Eur J Med Chem (1996) 31, 701–712

Molecular simulation of 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracils with anti-HIV-1 activity

PP Mager[1]*, E De Clercq[2], H Takashima[3], M Ubasawa[3], K Sekiya[3], M Baba[4], H Walther[1]

[1]*Research Group of Pharmacochemistry, Institute of Pharmacology and Toxicology of the University, Härtelstraße 16–18, 04107 Leipzig, Saxony, Germany;*
[2]*Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium;*
[3]*Research and Development Division, Yokohama Research Center, Mitsubishi Chemical Corporation, Yokohama 227;*
[4]*Division of Human Retroviruses, Center for Chronic Viral Diseases, Faculty of Medicine, Kagoshima University, Kagoshima 890, Japan*

(Received 8 November 1995; accepted 13 March 1996)

Summary — The dioxypyrimidine ring of 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracils is an extended 'partial π system' with ring distortion. PM3-MM+ geometry optimization suggested a lipophilic 'butterfly-like' orientation which was also found in other non-nucleoside inhibitors that interact with the HIV-1 reverse transcriptase. Multivariate QSAR has shown that discrimination between the antiviral response and undesired cytotoxicity is possible. Related to the C-6 thiophenyl ring substituents and to modifications at the C-5 position, the antiviral response depends on hydrogen-bonding forces, steric parameters, and electronic properties. The cytotoxicity depends on the lipophilicity and steric parameters.

1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine / anti-HIV-1 activity / non-nucleoside HIV-1 reverse transcriptase inhibitor / QSAR / molecular modelling / non-least-squares (NLS) regression / cluster analysis

Introduction

A major point of attack to treat chemically acquired immunodeficiency syndrome (AIDS) is the inhibition of human immunodeficiency virus type 1 (HIV-1), the causative agent. Among the points of a chemotherapeutic attack, viral reverse transcriptase (RT) plays a key role. This is a bifunctional enzyme containing a DNA polymerase activity which can copy either a DNA or RNA template, and a ribonuclease H (RNase H) activity. These two activities cooperate to convert the single-stranded genomic RNA of HIV-1 into the double-stranded DNA which subsequently becomes integrated into host cell chromosomes.

Inhibitors of RT may be discriminated into 2',3'-dideoxynucleoside (ddN) inhibitors [1] and non-nucleoside inhibitors [2–4]. Nucleoside-type inhibitors act as competitive antagonists at the substrate-binding site of the RT, and initial phosphorylation is a crucial step in the intracellular metabolisms of ddNs. Therefore, ddNs with low affinity to cellular nucleoside kinases are relatively ineffective in HIV-1 therapy. Non-nucleoside-type inhibitors interact non-competitively with the allosteric site (a hydrophobic pocket in the RT-DNA complex close to the active site catalytic centre). The non-nucleoside agents are taken up into cells without phosphorylation by cellular nucleoside kinases. Prolonged therapy with commercially available drugs often leads to the emergence of resistant HIV-1 strains. The mutability of the virus prompted clinical studies with combination therapies to replace antiviral monotherapy with nucleoside and non-nucleoside RT inhibitors, and antagonists that influence other sites of interactions such as the HIV-1 protease [5, 6]. On the other hand, the mutability discouraged further molecular optimization of the lead structures, in particular of non-nucleoside inhibitors.

Only recently, a high resolution crystal structure of intact unliganded HIV-1 RT and of the enzyme

*E-mail: ppm@pharma.phto.uni-leipzig.de

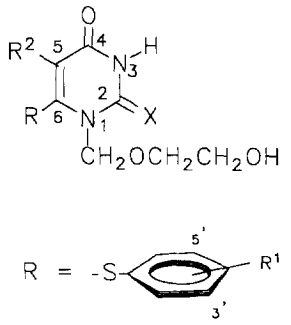


Fig 1. Basic structure of 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracils. The substituents R^1 and R^2 are given in table I.

cocrystallized with inhibitors became available [7] (atomic coordinates have been deposited with the Brookhaven Protein Data Bank). This makes HIV-1 RT inhibitors an appropriate class of molecules to investigate drug–receptor interactions with molecular modelling and structure–activity relationship analysis [7–18].

In this study, 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracil derivatives [4] were investigated. The aim was to find a common pharmacophore in this series and in other non-nucleoside HIV-1 RT inhibitors, and to delineate the antiviral and cytotoxic action by multivariate quantitative structure–activity relationship (QSAR) models.

Materials and methods

Chemistry and physicochemical descriptors

The synthetic routes of 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracil derivatives were described previously [4], together with the data that verify the structures (fig 1). The carbohydrate moieties of the nucleosides are replaced with an acyclic side chain with an intact free hydroxy function; the same side chain is also found in acyclovir.

The linear free energy related (LFER) and extrathermodynamic parameters of the substituents (fig 1) were taken from the literature [20–23]. Other parameters (distribution coefficients, pK_a values, molecular surface areas and volumes, dipole moments, Verloop's steric parameters, polarizability, molar refractivity and hydration energy) were estimated by computer-aided approaches using geometry-optimized structures or simplified molecular identification and line entry system notation [24–31]. The following descriptors were of particular interest: with respect to R^1 substituents (see fig 1), the position-dependent (*ortho*, *meta*, *para*) lipophilic substituent constant π_N (octanol/water) of Norrington [22], the hydrogen-bonding parameter pK_H derived from the equilibrium constant of hydrogen bonding between substituted benzenes and phenol as hydrogen donor [24] (in CCl_4, at 25 °C); and the molar refractivity MR which was scaled (MR/10). The reason for the interest in the latter is that MR/10 agrees approximately with the lipophilic substituent constant of apolar substituents [32, 33]. Furthermore, the Sterimol parameters L (length) and $B1$ (minimum width) were determined [30]. To parameterize the substituents in the R^2 position (fig 1), the passage of the frontier electron flow related to hydrogen (*Re*) was determined (Esaki [48] listed these electronic substituent constants). Also, the steric force-field substituent constant L_f was used [23]. A parameterization of substituent X of figure 1 and table I (carbonyl or thiocarbonyl group) was not made because the sample sizes of the resulting two classes are too different (leading to a danger of cluster correlation).

Molecular modelling and simulation

Different molecular mechanics (MM) methods [34–37] are used to compare the conformations of key molecules. The MM+ force field is an improved MM2 force field. It was chosen in this report as the routine method for the complete set of molecules, because it was designed for small organic molecules and then expanded to encompass amino acids, peptides, heterocyclics, and nucleic acids. The whole procedure was repeated with Coulomb's law functions using a quantum chemistry interface. The minima of the molecular electrostatic potentials (MEPs) were obtained by the semiempirical PM3 approximation [38]. The resulting geometry-dependent partial atomic charges, electrostatic energies and dipole moments [39] were compared with those determined by other semiempirical approaches, and the connectivity-based iterative partial equalization of orbital electronegativity (Gasteiger method) which does not depend on a particular geometry optimization [40].

