Selection of Molecular Descriptors with Artificial Intelligence for the Understanding of HIV-1 Protease Peptidomimetic Inhibitors-activity

S. Sirois[1,2,5,*], C.M. Tsoukas[1], Kuo-Chen Chou[4,5], Dongqing Wei[4,5,6], C. Boucher[1] and G.E. Hatzakis[1,3]

[1]Immune Deficiency Treatment Centre (IDTC), Montreal General Hospital, McGill University Health Centre, 1650 Cedar Avenue, Montréal, Québec, H3G 1A4, Canada, [2]Université du Québec à Montréal (UQAM), Chemistry Department, C.P. 8888 Succursale Centre-Ville, Montréal, Québec, Canada, H3C 3P8, [3]McGill University, Department of Medicine, Divisions of Clinical Immunology & Allergy, Clinical Epidemiology, [4]Tianjin Institute of Bioinformatics and Drug Discovery (TIBDD) and Tianjin Normal University, Tianjin, China 300074, [5]Gordon Life Science Institute, San Diego, CA 92130, USA, [6]Center for Research in Molecular Modeling (CERMM), Chemistry Department, Concordia University, Montréal, Canada

Abstract: Quantitative Structure Activity Relationship (QSAR) techniques are used routinely by computational chemists in drug discovery and development to analyze datasets of compounds. Quantitative numerical methods like Partial Least Squares (PLS) and Artificial Neural Networks (ANN) have been used on QSAR to establish correlations between molecular properties and bioactivity. However, ANN may be advantageous over PLS because it considers the interrelations of the modeled variables. This study focused on the HIV-1 Protease (HIV-1 Pr) inhibitors belonging to the peptidomimetic class of compounds. The main objective was to select molecular descriptors with the best predictive value for antiviral potency (Ki). PLS and ANN were used to predict Ki activity of HIV-1 Pr inhibitors and the results were compared. To address the issue of dimensionality reduction, Genetic Algorithms (GA) were used for variable selection and their performance was compared against that of ANN. Finally, the structure of the optimum ANN achieving the highest Pearson's-R coefficient was determined. On the basis of Pearson's-R, PLS and ANN were compared to determine which exhibits maximum performance. Training and validation of models was performed on 15 random split sets of the master dataset consisted of 231 compounds. For each compound 192 molecular descriptors were considered. The molecular structure and constant of inhibition (Ki) were selected from the NIAID database. Study findings suggested that non-covalent interactions such as hydrophobicity, shape and hydrogen bonding describe well the antiviral activity of the HIV-1 Pr compounds. The significance of lipophilicity and relationship to HIV-1 associated hyperlipidemia and lipodystrophy syndrome warrant further investigation.

Key Words: *In silico* Drug Design, Neural Networks, QSAR, HIV-1 Protease, peptidomimetic inhibitors, Molecular Descriptors.

INTRODUCTION

HIV-1 Pr inhibitors have been the subject of molecular modeling in the field of drug design in the last decade, as investigated [1-6, 75-80] and reviewed [7] by other researchers. A molecule exhibits inhibitor activity only when it is quantitatively determined. A molecule as an entity displays an identifiable property (e.g. inhibition of a target enzyme), only in relation to its measuring experiment. One way to reveal the inhibitory property is to measure the enzyme velocity at a variety of substrate concentration in the presence or absence of a potential inhibitor [8]. Thus, Ki is expressed as the inhibitory constant for inhibitor binding. The competitive molecule, the inhibitor, forms a complex with the target enzyme at the substrate binding site and prevents any further entry of the substrate thus stopping enzyme catalysis of the substrate. In order for a molecule to compete for the same "spot" it must possess features that are structurally and chemically similar to the substrate. For instance, peptidomimetic inhibitors - which mimic the transition state (TS) of the substrate during enzyme catalysis - have been successful as HIV-1 Pr inhibitors in preventing progression to AIDS. Representatives of this class of compounds are Saquinavir [9], Indinavir [10], Ritonavir [11], Lopinavir [12], Amprenavir [13] and Nelfinavir [14]. However, since HIV-1 Pr inhibitors were approved by the US FDA in 1996, beneficial therapeutic effects of HIV-1 Pr inhibitors have been shaded by adverse reactions [15] and resistance [16]. Treatment choices subsequent to initial therapeutic failure require a comprehensive assessment of several contributing factors including ADME/TOX [17], pharmacokinetics [18], drug exposure [19] and resistance [19]. Consideration of these factors is a complicated but necessary step required for the selection of the optimum regimen in HIV-1 treatment. Complexity in medical decision-making varies with the volume and diversity of the pharmacologic and clinical evidence during patient-treatment. Therefore, it is essential to identify the structural molecular factors that are important to the activities of inhibitors; QSAR was employed to address this task. The

*Address correspondence to this author at the Montreal General Hospital, Room A5-145, 1650 Cedar Avenue, Montreal, Quebec, H3G 1A4, Canada; Tel: +1 (514) 934-8035; Fax: +1 (514) 937-1424
E-mail: ssirois@muhc.mcgill.ca

1573-4064/05 $50.00+.00

knowledge gained through this research would assist in: 1) optimizing HIV-1 Pr inhibitors and 2) better understanding how to control the unwanted side effects and resistance patterns resulted from the drug-administration of protease inhibitors.

Intuitively, molecular entities should possess specific features, which classify them into drugs or non-drugs. First and foremost, these features have to be specific to their target. Thus, drugs are subdivided into antiviral, antiretroviral, antibiotics, antineoplastics etc. Secondly, the structure, physical and chemical features of a particular class of compounds should be related to its biological activity. Characterizing a class of molecule to a specific activity is another major challenge in drug identification and optimization [20]. The pharmaceutical industry routinely uses classical approaches based on: 1) High Throughput Screening (HTS) for identifying new class of compounds and 2) Linear Quantitative Structure Activity relationship for optimizing lead candidate. Knowing that development of any new drug costs over $500M [21], and that more than 80% of all new chemical entities (NCE) fail during clinical studies [22] it is imperative that new paradigms be introduced for drug identification and optimization. Information and knowledge from various domains need to be incorporated in every step of drug development.

Traditionally, QSAR statistical methods - like PLS and Principal Component Analysis (PCA) - are used in drug design and optimization. These methods address primarily correlations of an activity with a single point change made at a particular position of a molecule while keeping all the other positions constant. However, change of these points depends on non-linear effects in the overall molecule potency; PLS is limited in describing this non-linearity. Furthermore, PLS is of a reductionism nature and does not take into account interrelations between the various types of descriptors. In a similar manner, PCA, reduces representation from a higher to a lower dimension thus resulting in loss of physical meaning of the descriptors.

So far, a single silver bullet property correlating a specific class of compounds with an activity does not exist. For instance, a candidate molecule is established as an HIV-1 Pr inhibitor because experimentally it attains a threshold Ki value in the range of the nanomolar (nM). However, the structural, chemical and physical characteristics leading to this threshold Ki value have to be defined. Several characteristics can be considered individually as per the classical QSAR approach. However, the synergetic effect of these characteristics is greater than that of their individual contribution. The complexity of these interrelations differentiates a potent inhibitor versus a non-potent one. In this study, a set of characteristics describing peptidomimetic TS-analogues of HIV-1 Pr inhibitors was determined and the activity, adverse reactions and resistance patterns were discussed. ANN and nonlinear iterative PLS (NIPALS) were used to predict the inhibition activity of new molecules belonging to this class.

