Three-Dimensional Quantitative Structure-Activity Relationship (QSAR) of HIV Integrase Inhibitors: A Comparative Molecular Field Analysis (CoMFA) Study

K. Raghavan, John K. Buolamwini, Mark R. Fesen, Yves Pommier, Kurt W. Kohn, and John N. Weinstein*

Laboratory of Molecular Pharmacology, National Cancer Institute, Developmental Therapeutics Program, Division of Cancer Treatment, Building 37, Room 5C25, National Institutes of Health, Bethesda, Maryland 20892

Received April 1, 1994[⊗]

We present the results from a comparative molecular field analysis (CoMFA) of a set of flavone analogs that inhibit HIV-1 integrase-mediated cleavage (3'-processing step) and integration (strand transfer step) *in vitro*. The results indicate a strong correlation between the inhibitory activity of these flavones and the steric and electrostatic fields around them. CoMFA quantitative structure-activity relationship models with considerable predictive ability (cross-validated r^2 as high as 0.8) were obtained.

Introduction

The HIV-1 integrase mediates integration of reverse-transcribed viral DNA into the host genome, an essential step in the life cycle of the virus. This enzyme presents an attractive target for the development of agents active against AIDS. Recently, Fesen et al. identified families of compounds that inhibit the integrase.[1,2] These include topoisomerase inhibitors, antimalarial agents, DNA binders, naphthoquinones, flavones, and caffeic acid phenethyl ester (CAPE). In their assay, purified recombinant integrase (expressed in *Escherichia coli*[3]) is treated with a 21-mer oligonucleotide corresponding to the U5 end of HIV-1 proviral DNA. The integration reaction consists of two steps. The first results in nucleolytic cleavage of two bases from the 3' end next to the conserved CA dinucleotide (referred to as 3'-processing); the second is a strand transfer (i.e., integration) reaction in which the recessed ends are joined to the 5' end of an integrase-induced break in a second, identical oligonucleotide, which serves as the target DNA.[3-7]

The results obtained to date by Fesen et al. show no direct relationship between integrase inhibition and topoisomerase-II poisoning or DNA binding affinity.[1] For example, some topoisomerase inhibitors (doxorubicin, mitoxantrone, quercetin, and the ellipticines) are potent inhibitors of the integrase, whereas others (e.g., amsacrine, etoposide, teniposide, and camptothecin), are inactive. The fact that flavones are natural products, abundant in food and flowers, and relatively nontoxic, motivated Fesen et al. to study these compounds and their analogs in greater detail. They assessed the effect of hydroxylation, glycosylation, and methoxy substitution on the capacity of a set of flavones to inhibit HIV-1 integrase *in vitro*. Their study indicated that hydroxyl groups at certain positions are important for activity. Details of the assay and its results can be found in ref 2.

To obtain further insight into the relationship between the structure and function of these flavones as integrase inhibitors, we have carried out quantitative structure-activity relationship (QSAR) studies using the comparative molecular field analysis (CoMFA) method. CoMFA,[8] introduced by Cramer et al. in 1988, is one of the most used 3D-QSAR methods.[8-11] It has been applied to a number of different classes of compounds.[12-27] The method is based on the premise that ligand-receptor interaction reflects steric and electrostatic forces. CoMFA can be useful in the design of ligands when, as in this case, the three-dimensional structure of the receptor site is unknown.

CoMFA requires that a single conformation be selected for each molecule in the database. In the case of rigid molecules, there is no problem. In the case of conformationally flexible species, most studies have used the lowest energy conformation for each molecule. These minimum energy conformations are then superimposed either by the "field-fit" technique in Sybyl[11] or by matching atoms in the rigid fragment that are common to all molecules. The superimposed molecules are placed in a cubic lattice. Next, the steric and electrostatic interactions experienced by a probe atom placed at each lattice point are computed and stored in an array. Then, the partial least squares (PLS) technique[11,29] is used to derive a 3D-QSAR. The CoMFA technique was applied to a set of flavone analogs for which HIV-1 integrase inhibition activity has been determined by Fesen et al.[2] The results are presented in this paper.

Methods

Molecular Modeling and CoMFA. The structures of the flavones and their HIV-1 integrase inhibitory activities are given in Table 1. The data for both cleavage and integration were used in the CoMFA study. The three-dimensional structures of the molecules were constructed using the molecular modeling program Sybyl[11] on an Indigo Elan workstation (Silicon Graphics Inc., Mountain View, CA). Each structure was first energy-minimized using the Tripos force field, with distance-dependent dielectric function and the default convergence criteria. Partial atomic charges required for calculation of the electrostatic interaction were computed by a semiempirical molecular orbital method using the MOPAC program.[11] The charges were computed using the AM1 method.

The energy-minimized structures were then subjected to conformational analysis. The rotatable bonds in each molecule were examined using the systematic conformational search technique in Sybyl. All of the molecules studied have at least one rotatable bond (excluding the hydroxyl group rotation), the single bond between the phenyl group and carbon atom 2.

* Author to whom correspondence should be addressed.

[⊗] Abstract published in *Advance ACS Abstracts*, February 15, 1995.

The orientations of sugar rings in the substituents were selected from the minimum energy conformation. The conformations obtained in this way for each molecule were compiled in a molecular database for use in the CoMFA study described below.