Correlation-gradient geometry optimization was achieved by the following steps [41]. The structures were refined using a conjugate gradient minimizer (Fletcher–Reeves modification of the Polak–Ribière method). Convergence was obtained when the gradient root mean square (RMS) was < 0.01 kcal/Å·mol. The conformations were initially energy-minimized using force field without an electrostatic term. If not otherwise stated, we used the cooling-heating-cooling schedule [41]. After a short mechanics run (at 0 K), a fully optimized lowest-energy structure was obtained. We continued with short molecular dynamics (MD) runs (temperatures of res-

pectively 373, then 200, 100, 50, and 25 K) to overcome the energy barriers, followed again by the usual MM minimization (0 K) until the procedure converged to a minimum-energy structure. After including the partial charges, the resulting geometry was again optimized to include the molecular electrostatic potential and electrostatic energies.

Pharmacology

HIV-1 virus (HTLV-III_B strain) was used in the anti-HIV assays. Virus stocks were titrated in MT-4 human T-lymphoid cells and expressed as 50% cell culture infective dose. The assay was based on the inhibition of virus-induced cytopathic effect in MT-4 cells. After a 4-day incubation at 37 °C, the number of viable cells was determined by the 3-(4,5-dimethylthiazolyl-2-yl)-2,5-diphenyltetrazolium bromide (MTT) method [42]. The effective concentration (EC_{50}, μM) of a compound required to achieve 50% protection of MT-4 cells against the cytopathic effect of HIV-1 was determined. The cytotoxicity of the compounds was assessed in parallel with the antiviral response; this was based on the viability of mock-infected MT-4 cells as monitored by the MTT method [43]. The cytotoxic concentration (CC_{50}, μM) of a compound required to reduce the viability of mock-infected MT cells was determined. The selectivity index (SI) is defined by the ratio CC_{50}/EC_{50}. The numerical data were taken from the literature [4]. Not all compounds of this series could be used due to the missing, 'larger-than', or 'less-than' scores of the pharmacological affinities.

The parameters $Y_1 = \ln EC_{50}$, $Y_2 = \ln CC_{50}$, and $Y_3 = \ln SI$ were used in subsequent calculations to obtain normally distributed and variance-stabilized scores.

Mathematical and statistical analysis

The underlying statistical approach is based on the MASCA model [19, 44]. In brief, the model starts from a multivariate design constructed according to the rules of decision theory, collects the data to get summary statistics, tests the hypotheses using multivariate and simultaneous statistical inference, applies diagnostic statistics for examining design and model validity, and uses the Bayes approach for subsequent procedures. The matrix computations of the MASCA software are very stable because suitable software modules are used (strong reduction of the number of cumulative rounding errors and of annulments of near-zero scores [45, 46]; different results from regression equations estimated by commercially available packages are mainly due to the varying quality of software modules applied to matrix computations [47]).

Results and discussion

Pharmacological–toxicological parameters

It is remarkable that the compounds (fig 1) do not inhibit other retroviral, bacterial, or mammalian polymerases as well as HIV-2 RT [11]. They are potent inhibitors of HIV-1 reverse transcriptase [4, 12–13]. It was found that non-nucleoside HIV-1 RT inhibitors such as nevirapine and TIBO (see fig 6) bind specifically to an allosteric site of HIV-1 RT [12, 13]. They are unlikely to be phosphorylated by cellular nucleoside kinases (such as thymidine kinase [4]).

The original data from the pharmacological–toxicological assay are collected in table I. The linear correlation coefficients (table II) indicate that there is no simple relationship between antiviral activity and cytotoxicity. However, the partial correlation coefficients are highly significant (table II).

A nonlinear second-order model is used to fit the isobiological contours [48] of the logarithms of the selectivity index. Using an ordinary least-squares regression fit, we get:

$$Y_3 = -0.216 - 0.990Y_1 + 1.106Y_2 - 0.000758Y_1^2 - 0.0117Y_2^2 - 0.00184Y_1Y_2$$

For example, assume the target is to get $Y_3 = \ln SI \approx 4.7$. Then, $Y_1 = \ln EC_{50}$ should be within the range –3 to –2 while $Y_2 = \ln CC_{50}$ should be within the range 1.8–2.7 (fig 2).

Employing nearest-neighbour cluster analysis to the unstandardized scores of Y_1 and Y_2 leads to the result that there are two subgroups within the given series of compounds (fig 3). The first class (B) includes compounds where the antiviral action increases with toxicity (or ν•ν). It contains compounds where a modification at the C-5 position is the predominant factor. The second class (A) contains compounds where an increasing antiviral response is not closely related to increasing toxicity. It includes compounds where the C-5 methyl group is replaced by either an ethyl or isopropyl group.

Molecular modelling of doxypyrimidines

The dioxypyrimidines uracil (U) and thymine (T) are very weak acids (U: $pK_1 = 0.5$, $pK_2 = 9.45$, $pK_3 > 13$; T: $pK_1 \approx 0$, $pK_2 = 9.94$, $pK_3 > 13$). Absorption spectra indicated that the keto form of the lactam–lactim equilibrium is dominant [34] (fig 4). The largest energy component involves the out-of-plane motion of the exocyclic oxygens. Substantial distortion of the ring by the N1 mass including pyramidalization of the N1 nitrogen also occurs (fig 5, table III). Due to the electron delocalization effects, the N-3 atom adopts a planar conformation (table III). The ring distortion is

Table I. Inhibition of HIV-1 replication in MT-4 human T-lymphoid cells by 5,6-substituted 1-[(2-hydroxyethoxy)methyl]-uracils.

Compound	*X*	*R¹*	*R²*	EC_{50}	CC_{50}	*SI*
1	O	3-Me	Me	2.6000	420	162
2	O	3-Et	Me	2.7000	181	67
3	O	3-*tert*-Bu	Me	12.0000	75	6.3
4	O	3-CF_3	Me	45.0000	196	4.4
5	O	3-F	Me	3.3000	282	85
6	O	3-Cl	Me	13.0000	210	16
7	O	3-Br	Me	5.7000	141	25
8	O	3-I	Me	10.0000	106	11
9	O	3-NO_2	Me	34.0000	170	5
10	O	3-OH	Me	82.0000	446	5.3
11	O	3,5-diMe	Me	0.2600	243	935
12	O	3,5-diCl	Me	1.3000	130	110
13	S	3,5-diOMe	Me	0.2200	172	782
14	O	3-COOMe	Me	7.9000	221	28
15	O	3-Ac	Me	7.3000	228	35
16	O	3-CN	Me	10.0000	234	23
17	O	H	$CH_2CH{=}CH_2$	2.5000	183	73
18	S	H	Et	0.1100	148	1350
19	S	H	Pr	10.0000	230	23
20	S	H	*iso*-Pr	0.0590	400	6780
21	S	3,5-diMe	Et	0.0078	277	35500
22	S	3,5-diMe	*iso*-Pr	0.0050	52	10200
23	S	3,5-diCl	Et	0.0430	64	1490
24	O	H	Et	0.1200	400	3330
25	O	H	Pr	3.4000	244	72
26	O	H	*iso*-Pr	0.0630	231	3670
27	O	3,5-diMe	Et	0.0130	149	11500
28	O	3,5-diMe	*iso*-Pr	0.0027	128	47400
29	O	3,5-diCl	Et	0.0140	51	3640
HEPT[a]	O	H	Me	7	743	106
HEPT-S[a]	S	H	Me	0.98	123	126
AZT[a]				0.006	7.8	1300

EC_{50} (μM) is the effective concentration required to achieve 50% protection of MT-4 cells against HIV-1; CC_{50} (μM) is the cytotoxic concentration required to reduce the viability of mock-infected MT-4 cells by 50% and SI is the selectivity index [4]. [a]For comparison.

