Digest Journal of Nanomaterials and Biostructures, Vol. 2, No. 2, June 2007, p. 207 - 219

QSAR MODELLING OF ANTI-HIV ACTIVITY WITH HEPT DERIVATIVES

C. Duda-Seiman, D. Duda-Seiman[a], M. V. Putz[*], D. Ciubotariu[b]

University of West Timişoara, Department of Chemistry, Street Pestalozzi no. 16, RO-300115, Timişoara, Romania
[a]University of Medicine and Pharmacy "Victor Babeş", Department of Medical Ambulatory, Cardiovascular Rehabilitation, Medical Emergencies, Avenue C.D. Loga no. 49, RO-300020, Timişoara, Romania
[b]University of Medicine and Pharmacy "Victor Babeş", Faculty of Pharmacy, Avenue Eftimie Murgu no. 2, RO-300041, Timişoara, Romania

We performed studies upon an extended series of 79 HEPT ligands (1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine), inhibitors of HIV *reverse-transcriptase – RT*, with anti-HIV biological activity, using QSAR methods that imply analysis of correlations and multiple linear regression; a significant collection of molecular descriptors was used, e.g. the number of oxygen atoms, content of bond information, Broto-Moreau autocorrelation of topologic structure of orders 6 and 8 normalized to Sanderson electronegativities, or Moran autocorelation of order 5 normalized to atomic mass. The QSAR analysis of the 79 bioactive compounds indicated that there are three important zones to promote biologic activity. The obtained multi-parametric models when different classes of structural descriptors were used led to correlation coefficients closed to 0.9, especially while combining topologic parameters and hydrophobicity. Constitutive parameters did not bring significant improvements of the quality of the obtained statistics.

(Received May 4, 2007; accepted May 9, 2007)

Keywords: QSAR, Anti-HIV activity, HEPT ligands

1. Introduction

AIDS (Acquired Immunodeficiency Syndrome) is still an illness with lethal evolution. Still, there is no any complete effective chemotherapy. The disease is caused by a retrovirus of the Lentiviridae family with monocatenar RNA called the Human Immunodeficiency Virus – HIV [1]. The virus annihilates the organism's natural capacity for self-defense: it compromises the immune system and creates a favorable environment to develop various infective diseases (opportunistic infections) with increased lethal potential. The same mechanism is involved in development of various neoplasias. The nervous system represents also a target of HIV with severe neurological and psychiatrical consequences. Once the viral replication has reached an explosive rate, the chances to control the disease are minimal [2, 3].

Daily, 8000 people die of AIDS. It is estimated at global level that over 70 million individuals will die because of AIDS in the next 20 years. Although new anti-HIV pharmacological agents are permanently developed and hope for an effective therapy using vaccines is arising, mortality through AIDS is yet increased. The main possible strategy still comprises of measures in order to avoid the extension of the epidemic. In 2003, 2,8 bln. Euros were spent in research and health care only for this disease; UNO showed that at least 10 bln Euros are necessary yearly. HIV/AIDS research is a priority in the EURO/HIV program, financed by the European Committee [4].

[*] Corresponding aithor: mvputz@cbg.uvt.ro or mv_putz@yahoo.com

These data are strong motivations for research especially in the field of relationships among different classes of viral inhibitors and also for knowledge of HIV and its consequences at the molecular level. Thus, these concepts are a computational priority by means of *de nuovo* design of compounds with increased therapeutic properties and fewer side effects. Our work studies the role of different structural parameters in the case of a series of pyrimidinic congeners with anti-HIV activity: the objective was to assess electronic, transportation and topological effects of an ideal ligand to express while binding to target receptors.

2. Modelling

2.1 About HIV

The Human Immunodeficiency Virus is a retrovirus because it possesses an enzyme able to transform RNA into DNA (the reverse process compared to the common process in cellular biology) [5].

Known human retroviruses are: HTLV I, HTLV II and HIV. Pathogenic human retroviruses (HTLV I, HIV) have the same target – T lymphocytes. Thus, Robert Gallo et al. [6] named these agents human T-lymphotropic viruses.

High resolution electronic microscopy shows that HIV-1 is a 100 nm virus with a capsule [7]. The external layer is a double lipidic layer derived from the host cell during maturation and contains two major viral glicoproteins (gp): the transmembranar gp41 and outside gp120. Immediately below there is a protein associated to the membrane (p18) which provides the matrix for the viral structure and which is essential for the integrity of the virus. The matrix surrounds a dense cylindrical characteristic nucleoid which contains the p24 protein from the capside. Inside of the nucleoid there are 2 identical RNA strains; the viral RNA dependant DNA-polymerase (p66/p55) called reverse-transcriptase (RT) is related to p9 nucleoprotein, to p12 integrase protein and to components of p15 protease, see Fig. 1 [8].

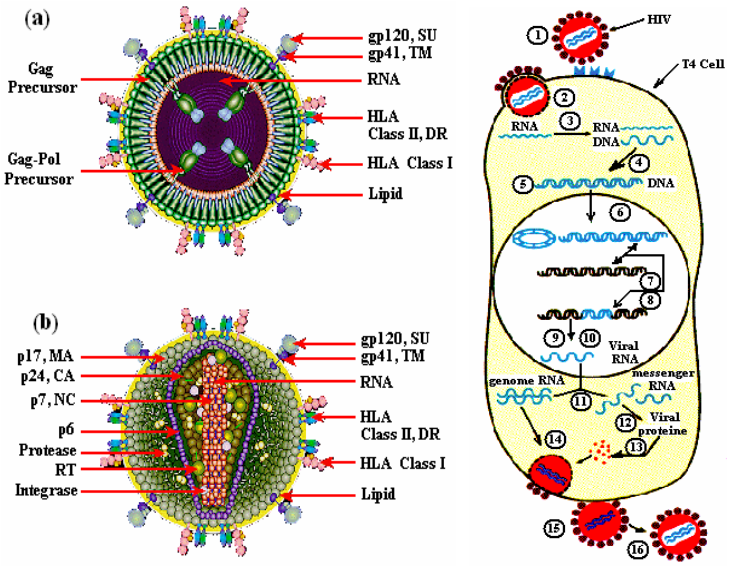

Fig. 1. Left: the human immunodeficiency virus (HIV) in its immature (a) and mature (b) stages [8]. Right: life cycle of HIV [9].

The main property is the variability of the viral envelope; mutations are responsible for changes in the envelope's structure providing new pathogenic viral properties which appear 100 times faster than in influenza viruses.