The atoms forming the fused ring system (nine carbon atoms and one oxygen atom in each molecule) were superimposed on the equivalent atoms in a template molecule (quercetagetin, in this case) using the "match" function in Sybyl. A three-dimensional cubic lattice ($11 \times 12 \times 9 = 1188$ grid points) with a spacing of 2 Å was then constructed around these molecules. We used two different probe atoms in the study, an sp^3 carbon with a charge of +1.0 (default probe atom in Sybyl) and an sp^3 oxygen with a charge of −0.4. The probe atom was placed at each lattice point, and its steric and electrostatic interactions with each atom in the molecule were computed and then saved in a CoMFA QSAR table. The calculations were done with different cutoff values, in the range 10−100 kcal/mol, for both steric and electrostatic energies. Then, partial least squares fitting[11,29] was used to obtain a 3D-QSAR. PLS is a regression technique for solving linear models. It is analogous to principal component regression in that it extracts an orthogonal set of explanatory variables that are linear combinations of the original variables. In principal component regression, however, extraction of orthogonal variables is independent of the target variables, and a subsequent multiple regression step determines the relationship between target and explanatory variables. In PLS, the orthogonal set of variables is constrained to maximize the communality of the predictor and response variable blocks.

Cross-validated PLS was used to find the set of reduced variables that yield the best predictive model. In cross-validation (a method for estimating the predictive ability of a statistical procedure) a randomly selected subset of observations is omitted and the model constructed with the remaining set. The predictive ability of the model is quantitated in terms of the cross-validated r^2 (c-v r^2)[11] which is defined as

$$\text{c-v } r^2 = 1.0 - \frac{\sum_y (y_{\text{pred}} - y_{\text{actual}})^2}{\sum_y (y_{\text{actual}} - y_{\text{mean}})^2} \quad (1)$$

where y_{pred}, y_{actual}, and y_{mean} are predicted, actual, and mean values of the target property, respectively. The summation is taken over all observations. As indicated by eq 1, c-v r^2 is equal to the proportional reduction in the sum of squared residuals due to the model. The c-v r^2 contrasts with the non-cross-validated r^2 (obtained using all observations in the model), which gives only a measure of how well the model fits the data, not its predictive ability. In this study, we used the "leave-one-out" method, in which each molecule's activity is predicted by a model derived from the rest of the molecules. While doing the PLS, one can specify a value for the maximum number of components to be used in the analysis. During cross-validation, PLS calculates the c-v r^2. The c-v r^2 initially increases with the number of components and then takes on an almost constant value after an appropriate number of components have been included in the model. During cross-validation, CoMFA columns (i.e., values at a given lattice point) were filtered so that the columns retained some specified variance (which in this study was 2.0 kcal/mol). This was done to speed up computation while determining the optimum number of components. Once the optimum set of reduced variables was determined, a final PLS analysis was carried out on the entire data set to obtain the final model to be used for predictions beyond the data set. This final PLS fit was done with no filtering.

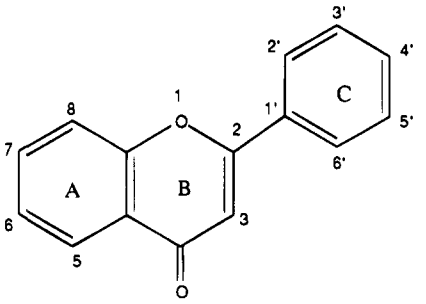
Prior to discussing the results, it will be useful to review some of the technical details of various options used in this CoMFA QSAR analysis. The steric and electrostatic energies computed for each molecule are the sums of all pairwise interactions between the probe atom and each atom in that molecule. The maximum value to be retained at each lattice point (i.e., if the point falls very near to, or within, the molecule) must be decided and the computed value "clipped" accordingly ("steric-energy-max" and "elec-energy-max" in CoMFA). One can choose not to include the electrostatic interaction at lattice points with steric values at the maximum. This choice is available through the "drop electrostatics" option in Sybyl. If this option is used, the electrostatic energy for the lattice point in the particular molecule is computed from the average of electrostatic interactions for the rest of the molecules at that lattice point (i.e., treating each missing value as the column average in the CoMFA QSAR table). We used the "switching function" option in treating the energy values at lattice points during the transition from inside the atom to outside. Also, one can scale both steric and electrostatic fields using CoMFA-STD scaling. In CoMFA-STD scaling, the CoMFA variables (steric and electrostatic fields) are scaled using the overall field mean and standard deviation, rather than the individual column mean and standard deviation. The selection of a specific option while deriving the QSAR will have some influence on the results obtained. These influences are discussed along with the results that follow.

Results and Discussion

In a preliminary study, when we first developed a CoMFA model for inhibition of the cleavage step using all 15 molecules in Table 1, the results indicated compound **14** (6-MeO-luteolin) to be an outlier. The c-v r^2 of the resulting model was −0.12, implying that the model lacked ability to predict. The residuals obtained are summarized in Table 2 under column a. The residual associated with 6-MeO-luteolin was the highest and greater than two standard deviations (std-dev = 0.67). Also, the residuals associated with predicted activities of other compounds were quite high. It is common practice in QSAR studies to omit outliers in the spirit of exploratory data analysis. Hence, we omitted 6-MeO-luteolin and derived a CoMFA model with the remaining 14 molecules. The residuals from this model are also given in Table 2 under column b. It is clear that in this case the residuals for almost all compounds were substantially reduced below the values obtained when 6-MeO-luteolin was included. The associated c-v r^2 was 0.81, as will be discussed in detail later.

