Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 14 (2006) 273–279
Bioorganic & Medicinal Chemistry

A 4D-QSAR study on anti-HIV HEPT analogues

Andrzej Bak and Jaroslaw Polanski*

Department of Organic Chemistry, Institute of Chemistry, University of Silesia, PL-40-006 Katowice, Poland

Received 24 May 2005; revised 31 July 2005; accepted 5 August 2005
Available online 26 September 2005

Abstract—We used the 4D-QSAR method coupled with the PLS analysis and uninformative variable elimination or its variants for the investigations of the antiviral activity of HEPT, a series of conformationally flexible molecules that bind HIV-1 reverse transcriptase. An analysis of several Hopfinger's and SOM-4D-QSAR models indicated that both methods yield comparable results. Generally, charge descriptors provide better modeling efficiency. We have shown that the method properly indicates the mode of interaction revealed by X-ray studies. It also allows us to calculate highly predictive QSAR models.

1. Introduction

Although new anti-HIV agents get better and better blocking the virus, novel inhibitors are still sought after.[1,2] Reverse transcriptase (RT) which controls the development of proviral DNA is one among possible targets for such investigations. A number of RT inhibitors[3] including various non-nucleoside RT inhibitors (NNRTIs)[4,5] have been discovered. Generally, these compounds are less toxic and more stable than nucleoside RT inhibitors. Lower metabolism and clearance rates are further advantages of NNRTIs. 1-[2-Hydroxyethoxy)methyl]-6-(phenylthio)-thymines (HEPT) form the NNRTI series that does not target an active site of polymerase but rather the enzyme allosteric site. The interactions of these compounds with RT have been thoroughly investigated and the crystal structures of several ligand–enzyme complexes have been determined.[6,7] In this context, the experimental studies have been published for two HEPT analogues, namely MKC-442 and TNK-651, complexed with HIV-1 RT.[8] This study revealed an active HEPT conformation that binds the p66 enzyme unit and pointed to the importance of so-called *switching effect*, which forces an enzyme Tyr181 to adopt a conformation that enhances the drug–receptor interactions.[9] Moreover, it was observed that NNRTI inhibitors bind in a common way. This fills the allosteric pocket in a so-called *butterfly mode*.[10] One of the *butterfly wings* binds with π electron-rich moiety with a hydrophobic pocket that includes mainly the side chains of the enzymes Tyr181, Tyr188, Phe227, Trp229, and Tyr318, while the other wing having the hydrogen donor/acceptor interacts with Lys101 and Lys108. A hydrophobic moiety of the butterfly body interacts with an enzyme receptor site formed by the side chains of Lys103, Val106, and Val 179.[11]

A number of QSAR studies have also been reported for these compounds.[12–14] This includes the traditional 2D as well as 3D and holographic (HQSAR) methods.[15] More recently, a computational docking protocol has been used for the investigations of the NNRTI inhibition of HIV-RT.[11] Generally, QSAR is a computational tool for the investigation of chemical compounds' space in a search for novel properties that generally analyzes the data in a receptor-independent mode. This method should work like a dictionary between molecular structures and properties, which clearly makes it an essential and irreplaceable method in molecular design. However, more and more sophisticated tools are needed for the efficient and robust transformation of the molecular structure space into the compound property space. Conformational flexibility that is generally overlooked by traditional QSAR is an example of such an issue that can significantly improve modeling efficiency.

The aim of the present study was the investigation of the structure–activity relationships of the HEPT series using the 4D-QSAR method. Basically, 4D-QSAR investigates the conformational space of the molecular objects.[16–19] Since HEPTs are conformationally labile compounds, the application of 4D-QSAR should provide

Keywords: 4D QSAR; SOM-4D-QSAR; Anti-HIV; Reverse transcriptase.
* Corresponding author. Tel.: +48 32 3591197; fax: +48 32 2599978; e-mail: Polanski@us.edu.pl

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.08.023

better insight into the activity of these molecules. In 4D-QSAR calculation, we generate for a single molecule the enormous number of conformers that investigates different spatial region. Actually, it is the likelihood of a formation of a common 3D pattern for a series of molecules that is sought after during the molecular dynamics simulations. Moreover, many-fold replication of molecular configuration (represented by a series of conformers) increases the chances for a proper pharmacophore mapping by the series of ligand structures. In this publication, we used traditional Hopfinger's 4D-QSAR and the self-organizing neural network version of this method, namely SOM-4D-QSAR.[20,21]

2. Methods

2.1. Data sets for the analysis

The chemical structures of the HEPT are shown in Table 1. The RT inhibition data are reported according to Ref. 22.

