Construction of a Molecular Shape Analysis–Three-Dimensional Quantitative Structure–Analysis Relationship for an Analog Series of Pyridobenzodiazepinone Inhibitors of Muscarinic 2 and 3 Receptors

Benjamin J. Burke,† William J. Dunn, III, and A. J. Hopfinger*

Laboratory of Molecular Modeling and Design, Department of Medicinal Chemistry and Pharmacognosy, M/C 781, College of Pharmacy, University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612-7231

Received April 29, 1994®

A generalized three-dimensional (3D) quantitative structure–property relationship (QSPR) formalism, based upon molecular shape analysis (MSA), has been applied to an analog series of pyridobenzodiazepinone inhibitors of muscarinic 2 (M2) and 3 (M3) receptors. The fundamental goal of this application is to establish MSA–3D-QSARs (P = A = inhibition activity) that are based upon identifying the active conformations of these flexible analogs. The repetitive use of partial least squares (PLS) analysis permits the construction of the MSA–3D-QSARs. In addition to molecular shape, the identification of the properties of a lipophilic binding site and specific nonallowed steric receptor sites govern the MSA–3D-QSARs. The M2 and M3 QSARs suggest receptor subtype specificity might be realized by targeting upon a specific nonallowed steric receptor site. One conformation, common to both M2 and M3 receptors, emerges as dominant in the optimum MSA–3D-QSARs. However, other similar conformations are also found to yield meaningful MSA–3D-QSARs.

Introduction

In the previous paper[1] we describe a general formalism to construct MSA–3D-QSPRs for a set of flexible molecules $\{M_u\}$ using multiple trial molecular alignments. In the case of the relative physicochemical feature representation, three classes of feature tensors can be defined as the basis for the QSPR:

$\mathbf{V}_{u,v}(s,\alpha,\beta)$ is the intrinsic steric molecular shape (SMS) tensor where s refers to the set of steric shape measures, α the set of conformations, and β the set of alignments considered in the analysis of $\{M_u\}$. Compound v is the reference compound[1] and u is any member of the data set.

$\mathbf{F}_{u,v}(p,\mathbf{r}_{j,k,l},\alpha,\beta)$ is the molecular field (MF) tensor where p are the set of probes used to generate the molecular field features[2] and the $\mathbf{r}_{j,k,l}$ are the spatial positions at which the potential field is evaluated for each p.

$\mathbf{H}_{u,v}(\alpha,\beta)$ is the tensor which contains all the remaining physicochemical features which may, or may not, depend on α and/or β.

These three tensors can be combined to yield the **VFH** tensor which is related to the property measures $\{P_u\} = P_u$ by

$$P_u = \mathbf{T}_{u,v} \otimes [\mathbf{V}_{u,v}(s,\alpha,\beta),\mathbf{F}_{u,v}(p,\mathbf{r}_{j,k,l},\alpha,\beta),\mathbf{H}_{u,v}(\alpha,\beta)] \quad (1)$$

where $\mathbf{T}_{u,v}$ is the transformation tensor that establishes the specific relationship between $\{P_u\}$ and **VFH**. The current approach to determining $\mathbf{T}_{u,v}$ is the repeated application of partial least squares (PLS) regression.[3]

Equation 1 can be decomposed into a variety of submodel QSPRs. In fact, all popular methods of QSPR analysis are contained within the framework of eq 1.[1] In this paper we apply eq 1 to the MSA–3D-QSAR analysis of an analog series of muscarinic antagonists for which both cardiac and ileal inhibitory activities have been measured.[4] In this MSA–3D-QSAR application the alignment is restricted to one choice and molecular field features are not employed. Consequently, eq 1 reduces to

$$P_u = \mathbf{T}_{u,v} \otimes [\mathbf{V}_{u,v}(s,\alpha,\beta_o),\mathbf{H}_{u,v}(\alpha,\beta_o)] \quad (2)$$

where β_o indicates that only one choice in molecular alignment is being considered. In essence, this application focuses upon establishing the optimum relationships between the P_u and conformational features.

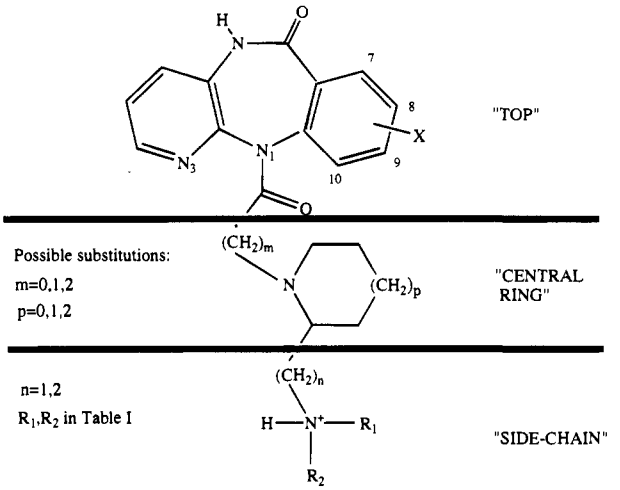


Figure 1. Core structure of the pyridobenzodiazepinone analog series. The structure is divided into three portions. The tricyclic portion of the molecule is referred to as the "top", the middle portion is referred to as the "central ring", and the lower portion as the "side chain" portion of the molecule.

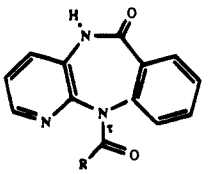
Methods

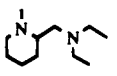
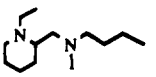
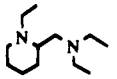
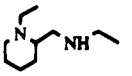
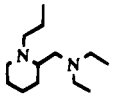
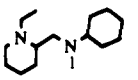
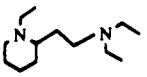
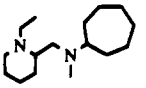
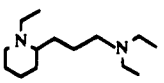
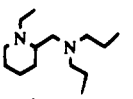
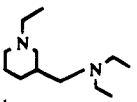
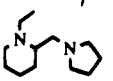
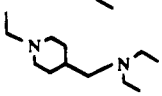
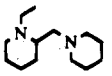
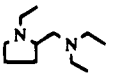
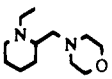
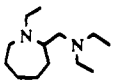
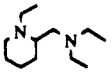
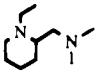
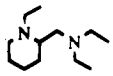
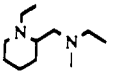
1. SAR Database. Engel et al.[4] report the cardiac (M2) and ileal inhibitory (M3) activities for an analog series of 26 muscarinic antagonists. The core structure of this series is given in Figure 1. Nineteen of the 26 compounds from the original data base[4] were selected for MSA–3D-QSAR analysis

* Author to whom correspondence should be addressed.
† Current address: Agouron Pharmaceuticals, Inc., 3565 General Atomics Court, San Diego, CA 92121.
® Abstract published in *Advance ACS Abstracts*, September 15, 1994.

0022-2623/94/1837-3775$04.50/0

Table 1. Some of the Substituted Tricyclic Pyridobenzodiazepinones Used in the 3D-QSAR Analysis and Their Binding Affinities to Cardiac (M2) and Glandular (M3) Tissue Given as IC_{50} (nM)[a]



No.	R	IC_{50}, nM M2	IC_{50}, nM M3	Selectivity M2/M3	No.	R	IC_{50}, nM M2	IC_{50}, nM M3	Selectivity M2/M3
1		5000	>10000	>2	12		100	200	2.0
2		140	6000	42.9	13		600	3000	5
3		220	600	2.7	14		60	150	2.5
4		330	1000	3.0	15		30	70	2.3
5		60	200	3.3	16		1500	6000	4.0
6		170	300	1.8	17		560	1500	2.7
7		800	4000	5	18		600	1500	2.5
8		150	2000	13	19		>1000	4000	<4
9		2000	15000	7.5	20		400	10000	25
10		140	1000	7.1	21		50	6000	120
11		120	1500	12.5					

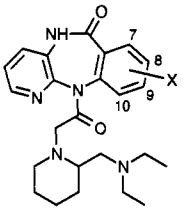
[a] Compound **2** is the racemic mixture; **20** and **21** are the enantiomers.

and are reported in Table 1. Compound **2** in the original report was resolved into its two enantiomers (compounds **20** (+) and **21** (+a)), which were reported as *R* and *S* isomers, respectively. Each of the other compounds was reported as a racemic mixture for consistency. Compounds **20** and **21** were not included in the analysis. The tricyclic portion of the molecules (see Figure 1) was substituted at positions 8 and 9 as defined in Figure 1 for six compounds (**22-27**) of Table 2. The major contribution to activity was postulated to be related to the conformational profiles of the analogues in Table 1 and corresponding molecular shapes of these analogues. The six compounds in Table 2 are unlikely to modify the conformational profile of this series of analogs. These compounds also showed no improvement in the M3/M2 selectivity over that of compound **2**. Hence, they were not considered in this study.

2. Biological Activity. The biological studies used rat tissue homogenates of atria (M2) and submandibular gland (M3). Binding curves for the different compounds were derived indirectly from competition against tritiated *N*-methylscopolamine (nonselective) for each homogenate type. The assays were done in quadruplicate, and nonspecific binding versus racemic 3-quinuclidinyl benzilate (QNB) was reported to average less than 30%. The values for activity are given as $-\log(IC_{50})$. Compound **2** of Table 1 exhibits the highest selectivity of M3/M2 activity in the series and has entered clinical trials. The range of activity is about 2.5 log concentration units for both the M2 and M3 receptors. The ratio of M3/M2 selectivity (which is given as $(IC_{50})_{nM}/(IC_{50})_{nM}$) ranges from 2.3 to 43 for the analogues reported in Tables 1 and 2.