We wanted to see if any of the other 14 molecules could have been eliminated as outliers. Table 3 summarizes the results of 15 different models in which each of the 15 molecules was assumed to be an "outlier". In such an experiment one would expect, a priori, that the model with the "true outlier" removed would show a high c-v r^2 compared to others. Indeed, Table 3 shows such a result. When 6-MeO-luteolin was eliminated, the c-v r^2 of the model was 0.81. Otherwise, it was poor. For these reasons, we decided to omit 6-methoxyluteolin and focus only on the remaining 14 molecules for further study. It must be remembered, however, that this sort of selection process affects statistical validation and interpretation of the results. A Bonferroni correction[28] for the omission of 6-methoxyluteolin will be considered later in this section. It can be seen from Table 1 that this is the only molecule that has a methoxy substituent at the 6 position. The CoMFA model appears to associate 6-methoxyluteolin with quercetagetin and baicalein on the basis of structural similarity with respect to a substituent at the 6 position, despite the fact that its activity is quite different. Apparently, the CoMFA approach does not sufficiently distinguish the methoxy and hydroxyl moieties in this context.

Table 1. Structures of the Flavones and Their Integrase Inhibition Activity



flavones	IC_{50} (μM) cleavage	IC_{50} (μM) integration	ring substitutions[a] 3	5	6	7	8	2′	3′	4′	5′
1, quercetagetin	0.8	0.1	OH	OH	OH	OH			OH	OH	
2, baicalein	1.2	4.3		OH	OH	OH					
3, robinetin	5.9	1.6	OH			OH			OH	OH	OH
4, myricetin	7.6	2.5	OH	OH		OH			OH	OH	OH
5, quercetin	23.6	13.6	OH	OH		OH			OH	OH	
6, fisetin	28.4	8.5	OH			OH			OH	OH	
7, luteolin	32.9	25.0		OH		OH			OH	OH	
8, myricetrin	39.6	10.3	RH	OH		OH			OH	OH	OH
9, quercetrin	60.0	38.5	RH	OH		OH			OH	OH	
10, rhamnetin	61.6	28.7	OH	OH		MeO			OH	OH	
11, avicularin	66.3	25.1	AR	OH		OH			OH	OH	
12, gossypin	69.7	22.5	OH	OH		OH	GL		OH	OH	
13, morin	76.5	31.7	OH	OH		OH		OH		OH	
14, 6-MeO-luteolin	94.3	39.1		OH	MeO	OH			OH	OH	
15, kaempferol	97.8	64.7	OH	OH		OH				OH	

[a] RH = rhamnose; AR = arabinose; GL = glucose.

Table 2. Comparison of Residuals from Cross-Validated Predictions for the Cleavage Model: (a) Residuals from the Model Including All 15 Compounds and (b) Residuals from the Model after Excluding Compound **14**

compound	actual $\log(1/IC_{50})$	residual a	residual b
1, quercetagetin	6.097	1.164	0.467
2, baicalein	5.921	1.176	0.220
3, robinetin	5.229	0.436	0.203
4, myricetin	5.119	0.471	−0.057
5, quercetin	4.627	−0.071	0.078
6, fisetin	4.547	−0.387	−0.228
7, luteolin	4.483	−0.099	−0.122
8, myricitrin	4.402	0.232	−0.237
9, quercetrin	4.222	0.039	0.133
10, rhamnetin	4.210	−0.548	−0.458
11, gossypin	4.157	−0.565	−0.485
12, avicularin	4.178	−0.318	−0.329
13, morin	4.116	−0.601	0.106
14, 6-MeO-luteolin	4.025	−1.430	–
15, kaempferol	4.010	−0.739	−0.289

Table 3. Effect of Elimination of Each Compound on the Cross-Validated r^2

compound excluded	c-v r^2
1, quercetagetin	−0.570
2, baicalein	−0.259
3, robinetin	−0.183
4, myricetin	−0.115
5, quercetin	−0.134
6, fisetin	−0.112
7, luteolin	−0.163
8, myricitrin	−0.136
9, quercetrin	−0.185
10, rhamnetin	−0.135
11, gossypin	−0.126
12, avicularin	−0.137
13, morin	−0.141
14, 6-MeO-luteolin	0.813[a]
15, kaempferol	−0.141

[a] Compound eliminated from final QSAR.

Figure 1 shows 14 flavone molecules superimposed on each other in the CoMFA lattice. Figure 2 summarizes results obtained using the cleavage data with an sp^3 carbon (charge = +1.0) as the probe atom. The effect of using different values for the steric and electrostatic energy cutoffs is shown in this figure. As one can see, the CoMFA approach yields reasonably predictive QSAR models for the system of flavone analogs studied. There seems to be some dependence of the c-v r^2 on the choice of cutoff value used when the electrostatics are dropped at steric maximum. In this case, as the cutoff value changes from 10 to 100 kcal/mol, the c-v r^2 varies from 0.32 to 0.72. However, the variation is less pronounced when no scaling is used for the steric and electrostatic fields. Higher cutoff values emphasize the effect of interactions at lattice points close to atoms in the molecules. It is interesting to note that when the electrostatics are not dropped at steric maximum, the models derived are substantially better in their predictive ability, as indicated by c-v r^2 in the range 0.78−0.82. Also the c-v r^2 remains almost constant as the cutoff value is varied. Use of CoMFA-STD scaling has little effect on the c-v r^2. On the basis of the results summarized in Figure 2, we have decided not to drop electrostatics at steric maximum (i.e., electrostatics were included at all lattice points). The cutoff value we used for both steric and electrostatic interactions is 10 kcal/mol.