2.2. Molecular alignment

All modeling work was performed using the Sybyl 6.2 software package run on Silicon Graphics O2 workstation.[23] The initial geometry was optimized using the standard Tripos force field (POWELL method) with 0.005 kcal/mol energy gradient convergence criterion and a distant dependent dielectric constant. Charges were calculated using the Gasteiger–Marsilli method implemented in Sybyl. After the optimization, the molecules were used as the initial structures in the molecular dynamics simulations (MDs). Each 3D structure is the starting point in generating the conformational ensemble profile (CEP). The temperature was set to 300 K. The atomic coordinates of each conformation and its total energy were recorded every 0.1 ps. One thousand conformations were sampled for each analogue. Partial atomic charges were calculated using the semiempirical AM1 Hamiltonian (HyperChem package).[24] All molecules were superimposed before the calculation of 4D-QSAR descriptors. The superimposition was performed by covering all non-hydrogen atoms of the pyrimidine ring.

2.3. Hopfinger's and SOM-4D-QSAR analysis

For the calculation of 4D-QSAR descriptors we applied Hopfinger's[16–18] and SOM formalisms that are described in detail in our previous publications.[20–22] Thus, each 3D molecular representation is placed in its own virtual cubic grid. Different types of grid cell occupancy descriptors (GCODs) were considered and calculated for indicating atoms referred to as interaction pharmacophore elements (IPE-all). Apart from the GCODs used by Hopfinger, we applied the absolute charge occupancy (A_q) for the chosen IPE atoms of compound c defined as

$$A_q(c,i,j,k,N) = \sum_{t=0}^{T} O_t(c,i,j,k) \times q/m, \qquad (1)$$

where m means the number of the atoms of compounds, c present in the cell (i,j,k) at time t, q means the sum of partial atoms of charges present in some cell at time t, and T is the length of the time in MDs. N is the number of sampling MD steps. The joint (J_q) and self-charge occupancy (S_q) with the most active reference compound R were defined after the following equations:

$$J_q(c,i,j,k,N) = \sum_{t=0}^{T} O_t(c,i,j,k) \cap O_t(R_q,i,j,k) \times q/m, \qquad (2)$$

$$S_q(c,R,i,j,k,N) = \sum_{t=0}^{T} \left\{ O_t(c,i,j,k) - \left[\sum_{t=0}^{T} O_t(c,i,j,k) \cap O_t(R,i,j,k) \right] \right\} \times q/m. \qquad (3)$$

Grid cells are unfolded into vectors and vectors describing all molecules of the series are aligned into a matrix. Grid cells that are empty for all molecules in the series analyzed are eliminated and the resulting matrix was used for further calculations using the PLS method.[25–27] Alternatively, a neural network version with Kohonen self-organizing neural network (SOM-4D-QSAR) was used for the comparison.

2.4. PLS analysis

The obtained vectors were processed by the PLS analysis with a leave-one-out cross-validation procedure (LOO-CV). The PLS procedures were programmed within the MATLAB environment (MATLAB).[28]

A PLS model was constructed for the centered data and its complexity was estimated on the basis of the leave-one-out cross-validation procedure. In the leave-one-out CV, one repeats the calibration m times, each time treating the ith left-out object as the prediction object. The dependent variable for each left-out object is calculated on the basis of the model with one, two, three, etc, factors. The root-mean-square Error of CV for the model with j factors is defined as:

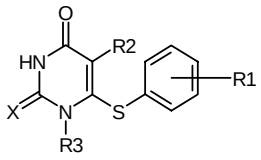
$$\mathrm{RMSECV}_j = \sqrt{\frac{\sum_i (\mathrm{obs}_i - \mathrm{pred}_{i,j})^2}{m}}, \qquad (4)$$

where obs denotes the assayed value; pred is the predicted value of dependent variable and i refers to the object index, which ranges from 1 to m. A model with k factors, for which RMSECV reaches a minimum, is considered as an optimal one.