Structural (and antigenic) dynamics in certain portions of the envelope provides to HIV viruses the possibility to avoid the neutralizing action of formed antibodies. The viral genetic patrimony comprises three genes encoding viral components: gag- determines the synthesis of nucleotide proteins, pol- determines the synthesis of the reverse transcriptase enzyme, while env- determines the synthesis of envelope glycoproteins.

HIV life cycle starts with a high affinity of its membrane via gp120 for the CD4 receptor (proteic molecule) at the host cell's surface - see Fig. 1 - right [9]. CD4 receptor is a proteic molecule present especially on the surface of T lymphocytes that are *helper* and *inducer*, respectively, of the immune response (T_4 lymphocytes modulate the immune response). The next stage, the fusion between the virus and the membrane of the host cell, emerges through gp41 and the HIV genomic RNA. RT and the genomic RNA enter the double DNA collaring. Thus, DNA migrates towards the nucleus in order to be integrated in the host cell's chromosomes by the viral encoded integrase enzyme. The incorporation of the formed "provirus" into the host cell genome is permanent [10, 11].

The activation of the provirus is accomplished via the membrane of the host cell and the virus gains its external envelope. During the final activation process, the HIV protease contributes to the cleavage of the gag-pol glycoprotein precursor. In this way, the maturation of the viral particle is achieved. These stages of the replication cycle of HIV are viable targets for anti-HIV chemotherapyc agents in order to stop the infection of healthy host cells [1].

2.2 About HEPT

An ideal anti-HIV agent should stop the virus' rankness and also the infection of healthy host cells with now toxicity against normal cell physiology.

An anti-HIV agent can exert its biological activity in different stages of the viral life cycle inhibiting them. Studies were limited to those stages and phenomenon that appear during viral replication: viral binding to the target cell, viral fusion with the host cell by viral penetration into the host cell's membrane, viral uncovering in the host cell, reverse genomic RNA transcription, integration of the new viral DNA into the host cell's chromosomes, provirus activation producing mRNA, viral detachment from the host cell, viral maturation.

Reverse transcription of viral genomic RNA into double strained DNA by the RT enzyme is essential for HIV replication. Thus, the inhibition of this essential phase of HIV life cycle provides the most attractive target in order to develop a compound with biological anti-HIV potential. For example, most drugs approved by the FDA for HIV infection treatment are RT inhibitors.

By these means, HEPT (1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine) derivatives are *non-nucleosidic reverse transcriptase inhibitors (NNRTI)*, see Figure 2, and they are analogues of the natural substrate. HEPT derivatives don't interact with the binding site of the DNA or RNA dependant DNA-polymerase. This is why it is expected these ligands not to determine side effects. HEPT ligands interact uncompetitive with an allosteric site of the enzyme and don't affect in a direct way the substrate binding. Actually, *NNRTI* have a higher binding affinity to the ligand-enzyme complex than to the free enzyme. The HEPT ligand – enzyme interaction leads to enzymatic conformational variations meaning that the enzyme's active site has a decreased affinity to the natural substrate. This property is valid only regarding the HIV-1 RT; HEPT ligands are inactive against HIV-2 or other retroviruses. The *NNRTI* exclusive specificity for the HIV-1 RT is due to the presence at the level of this enzyme (and not in the case of other RT or DNA-polymerases) of a flexible extreme hydrophobic pocket in which HEPT derivatives (different from natural substrate analogues) fit and can be bound.

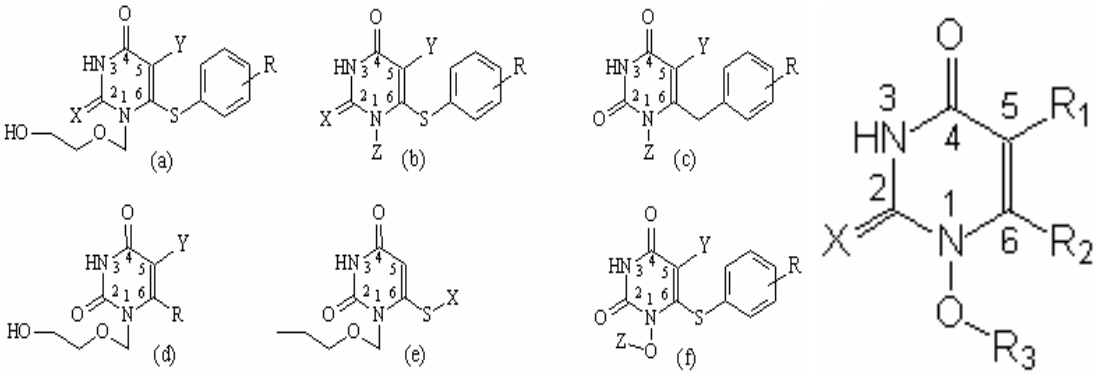

Fig. 2. Left: Typical examples of HEPT **(1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine)** derivatives. Right: The reference structure of HEPT derivatives under study.

Literature data regarding quantitative studies upon chemical structure – biological activity relationships (QSAR) [12-21] in the class of HEPT derivatives are strictly related to the presence or absence of atomic groups in certain positions on the HEPT general structure, see Fig. 2 - right. Our research is directed towards an extended HEPT series and is performed using the Hansch model with structural descriptors adapted to multi-linear regression technique (MLR).

2.3. Background of QSAR modelling

Actual drug design methods quantify biological activity depending on the molecular structure. Usually it is accomplished by modifying a reference structure through grafting X substituents. This leads to a series of bioactive compounds called effectors (E), or more recently ligands (L). L biological activities are determined and then correlated with the structure using correlation analysis methods, multi-linear regression equations like:

$$A_x = a + b_1 t_x + b_2 t_x^2 + b_3 \sigma_{I,x} + b_4 \sigma_{D,x} + b_5 S \qquad (1)$$

where A represents bioactivity, t is a transportation parameter, σ_I and σ_D express localized (field and/or inductive) and not localized (resonance) electronic effects, respectively; S measures steric effects; bioactivity A is usually related to log (1/C), where C is the molar concentration which determines a constant biological response.

Correlation equation (1) or other derived relations, statistically validated, are named quantitative structure – biological activity relationship (QSAR). If on the molecular structure of the ligand L more X substituents are grafted, their effects can be regarded in an additive way or, on the contrary, each of substituents is treated separately. The difficulties to characterize quantitatively a certain structure using the t, σ and S parameters explain the high diversity of developed parametric scales.

