

Quantitative Structure–Activity Relationships (QSARs) of N-Terminus Fragments of NK1 Tachykinin Antagonists: A Comparison of Classical QSARs and Three-Dimensional QSARs from Similarity Matrices

David C. Horwell, William Howson, Michael Higginbottom, Dorica Naylor,* Giles S. Ratcliffe, and Sophie Williams

Parke-Davis Neuroscience Research Centre, Cambridge University Forvie Site, Robinson Way, Cambridge CB2 2QB, U.K.

Received March 31, 1995[§]

The ability of three-dimensional quantitative structure–activity relationships (QSARs) derived from classical QSAR descriptors and similarity indices to rationalize the activity of 28 N-terminus fragments of tachykinin NK1 receptor antagonists was examined. Two different types of analyses, partial least squares and multiple regression, were performed in order to check the robustness of each derived model. The models derived using classical QSAR descriptors lacked accurate quantitative and predictive abilities to describe the nature of the receptor–inhibitor interaction. However models derived using 3D QSAR descriptors based on similarity indices were both robust and significantly predictive. The best model was obtained through the statistical analysis of molecular field similarity indices ($n = 28$, $r^2 = 0.846$, $r(\text{CV})^2 = 0.737$, $s = 0.987$, PRESS = 7.102) suggesting that electronic and size-related properties are the most relevant in explaining the affinity data of the training set. The overall quality and predictive ability of the models applied to the test set appear to be very high, since the predicted affinities of three test compounds agree with the experimentally determined affinities obtained subsequently within the experimental error of the binding data.

Introduction

Substance P (SP),^{1,2} an undecapeptide neurotransmitter, is one of five peptides belonging to the tachykinin family, isolated from mammals. They all share a common C-terminal sequence essential for biological activity^{3–5} and are found throughout the peripheral and central nervous systems.

SP interacts preferentially with the NK1 receptor, one of at least three pharmacologically distinct tachykinin receptor types,^{6,7} and it is thought to be involved in a variety of biological actions including pain transmission and neurogenic inflammation,^{8,9} vasodilation, smooth muscle contraction, bronchoconstriction, saliva secretion,¹⁰ and activation of the immune system. Therefore, pain, inflammation, and asthma are potential therapeutic targets for SP antagonists.

The recognition of the key amino acid residues responsible for antagonist activity^{11,12} recently led to the identification of several high-affinity small peptidic SP antagonists.^{13–15} However a number of potent nonpeptide NK1 tachykinin receptor antagonists^{16–20} have been identified primarily as a result of database screening. The rational design strategy²¹ involved the independent optimization of the N- and C-terminal structure–activity relationships (SAR) of the compound shown in Figure 1. Such an approach assumes that the binding energies of the N- and C-terminus moieties are additive as the ligand interacts with the receptor active site.²² Following on from the SAR studies,²¹ this paper describes several quantitative structure–activity relationship (QSAR) studies involved in defining and understanding the relationship between the properties of N-terminus substituents of NK1 antagonists and their overall affinities.

Firstly, models derived from classical QSAR descriptors are presented and discussed. Although there are

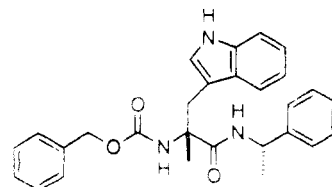


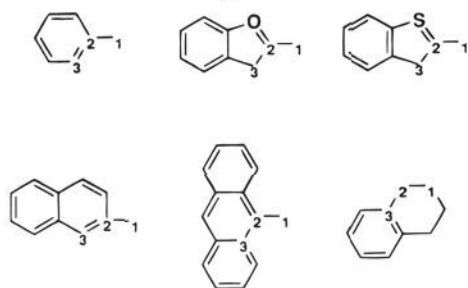
Figure 1. Chemical structure of compound 12.

many examples in the literature^{23–26} where classical QSAR descriptors have been used very successfully, such techniques necessarily are limited by their inadequate treatment of the three-dimensional nature of the compounds and their restricted applicability beyond the limit of a congeneric series.^{25,26}

The biological action of drugs results essentially from their chemical structure which adapts itself to the three-dimensional structure of the receptor by forming a complex with it. The receptor sees the drug in terms of the intermolecular forces (mainly electrostatic and steric) between them. Hence molecular descriptors based on similarity of shape and/or electrostatic properties may be expected to correlate with binding affinity of drugs. Therefore in the second set of QSAR analyses, similarity indices representing steric, electrostatic and lipophilic properties were introduced due to their topographical nature.^{27–29}

A measurement of how similar one molecule is to another in respect of a chosen 3D property has proved to be a useful parameter in studies of relationships between molecular structure and binding affinity.^{27,30–33} Correlations were derived mainly by relating binding affinity to the similarity data obtained from comparison to a single compound usually that with the highest binding affinity.^{36–38} An alternative approach is to use partial least squares (PLS)^{39,40} to analyze the correlations between the affinity data and a data matrix obtained by calculating similarity indices between each

[§] Abstract published in *Advance ACS Abstracts*, October 1, 1995.