We used performance metrics that are widely accepted and used in CoMFA analyses, that is, cross-validated q^2_{CV}

$$q^2_{CV} = 1 - \frac{\sum(\mathrm{obs}_i - \mathrm{pred}_i)^2}{\sum(\mathrm{obs}_i - \mathrm{mean}(obs))^2}, \qquad (5)$$

Table 1. Chemical structures with the observed values of the anti-HIV activity for the HEPT derivatives



No.	R1	R2	R3	X	Obs.
1	2-Me	Me	$CH_2OCH_2CH_2OH$	O	4.15
2	2-NO_2	Me	$CH_2OCH_2CH_2OH$	O	3.85
3	2-OMe	Me	$CH_2OCH_2CH_2OH$	O	4.72
4	3-Me	Me	$CH_2OCH_2CH_2OH$	O	5.59
5	3-Et	Me	$CH_2OCH_2CH_2OH$	O	5.57
6	3-*t*-Bu	Me	$CH_2OCH_2CH_2OH$	O	4.92
7	3-CF_3	Me	$CH_2OCH_2CH_2OH$	O	4.35
8	3-F	Me	$CH_2OCH_2CH_2OH$	O	5.48
9	3-Cl	Me	$CH_2OCH_2CH_2OH$	O	4.89
10	3-Br	Me	$CH_2OCH_2CH_2OH$	O	5.24
11	3-I	Me	$CH_2OCH_2CH_2OH$	O	5.00
12	3-NO_2	Me	$CH_2OCH_2CH_2OH$	O	4.47
13	3-OH	Me	$CH_2OCH_2CH_2OH$	O	4.09
14	3-OMe	Me	$CH_2OCH_2CH_2OH$	O	4.66
15	3,5-Me_2	Me	$CH_2OCH_2CH_2OH$	O	6.59
16	3,5-Cl_2	Me	$CH_2OCH_2CH_2OH$	O	5.89
17	3,5-Me_2	Me	$CH_2OCH_2CH_2OH$	S	6.66
18	3-COOMe	Me	$CH_2OCH_2CH_2OH$	O	5.10
19	3-COMe	Me	$CH_2OCH_2CH_2OH$	O	5.14
20	3-CN	Me	$CH_2OCH_2CH_2OH$	O	5.00
21	H	$CH_2CH{=}CH_2$	$CH_2OCH_2CH_2OH$	O	5.60
22	H	Et	$CH_2OCH_2CH_2OH$	S	6.96
23	H	Pr	$CH_2OCH_2CH_2OH$	S	5.00
24	H	*i*-Pr	$CH_2OCH_2CH_2OH$	S	7.23
25	3,5-Me_2	Et	$CH_2OCH_2CH_2OH$	S	8.11
26	3,5-Me_2	*i*-Pr	$CH_2OCH_2CH_2OH$	S	8.30
27	3,5-Cl_2	Et	$CH_2OCH_2CH_2OH$	S	7.37
28	H	Et	$CH_2OCH_2CH_2OH$	O	6.92
29	H	Pr	$CH_2OCH_2CH_2OH$	O	5.47
30	H	*i*-Pr	$CH_2OCH_2CH_2OH$	O	7.20
31	3,5-Me_2	Et	$CH_2OCH_2CH_2OH$	O	7.89
32	3,5-Me_2	*i*-Pr	$CH_2OCH_2CH_2OH$	O	8.57
33	3,5-Cl_2	Et	$CH_2OCH_2CH_2OH$	O	7.85
34	4-Me	Me	$CH_2OCH_2CH_2OH$	O	3.66
35	H	Me	$CH_2OCH_2CH_2OH$	O	5.15
36	H	Me	$CH_2OCH_2CH_2OH$	S	6.01
37	H	I	$CH_2OCH_2CH_2OH$	O	5.44
38	H	$CH{=}CH_2$	$CH_2OCH_2CH_2OH$	O	5.69
39	H	CH=CHPh	$CH_2OCH_2CH_2OH$	O	5.22
40	H	CH_2Ph	$CH_2OCH_2CH_2OH$	O	4.37
41	H	$CH{=}CPh_2$	$CH_2OCH_2CH_2OH$	O	6.07
42	H	Me	$CH_2OCH_2CH_2OMe$	O	5.06
43	H	Me	$CH_2OCH_2CH_2OAc$	O	5.17
44	H	Me	$CH_2OCH_2CH_2OCOPh$	O	5.12
45	H	Me	CH_2OCH_2Me	O	6.48
46	H	Me	$CH_2OCH_2CH_2Cl$	O	5.82
47	H	Me	$CH_2OCH_2CH_2N_3$	O	5.24
48	H	Me	$CH_2OCH_2CH_2F$	O	5.96
49	H	Me	$CH_2OCH_2CH_2Me$	O	5.48
50	H	Me	CH_2OCH_2Ph	O	7.06
51	H	Et	CH_2OCH_2Me	O	7.72
52	H	Et	CH_2OCH_2Me	S	7.58
53	3,5-Me_2	Et	CH_2OCH_2Me	O	8.24
54	3,5-Me_2	Et	CH_2OCH_2Me	S	8.30
55	H	Et	CH_2OCH_2Ph	O	8.23
56	3,5-Me_2	Et	CH_2OCH_2Ph	O	8.55
57	H	Et	CH_2OCH_2Ph	S	8.09
58	3,5-Me_2	Et	CH_2OCH_2Ph	S	8.14
59	H	*i*-Pr	CH_2OCH_2Me	O	7.99

