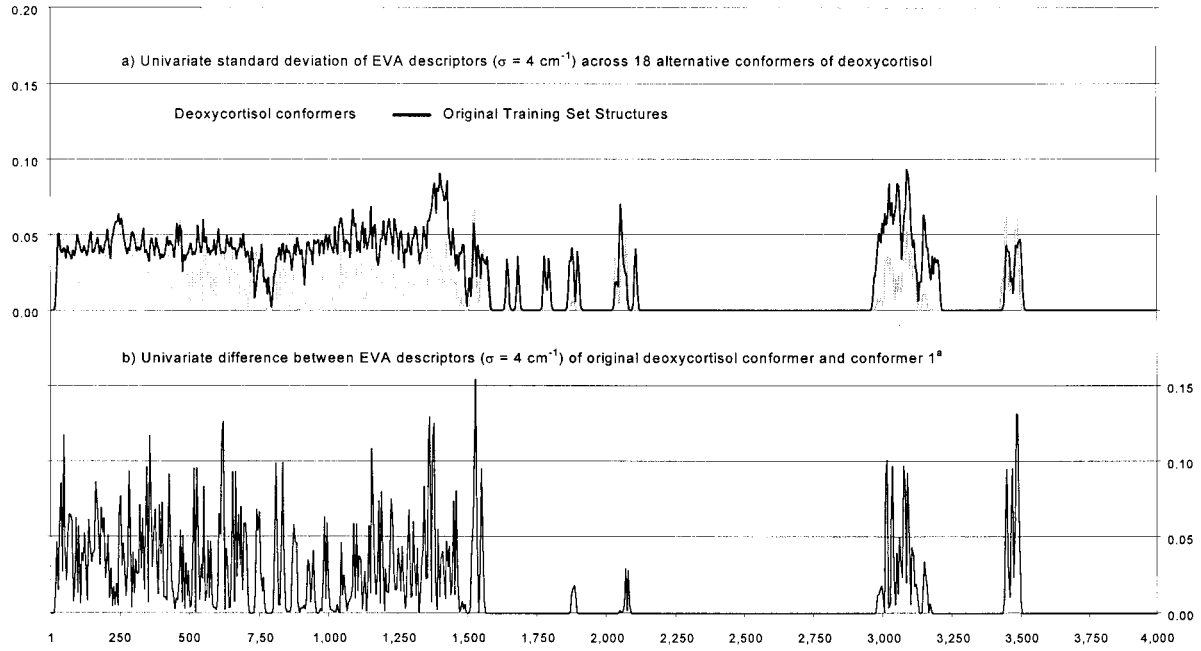


Figure 11. Univariate detail of the EVA descriptor for the 18 deoxycortisol conformers (see main text for further information). [a] See Table 12 for further detail about the various conformers.

Table 12. Detail of deoxycortisol conformations and predictions

Conformer	Torsion (ω°)[a]	C=O Stretching frequencies[b]		Activity predictions[c]	
		C20=O	C3=O	EVA 4 cm^{-1}	CoMFA
Original	0.0	2074.8	2055.3	7.69	7.67
1	−57.6	2076.9	2055.4	7.26	6.69
3	−92.0	2064.0	2055.3	7.35	6.85
4	−76.4	2064.2	2055.5	7.35	6.84
5	−61.4	2059.5	2055.8	7.70	6.94
6	−46.4	2074.8	2055.7	7.67	7.04
7	−31.4	2074.9	2055.6	7.67	7.15
8	−16.4	2074.8	2055.7	7.67	7.24
9	−1.4	2074.8	2055.6	7.67	7.34
10	13.6	2074.8	2055.7	7.67	7.40
11	29.8	2074.7	2056.1	7.66	7.18
12	43.6	2074.8	2055.6	7.67	7.25
13	58.6	2074.8	2055.6	7.67	7.16
16	30.2	2062.1	2056.1	7.89	7.24
17	35.9	2059.2	2056.3	7.70	7.19
18	37.0	2060.1	2056.4	7.53	6.85
22	−135.3	2037.3	2056.2	7.18	6.83
23	−143.2	2037.7	2056.3	7.13	6.87
24	−136.4	2038.0	2055.9	7.18	6.72
Maximum		2076.9	2056.4	7.89	7.67
Minimum		2037.3	2055.3	7.13	6.69
StDev		13.6	0.3	0.22	0.26
Mean		2064.9	2055.8	7.54	7.08

[a] C13-C17-C20-O26 torsion angle relative to that of the original matched conformation of deoxycortisol after MOPAC 6.0 AM1 geometry optimisation (redundant conformers have been discarded).

