Evaluation of a Novel Electronic Eigenvalue (EEVA) Molecular Descriptor for QSAR/QSPR Studies: Validation Using a Benchmark Steroid Data Set

Kari Tuppurainen,* Marja Viisas, Reino Laatikainen, and Mikael Peräkylä

Department of Chemistry, University of Kuopio, P.O. Box 1627, FIN-70211 Kuopio, Finland

Received October 8, 2001

A novel electronic eigenvalue (EEVA) descriptor of molecular structure for use in the derivation of predictive QSAR/QSPR models is described. Like other spectroscopic QSAR/QSPR descriptors, EEVA is also invariant as to the alignment of the structures concerned. Its performance was tested with respect to the CBG (corticosteroid binding globulin) affinity of 31 benchmark steroids. It appeared that the electronic structure of the steroids, i.e., the "spectra" derived from molecular orbital energies, is directly related to the CBG binding affinities. The predictive ability of EEVA is compared to other QSAR approaches, and its performance is discussed in the context of the Hammett equation. The good performance of EEVA is an indication of the essential quantum mechanical nature of QSAR. The EEVA method is a supplement to conventional 3D QSAR methods, which employ fields or surface properties derived from Coulombic and van der Waals interactions.

INTRODUCTION

Recently, we have put forward a new QSAR/QSPR method, EEVA (electronic eigenvalue), in which calculated molecular orbital energies are employed to derive high-dimensional descriptor vectors for use in PLS (partial least-squares) models.[1,2] In these preliminary works, we have shown that EEVA can be used to generate quite good models for a number of biological data sets. In fact, the overall performance of the EEVA method seems to be comparable to that of the popular CoMFA[3] method despite the fact that EEVA does not require structural alignment of molecules. In broad outline, EEVA belongs to the fast-growing family of spectroscopic QSAR descriptors. The other members are EVA[4] and CoSA,[5] in which vibrational frequencies, NMR chemical shifts, and other spectroscopic quantities are used as eigenvalues. The aim of the present work was to test EEVA's performance in detail employing CBG (corticosteroid binding globulin) affinity of a steroid data set. The predictive ability of EEVA was compared to several other well-established 3D QSAR methods, which have also been tested with the same benchmark steroid data set.

In general, biochemical activity of steroids varies considerably with seemingly small structural changes. Thus, this important molecular family presents a very challenging problem for any prediction method, and in the early days of QSAR the success was limited. However, after the advent of CoMFA[3] the situation has been changed completely, and it is now well established that many (pseudo-) 3D QSAR methods, employing a very diverse set of descriptors, algorithms, and multivariate methods, can provide good predictive models for the CBG activity of steroids.[6]

COMPUTATIONAL METHODS

EEVA Descriptor. The EEVA method involves the following key steps: (i) the determination of the orbital energies of a molecule by some feasible computational method employing semiempirical or ab initio MO theory, (ii) the transformation of MO energies to a bounded scale (e.g. −45−10 eV is appropriate for AM1 energies of "common" organic molecules), (iii) the placement of a Gaussian kernel of fixed standard deviation σ (e.g. 0.1 eV) over each energy value, and (iv) the summation of the overlaid kernels at intervals of L eV (usually L is set at $\sigma/2$). The value of the EEVA descriptor at a sampling point (x) is the sum of the N overlapping Gaussian curves, i.e.

$$\mathrm{EEVA}(x) = \sum_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x - E_i)^2/2\sigma^2) \qquad (1)$$

where E_i is the ith MO energy of the molecule in question. The procedure provides high-dimensional QSAR/QSPR descriptors that are feasible for PLS analyses.

Three-dimensional structures of 31 steroids (Table 1, Figure 1) were modeled with the HYPERCHEM program package (Hypercube, Inc.). Since the steroid skeleton is relatively rigid, conformational search was only performed for the compounds with a bulky C17 side chain (**1**, **6**−**8**, **10**-**11**, **17**−**20**, **22**−**25**, **27**−**31**). Dihedral angle C13−C17−C═O was rotated in steps of 30° with subsequent minimizations employing the MM+ force field as implemented in the HYPERCHEM. The structure of the lowest energy conformer of each steroid was then minimized with the AMPAC program package (QCPE No. 506, version 2.11) by employing the AM1[7] Hamiltonian. All the geometric variables were fully optimized for each compound keeping all settings and options of AMPAC at their default values, except the keyword PRECISE was used. The FORTRAN code of AMPAC was slightly modified to output MO energies to separate disk files for further calculations. The conformational sensitivity of the EEVA descriptor will be discussed briefly below.

* Corresponding author phone: +358-17-162225; fax: +358-17-163259; e-mail: Kari.Tuppurainen@uku.fi.