EXPERIMENTAL DETAILS

This research project was materialized through the following steps:

1) HIV-1 Pr inhibitor compounds were selected on the basis of Ki inhibition constant,

2) The molecular descriptors were calculated for each compound [23],

3) Approaches were developed to:

3a) Select the optimal descriptors (GA-ANN),

3b) Predict and validate the Ki inhibition constant (ANN and PLS).

Selection of HIV-1 Protease Inhibitors

A set of HIV-1 Pr inhibitor complexes with known chemical structures in SDF format [24], binding properties and Ki activities were obtained from the NIAID [25] database. Compounds were selected on the basis that inhibitors should reflect the diversity of the overall peptidomimetics class of compounds. A master dataset was compiled with HIV-1 Pr subset inhibitors including 231 structures for the peptidomimetics class of compounds (Table **1**). The subset was divided into 21 subclasses of HIV-1 Pr inhibitors ranging from allophenylnorstatine to urethanes. Compounds were retained representing a spectrum of Ki values that ranged from 0.001 nM to 2,000,000 nM. Ki values were preferred over IC50 values because the inhibition constant (Ki) can be directly related to the free energy of interaction of the drug [26]. The target of this interaction can be defined through the equations listed below where K_d denotes the equilibrium constant expressing the binding affinity between the receptor *R* and the ligand *L*. The thermodynamics parameters are related to each other where ΔG denotes the free energy of binding, ΔH is the enthalpy and $T\Delta S$ represents the entropy:

$$Ki \quad Kd = \frac{[R][L]}{[RL]} \qquad \Delta G_{bind} = \Delta H - T\Delta S$$

$$\Delta G_{bind}{}^{0} = RT\ln Kd$$

$$Kd = \frac{Kon}{Koff}$$

The Ki values were also transformed into logKi (log base e) - the value range was from -6.8 to 14.5 nM.

Calculation of Molecular Descriptors

Numerical representation of molecules was described with *n*-vectors of numbers called molecular descriptors. In general, a descriptor can be any quantity that describes a molecule. Molecular descriptors were generated from the molecular structure on the basis of 1D, 2D and 3D formulas. The descriptor types used to represent the molecular structure were topological [27], structural [28], physical [29], and chemical [29, 30]. A typical QSAR table is composed of rows representing the molecules and columns representing descriptor values. The respective database with compounds was compiled by:

1 Downloading individual compound-structure from NIAID database in SD file format [24] (see section Selection of HIV-1 Pr inhibitors),

Table 1. List of Peptidomimetic Subclasses of Compound Used for the Current Study

Allophenylnorstatine	Phosphinate
Dihydroxy ethylene; propylene	Phosphinic acid analog
Dihydroxy isostere P1/P1'; P2/P2'; P1/P1'P2/P2'	Pseudophenylalanine hydroxyethylproline
Dimerisation inhibitor	Pyridines
Epoxydes	Reduced amide
Hydroxy ethylamine; ethylene; propylene urethanes; indanes; aminosulfonamide	Reduced PHE(R)PRO or TYR(R)PRO dipeptide isostere
Hydroxy isostere P1/P1'P2/P2'	Statine analog
Isoquinolines	Symmetry-based dihydroxy
Peptide suicide inhibitor	Thiazoles
Phenylpropanoic acid	Urethanes
Phenylstatine	

2 Pre-processing the database by consecutively:

   a Correcting 2D structural representation; adding explicit hydrogen atoms; setting atom ionization to formal charge; deprotonating acids and protonating bases,

   b Removing from the database: water, counter ions (belonging to IA, IIA, and VIIA groups in the chemical table), oxygen ions OH^-, O^{2-}, solvents, salts, redundant compounds,

1 Minimizing the 3D structure,

2 Calculating molecular descriptors.

Genetic Algorithms (GA)

GA algorithms, in general, use a "survival of the fittest" approach to determine if a selection (i.e. potential solution) would make it to the next step. In this study, a genetic algorithm was used to identify "the most dominant" descriptors. An elaborate description of this algorithm and its theoretical foundations is beyond the scope of the present manuscript and could be sought in the literature [31-33].

Partial Least Squares

The nonlinear iterative partial least squares [34] (NIPALS) algorithm was employed to compute PLS regression components. Several variants of the NIPALS exist depending on whether certain vectors are normalized. In the current implementation both X and Y variables were transformed to have a zero average (similarly to the *z-score*). Concisely, i=1,..,n and $A_0=X'Y$, $M_0=X'X$, $C_0=I$ and for a known c:

1. *Compute* the eigenvector q_i of $Ai'Ai$
2. *Compute* the $w_i=C_iA_iq_i$, then *normalize* $w_i=w_i/\|w_i\|$ and *store* the w_i column in the W matrix
3. *Compute* the $p_i=M_iw_i$ and $c_i=w_i'M_iw_i$, then *normalize* $p_i=p_i/c_i$ and *store* the p_i column in the P matrix
4. *Compute* the $q_i=A_i'w_i/c_i$ and *store* the q_i column in the Q matrix
5. *Compute* the $A_{i+1}=A_i-c_ip_iq_i'$ and $M_{i+1}=M_i-c_ip_ip_i'$
6. *Compute* the $C_{i+1}=C_i-w_ip_i'$
7. *Compute* the $T=XW$ and $B=WQ'$ where T denotes the factor score matrix and B the PLS coefficients

Artificial Neural Networks (ANN)

ANN approximation in general is achieved through an initial training session at the end of which internal associations - relating patterns of inputs and outputs - are built. Subsequently, ANN can make a prediction based on new inputs or "unknowns". The study dataset was characterized by non-linearities and the interrelations of included variables were poorly understood. ANN is a well-defined numerical approach for this type of problems. To address the non-linear nature of the study dataset, different structures of ANN were tested i.e. multi layer perceptron (MLP) with 3 or 4 layers, radial basis function (RBF), and generalized regression neural networks (GRNN). The MLP [35], in general, is the most popular network structure. It is dependent on iterative training (slow in cases) but it produces networks that are compact and fast in their execution (following training). MLPs with 3 or 4 layers were tested featuring the hyperbolic tangent as activation function interconnecting these layers. Radial basis function [36] (RBF) networks originate from the regularization theory for solving ill-conditioned problems. They tend to perform slower than MLPs but they train fast. Conversely to the MLPs, the effectiveness of RBF is inversely proportional to the increasing number of input variables. However, inclusion of unnecessary inputs makes them more sensitive which may have an indicative value in unexploited datasets like the study set. Generalized regression neural networks [37] (GRNN) train fast (when N<1000 approximately) and perform satisfactorily, yet they execute slowly. GRNNs augment the pros and cons of RBFs thus contrasting MLPs.

GRNNs use Bayesian techniques to estimate the expected value of an output variable dependent on a given input.

RESULTS

Training and Validating Set Design

A flat dataset (master) was compiled consisting of 231 compounds (rows) with 192 descriptors each (columns). To assess the performance of the ANN a random number generator [38] was employed to randomly partition 15 splits of the master dataset into training (T) and validating (V) subsets in the proportion 186(T):45(V) compounds, respectively. It is noted that the number of split sets as well as the proportion T:V was arbitrarily chosen.

Predict Ki with NIPALS and ANN Models

The 15 training split sets were used, individually, to compute the coefficients for the NIPALS models as well as the weight factors for the ANN models, respectively. These coefficients and weight factors were then used to predict the "unknown" Ki, individually, in the 15 validating split sets for the NIPALS and ANN, respectively. In each of the 15 split sets, training and validating, the predicted Ki was compared to the known experimentally-determined Ki values and the Pearson's R correlation coefficient was computed for both models PLS and ANN (Table **2**). This procedure was repeated for the logKi in all 15 runs (Table **2**). Findings suggested that ANN predicted Ki activity more efficiently than NIPALS. GRNN-structure generated predictions with the highest Pearson's R for the runs 3, 5 and 11 and RBF-structure for runs 6, 7 and 8. Ki was predicted at a higher accuracy than logKi, from both models NIPALS and ANN thus suggesting that Ki values are a better representation of the activity of a compound than logKi.