[b] Carbonyl group stretching frequencies (100% radial stretch) – see Figure 6 for atom-numbering scheme.

[c] Predictions made using EVA and CoMFA 1 Å two-LV optimal models (Tables 2 and 4).

(and improves) the predictions – such discontinuities are apparent, for example, where $\sigma = 12$, 25 and 50 cm^{-1} (Figure 9, lines a and b). It is also clear from this figure that choosing the EVA model (i.e., σ and LV_{opt}) using CV (line e) and favouring model parsimony quite definitely provides the best models as judged by the maxima (at $\sigma \approx 4$ cm^{-1}) in the test set predictions (lines f and g).

If the tweaked deoxycortisol conformers are now considered in relation to the sensitivity of EVA at various σ then it is apparent that at low σ the activity predictions for the 18 conformers exhibit a great deal of variance (Figure 12, line a); as σ is raised from 1 cm^{-1} the sensitivity to these differences is gradually reduced up to where $\sigma \approx 15$ cm^{-1}, after which there is little change in conformer discrimination. At the same time the overall quality (SE) of the predictions improves up to $\sigma \approx 8$ cm^{-1} (Figure 12, line b) and remains at roughly this level until $\sigma \approx 30$ cm^{-1}, beyond which there a deterioration in prediction quality. As stated above the C13-C17-C20-O26 torsion angle (ω) can be used as a crude measure of conformational differences between the conformers. A plot of ω vs. log [K] prediction at various σ (Figure 10) indicates that as σ is raised from 1 to 10 cm^{-1} the predictions for the conformers with the largest ω significantly improve. The various discontinuities in the plots reflect the fact that ω is only a rough guide to conformer differences over the whole structure.

Overall, these results indicate that the EVA σ term can be used to alter the sensitivity of the EVA descriptor to conformational differences although for this dataset there is a concomitant reduction in the quality of these predictions as σ is raised beyond an optimum