10.1021/ci0103830 CCC: $22.00
Published on Web 04/17/2002

Table 1. Benchmark Data Set of Steroids with Observed CBG Affinities

no.	compound	activity
1	aldosterone	6.279
2	androstanediol	5.000
3	androstenediol	5.000
4	androstenedione	5.763
5	androsterone	5.613
6	corticosterone	7.881
7	cortisol	7.881
8	cortisone	6.892
9	dehydroepiandrosterone	5.000
10	deoxycorticosterone	7.653
11	deoxycortisol	7.881
12	dihydrotestosterone	5.919
13	estradiol	5.000
14	estriol	5.000
15	estrone	5.000
16	etiocholanolone	5.255
17	pregnenolone	5.255
18	17-hydroxypregnenolone	5.000
19	progesterone	7.380
20	17-hydroxyprogesterone	7.740
21	testosterone	6.724
22	prednisolone	7.512
23	cortisol 21-acetate	7.553
24	4-pregnene-3,11,20-trione	6.779
25	epicorticosterone	7.200
26	19-nortestosterone	6.144
27	16α,17-dihydroxy-4-pregnene-3,20-dione	6.247
28	16-methyl-4-pregnene-3,20-dione	7.120
29	19-norprogesterone	6.817
30	11*β*,17,21-trihydroxy-2α-methyl-4-pregnene-3,20-dione	7.688
31	11*β*,17,21-trihydroxy-2α-methyl-9α-fluoro-4-pregnene-3,20-dione	5.797

Data Analysis. All PLS analyses were done using MATLAB (The MathWorks, 24 Prime Park Way, Natick, MA 01760-1500) scripts written by the authors. The scripts employ an efficient modification of PLS algorithm, SVD-PLS[8] (singular value decomposition PLS), which facilitates very rapid cross-validation runs.

The following nomenclature will be used in PLS analyses throughout the paper. For the model building phase CV = cross-validation, LOO = leave-one-out CV, *PRESS* = predictive residual sum of squares, S_{press} = cross-validated standard error of prediction (eq 2), and q^2 = cross-validated correlation coefficient (eq 3)

$$S_{press} = \sqrt{PRESS/(n - c - 1)} \tag{2}$$

$$q^2 = 1 - \frac{\sum(y_{obs} - y_{pred})^2}{\sum(y_{obs} - y_{mean})^2} = 1 - \frac{PRESS}{\sum(y_{obs} - y_{mean})^2} \tag{3}$$

where n is the number of compounds and c is the number of principal components extracted. The S_{press} value is weighted so that it penalizes models with a large number of principal components. For fitted models r^2 = squared correlation coefficient, *SE* = standard error, and *F* = Fischer test for significance. For external test sets, conventional squared correlation coefficients (r^2_{ex}), mean absolute deviations ($|\Delta|_{av}$), predictive r^2-scores ($pr\text{-}r^2$, eq 4), and *SDEP* values (eq 5) were calculated

$$pr - r^2 = \frac{SD - PRESS}{SD} \tag{4}$$

$$SDEP = \sqrt{PRESS/n} \tag{5}$$

where *SD* is the sum of the squared deviations between the activities of molecules in the test set and the mean affinity of the training set molecules.

CBG Activity of Steroids. The data set of 31 steroids with corticosteroid binding globulin (CBG) affinity (Table 1) is taken from the publications by Dunn et al.[9] and Mickelson et al.[10] This set has repeatedly served as a benchmark in evaluating the performance of new QSAR methods. However, as has been emphasized by many authors,[6,11,12] most early publications employing this set include errors in the structures of the steroids, and it is very likely that some new publications are still not error-free. The structures used in this work (Figure 1) have been carefully checked in order to avoid any further propagation of errors.

Qualitatively, molecules with light substituents such as oxygen and hydroxyl at position 17 of steroid skeleton lead to low CBG activity, whereas the presence of the bulky chain such as $COCH_2OH$ enhances the activity. In contrast, the degree of aromaticity of the A ring is not important for biological activity.

RESULTS

Model Development. In developing PLS model, one should keep an eye on both internal and external predictability; in this work we have put special emphasis on the latter. The model development and validation have been performed in four phases: (i) the optimization of the Gaussian width σ in eq 1, (ii) the validation of the EEVA descriptor employing the optimum value of σ together with a large number of randomized training and test sets, (iii) the verification of the reliability of the EEVA descriptor by scrambling tests, and (iv) the assessment of the conformational sensitivity of the EEVA descriptor.

Previously, it has been found that EVA PLS models[13] may be sensitive to the value of σ, and the optimum value is a data set dependent feature. Accordingly, one must always investigate this sensitivity. In this work, three different ranges have been examined: 0.05−0.15 eV was scanned in steps of 0.01 eV, 0.15−0.50 in steps of 0.05, and 0.50−1.0 in steps of 0.1, resulting in a total of 23 σ values. A fairly large number (100) of PLS models were derived for each σ value by choosing randomly 21 steroids for the training set and placing the remaining 10 in the test set, resulting in 2300 different models. For each model, the internal predictability was assessed with LOO CV, and the optimum number of principal components was selected on the basis of the first S_{press} minimum, with the constraint[14] that the maximum number of principal components does not exceed $n/4$. To assess the external predictability, r^2_{ex}, $|\Delta|_{av}$, $pr\text{-}r^2$, and *SDEP* indices were computed for each model. The best performing σ value was selected by examining the average statistics of the PLS performance indicators S_{press}, q^2, *SDEP*, and $pr\text{-}r^2$ as a function of σ (Figure 2). Obviously, the range of 0.09−0.13 eV is appropriate in this case, i.e., the optimum value of σ depends slightly on the indicator variable used. In our previous study[2] with polychlorinated dibenzo-*p*-dioxins, dibenzofurans, and biphenyls the optimum value of σ was much larger, about 0.5 eV. This difference is probably due to large substituent effect of chlorine, i.e., MO energies of molecules vary considerably with halogen substituents (for more details, see Discussion) and thus a broader Gaussian width σ is appropriate for organohalogen compounds.