Dimensionality Reduction

Often a number of variables may carry - to some extent - the same information as other variables. This problem is known as multi-colinearity and the only remedy known is to decrease the dimensionality of the problem in question by selecting only the most significant variables in relation to the predicted outcome. Genetic algorithms and neural networks were used, separately, to identify the most significant descriptors of the compounds in the training and validating split datasets. As anticipated, separate applications of GA and ANN on the 15 split datasets suggested a ranking of significance for each one of the chemical descriptors. From the original set of 192 descriptors a set of 25 significant descriptors was identified through GA. However, when using these 25 descriptors as inputs to predict Ki, NIPALS demonstrated very poor performance (up to 50% reduction of Pearson's-R on the basis of the results presented - Table **2**). Separately, ANN was used to identify significant descriptors. ANN suggested a ranking of descriptors varying from 1 (most significant) to 192 (less significant) in each of the 15 runs. Variations were noted in this ranking order of significance amongst the 15 runs. However, no statistically

Table 2. Comparison of Pearson's R Correlation Coefficient for Both Models NIPALS and ANN for 15 Random Splits of the Master Dataset into *Training (T)* and *Validating (V)* with the Uses of 192 Descriptors for ANN and 25 Descriptors for NIPALS

Network	Run	ANN Ki *Training*	ANN Ki *Validating*	ANN logKi *Training*	ANN logKi *Validating*	PLS Ki *Training*	PLS Ki *Validating*	PLS logKi *Training*	PLS logKi *Validating*
RBF	1	0,9990	0,8772	0,9908	0,5242	0,8651	0,8632	0,8309	0,7668
GRNN	2	0,9955	0,4016	0,9888	0,7546	0,4674	0,2054	0,5057	0,5140
GRNN	3	0,9966	0,9914	0,9872	0,6598	0,7760	0,9410	0,9115	0,7741
RBF	4	0,9957	0,8011	0,9199	0,6873	0,3102	0,7115	0,2838	0,6082
GRNN	5	0,9907	0,9996	0,9835	0,7796	0,8412	0,8677	0,8808	0,7435
RBF	6	0,9890	0,9788	0,9313	0,7149	0,8299	0,8237	0,8244	0,7781
RBF	7	0,9916	0,9839	0,8790	0,7870	0,8111	0,8272	0,8853	0,7872
RBF	8	0,9912	0,9563	0,8905	0,8498	0,9113	0,8748	0,7314	0,7456
RBF	9	0,9663	0,8300	0,9671	0,7295	0,6899	0,8157	0,7937	0,7735
RBF	10	0,9878	0,7557	0,8879	0,7378	0,2649	0,7149	0,4007	0,5652
GRNN	11	0,9972	0,9551	0,9886	0,6801	0,2297	0,6000	0,0989	0,4469
RBF	12	0,9644	0,8716	0,8776	0,7546	0,8736	0,8505	0,6379	0,7400
RBF	13	0,9734	0,9447	0,8926	0,7105	0,7868	0,8335	0,8107	0,7102
RBF	14	0,9815	0,6725	0,8851	0,8034	0,2073	0,5658	0,2186	0,6133
RBF	15	0,9984	0,9209	0,9915	0,7560	0,7892	0,8449	0,8021	0,7590

significant difference was noted (ANOVA-P<0.05). The 60 (out of 192) most significant descriptors identified through ANN for the runs 3, 5, 6, 7, 8 and 11 are listed (Table **3**). These runs have Pearson's-R correlation coefficient values greater than 0.95 for both training and validating. This indicates a consistency amongst the ranking of the descriptors; this ranking is a good indication of the compounds' universal properties in relation to their activities.

Table 3. Relative Importance of the First 60 Molecular Descriptors Out of 192 Obtained from the Run 3, 5, 6, 7,8 and 11 Using ANN

Molecular Descriptor	Description	run 3	run 5	run 6	run 7	run 8	run 11	Median	Sum-Max/6	Min	Max
STDDIM3	Standard dimension 3: the square root of the third largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	1	1	1	20	7	1	1	2	1	20
MC_108	CH3 4 bonds from a CH2	3	3	5	14	8	2	4	4	2	14
DLI_18	# of 3-level bonding patterns	5	6	12	1	6	7	6	5	1	12
MC_86	#CH2 or CH3 separated by non-C	6	7	18	47	3	8	8	8	3	47
MC_97	#O 4 bonds from an N	2	2	54	8	27	3	6	8	2	54
MC_116	#CH3 3 bonds from a CH2	4	4	127	12	21	6	9	9	4	127
MC_139	#OH groups	14	11	16	5	22	10	13	11	5	22
MC_80	#N separated by 4 bonds	15	10	9	90	16	14	15	13	9	90
MC_49	'#charged atoms'	22	5	56	53	4	4	14	18	4	56
MC_115	#CH3 2 bonds from a CH2	7	13	29	191	37	13	21	20	7	191
MC_113	#O in non-aromatic bonds[46]	54	9	22	54	10	5	16	20	5	54
MC_111	#N 2 bonds from a CH2	13	32	34	9	12	38	23	20	9	38
MC_23	#C bonded 1 N and 2 O	42	30	15	3	81	28	29	24	3	81
MC_141	#CH3 groups -2 if key(160)>2; else 0)	12	27	45	11	77	25	26	24	11	77
MC_128	#CH2s separated by 4 bonds	20	22	82	24	25	31	25	24	20	82
PETITJEA	Petitjean graph Shape Coeffecient as defined in: (diameter - radius) / radius.	10	20	119	55	32	20	26	27	10	119
PETITJEA	Value of (diameter - radius) / diameter	9	17	90	77	24	19	22	29	9	90
MC_89	#O separated by 4 bonds	107	24	40	18	42	34	37	32	18	107
MC_143	#non ring O connected to a ring	17	16	57	161	55	15	36	32	15	161
MC_149	(#CH3 groups)-1 if key(160)>1; else 0	21	39	55	17	167	32	36	33	17	167
LOGPOW	Log of the octanol/water partition coefficient (including implicit hydrogens). This property is an atomic contribution model[66] that calculates logP from the given structure; i.e., the correct protonation state (washed structures). Results may vary from the logP(o/w) descriptor. The training set for SlogP was ~7000 structures.	34	42	36	75	11	50	39	35	11	75
RGYR_	Radius of gyration	49	73	13	22	18	78	36	35	13	78
MC_150	#X separated by (!r)-r-(!r)	18	18	185	16	118	12	18	36	12	185
MC_96	#atoms in 5-rings	26	28	47	74	192	9	38	37	9	192
MC_157	#O in C-O single bonds	44	41	50	41	36	27	41	38	27	50

(Table 3. Contd….)