Figure 3 shows results from the CoMFA model for cleavage data (c-v r^2 = 0.81; cutoff = 10 kcal/mol; number of components = 9). The model has 21% contribution from the steric field and 79% contribution from the electrostatic field. The percentage contribution indicates which explanatory variables influence the QSAR. In this case, the final QSAR model is mostly influenced by the electrostatic field (79% contribution)

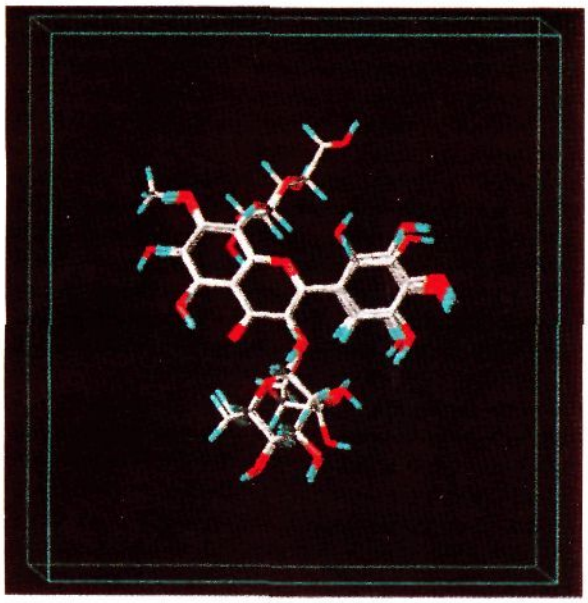


Figure 1. Superimposition of 14 flavones in the CoMFA lattice.

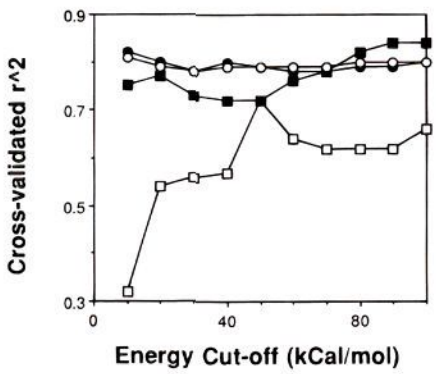


Figure 2. Effect of energy cutoff on the predictive ability of the CoMFA model for DNA cleavage inhibition. Squares: electrostatics dropped at steric-max. Circles: electrostatics not dropped at steric-max. Open symbols: CoMFA-STD. Filled symbols: no scaling. Note: The fact that the open and closed squares at an energy cutoff of 50 kcal/mol are at essentially the same point is a computational coincidence, not an error in graphing.

around the flavones relative to the steric field (21% contribution). Figure 3 shows major features of the steric and electrostatic maps from the final QSAR model. For reference, quercetagetin is displayed inside the field. The model suggests that activity would be favored by the presence of a bulky group near the volume colored green (around position 6 of the flavone ring) and by the lack of a bulky group near those colored yellow. Similarly, positive charge near the blue regions (e.g., the region near the 3′ and 4′ positions) and negative charge near the red regions (e.g., near the 6 and 5′ positions) would favor increased activity. It should be emphasized that though the contributions from steric and electrostatic fields are usually separated during CoMFA analysis for ease of interpretation and visualization, their interplay in determining the activity should not be forgotten.

The number of components in the model denotes the optimum number of components (which are linear combinations of the original variables) that gives the best predictive CoMFA model. As mentioned earlier in the Methods section, the c-v r^2 initially increases with the number of components and then reaches an almost constant value after the optimum number of components is attained. This is shown in Figure 4 for the cleavage model. The c-v r^2 increases from a value of 0.2 and reaches nearly a plateau at a value of 0.8 for nine components. The conventional (i.e. non-cross-validated) r^2 for the final model is 0.999. Other statistics associated with this model are as follows: standard error of estimate = 0.036, F = 499.8, probability (P) of obtaining this value of F if r^2 were actually zero (prob of $r^2 = 0$) < 0.001. With a Bonferroni correction[28] for leaving out 6-methoxyluteolin, $P_{critical}$ (two tails) = $P/C(15,1) = 0.025/15 = 0.0017$. $C(15,1)$ is the number of ways that one of the 15 compounds can be omitted. Thus, the result appears still to be statistically significant at the 5% level, even given the stringency of the Bonferroni correction.

Figure 5 shows the cross-validated prediction obtained from this model for each molecule. The activity of each molecule shown in this figure was predicted by a model constructed using the rest of the molecules in the study. In other words, the calculation did not include the molecule for which the activity was predicted. Also shown in the figure is the activity value for 6-MeO-luteolin predicted by the final CoMFA model derived using the remaining 14 molecules. The predicted value is similar to that of other 6-substituted flavones, quercetagetin and baicalein. This finding emphasizes that 6-MeO-luteolin can be considered an outlier and gives further reason for leaving it out in additional analyses. It should be mentioned that the activities of the molecules considered in this study fall into three or four groups as revealed by Figure 5. At this time, there are no data available on molecules that have activities in the intermediate range between these groups. More data in these intermediate ranges would be useful and would increase the generalizability of this QSAR.

Since the lattice spacing used in these calculations was 2 Å, we were concerned that the results obtained might be highly sensitive to the discretization. Hence, we investigated the effect of offsetting the lattice. The results obtained for CoMFA models (not dropping electrostatics at steric-max; no scaling) with different offsets are summarized in Table 4. It is encouraging to see that the predictive abilities of these models, as indicated by the c-v r^2, are affected only moderately. Overall, we examined 180 permutations of lattice shifts with different CoMFA and PLS options (not shown). In only four of these were poor values of c-v r^2 obtained.