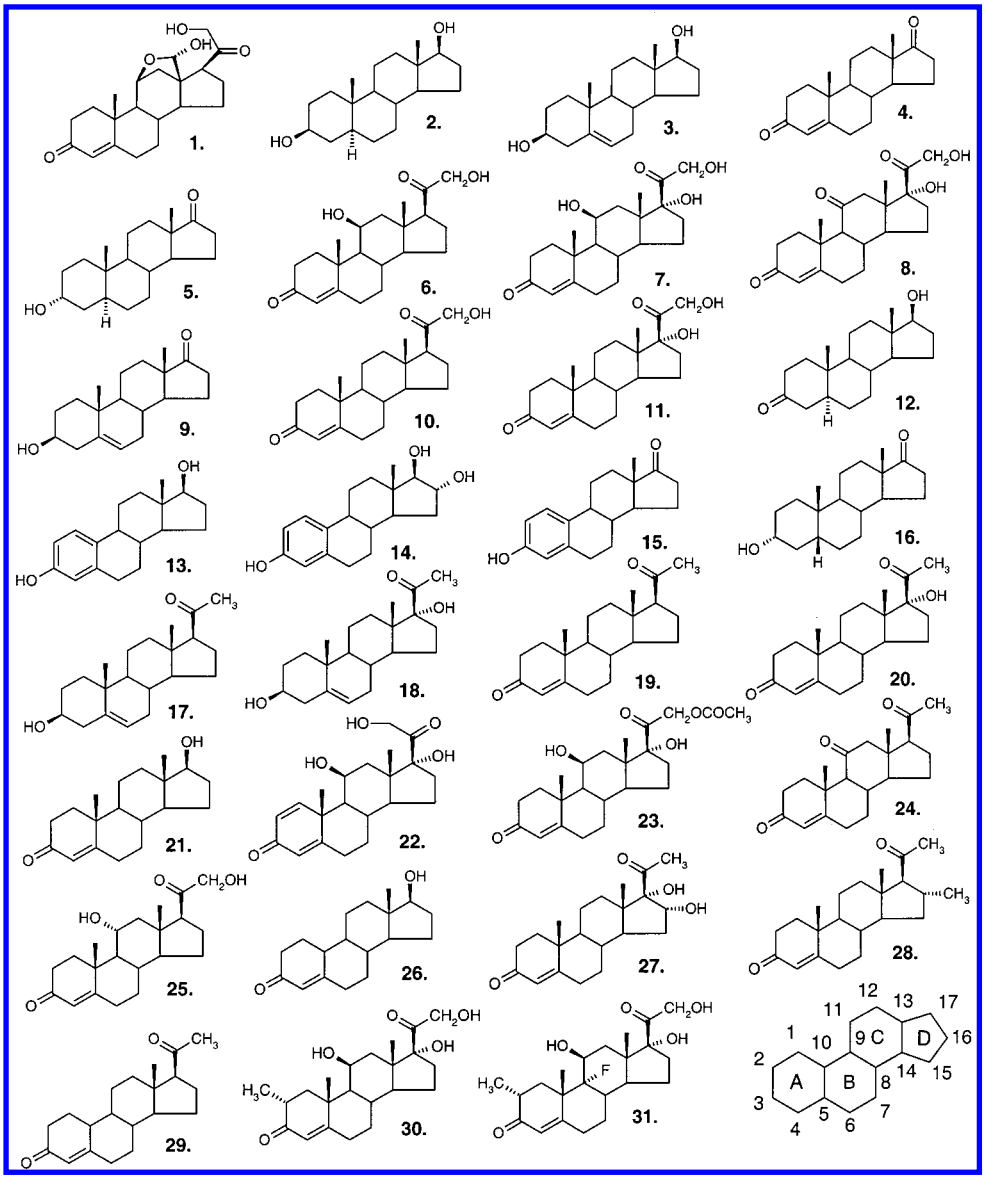

Figure 1. Structures of the 31 steroids of the benchmark data set. The numbering of the atoms of the steroid backbone is given on the lower right-hand corner.

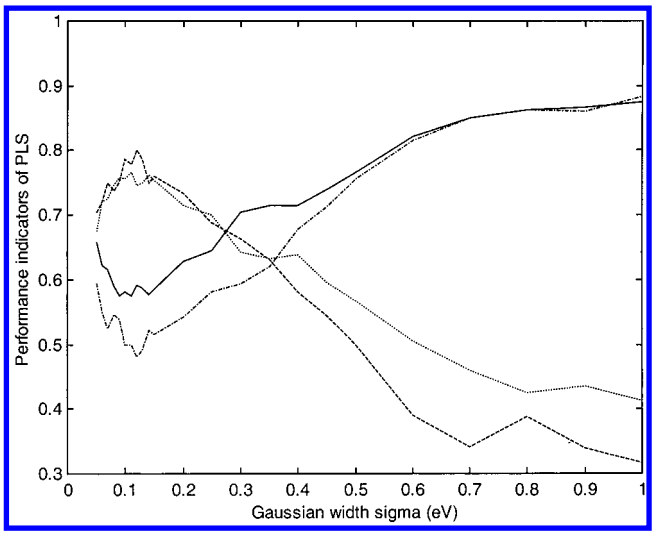

Figure 2. The variation of internal and external PLS performance indicators as a function of σ (the values are averages over 100 randomized PLS models): S_{press} (solid line), q^2 (dotted line), *SDEP* (dashdot line), and $pr\text{-}r^2$ (dashed line).