Usually the transportation parameter is used as the logarithm of the partition coefficient of the bioactive compound (L), almost in all cases determined between water and 1-octanol (it simulates the lipidic hydrophobic membrane), or other derived parameters. Recently, chromatographic obtained parameters are more utilized. The t^2 term was introduced to take in account the frequent noticed parabolic dependence between bioactivity and the transportation parameter. These non-linear behaviors in some QSAR studies were explained by models that consider that when an effector passes a biological membrane it is first transferred from the aqueous phase into the lipidic phase of the membrane and then retransferred on the other side into the aqueous phase, etc.

The electronic structure of bioactive compounds is usually quantified starting from the extra-thermodynamical model using the well known σ-Hammett substituent constants [22] together with further refining [23]. This was and remains in most cases the most proper parameterization possibility of electronic effects; there are large computer data bases comprising such constants concerning very numerous substituents [24]. An alternative approach consists in using quantum mechanics methods providing quantum reactivity indices like bonding orders, charges on an atom series, etc. The latter direction will be followed in this study.

3. Results and discussion

3.1. Computational details

The considered structures to be analyzed starting from the general structure shown in Fig. 2 - right together with experimental anti-HIV activities are presented in Table 1.

Table 1. The series of HEPT derivatives employed in this study. In table are given the group of substituents considered on the general structure of Fig. 2 - right. Biological activities are given as A_{exp} = log (1/ EC50), where EC50 represents the concentration which produces a 50% protection of MT-4 cells against the direct toxic HIV-1 effect [1, 3, 4].

No.	R_1	R_2	R_3	X	A_{exp}
1	methyl	2-methylphenylthio	2-hidroxyethyl	O	4.15
2	methyl	2-nitrophenhylthio	2-hidroxyethyl	O	3.85
3	methyl	2-methoxyphenylthio	2-hidroxyethyl	O	4.72
4	methyl	3-metilfeniltio	2-hidroxietil	O	5.59
5	methyl	3-ethylphenylthio	2-hidroxyethyl	O	5.57
6	methyl	3-*tert*buthylphenylthio	2-hidroxyethyl	O	4.92
7	methyl	3-trifluoromethylphenylthio	2-hidroxyethyl	O	4.35
8	methyl	3-fluorophenylthio	2-hidroxyethyl	O	5.48
9	methyl	3-clorophenylthio	2-hidroxyethyl	O	4.89
10	methyl	3-bromophenylthio	2-hidroxyethyl	O	5.24
11	methyl	3-iodophenylthio	2-hidroxyethyl	O	5.00
12	methyl	3-nitrophenylthio	2-hidroxyethyl	O	4.47
13	methyl	3-hidroxyphenylthio	2-hidroxyethyl	O	4.09
14	methyl	3-metoxyphenylthio	2-hidroxyethyl	O	4.66
15	methyl	3,5-dimetilphenylthio	2-hidroxyethyl	O	6.59
16	methyl	3,5-diclorophenylthio	2-hidroxyethyl	O	5.89
17	methyl	3,5-dimethylphenylthio	2-hidroxyethyl	S	6.66
18	methyl	3-metoxycarbonylphenylthio	2-hidroxyethyl	O	5.10
19	methyl	3-acetylphenylthio	2-hidroxyethyl	O	5.14
20	methyl	3-cyanophenylthio	2-hidroxyethyl	O	5.00
21	allyl	phenylthio	2-hidroxyethyl	O	5.60
22	ethyl	phenylthio	2-hidroxyethyl	S	6.96
23	propyl	phenylthio	2-hidroxyethyl	S	5.00
24	*iso*propyl	phenylthio	2-hidroxyethyl	S	7.23
25	ethyl	3,5-dimethylphenylthio	2-hidroxyethyl	S	8.11
26	*iso*propyl	3,5-dimethylphenylthio	2-hidroxyethyl	S	8.30
27	ethyl	3,5-diclorophenylthio	2-hidroxyethyl	S	7.37
28	ethyl	phenylthio	2-hidroxyethyl	O	6.92
29	propyl	phenylthio	2-hidroxyethyl	O	5.47
30	*iso*propyl	phenylthio	2-hidroxyethyl	O	7.20
31	ethyl	3,5-dimetylphenylthio	2-hidroxyethyl	O	7.89
32	*iso*propyl	3,5-dimetylphenylthio	2-hidroxyethyl	O	8.57
33	ethyl	3,5-diclorophenylthio	2-hidroxyethyl	O	7.85
34	methyl	4-metylphenylthio	2-hidroxyethyl	O	3.66
35	methyl	phenylthio	2-hidroxyethyl	O	5.15
36	methyl	phenylthio	2-hidroxyethyl	S	6.01
37	iodo	phenylthio	2-hidroxyethyl	O	5.44
38	ethenyl	phenylthio	2-hidroxyethyl	O	5.69
39	2-phenylethenyl	phenylthio	2-hidroxyethyl	O	5.22
40	benzyl	phenylthio	2-hidroxyethyl	O	4.37
41	methyl	phenylthio	2-metoxyethyl	O	5.06
42	methyl	phenylthio	2-acetyloxyethyl	O	5.17
43	methyl	phenylthio	2-benzoyloxyethyl	O	5.12
44	methyl	phenylthio	ethyl	O	6.48
45	methyl	phenylthio	2-cloroethyl	O	5.82
46	methyl	phenylthio	2-azidoethyl	O	5.24