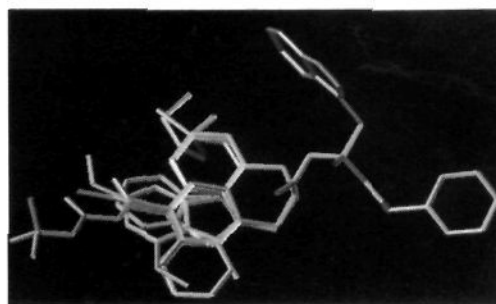
Chart 1. Three Alignment Points for Six Representative Classes of N-Terminus Fragments

molecule and every other. A similar approach was used initially within the comparative molecular field analysis (CoMFA) methodology³⁵ which gave an excellent correlation with the binding data for a steroid data set. Recent results suggest that similarity matrices can be used also to derive good QSAR models for many different systems.^{36,41} The results obtained using similarity indices compare well with those achieved using CoMFA.⁴¹

Methods

All molecular modeling, calculations, and optimization were performed using SYBYL⁴² versions 6.00 and 6.10 running on a Silicon Graphics Iris 4D/310GTX workstation. Initially the assumption was made that the overall 3D structure is the same for each N-terminus substituent. Therefore only the N-terminus fragments were used for the following reasons: (i) all the recent and previously reported biological data²¹ are consistent with a single competitive mode of binding; (ii) the central and C-terminal parts of all molecules are identical, hence differences in binding should be associated exclusively with the N-terminus part of the molecule; and (iii) since some properties (similarity indices) used in the QSAR analyses were associated with 3D-molecular properties, small differences in such properties would be emphasized by using the smaller N-terminus fragments rather than being lost in the larger structures. Additionally, the calculation of the properties for whole molecules would be significantly slower without providing any additional information. The starting fragments were extracted from the SYBYL fragment library and energetically optimized using the TRIPOS force field.⁴³ The library is a database of small molecules which represent averaged geometries from the Cambridge crystallographic data files. Any additional atoms or groups were added using the modeling program SYBYL with standard bond lengths and angles, ensuring identical starting geometry for any given substituent. Finally the resultant structures were superimposed by least squares fitting as indicated in Chart 1 using the FIT option within SYBYL. The chosen alignment was defined by the first two atoms as an obvious choice because they are common to all molecules under study. The third atom was chosen such that the dipole of the fitted molecule corresponded most closely to the dipole of the most active compound.

Charges were calculated using the Charge-2 method. Charge-2 is an empirical method which has been used successfully in a wide variety of chemical classes.⁴⁴ It is based on two fundamental chemical concepts: (i) inductive effect in saturated molecules operating via the atomic electronegativity and polarizability and (ii)

**Figure 2.** Superimposed N-terminus molecular fragments of the 28 NK1 antagonists.**Table 1.** Definition of Variables Calculated by TSAR

variable	definition of variables ^a
X1	molecular volume
X2	log <i>P</i>
X3	square of log <i>P</i>
X4	total dipole
X5	number of methyl groups
X6	surface area
X7	moments of inertia X-moment
X8	moments of inertia Y-moment
X9	moments of inertia Z-moment
X10	ellipsoidal volume
X11	molar refractivity
X12	molecular mass
X13	Kier shape index Kappa α 1
X14	Kier shape index Kappa α 2
X15	Kier shape index Kappa α 3
X16	flexibility (ϕ)
X17	total number of atoms
X18	number of C atoms
X19	number of heteroatoms
X20	total lipole
X21	moments of inertia x-axis length
X22	moments of inertia y-axis length
X23	moments of inertia z-axis length
X24	Kier connectivity index ChiV0
X25	Kier connectivity index ChiV1 (path)
X26	Kier connectivity index ChiV2 (path)
X27	Kier connectivity index ChiV3 (cluster)
X28	Kier connectivity index ChiV4 (cluster)
X29	Kier connectivity index ChiV5 (cluster/path)
X30	Kier connectivity index ChiV3 (path)
X31	Kier connectivity index ChiV4 (path)
X32	Kier connectivity index ChiV5 (path)
X33	Kier connectivity index ChiV5 (ring)
X34	Kier connectivity index ChiV6 (ring)
X35	Randic topological index
X36	Balaban topological index
X37	Wiener topological index

^a The variables are defined as described in ref 45.