Table II. Linear and partial correlation coefficients of the bioassay parameters.

Coefficient	Y_1	Y_2	Y_3
Simple			
Y_1	1	0.325	–0.982
Y_2		1	–0.141
Y_3			1
Partial			
Y_1	1	0.99874	–0.99995
Y_2		1	0.99862
Y_3			1

enhanced by introduction of a heavy exocyclic substituent in the N1 position. In contrast, the forces maintaining the planarity of the adenine ring and xanthine ring system are fairly strong [49–52].

The bonding distances indicate that the bonds N3–C2, N3–C4, C4–C5, and C5–C6 have at least 50% double-bond character. The double-bond character of the N1–C2 bond is < 50% (table III). Therefore, some aromatic resonance occurs.

The atomic charges of the molecule S2 (see fig 1; X = O, R = H, X^2 = Me) were determined (table IV). Three neglect of differential overlap (NDO) methods were applied: CNDO (complete NDO), INDO (intermediate NDO), and MINDO/3 (modified INDO, version 3); and three neglect of diatomic differential overlap (NDDO) methods were used for comparisons: MNDO, AM1, and PM3. It can be concluded that the

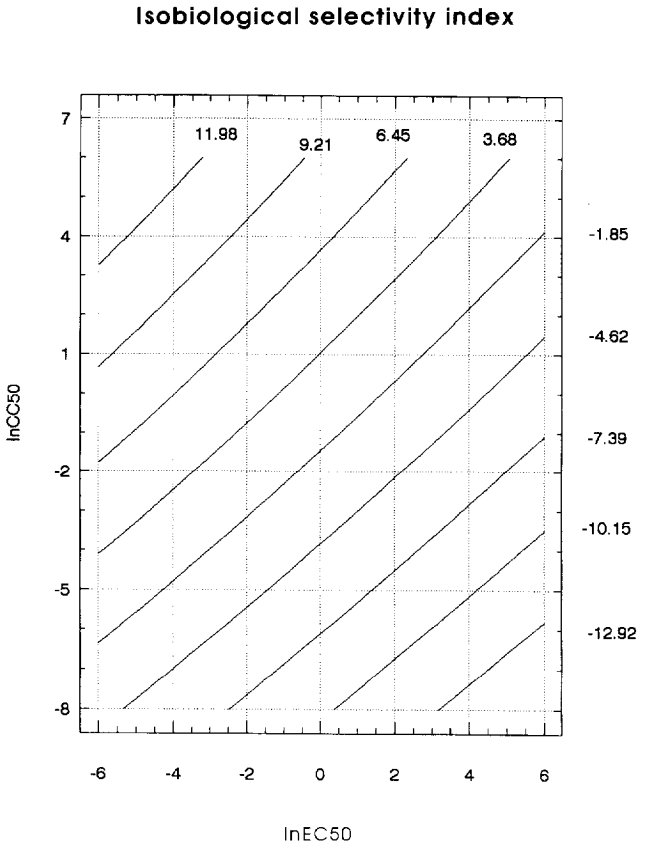


Fig 2. Response-surface analysis using a second order model in order to get scores of isobiological selectivity indices as a function of the natural logarithms of the antiviral response and cytotoxicity.

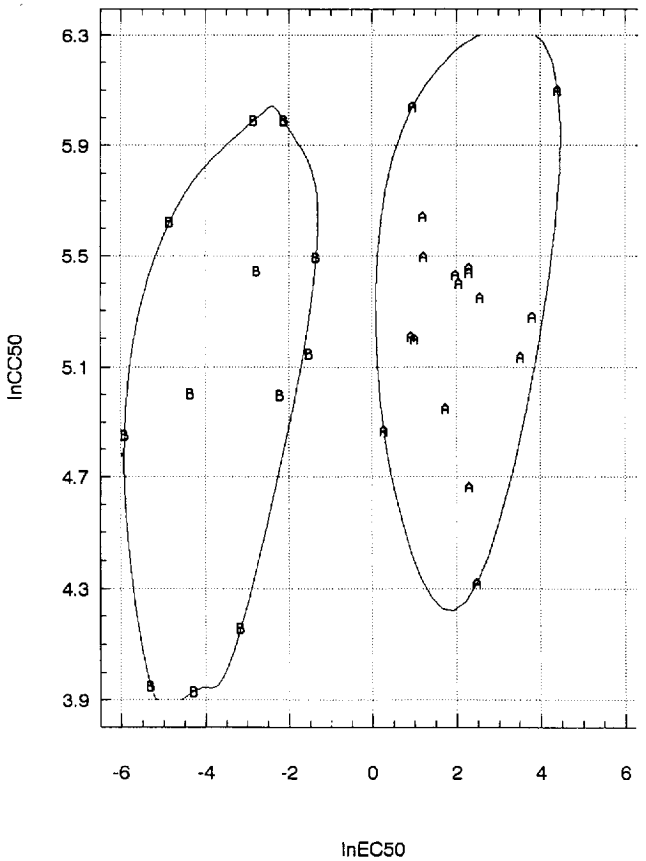


Fig 3. Cluster analysis using parameters of the antiviral response and cytotoxicity. Two compounds were incorrectly classified into group A (compounds **19** and **25**), and two agents were incorrectly categorized into group B (compounds **11** and **13**). The clustering of physicochemical descriptors are shown in table VIII.

charges are transferable. However, only the PM3 approach led to a positive N1 point charge (table IV). Also, the nitrogen of pyrrole is positive (0.31) if a PM3 Hamiltonian is employed. Similar results (ie, PM3 produced a positive charge at N1 while AM1 and MNDO produced negative charges) were observed for the N1 of a 9*H*-thioxanthen-9-one [53], for the N1, N3 and N7 atoms of 8-styrylxanthines [52], and for the N7, N14 and N21 nitrogens of *N*-acetylglycylglycine *N'*-methylamide [54]. Although it is often stated that MNDO is the best semiempirical method to obtain charges, nitrogen is a problem for the MNDO method [55]. The PM3 result becomes acceptable if one anticipates some shift of lone-pair density from N1 into the C6=C5 and C2=O2 double bonds [56, 57].