As a further test of robustness of the CoMFA models, we randomized the target values (cleavage and integration inhibition data) for the set of 14 molecules analyzed and derived CoMFA models with different combinations of options and cutoff values as before. None of those models had significant c-v r^2. The c-v r^2 obtained were in the range −0.02 to −0.5. This indicates that the c-v r^2 in the CoMFA models with original data are not due to chance correlations.

Fesen et al.[3] suggested that the presence of two hydroxyl substituents adjacent to each other in either aromatic ring A or C is needed for activity against the integrase. Presence of a third hydroxyl group next to

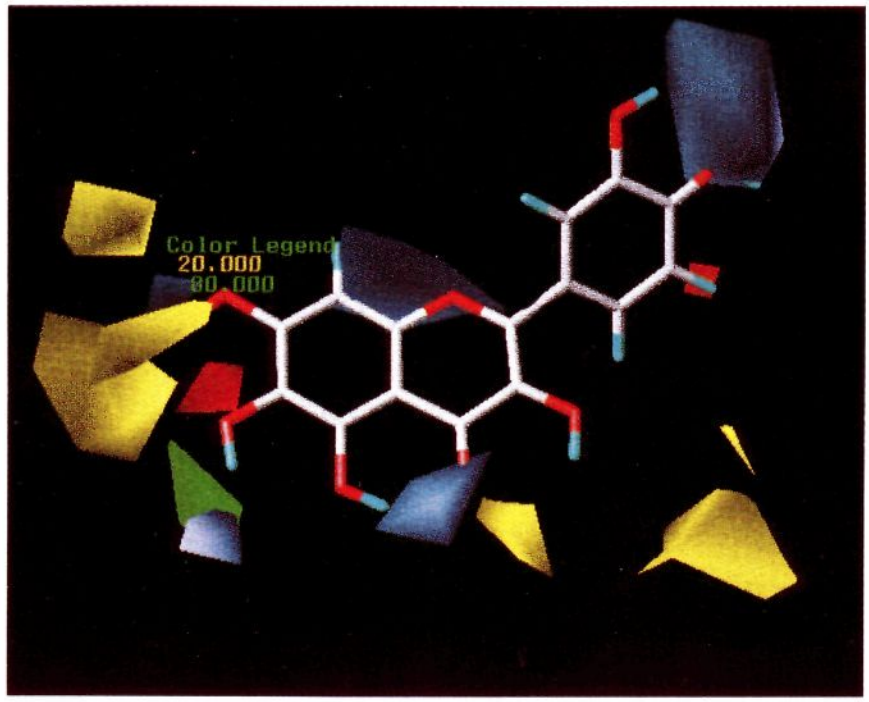


Figure 3. Steric and electrostatic maps from the CoMFA model for DNA cleavage inhibition. Quercetagetin shown inside the field. Favoring activity: green, bulky group; yellow, less bulky group; blue, positive charge; red, negative charge.

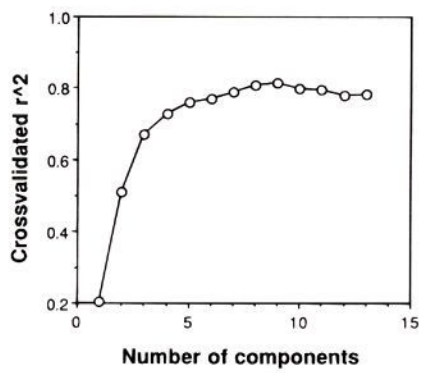


Figure 4. Cross-validated r^2 as a function of number of components for the DNA cleavage model.

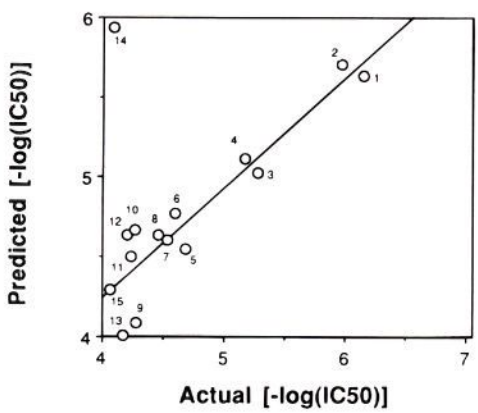


Figure 5. Cross-validated prediction for DNA cleavage data (c-v r^2 = 0.81). Compound **14**, which was omitted from this QSAR calculation, appears as an outlier.

Table 4. Comparison of CoMFA Models for DNA Cleavage Inhibition by Flavones Using Different Offsets of Lattice

lattice offset (Å)			steric−max/electrostatic−max (kcal/mol)				
X	*Y*	*Z*	10	20	30	40	50
0.0	0.0	0.0	0.81	0.79	0.78	0.79	0.79
0.5	0.0	0.0	0.81	0.79	0.77	0.76	0.77
0.0	0.5	0.0	0.75	0.71	0.71	0.69	0.70
0.0	0.0	0.5	0.82	0.76	0.73	0.70	0.65
0.5	0.5	0.5	0.81	0.78	0.75	0.75	0.76
−0.5	0.0	0.0	0.79	0.81	0.81	0.80	0.81
0.0	−0.5	0.0	0.86	0.86	0.83	0.83	0.85
0.0	0.0	−0.5	0.74	0.75	0.73	0.71	0.72
−0.5	−0.5	−0.5	0.77	0.83	0.82	0.77	0.75
	mean		0.80	0.78	0.77	0.76	0.76
	std−dev		0.04	0.04	0.04	0.05	0.06