Second, a large number of validation runs (500) were done (data randomized as above) with the optimum value of σ (0.09 eV). This value was chosen in order to get the best possible "resolution" in the EEVA descriptor. It should be emphasized, however, that this choice is not critical: test calculations with the σ values of 0.10, 0.11, 0.12, and 0.13 eV gave almost identical results (data not shown). The validation runs were done in order to become completely convinced that change correlations (i.e., fortuitous correlations without any predictive ability—typical for all regression models which employ a large number of independent variables) are absent and to explore how data set-dependent the PLS results are. The results indicate that both internal (Figure 3a,b) and external (Figure 3c,d) predictability of the EEVA models are good, irrespective of how the compounds **1**−**31** were divided into the training and test sets. However, the scattering of PLS statistics is considerable (Figure 3a−d), a reminder of how important well-balanced data sets are in QSAR studies (see below).

Third, the reliability of EEVA PLS models was verified by scrambling 500 times the *y* variable. This means that the

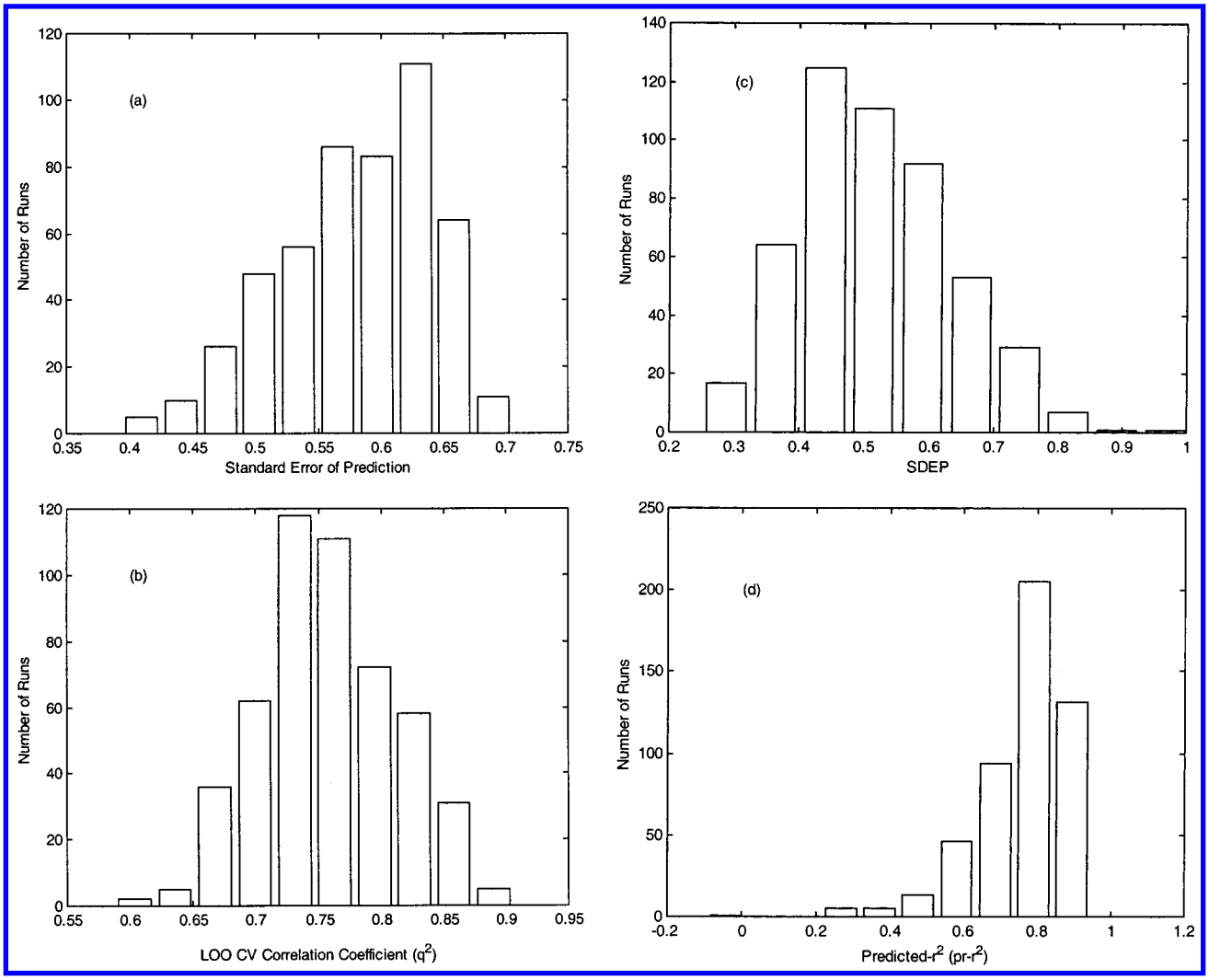

Figure 3. The histograms of internal and external PLS performance indicators of 500 randomized EEVA PLS models with the optimum value of σ (0.09 eV): S_{press} (a), q^2 (b), *SDEP* (c), and *pr-r*^{2} (d).

activities of the training set compounds were mixed so that each *y* value is no longer assigned to the right EEVA descriptor and repeating the LOO CV run. It appeared that no random combination yielded PLS statistics even close to the correct one: usually $q^2 < 0$ and only in very few cases q^2 values were above zero (mean = −0.27).