47	methyl	phenylthio	2-fluoroethyl	O	5.96
48	methyl	phenylthio	propyl	O	5.48
49	methyl	phenylthio	benzyl	O	7.06
50	ethyl	phenylthio	ethyl	O	7.72
51	ethyl	phenylthio	ethyl	S	7.58
52	ethyl	3,5-dimethylphenylthio	ethyl	O	8.24
53	ethyl	3,5-dimethylphenhylthio	ethyl	S	8.30
54	ethyl	phenylthio	benzyl	O	8.23
55	ethyl	3,5-dimethylphenylthio	benzyl	O	8.55
56	ethyl	phenylthio	benzyl	S	8.09
57	ethyl	3,5-dimethylphenylthio	benzyl	S	8.14
58	*iso*propyl	phenylthio	ethyl	O	7.99
59	*iso*propyl	phenylthio	benzyl	O	8.51
60	*iso*propyl	phenylthio	ethyl	S	7.89
61	*iso*propyl	phenylthio	benzyl	S	8.14
62	methyl	phenylthio	methyl	O	5.68
63	methyl	phenylthio	butyl	O	5.33
64	methyl	phenylthio	methyl	S	5.66
65	methyl	phenylthio	propyl	S	5.92
66	ethyl	3,5-diclorophenylthio	ethyl	S	7.89
67	ethyl	phenylthio	*iso*propyl	S	6.66
68	ethyl	phenylthio	cyclohexyl	S	5.79
69	ethyl	phenylthio	cyclohexylmethyl	S	6.45
70	ethyl	phenylthio	4-methylbenzyl	S	7.11
71	ethyl	phenylthio	4-clorobenzyl	S	7.92
72	ethyl	phenylthio	2-phenylethyl	S	7.04
73	ethyl	3,5-diclorophenylthio	ethyl	O	8.13
74	ethyl	phenylthio	*iso*propyl	O	6.47
75	ethyl	phenylthio	cyclohexyl	O	5.40
76	ethyl	phenylthio	cyclohexylmethyl	O	6.35
77	ethyl	phenylthio	2-ciclohexylethyl	O	7.02
78	cyclopropyl	phenylthio	ethyl	S	7.02
79	cyclopropyl	phenylthio	ethyl	O	7.00

Upon these structures, *molecular modeling* was performed using the HyperChem 7.01 programme package (MM+ programme) [25] with 0.05 kcal/mol RMS gradient. Optimization process comprised the Polak-Ribiere algorithm with conjugated gradient. *Conformational analysis* was performed with Conformational Search from HyperChem 7.01 package program [25].

Table 2. Structural descriptors for biological activity considered in this work.

Descriptor code	**Description**
nO	Number of oxygen atoms
PW4	Randic Shape index - replacement of order 4
CIC4	Content of complementary information (adjacent symmetry of order 4)
BIC4	Content of bond information (adjacent symmetry of order 4)
BIC5	Content of bond information (adjacent symmetry of order 5)
ATS6e	Broto-Moreau autocorrelation of topologic structure of order 6, normalized to Sanderson electronegativities
ATS8e	Broto-Moreau autocorrelation of topologic structure of order 8, normalized to Sanderson electronegativities
MATS5e	Moran autocorelation of order 5 normalized to atomic mass
MATS6v	Type 6 Moran autocorrelation, normalized to atomic polarizabilities
nHAcc	Number of accepting atoms for hydrogen bonds (N, O, F)

Anti-HIV activities from Table 1 will be correlated with structural and topological indices from Table 2. To calculate these descriptors, DRAGON programme [26] was used; it allows importing *3D* structures from the HyperChem package [25]. Results are shown in Table 3.

3.2. QSAR mono-linear correlations

Mono-parametric correlations of biologic activity in the series of 79 studied ligands of Table 1 are shown in equations (2)-(9) by means of descriptors determined in Table 3.

Table 3. Structural parameters used for mono- and multi- linear QSAR regressions for studing the HEPT series of Table 1. The descriptions of the codes are given in Table 2. All computations are within Dragon and HyperChem environments [25,26].

No.	*nO*	*PW4*	*CIC4*	*BIC5*	*ATS6e*	*MATS5e*	*MATS6v*	*nHAcc*	*logP*
1	4	0.179	0.365	0.887	1.066	-0.047	-0.195	6	2.14
2	6	0.184	0.299	0.892	1.109	-0.208	-0.265	9	1.62
3	5	0.187	0.356	0.891	1.070	-0.054	-0.23	7	1.42
4	4	0.179	0.365	0.887	1.066	-0.047	-0.195	6	2.14
5	4	0.181	0.388	0.888	1.052	-0.044	-0.259	6	2.53
6	4	0.167	0.914	0.805	1.037	0.013	-0.284	6	3.3
7	4	0.167	0.365	0.887	1.102	-0.139	-0.261	9	2.55
8	4	0.178	0.257	0.902	1.089	-0.143	-0.312	7	1.81
9	4	0.178	0.257	0.902	1.083	-0.147	-0.286	6	2.19
10	4	0.178	0.257	0.902	1.08	-0.143	-0.25	6	2.46
11	4	0.178	0.257	0.902	1.075	-0.124	-0.209	6	2.93
12	6	0.175	0.299	0.892	1.091	-0.155	-0.299	9	1.62
13	5	0.178	0.250	0.905	1.08	-0.134	-0.294	7	1.39
14	5	0.181	0.356	0.891	1.068	-0.072	-0.306	7	1.42
15	4	0.182	0.807	0.813	1.038	-0.102	-0.073	6	2.61
16	4	0.182	0.493	0.858	1.093	-0.18	-0.265	6	2.71
17	3	0.182	0.807	0.813	1.016	-0.064	-0.132	5	3.25
18	6	0.177	0.338	0.892	1.061	-0.101	-0.203	8	1.4
19	5	0.175	0.346	0.889	1.064	-0.058	-0.282	7	0.98
20	4	0.181	0.250	0.894	1.077	-0.129	-0.29	7	1.71
21	4	0.191	0.421	0.874	1.077	-0.14	-0.209	6	2.25
22	3	0.192	0.507	0.861	1.042	-0.193	-0.239	5	2.72
23	3	0.191	0.519	0.865	1.054	-0.128	-0.224	5	3.11
24	3	0.191	0.738	0.825	1.04	-0.295	-0.123	5	3.05
25	3	0.193	0.797	0.82	1.017	-0.178	-0.044	5	3.65
26	3	0.192	0.979	0.793	1.018	-0.267	0.025	5	3.98
27	3	0.193	0.507	0.861	1.059	-0.249	-0.215	5	3.75
28	4	0.192	0.507	0.861	1.067	-0.225	-0.209	6	2.07
29	4	0.191	0.519	0.865	1.076	-0.166	-0.2	6	2.46
30	4	0.191	0.738	0.825	1.062	-0.308	-0.135	6	2.4
31	4	0.193	0.797	0.820	1.037	-0.198	-0.026	6	3
32	4	0.192	0.979	0.793	1.036	-0.273	0.011	6	3.33
33	4	0.193	0.507	0.861	1.084	-0.282	-0.168	6	3.1
34	4	0.173	0.581	0.848	1.072	-0.053	-0.31	6	2.14
35	4	0.180	0.493	0.858	1.073	-0.113	-0.31	6	1.67
36	3	0.180	0.493	0.858	1.044	-0.054	-0.398	5	2.32
37	4	0.180	0.387	0.87	1.085	-0.091	-0.263	6	2.03
38	4	0.192	0.400	0.872	1.072	-0.19	-0.197	6	1.85
39	4	0.180	0.489	0.865	1.064	-0.072	-0.123	6	3.42
40	4	0.183	0.545	0.858	1.076	-0.048	-0.159	6	3.28
41	4	0.175	0.581	0.848	1.076	-0.124	-0.228	6	1.95
42	5	0.163	0.552	0.852	1.075	-0.012	-0.113	7	1.8
43	5	0.176	0.582	0.851	1.063	-0.11	-0.247	7	3.71
44	3	0.189	0.591	0.837	1.051	-0.086	-0.302	5	2.46
45	3	0.180	0.508	0.852	1.073	-0.105	-0.25	5	2.82