As an example of substituted nucleoside derivatives, compound **28** is chosen (table I). For comparison, apparently rigid non-nucleoside inhibitors of HIV-1 RT such as nevirapine and TIBO (R82913), and a conformationally more flexible agent, namely Merck L-969,229 (compound **9** of a study of the Merck company, [2]), were chosen (fig 6). The two-dimensional basis exhibited no structural similarities.

Different spatial orientations of compound **28**, as illustrated in figures 7–9, show a hydrophobic 'butterfly-like' configuration, consisting of the previously described 'partial resonance' system of the dioxypyrimidine ring (left side of the figure), a sulfur bridge (bottom), and the π system of the lipophilic phenyl ring (right side). Using the atomic numbering of figure 1, the following conclusions can be drawn. The N(H)–C2(O2) moiety (or a related group) must be envisaged as the electrostatic centre. The circles mark the C4, S, and C3' atoms (C3' of the phenyl ring in the *meta* position which points up and away from the sul-

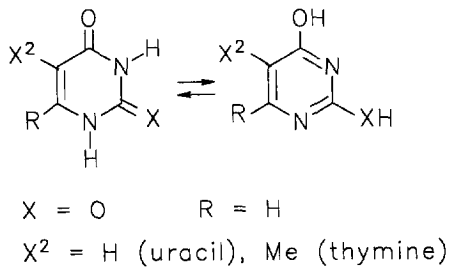


Fig 4. Lactam-lactim equilibrium.

fur). The resulting angle is 94°, and the C4-C3' distance 5.94 Å. This 'butterfly-like' configuration has also been proposed [58, 59] for nevirapine and TIBO (R82913). With respect to nevirapine, the circles denote the C4, N11 and C7 atoms (fig 6). The angle is 96°, and the C4–C7 distance 5.41 Å (fig 10). With respect to TIBO (R82913), the circles denote the Cβ, N6 and C11 atoms (fig 6). The conformationally flexible side chain of TIBO (R82913) was adjusted in the first optimization step by the frozen reaction coordinate (ie, the N6–Cα–Cβ–Cγ dihedral angle). The angle is 142°, and the Cβ-C11 distance 5.80 Å (fig 11).

As sulfur is isosteric with selenium, organometallic nucleoside analogues were synthesized but the results suggest [60] that the compounds are not inhibitors of HIV-1 RT. Thus, we believe that the phenylthio group and a central nitrogen of the 'butterfly-like' configuration are essential for the pharmacodynamic action. In L-969,229 (fig 6), which also assumes a 'butterfly-like' structure (fig 12), the selected atomic points C2, C7 (linker atom) and C11 do not include a central nitrogen like in nevirapine and TIBO (R82913), or a sulfur atom (as in the present series). The angle is 84°, and the C2–C11 distance 5.65 Å. The nature of the 'best' linker group therefore deserves further consideration.

Table III. Evidence of aromatic resonance and ring distortion in dioxypyrimidines.

Atomic code	*Observed distance (Å)*	*'Idealized distances' of pure bonds*		
		Single bond	*Double bond*	*'Mean'*
N1–C2	1.42	1.47	1.27	1.37
N1–C6	1.44	1.47	1.27	1.37
N3–C2	1.35	1.47	1.27	1.37
N3–C4	1.35	1.47	1.27	1.37
C4–C5	1.42	1.54	1.34	1.44
C5–C6	1.42	1.54	1.34	1.44

Atomic code	*Observed angle (°)*
N1–C2–O2	119
C2–N3–C4[a]	123
O2–C2–N3	120
N3–C4–O4	122
N3–C4–C5	117
C4–C5–C6	109
C5–C6–N1	117
C2–N1–C6[a]	112

[a]The mean angle of a trigonal-pyramidal nitrogen is 103–110°; the mean angle of a trigonal-planar nitrogen is 120°.

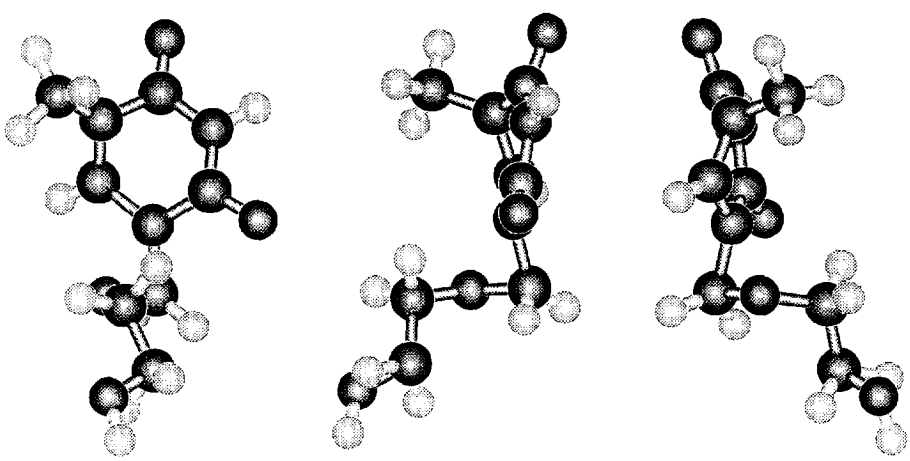


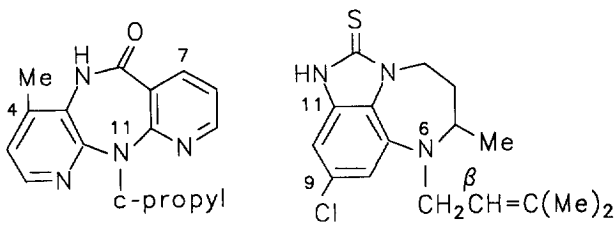
Fig 5. Left: Optimized ball-and-stick conformation of compound S2 (see fig 1; X = O, R = H, X^2 = Me). Centre and right: different views of S2. Distortion of the ring occurs by the N1 mass including pyramidalization of the nitrogen, and by introducing the acyclic side chain.

Table IV. Compound S2 (see fig 1; X = O, R = H, R^2 = Me): comparison of net atomic charges determined by various semi-empirical methods.

Code	*CNDO*	*INDO*	*MINDO/3*	*MNDO*	*AM1*	*PM3*	*Gasteiger*
N1	–0.17	–0.18	–0.18	–0.33	–0.25	0.15	–0.15
C2	0.44	0.52	0.67	0.48	0.41	0.20	0.25
O2	–0.37	–0.42	–0.58	–0.37	–0.35	–0.40	–0.39
N3	–0.25	–0.27	–0.29	–0.39	–0.38	0.03	–0.18
C4	0.37	0.45	0.65	0.41	0.36	0.28	0.19
O4	–0.36	–0.42	–0.59	–0.35	–0.34	–0.38	–0.41
C5	–0.11	–0.14	–0.24	0.28	–0.25	–0.25	0.05
C6	0.16	0.20	0.20	0.19	0.05	–0.08	0.03

Multivariate QSAR analysis

The R^1 and R^2 substituents (see fig 1) were treated independently. Simulation of the R^1 effect occurred by the position-dependent lipophilic substituent constant $\pi_{\bar{N}}$ (octanol/water), the hydrogen-bonding parameter pK_H, and the scaled molar refractivity MR. Furthermore, the Sterimol parameters L (length) and $B1$ (minimum width) were used. To parameterize the substituents in the R^2 position, the electronic substituent constant Re and the steric force-field substituent constant L_f were used. The physicochemical descriptors were defined as follows (table V):



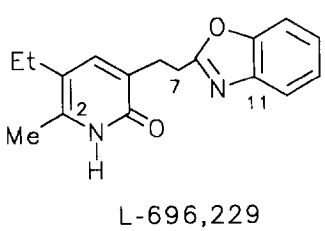


Fig 6. Examples of non-nucleoside HIV-1 RT inhibitors. Numbering indicates the marked points in figures 7–12.