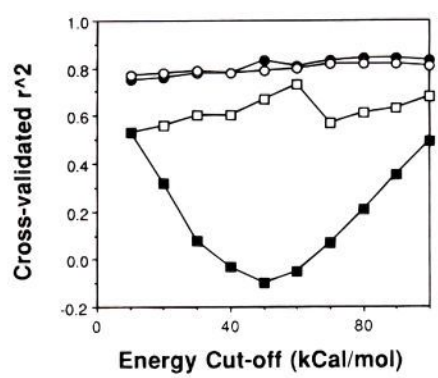


Figure 6. Effect of energy cutoff on predictions of the CoMFA model for inhibition of integration. Squares: electrostatics dropped at steric-max. Circles: electrostatics not dropped at steric-max. Open symbols: CoMFA-STD. Filled symbols: no scaling.

them, as in quercetagetin, appears to make the compound more active. Here, and elsewhere in this paper, the term "active" refers to observed activity at less than 100 μM. We wanted to see if our CoMFA model predicted those characteristics. The model predicts that an additional OH group at the 5′ position in quercetagetin will increase the activity. This agrees with the trend observed experimentally that three OH groups

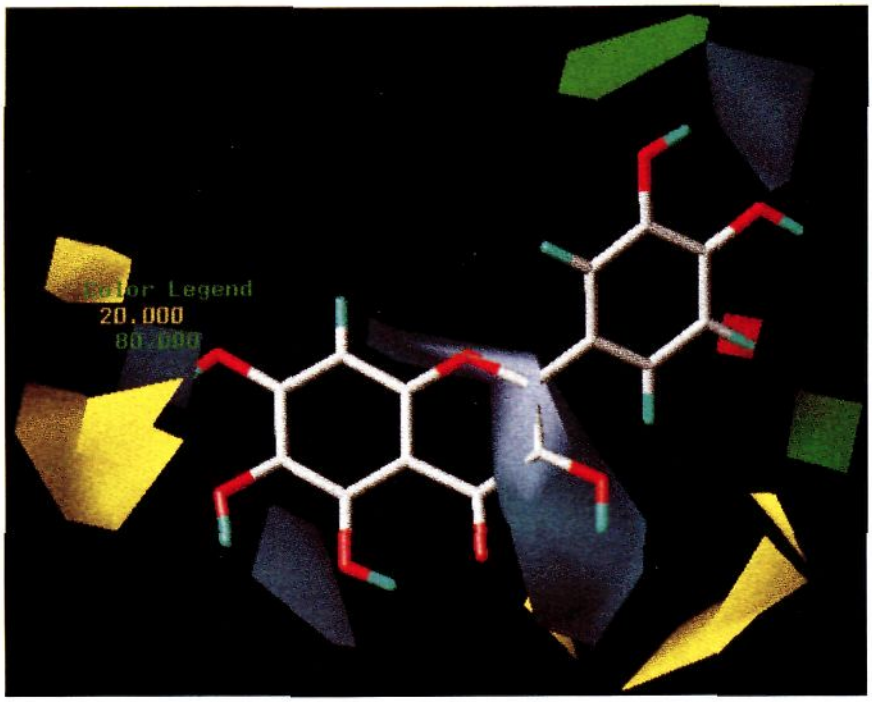


Figure 7. Steric and electrostatic maps from the CoMFA model for inhibition of integration. Quercetagetin shown inside the field. Favoring activity: green, bulky group; yellow, less bulky group; blue, positive charge; red, negative charge.

adjacent to each other in either aromatic ring enhance the activity.

Further, we used the model to predict the cleavage data for molecules representing modifications of quercetagetin. We found that if the OH group at position 5 is removed from quercetagetin, the predicted IC_{50} value for that molecule is 1.44 μM, which is somewhat higher than the value of 0.8 for quercetagetin. Removal of an additional OH group (from the 6 position) leads to the compound fisetin, which has an experimental value of 28.4 μM and a predicted value of 26.1 μM. This is also in agreement with the empirical observation that compounds with adjacent hydroxyl groups are more active than those without them.

Since the CoMFA approach encodes only steric and electrostatic factors, it predicted that a chlorine substituent would have qualitatively the same effect as a hydroxyl substituent. Unfortunately, there were no halogenated compounds in the current database with which to test this proposition. We were able, however, to identify in the National Cancer Institute's Drug Information System (DIS) an analog of one of the active compounds but with two hydroxyl groups replaced by chlorines. (Since the compound was submitted confidentially, its structure cannot be specified.) It was tested in the integrase assay and found to be inactive with respect to both cleavage and integration at 100 μM. This finding emphasizes the danger of extrapolating beyond what the parametrization and method of analysis can distinguish. Viewed from another angle, this observation suggests that the hydroxyl groups are playing a role beyond their electronegativity, perhaps by hydrogen bonding.

We have also used different probe atoms to investigate the robustness of the CoMFA model. The results (not shown) indicated further increase in the c-v r^2 when an sp^3 oxygen (O.3) probe with a charge of -0.4 was used instead of the sp^3 carbon (C.3) with a charge of +1.0. The effect on c-v r^2 of using different cutoff values was similar to that seen for the C.3 probe. The c-v r^2 varied in the range 0.44−0.88.