3. Building the Molecules. The crystal structure reported for the tricyclic portion of pirenzipine was fully optimized using MNDO[5] and used for the "top" of the molecules, see Figure 1. The "central ring" structures were built (by fragments from the CHEMLAB-II library[6]) to reflect equatorial–equatorial substitution at the nitrogen and substituted carbon, respectively. The nitrogen in the "side chain" was also protonated to allow for quaternary nitrogen inversion products. When nitrogen substitution is symmetrical, building the inversion product is not necessary. Therefore, unsymmetrically substituted compounds **11, 13, 14, 15** of Table 1 were built with both the *R* and *S* configurations about the side chain nitrogen. The lowest energy ring conformers were chosen as the only ring conformers when building the side chain portion of the compounds. In compounds **14, 18,** and **19**, the six-membered rings in the "side chain" are assumed to be the chair form, while in compound **15** the seven-membered ring is assumed to be in the twist-chair form. The seven-membered

Table 2. Compounds Are Substituted Tricyclic Pyridobenzodiazepinones Not Used in the MSA–3D-QSAR Analysis[a]



no.	X	IC_{50}, nM M2	M3	M3/M2
22	9-Cl	40	1500	38
23	8-Cl	100	4000	40
24	9-CH_3	70	2000	29
25	8-CH_3	30	1000	33
26	8-C_2H_5	200	2000	10
27	8-Br	150	4000	27
2	H	140	6000	43

[a] Their binding affinities to cardiac and glandular tissue are given as IC_{50} (nM). Compound **2** is also listed for reference.

central ring was chosen as the twist-chair form because it is the lowest energy form and mimics the six-membered ring.

The remaining substructures of the compounds were built with fragments from the CHEMLAB-II library (protonated piperidine chair ring, methane, ethane, amine function) and joined with standard bond lengths and angles using the BUILD option in the CHEMLAB-II molecular modeling package. Although many low-energy conformer states have been reported for seven- and five-membered rings,[7] compounds **8** and **9** were built such that the common overlap volume was maximized with respect to compound **15** which turns out to be the best reference compound. These structures were then optimized using MMFF, an extended MM2 force field.[6,8] It was assumed that if the compound could adopt a best fit conformation with the reference compound, then the corresponding ring conformation would be adopted.

4. Conformational Analysis. The SCAN option of CHEMLAB-II was used to perform a fixed valence geometry conformational energy scan at 30° increments for all torsion angles except for 180° increments for τ. The torsion angles are defined in Figure 2. The three principal bonds used to define each torsion angle are represented by thickened lines. The central bond of each torsion angle is defined by one of the letters τ, ϕ, θ, or ψ. The zero-angle values refer to *antiperiplanar* conformations about the corresponding principal bonds.

For the most flexible molecules, the decomposition–recomposition method[9] was used to reduce the dimensionality of the systematic conformational search. Some analogs were decomposed and recomposed as shown in Figure 3. The analogs treated in this fashion were **3**, **4**, **5**, and **16** of Table 1.

A fixed valence geometry molecular mechanics force-field composed of a dispersion/steric, electrostatic, and, where applicable, hydrogen-bonding contribution was used to estimate the conformational energy. The nonbonded steric MMFF parameters, which are an extended set of MM2 parameters of Allinger,[8] were used to compute the dispersion/steric interactions. The electrostatic interactions were calculated by using a Coulombic monopole representation with a molecular dielectric of 3.5 and atomic charges calculated by the CNDO/2 method.[10] The hydrogen bonding potential developed by Hopfinger[11] was used. Additionally, the hydration shell model of Hopfinger[11] was used to evaluate the influence of aqueous solvation upon free-space conformational energy profiles. The apparent global energy minimum was used to define the relative stability of each conformational state sampled.

5. Selection of Conformations. Molecular comparisons were carried out on the basis of relative feature representations. This requires the assignment of a shape reference compound. In addition, it is necessary to generate the set of QSAR conformations that are used in the construction of trial MSA–3D-QSARs. One of the QSAR conformations of the set

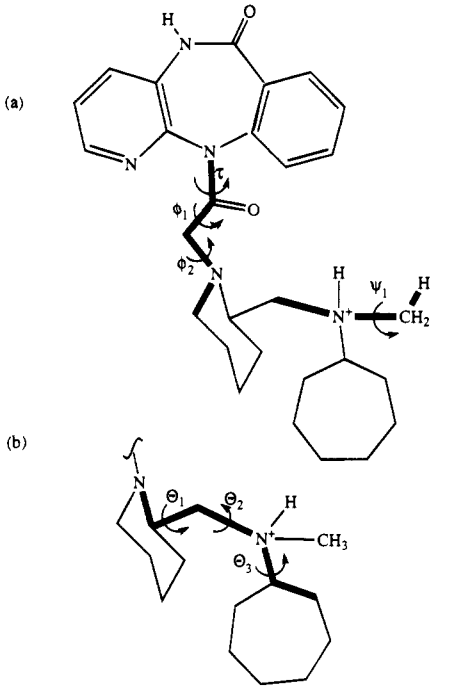


Figure 2. The definition of torsion angles for compound **15**. The three principal bonds used to define each torsion angle are darkened and the central bond of each torsion angle is defined by one of the Greek letters τ, ϕ, θ, or ψ. The zero-angle values refer to antiperiplanar conformations about the principal bonds. In part a, the torsion angle τ is drawn as the 0° conformation, and the three bonds defining the ϕ and ψ torsion angles are illustrated. In part b, the three bonds defining each θ torsion angle are shown.

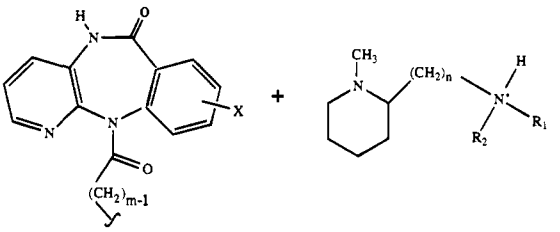


Figure 3. The decomposition–recomposition technique was used to reduce the dimensionality of a conformational search. Analogs **3**, **4**, **5**, and **16** were decomposed–recomposed into fragments as drawn. R_1 and R_2 are given in Table 1. The fragment on the right-hand side is the most conformationally flexible.

of v corresponds to the action conformation, or in this application, the *active* conformation, and optimizes the relationships expressed by eqs 1 and 2.[1] The QSAR conformation of an analog, in turn, refers to that conformation whose properties, relative to the shape reference compound, are most similar to the reference compound in its trial reference conformation. For the MSA 3D-QSAR models the set of candidate QSAR conformations for each analog was generated at 15° resolution grids in torsion angle space, but were generally limited to those states having conformational energies within 6 kcal/mol of the apparent global minimum energy conformation. Two conformations of compound **8** at 7–8 kcal/mol above the apparent global minimum were permitted in the analysis in order to include equivalent five-membered ring inversions for this analog.

Candidate shape reference conformations were considered at energies less than 6 kcal/mol relative to the apparent global energy conformation. The shape reference compound in a specific conformation is designated (v, α) where v is the number of the shape reference compound and α is the index number used to identify the conformer. The set of v considered are given by the compound numbers given in Tables **1** and **2**.

6. Molecular Alignment. The criterion for molecular alignment was to place three atoms in the tricycle identically

upon each other. The three atoms selected were the tricyclic nitrogen N1 of Figure 1 and the two adjacent ring carbons. This criterion was selected because the tricycle is a large, invariant part of each molecule in the series.

7. Measures of Steric Molecular Shape (SMS). The COSV (V_o)[12] was used to construct the SMS tensor, $\mathbf{V}_{u,v}$.[1]

8. Measures of Molecular Field (MF). No MF features were used in constructing the MSA−3D-QSAR.

9. Other Molecular Descriptions ($\mathbf{H}_p$). In addition to COSV, the other physicochemical features belonging to the $\mathbf{H}_p$ tensor, $\mathbf{H}_{u,v}$,[1] that were considered in the construction of the MSA−3D-QSARs are described below.

(a) Lipophilicity: LOGP. The relative lipophilicities of the analogs, as measured by the 1-octanol/water partition coefficient, log P, were determined using the program, CLOGP.[13] The descriptor LOGP is the lipophilicity of the entire molecule.

Scheme for Estimating a Lipophilic Contribution of a Part of the Molecule. An examination of the SAR in Table 1 suggested that the lengths of the lipophilic chains directly correlate with increased activity and provide an indication that hydrophilic additions (compound **19** is the only example) are associated with a decrease in activity. This suggested that there might be a receptor-specific lipophilic area with which the ligands interact. Two descriptors were constructed to reflect both the lipophilic character of the substructures, as well as the solvent accessible surface area of the substructures which might interact at a specific hydrophobic area of the receptor.

The two descriptors, the solvent accessible surface area, SASA, and the sum of the π constants, $\sum\pi$, were developed based upon certain assumptions. The premises were that (1) the location of the lipophilic pocket can be based upon the relative location of the atoms in the side chain (see Figure 1) of the most active compound in the series and that (2) the thermodynamic character of the atoms can be estimated by summing up atomic π constants and/or SASA contributions of atoms near that lipophilic pocket.

On the basis of the two assumptions above, the following procedure to estimate the relative local lipophilic effect of the ligand was devised.

i. One at a time, all of the reference compounds, v, in their reference conformations were selected.

ii. One at a time, all comparison conformations from each of the nonreference compounds, u, were chosen.

iii. One pair at a time, the reference, v, and comparison, u, molecules were superimposed.

iv. A sphere was generated about the geometric center of the seven-membered ring of the side chain of the reference compound, v,[15] where the radius was varied in increments of 0.5 Å between 2.0 and 6.0 Å.

v. Every atom which had at least its geometric center within the sphere was assigned the appropriate group π constant and SASA. The sums of these two descriptors, over the number of atoms, was computed for each analogue.

The notation for the $\sum\pi$ term is $\sum\pi(n)r$ Å where n indicates the value of the π constant term used for the side-chain nitrogen and r indicates the radius of the sphere. Likewise, the notation for the summed contribution SASA is $\sum$SASA(R) Å, where R indicates the radius of the sphere.