Finally, the conformational sensitivity of EEVA was investigated by predicting the activities of several conformations of one steroid. The most suitable compound for this test is cortisol 21-acetate (compound **23**), having the longest C17 side chain and thus the largest conformational flexibility. A standard LOO CV PLS model ($q^2 = 0.806$, $S_{press} = 0.511$, $c = 4$) was derived by using 30 steroids for the training set and predicting the activity of **23** for 18 different conformations (C13−C17−C═O dihedral angle was rotated with 20° increments). The results indicate that there is some conformational sensitivity present in the predicted values (max = 8.160, min = 7.809, mean = 7.958, std = 0.104; cf. prediction with the minimum energy conformation = 7.885 and experimental value = 7.553). Interestingly, EEVA seems to give reasonable predictions even if energetically disfavored conformations are used to compute the descriptor vector.

Standard Benchmark Tests for Steroids. In line with most previous publications, the binding affinities for CBG of steroids **1**−**21** were used to derive the standard model, and steroids **22**−**31** formed the corresponding test set (Model 1). The predicted affinities and PLS statistics of Model 1 are given in Table 2. A detailed comparison of EEVA with some established QSAR methods[12,15−24] is presented in Table 3. Overall, EEVA compares favorably with most previous 3D QSAR methods, which employ field and surface proper-

Table 2. Observed and Predicted CBG Affinities for the EEVA Model 1 (Training Set **1**−**21**, Test Set **22**−**31**) and the Corresponding PLS Statistics[a]

	CBG affinity		
molecule	observed	predicted	deviation
22	7.512	7.059	0.453
23	7.553	7.978	0.425
24	6.779	6.839	0.060
25	7.200	7.509	0.309
26	6.144	6.107	−0.037
27	6.247	7.126	0.879
28	7.120	7.301	0.181
29	6.817	6.473	−0.344
30	7.688	7.749	0.061
31	5.797	7.182	1.385

[a] PLS statistics for the LOO CV: $S_{press} = 0.52$ $q^2 = 0.84$, $c = 3$; for the fitted model: $r^2 = 0.97$, $SE = 0.24$, $F = 162$; for the external prediction: $r^2_{ex} = 0.36$ (0.58), $|\Delta|_{av} = 0.41$ (0.31), $SDEP = 0.58$ (0.40), $pr\text{-}r^2 = 0.64$ (0.85). The values in parentheses represent predictions without **31**.

Table 3. Comparison of EEVA with Other 3D QSAR Models for Test Set **22**−**31**[a,d]

method[ref]	r^2_{ex}	$\|\Delta\|_{av}$	SDEP	$pr\text{-}r^2$
COMPASS[15]	0.16 (0.69)	0.46 (0.29)	0.70 (0.34)	0.46 (0.89)
MS-WHIM[16]	0.28 (0.63)	0.44 (0.30)	0.66 (0.41)	0.52 (0.83)
PARM[17]	0.33 (0.30)	0.52 (0.56)	0.71 (0.74)	0.45 (0.45)
TQSAR[18]	0.16 (0.36)	0.59 (0.46)	0.76 (0.56)	0.37 (0.69)
SOMFA[19]	0.20 (0.62)	0.43 (0.32)	0.58 (0.36)	0.63 (0.87)
EVA[20]	0.36 (0.34)	0.42 (0.39)	0.53 (0.51)	0.69 (0.74)
CoMFA[20]	0.25 (0.75)	0.46 (0.30)	0.71 (0.40)	0.45 (0.84)
GRIND[21,b]	− (0.88)	− (0.23)	− (0.26)	− (0.93)
MFTA[22]	0.87 (0.82)	0.21 (0.23)	0.30 (0.31)	0.90 (0.90)
COMSA[12]	0.09 (0.41)	0.52 (0.38)	0.70 (0.44)	0.47 (0.81)
MEDV[23]	0.45 (0.57)	0.54 (0.48)	0.65 (0.59)	0.54 (0.66)
QS-SM[24]	0.36 (0.22)	0.47 (0.42)	0.54 (0.49)	0.68 (0.76)
EEVA[c]	0.36 (0.58)	0.41 (0.30)	0.58 (0.40)	0.64 (0.85)