Values of c-v r^2 for CoMFA models based on the integration data are shown in Figure 6. When all 14 molecules were included, the models derived were only moderately predictive. The highest c-v r^2 obtained was 0.5, obtained when the electrostatics were not dropped at steric maximum. The c-v r^2 for the corresponding model with electrostatics dropped was 0.1. This kind of large difference can arise from poor prediction of a particular molecule during the cross-validation. In fact, the activities of quercetagetin and baicalein were poorly predicted in the cross-validation. It has been observed[8,11] that CoMFA is very sensitive to either a unique structure or a unique value for activity in the data set. In this particular case, these are the only two molecules in the set that have a substituent in the 6 position. So, we excluded these two molecules from the data set and carried out CoMFA analysis on the remaining 12 molecules. This approach is not very satisfying, since these are among the more potent agents with respect to integration. Note that 6-methoxyluteolin, the compound originally excluded from consideration, also has a substituent at the 6 position. Clearly, more analysis will be required to determine why we do not get better QSAR predictions for integration with respect to this position.

CoMFA results for integration data using the 12 molecules are summarized in Figure 6. As seen earlier for cleavage, the models that include electrostatics at all lattice points seem to be quite stable, with very good c-v r^2 (in the range 0.7−0.8) throughout. When the electrostatics are dropped, the effect of cutoff (both steric and electrostatic) on c-v r^2 is much more pronounced than for cleavage data. Steric and electrostatic features of the CoMFA model for integration data are shown in Figure 7 (c-v $r^2 = 0.77$; cutoff = 10 kcal/mol; number of

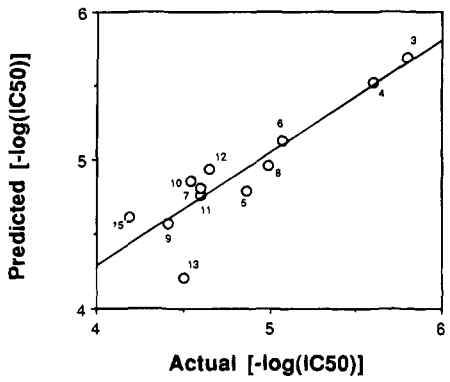


Figure 8. Cross-validated prediction for integration data (c-v $r^2 = 0.77$).

components = 8). The corresponding cross-validated predictions are given in Figure 8. Steric and electrostatic field contributions to this model are 20.5% and 79.5%, respectively. The conventional r^2 for the final non-cross-validated model is 1.0, the F value is 804.2, and the probability of $r^2 = 0$ is <0.001. However, this must be considered an exercise in exploratory data analysis; with three selected molecules excluded, the statistical properties are poor. The Bonferroni correction (which assumes independent effects) is presumably much too conservative, but it would indicate a critical P value reduced by a factor of $C_{15,12} = 15!/(12!\,3!) = 455$. Comparison of Figures 3 and 6 indicates several regions around the flavones that are predicted to be important for both cleavage and integration.

We have also carried out another QSAR study on the same set of flavones (Buolamwini et al., in preparation[30]) using a newly developed set of structural descriptors, the electrotopological (E-state) indices formulated by Kier and Hall.[31] The E-state indices merge topological information on each molecule with information on electronic states of its atoms as determined from the valence electrons. It is encouraging to note that results from that study also identify similar regions around the flavones as important for activity against HIV-1 integrase. Those findings, to be presented separately, greatly increase our confidence in the results obtained heere with the CoMFA approach.

Summary and Conclusion

We have derived 3D-QSAR models using the CoMFA methodology for a set of flavones that are known to inhibit HIV-1 integrase *in vitro*. The results indicate a correlation between the inhibitory activity of these flavones and the steric and electrostatic fields around them. The QSAR models reveal regions in three-dimensional space around these flavones that are important for HIV-1 integrase inhibition. The models obtained in this study are reasonably predictive, as indicated by the cross-validated r^2 values. The CoMFA QSAR models derived will be used in the design of new flavone analogs that may be more potent inhibitors of HIV-1 integrase. The results obtained in this study also indicate possible limitations, including those that arise from small size of the sample and from attempts to extrapolate outside of the immediate domain of the data.

Acknowledgment. We would like to thank Dr. Shaomeng Wang of the Laboratory of Medicinal Chemistry at the National Cancer Institute, NIH, for many insightful discussions and critical reading of this manuscript. We also thank Dr. Marc C. Nicklaus of the Laboratory of Medicinal Chemistry at the National Cancer Institute, NIH, and Drs. Mike Lawless and Roy J. Vaz of Tripos Associates, Inc. for many insightful discussions. This work was supported in part by the NIH Intramural AIDS Targeted Antiviral Program.