(b) Nonallowed Space (NAS). The conformational profiles of compounds **2**, **9**, and **16** indicated that the required reference shape of the most active compounds could also be adopted at low energies by inactive analogs. Hence, the loss in biological activity is not due to intramolecular conformational restrictions. This led to considering whether specific intermolecular interactions were responsible for the differences in activity. Further, it seemed desirable to quantify the effects of the intermolecular interactions (both positive and negative) in order to develop MSA−3D-QSARs and structural hypotheses about the receptor.

Indicator variables for nonallowed space occupancy for compounds **2**, **9**, and **16** were used in the initial stages of the MSA−3D-QSAR analysis. Differences in occupied space between the shape reference compound, v, and compounds **2**, **9**, and **16** were located for those indicator variables which were found to be significant in trial MSA−3D-QSARs. The nonallowed space indicator variable is termed NAS-#, where NAS stands for nonallowed space and # refers to the compound to which the NAS applies.

(c) Relative Energy, ΔE. The relative conformational stability of a compound is defined as the difference in energy between a particular conformation and the apparent global minimum energy conformation. The energy difference, ΔE, was used as a $\mathbf{H}_p$ feature. The ΔE values were combined with the COSV (V_o) to construct the shape commonality index, I_c.[1,14] I_c was then used in the $\mathbf{V}_{u,v}$ tensor.

(d) N-to-C Internal Distance, DistNC. The distance between the central tricyclic nitrogen, N1 (see Figure 1), and the side-chain carbon which is most distant from that nitrogen (the N-to-C internal distance) was used as a $\mathbf{H}_p$ feature. The purpose of this feature is to describe how far a compound might extend into a binding site.

(e) Nchg. The binding of a ligand to a receptor may be due, in part, to charge density and its distribution over the ligand. Hence, the computed charge on the side-chain nitrogen was used as a feature.

10. Partial Least Square Regression (PLS). PLS regression has only recently been developed and discussed in the chemical literature.[3,20,21] It is a generalization of multiple linear regression which yields a robust regression model in the cases in which the number of independent variables approaches or exceeds the number of compounds in the QSAR study. It also avoids the matrix singularity problem that occurs with multiple linear regression when there is collinearity in the independent variable matrix. Both are very common problems in QSAR model development.

With PLS, the dependent variables, y, are modeled as

$$y_{ij} = \mathbf{y}_j + \sum_{i=1}^{A} t_{ia} p_{aj} + e_{ij} \tag{3}$$

Here, i is the compound index and j is the biological activity index. $\mathbf{y}_j$ is the average of the biological activity vector and t is a "latent" variable with p the PLS weight or loading for t. A is the number of significant latent variables or PLS components required to reduce the residuals, e_{ij}, to an acceptable value.

The independent variables, x, are modeled as

$$x_{ik} = \mathbf{x}_k + \sum_{i=1}^{A} q_{ia} u_{ak} + e_{ik} \tag{4}$$

The index, k, is independent variable dependent. q is a latent variable corresponding to the independent variables, u is the corresponding loading and the e_{ik}'s are the residuals. The t's and q's are computed (1) along the axes of greatest variance in $\mathbf{X}$ and $\mathbf{Y}$ and (2) to make the e's as small as possible in the least squares sense. The t's and u's are related through the inner relation

$$\hat{u} = b^* t \tag{5}$$

The computation of PLS models has been discussed.[20]

Determination of the number of PLS components, A, sufficient to model the data is of critical importance to 3D-QSAR studies. PLS models are optimized for the estimation of the y's using least squares and a cross-validation method. Cross-validation begins with $A = 0$ or estimation of y from the mean. A preselected number of data points is deleted, and the deleted data are estimated from a model derived from the mean of the retained points. This is continued until all points are deleted and predicted once and only once. This results in a predicted sum of squares, or PRESS_0 where the subscript refers to a model with $A = 0$ for the deleted points. A model with $A = 1$ is derived and the process of deletion is applied as before. PRESS_1, where the subscript refers to $A = 1$, is computed and the ratio, after correcting for differences in degress of freedom, PRESS_1:PRESS_0 is computed. If the ratio is less than 1 the component is retained and the process is continued until the ratio becomes greater than 1. At this point the process is stopped and the model is adopted. The x and y variance explained by PLS models is also a useful diagnostic.

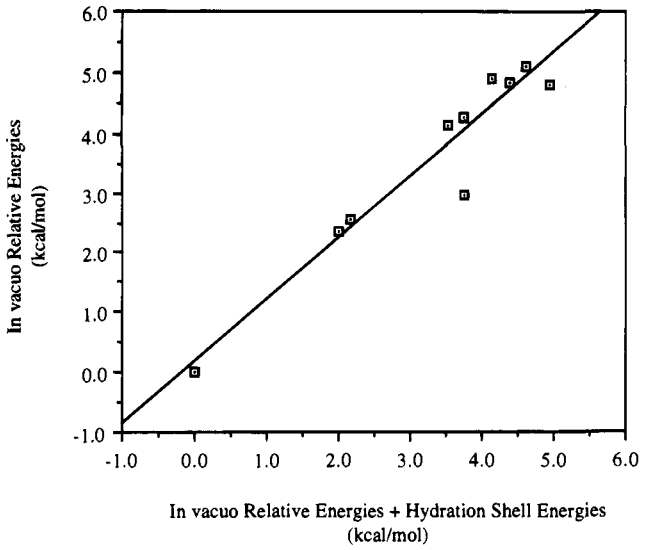


Figure 4. *In vacuo* relative energies (Y) versus the sum of *in vacuo* relative energies and hydration shell relative energies (X) for conformations of compound **15**. The effect of solvent on the conformational ordering is minimal as evidenced by the equation of fit, $y = 0.17 + 1.03x$, with $R^2 = 0.923$ where y is the *in vacuo* relative energies and x is the *in vacuo* relative energies and hydration shell relative energies.

The variance of each, explained by each PLS component extracted, is computed. The p's and q's are normalized to length 1 so the sum of squares of the loadings is the relative variance explained by each component. This result can be used to assess the significance of the variables.

PLS models are also evaluated with a cross-validation technique which computes R^2, the correlation coefficient. Here, after a model has been selected for validation, each point is left out one-at-a-time and predicted. The R^2 of the left out data is computed. This statistic, which is sensitive to outliers, is generally less than the conventional R^2, and a value in the range 0.60 or greater is considered significant. In the conformation selection process, the variables were not scaled. In the 3D-QSAR, they were autoscaled. The PLS analyses reported here were carried out with the UNIPALS software.[20,21]

Results

The hydration shell model of Hopfinger[11] was used to estimate the effect of aqueous solvation on the conformational energetics. The difference in the conformational profiles of Table 1 under *in vacuo* and estimated solvated conditions for the shape reference molecule, number **15**, was investigated. When the conformational scans were performed under *in vacuo* or aqueous solvation conditions, the relative conformational energy difference between *in vacuo* and solvent conditions for any one conformer was not more than 0.8 kcal/mol. The stability ranking of the conformers is similar for both environments (see Figure 4). Overall, the aqueous solvation effect on conformation is minimal for this set of compounds.

To test the fidelity of the force field chosen for the conformational analysis, a check was made to see if the two reported[15] crystal conformations for pirenzipine are low-energy conformers. Both of the crystal conformations are similar to conformations generated by the fixed valence geometry scan. Most of the torsion angle differences between the crystal structures and the fixed valence geometry structures are within 10°. However, between the pirenzipine monohydrochloride crystal structure and closest computed (fixed-valence geometry scan) structure, the ϕ_2 torsion angle differs by 23°. This crystal structure conformation for pirenzipine monohydrochloride is computed to be about 4 kcal/mol higher in energy than the nearest fixed-valence energy structure, and the computed energy is in agreement with past MNDO calculations.[5] However, small differences in ϕ_2 rapidly reduce the computed conformational energy of the pirenzipine monohydrochloride structure, such that a difference in ϕ_2 of 4° from the crystal structure conformation reduces the difference in energy to less than 2 kcal/mol. Another 3° in ϕ_2 reduces the energy to less than 1 kcal/mol. The other crystal structure conformation, for pirenzipine dihydrochloride, is quite similar in conformation, and the corresponding energy difference is much smaller (0.3 kcal/mol higher).

Table 3. Number of Conformers with Relative Energy below 6 kcal/mol Cutoff by Fixed Valence Conformational Scan

compd v	number of conformers less than 6 kcal/mol (not minima)	compd v	number of conformers less than 6 kcal/mol (not minima)
1	9	11	110
2	61	12	359
3	151	13	264
4	87	14	30
5	189	15	12
6	53	16	706
7	33	17	26
8	33	18	12
9	41	19	13
10	35		

Twenty pyridobenzodiazepinone inhibitors of M2 and M3 receptors, given in Table 1, were analyzed within the framework of eq 1 in order to construct MSA–3D-QSARs. No MF physicochemical descriptors were considered, and only those $\mathbf{H}_p$ features described in the Methods section were employed in building $\mathbf{H}_{u,v}$. Multiple conformations, α, and shape reference compounds, v, were considered. However, only one v was used at a time in constructing a series of $\mathbf{V}_{u,v}$, based either on the COSV, V_o, or the shape commonality index, I_c.[1,14] All trial MSA–3D-QSARs were restricted to a single alignment involving the tricyclic N1 and two adjacent ring carbons (see Figure 1).