Molecular Descriptor	Description	run 3	run 5	run 6	run 7	run 8	run 11	Median	Sum-Max/6	Min	Max
MC_90	#het. 3 bonds from a CH2	19	26	49	95	63	35	42	38	19	95
MC_147	#CH2 attached to CH2	16	12	93	183	64	17	41	40	12	183
MC_66	#CX4 bonded to >=3 carbons	25	8	27	119	182	24	26	41	8	182
MC_114	#CH3 attached to CH2	11	14	114	169	54	11	34	41	11	169
MC_153	#non-carbons attached to CH2	32	29	88	13	189	47	40	42	13	189
MC_100	#N attached to CH2	28	21	113	118	29	30	30	44	21	118
MC_82	#heteratoms attached to a CH2	40	33	48	50	127	54	49	45	33	127
MC_106	#X bonded to >= 3 non-C	52	37	28	79	169	29	45	45	28	169
STDDIM2	Standard dimension 2: the square root of the second largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	29	25	6	135	89	83	56	46	6	135
MC_104	#hets. 2 bonds from a CH2	8	15	58	163	132	21	40	47	8	163
MC_140	(#oxygens)-3 if key(164)>3; else 0)	35	40	130	76	49	40	45	48	35	130
MC_53	#QH 4 bonds from another QH'	23	19	103	170	79	16	51	48	16	170
STDDIM1	Standard dimension 1: the square root of the largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	37	58	10	125	75	67	63	49	10	125
MC_160	#CH3 groups	27	44	60	80	139	39	52	50	27	139
PC		24	97	23	127	91	22	58	51	22	127
MC_152	#C bonded to >=2 C and 1 O	58	46	129	37	72	45	52	52	37	129
SLOGP	Log of the octanol/water partition coefficient (including implicit hydrogens). This property is an atomic contribution model[66] that calculates logP from the given structure; i.e., the correct protonation state (washed structures). Results may vary from the logP(o/w) descriptor. The training set for SlogP was ~7000 structures.	41	53	31	162	76	59	56	52	31	162
MC_54	#QH 3 bonds from another QH	53	54	62	114	33	63	58	53	33	114
MC_129	#CH2s separated by 3 bonds	39	65	191	71	26	66	66	53	26	191
MC_112	#atoms with coordination number >= 4	68	43	96	91	44	23	56	54	23	96
MC_155	#non-ring CH2	33	34	118	164	35	49	42	54	33	164
MC_163	#atoms in 6 rings	71	76	35	139	19	70	71	54	19	139
DLI_17	# of 2-level bonding patterns	81	36	86	39	114	42	62	57	36	114
DLI_10	# of H-bond acceptors	62	60	64	26	188	72	63	57	26	188
KIERA1	First alpha modified shape index: s (s-1)2 / m2 where s = n + a . The Kier and Hall kappa molecular shape indices compare the molecular graph with minimal and maximal molecular graphs, and are intended to capture different aspects of molecular shape.	82	104	61	30	43	73	67	58	30	104

(Table 3. Contd….)

Molecular Descriptor	Description	run 3	run 5	run 6	run 7	run 8	run 11	Median	Sum-Max/6	Min	Max
MC_74	#dimethyl substituted atoms	50	77	133	36	45	85	64	59	36	133
DLI_20	# of aromatic systems	43	47	81	186	20	108	64	60	20	186
MC_123	#O separated by 1 C	36	38	192	73	113	43	58	61	36	192
MC_38	#C bonded >= 2 N and 1 C	147	146	7	6	5	145	76	62	5	147
MC_132	#O 2 bonds from CH2	56	35	78	112	177	37	67	64	35	177
MC_164	#oxygens	48	48	126	174	47	51	50	64	47	174
MC_159	'Key(164)-1 if key(164)>1; else 0'	46	49	125	172	48	53	51	64	46	172
MC_145	#6M RING > 1	160	191	2	2	1	157	80	64	1	191
MC_146	Key(164)-2 if key(164)>2; else 0	47	50	124	171	52	52	52	65	47	171

DISCUSSION

Properties of Descriptors

This study attempted to tackle the problem of identifying and selecting a set of molecular descriptors for a universal peptidomimetic class of HIV-1 Pr inhibitors rather than focusing on a particular subclass. Based on the ranking of the most important descriptors, suggested by ANN, the chemical, biological and clinical activity of the HIV-1 Pr inhibitors was considered to analyze the: 1) shape of chemical compound in relation to the catalytic cavity, 2) hydrophobicity (lipophilicity) in relation to entropy and its possible association to hyperlipidemia, 3) hydrogen bond donor and acceptor in relation to enthalpy, 4) presence of a central hydroxyl group and 5) significance of other descriptors which contribute less to the biological activity (on the basis of the ANN-ranking established in the present study).

Shape

Topological and geometrical shapes of chemical compounds are important factors in assessing molecular activity. Molecular topology describes the way that the atoms in a molecule are bonded together. Whereas, molecular shape is an attribute of a molecule dealing with spatial extension, form, framework, or geometry. Study findings suggested that shape descriptors were important contributors. For instance STD_DIMM-1 to 3 ranked in positions 38, 34, and 1 (Table **3**), respectively. Also, ANN ranks radius of gyration (RGYR) at 22 and both PetitJean at 16 and 17, respectively. STD_DIMM-3 is a standard dimension equivalent to the standard deviation along a principal component axis. STD_DIMM3 was more important than the other dimension. This would suggest that peptidomimetic compounds correlate with a certain size in dimension 3, thus indicating that the potential inhibitor has certain size in this dimension which expresses an exclusive volume effect to fit the catalytic site cleft of 24 Á long. Accordingly, Nelfinavir has the smallest molecular structure amongst the six approved HIV-1 Pr inhibitors; its predictive Ki value is 100-fold higher than the experimental value. It is noted that not many molecules of Ki activity comparable to that of Nelfinavir were included in the modeling. Therefore, the development of a more universal model of HIV-1 Pr peptidomimetics-class would require the consideration of additional compounds compatible (in terms of structure and Ki activity) to Nelfinavir.

Hydrophobicity (Lipophilicity)

Hydrophobicity and water solubility are properties which are used as early as ADME screens [39] to reject probable development failures early on stage. Study findings suggested that hydrophobicity is an important contributor that controls the activity of HIV-1 Pr inhibitors. The hydrophobic feature of HIV-1 Pr inhibitor comes from descriptors representing aliphatic hydrocarbon chains, aromatic groups and calculated logP, where logP denotes the log of the octanol/water partition coefficient - including implicit hydrogen. The MACC_108, 166, 115, 141, 149, 114, 160 keys are all related to the number of methyl CH_3 groups and to the length of the aliphatic chain. They were all amongst the first 40 most significant descriptors at the ranking order 2, 6, 10, 14, 20, 29 and 39, respectively. Furthermore, descriptors representing aromaticity were ranked at positions 24, 47 and 52 and logP was ranked at position 21. As aforementioned, logP values represent the hydrophobic binding of a drug to a receptor and its water solubility and permeability. Typical values range from -3 (very hydrophilic) to +7 (very hydrophobic). However, in the vast majority of drugs logP values are in the 2-4 range. The hydrophobic contribution to the structure of HIV-1 Pr inhibitors is dictated, in general, by the structure of the protease itself. In its mature form, the viral protease exists as a dimer with two subunits consisting of 99 amino acids. The folded subunits interact together to form a core hydrophobic catalytic cavity, which measures 24 Á long and 6-9 Á in diameter. Within the hydrophobic active site pocket there are two catalytic aspartyl residues named Asp25 and Asp25´. The latter are involved in the hydrolysis of the scissile peptide bond of the substrate with the participation of a catalytic water molecule. Thus, the hydrophobic cavity limits the freedom of water molecules and permits the entrance of hydrophobic groups. Since HIV-1 Pr cavity is composed of

hydrophobic groups, HIV-1 Pr inhibitors were designed to complement non-convalent interactions at various positions of the inhibitors based on the natural substrate of the HIV-1 Pr. Due to this hydrophobic requirement HIV-1 Pr drugs show low water-solubility.