(*continued on next page*)

Table 1. (*continued*)

No.	R1	R2	R3	X	Obs.
60	H	*i*-Pr	CH_2OCH_2Ph	O	8.51
61	H	*i*-Pr	CH_2OCH_2Me	S	7.89
62	H	*i*-Pr	CH_2OCH_2Ph	S	8.14
63	H	Me	CH_2OMe	O	5.68
64	H	Me	CH_2OBu	O	5.33
65	H	Me	Et	O	5.66
66	H	Me	Bu	O	5.92
67	3,5-Cl_2	Et	CH_2OCH_2Me	S	7.89
68	H	Et	CH_2O-*i*-Pr	S	6.66
69	H	Et	CH_2O-c-Hex	S	5.79
70	H	Et	CH_2OCH_z-c-Hex	S	6.45
71	H	Et	$CH_2OCH_2C_6H_4$(4-Me)	S	7.11
72	H	Et	$CH_2OCH_2C_6H_4$(4-Cl)	S	7.92
73	H	Et	$CH_2OCH_2CH_2Ph$	S	7.04
74	3,5-Cl_2	Et	CH_2OCH_2Me	O	8.13
75	H	Et	CH_2O-*i*-Pr	O	6.47
76	H	Et	CH_2O-c-Hex	O	5.40
77	H	Et	CH_2OCH_z-c-Hex	O	6.35
78	H	Et	$CH_2OCH_zCH_2Ph$	O	7.02
79	H	c-Pr	CH_2OCH_2Me	S	7.02
80	H	c-Pr	CH_2OCH_2Me	O	7.00
81	H	Me	$CH_2OCH_2CH_2OC_5H_{11}$-n	O	4.46
82	2-Cl	Me	$CH_2OCH_2CH_2OH$	O	3.89
83	3-CH_2OH	Me	$CH_2OCH_2CH_2OH$	O	3.53
84	4-F	Me	$CH_2OCH_2CH_2OH$	O	3.60
85	4-Cl	Me	$CH_2OCH_2CH_2OH$	O	3.60
86	4-NO_2	Me	$CH_2OCH_2CH_2OH$	O	3.72
87	4-CN	Me	$CH_2OCH_2CH_2OH$	O	3.60
88	4-OH	Me	$CH_2OCH_2CH_2OH$	O	3.56
89	4-OMe	Me	$CH_2OCH_2CH_2OH$	O	3.60
90	4-COMe	Me	$CH_2OCH_2CH_2OH$	O	3.96
91	4-COOH	Me	$CH_2OCH_2CH_2OH$	O	3.45
92	3-$CONH_2$	Me	$CH_2OCH_2CH_2OH$	O	3.51
93	H	COOMe	$CH_2OCH_2CH_2OH$	O	5.18
94	H	CONHPh	$CH_2OCH_2CH_2OH$	O	4.74
95	H	SPh	$CH_2OCH_2CH_2OH$	O	4.68
96	H	CCH	$CH_2OCH_2CH_2OH$	O	4.74
97	H	CCPh	$CH_2OCH_2CH_2OH$	O	5.47
98	3-NH_2	Me	$CH_2OCH_2CH_2OH$	O	3.60
99	H	$COCHMe_2$	$CH_2OCH_2CH_2OH$	O	4.92
100	H	COPh	$CH_2OCH_2CH_2OH$	O	4.89
101	H	CCMe	$CH_2OCH_2CH_2OH$	O	4.72
102	H	F	$CH_2OCH_2CH_2OH$	O	4.00
103	H	Cl	$CH_2OCH_2CH_2OH$	O	4.52
104	H	Br	$CH_2OCH_2CH_2OH$	O	4.70
105	H	Me	$CH_2OCH_2CH_2OCH_2Ph$	O	4.70
106	H	Me	H	O	3.60
107	H	Me	Me	O	3.82

where obs is the assayed values; pred is the predicted values, mean is the mean value of obs, and i refers to the object index, which ranges from 1 to m; and cross-validated standard error s

$$s = \sqrt{\frac{\sum(\mathrm{obs}_i - \mathrm{pred}_i)^2}{m - k - 1}}, \tag{6}$$

where m is the number of objects and k is the number of the PLS factors in the model.