[a] The values in parentheses are predictions without **31**. [b] The predicted value of **31** was not reported (stated as an outlier). [c] This work, Model 1. [d] Abbreviations: COMPASS, comparative surface similarity; MS-WHIM, molecular surface − weighted holistic invariants of molecules; PARM, pseudo atomic receptor model; TQSAR, tuned QSAR; SOMFA, self-organizing molecular field analysis; EVA, eigenvalue, i.e., vibrational normal mode in this context; CoMFA, comparative molecular field analysis; GRIND, grid independent descriptors; MFTA, molecular field topology analysis; COMSA, comparative surface analysis; MEDV, molecular electronegativity distance vector; QS-SM, quantum self-similarity measures.

ties derived from steric and Coulombic interactions. The results indicate that EEVA provides a promising alternative and supplement to other QSAR descriptors, which often employ highly sophisticated physical, mathematical, and statistical tools such as (quantum) similarity matrices, neural networks and genetic algorithms. Interestingly, the success of the MFTA method,[22] which is essentially a 2D QSAR approach, indicates that 3D information is not necessary for a predictive QSAR model with the steroid data set. The predictions of QS-SM[24] (quantum self-similarity measures) and EEVA are almost identical; this is understandable as both methods rely on quantum mechanical descriptors.

There is still much room for improvement. In particular, the notorious outlier in most previous QSAR studies, compound **31**, is poorly predicted also with EEVA (Table 2). Furthermore, the difference between the observed and predicted activity is quite large for compound **27**. However, the exceptional structural features of these compounds give a possible explanation: compound **31** is the only steroid having a fluoro functionality, and compound **27** has two hydroxyl groups attached to adjacent carbons, a feature that is absent in the training set. In fact, as emphasized by Kubinyi,[25] the molecules repeatedly used in the training set (compounds **1**−**21**) do not cover all structural features within the series. Instead, the recommended training set comprises molecules **1**−**12**, **23**−**31**, and test set molecules **13**−**22**. This model will be referred to as EEVA Model 2 (Table 4). It is evident that the performance of EEVA is better with this new training set composition, in particular when the external predictability is considered. Recently, it has been shown by employing principal components analysis that molecules not representative of training set molecules should not be included in any test set used for training set validation, at least for this data set.[26] Consequently, it seems that the proposition of Kubinyi should be accepted for future work with the benchmark steroid data set. Furthermore, the performance of the previous QSAR models should be reevaluated. Taken overall, the results of this study suggest that EEVA may be used for predictive purposes, facilitating a rapid screening of steroids for CBG activity.

Table 4. Observed and Predicted CBG Affinities for the EEVA Model 2 (Training Set **1**−**12**, **23**−**31**, Test Set **13**−**22**) and the Corresponding PLS Statistics[a]

	CBG affinity		
molecule	observed	predicted	deviation
13	5.000	4.644	−0.356
14	5.000	5.090	0.090
15	5.000	4.725	−0.275
16	5.255	5.881	0.626
17	5.255	5.851	0.596
18	5.000	5.963	0.963
19	7.380	7.268	−0.112
20	7.740	7.215	−0.525
21	6.724	6.268	−0.457
22	7.512	7.092	−0.420

[a] PLS statistics for the LOO CV: $S_{press} = 0.58$, $q^2 = 0.73$, $c = 4$; for the fitted model: $r^2 = 0.99$, $SE = 0.12$, $F = 353$.; for the external prediction: $r^2_{ex} = 0.81$, $|\Delta|_{av} = 0.44$, $SDEP = 0.51$, $pr\text{-}r^2 = 0.84$.

DISCUSSION

The results of the present work are probably sufficient to demonstrate the applicability of EEVA to QSAR problems. But what is it about an EEVA descriptor, derived computationally from MO energies that leads to such good correlations with biological activity? The good performance of EVA[4] and CoSA[5,27,28] is in fact more easily understandable, since the 3D structure of a molecule and its IR or NMR spectra are closely related. IR spectra reflect vibrational motions and the spatial arrangements of molecular functional groups, and the relationship between the ^{13}C NMR chemical shifts of a molecule and its 3D conformations is well-known and routinely employed.

It should be emphasized that MO energies, together with squared MO coefficients, are the most fundamental physical quantities of molecules. They can be directly calculated from the Schrödinger equation ($H\Psi_n = E_n\Psi_n$), and all other physical parameters of molecules can be derived from them, at least in principle. The Hamiltonian operator (matrix) H in the Schrödinger equation involves all electronic interactions of a molecule, and its diagonalization gives the eigenvalues E_n (MO energies) and eigenvectors Ψ_n (state vectors or LCAO MO coefficients). It is well-known that changes in MO coefficients and MO energies are correlated

if electron−electron repulsion is neglected.[29,30] Thus their information content is equivalent in many respects, but there is an important exception: the regioselectivity of chemical reactivity can only be characterized by means of examining the MO coefficients. However, if the knowledge of regioselectivity is not important, it is merely a matter of taste which of two is considered as a potential QSAR descriptor. The MO energy spectrum facilitates a development of computationally simple QSAR models, and it was employed in this work to derive the EEVA descriptor. In the spectroscopy vernacular, the EEVA descriptor is very distantly related to the ESCA (Electronic Spectroscopy for Chemical Analysis) spectrum.