Hyperlipidemia and Lipodystrophy Syndrome

The relation of hydrophobicity property to drug side effects (such as hyperlipidemia and lipodystrophy syndrome) is unknown; yet it constitutes a major issue [40]. HIV-1 Pr inhibitors are lipophilic and thus are absorbed with difficulty by the human body. Further examination of the magnitude of the calculated logP values and correlation with marketed HIV-1 Pr inhibitors would suggest a clear trend between high logP and hyperlipidemia. The ranking of severity of hyperlipidemia is Ritonavir (RTV - highest severity), Lopinavir (LPV), Nelfinavir (NFV), Indinavir (IDV) and Amprenavir (AMP - relatively benign) [41]. Their associated calculated logP values are 6.63, 5.71, 5.33, 2.39 and 1.82, respectively. Saquinavir (FTV) with a logP=6.16 does not seem to follow this trend. In the case of Ritonavir it is known that a dose of 400 mg bid or higher adversely affects the serum lipid profiles similarly to the activity of RTV combined with Lopinavir [42]. Furthermore, as a class of drugs, protease inhibitors are highly hydrophobic and hence may be concentrated in fatty tissues. Also, the daily administration and chronic exposure to these medications may impact on the mitochondria in fatty tissues [43].

Hydroxyl Group and Bonding Level Pattern

HIV-1 Pr catalytic mechanism is believed to involve a general acid general base (GAGB) assisted nucleophilic attack of a catalytic water molecule upon the carbonyl carbon leading to a non-covalently bonded intermediate [44]. In the design of the peptidomimetic TS analogues, the inhibitor has to avoid hydrolysis by the protease. The replacement of the peptide C=O with a hydroxyl OH group creates a non-hydrolysable TS mimic. Study findings suggested that the presence of OH groups was important and ranked at positions 7 and 25.

Hydrogen Bond Acceptor and Donor

Study findings suggested the significance of the bonding level pattern between C=O and O. DLI_18 ranked at 3rd place typifying the # of a 3-level pattern as defined by Xu *et al.* [28] (Fig. (**1**)). It represents a three-bond distance between a carbonyl group C=O and an oxygen or nitrogen

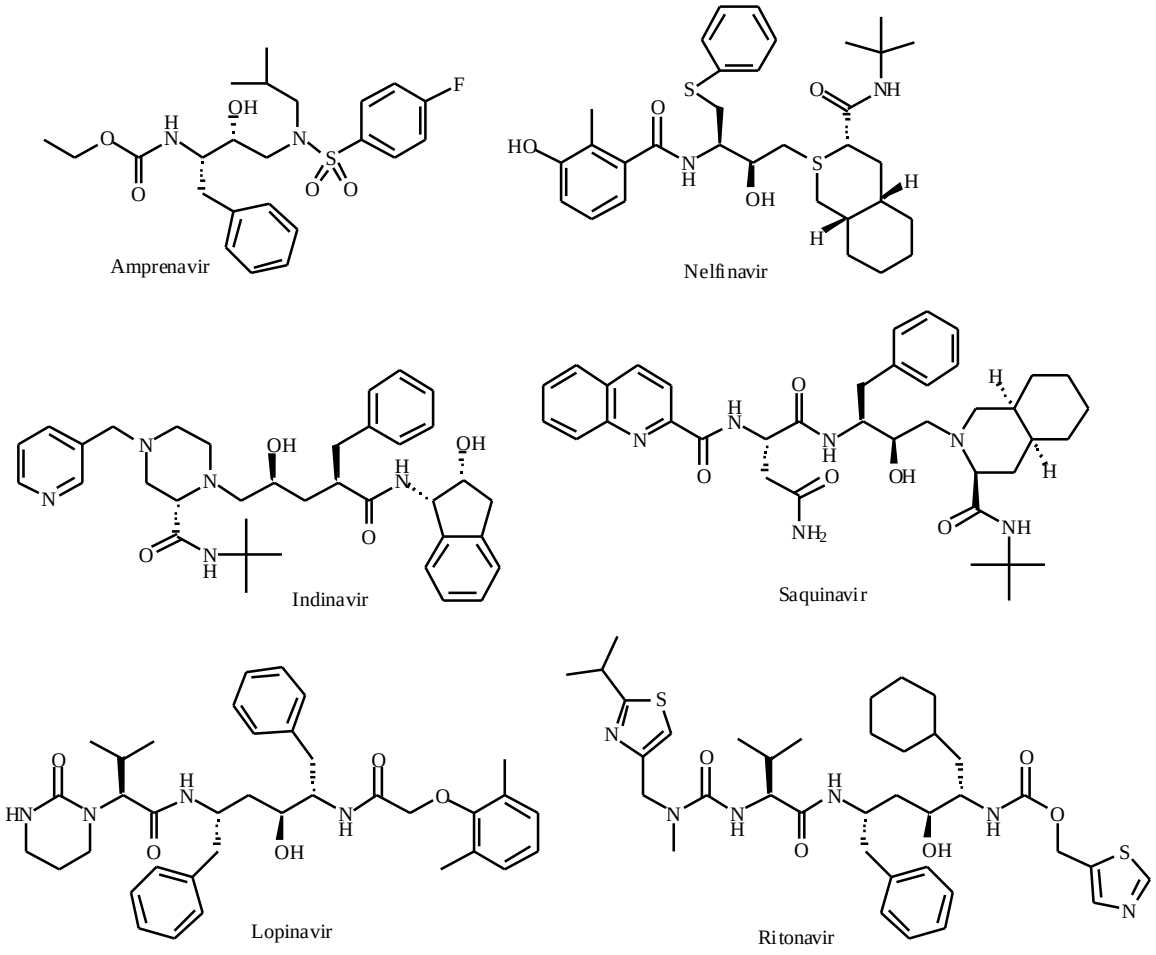

Fig. (1). FDA approved HIV-1 Protease inhibitors.

atom. According to the structures of HIV-1 marketed drugs (Table **4**), all compounds contain a carbonyl oxygen are at a three-bond distance from the central hydroxyl group. This means that a C=O group, positioned three bonds away from the central hydroxyl bond, is an important hydrogen-bond acceptor contributing to the activity of the enzyme. The total number of hydrogen bonds acceptor ranked at position 49. They are represented by the number of atoms bearing a lone pair electron such as C=O atoms. Descriptors that ranked at positions 5, 11, 18, 25, 36 and 53 fit the description of hydrogen bonds acceptor. The presence of NH groups is also significant. These represent the hydrogen bond donor interactions with the viral enzyme. They ranked at positions 8, 12, and 31. Hydrogen bond donor and acceptor as well as logP are characteristics that belong to - what is commonly known as Lipinski's - *rule-of-five* (ROF) [45]. Lipinski's ROF is a heuristic guide for determining if a compound will be orally bio-available. These rules were derived from the World Drug Index [46] following an analysis of 2245 approved compounds. Compounds tested today in clinical trials in humans follow the aforementioned characteristics.

Other Descriptors

Within the set of descriptors considered in this study, the following seem to be less important for the activity of the HIV-1 Pr inhibitor: 1) Physical and electronic properties descriptors such as sum of the atomic polarizabilities, total charge of the molecule (sum of formal charges), molecular refractivity, molecular weight, dipole moment, 2) Van der Waals (VdW) volume, area of Van der Waals surface and subdivided surface areas which are descriptors based on an approximate accessible van der Waals surface area calculation for each atom. These VdW descriptors are related to binding, transport, and solubility, 3) Topological descriptors such as Wiener Index, Zagreb index, Kier & Hall connectivity index, Kappa shape indices, atomic and carbon connectivity index, atomic and carbon valence connectivity index, 4) All atom and bond counts partial charge descriptors, 5) Energy descriptors were not considered in the study, however, preliminary results suggested that ROF, adaptability and shape of the molecule account for the major activity of the compound class.