R^1: $X_1 = \pi_{\bar{N}}$, $X_2 = pK_H$, $X_3 = MR/10$, $X_4 = L$, $X_5 = B1$

R^2: $X_6 = Re$, $X_7 = L_f$

The resulting correlation matrix is listed in table VI. Several of the simple correlation coefficients between the two sets **Y** (biological variables) and **X** (physicochemical parameters) are significant.

The selectivity index must be excluded because it depends on Y_1 and Y_2. The squared multivariate correlation coefficient of the two sets

$\mathbf{Y} = (Y_1 \quad Y_2)$

$\mathbf{X} = (X_1 \quad \ldots \quad X_7)$

becomes $R^2_{can} = 0.972$, the maximum-likelihood quantile is 0.694 at the 5% significance level. The squared canonical correlation coefficient is $\theta = 0.908$ (first eigenvalue); the critical quantile (largest-root crite-

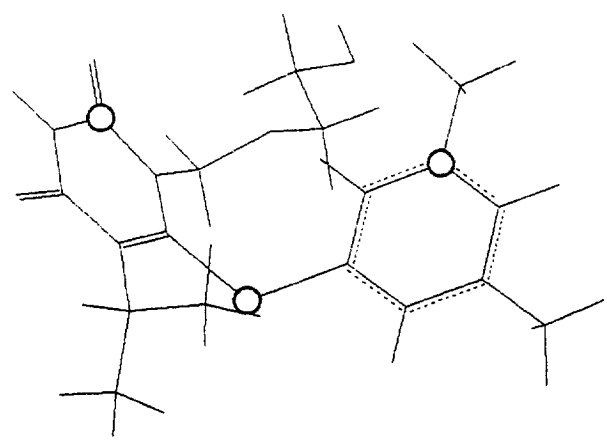


Fig 7. Energy-minimum PM3-MM+ conformation of compounds **28** (fig 1, table I): first screen view. The circles denote the C4, S ('linker') and C3, atoms (see fig 1).

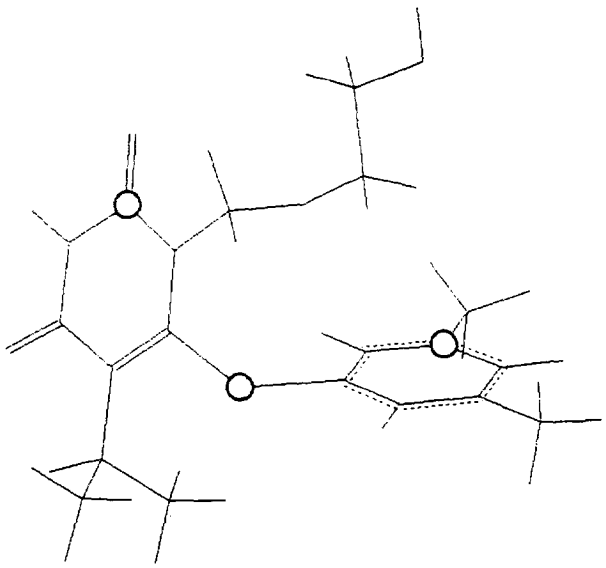


Fig 8. Like figure 7 but another screen view after rotating the molecule.

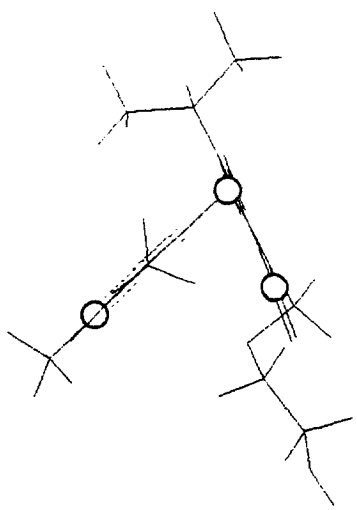


Fig 9. Like figure 7 but another screen view after rotating the molecule. The molecular conformation may be interpreted as a butterfly-like orientation. The conformational flexibility of the molecule allows an adaptation to the receptor site without steric strain.

rion) is 0.548 at the 5% significance level. The canonical correlation plot (first canonical variate) is illustrated in figure 13. Table VII lists the coefficients of the canonical variates. The squared multiple correla-

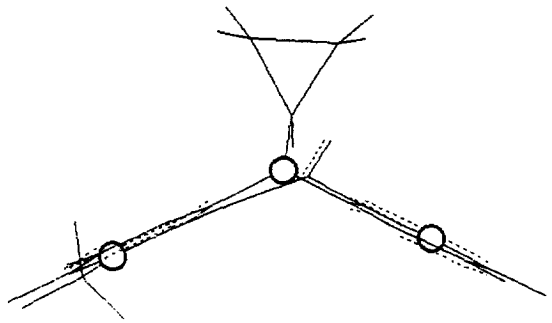


Fig 10. Energy-minimum PM3-MM+ conformation of nevirapine. The circles denote the C4, N11 ('linker') and C7 atoms in figure 6. The butterfly-like orientation is obtained from a conformationally rigid molecule.

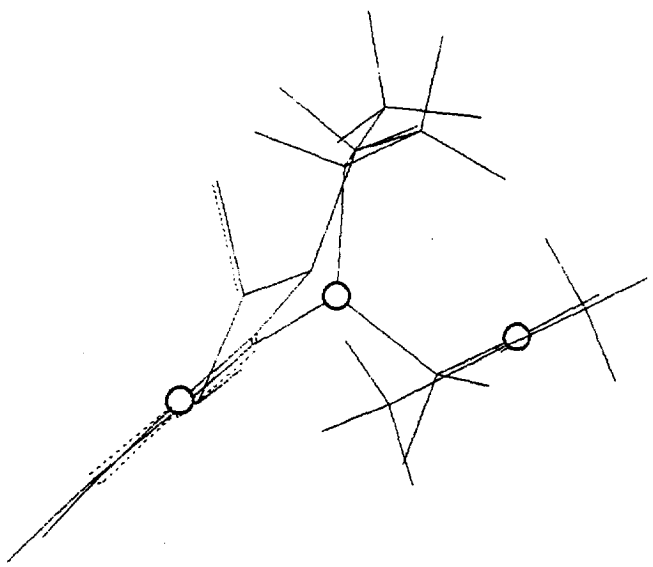


Fig 11. Energy-minimum PM3-MM+ conformation of TIBO (R82913). The circles denote the C11, N6 ('linker') and Cβ atoms in figure 6. There is a more or less great departure from the 'idealized butterfly-like' orientation although the molecule is conformationally rigid.

tion coefficients are $R_1^2 = 0.882$ and $R_2^2 = 0.782$ (tested using the largest-root criterion). Taking these results together, it can be concluded that the multivariate and multiple QSAR equation systems are significantly different from zero.