The quality of external predictions was measured by the standard deviation of error of prediction (SDEP) parameter:

$$\mathrm{SDEP} = \sqrt{\frac{\sum_i(\mathrm{pred}_i - \mathrm{obs}_i)^2}{n}}, \tag{7}$$

where pred is the predicted value and obs is the observed value.

2.5. Iterative variable elimination

In our previous publications, we have shown that uninformative variable elimination (UVE)[29,30] as well as its modifications, that is, modified UVE (m-UVE) and iterative variable elimination (IVE), can be used in 3D and 4D-QSAR schemes.[20,22,31] This allowed us to identify the molecular areas important for the interactions with

biological receptors or enzymes. In the current calculation, we used iterative variable elimination (IVE-PLS) that is a modification of the UVE algorithm based on the analysis of the regression coefficients calculated by the PLS method. PLS enables us to present the relation between the Y answer and X predictors in the form of

$$Y = Xb + e, \quad (8)$$

where b is a vector of the regression coefficients and e is the vector of the errors.

Thus, the UVE algorithm analyzes the reliability of the mean $(b)/s(b)$ ratio (where $s(b)$ means standard deviation of b). Then, only the variables of the 'relative' high mean $(b)/s(b)$ ratio are included into the final PLS model. Instead of a single-step UVE procedure, we used here an iterative algorithm based on the abs (mean $(b)/s(b)$) criterion to identify variables to be eliminated. This procedure includes:

1. Standard PLS analysis applied to analyze the matrices yielded from the s-CoMSA procedure with the leave-one-out cross-validation to estimate the performance of the PLS model (q^2_{CV}),
2. Elimination of the matrix column of the lowest abs (mean $(b)/s(b)$) value,
3. Standard PLS analysis of the new matrix without the column eliminated in step 2,
4. Iterative repetition of the steps 1–3 to maximize the LOO q^2_{CV} parameter.

All MATLAB functions and *m*-scripts are available from the authors on request.

3. Results and discussion

It is interesting to observe that atom/fragmental level descriptors provide better QSAR models for the HEPT compounds than molecular descriptors calculated for the whole molecules.[13] These results probably form the conformational flexibility of the series. In the traditional 3D-QSAR, the conformational flexibility of HEPT makes an important complication. Thus, the efficiency of the fragmental level descriptors can probably be explained by the fact that such a description does not need to define a molecule as a single conformational representation. Instead, it uses smaller molecular fragments. This significantly simplifies the problem of modeling molecular interaction pattern, but it also cannot give us any information on the hypothetical active conformation. Alternatively, a series of equations has been modeled for different subsets of the HEPT molecules in another 3D-QSAR study.[8] In our investigations we address the problem of the molecular flexibility of HEPTs using the 4D-QSAR method.

Table 2 reports the predictive power of some selected 4D-QSAR and SOM-4D-QSAR models measured by the *leave-one-out cross-validation* procedure. These values range from $q^2_{cv} = 0.77$, $s = 0.68$ (4D-QSAR—J_q) to $q^2_{cv} = 0.95$, $s = 0.32$ (4D-QSAR-J_q-IVE) versus $q^2_{cv} = 0.76$, $s = 0.69$ (SOM-4D-QSAR$_o$) to $q^2 = 0.98$, $s = 0.26$ (SOM-4D-QSAR$_q$-IVE), respectively. Generally, charge descriptors give better models (slightly higher q^2_{cv} values and significantly lower number of latent PLS variables) than the occupancy descriptors. Figure 1 gives molecular illustration of the SOM-4D-QSAR$_q$ model for the most active molecule (compound **32**). The areas indicating possible location of atoms are blurred in relatively many sections indicating conformational flexibility of the molecule. In order to select these areas that would pro-

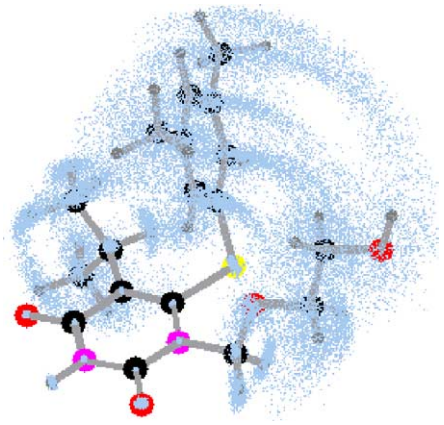


Figure 1. The points indicate all locations of atoms for the most active compound **32** during the MD simulations.