Table 4. Molecular Structure of HIV-1 Pr Inhibitors with Experimentally Determined Ki and Calculated logP and Predicted Ki Values

Compound name (Fig. (1))	Molecular Class	M.W.	Hdon	Hacc	logP	Ki exp (nM)	Ki predict
Amprenavir Agenerase® VX-478 141W94 AIDS006080	Hydroxyethylamine	505.63	3	9	1.82	0.6 [67] 0.17 [68] 0.11 [69]	0.012
Indinavir Crixivan® MK-63 AIDS005824	Hydroxypropylene	613.8	4	9	2.39	0.358 [70]	0.041
Lopinavir Aluviran Kaletra® ABT-378 AIDS032937	Hydroxyethylene	628.81	4	9	5.71	0.0013 [71]	N.A.
Nelfinavir Viracept AG1343 AIDS028590	Decahydroisoquinoline	567.79	4	8	5.33	0.17 [72] 2.3 [73]	221
Saquinavir Invirase® Fortovase® Ro-31-8959 AIDS000640	Decahydroisoquinoline	670.85	5	11	6.16	0.14 [74]	0.172
Ritonavir Norvir® ABT538 AIDS028478	Thiazoles	720.96	4	13	6.63	0.37 [68] 0.0017	0.012

Free Energy, Enthalpy and Entropy

Non-covalent interactions such as lipophilicity, hydrogen-bond donor and acceptor and shape of molecule can account for the overall antiviral activity of the HIV-1 Pr compounds. In drug design the thermodynamic binding of a drug to its therapeutic target includes free energy of binding, enthalpy and entropy. Enthalpy contribution is expressed through the drug-target interaction relative to the solvent. The primary contribution comes from hydrogen bonding and Van der Waals interactions. Entropy contribution is primarily due to hydrophobic interactions caused by an increase in the solvent entropy from burial of hydrophobic groups of the drugs and by release of water molecules upon binding and also from a small loss of conformational degrees of freedom of the candidate molecule. A drug with favourable entropy indicates that the binding is driven by hydrophobic interactions and low hydrogen bound formation. This type of drugs are hydrophobic and poorly water soluble and are also conformationaly restraint. This entails that they lack a potential of adaptability and consequently are highly susceptible to cause drug resistance [47]. The challenge would be to fine-tune the hydrogen bond contribution, hydrophobicity and flexibility of this particular class of compounds. One strategy to alleviate clinical adverse reactions and overcome suboptimal pharmacokinetics is to modify HIV-1 Pr inhibitors into prodrugs [48-51] which are then subsequently converted to their parent drug by *in vivo* chemical reaction. Hydrophilic prodrug such as fosamprenavir [52, 53] (Lexiva®) a phosphate ester prodrug of amprenavir [48, 53-56] seems to be a major alternative to the problem of hydrophobicity and thus entropy factor.

Similar studies - applied to the non-nucleoside HIV-1 reverse transcriptase inhibitors-have shown that nonconvalent interactions such as lipophilic, steric, hydrogen-bonding and inductive forces are responsible for the antiviral and cytotoxic actions [57]. Thus, a comprehensive understanding of biological activity through the respective molecular descriptors establishes the grounds for conceptualization of additional targeted strategies for the discovery of NCE.

Virtual HTS (vHTS)

The current drug discovery processes in many pharmaceutical companies require large collections of small molecules for use in HTS assays [58]. Identifying new class of compounds with HTS is a costly procedure because appropriate inhibition assays have to be developed, small molecules compounds have to be purchased and the results obtained should be environment dependent (pH, ionic strength and buffer conditions). Moreover, hit lists generated by HTS typically contain a large percentage of false positives [58], making follow-up assays necessary to distinguish between active from inactive substances [59]. Presently, the availability of the 3D structure of the target and knowledge of its active site provide sufficient accuracy to conduct virtual HTS (vHTS). This approach is cost-efficient since there is no need to buy costly HTS library of compounds and set up biological assays for the experimental screening. For instance, a successful application of vHTS was the Anthrax Research Project, in which 300,000 molecules approximately were identified (from the pool of 3.5 billion) as possible candidates for further research [60, 61]. That virtual screening project was completed in 24 days, a process which traditionally would have taken years to complete. Amongst others, the authors [62] were able to demonstrate that pharmacophore modeling - based on hydrogen bond donors and acceptors - can be used to effectively screen small molecule databases to assist drug-discovery process. This task was attained through the use of vHTS which facilitated the identification of SARS-CoV Protease [62, 63] through screening of over 3 million compounds [64] based on a gamut of molecules readily purchasable. An example demonstrating the increasing complexity of similar experiments is the Anthrax project: the respective database comprised of 3.5 billion virtual molecules. As a future research step would be to combine both experiences: HIV-1 Pr descriptors identification through ANN and SARS-CoV Pr vHTS to choose features *for in silico* modeling such as hydrogen bond donor and acceptor, hydrophobic aromatic and aliphatic, positive and negative charges, positively and negatively ionisable groups, shape and excluded volume. Applications of interest would be: the search for compounds that inhibit dimer formation, inhibition of stable monomers of HIV-1 Pr, and a comprehensive understanding of C2 symmetric inhibitors.

CONCLUSION

A consistent set of descriptors was determined to predict the activity of HIV-1 Pr inhibitors; 15 different training and validating sets were used for this task. The best structures of ANN providing very good predictions were GRNN and RBF. The set of descriptors suggested by the ANN reflects fairly well the characteristics and biological activity of HIV-1 Pr inhibitors. These account for enthalpy and entropy contribution to the binding affinity of the compounds with descriptors representing hydrogen bonding, hydrophobicity, lipophilicity and shape contribution. Amongst the 192 descriptors tested, it would appear that the category of descriptors representing non-covalent interactions accounts for the antiviral activity of the HIV-1 Pr compounds. Due to consistent contribution to hydrophobicity, the significance of lipophilicity and relationship to HIV-1 associated hyperlipidemia and lipodystrophy syndrome warrants further investigation. The distinctive difference between cause-effect and determining cause should be highlighted. When environmental conditions are met, a response will generally occur as a result of various sets of determining rules. For instance, there is a direct relationship between HIV-1 Pr inhibitors and adverse serum lipid profiles resulting in hyperlipidemia. The challenge is to design drugs with a: 1) binding adaptability and favourable enthalpy in order to reduce drug resistance and 2) reduced level of entropy contribution which in turn is related to the hydrophobicity of the compounds. Presently, hydrophilic prodrug seems to be an alternative to the lipophilicity of the HIV-1 Pr inhibitors.

As a future direction, the authors are investigating descriptors for the non-peptidomimetic C2-symmetric compounds - such as the recently approved Tipranavir [65]. Modeling of: 1) hydrogen bond donor and acceptor, 2) hydrophobic aromatic and aliphatic, 3) positive and negative

charges, 4) positively and negatively ionisable groups, 4) shape and excluded volume, within the concept of a comparison between the peptidomimetic and non-peptidic class of compounds, is expected to impact the drug discovery process.

ACKNOWLEDGEMENTS

This work was supported by research funding from the Canadian Institutes of Health Research (CIHR) grant 200303HHP-118726-HSR-CFAC-26137 (funding decision HHP-64516), and from the Fonds de la Recherche en Sante du Quebec (FRSQ). Dr. Hatzakis was also supported by a Chercheur Boursier Career Award from the FRSQ. The authors would like to thank Mr James Vatistas, Administrative Coordinator of IDTC for fruitful discussion. Dr. Wei was supported by grants from the Tianjin Commission of Sciences and Technology under the contract number 023618211 and the Chinese National Science Foundation under the contact No. 20373048. Dongqing Wei is supported by the grants from Chinese National Foundation under the contract 20373048 and Tianjin Scientific Commission under contract 033801911.