As expected, there is no collinearity between the variables that were chosen to model the R^1 and R^2 substituent effects. However, within the R^1 set there is a collinearity; in particular the correlation between the two Sterimol parameters (0.97) is remarkable (table VI).

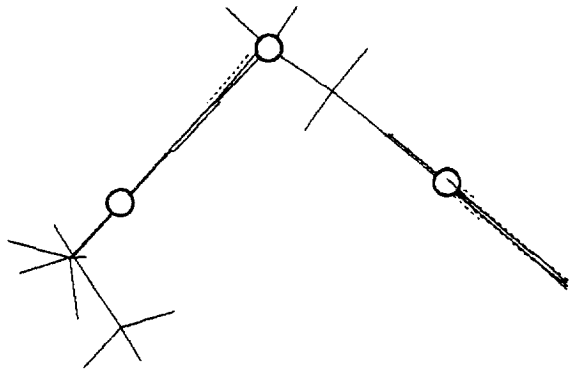


Fig 12. Energy-minimum PM3-MM+ conformation of (Merck L-969,229). The circles denote the C2, C7 ('linker') and C11 atoms in figure 6. The butterfly-like orientation is ideal, and there is a 'compromise' between conformationally flexible and rigid moieties of the molecule.

The internal determination coefficients are significant ($D_1 = 0.78$, $D_2 = 0.63$, $D_3 = 0.72$, $D_4 = 0.96$, $D_5 = 0.97$, $D_6 = 0.65$, $D_7 = 0.59$; the critical quantile is 0.41 at the 5% significance level). Therefore, a multicollinearity exists and ordinary least-squares regression must be replaced by an alternative approach such as non-least-squares (NLS) regression. The diagnosis of the type of multicollinearity, and the selection of the significant regressors and of their components, was based on the described algorithms of the MASCA model [10, 41, 52]. Only the first principal component of the principal-component regression equations (first pseudo regressor) was significant at the 5% level or less.

Retransformation of the resulting principal-component regression equation led to:

$$Y_1 = 1.937 + 1.388\mathrm{p}K_H - 0.811B1 - 0.900Re - 1.238L_f$$

(fig 14). The multiple correlation coefficient $R = 0.91$ is highly significant. The cytotoxicity is modelled by:

$$Y_2 = 5.999 - 0.141\pi_N^{\sim} - 0.208\mathrm{MR}/10 - 0.067L - 0.129B1$$

The multiple correlation coefficient $R = 0.738$ is significantly different from zero at the 5% level or less.

Therefore, the antiviral response depends on position-dependent hydrogen-bonding forces (R^1 substituents), steric parameters (R^1: $B1$, R^2: L_f), and the electronic term Re of the C-5 substituents in the R^2 position.

Hydrogen-bonding between the R^1 region of the compounds and the enzyme are an important part in drug–receptor affinity. The lipophilic pocket region of RT also contains a few amino residues (Lys101, Lys103, Ser105, Asp192, Glu224, Glu138) and backbone atoms that may form hydrogen bonds to the inhibitors.

The negative signs of the regression coefficients associated with the steric effects of the R^1 and R^2 substituents show that less bulky groups enhance the antiviral response. The electronic parameter Re is also significant, ie, substituents with a positive inductive effect in the R^2 position improve the antiviral action.

The lack of significance of a hydrophobic-substituent constant does not automatically imply the insignificance of hydrophobic drug–receptor interactions if other areas of the molecules (not the substituent regions) are involved. It might be expected that the hydrophobic 'butterfly-like' orientation of the non-nucleoside RT inhibitors interacts with the apolar pocket of RT into which the inhibitors must fit [7, 11–17].

The cytotoxicity depends on a position-dependent lipophilicity ($\pi_N^{\sim}$) and steric parameters (MR/10, L, $B1$) of the R^1 substituents. Due to the lack of significance of hydrogen-bonding forces and Re, it appears that the substituent effects are less specific. In addition, with exception of the Sterimol parameter $B1$, the chemical variables of the two QSAR equations differ. Therefore, discrimination between the antiviral response and undesired cytotoxicity is possible by molecular manipulation.

As there was considerable multicollinearity among the physicochemical descriptors (listed in table V), the furthest-neighbour method of cluster analysis was applied to the standardized chemical variables. The first class includes the same members as grouped in class B of the biological-activity profiles (fig 3). The second class, A, includes the same members as grouped in class A of the biological activity profiles (fig 3) if two subgroups (A1 and A2) are assumed (table VIII). Thus, there is sufficient agreement between the results of the two cluster analyses.

Conclusions

The dioxypyrimidine ring of 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracils consists of an extended 'partial π system' with ring distortion mainly due to the N1 atom. Cluster analysis of the biological activity profile and the chosen set of physicochemical descriptors, and multivariate QSAR using non-least-squares regression, have shown that maximization of the antiviral response and minimization of the undesired cytotoxicity is possible. Related to the C-6 thiophenyl ring substituents and modifications at the C-5 position, the antiviral response depends on hydrogen-bonding forces, steric and electronic parameters. The cytotoxicity depends on the lipophilicity and steric parameters.

Table V. Physicochemical descriptors of 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracils.