46	3	0.171	0.479	0.857	1.073	-0.063	-0.285	8	3.15
47	3	0.180	0.508	0.852	1.085	-0.096	-0.299	6	2.27
48	3	0.180	0.597	0.843	1.046	-0.119	-0.293	5	2.93
49	3	0.177	0.619	0.838	1.050	-0.126	-0.114	5	3.89
50	3	0.201	0.597	0.843	1.048	-0.215	-0.202	5	2.85
51	2	0.201	0.597	0.843	1.024	-0.184	-0.246	4	3.5
52	3	0.201	0.881	0.803	1.021	-0.194	-0.021	5	3.79
53	2	0.201	0.881	0.803	1.002	-0.178	-0.049	4	4.44
54	3	0.187	0.622	0.843	1.047	-0.239	-0.047	5	4.29
55	3	0.188	0.822	0.811	1.023	-0.22	0.103	5	5.22
56	2	0.187	0.622	0.843	1.028	-0.219	-0.007	4	4.94
57	2	0.188	0.822	0.811	1.007	-0.21	0.142	4	5.87
58	3	0.199	0.827	0.807	1.046	-0.309	-0.13	5	3.18
59	3	0.186	0.815	0.815	1.045	-0.327	0.002	5	4.62
60	2	0.199	0.827	0.807	1.024	-0.304	-0.129	4	3.83
61	2	0.186	0.815	0.815	1.028	-0.33	0.073	4	5.27
62	3	0.189	0.584	0.83	1.042	-0.146	-0.339	5	2.11
63	3	0.175	0.603	0.848	1.05	-0.134	-0.217	5	3.32
64	3	0.189	0.584	0.83	1.042	-0.146	-0.339	5	2.11
65	3	0.180	0.597	0.843	1.046	-0.119	-0.293	5	2.93
66	2	0.201	0.597	0.843	1.041	-0.245	-0.219	4	4.54
67	2	0.190	0.827	0.807	1.026	-0.138	-0.282	4	3.91
68	2	0.191	0.799	0.821	1.025	-0.157	-0.149	4	4.74
69	2	0.187	0.791	0.827	1.025	-0.173	-0.011	4	5.06
70	2	0.182	0.642	0.837	1.027	-0.165	-0.088	4	5.4
71	2	0.182	0.576	0.843	1.039	-0.223	-0.089	4	5.45
72	2	0.187	0.625	0.847	1.027	-0.144	-0.050	4	5.19
73	3	0.201	0.597	0.843	1.065	-0.275	-0.157	5	3.89
74	3	0.190	0.827	0.807	1.054	-0.157	-0.191	5	3.27
75	3	0.191	0.799	0.821	1.051	-0.164	-0.13	5	4.1
76	3	0.187	0.791	0.827	1.044	-0.187	-0.041	5	4.41
77	3	0.187	0.784	0.832	1.04	-0.115	0.006	5	4.17
78	2	0.193	0.669	0.827	1.026	-0.222	-0.143	4	3.33
79	3	0.193	0.669	0.827	1.049	-0.246	-0.136	5	2.68

Considering constitutional parameters, the one which counts the number of oxygen atoms (*nO*) led to the best mathematical model ($r = 0.640$). Biologic activity decreases with the increase of the number of atoms of oxygen in the molecule eq. (2). The same dependency is noticed in case of *Ms* (Mean electro-topological state) parameters, being in fact inter-correlated ($r_{inter\text{-}correlation} = 0.790$).

$$A_i = 9.248(\pm 0.421) - 0.865(\pm 0.118) \cdot nO_i$$
$$(n = 79;\ r = 0.640;\ s = 1.040;\ F = 53.39) \quad (2)$$

The shape index (*PW4)* determined the best mono-linear correlation while using simple topologic parameters, see eq. (3). From the other calculated topologic parameters, those which count the information content regarding binding and complementarity aspects (proximity symmetry) were the most proper to obtained mono-parametric models. Thus, *BIC* parameters (binding symmetry) determined the best equations with correlation coefficients above 0.69, see equation (4); *CIC* parameter regarding structural and complementary symmetry led to $r > 0.6$, see equation (4). Higher values of parameters that define binding and structural symmetries determine the decrease of biologic activity, while the opposite situation is described for complementary symmetry parameters. An increased biologic activity implies a more asymmetrical molecular shape.

$$A_i = -13.789(\pm 2.436) + 108.338(\pm 13.132) \cdot PW4_i$$
$$(n = 79;\ r = 0.685;\ s = 0.986;\ F = 68.06) \quad (3)$$
$$A_i = 32.341(\pm 3.035) - 30.758(\pm 3.581) \cdot BIC4_i$$
$$(n = 79;\ r = 0.700;\ s = 0.968;\ F = 73.76) \quad (4)$$
$$A_i = 3.518(\pm 0.358) + 4.717(\pm 0.578) \cdot CIC4_i$$

(n = 79; r = 0.681; s = 0.992; F = 66.56) (5)

$$A_i = 4.315(\pm 0.237) - 12.459(\pm 1.340) \cdot MATS5e_i$$

(n = 79; r = 0.727; s = 0.929; F = 86.45) (6)

$$A_i = 7.728(\pm 0.200) + 8.156(\pm 0.952) \cdot MATS6v_i$$

(n = 79; r = 0.700; s = 0.969; F = 73.43) (7)

Table 4 presents the *r* inter-correlation coefficients for the discussed parameters. All symmetry parameters have inter-correlation coefficients above 0.9, but they do not inter-correlate significantly with simple topologic indices ($r < 0.600$). Most of the *2D* (auto-correlation) parameters show no inter-correlations. These parameters led to the best results when linear mono-parametric models were considered, especially when following parameters were used: the auto-correlation Moran parameters (normalized to electronegativities), see equation (6), or the van der Waals volumes, see equation (7), relating polarizability. However, the electronic component intervenes also in these last situations (by normalizing to electronegativities); thus, biologic activity decreases while the so normalized auto-correlation parameter increases.