ABBREVIATIONS

QSAR	=	Quantitative Structure Activity Relationship
PLS	=	Partial Least Squares
ANN	=	Artificial Neural Networks
GA	=	Genetic Algorithms
HIV-1	=	Human Immunodeficiency Virus-1
HIV-1 Pr	=	HIV-1 Protease
Ki	=	Constant of Inhibition
HTS	=	High Throughput Screening
PCA	=	Principal Component Analysis
NIPALS	=	Nonlinear Iterative Partial Least Squares
RBF	=	Radial Basis Function Artificial Neural Network
GRNN	=	Generalized Regression Artificial Neural Network
RGYR	=	Ranks Radius of Gyration
RTV	=	Ritonavir, HIV-1 Pr Inhibitor, Abbott Labs®
LPV	=	Lopinavir, HIV-1 Pr Inhibitor, Abbott Labs®
NFV	=	Nelfinavir, HIV-1 Pr Inhibitor, Agouron®
IDV	=	Indinavir, HIV-1 Pr Inhibitor, Merck®
AMP	=	Amprenavir, HIV-1 Pr Inhibitor, GlaxoSmithKline®
FTV	=	Saquinavir, HIV-1 Pr Inhibitor, Roche®
GAGB	=	General Acid General Base (GAGB)
HTS	=	High Throughput Screening
VHTS	=	Virtual HTS
ROF	=	Lipinski's rule-of-five
WDI	=	World Drug Index
VdW	=	Van der Waals

REFERENCES

[1] Sirois, S.; Proynov, E. I.; Truchon, J. F.; Tsoukas, C. M.; Salahub, D. R. *J. Comput. Chem.* **2003,** *24*, 1110-1119.
[2] Chou, K. C.; Zhang, C. T. *J. Protein Chem.* **1993,** *12*, 709-724.
[3] Senese, C. L.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **2003,** *43*, 1297-1307.
[4] Kiralj, R.; Ferreira, M. M. *J. Mol. Graph Model* **2003,** *21*, 499-515.
[5] Perez, J. J.; Corcho, F.; Llorens, O. *Curr. Med. Chem.* **2002,** *9*, 2209-2229.
[6] Weber, I. T.; Miller, M.; Jaskolski, M.; Leis, J.; Skalka, A. M.; Wlodawer, A. *Science* **1989,** *243*, 928-931.
[7] Chou, K. C. *Anal. Biochem.* **1996,** *233*, 1-14.
[8] Chou, K. C.; Kezdy, F. J.; Reusser, F. *Anal. Biochem.* **1994,** *221*, 217-230.
[9] Kitchen, V. S.; Skinner, C.; Ariyoshi, K.; Lane, E. A.; Duncan, I. B.; Burckhardt, J.; Burger, H. U.; Bragman, K.; Pinching, A. J.; Weber, J. N. *Lancet* **1995,** *345*, 952-955.
[10] Lacy, M. K.; Abriola, K. P. *Conn. Med.* **1996,** *60*, 723-727.
[11] Markowitz, M.; Saag, M.; Powderly, W. G.; Hurley, A. M.; Hsu, A.; Valdes, J. M.; Henry, D.; Sattler, F.; La Marca, A.; Leonard, J. M. *N. Engl. J. Med.* **1995,** *333*, 1534-1539.
[12] Mangum, E. M.; Graham, K. K. *Pharmacotherapy* **2001,** *21*, 1352-1363.
[13] Adkins, J. C.; Faulds, D. *Drugs* **1998,** *55*, 837-842; discussion 843-834.
[14] Kuschak, D.; Mauss, S.; Schmutz, G.; Gantke, B. *Clin. Lab.* **2001,** *47*, 471-477.
[15] Manchanda, T.; Schiedel, D.; Fischer, D.; Dekaban, G. A.; Rieder, M. J. *Can. J. Clin. Pharmacol.* **2002,** *9*, 137-146.
[16] Buckheit, R. W., Jr. *Expert Opin. Investig. Drugs* **2004,** *13*, 933-958.
[17] Koppal, T. *Drug Discov. Developm.* **2004,** *7*, 30-32.
[18] van Heeswijk, R. P.; Veldkamp, A.; Mulder, J. W.; Meenhorst, P. L.; Lange, J. M.; Beijnen, J. H.; Hoetelmans, R. M. *Antivir. Ther.* **2001,** *6*, 201-229.
[19] Scudeller, L.; Torti, C.; Quiros-Roldan, E.; Patroni, A.; Lo Caputo, S.; Moretti, F.; Mazzotta, F.; Donati, E.; Vivarelli, A.; Carosi, G. *J. Antimicrob. Chemother.* **2003,** *52*, 776-781.
[20] Burke, M. D.; Berger, E. M.; Schreiber, S. L. *Science* **2003,** *302*, 613-618.
[21] Kermani, F.; Findlay, G. *The Pharm. R&D Compend.* **2000**.
[22] Prentis, R. A.; Lis, Y.; Walker, S. R. *Br. J. Clin. Pharmacol.* **1988,** *25*, 387-396.
[23] CCG Chemical Computing Group, Molecular Operating Environment (MOE), Chemical Computing Group: Montreal, Canada, 2004.
[24] Dalby, A.; Nourse, J. G.; Hounsell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. *J. Chem. Inf. Comput. Sci.* **1992,** *32*, 244-255.
[25] NIAID. **2004**: http://www.niaid.nih.gov/daids/dtpdb/.
[26] Gossas, T.; Danielson, U. H. *J. Mol. Recognit.* **2003,** *16*, 203-212.
[27] Katritzky, A. R.; Gordeeva, E. V. *J. Chem. Inf. Comput. Sci.* **1993,** *33*, 835-857.
[28] Xu, J.; Stevenson, J. *J. Chem. Inf. Comput. Sci.* **2000,** *40*, 1177-1187.
[29] Oprea, T. I. *J Comput. Aided Mol. Des.* **2000,** *14*, 251-264.
[30] Rishton, G. M. *Drug Discov. Today* **2003,** *8*, 86-96.
[31] Forrest, S. *Science* **1993,** *261*, 872-878.
[32] Goldberg, D. ed.; Addison-Wesley: Reading, MA, **1989**.
[33] Mitchell, M. ed.; MIT Press: Cambridge, MA, **1996**.
[34] Israelsson, A. ed.; Wold H & Lyttkens: 1969; '*Vol.*' pp. 29-51.
[35] Lippmann, R. P.; Shahian, D. M. *Ann. Thorac. Surg.* **1997,** *63*, 1635-1643.
[36] Bang, S. Y.; Hwang, Y. S. *Neural Netw.* **1997,** *10*, 1495-1503.
[37] Mosier, P. D.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2002,** *42*, 1460-1470.