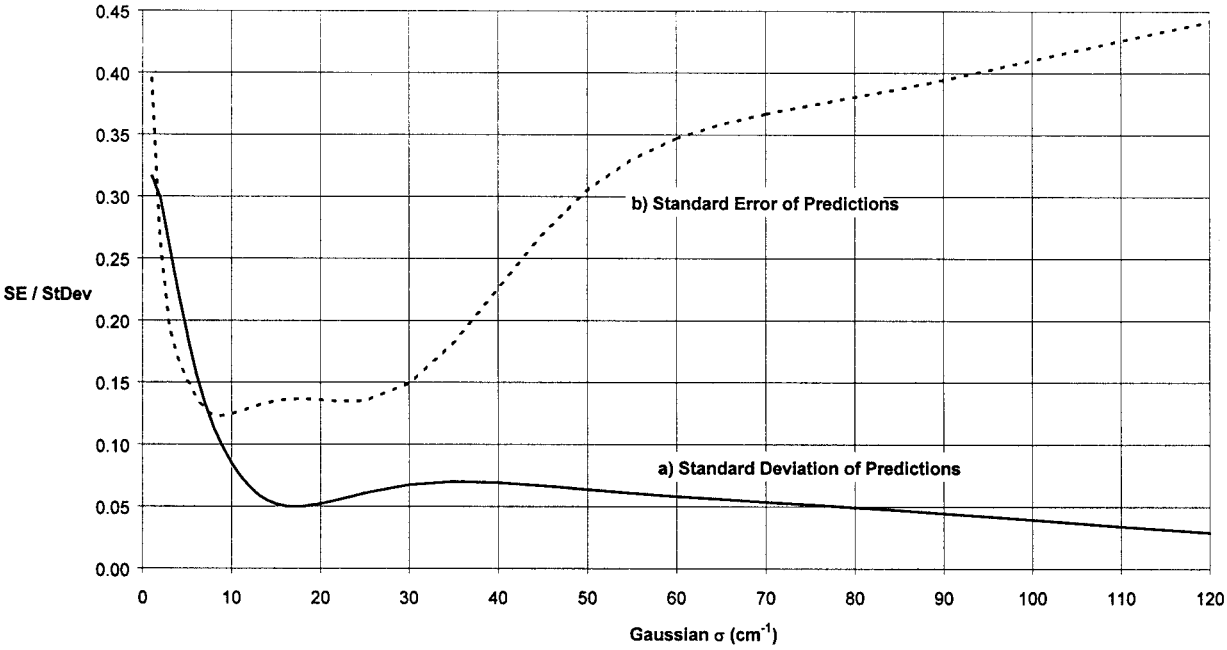


Figure 12. Predictive error and standard deviation of predictions for the 18 deoxycortisol conformers using EVA training set models derived using a range of different Gaussian σ values (see main text for further information).

value (here ∼4 cm^{-1}). Models based upon very low σ are quite sensitive to conformation while relaxation of the model through the use of larger σ means that progressively larger conformational differences can be better accounted for although this can only be the case where such differences result in a sufficiently small wave number change relative to a chosen σ value; the identification of such thresholds is not straightforward given that there is complex relationship between normal mode frequencies and measures of conformation difference.

Conclusions

This study has established that EVA can be used to develop highly predictive and robust QSAR models with a benchmark steroid dataset. The technique has been extensively validated in terms of both internal and external predictivity coupled with random permutation tests. In addition, the EVA method is found to be sensitive to inconsistencies in those features where, for example, there is a topological coding error for a structure. An exploratory and somewhat crude selection of the most important (heavily weighted) variables in an EVA PLS analysis has indicated that the high weighting of at least some of the underlying structural features can be straightforwardly related to those previously thought to be correlated with binding activity (Figure 6). With this steroid dataset there appears to be no predictive advantage to be gained by the use of higher levels of theory to geometry optimise compounds although, as discussed above, this is expected to be less likely the case with more heterogeneous sets of structures that exhibit a wider variety of functional groups.

The EVA descriptor has been shown to be conformationally sensitive whilst at the same time completely invariant to the relative rotation/translation of the structures concerned. A priori it is important to match conformations as closely as possible when developing a QSAR since this will mean that there is minimum variation in the frequency values resulting from conformational differences; this will permit PLS to focus upon those (structural) differences that are correlated with binding activity rather than conformation. However, it has been shown that the conformational sensitivity of the EVA technique can be adjusted though the appropriate selection of the Gaussian kernel width (determined by σ). At least with this steroid dataset a larger σ term enables the effect of conformational differences on predictivity to be reduced, albeit at the cost of some degradation to overall model quality when σ becomes quite large. This may be a

particularly useful feature with conformationally flexible and/or heterogeneous datasets where the choice of conformation to use may not be obvious; this is a topic for which further investigation is required.

In comparison to CoMFA, EVA has provided statistically comparable results, but without the difficulty and ambiguity associated with structural overlay. In this instance, EVA has been found to provide significantly better predictions than CoMFA for what has previously been considered a structural outlier in the external test set. This suggests that the EVA descriptor provides a different view of molecular structure to that of steric and electrostatic fields and EVA may therefore be a more effective descriptor in certain situations. Furthermore, EVA does not suffer from the sampling problems associated with CoMFA grids which has triggered the development of algorithms such as crossvalidation R^2-guided region selection [8]. With EVA, provided that the sampling increment, L, does not exceed critical threshold values (L_{crit}) for a given σ [2], then problems related to sampling do not arise. The essential difference between the two techniques is that with EVA a single set of conformations can only result in one statistical model whilst with CoMFA there will always be more than one possible overlay of the chosen conformations and thence more than one statistical model. However, with EVA the Gaussian kernel width (σ) needs to be selected through the repeated application of crossvalidation to various descriptor sets; this process is readily automated and extremely rapid when SAMPLS [26] is used.