The PLS method of determining the assignment of conformations for the MSA–3D-QSARs[1] was carried out using the set of conformers, {α}, of each compound satisfying the $\Delta E = 6$ kcal/mol energy constraint. The number of conformations satisfying the energy constraint for each analogue, based on fixed valence geometry scans, are given in Table 3. The specific conformations, in terms of the torsion angle definitions in Figure 2, are given in Table 4 for compound **15**, which turns out to be the preferred shape reference compound, v. Compounds **2, 9, 14, 15,** and **16** were considered as candidates for v based upon their M2, M3, M2/M3, and/or structural features. All conformations of compound **15** which satisfy the energy constraints are also energetically acceptable to both active and inactive analogs. The most active compound at both the M2 and M3 receptors is compound **15**. Compound **2** has reasonable activity at the M2 receptor and considerably poorer activity at the M3 receptor, making it the most M2/M3 selective of the analogs in Table 1. The conformational searches, and subsequent pairwise molecular superpositions, indicate that compound **2** can overlap compound **15** quite well. There are seven cases in which compound **2** cannot adopt the conformational state of compound **15** at an energy expenditure equal to, or lower, than that of compound **15** (see Table 5). However, in two of

Table 4. Torsion Angle Values and Corresponding Relative Energies of 12 Proposed Shape Reference Conformations of Compound **15** Which Have Relative Energies below the 6 kcal/mol Cutoff

shape reference conformation (v,α)	relative energy, kcal/mol	torsion angle,[a] deg						
		τ	ϕ_1	ϕ_2	θ_1	θ_2	θ_3	ψ
15,1	4.3	0	−120	0	60	−30	90	0
15,2	5.1	0	30	0	60	−30	90	0
15,3	4.8	180	30	0	60	−30	90	0
15,4	4.9	0	60	0	60	−30	90	0
15,5	2.7	0	30	30	60	−30	90	0
15,6	2.6	180	30	30	60	−30	90	0
15,7	4.1	0	60	30	60	−30	90	0
15,8	3.0	0	30	60	60	−30	90	0
15,9	0.0	180	30	60	60	−30	90	0
15,10	4.8	0	60	60	60	−30	90	0
15,11	5.8	0	−120	0	90	90	0	0
15,12	5.9	180	30	0	90	90	0	0

[a] Initial torsion angles shown in Figure 5.

Table 5. A Comparison of the Relative Energies of the Active (**14**, **15**) Compounds and the Less Active (**2**, **9**, **16**) Compounds

shape reference	relative energies, ΔE, kcal/mol active compounds			differences in relative energies, ΔE, kcal/mol less active compounds		
	15	**14**	**15** or **14** (largest ΔE between **15** and **14**)	**2** ΔE difference between **2** and (**15** or **14**)	**9** (ΔE difference between **9** and (**15** or **14**))	**16** ΔE difference between **16** and (**15** or **14**)
15,1	4.26	4.34	4.34	0.78	5.2	−0.50
15,2	5.11	4.97	5.11	1.26	−2.11	0.06
15,3	4.82	4.72	4.82	0.93	0.54	0.12
15,4	4.90	5.41	5.41	0.44	−5.41	−0.80
15,5	2.34	2.93	2.93	−0.16	−2.10	0.49
15,6	2.55	2.93	2.93	−0.46	−0.10	−2.04
15,7	4.11	3.89	4.11	−0.08	−0.26	−2.22
15,8	2.97	2.59	2.97	−1.72	−2.46	−0.85
15,9	0.00	3.05	3.05	−1.32	0.38	−1.20
15,10	4.80	2.59	4.80	1.14	−0.53	0.95
15,11	5.83	3.30	5.83	−1.59	3.8	−1.89
15,12	5.90	4.72	5.90	−0.09	−0.54	−0.91

these cases compound **2** can adopt conformations of equal, or lower, relative energy than that of the second most active compound, **14**. For the remaining five cases, the difference in relative energies is at most 1.3 kcal (see Table 5).

The loss in intramolecular energy by some of the less active compounds, in order to adopt a reference conformation, is sufficient to explain the corresponding loss in activity. However, differences in the conformational energies between the least active compounds (**2**, **9**, **16**) and the two active reference candidate compounds, **14** and **15**, are insufficient to account for the loss in enzyme inhibition potency when each of the compounds adopts one common reference conformation. Also, the selectivity between M2 and M3 cannot be accounted for solely by internal conformational energies. Overall, the reference conformation yielding features most in line with the trend in selectivity is (v,α) = (**23**,**3**). This is the only reference state which has positive internal conformational energy differences for all three of the less active compounds **2**, **9**, and **16**.

A similar analysis has been performed for compounds **9** and **16**, and the results are also presented in Table 5. For compound **9**, the results are different than those for compound **2**. Compound **9** can adopt maximum COSVs at energies equal to, or less than, that of the most active compounds in the series. There are four exceptions. For shape reference (v,α) = (**15**,**9**), the difference in relative energies between compound **14** and compound **9** is about 0.5 kcal/mol. A more drastic difference is for shape reference (**15**,**1**) where the difference is over 5 kcal/mol. Compound (**15**,**11**) is the inverted analog of (**15**,**1**) where the energy difference is over 5 kcal/mol. Compound (**15**,**11**) is the inverted analog of (**15**,**1**) and is similarly high at a difference of 3.8 kcal/mol.

Compound **16** is similar to compound **2**, where the difference in relative energies ranges up to about 1.0 kcal/mol. The only shape reference conformations for which a large (>1.25 kcal/mol) positive difference in relative energies occurs are (**15**,**1**) and (**15**,**11**). This difference in energy suggests that (**15**,**1**) (and perhaps (**15**,**11**)) are important candidates for the active conformation since **9** cannot easily adopt these conformations, and **9** is one of the least active compounds in the series.

The optimum MSA−3D-QSAR was found with compound **15** as the shape reference compound, v. The preferred shape reference compound, from non-PLS MSA investigations,[1,2,12,16−19] has usually been either the most active or the largest analog, and compound **15** is both. If the shape reference compound is the largest compound, then the volume of smaller analogs will be fully accounted for provided the conformational profiles are similar. When the most active compound is selected, it is known that this compound has both the pharmacophoric features necessary for activity and a conformational profile which can adopt the active conformation. The set of trial reference conformations of compound **15** for the MSA−3D-QSAR models are reported in Table 4.

The NAS is defined relative to v in its shape reference conformation. A typical example of part of the defined NAS is given for (**15**,**8**) in Figure 5. The circled area indicates the proposed NAS for each compound (**2**, **9**, **16**) as labeled. The NAS for compound **2** occupies only one region of space with respect to v, that volume

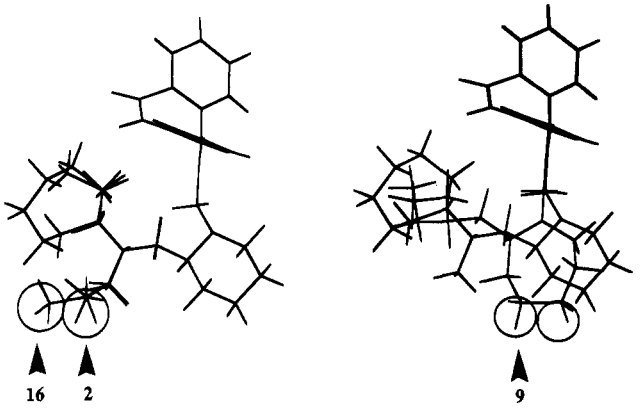


Figure 5. Locations of part of the NAS for (**15,8**). The circled area indicates the NAS for each compound (**2, 9, 16**) as labeled.

adjacent to the terminal end of the side-chain methyl as illustrated in Figure 5. The central ring of compound **9** can fit the shape reference compound in two similar conformations; the NAS-9 site located as illustrated in Figure 5. Finally, compound **16** does not occupy common space with the reference compound in the same general area as NAS-2 as illustrated in Figure 5. The NAS partially defines the shape of the receptor pocket because it indicates receptor occupied space.

Two classes of MSA–3D-QSARs, based upon PLS selection of conformational states, were developed. The two classifications arise because it was not clear how to assign the π constant term for the protonated nitrogen in the side chain. It was decided that MSA–3D-QSARs would be developed for two limiting cases regarding the π constant for the protonated nitrogen. In the first case the $\Sigma\pi$ feature was computed by neglecting the nitrogen term, $\pi_N = 0.0$, model 1. In the second case the protonated nitrogen for salts[22] was considered, $\pi_N = -3.8$, model 2. The range between these two $\Sigma\pi$ features bracket a considerable π value range (−3.8, 0) and, therefore, afford a reasonable basis to evaluate the role of the nitrogen π constant on the form and quality of the corresponding MSA–3D-QSARs.

The correlation coefficient, R^2, for the MSA–3D-QSARs, in which the nitrogen π constant is neglected, $\pi_N = 0$, model 1, with the highest PLS regression correlation coefficients are listed in Table 6. The R^2 terms for these MSA–3D-QSARs having the five best shape reference conformations for compound **15** are listed under the "PLS selected conformations" column in Table 6. The number in parentheses next to each R^2 term is the number of significant cross-validated components used to construct that MSA–3D-QSAR model. For example, the R^2 for the MSA–3D-QSAR for (**15,3**) for the M2 activity is 0.85. Three components were significant by cross-validation, and this is indicated by the "(3)" next to "85" in Table 6. The NAS was evaluated for each of the PLS-selected conformers. Conformers which violated NAS space were replaced with the highest ranked conformer which did not violate the proposed NAS space. New QSARs were developed with data from these conformers. The R^2 terms for these MSA–3D-QSARs are reported under the "NAS adjusted conformations" column in Table 6. The NAS conformer replacements improved the (**15,6**) and (**15,8**) models by explaining an additional 10% or more of the M2 activity. However, when the NAS conformers replacements are made to the (**15,9**) model, the resulting MSA–3D-QSAR explains 60% less M2 variance than the MSA–3D-QSAR composed of the original PLS selected conformers.