Steric and simple functional parameters, calculated with DRAGON [26] or with QSAR properties [25], had statistical significance, but they did not lead to very good mathematical models. The exception was the parameter that counts the number of atoms accepting hydrogen bindings (N, O, F), *nHAcc*, see equation (8), which shows a quite good inter-correlation with the similar parameter which counts the number of oxygen atoms in the molecule, *nO* ($r = 0.9$). When using the transport parameter log*P* in a mono-linear equation, the two available methods, i.e the Moriguchi octanol/water partition coefficient (*MLOGP*) within DRAGON programme [26]) and those computed upon atomic contributions (log*P*) within HyperChem environment [25] leads the inter-correlation coefficient between them as 0.96. However, the best equation was obtained when log*P* was calculated through the second way ($r = 0.67$), see equation (9).

Table 4. Intercorrelation coeficients between parameters used in molecular QSAR corelations (2)-(12).

	nO	***PW4***	***CIC4***	***BIC5***	***ATS6e***	***MATS5e***	***MATS6v***	***nHAcc***	***logP***
nO	1.00	-0.15	-0.59	0.63	0.73	0.38	-0.40	-0.90	-0.77
PW4		1.00	0.45	-0.47	-0.49	-0.67	0.36	-0.57	0.36
CIC4			1.00	-0.98	-0.79	-0.41	0.66	-0.62	0.66
BIC5				1.00	0.78	-0.42	-0.59	-0.64	-0.63
ATS6e					1.00	0.29	-0.60	0.78	-0.70
MATS5e						1.00	-0.50	0.37	-0.44
MATS6v							1.00	-0.44	0.73
nHAcc								1.00	-0.72
logP									1.00

There gets out that the anti-HIV biologic activity increases with the increase of the octanol/water partition coefficient. This indicates that the transport of the bioactive compound to the target (receptor site) is important in the studied series of ligands.

$$A_i = 10.401(\pm 0.549) - 0.740(\pm 0.097) \cdot nHAcc_i$$

(n = 79; r = 0.657; s = 1.020; F = 58.62) (8)

$$A_i = 3.838(\pm 0.327) + 0.790(\pm 0.099) \cdot \log P_i$$

(n = 79; r = 0.673; s = 1.001; F = 63.85) (9)

In most cases correlation coefficient above 0.600, some of the obtained models being closed to this value. In other cases simple linear correlation led to correlation coefficients above 0.700 (*MATS5e*). HyperChem parameters led to simple linear models with relatively small *r*

correlation coefficients, but statistically significant. Only log*P* determined a better correlation ($r > 0.670$).

3.3. QSAR multi-linear correlations

The attempt to obtain multi-linear mathematical models using different classes of molecular descriptors led to improvement of correlation coefficients and of other statistic parameters while the coefficient of inter-correlation between BIC5 and ATS6e is 0.78 in Table 4.

$$A_i = 13.321(\pm 3.847) - 0.276(\pm 0.104)\cdot nO_i + 25.709(\pm 13.137)\cdot PW4_i - \\ -14.101(\pm 3.400)\cdot BIC5_i - 6.953(\pm 1.345)\cdot MATS5e_i \\ (n = 79;\ r = 0.870;\ s = 0.681;\ F = 57.61) \quad (10)$$

$$A_i = 7.699(\pm 4.144) + 37.502(\pm 12.301)\cdot PW4_i - 12.083(\pm 3.348)\cdot BIC5_i - \\ - 5.717(\pm 1.337)\cdot MATS5e_i + 0.315(\pm 0.087)\cdot \log P_i \\ (n = 79;\ r = 0.880;\ s = 0.656;\ F = 63.53) \quad (11)$$

$$A_i = 31.979(\pm 4.667) - 26.179(\pm 4.249)\cdot ATS6e_i - \\ - 9.332(\pm 1.015)\cdot MATS5e_i + 0.132(\pm 0.093)\cdot \log P_i \\ (n = 79;\ r = 0.888;\ s = 0.631;\ F = 93.31) \quad (12)$$

Multi-parametric equations with $r > 0.850$, see equation (10), were obtained while using following parameter types: parameters that count the number of oxygen atoms in the molecule (*nO*), the Randić shape index (*PW4*), parameters that quantify structural or binding symmetry (*BIC*), auto-correlation parameters (*ATS, MATS*), especially normalized to electronegativities. A slight increase of the correlation coefficient is noticed if the transport parameter log*P* is introduced in the multi-linear correlations as shown by equations (11) and (12).

3.4. Discussion

To have a more complete image of the present analysis, it is worth to mention another extensive study upon HEPT derivatives with structure (b) in Figure 2; this study was performed by forming a series of 79 compounds starting from ligands used in previous linear models deriving and also from congeners reported by other sources [27-29]. The considered biological activity was that considered for the series in Table 1. Under these conditions, the best obtained QSAR model is [18]:

$$\log(1/C) = 35.99(\pm 2{,}41) + 0.48(\pm 0{,}12)\Sigma\pi(R+Y) + 162.5(\pm 17.6)\,{}^{1}\chi^{N}(Y) - \\ - 239.8(\pm 25.1)[{}^{1}\chi^{N}(Y)]^2 - 0.85(\pm 0{,}19)B1(R,3) + 1.52(\pm 0.28)E_S(R,2) + \\ + 52.06(\pm 7{,}54)\,{}^{4}\chi_P^{N} + 0.78(\pm 0.34)MgVol - 2.22(\pm 0{,}28)I(Z) - \\ - 0.56(\pm 0{,}21)\,{}^{0}\Delta_{\chi}(Z) \\ (n = 79;\ r^2 = 0.949;\ s = 0.44;\ q^2 = 0.745) \quad (13)$$