Table 2. The comparison of 4D-QSAR models for the HEPT analogues

Entry	Model[a]	q^2_{CV} (onc)[b]	s	SDEP
1	4D-QSAR-J_o[c,d]	0.78(10)	0.67	1.41
2	4D-QSAR-J_o-IVE	0.88(10)	0.49	1.44
3	SOM-4D-QSAR$_o$[d,e]	0.76(10)	0.69	1.46
4	SOM-4D-QSAR$_o$-IVE	0.92(10)	0.38	1.39
5	4D-QSAR-J_q[c,d]	0.77(9)	0.68	1.46
6	4D-QSAR-J_q-IVE	0.95(10)	0.32	1.48
7	SOM-4D-QSAR$_q$[d,f]	0.77(7)	0.66	1.76
8	SOM-4D-QSAR$_q$-IVE	0.98(9)	0.26	1.68

[a] Model: compounds **1**–**80** training set and **81**–**107** test set.
[b] (onc)—optimal number of PLS components.
[c] Box 30 Å:30 Å:30 Å.
[d] Compound **32** was used as reference compound **R**.
[e] sum_occupancy descriptor.
[f] mean_charges descriptor.

vide the best QSAR model, we eliminated uninformative variables using different variants of the so-called UVE or IVE method. In fact, all variants of this procedure improve model predictivity (Table 2). The *SDEP* values compare well to those reported in previous publications. For a detailed discussion of the statistical importance of this parameter in previous reports compare Ref. 31. Figure 2 indicates the molecular illustration of SOM-4D-$QSAR_q$ model for the most active molecule. In order to make this illustration more clear, we indicated only 10% of the points of the highest positive and negative contribution into the model. As variable elimination is stochastic in nature, the areas indicated somewhat depend upon the individual method used.

It is worth mentioning that our 4D-QSAR model conveys the most subtle shades of the HEPT-RT interactions, properly indicating an active conformation that is consistent with the one that has been actually observed in ligand–receptor X-ray structures (Fig. 3, conformation **b**—1,6-*trans*).[8] This compares advantageously to the previous 3D-QSAR study that was based on 1,6-*cis* conformation (Figure 3, conformation **a**). Our model also satisfies the *butterfly wing* pattern[13] properly indicating the interactions of the side chains, that is, hydrogen-acceptor/donor OH group (R^3) and R^2 group. The high predictive ability of the models as measured by q^2_{cv} is also important. Moreover, as recent 3D-QSAR investigations indicated that q^2 alone could not be a single parameter that tests model quality we always inspected the so-called external predictive ability of the models. Thus, the series is divided into two sub-series, the training set that is used to model basic equation on the basis of the cross-validated q^2_{cv} value and the test set that is used for predictions only. We used two different training/test set samplings. The first one (training: **1**–