All QSAR methods, regardless of the physical, mathematical, or statistical complexity, are basically related to the substituent effect, i.e., the Hammett equation—an empirical relationship whose almost universal applicability is not completely understood yet. When different substituents are attached to a molecular backbone, MO energies will vary with the chemical properties of the substituents (electron donor vs acceptor, changes in polarizability, possible through-resonance etc.), resulting in changes in both the positions of individual MO energy "peaks" and in the shapes of "peak" patterns. The results of this study indicate that the changes can be utilized in a quantitative and predictive manner employing the EEVA descriptor. Thus it seems reasonable to propose that the success of EEVA is a quantum-mechanical manifestation of the Hammett equation. In fact, we have previously shown that the Hammett σ constants can be predicted with the EEVA method with considerable success.[1] In this particular case, the results suggest that even formally inert σ-electron molecular backbone of steroids can play an important role in the transmission of the electronic effects of the substituents, in accordance with the "ribbon-orbital effect" described by Novak and Kovač.[31]

It should be emphasized that the electronic descriptor that gives a good fit with the data may not be the best for understanding the mechanism, even if the predictive ability is good. In fact, it can be anticipated that in the case of EEVA the inverse problem of QSAR will be difficult to solve. At this moment, the practical value of EEVA relies mainly on its predictive ability, which should be carefully tested by comprehensive cross-validation runs following, for example, the guidelines given above.

In the future, the performance of EEVA may be improved in many ways. For example, it is possible to use ab initio MO energies in order to get a better "resolution" in the MO energy spectrum, leading probably to better predictions. Second, the presence of the solvent reaction field can significantly alter the electronic structure of a molecule, including its MOs, especially in aromatic systems. Solvation effects can be modeled by using e.g. SCRF (self-consistent reaction field) methods, and it is of interest to see whether this approach would provide better QSAR models.

In general, this comparative study corroborates that a large number of 3D QSAR descriptors can be employed to predict the CBG activity of steroids (Table 3). However, it is difficult to determine which of them is most closely related to the physicochemical mechanisms behind the activity. This is probably due to the high collinearity between seemingly different 3D QSAR descriptors. The substituent effects obviously cause simultaneous and correlated changes in several steric and electronic indices, any of which may give a statistically significant and predictive correlation with biological activity. This implies that the combination of different 3D QSAR descriptors would hardly lead to an improved performance. In this respect 3D QSAR descriptors contrast sharply to Hansch-type descriptors that are often orthogonal enough to allow the use of multiple linear regression. The origin of the collinearity is of considerable theoretical interest, as it may ultimately lead to a unification of different 3D QSAR methods.

CONCLUSION

The results of this study indicate that the EEVA method provides robust predictive models for the CBG activity of steroids. In general, EEVA provides a promising alternative and supplement for other QSAR methods. It is computationally simple, easy to use, quite insensitive to molecular conformations, and it gives useful results at a semiempirical level of theory. From a theoretical point of view, EEVA's good performance is of considerable interest, but its sound physical interpretation needs further elucidation in the future. In any case, the evident success of EEVA is yet another manifestation of the quantum mechanical origin of QSAR.[32]

ACKNOWLEDGMENT

This work was supported by the Academy of Finland.

REFERENCES AND NOTES

(1) Tuppurainen, K. EEVA (Electronic Eigenvalue): A New QSAR/QSPR Descriptor for Electronic Substituent Effects Based on Molecular Orbital Energies. *SAR QSAR Environ. Res.* **1999**, *10*, 39−46.

(2) Tuppurainen, K.; Ruuskanen, J. Electronic Eigenvalue (EEVA): A New QSAR/QSPR Descriptor for Electronic Substituent Effects Based on Molecular Orbital Energies. A QSAR Approach to the *Ah* Receptor Binding Affinity of Polychlorinated Biphenyls (PCBs), Dibenzo-*p*-dioxins and Dibenzofurans. *Chemosphere* **2000**, *41*, 843−848.

(3) Cramer, R. D., III.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc*. **1988**, *110*, 5959−5967.

(4) Turner, D. B.; Willett, P. The EVA Spectral Descriptor. *Eur. J. Med. Chem*. **2000**, *35*, 367−375.

(5) Bursi, R.; Dao, T.; van Wijk, T.; de Gooyer, M.; Kellenbach, E.; Verwer, P. Comparative Spectra Analysis (CoSA): Spectra as Three-Dimensional Molecular Descriptors for the Prediction of Biological Activities. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 861−867.

(6) Coats, E. The CoMFA Steroids as a Benchmark Dataset for Development of 3D QSAR Methods. *Perspect. Drug Discov. Design*. **1998**, *12/13/14*, 199−213.

(7) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum-Mechanical Molecular Model. *J. Am. Chem. Soc*. **1985**, *107*, 3902−3909.

(8) Wang, T. W.; Khettry, A.; Berry, M.; Batra, J. SVDPLS: An Efficient Algorithm for Performing PLS. *The First International Chemometrics InterNet Conference, InCINC'94.* (The MATLAB code of SVDPLS can be found from the WWW site http://www.emsl.pnl.gov: 2080/docs/incinc/papers/wang_pls/sectionstar3_9.html).

(9) Dunn, J. F.; Nisula, B. C.; Rodbard, D. Transport of Steroid Hormones: Binding of 21 Endogeneous Steroids to Both Testosterone-Binding Globulin and Corticosteroid-Binding Globulin in Human Plasma. *J. Clin. Endocrin. Metab*. **1981**, *53*, 58−68.