The importance of a descriptor is related to the relative variance explained by that descriptor in each PLS component. Plots of the loadings of the NAS adjusted conformations for the M2 activity for the (**15,8**) MSA–3D-QSAR are given in Figure 6. The columns are labeled with the names of the features corresponding to each column in the bar graph. The sign of the loading is the direction of correlation and the height of the bar indicates the square root of the variance explained by that descriptor in that component. The indicated (*) bars in the figure are those variables with the highest amount of variance explained by each feature (for example, $\Sigma\pi$). In component 1, with the exception of variables 14–16, the first 20 variables are essentially equivalent and not correlated with 21–26.

The starred (*) features for each of the five shape reference conformers of compound **15** with the highest correlation coefficients, using the conformers adjusted for NAS, are listed in Table 7. By way of example, the first component for (**15,8**) is optimally related to V_o, $\Sigma\pi$-($N = 0.0$) 4.5 Å, $\Sigma\pi(N = 0.0)$ 5.0 Å, $\Sigma\pi(N = 0.0)$ 5.5 Å, and ΣSASA 4.0 Å for the M2 activity. The second component is mainly composed of ΣSASA 3.0 Å, DistNC, ΔE, and NAS-9. The third component contains the previously unselected NAS-16 term which is also negatively correlated with activity. A range of features is given if the loadings are about the same within a type of descriptor. For example, if for the $\Sigma\pi$ feature, the loadings at three radii are all about the same, then $\Sigma\pi$ terms for all three radii are used to create the reduced feature data matrix.

In general, for each of the five shape reference conformers of compound **15**, the first component of each MSA–3D-QSAR is related to the volume/shape of the

Table 6. PLS Regression Results Using $\Sigma\pi$ ($N = 0.0$) Data Matrices for Reference Conformations with the Highest R^2

	cross-validated coefficient of correlation squared (×100%) (number of cross-validated components in parentheses)									
	PLS selected conformations		NAS adjusted conformations		reduced features		reduced features no. 2		additional features in reduced features no. 2	
reference conformation (α,v)	M2	M3	M2	M3	M2	M3	M2	M3	M2	M3
15,3	85 (3)	95 (4)	84 (3)	95 (4)						
15,6	78 (2)	96 (4)	88[a] (3)	97 (5)	71 (2)	84 (2)	88 (3)	94 (3)	$\Sigma\pi$ 5.5 Å	$\Sigma\pi$ 5.5 Å, NAS-10
15,7	97 (6)	96 (4)	93 (4)	91 (2)						
15,8	77 (2)	88 (2)	93 (3)	91 (2)	91 (3)	92 (3)[a]				
15,9	91 (5)	95 (5)	34 (1)	88 (2)						
	number of features in data matrices									
15,6	26	26	26	26	14	14	15	16		
15,8	26	26	26	26	10	9				

[a] Includes NAS-10 and NAS-24.

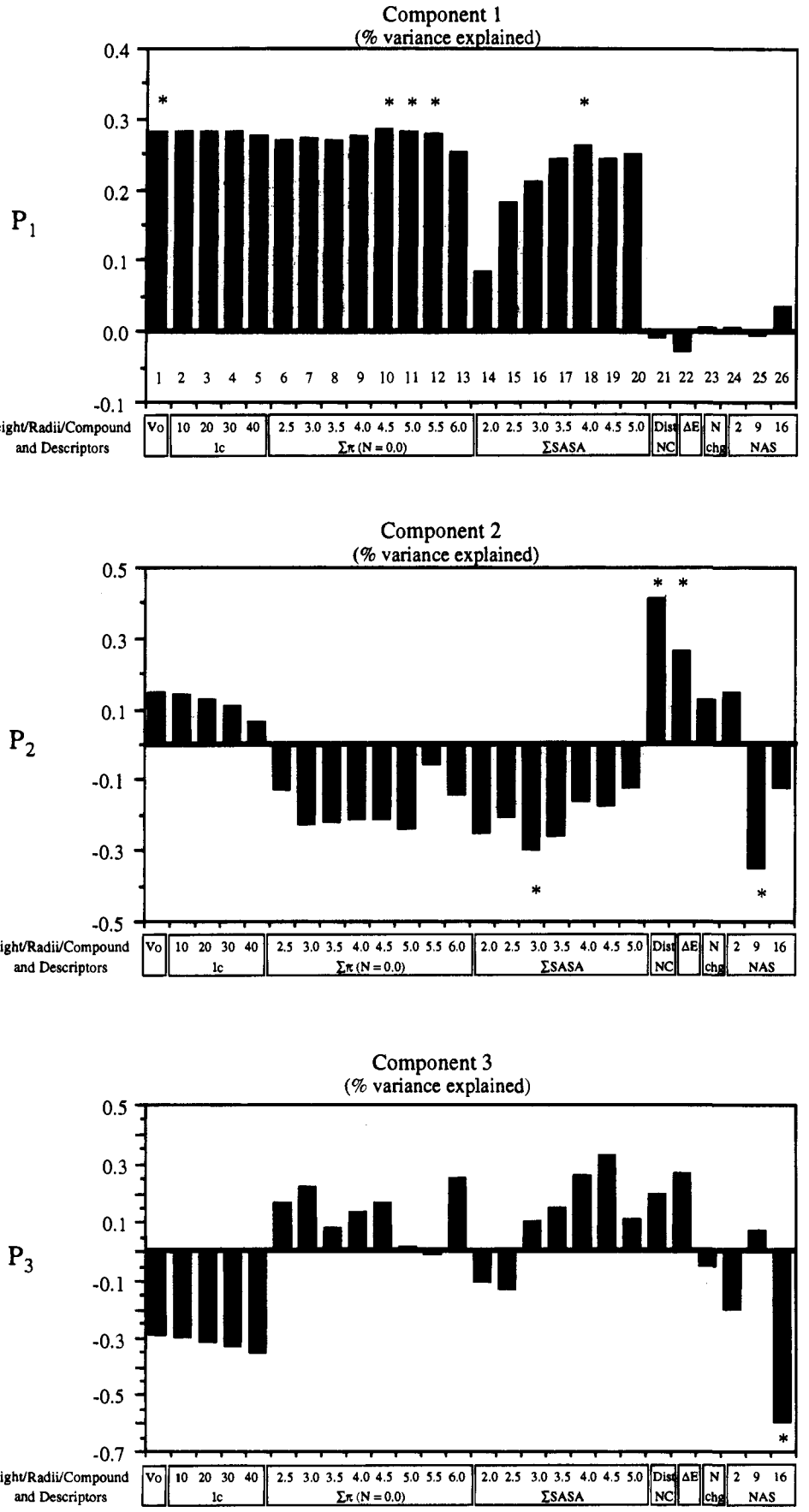


Figure 6. A plot of loadings for the NAS adjusted conformations for the M2 activity of the (**15,8**) MSA–3D-QSAR.

compound. The second component is a consistent mixture of ΣSASA, NAS, and other features. The internal distance measure, DistNC, and the NAS terms constitute the bulk of the remaining terms. A feature of the third component is the repetition of previously selected physicochemical features, such as V_o, NAS, and DistNC terms. The presence of significant fourth and fifth components is relatively infrequent (30% of the cases), and they exhibit no apparent pattern of feature distribution. For the complete data matrix (NAS ad-

Table 7. Variables Most Highly Correlated with Activity for Some $\Sigma\pi$ (N = 0.0) Data Matrices

		variables with the largest squared loading term													
		component 1				component 2			component 3			component 4			
conformation	receptor	shape	$\Sigma\pi$ R (Å)	ΣSASA R (Å)	other	ΣSASA R (Å)	other	NAS no.	shape	$\Sigma\pi$ R (Å)	other	NAS no.	ΣSASA R (Å)	other	NAS no.
15,3	M2	V_o	3.0−5.0	4.5−5.0		2.0	DistNC, ΔE	9, 16			ΔE	2, 9, 16	5.0		
	M3	V_o	3.0−4.0, 5.0	4.5		2.0	DistNC	9, 16			DistNC, ΔE				
15,6	M2	V_o	2.5−4.0, 5.0	3.5−5.0		2.5	DistNC	9, 16			ΔE	16			
	M3	V_o	2.5−4.0	3.5, 5.0		2.0	DistNC	2, 9, 16	V_o		DistNC		2.0, 2.5	ΔE, Nchg	
15,7	M2	V_o	4.5−5.5	3.5			DistNC, ΔE, Nchg	9, 16	V_o	4.0		16	2.0	DistNC	16
	M3	V_o	4.5	3.5, 5.0		2.0	DistNC, ΔE, Nchg	9, 16							
15,8	M2	V_o	4.5−5.5	4.0		3.0	DistNC, ΔE	9				16			
	M3	V_o	4.5	4.0		2.0	DistNC, ΔE	9							
15,9	M2	I_c (w80)	3.5, 5.0, 5.5	4.0, 5.0	ΔE										
	M3	I_c (w80)	3.5, 4.5	4.5	ΔE		DistNC	9, 16							

justed), the range in percent of dependent variable variance explained by each component is approximately as follows: (1) 30−50%, (2) 30−50%, (3) 3−10%, and (4) 2−10%.

An examination of Table 4 reveals that the shape reference conformers (**15,3**), (**15,6**), and (**15,9**) occupy similar space. Likewise, the shape reference conformers (**15,7**) and (**15,8**) are similar in conformation. Each set of related conformers form a family of conformations which are similar to each other. In Table 6, the shape reference conformers with the highest cross-validated R^2 (NAS adjusted conformers) from each of these two families are reported. The conformer with the highest R^2 values in the first family is (**15,6**). In the second family, (**15,7**) and (**15,8**) are equivalent within the highest R^2 value. For (**15,6**) and (**15,8**), a reduced data matrix was constructed using only those features with large absolute values for the loadings as given in Table 7. Again, these are features which are selected as most highly correlated with activity. PLS regression was performed on the reduced-descriptor data matrices. The cross-validated R^2 values are reported under the "reduced features" column in Table 6. The number of features used in each PLS regression is listed in the lower portion of Table 6.