The main deficiency of the EVA technique in comparison to CoMFA is that a means of effectively interpreting and displaying the structural features (contra-)indicated by a PLS model has yet to be developed to the extent that it has in CoMFA. However, there remain difficulties with interpretation of the CoMFA 3D iso-contour plots since it is not always obvious which atoms/features should be changed to alter activity [45]. The development of EVA model interpretation techniques is the object of current study. This may be aided both by the use of a localised Gaussian kernel width, so as to minimize as far as possible intra-structural vibrational frequency overlap in different parts of the spectrum [46], and by the application of variable-selection techniques to reduce the number of variables and hence normal mode vibrations to consider. However, as demonstrated herein it is the densely populated skeletal region of the spectrum that provides the most internally and externally predictive EVA variables, a fact that makes decomposition of significant variables all the more complex.

Acknowledgements

We thank Shell Research Centre, the Science and Engineering Research Council and the Biotechnology and Biological Sciences Research Council for funding; Tripos Inc. for the provision of hardware and software; Prof. Johann Gasteiger (Universität Erlangen-Nürnberg, Germany) for providing the steroid dataset [6]; Bruno Bienfait (National Cancer Institute, Bethesda, U.S.A.) for providing the AMSOL PM3 normal mode frequencies; and Greg Paris (Novartis Research Institute, New Jersey, U.S.A.) for performing the ab initio calculations. This paper is a contribution from the Krebs Institute for Biomolecular Research, which is a designated Biomolecular Sciences Centre of the Biotechnology and Biological Sciences Research Council.

References

1. Ferguson, A.M., Heritage, T., Pack, S.E., Philips, L., Rogan, J. and Snaith, P.J., J. Comput.-Aided Mol. Design, 11 (1997) 143.
2. Turner, D.B., Willett, P., Ferguson, A.M. and Heritage, T., J. Comput.-Aided Mol. Design, 11 (1997) 409.
3. Cramer, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
4. Ginn, C.M.R., Turner, D.B., Willett, P., Ferguson, A.M. and Heritage, T.W., J. Chem. Inf. Comput. Sci., 37 (1997) 23.
5. Wold, S. and Eriksson, L., In van de Waterbeemd, H. (Ed.), Methods and Principles in Medicinal Chemistry, Vol. 2, Chemometric Methods in Molecular Design, VCH, Weinheim, Germany, 1993, pp. 309–318.
6. Wagener, M., Sadowski, J. and Gasteiger, J., J. Am. Chem. Soc., 117 (1995) 7769.
7. Klebe, G., Abraham, U. and Mietzner, T., J. Med. Chem., 37 (1994) 4130.
8. Cho, S.J. and Tropsha, A., J. Med. Chem., 38 (1995) 1060.
9. Sybyl 6.3, Tripos Associates Inc., St. Louis, MO, U.S.A.
10. Wold, S., Ruhe, A., Wold, H. and Dunn III, W.J., SIAM J. Sci. Stat. Comput., 5 (1984) 735.
11. Dunn, J.F., Nisula, B.C. and Rodbard, D.J., Endocrinol. Metab., 53 (1981) 58.
12. Good, A.C., So, S.-S. and Richards, W.G., J. Med. Chem., 36 (1993) 433.
13. Oxford Molecular Ltd., Oxford, U.K.
14. Jain, A.N., Koile, K. and Chapman, D., J. Med. Chem., 37 (1994) 2315.
15. Oprea, T.I., Ciubotariu, D., Sulea, T.I. and Simon, Z., Quant. Struct.-Act. Relatsh., 12 (1993) 21.
16. Hahn, M. and Rogers, D., J. Med. Chem., 38 (1995) 2091.
17. Silverman, B.D. and Platt, D.E., J. Med. Chem., 39 (1996) 2129.