(10) Mickelson, K. E.; Forsthoefel, J.; Westphal, U. Steroid-Protein Interactions. Human Corticosteroid Binding Globulin: Some Physicochemical Properties and Binding Specificity. *Biochemistry* **1981**, *20*, 6211−6218.

(11) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of Molecular Surface Properties for Modeling *Corticosteroid Binding Globulin* and Cytosolic *Ah* Receptor Activity by Neural Networks. *J. Am. Chem. Soc*. **1995**, *117*, 7769−7775.

(12) Polanski, J.; Walczak, B. The Comparative Molecular Surface Analysis (COMSA): a novel Tool for Molecular Design. *Comput. Chem*. **2000**, *24*, 615−625.
(13) Turner, D. B.; Willett, P.; Ferguson, A. M.; Heritage, T. Evaluation of a Novel Infrared Range Vibration-Based Descriptor (EVA) for QSAR Studies. 1. General Application. *J. Comput.-Aided Mol. Design* **1997**, *11*, 409−422.
(14) Wold, S.; Johansson, E.; Cocchi, M. PLS-partial least squares projections to latent structures. In *3D QSAR in Drug Design: Theory Method and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 523−550.
(15) Jain, A. N.; Koile, K.; Chapman, D. Compass: Predicting Biological Activities from Molecular Surface Properties. Performance Comparisons on a Steroid Benchmark. *J. Med. Chem*. **1994**, *37*, 2315−2327.
(16) Bravi, G.; Gancia, E.; Mascagni, P.; Pegna, M.; Todeschini, R.; Zaliani, A. MS−WHIM, New 3D Theoretical Descriptors Derived from Molecular Surface Properties: A Comparative 3D QSAR Study in a Series of Steroids. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 79−92.
(17) Chen, H.; Zhou, J.; Xie, G. PARM: A Genetic Algorithm to Predict Bioactivity. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 243−250.
(18) Robert, D.; Amat, L.; Carbo-Dorca, R. Three-Dimensional Quantitative Structure−Activity Relationships from Tuned Molecular Quantum Similarity Measures: Prediction of the Corticosteroid-Binding Globulin Binding Affinity for a Steroid Family. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 333−344.
(19) Robinson, D. D.; Winn, P. J.; Lyne, P. D.; Richards, W. G. Self-Organizing Molecular Field Analysis: A Tool for Structure−Activity Studies. *J. Med. Chem*. **1999**, *42*, 573−583.
(20) Turner, D. B.; Willett, P.; Ferguson, A. M.; Heritage, T. W. Evaluation of a Novel Molecular Vibration-Based Descriptor (EVA) for QSAR Studies: 2. Model Validation Using a Benchmark Steroid Dataset. *J. Comput.-Aided Mol. Design* **1999**, *13*, 271−296.
(21) Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S. Molecular Field Topology Analysis in QSAR Studies of Organic Compounds. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 659−667.
(22) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem*. **2000**, *43*, 3233−3243.
(23) Liu, S.-S.; Yin, C.-S.; Li, Z.-L.; Cai, S.-X. QSAR Study of Steroid Benchmark and Dipeptides Based on MEDV-13. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 321−329.
(24) Amat, L.; Besalu, E.; Carbo-Dorca, R.; Ponec, R. Identification of Active Molecular Sites Using Quantum-Self-Similarity Matrices. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 978−991.
(25) Kubinyi, H. A general view on similarity and QSAR studies. In *Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry;* van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; VHCA: Basel, Switzerland, and Wiley-VCH: Weinheim, Germany, 1997; pp 7−28.
(26) Silverman, D. B. The Thirty-one Benchmark Steroids Revisited: Comparative Molecular Moment Analysis (CoMMA) with Principal Component Regression. *Quant. Struct.-Act. Relat*. **2000**, *19*, 237−246.
(27) Beger, R. D.; Freeman, J. P.; Lay, J. O., Jr.; Wilkes, J. G.; Miller, D. W. Use of ^{13}C NMR Spectrometric Data to Produce a Predictive Model of Estrogen Receptor Binding Affinity. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 219−214.
(28) Beger, R. D.; Wilkes, J. G. Models of Polychlorinated Dibenzodioxins, Dibenzofurans, and Biphenyls Binding Affinity to the Aryl Hydrocarbon Receptor Developed Using ^{13}C NMR Data. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1322−1329.
(29) Fukui, K.; Yonezawa, T.; Nagata, C. Interrelations of Quantum-Mechanical Quantities Concerning Chemical Reactions of Conjugated Molecules. *J. Chem. Phys*. **1957**, *27*, 831−841.
(30) Dewar, M. J. S.; Dougherty, R. C. *The PMO Theory of Organic Chemistry*; Plenum Press: New York, 1975.
(31) Novak, I.; Kovač, B. Electronic Structure and Biological Activity of Steroids. *Biophys. Chem*. **1999**, *78*, 233−240.
(32) Carbo-Dorca, R.; Amat, L.; Besalu, E.; Girones, X.; Robert, D. Quantum Mechanical Origin of QSAR: Theory and Applications. *J. Mol. Struct. (THEOCHEM)* **2000**, *79*, 181−228.

CI0103830