The MSA−3D-QSARs with the reduced feature data matrices explain less variance than the complete feature data matrix for (**15,6**) and about the same amount as (**15,8**). The amount of M2 activity explained is 17% lower and the amount of M3 activity explained is 13% lower for (**15,6**). However, the amount of M2 activity explained is 2% lower, and the amount of M3 activity explained is 1% higher for (**15,8**). The additional variance explained by the reduced descriptor matrix for (**15,8**) is due to an additional component in the MSA−3D-QSAR. Upon examination of the features selected for (**15,8**), $\Sigma\pi(N = 0.0)$ 5.5 Å and NAS-10 were added to the (**15,6**) data matrix. The variance explained for the MSA−3D-QSAR is reported under the "reduced features no. 2" column of Table 6. The addition of these terms explain an additional 10%, or more, of M2 and M3 activity above that which is explained under the "reduced features" column. In general, the reduced features matrices (composed of 9−16 features) explain most of the variance taken into account by the entire data matrix (26 features). The MSA−3D-QSAR, using the reduced feature data matrix no. 2 for (**15,6**), explains about the same amount of variance in both the M2 and the M3 activity as the reduced feature MSA−3D-QSAR for (**15,8**).

Both of the MSA−3D-QSARs are consistent with the hypothesis that there is a secondary binding site available only to the compounds that can adopt extended conformations, that it is a lipophilic site, and that there are key intermolecular effects which are responsible for the reduced activity of compounds **2, 9**, and **16**. In particular, the length of the side chain is a significant variable in both MSA−3D-QSARs. It is difficult to postulate the size of the receptor pocket because the $\Sigma\pi$ terms are similarly correlated with activity over a wide range of radii.

The set of MSA−3D-QSARs constructed for π_N = −3.8, model 2, are, overall, very nearly the same as those having π_N = 0.0. The conformers selected by the PLS selection method are almost the same for every shape reference conformer for MSA−3D-QSAR models

Table 8. PLS Conformer Selection Dependence Based Upon $\Delta\pi$ Representation in the Data Matrix

shape reference conformation (of **15**)	compound	number of times selected conformers were different[a]
1	–	0
2	–	0
3	**4**	1
4	**3**	1
5	–	0
6	–	0
7	**13**	1
8	**13**	1
9	**16**	1
10	**16**	1
11	**16**	1
12	**9, 16**	2
		9 total cases

[a] The number of times for each compound where different conformers were selected by the PLS selection procedure based upon the $\Sigma\pi$ representation organized by shape reference conformation.

1 and 2. Even though the PLS conformers selection matrices are different (one contains $\Sigma\pi$ (N = 0.0) and the other $\Sigma\pi$ (N = −3.8)), the $\Sigma\pi$ representation makes little difference in the overall conformer selection. Out of the 216 (12 × 18) conformers selected by the PLS selection procedure, in only nine cases are the two sets of PLS selected conformers different between the two $\Sigma\pi$ models (see Table 8).

The R^2 terms for the MSA−3D-QSARs, for model 2, with the highest PLS correlation coefficients are listed in Table 9. The columns in this table are the same as Table 6 and reflect the same treatment of the data used to construct the MSA−3D-QSARs of model 1. As with the MSA−3D-QSARs of model 2, NAS conformer replacements improve the ability of the (**15,8**) shape reference state to account for additional (12%) M2 activity. NAS conformer replacements for (**15,9**) yield a MSA−3D-QSAR which explains less (15%) M2 variance than not correcting for NAS interactions. This behavior is the same as for the corresponding MSA−3D-QSAR, model 1.

The general trends found for MSA−3D-QSAR model 2 are similar to those seen in MSA−3D-QSAR model 1 (see Tables 7 and 10). Shape (V_o and I_c) terms are always found in the first component. In general, the MSA−3D-QSAR is insensitive to ΔE_u. There is little change in the correlations with activity between V_o and I_c. The exception is (**15,9**) in MSA−3D-QSAR model 2. The second component consists primarily of $\Sigma\pi$, ΣSASA, DistNC, and NAS terms. NAS terms consistently include NAS-9 and NAS-16 for M2 regressions, and usually NAS-2, NAS-9, and NAS-16 for M3 regressions. The exception is for (**15,1**), which includes only NAS-16 in component 2 and NAS-9 in component 3. The third component contains $\Sigma\pi$ terms and ΣSASA. Nchg is seen most often in this component, and some NAS terms are repeated in this component. The fourth component is significant in one-third of the cases and shows no apparent pattern of descriptor distribution. The ranges of variance in activity explained by components 1−4 are 30−70%, 20−40%, 2−10%, and 1−7%, respectively, for the NAS adjusted data matrices which contain all of the descriptors.

Conformational families are defined for MSA−3D-QSAR model 2 as for MSA−3D-QSAR model 1. The (**15,1**) reference conformation represents a third class of conformations for model 2. Using (**15,1**) as the shape reference conformer, the MSA−3D-QSAR using a reduced descriptor data matrix describes the same amount of variance as the full descriptor matrix.

In the case of (**15,8**), the reduced descriptor data matrix regression explains slightly more (3%) of the M2 variance in activity than the complete descriptor data matrix. This can be attributed to the additional component allowed by cross-validation. The amount of M3 activity explained is 4% lower for (**15,8**). However, the amount of M2 activity explained (70%) is 16% lower and the amount of M3 activity explained is 2% lower for (**15,3**). Upon examination of the descriptors selected for (**15,3**), compared to those for (**15,8**), ΔE was added to the (**15,3**) data matrix. The M2 variance included in this MSA−3D-QSAR is 76% and is listed under the "reduced descriptors no. 2" column of Table 9. Notice that the reduced data matrices for (**15,8**) contain fewer descriptors (8 and 11) than the other reduced data matrices (12 or 13 descriptors), while explaining more variance in the activity. The MSA−3D-QSAR for (**15,8**) explains 21% more variance in the M2 activity, and 2% more in the M3 activity than the MSA−3D-QSAR for (**15,3**). The MSA−3D-QSAR for (**15,1**) with reduced descriptors explains the same amount of variance as (**15,8**) for the M2 data, but does not explain the M3 activity. Overall, the reduced descriptor matrices (composed of 8−13 descriptors) explain most of the variance taken into account by the entire data matrix (27 descriptors).

Plots of predicted (using the reduced descriptor matrix) versus observed M2 and M3 activity for MSA−3D-QSAR model 2 for (**15,8**) are given in Figure 7. An illustration of a stick representation of a shape reference

Table 9. PLS Regression Results Using $\Sigma\pi$ (N = −3.8) Data Matrices for Reference Conformations with the Highest R^2

	cross-validated coefficient of correlation squared (×100%) (number of cross-validated components in parentheses)									
	PLS selected conformations		NAS adjusted conformations		reduced features		reduced features no. 2		additional features in reduced features no. 2	
reference conformation	M2	M3	M2	M3	M2	M3	M2	M3	M2	M3
15,1	90 (2)	48 (1)	90 (3)	47 (1)	90 (4)					
15,3	87 (4)	98 (5)	86 (4)	94 (4)	70 (3)	92 (4)	76 (3)		ΔE	
15,6	71 (2)	94 (4)	71 (2)	93 (3)						
15,7	81 (2)	93 (3)	81 (2)	97 (4)						
15,8	77 (2)	99 (6)	89 (2)	98 (4)	91 (3)	94 (3)				
15,9	76 (3)	92 (3)	61 (2)	87 (3)						
	number of features in data matrices									
15,1	27		27		12					
15,3	27	27	27	27	12	13	13			
15,8	27	27	27	27	8	11				

Table 10. Variables Most Highly Correlated with Activity for Some $\Sigma\pi$ ($N = -3.8$) Data Matrices

			variables with the largest squared loading term													
			component 1			component 2				component 3			component 4			
conformation	receptor	shape	$\Sigma\pi$ R (Å)	ΣSASA R (Å)	other	$\Sigma\pi$ R (Å)	ΣSASA R (Å)	other	NAS no.	$\Sigma\pi$ R (Å)	ΣSASA R (Å)	other	NAS no.	$\Sigma\pi$ R (Å)	other	NAS no.
15,1	M2	V_o		3.5, 5.0	DistNC, ΔE	4.0	4.0, 4.5, 5.0	DistNC, ΔE	9	6.0	4.5, 5.0		9, 16			
15,3	M2	V_o		3.0, 4.5, 5.0		3.0, 3.5	2.0	DistNC	9, 16	6.0		Nchg	9, 16	2.5		
	M3	V_o	5.0	3.0, 4.5, 5.0		3.0, 3.5	2.0	DistNC	2, 9, 16	5.5		Nchg	9, 16	2.5		
15,6	M2	V_o	5.5	3.0, 3.5		2.5	2.0		9, 16	5.0, 6.0			16			
	M3	V_o	5.5	3.5		3.0	2.0		2, 9, 16	4.0, 6.0	2.5					
15,7	M2	V_o		3.5		3.5	2.5, 3.0	DistNC	9, 16	3.0, 4.5, 6.0			16			
	M3	V_o	5.5	3.5, 5.0		3.0, 6.0	2.0	DistNC	2, 9, 16	4.5, 6.0		Nchg			ΔE	16
15,8	M2	V_o	5.5	4.0			3.0	DistNC, ΔE	9, 16	4.0			16			
	M3	V_o	4.5, 5.5	4.0			2.0, 3.0	DistNC, ΔE	2, 9, 16	4.0, 4.5, 6.0	2.5		9, 16	3.0	Nchg	2, 9, 16
15,9	M2	I_c (w80)	5.5	3.5, 5.0	ΔE		2.0	DistNC	9, 16							
	M3	I_c (w80)	5.0	3.5, 4.0, 5.0	ΔE	4.5	2.0	DistNC	2, 9, 16	3.0	2.0, 2.5	Nchg	2, 9, 16			