18. Kellogg, G.A., Kier, L.B., Gaillard, P. and Hall, H.H., J. Comput.-Aided Mol. Design, 10 (1996) 513.
19. Anzali, S., Barnickel, G., Krug, M., Sadowski, J., Wagener, M., Gasteiger, J. and Polanski, J., J. Comput.-Aided Mol. Design, 10 (1996) 521.
20. Bravi, G., Gancia, E., Mascagni, P., Pegna, M., Todeschini, R. and Zaliani, A., J. Comput.-Aided Mol. Design, 11 (1997) 79.
21. Schnitker, J., Gopalaswamy, R. and Crippen, G.M., J. Comput.-Aided Mol. Design, 11 (1997) 93.
22. Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C., Mitchell, E.M., Mitchell, G.F., Smith, J.M. and Watson, D.G., J. Chem. Inf. Comput. Sci., 31 (1991) 187.
23. Austel, V., In van de Waterbeemd, H. (Ed.), Methods and Principles in Medicinal Chemistry, Vol. 2, Chemometric Methods in Molecular Design, VCH, Weinheim, Germany, 1993, pp. 49–62.
24. Sjöström, M. and Eriksson, L., In van de Waterbeemd, H. (Ed.), Methods and Principles in Medicinal Chemistry, Vol. 2, Chemometric Methods in Molecular Design, VCH, Weinheim, Germany, 1993, pp. 63–90.
25. MOPAC version 6.0., Quantum Chemistry Program Exchange (QCPE), Indiana University, Bloomington, IN, U.S.A.
26. Bush, B.L. and Nachbar Jr., R.B., J. Comput.-Aided Mol. Design, 7 (1993) 587.
27. Sadowski, J. and Gasteiger, J., J. Chem. Inf. Comput. Sci., 34 (1994) 1000.
28. Wold, S., In van de Waterbeemd, H. (Ed.), Methods and Principles in Medicinal Chemistry, Vol. 2, Chemometric Methods in Molecular Design, VCH, Weinheim, Germany, 1993, pp. 195–218.
29. Shao, J., J. Am. Stat. Assoc., 88 (1993) 486.
30. Topliss, J.G. and Edwards, R.P., J. Med. Chem., 22 (1979) 1238.
31. Wold, S., Johansson, E. and Cocchi, M., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design: Theory Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 523–550.
32. Jonathan, P., McCarthy, W.V. and Roberts, A.M.I., J. Chemometrics, 10 (1996) 189.
33. Mickelson, K.E., Forsthoefel, J. and Westphal, U., Biochemistry, 20 (1981) 6211.
34. Scott, A.P. and Radom, L., J. Phys. Chem., 100 (1996) 16502.
35. MM3(94) Manual (Version 1.0), Tripos Associates Inc., St. Louis, MO, U.S.A. This contains numerous references to the MMx series of programs developed by Norman Allinger and co-workers at the University of Georgia.
36. AMSOL (version 6.1.1), Oxford Molecular Ltd., Oxford, U.K.
37. SPARTAN (release 4.0 2b), Wavefunction Inc., Irvine, CA, U.S.A.
38. GAUSSIAN 92/DFT (Revision G.4), Gaussian Inc., Pittsburgh, PA, U.S.A.
39. Cruciani, G. and Clementi, S., In van de Waterbeemd, H. (Ed.), Methods and Principles in Medicinal Chemistry, Vol. 3, Advanced Computer-Assisted Techniques in Drug Discovery, VCH, Weinheim, Germany, 1995, pp. 61–88.
40. Lindgren, F., Geladi, P., Rannar, S. and Wold, S., J. Chemometrics, 8 (1994) 349.
41. Clementi, S. and Wold, S., In van de Waterbeemd, H. (Ed.), Methods and Principles in Medicinal Chemistry, Vol. 2, Chemometric Methods in Molecular Design, VCH, Weinheim, Germany, 1993, pp. 319–338.
42. Kroemer, R.T. and Hecht, P., J. Comput.-Aided Mol. Design, 9 (1995) 205.
43. Muresan, S., Sulea, T., Ciubotariu, D., Kurunczi, L. and Simon, Z., Quant. Struct.-Act. Relatsh., 15 (1996) 31.
44. CONCORD version 2.9.3., TRIPOS Inc., St. Louis, MO, U.S.A.
45. Waszkowycz, B., Clark, D.E., Frenkel, D., Li, J., Murray, C.W., Robson, B. and Westhead, D.R., J. Med. Chem., 37 (1994) 3994.
46. Turner, D.B. and Willett, P., J. Comput.-Aided Mol. Design, manuscript submitted.