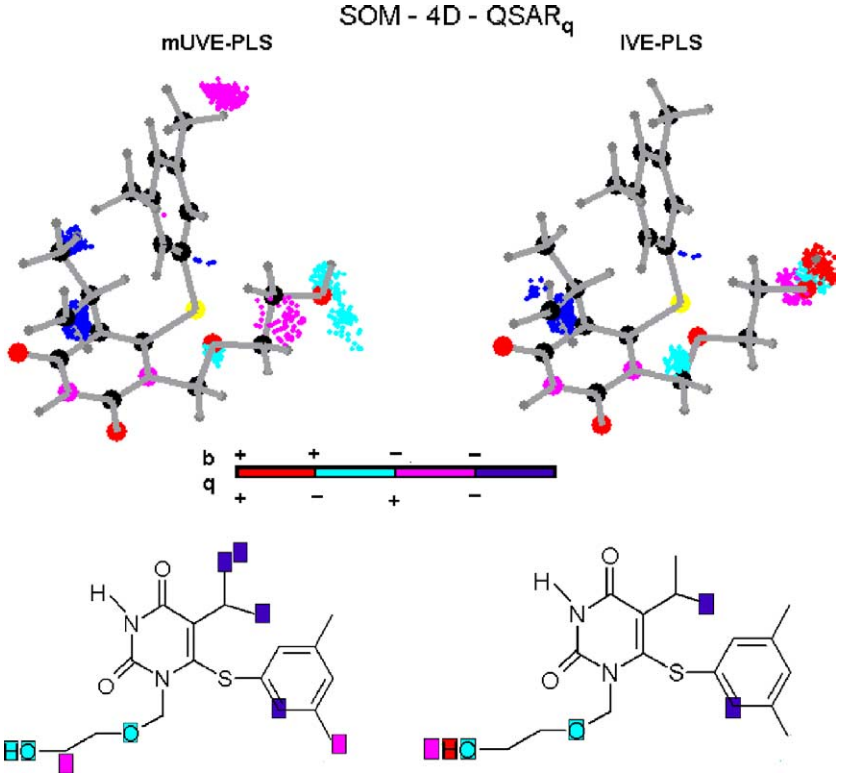


Figure 2. Molecular coordinates indicated by SOM-4D-QSAR models coupled with variable elimination methods for the compound of the highest activity. Colors code the combination of the mean partial atom charge sign and the sign of the *b* weight in the model: +/+ (red, increases the activity), −/+ (cyan, decreases the activity), +/− (magenta, decreases the activity), and −/− (blue, increases the activity).

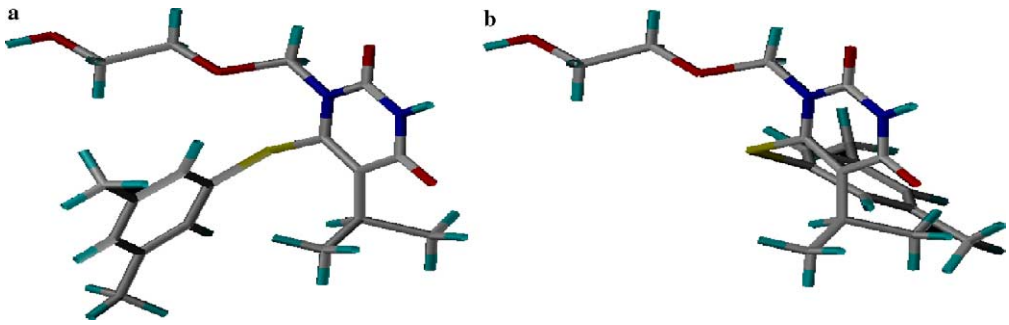


Figure 3. Two most stable conformational families found for the HEPT analogues.

80, test: **81**–**107**) complies with that used in previous QSAR and the second (training: **1**, **2**, **3**, **5**, **6**, **7**, **8**, **9**, **10**, **11**, **13**, **14**, **15**, **16**, **17**, **18**, **19**, **21**, **22**, **24**, **25**, **26**, **27**, **28**, **29**, **30**, **31**, **32**, **33**, **34**, **35**, **36**, **37**, **38**, **39**, **40**, **41**, **42**, **44**, **45**, **46**, **48**, **50**, **51**, **52**, **55**, **56**, **57**, **58**, **59**, **60**, **63**, **64**, **65**, **66**, **69**, **70**, **71**, **72**, **73**, **76**, **77**, **78**, **80**, **81**, **82**, **83**, **84**, **86**, **88**, **90**, **91**, **92**, **93**, **94**, **95**, **102**, **103**, **104**, **107**, test: **4**, **12**, **20**, **23**, **43**, **47**, **49**, **53**, **54**, **61**, **62**, **67**, **68**, **74**, **75**, **79**, **85**, **87**, **89**, **96**, **97**, **98**, **99**, **100**, **101**, **105**, **106**) was precisely calculated by the statistical rules of the Kennard–Stone criterion.[32] The respective data (given in Supporting Information Paragraph, Tables 3–6) indicate that the second method enables us to obtain the models of the better predictive ability. Moreover, these tests clearly indicated also that data elimination does not increase the SDEP error.

4. Conclusions

We used the 4D-QSAR method coupled with the PLS analysis and UVE or its variants for the investigations of the antiviral activity of HEPT, a series of conformationally flexible molecules that bind HIV-1 reverse transcriptase (RT). An analysis of a series of Hopfinger's and SOM-4D-QSAR models indicated that both methods yield comparable results. Generally, charge descriptors provide better modeling efficiency. We have shown that the method properly indicates the mode of interaction revealed by X-ray studies. It also allows us to calculate highly predictive QSAR models.