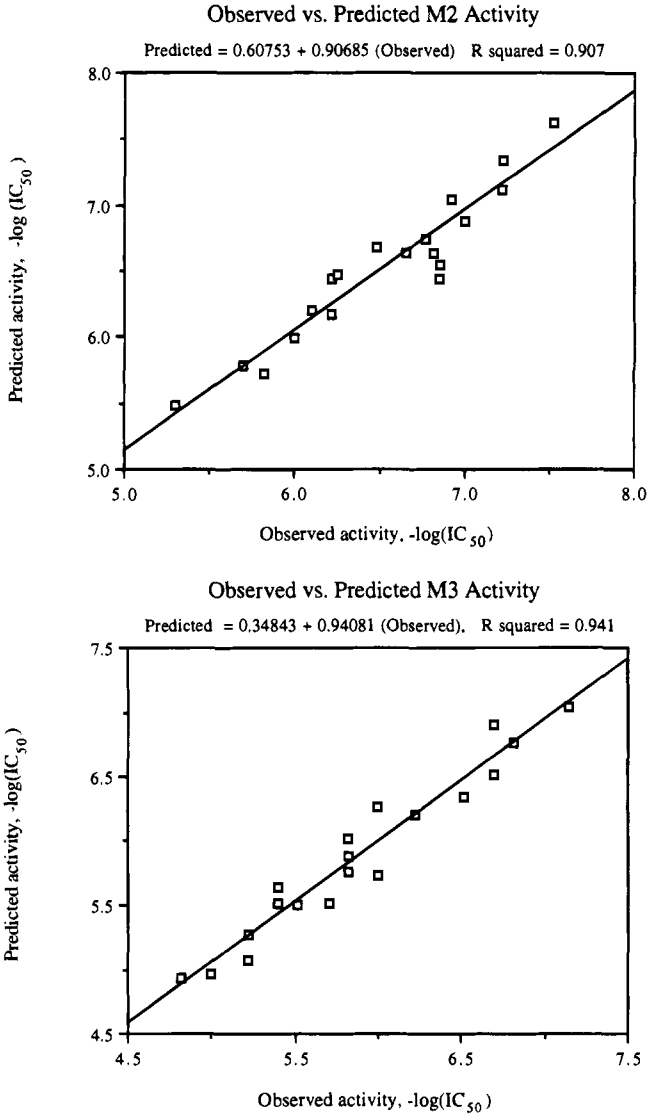


Figure 7. A plot of the observed versus predicted activity given as -log(IC_{50}) using the reduced descriptor data matrix for (**23,8**).

compound in a proposed shape reference conformation, along with the other analogs in their proposed QSAR conformations, using (**15,7**) as the references state, is given in Figure 8.

Not all analogs adopt conformations which best fit to the shape reference compound. (**15,1**) is a notable exception, with compound **9** (at conformations below 6 kcal/mol) fitting differently than illustrated in Figure 8 because ϕ_1 and ϕ_2 of the shape reference state cannot be adopted by **9** below the 6 kcal/mol cutoff.

Each of the two MSA–3D-QSAR models with respect to selection of the nitrogen π value are consistent with the hypothesis that there is a secondary binding site available only to the compounds that can adopt extended conformations, that it is a lipophilic site, and that there are also key intermolecular effects which are responsible for the reduced activity of compounds **2, 9,** and **16**. The V_o, $\Sigma\pi$, ΣSASA features are the primary independent variables in the MSA–3D-QSARs. The length of the side chain (DistNC) is also a significant variable.

"Classic" MSA–3D-QSARs were constructed using multidimensional linear regression (MLR) analysis and the most significant $\mathbf{V}_{u,v}$ and $\mathbf{H}_{u,v}$ physicochemical features identified in the PLS analyses, see Figure 6.

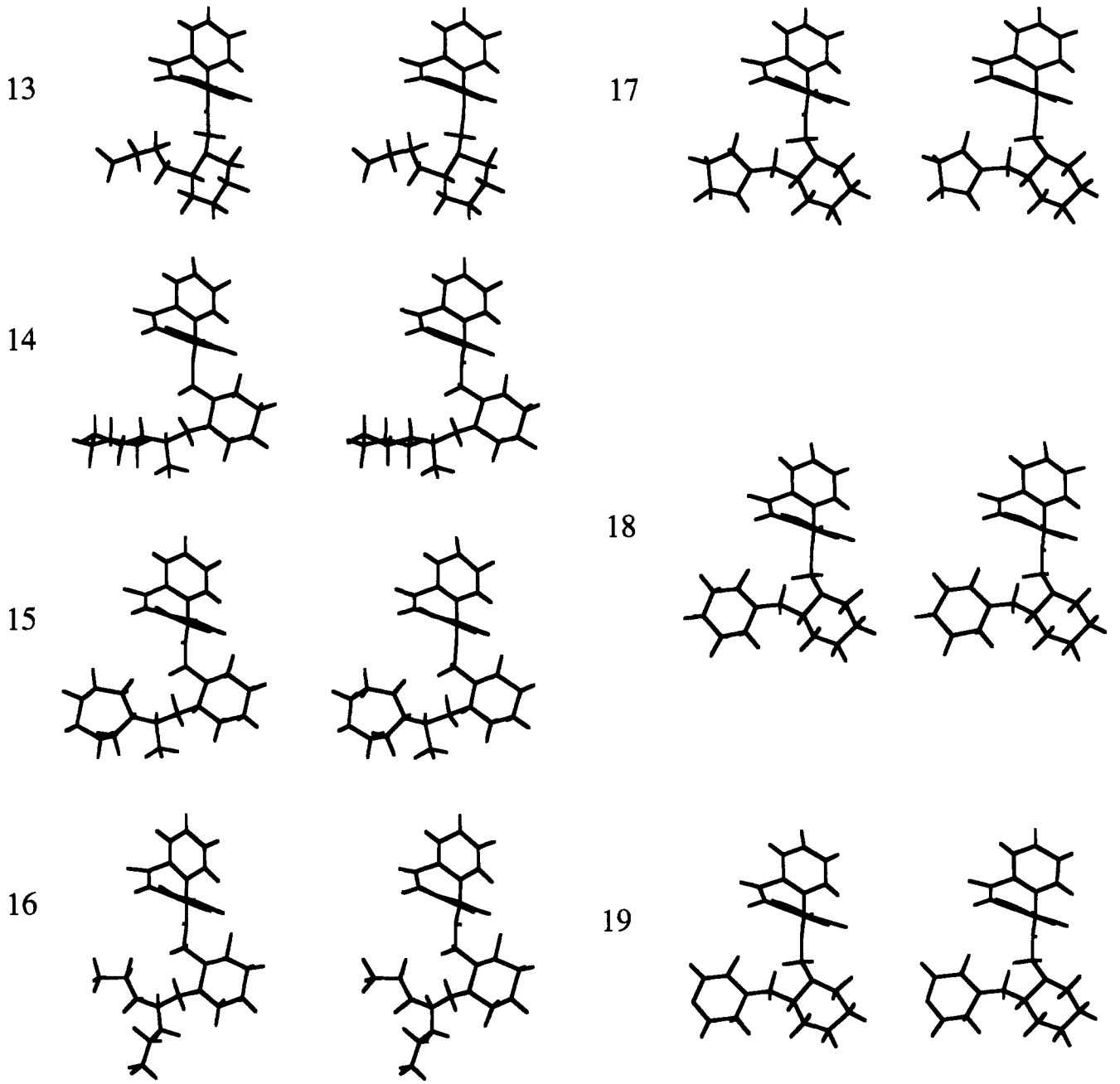


Figure 8. Stick representations of the PLS-selected, NAS-adjusted conformations of the analogues for the MSA–3D-QSAR using (**23,7**).

The optimum correlation equations found for the case in which $\pi_N = 0.0$, model 1, are

$$-\log(IC_{50})\ M2 = 0.0048(0.0018)[V_o] + 0.319(0.085)[\sum\pi(N = 0.0)\ 2.5\ \text{Å}] - 1.02(0.32)[\text{NAS-9}] - 1.13(0.32)[\text{NAS-16}] + 4.47 \quad (6)$$

$$R^2 = 0.79, \quad F = 13.1, \quad n = 19, \quad \text{SD} = 0.30, \text{ shape reference} = (\mathbf{15,8})$$

$$-\log(IC_{50})\ M3 = 0.0049(0.0014)[V_o] + 0.389(0.068)[\sum\pi(N = 0.0)\ 2.5\ \text{Å}] - 1.16(0.16)[\text{NAS-2,9,16}] + 3.78 \quad (7)$$

$$R^2 = 0.88, \quad F = 37.8, \quad n = 19, \quad \text{SD} = 0.25, \quad \text{shape reference} = (\mathbf{15,8})$$

where R^2 is the correlation coefficient squared, F is the F statistic, n is the number of compounds in the data set, and SD is the standard deviation. The terms in parentheses are the 95% confidence intervals for the respective terms.

For model 2, $\pi_N = -3.8$, the most significant MSA–3D-QSAR equations are

$$-\log(IC_{50})\ M2 = 0.0042(0.0016)[V_o] + 0.41(0.10)[\sum\pi(N = -3.8)\ 5.5\ \text{Å}] - 0.25(0.081)[\sum\pi(N = -3.8)\ 3.5\ \text{Å}] - 1.51(0.34)[\text{NAS-16}] + 4.475 \quad (8)$$

$$R^2 = 0.79, \quad F = 13.1, \quad n = 19, \quad \text{SD} = 0.30, \quad \text{shape reference} = (\mathbf{15,1})$$

$$-\log(IC_{50})\ M3 = 0.01(0.0014)[V_o] + 0.21(0.044)[\sum\pi(N = -3.8)\ 4.0\ \text{Å}] - 1.05(0.16)[\text{NAS-2,9,16}] + 2.575 \quad (9)$$

$$R^2 = 0.88, \quad F = 34.9, \quad n = 19, \quad \text{SD} = 0.26, \quad \text{shape reference} = (\mathbf{15,8})$$

There are two significant differences between eqs 6 and 8 that presumably arise because of $\pi_N = 0.0$ versus $\pi_N = -3.8$, respectively. The first difference is that the shape reference for eq 6 is (**15,8**), but (**15,1**) for eq 7. Interestingly, the "reduced features" column in Table 9 indicates a significant MSA–3D-QSAR for M2 activity using (**15,1**) for model 2, but no significant correlation using (**15,1**) for model 1 (see Table 6). The second difference between eqs 6 and 8 is that eq 6 suggests that the radius of the lipophilic center is 2.5 Å. Equation 7 indicates that the center is 5.5 Å in radius with only lipophilic atoms in the 3.5–5.5 Å shell contributing to specifying M2 activity. However, as already noted in the PLS analyses, the fidelity of a MSA–3D-QSAR is not very sensitive to the size and representation of the lipophilic center.