[38] Marsaglia, G.; Zaman, A.; Tsang, W. W. *Stat. Prob. Letters* **1990,** *8*, 35-39.
[39] Li, A. P. *Drug Discov. Today* **2001,** *6*, 357-366.
[40] Distler, O.; Cooper, D. A.; Deckelbaum, R. J.; Sturley, S. L. *Curr. Opin. Clin. Nutr. Metab. Care* **2001,** *4*, 99-103.
[41] Nolan, D.; Upton, R.; McKinnon, E.; John, M.; James, I.; Adler, B.; Roff, G.; Vasikaran, S.; Mallal, S. *Aids* **2001,** *15*, 1275-1280.
[42] Shafran, S. D.; Mashinter, L. D.; Roberts, S. *Int. Cong. Drug Therapy HIV* **2002**, 17-21; 6 abstract No. P129.
[43] Mukhopadhyay, A.; Wei, B.; Zullo, S. J.; Wood, L. V.; Weiner, H. *Mitochondrion* **2002,** *1*, 511-518.
[44] Hyland, L. J.; Tomaszek, T. A. Jr.; Meek, T. D. *Biochemistry* **1991,** *30*, 8454-8463.
[45] Lipinski, C. A. *J. Pharmacol. Toxicol. Methods* **2000,** *44*, 235-249.
[46] WDI World Drug Index: 14 Great Queen Street, London, WC2B 5DF, UK.
[47] Ohtaka, H.; Velazquez-Campoy, A.; Xie, D.; Freire, E. *Protein Sci.* **2002,** *11*, 1908-1916.
[48] Chaudry, N. I.; Eron, J. J.; Naderer, O. J.; Pereira, A. S.; Wire, M. B.; Fiscus, S. A.; Kashuba, A. D. *Clin. Infect. Dis.* **2002,** *35*, 760-762.
[49] Farese-Di Giorgio, A.; Rouquayrol, M.; Greiner, J.; Aubertin, A. M.; Vierling, P.; Guedj, R. *Antivir. Chem. Chemother.* **2000,** *11*, 97-110.
[50] Gaucher, B.; Rouquayrol, M.; Roche, D.; Greiner, J.; Aubertin, A. M.; Vierling, P. *Org. Biomol. Chem.* **2004,** *2*, 345-357.
[51] Rouquayrol, M.; Gaucher, B.; Roche, D.; Greiner, J.; Vierling, P. *Pharm. Res.* **2002,** *19*, 1704-1712.
[52] Wire, M. B.; Ballow, C.; Preston, S. L.; Hendrix, C. W.; Piliero, P. J.; Lou, Y.; Stein, D. S. *Aids* **2004,** *18*, 897-907.
[53] Gathe, J. C. Jr.; Ive, P.; Wood, R.; Schurmann, D.; Bellos, N. C.; DeJesus, E.; Gladysz, A.; Garris, C.; Yeo, J. *Aids* **2004,** *18*, 1529-1537.
[54] Corbett, A. H.; Kashuba, A. D. *Curr. Opin. Investig. Drugs* **2002,** *3*, 384-390.
[55] Falcoz, C.; Jenkins, J. M.; Bye, C.; Hardman, T. C.; Kenney, K. B.; Studenberg, S.; Fuder, H.; Prince, W. T. *J. Clin. Pharmacol.* **2002,** *42*, 887-898.
[56] Furfine, E. S.; Baker, C. T.; Hale, M. R.; Reynolds, D. J.; Salisbury, J. A.; Searle, A. D.; Studenberg, S. D.; Todd, D.; Tung, R. D.; Spaltenstein, A. *Antimicrob. Agents. Chemother.* **2004,** *48*, 791-798.
[57] Mager, P. P. *Curr. Med. Chem.* **2003,** *10*, 1643-1659.
[58] Rishton, G. M. *Drug Discov. Today* **1997,** *2*, 382-384.
[59] Jenkins, J. L.; Kao, R. Y.; Shapiro, R. *Proteins* **2003,** *50*, 81-93.
[60] Anthrax.
[61] Richards, W. D. *Nat. Rev. Drug Discov.* **2002,** *1*, 551-555.

Received: 16 August, 2004 Accepted: 19 October, 2004

[62] Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. *J. Chem. Inf. Comput. Sci.* **2004,** *44*, 1111-1122.
[63] Chou, K. C.; Wei, D. Q.; Zhong, W. Z. *Biochem. Biophys. Res. Commun.* **2003,** *308*, 148-151.
[64] Sirois, S.; Hatzakis, G.; Wei, D.; Du, Q.; Chou, K. *Computational Biology and Chemistry* **2004,** *(Accepted for Publication)*.
[65] Cheonis, N. *Beta* **2004,** *16*, 15-17.
[66] Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1999,** *39*, 868-873.
[67] Navia, M. A.; Sato, V.; Tung, R. D. *Int. Antiviral News* **1995,** *3*, 143-145.
[68] Wilkerson, W. W.; Akamike, E.; Cheatham, W. W.; Hollis, A. Y.; Collins, R. D.; DeLucca, I.; Lam, P. Y.; Ru, Y. *J. Med. Chem.* **1996,** *39*, 4299-4312.
[69] Ala, P. J.; Huston, E. E.; Klabe, R. M.; Jadhav, P. K.; Lam, P. Y.; Chang, C. H. *Biochemistry* **1998,** *37*, 15042-15049.
[70] Vacca, J.; Dorsey, B. D.; Schleif, W. A.; Levin, R. B.; McDaniel, S. L.; Darke, P. L.; Zugay, J.; Quintero, J. C.; Blahy, O. M.; Roth, E.; Sardana, V.; Schlabach, A. J.; Graham, P. I.; Condra, J. H.; Gotlib, L.; Aanderson, P. S.; Emini, E. A.; Huff, J. R.; *et al. Proc. Natl. Acad. Sci. USA* **1994,** *91*, 4096-4100.
[71] Sham, H. L.; Kempf, D. J.; Molla, A.; Marsh, K. C.; Kumar, G. N.; Chen, C. M.; Kati, W.; Stewart, K.; Lal, R.; Hsu, A.; Betebenner, D.; Korneyeva, M.; Vasavanonda, S.; McDonald, E.; Saldivar, A.; Wideburg, N.; Chen, X.; Niu, P.; Park, C.; Jayanti, V.; Grabowski, B.; Granneman, G. R.; Sun, E.; Japour, A. J.; Norbeck, D. W. *Antimicrob. Agents Chemother.* **1998,** *42*, 3218-3224.
[72] Patick, A. K.; Mo, H.; Markowitz, M.; Appelt, K.; Wu, B.; Musick, L.; Kalish, V.; Kaldor, S.; Reich, S.; Ho, D.; Webber, S. *Antimicrob. Agents Chemother.* **1996,** *40*, 292-297.
[73] Pyring, D.; Lindberg, J.; Rosenquist, A.; Zuccarello, G.; Kvarnstrom, I.; Zhang, H.; Vrang, L.; Unge, T.; Classon, B.; Hallberg, A.; Samuelsson, B. *J. Med. Chem.* **2001,** *44*, 3083-3091.
[74] Chen, C. A.; Sieburth, S. M.; Glekas, A.; Hewitt, G. W.; Trainor, G. L.; Erickson-Viitanen, S.; Garber, S. S.; Cordova, B.; Jeffry, S.; Klabe, R. M. *Chem. Biol.* **2001,** *8*, 1161-1166.
[75] Chou, K. C.; Zhang, C. T.; Kezdy, F. J. *Proteins: Struct. Funct. and Genet.* **1993**, *16*, 195-204.
[76] Chou, J. J. *Journal of Protein Chemistry*, **1993**, 12, 291-302.
[77] Chou, K. C. *Journal of Biological Chemistry*, **1993**, *268*, 16938-16948.
[78] Thompson, T. B.; Chou, K. C.; Zheng, C. *Journal of Theoretical Biology* **1995**, *177*, 369-379.
[79] Chou, K. C.; Tomasselli, A. L.; Reardon, I. M.; Heinrikson, R. L. *Proteins: Struct. Funct. and Genet.* **1996**, *24*, 51-72.
[80] Yang, Z. R.; Chou, K. C. *Bioinformatics* **2004**, *20*, 735-741.