Acknowledgments

The authors thank Professor Johann Gasteiger of the University of Erlangen-Nürnberg, BRD, for facilitating access to the CORINA, PETRA, SURFACE, and KMAP programs. The financial support of the KBN Warsaw under Grant nos: KBN 4T09A 088 25 and PBZ 040 P04/08 is gratefully acknowledged.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2005.08.023.

References and notes

1. Makhija, M.; Kasliwal, R.; Kulkarni, V.; Neamati, N. *Bioorg. Med. Chem.* **2004**, *12*, 2317.
2. Pontikis, R.; Benhida, R.; Aubertin, A.-M.; Grierson, D. S.; Mohnert, C. *J. Med. Chem.* **1997**, *40*, 1845.
3. Jalali-Heravi, M.; Parastar, F. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 147.
4. De Clerq, E. *Med. Res. Rev.* **1993**, *13*, 229.
5. Young, S. D. *Perspect. Drug Discovery Des.* **1993**, *1*, 181.
6. Miyasaka, T.; Tanaka, H.; Baba, M.; Hayakawa, H.; Walker, T.; Balzarini, J.; De Clercq, E. *J. Med. Chem.* **1989**, *32*, 2507.
7. Hannongbua, S.; Nivesanond, K.; Lawtrakul, L.; Pungpo, P.; Wolschann, P. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 848.
8. Kireev, D.; Chrétien, J.; Grierson, D.; Monneret, C. *J. Med. Chem.* **1997**, *40*, 257.
9. Hopkins, L.; Ren, J.; Esnouf, M.; Willcox, E.; Jones, Y.; Ross, C.; Miyasaka, T.; Walker, T.; Tanaka, H.; Stammers, K.; Stuart, I. *J. Med. Chem.* **1996**, *39*, 1589.
10. Schaefer, W.; Friebe, G.; Leinert, H.; Mertens, A.; Poll, T.; Van der Saal, W.; Zilch, H.; Nuber, B.; Ziegler, L. *J. Med. Chem.* **1993**, *36*, 726.
11. Ragno, R.; Frasca, S.; Mannetti, F.; Brizzi, A.; Massa, S. *J. Med. Chem.* **2005**, *48*, 200.
12. Luco, M.; Ferretti, H. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 392.
13. Gayen, S.; Debnath, B.; Samanta, S.; Jha, T. *Bioorg. Med. Chem.* **2004**, *12*, 1493.
14. Douali, L.; Villemin, D.; Charquaoui, D. *Curr. Pharm. Des.* **2003**, *9*, 1817.
15. Pungpo, P.; Hannongbua, S.; Wolschann, P. *Curr. Med. Chem.* **2003**, *17*, 1661.
16. Hopfinger, A.; Wang, S.; Tokarski, S.; Jin, B.; Albuquerque, M.; Madhav, J.; Duraiswami, C. *J. Am. Chem. Soc.* **1997**, *119*, 10509.
17. Venkatarangan, P.; Hopfinger, A. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1141.
18. Krasowski, M.; Hong, X.; Hopfinger, A.; Harrison, N. *J. Med. Chem.* **2002**, *45*, 3210.
19. Hong, X.; Hopfinger, A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 324.
20. Polanski, J.; Bak, A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2081.
21. Polanski, J.; Bak, A.; Gieleciak, R.; Magdziarz, T. *Molecules* **2004**, *9*, 1148.
22. Polanski, J.; Gieleciak, R.; Magdziarz, T.; Bak, A. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1423.
23. Sybyl 6.5 program available from the Tripos Inc., St. Louis, MO, USA, http://www.tripos.com.
24. HyperChem 5.0 program available from the HyperCube Inc., Gaines-Ville, FL, USA, http://www.hyper.com.
25. Wold, S.; Sjöström, M.; Erikson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109.
26. Lindgren, F.; Geladi, P.; Wold, S. *J. Chemometrics* **1993**, *7*, 45.
27. Helland, J. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 97.
28. Matlab 6.5 program available from the MathWorks Inc., Notich, MA, USA, http://www.mathworks.com.
29. Centner, V.; Massart, L.; De Noord, E.; De Jong, S.; Vandeginste, V.; Sterna, C. *Anal. Chim. Acta* **1996**, *330*, 1.
30. Koshoubu, J.; Iwata, T.; Minami, S. *Anal. Sci.* **2001**, *17*, 319.
31. Polanski, J.; Gieleciak, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 656.
32. Kennard, W.; Stone, A. *Technometrics* **1969**, *11*, 137.