The major difference between eqs 6 and 7 and between eqs 8 and 9 is in the NAS terms. Only the NAS of compounds **9** and **16** are present in eqs 6 and 8 whereas the NAS of compounds **2, 9,** and **16** occur in eqs 7 and 9. Thus M2/M3 selectivity might be achieved by analogs which occupy the NAS of compounds **2** but not **9** and **16**. The PLS MSA–3D-QSARs only partially support this NAS hypothesis.

Discussion

PLS permits an estimation of the $\mathbf{T}_{u,v}$ for various M2 and M3 MSA–3D-QSARs of the pyridobenzodiazepinone inhibitors using eq 2. In essence, conformational flexibility, as expressed by α, has been explored as part of the QSAR construction process. A criticism of the use of PLS is that it makes the MSA–3D-QSARs difficult to interpret since these models are constructed in terms of components each of which contain multiple features. However, this is not a serious consideration as PLS models can easily be expressed in terms of the original data.[20] In contrast, MLR analysis is of no value when high degrees of collinearity are found among the features. In this study the most significant features found in the PLS analyses were used as features in subsequent MLR studies, see eqs 6–9. In a sense, the PLS analyses uncovered the key features, and the corresponding MLR studies allowed an evaluation, and an interpretation, of these key features. In other words, the PLS study served as a feature filter for subsequent MLR calculations. The lower R^2 values for eqs 6–9 compared to some of the PLS MSA–3D-QSARs, see Tables 6 and 9, show that PLS gives little predictive models than MLR and that the minor PLS components contribute to model stability and specificity.

The shape reference conformation with the lowest intramolecular energy, (**15,8**), is not the best shape reference compound to explain the activity of the series (see Tables 6 and 9) based on the NAS model. This demonstrates the utility of allowing shape reference compounds to explore low-energy conformations which are neither the global, nor even local, intramolecular energy minima. An alternative explanation of this finding is that the NAS model is not an accurate representation of the phenomenon responsible for the loss of biological activity. However, this is difficult to rationalize in light of the similarity between the shape reference compounds (**15,6**) and (**15,9**).

Only one molecular superposition criterion was tried. It might be important to explore whether optimization of the location of the protonated nitrogen is important by different superposition criteria. Alternate superposition criteria were not used since the variances explained by MSA–3D-QSARs from PLS are each significant, especially given the limitations inherent to estimating the presence of a hydrophobic pocket.

Other shape measures of molecular similarity, besides V_o and I_c, could have been explored in this study. The sum of squares of intermolecular distances between key groups in superimposed compounds, measures related to electrostatic potential such as the integrated spatial difference in field potential (ISDFP),[21] and combinations of ISDFP and COSV could have been used.

All of the MSA–3D-QSARs reported here have been independently constructed for M2 and M3 inhibition potency measures. One relatively singular advantage of PLS is that the $\mathbf{T}_{u,v}$ can be determined for multiple property P_u measures. In other words, a matrix of property measures P_u can be related to the features. In the current application both M2 and M3 $-\log(IC_{50})$ measures could have simultaneously been regressed against the feature tensors $\mathbf{V}_{u,v}(s,\alpha,\beta_o)$ and $\mathbf{H}_{u,v}(\alpha,\beta_o)$ using PLS.

Overall, the MSA–3D-QSARs suggest that the active shape of the pyridobenzodiazepinones at both the M2 and M3 receptor sites is best represented by (v,α) = (**15,8**). The best opportunity for M2/M3 specificity is suggested to come from the occupancy of the NAS associated with compound **2**. Still it should be remembered that both the depth and breadth of the SAR of the data base is relatively limited. Most of the compounds have hydrophobic side chains. Only one compound has a hydrophilic side chain (see Table 1). Also, the location of the nitrogen in the side chain could have been more widely varied.

Acknowledgment. This work was performed using the resources of the Laboratory of Molecular Modeling and Design at UIC and computational resources of the Electronic Visualization Laboratory at UIC. B.J.B. gratefully acknowledges fellowships and citation fellowships from the American Foundation for Pharmaceutical Education. Also, the Graduate College of UIC provided fellowship and award support for this work.

References

(1) Hopfinger, A. J.; Burke, B. J.; Dunn, W. J., III. A Generalized Formalism of Three-Dimensional Quantitative Structure–Property Relationship Analysis for Flexible Molecules Using Tensor Representation. *J. Med. Chem.* previous paper in this issue.

(2) Hopfinger, A. J. Theory and Application of Molecular Potential Energy Fields in Molecular Shape Analysis: A Quantitative Structure-Activity Study of 2,4-Diamino-5-benzylpyrimidines as Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1983** *26*, 990–996.

(3) Wold, S.; Wold, H.; Ruhe A.; Dunn, W. J., III. The Colinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comp.* **1984**, *5*, 735–743.

(4) Engel, W. W.; Eberlein, W. G.; Mihm, G.; Hammer, R.; Trummlitz, G. Tricyclic Compounds as Selective Muscarinic Receptor Antagonists. 3. Structure-Selectivity Relationshhips in a Series of Cardioselective (M2) Antimuscarinics. *J. Med. Chem.* **1989** *32*, 1718–1724. (a) Engel, W. W. Personal communication.

(5) Dewar, M. J. S.; Thiel, W. Ground States of Melecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **1977** *99*, 4899–4907.
(6) *CHEMLAB-II Users Guide, Version 11.0*; Molecular Simulations Inc.: Burlington, MA, 1991.
(7) (a) van de Waterbeemd, H.; Testa, B. Theoretical Conformational Studies of Some Dopamine Antagonistic Benzamide Drugs: 3-Pyrrolidyl- and 4–Pypiridyl Deravatives. *J. Med. Chem.* **1983** *26*, 203–207. (b) Fuchs, B.; Wechsler, P. S. Steric Effects of 5-Membered Rings-VIII: A Conformational Analytical Study of cis- and trans-1,3–Dichlorocyclopentane. *Tetrahedron* **1977**, *33*, 57–64.
(8) Allinger, N. L. Conformational Analysis. 130. MM2. A Hydrocarbons Force Field Utilizing V1 and V2 Torsional Terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.
(9) Hopfinger, A. J.; Malhotra, H. D.; Battershell, R. D.; Ho, A. W. Conformational Behavior and Thermodynamic Properties of Phenothrin Analog Insecticides. *Pest. Sci.* **1984**, *9*, 381–391.
(10) Pople, J. A.; Beveridge, D. C. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.
(11) Hopfinger, A. J. *Conformational Properties of Macromolecules*; Academic Press: New York, 1973.
(12) Hopfinger, A. J. A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines based upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7196–7206.
(13) Pomona Medicinal Chemistry Project, Pomona College, Claremont, CA.
(14) Burke, B. J.; Hopfinger, A. J. 1-(Substituted-benzyl)imidazole-2(3H)-thione Inhibitors of Dopamine β-Hydroxylase. *J. Med. Chem.* **1990** *33*, 274–281.
(15) Trummlitz, G.; Schmidt, G.; Wagner, H.-U.; Luger, P. Conformational Analysis of the Antiulcer Drug Pienzepine: x-Ray Investigations, Molecular Mechanics and Quantum Mechanical Calculations and Comparisons with Structurally or Pharmacologically Related Compounds. *Arzneim.-Forsch./Drug Res.* **1984** *34* (II), 849–859.
(16) Hopfinger, A. J. Inhibition of Dihydrofolate Reductase: Structure-Activity Correlations of 2,4-Diamino-5-benzylpyrimidines Based opon Molecular Shape Analysis. *J. Med. Chem.* **1981** *24*, 818–822.
(17) Walters, D. E.; Hopfinger, A. J. Case Studies of the Application of Molecular Shape Analysis to Elucidate Drug Action. *J. Mol. Struct. (Theochem)* **1986**, *134*, 317–323.
(18) Koehler, M. G.; Rowberg-Schaefer, K. L.; Hopfinger, A. J. A Molecular Shape Analysis and Quantitative Structure-Activity Relationship Investigation of some Triazine-Antifolate Inhibitors of Leishamenia Dihydrofolate Reductase. *Arch. Biochem. Biophys.* **1988**, *266*, 152–161.
(19) Hopfinger, A. J.; Burke, B. J. Molecular Shape Analysis of Structure-Activity Tables. In *Quantitative Structure-Activity Relationships in Drug Design*; Fauchére, J. L., Ed.; Alan R. Liss: New York, 1989; p 151.
(20) Glen, W. G.; Dunn, W. J., III; Scott, D. R. Principal Components Analysis and Partial Least Squares Regression. *Tetrahedron Comput. Methodol.* **1989**, *2*, 349–376.
(21) Glen, W. G.; Dunn, W. J., III; Sarker, M.; Scott, D. R. UNIPALS: Software for Principal Components Analysis and Partial Least Squares Regression. *Tetrahedron Comput. Methodol.* **1989**, *2*, 377–396.
(22) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley-Interscience: New York, 1979.