Compound	*R1 position*					*R2 position*	
	π_N	pK_H	*MR/10*	*L*	*B1*	*Re*	$\mathcal{L}_f$
1	0.50	–0.56	0.57	3.00	1.52	–1.64	0.00
2	0.94	–0.61	1.03	4.11	1.52	–1.64	0.00
3	1.72	–0.62	1.96	4.11	2.60	–1.64	0.00
4	1.49	0.47	0.50	3.30	1.90	–1.64	0.00
5	0.42	–0.07	0.09	2.65	1.35	–1.64	0.00
6	1.04	0.50	0.59	3.52	1.80	–1.64	0.00
7	1.17	–0.53	0.89	3.82	1.95	–1.64	0.00
8	1.47	–0.56	1.40	4.23	2.15	–1.64	0.00
9	0.54	0.74	0.74	3.44	1.70	–1.64	0.00
10	–0.66	–0.19	0.29	2.74	1.35	–1.64	0.00
11	1.00	–1.12	1.14	6.00	3.04	–1.64	0.00
12	2.08	–1.00	1.18	7.04	3.60	–1.64	0.00
13	1.00	–1.12	1.14	6.00	3.04	–1.64	0.00
14	–0.07	1.21	1.29	4.73	1.64	–1.64	0.00
15	–0.07	1.22	1.12	4.06	1.60	–1.64	0.00
16	0.24	0.72	0.63	4.23	1.60	–1.64	0.00
17	0.00	0.00	0.10	2.06	1.00	0.00	0.00
18	0.00	0.00	0.10	2.06	1.00	0.72	0.86
19	0.00	0.00	0.10	2.06	1.00	0.57	0.89
20	0.00	0.00	0.10	2.06	1.00	0.42	2.29
21	1.00	–1.12	1.14	6.00	3.04	0.72	0.86
22	2.08	–1.00	1.18	7.04	3.60	0.42	2.29
23	2.08	–1.00	1.18	7.04	3.60	0.72	0.86
24	0.00	0.00	0.10	2.06	1.00	0.72	0.86
25	0.00	0.00	0.10	2.06	1.00	0.57	0.89
26	0.00	0.00	0.10	2.06	1.00	0.42	2.29
27	1.00	–1.12	1.14	6.00	3.04	0.72	0.86
28	1.00	–1.12	1.14	6.00	3.04	0.42	2.29
29	2.08	–1.00	1.18	7.04	3.60	0.72	0.86

Table VI. Correlation coefficients between the two sets **Y** (biological profile, table I) and **X** (physicochemical profile, table V) used to examine simple linear dependence, and within the variables of **X** to examine collinearity.

Variable	*R1 position*				*R2 position*		
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Y_1	–0.31	0.62	–0.15	–0.49	–0.51	–0.74	–0.74
Y_2	–0.79	0.41	–0.62	–0.64	–0.68	–0.18	–0.15
Y_3	0.16	–0.57	0.03	0.39	0.39	0.73	0.75
X_1	1	–0.61	0.69	0.77	0.85	–0.03	0.01
X_2		1	–0.39	0.60	–0.70	–0.30	–0.26
X_3			1	0.79	0.77	–0.28	–0.19
X_4				1	0.97	0.02	0.05
X_5					1	0.06	–0.09
X_6						1	0.76
X_7							1

Table VII. Canonical QSAR results.

Notation		*First canonical set*	*Second canonical set*
Eigenvalues		0.908	0.647
Standardized eigenvectors:			
Chemical variable	X_1	–0.218	1.130
	X_2	0.176	0.548
	X_3	–0.120	0.488
	X_4	–0.639	–0.569
	X_5	0.375	0.040
	X_6	–0.454	0.210
	X_7	–0.325	–0.340
Biological terms	Y_1	0.820	0.668
	Y_2	0.365	–0.992

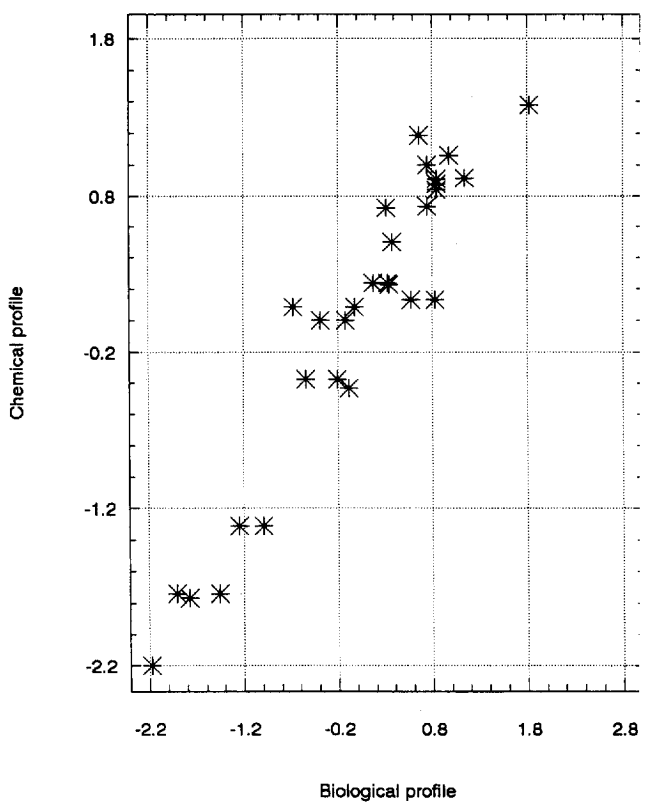


Fig 13. Result of multivariate subset selection: canonical correlation plot of the antiviral and cytotoxic response against the set of physicochemical variables.

Table VIII. Cluster analysis using the physicochemical descriptors.

Class	*Assigned compounds*	*Correct classification*
Class A	**17–29**	100%
Subgroup A1	**17–20, 24–26**	
Subgroup A2	**21–23, 27–29**	
Class B	**1–16**	100%

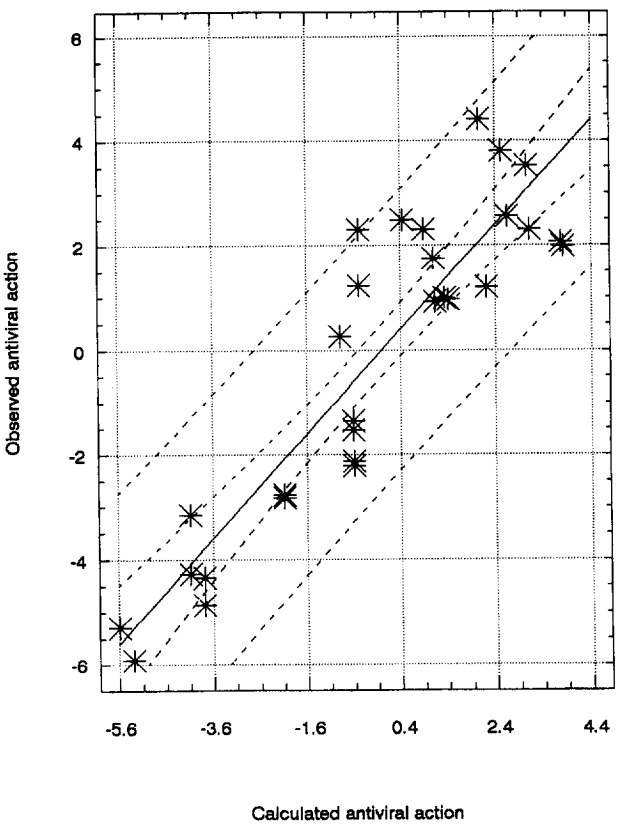


Fig 14. Experimentally obtained antiviral action versus the theoretically calculated antiviral response. The 95% confidence and prediction limits against the regression line are also shown.

The present results are evidence of the potency of multivariate QSAR methodology to discriminate between various biological responses, such as antiviral action and cytotoxicity, within a closely related series of antiviral compounds. Extrapolations outside the given chemical parameter set of a series are difficult,

however. Furthermore, recent results based on a structurally diverse series of antiviral compounds cannot be applied to validate the present model obtained by QSAR studies of 5,6-substituted 1-[(2-hydroxyethoxy)methyl]uracils.