In equation (13), a series of topological indices [19] derived from the graph theory was used, besides the π-Hansch hydrophobic substituent constant, the E_S – Taft steric constant and the B1 (STERIMOL) parameter which measures the minimal thickness of a substituent. So, ${}^{m}\chi^{N}$ is the molecular connectivity index of order *m* (${}^{m}\chi$) [19] divided through the number of atoms (N) implied in its calculation. The aim of dividing ${}^{m}\chi$ through N was to minimize the effect of the substituent's size in order that the new index ${}^{m}\chi^{N}$ to encode, relatively, in an improved way the substituent's offset degree. Thus, the presence of ${}^{m}\chi^{N}$ in equation (13) shows that biological activity is strongly related to Y substituents structural variations. ${}^{4}\chi_P^{N}$ is also a connectivity index which expresses the balanced number of all fragments (or graphs) with four interfacing bonds (edges) (meaning ways of length 4) [19]. MgVol is the McGowan volume which is easy to

calculate for any solved substance if its molecular structure is known [20]. The direct dependence between biological activity and above described topological parameters points out the importance of the size and of some aspects concerning molecular shape over the anti-HIV potential in the case of the analyzed HEPT ligands. Finally, ${}^{0}\Delta_{\chi}(Z)$, differential molecular connectivity index of order zero, was developed as an electronic parameter based on the graphs theory [21] and the variable indicator I_Z shows the presence ($I_Z = 1$) or the absence ($I_Z = 0$) of a saturated hexacycle in the structure of the Z radical. The inverse dependence of bioactivity upon these two structural parameters suggests a detrimental effect derived from the electronic character of Z substituents as measured by the topological index ${}^{0}\Delta_{\chi}$.

Using the parameters from model (13), Luco and Ferretti [18] performed a QSAR analysis by means of the PLS method (partial least-squares). A significant PLS model was obtained, with 3 components, described by following statistical parameters: $r^2 = 0{,}889$; $r_{CV} = 0{,}927$, $s = 0{,}44$ and $F = 202{,}3$ where r is the correlation coefficient, F represents the Fisher test and r_{CV} is the cross-validation coefficient which describes the predictive capacity of the model [30].

The predictive value of models (13) and PLS is quite good although they don't allow to obtain conclusive data regarding the way of interaction between ligand and receptor. The attempt to reduce the number of structural parameters from 9 to 4 led to correlation equations (10) and (11).

A significant collection of descriptors from different classes (constitutional, topologic, geometric, functional groups, molecular properties, etc.) were calculated for the 79 structures of minimal energy. However, the absence of statistically significant multi-linear correlations is explained by the powerful inter-correlations between topologic parameters, obtained through various changes in the adjacency and distance matrices. The most important are the descriptors defining structural information (proximity symmetry) and *2D* auto-correlation parameters. Linear mono-parametric mathematical models showed correlation coefficients closed to 0.7 for molecular shape and symmetry parameters (*BIC*, *CIC* and *PW4*) and above 0.7 using Moran auto-correlation parameters normalized to electronegativities or to van der Waals atomic volumes. Bioactive molecules are less symmetrical shapes (as it results from the linear mathematical models which use the complementary parameters *BIC* and *CIC*). Electronegativities on different atoms in the molecule are also important; the auto-correlation parameter normalized to these electronegativities led to the best mono-parametric mathematical models.

The molecules transport to the receptor site seems to be very important in the studied series. This fact is proven by the statistically significant mono-parametric model using the hydrophobicity parameter log*P* ($r = 0.67$). Biologic activity increases significantly with the increase of the logarithm of the octanol/water partition coefficient (the activity increases by almost one logarithmical unit when this parameter increases by one unit).

Since most HIV drugs works by inhibiting the DNA polymerase activity of reverse transcriptase or the virus's protease or integrase enzymes, the reverse transcriptase's RNase H (RNA cleaving) activity has to be an important direction to detect the optimal drug target. Nevertheless, among the other direction of studying the DNA polymerase (DNA-synthesizing) activity, give the important hints on how the HIV's life cycle can be broken in both its activities. However, such cooperative studies are now in progress [31] and their associate QSAR analysis will be reported in the foregoing papers.

4. Conclusions

Nowadays research in the field of discovering, developing and promoting new anti-HIV pharmacological agents is based on identifying the essential critical stages of HIV-1 replication. The inverse enzymatic transcription (by RT) of genomic RNA into double strained DNA which migrates towards the host cell's nucleus being then integrated in the host cell's chromosomes (by the viral encoded integrase) was chosen as target for the development of effective anti-HIV agents which act as RT inhibitors. The HEPT (1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine) congeners express *in vitro* anti-HIV-1 activity and no effect against other retroviruses including HIV-2. HEPT antiviral activity was experimentally assessed as EC_{50} – the effective concentration

of the compound in order to obtain a 50% protection of NT-4 cells against the HIV-1 cytopathic effect.

Qualitative chemical structure – biological activity (SAR) studies demonstrated that the presence of methyl group in position 5, the cyclic structure in position 6 and the HN-3 biatomic constellation are indispensable structural requirements for the anti-HIV activity. Structural analysis was extended and the effect of substitution in position C-5 and in the aromatic nucleus bound by an oxygen or sulphur bridge at C-6 was studied.

This work shows an extensive study performed by means of molecular modeling upon a series of 79 HEPT [27(a)] ligands with anti-HIV activity [27(b)]. Molecular modeling methods (molecular mechanics and conformational analysis) and QSAR methods based on correlation analysis and multiple linear regression use a large number of molecular descriptors calculated with the HyperChem and DRAGON programme packages. They allowed obtaining essential data regarding the imposed structural requirements at molecular level [32, 33] in order to improve the anti-HIV potentiality of the studied compounds. Molecular modeling performed upon the 79 bioactive compounds indicated that there are three important zones in order to promote biologic activity: (i) R_1 substitute zone of the primary common pirymidin-2,4(1H,3H)-dione structure (corresponding to position 5 on the primary structure), with substitutes with relatively small volumes (alkyl or aryl-alkyl); (ii) R_2 substitute zone of the primary common structure (corresponding to position 6 on the primary structure) with un-substituted or substituted phenyl-thyo rests; (iii) R_3 substitute zone of the primary common structure (corresponding to position N^1 on the primary structure) with the highest structural variability.

The most stable conformations obtained through conformational analysis showed an orientation towards the R_3 of the R_2 radicals; voluminous substitutes from the R_3 zone (phenyl, chlorine-phenyl, tolyl, cyclohexyl) indicated an orientation whether towards the R_2 zone, in parallel to phenyl nucleus from this zone, or towards the primary structure in a pseudo-parallel orientation, meaning that there is the possibility of van der Waals interactions between these more or less aromatic rests.

Therefore, obtained data by adequate designed QSAR studies allow observing aspects and essential structural characteristics to have an increased biological activity, suggesting certain structural requirements for an increased anti-HIV potential. Our results open very interesting perspectives regarding HEPT derivatives with anti-HIV activity.