References

(1) Fesen, M. R.; Kohn, K. W.; Leteurtre, F.; Pommier, Y. Inhibitors of Human Immunodeficiency Virus Integrase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2399–2403.
(2) Fesen, M. R.; Leteurtre, F.; Hiroguchi, S.; Yung, J.; Kohn, K. W.; Pommier, Y. Inhibition of HIV Integrase by Flavones and Caffeic Acid Phenethyl ester. *Biochem. Pharmacol.* **1994**, *48*, 595–608.
(3) Bushman, F. D.; Fujiwara, T.; Craigie, R. Retroviral DNA Integration Directed by HIV Integration Protein in vitro. *Science* **1990**, *249*, 1555–1558.
(4) Brown, P. O. Integration of Retroviral DNA. *Curr. Top. Microbiol. Immunol.* **1990**, *157*, 19–48.
(5) Varmus, H.; Brown, P. Retroviruses. In *Mobile DNA*; Berg, D., Howe, M., Eds.; Am. Soc. Microbiol.: Washington, DC, 1989; pp 53–108.
(6) Goff, S. P. Genetic of Retroviral Integration. *Annu. Rev. Genet.* **1992**, *26*, 527–544.
(7) Kulkovsky, J.; Skalka, A. M. HIV DNA integration: Observations and Inferences. *J. Acquired Immune Defic. Syndr.* **1990**, *3*, 839–851.
(8) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(9) Marshall, G. R.; Cramer, R. D., III. Three-dimensional Structure-Activity Relationships. *TIPS Rev.* **1988**, *9*, 285–289.
(10) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Boostrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.
(11) Sybyl Molecular Modeling Software, version 6.02, 1993. Tripos Associates, Inc., St. Louis, MO 63144; Sybyl theory manual, 1993.
(12) Nicklaus, M. C.; Milne, G. W. A.; Burke, T. R., Jr. QSAR of Conformationally Flexible Molecules: Comparative Molecular Field Analyses of Protein-tyrosine Kinase Inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 487–504.
(13) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of Angiotensin-converting Enzyme and Thermolysin Inhibitors: A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active Site Geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372–5384.
(14) McFarland, J. W. Comparative Molecular Field Analysis of Anticoccidial Triazines. *J. Med. Chem.* **1992**, *35*, 2543–2550.
(15) Kim, K. H. Nonlinear Dependence in Comparative Molecular Field Analysis. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 71–82.
(16) Calder, J. A.; Wyatt, J. A.; Frenkel, D. A.; Casida, J. E. CoMFA Validation of the Superposition of six Classes of Compounds Which Block GABA Receptors Non-competitively. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 45–60.
(17) Loughney, D. A.; Schwender, C. F. A Comparison of Progestin and Androgen Receptor Binding Using the CoMFA Technique. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 569–581.
(18) Rault, S.; Bureau, R.; Pilo, J. C.; Robba, M. Comparative Molecular Field Analysis of CCK-A Antagonists using Field-fit as an Alignment Techniques. A Convenient Guide to Design new CCK-A Ligands.*J. Comput.-Aided Mol. Des.* **1992**, *6*, 553–568.
(19) Harpalani, A. D.; Snyder, S. W.; Subramanyam, B.; Egorin, M. J.; Callery, P. S. Alkylamides as Inducers of Human Leukemia Cell Differentiation: A Quantitative Structure-Activity Relationship Study Using Comparative Molecular Field Analysis. *Cancer Res.* **1993**, *53*, 766–771.
(20) Klebe, G.; Abraham, U. On the Prediction of Binding Properties of Drug Molecules by Comparative Molecular Field Analysis. *J. Med. Chem.* **1993**, *36*, 70–80.
(21) Langlois, M.; Bremont, B.; Rousselle, D.; Gaudy, F. Structural Analysis by the Comparative Molecular Field Analysis Method of the Affinity of Beta-adrenoreceptor Blocking Agents for 5-HT1A and 5-HT1B Receptors. *Eur. J. Pharmacol.* **1993**, *244*, 77–87.
(22) Waller, C. L.; McKinney, J. D. Comparative Molecular Field Analysis of Polyhalogenated Dibenzo-p-dioxins, Dibenzofurans, and Biphenyls. *J. Med. Chem.* **1992**, *35*, 3660–3666.
(23) McFarland, J. W. Comparative Molecular Field Analysis of Anticoccidial Triazines. *J. Med. Chem.* **1992**, *35*, 2543–2550.

(24) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: A new Method of Empirical Hydrophobic Calculation for CoMFA. *J. Comput. Aided Mol. Des.* **1991**, *5*, 545–552.
(25) Diana, G. D.; Kowalczyk, P.; Tresaurywala, A. M.; Oglesby, R. C.; Peaver, D. C.; Dutko, F. J. CoMFA Analysis of the Interactions of Antipicornavirus Compounds in the Binding Pocket of Human Rhinovirus-14. *J. Med. Chem.* **1992**, *35*, 1002–1008.
(26) Carroll, F. I.; Gao, Y. G.; Rahman, M. A.; Abraham, P.; Parham, K.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. Synthesis, Ligand Binding, QSAR, and CoMFA Study of 3 Beta(p-substituted phenyl) tropane-2 Beta-Carboxylic Acid Methyl Esters. *J. Med. Chem.* **1991**, *34*, 2719–2725.
(27) Kim, K. H.; Martin, Y. C. Direct Prediction of Dissociation Constants (pKa's) of Clonidine-like ImidaZolines, 2-substituted Imidazoles, and 1-methyl-2-substituted-imidazoles from 3D Structures Using a Comparative Molecular Field Analysis (CoMFA). *J. Med. Chem.* **1991**, *34*, 2056–2060.
(28) Kleinbaum, D. G.; Kupper, L. L.; Muller, K. E. *Applied Regression Analysis and Other Multivariable Methods*; PWS-KENT Publishing Co.: Boston, 1988.
(29) Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjostrom, M. Multivariate Data Analysis in Chemistry. In *CHEMOMETRICS: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984.
(30) Buolamwini, J. K.; et al. Application of the Electrotopological State Index to QSAR Analysis of Flavone Derivatives as HIV-1 Integrase inhibitors. In review at *Pharm. Res.*
(31) Kier, L. B.; Hall, L. H. Atom Description in QSAR Models: Development and use of an Atom Level Index. *Adv. Drug Res.* **1992**, *22*, 1–38.

JM940201B