A butterfly-like orientation was found in seemingly structurally distinct classes of non-nucleoside HIV-1 inhibitors. Meanwhile, other classes of non-nucleosides were also studied, and the result was validated (PPM, unpublished). The butterfly-like configuration fits well into the lipophilic pocket of the allosteric area of the unoccupied HIV-1 reverse transcriptase, while the different sites of drug–receptor interactions seem to determine the potency, selectivity and resistance development.

Acknowledgment

We thank the technical assistance of R Wolfram and M Bretschneider, and the referees for providing numerous improvements and suggestions.

References

1 Mitsuya H, Weinhold KJ, Furman PA et al (1985) *Proc Natl Acad Sci USA* 82, 7096–7100
2 De Clercq E (1990) *Med Res Rev* 13, 229–258
3 Hoffman JM, Smith AM, Rooney CS et al (1990) *J Med Chem* 36, 953–966
4 Tanaka H, Takashima H, Ubasawa M et al (1992) *J Med Chem* 35, 337–345
5 Ghosh AK, Thompson WJ, Holloway MK et al (1993) *J Med Chem* 36, 2300–2310
6 Kohl NE, Emini EA, Schleif WA et al (1988) *Proc Natl Acad Sci USA* 85, 4686–4690
7 Ren JS, Esnouf R, Garman E et al (1995) *Nature Struct Biol* 2, 293–302
8 Bailey TR, Diana GD, Kowalczyk PJ et al (1992) *J Med Chem* 35, 4628–4633
9 Wang PH, Keck JG, Lien EJ, Lai MMC (1990) *J Med Chem* 33, 608–614
10 Antonucci T, Warmus JS, Hodges JC, Nickell DG (1995) *Antiviral Chem Chemother* 6, 98–108
11 De Clercq E (1995) *J Med Chem* 38, 2492–2517
12 Mertens A, Zilch H, König B et al (1993) *J Med Chem* 36, 2526–2535
13 Wu JC, Warren TC, Adams J et al (1991) *Biochemistry* 30, 2022–2026
14 Spence RA, Kati WM, Anderson KS, Johnson KA (1995) *Science* 267, 988–993
15 Pauwels R, Andries K, Janssen PAJ, Arnold E (1994) *J Mol Biol* 243, 369–387
16 Ding J, Das K, Tantillo C et al (1995) *Structure* 3, 365–379
17 Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA (1992) *Science* 256, 1783–1790
18 Kireev DB, ChrÇtien JR, Raevsky OA (1995) *Eur J Med Chem* 30, 395–402
19 Mager PP (1991) *Design Statistics in Pharmacochemistry*. Wiley & Sons, New York
20 Hansch C, Leo A (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley & Sons, New York
21 Hansch C (1971) In: *Drug Design* (Ariens EJ, ed) Academic Press, New York, Vol I, 271–342
22 Unger SH (1980) In: *Drug Design* (Ariens EJ, ed), Academic Press, New York, Vol X, 47–119
23 Beckhaus HD (1978) *Angew Chem Ger Ed* 90, 633–635
24 Mager PP, Rothe H (1990) *Pharmazie* 45, 758–764
25 Pacios LF (1994) *Comput Chem* 18, 377–385
26 Wong MW, Wibergg KB, Frisch MJ (1995) *J Comput Chem* 16, 385–394
27 Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) *J Comput Chem* 16, 273–284
28 Klopman G, Fercu D (1994) *J Comput Chem* 15, 1041–1050
29 Dixon SL, Jurs PJ (1993) *J Comput Chem* 14, 1460–1467
30 Verloop A, Hoogenstraaten W, Tipker J (1976) In: *Drug Design* (Ariens EJ, ed), Academic Press, New York, Vol VII, 271–342
31 Meylan WM, Howard PH (1995) *J Pharm Sci* 84, 83–92
32 Mager PP (1978) *Multivariate Chemometrics in QSAR: A Dialogue*, Wiley, New York, 303–305
33 Selassie CD, Fang ZX, Li R, Hansch C, Klein T, Langridge R, Kaufman BT (1986) *J Med Chem* 29, 621–626
34 Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) *J Comput Chem* 7, 230–252
35 Allinger NL, Yan L (1993) *J Am Chem Soc* 115, 11918–11925
36 Tai JC, Yang L, Allinger N L (1993) *J Am Chem Soc* 115, 11906–11917
37 Clark M, Cramer III RD, Van Opdenbosch N (1989) *J Comput Chem* 10, 982–1012
38 Aléman C, Lugue FJ, Orozco M (1993) *J Comput Chem* 7, 799–808
39 Dinur U, Hagler AT (1995) *J Comput Chem* 16, 154–170
40 Gasteiger J, Marsili M (1980) *Tetrahedron* 36, 3219–3288
41 Mager PP (1994) *Med Res Rev* 14, 75–126
42 Pauwels R, De Clercq E, Desmyter J et al (1987) *J Virol Methods* 16, 171–185
43 Pauwels R, Balzarini J, Baba M et al (1988) *J Virol Methods* 20, 309–322
44 Mager, PP (1975) *Drug Res* 25, 1006–1008, 1270–1272, 1355–1356, 1475–1476, 1745, 1864–1865
45 Kagiwada H, Kalaba R, Mease K (1977) In: *Applications of Statistics* (Krishnaiah PR, ed), North-Holland, Amsterdam, 289–299
46 Wilkinson JH (1963) *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, NJ, 28, 62, 79, 138
47 Pöppe C (1993) *Spektrum der Wissenschaft/Spectrum of Science*, Nov 1993, 18–22
48 Esaki T (1980) *J Pharmacol Dyn* 3, 562
49 Mager PP (1994) *Eur J Med Chem* 29, 369–380
50 Mager PP (1995) *J Chemometrics* 9, 211–221
51 Mager PP, Richter M, Reinhardt R (1995) *Eur J Med Chem* 30, 15–25
52 Mager PP, Reinhardt R, Richter M, Walther H, Rockel B (1995) *Drug Design Discovery*, 13, 89–107
53 Horwitz JP, Massova I, Wiese TE, Besler BH, Corbett TH (1994) *J Med Chem* 37, 781–786
54 Böhm HJ (1993) *J Am Chem Soc* 115, 6152–6158
55 Merz KM (1992) *J Comput Chem* 13, 749–767
56 Beck B, Rauhut G, Clark T (1994) *J Comput Chem* 15, 1064–1073
57 Wiberg KB, Rablen PR (1993) *J Comput Chem* 14, 1504–1518
58 Mui PW, Jacober SP, Hargrave KD, Adams J (1992) *J Med Chem* 35, 201–202
59 Schäfer W, Friebe WG, Leinert H et al (1993) *J Med Chem* 36, 726–732
60 Goudgaon NM, Schinazi SF (1991) *J Med Chem* 34, 3305–3309
61 Mager PP (1994) *Med Res Rev* 14, 533–588