Acknowledgement

MVP kindly appreciates the financial assistance of the Romanian National Council of Scientific Research in Universities – CNCSIS through the Grant AT/54/2006-2007.

References

[1] R. R. Redfield, W. A. Blattner, R. C. Gallo, in Textbook of AIDS Medicine, ed. by Merigan Jr. T. C., Bartlett J. G. and Bolognesi D., Williams & Wilkins (1999).
[2] A. C. Carne, Brit. Med. J. **294**, 1399 (1987).
[3] M. Smith, M. Spitle, Brit. Med. J. **294**, 1274 (1987).
[4] N. Suceveanu, R. Nuțiu, C. D. Duda-Seiman, D. Ardelean, A. Ardelean Compuşi pirimidinici - potențiale medicamente anti-HIV, „Vasile Goldiş" Univ. Press, Arad (2004).
[5] F. Vézinet-Brun, Ann. Pédiatr. **34**, 560 (1987).
[6] R. C. Gallo, S.Z. Salahuddin, M. Popovic, G. E. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Readfield, J. Oleske, B. Safai, G. White, P. Foster, P. Markham, Science **224**, 500 (1984).
[7] P. R. Clapham, R. A. Weiss, In Textbook of AIDS Medicine, ed. by Merigan Jr. T. C., Bartlett J.G. and Bolognesi D., Williams & Wilkins (1999).
[8] Viral Targets. http://www.niaid.nih.gov/daids/dtpdb/VIRTARG.asp (2007).
[9] HIV life cycle. http://www.niaid.nih.gov/daids/dtpdb/virpage1.asp. (2007); modified from M. Nasr, J. Cradock, M. Johnston Drug News & Perspectives **6**, 338 (1993).
[10] D. Klatzmann, E. Champagne, S. Chamaret, J. Gruest, D. Guetard, T. Hercend, J. C. Gluckman, L. Montagnier Nature **312**, 767 (1984).

[11] Rossio J. L., Bess Jr. L., Henderson L. E., Cresswell P. and Arthur L.O. AIDS Res. Hum. Retroviruses **11**, 1433 (1995).
[12] R. Garg, S. P. Gupta, H. Gao, M. S. Babu, A. K. Debnath, C. Hansch, Chem. Rev. **99**, 3525 (1999).
[13] H. Tanaka, H. Takashima, M. Ubasawa, K. Sekiya, I. Nitta, M. Baba, S. Shigeta, S. T. Walker, E. De Clercq, T. Miyasaka, J. Med. Chem. **35**, 337 (1992).
[14] H. Tanaka, H. Takashima, M. Ubasawa, K. Sekiya, I. Nitta, M. Baba, S. Shigeta, S. T. Walker, E. De Clercq, T. Miyasaka, J. Med. Chem. **35**, 4713 (1992).
[15] H. Tanaka, M. Baba, H. Hayakawa, T. Sakamaki, T. Miyasaka, M. Ubasawa, H. Takashima, K. Sekiya, I. Nitta, S. Shigeta, R.T. Walker, J. Balzarini, E. De Clercq, J. Med. Chem. **34**, 349 (1991).
[16] R. Pontkis, R. Benhida, A. M. Aubertin, D.S. Grierson, C. Monneret, J. Med. Chem. **40**, 1845 (1997).
[17] D. K. Kim, J. Gam, Y. W. Kim, J. Lim, H. T. Kim, K. H. Kim, J. Med. Chem. **40**, 2363 (1997).
[18] J. M. Luco, F. H. Ferretti, J. Chem. Inf. Comput. Sci. **37**, 392 (1997).
[19] L. B. Kier, L. H. Hall, chapter 3, in Molecular Connectivity in Chemistry and Drug Research, ed. by Bawden D., Research Studies Press Ltd. Letchworth, Hertfordshire, (1986).
[20] M. H. Abraham, Chem. Soc. Rev. **22**, 73 (1993).
[21] L. B. Kier, L. H. Hall, Quant. Struct.-Act. Relat. **10**, 134 (1991).
[22] L. P. Hammett, J. Am. Chem. Soc. **59**, 96 (1937).
[23] N. B. Chapman, J. Shorter, Correlation Analysis in Chemistry, Plenum, New York (1978).
[24] C. Hansch, A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley-Interscience, New York (1979).
[25] HyperChem 7.01 Programme Package, Hypercube Inc. (2002).
[26] DRAGON 2.1 Programme, R. Todeschini, V. Consonni, M. Pavan, Milano Chemometrics & QSAR Group (2002).
[27] (a) T. Miyasaka, H. Tanaka, M. Baba, H. Hayakawa, R. T. Walker, J. Balzarini, E. De Clercq, J. Med. Chem. **32**, 2507 (1989); (b) M. Baba, H. Tanaka, E. De Clercq, R. Pauwels, J. Balzarini, D. Schols, H. Nakashima, C. F. Perno, R. T. Walker, T. Miyasaka, Biochem. Biophys. Res. Commun. **165**, 1375 (1989).
[28] H. Tanaka, M. Baba, M. Ubasawa, H. Takashima, K. Sekiya, I. Nitta, S. Shigeta, R. T. Walker, E. De Clercq, T. Miyasaka, J. Med. Chem. **34**, 1394 (1991).
[29] H. Tanaka, M. Baba, S. Saito, T. Miyasaka, H. Takashima, K. Sekiya, M. Ubasawa, I. Nitta, R. T. Walker, H. Nakashima, E. De Clercq, J. Med. Chem. **34**, 1508 (1991).
[30] D. B. Kireev, J. R. Chrétien, D.S. Grierson, C. Monneret, J. Med. Chem. **40**, 4257 (1997).
[31] D. M. Himmel, S. G. Sarafianos, S. Dharmasena, M. M. Hossain, K. McCoy-Simandle, T. Ilina, A. D. Clark Jr, J. L. Knight, J. G. Julias, P. K. Clark, K. Krogh-Jespersen, R. M. Levy, S.H. Hughes, M.A. Parniak, E. Arnold, ACS Chem. Biol. **1**, 702 (2006).
[32] D. Ciubotariu, M. Medeleanu, V. Vlaia, T. Olariu, C. Ciubotariu, D. Dragoş, C. Seiman, Molecules **9**, 1053 (2004).
[33] C. Duda-Seiman, D. Duda-Seiman, D. Dragoş, M. Medeleanu, V. Careja, M. V. Putz, A.-M. Lacrămă, A. Chiriac, R. Nuțiu, D. Ciubotariu, Int. J. Mol. Sci. **7**, 537 (2006).