Hückel molecular orbital calculations for π systems operating through the appropriate Coulomb and resonance integrals.

All superpositioned molecular fragments shown in Figure 2 were then loaded via SYBYL mol2 files into a data table within the graphics-based QSAR program TSAR⁴⁵ version 2.20. The original affinity data (IC₅₀, nM) for each molecule was entered into the table, and a new affinity column was then generated by taking the negative logarithm of the original data (pIC₅₀). The entire range of molecular properties was calculated using TSAR (Table 1). The molecular property calculation in TSAR uses 3D-structural information stored in the data table.

The concept of bioisosterism was used as the basis for calculating similarity indices.²⁹ Similarity indices represent a quantitative measure of the similarity between two molecules on the basis of their size, shape,

electronic distribution, lipid solubility, water solubility, or chemical reactivity.⁴⁶ There are two approaches to calculating similarity indices, a grid-based method^{28,29} and a Gaussian approximation^{37,38}. For grid-based methods, a 3D grid of points surrounding the two molecules is compared. The similarity calculations are performed by numerical integration of the potential and field across the 3D grid using the Carbo or Hodgkin equations.^{28,29} The Carbo equation²⁸ is sensitive to the shape of a property's distributions but not to its overall magnitude. The Hodgkin index²⁹ was introduced in order to increase the sensitivity of the formula to a property's magnitude. Both indices are highly correlated, and there is hardly any difference in their overall performance.⁴¹ The shape similarity indices are calculated using Meyer's⁴⁷ modification of the Carbo and Hodgkin equations based on the degree of overlap between the two molecules. Grid-based potential, field, and shape similarity evaluations are both time consuming and dependent upon the grid size and increment.^{37,38} The analytical method fits Gaussian curves to reproduce the $1/r$ term (electrostatic potential)³⁷ or the electron density function for different atom types (shape).³⁸ It has been shown that the behavior of the Gaussian functions in similarity calculations closely mirrors that of the grid-based calculations but is much faster.^{37,38}

Therefore the following potential and field $N \times N$ similarity calculations were performed (i) Carbo potential similarity indices with the Gaussian function approximation and (ii) Carbo atomic field indices—single-point grid method.

All derived data were standardized by mean (zero) and standard deviation (unity) and then analyzed using various statistical techniques. Principal component analyses (PCA) were performed on all similarity data extracting only the components that explained up to 95% of the variance. Also correlations were derived using PLS analysis^{39,40} with different number of principal vectors⁴⁰ and multiple regression (MR) analysis with or without stepping procedures.⁴⁸ For a well-defined problem, both PLS and MR techniques should give similar predictions.⁴⁹ Cross-validation of all PLS and MR analyses was performed to indicate if the chance correlations⁵⁰ were obtained. It gives an estimate of the true predictive power of the model. The model is judged based on these predictions by comparing predicted values with the exact values from each compound that has been held out. The predictions are then summed for the n th PLS (MR) dimension to obtain a value for the predictive sum of squares (PRESS _{m}) of the model (uncertainty of the prediction).

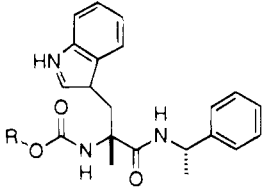
Cross-validation is also used in the jackknife procedure^{39,49} to estimate the standard deviation of the regression coefficients. The procedure uses each of the reduced matrices during the cross-validation to derive a new set of regression coefficients. The spread in the values of these coefficients is a conservative measure of the standard deviation for each coefficient. The cross-validation method used in each analysis was a fixed pattern of three subsets of data.

The regression coefficients r^2 for the models derived using PLS and MR analysis were compared together with their corresponding $r(\text{cv})^2$ and PRESS values.

Results and Discussion

The N-terminus training set shown in Table 2. was

Table 2. Binding Affinity Data for Compounds in Training Set



compd	R	pIC ₅₀
1	CH ₂ -(2-benzofuran)	9.26
2	CH ₂ -(2,4-diF-C ₆ H ₃)	8.72
3	CH ₂ -(2-F-4-CH ₃ -C ₆ H ₃)	8.70
4	CH ₂ -(2,5-diF-C ₆ H ₃)	8.67
5	CH ₂ -(2-benzothiophene)	8.62
6	CH ₂ -(2-F-C ₆ H ₄)	8.50
7	CH ₂ -(2-naphthalene)	8.00
8	CH ₂ -[2-(7-OCH ₃ -benzofuran)]	7.89
9	CH ₂ -(4-OCH ₃ -C ₆ H ₄)	7.80
10	CH ₂ -(3,4-diCH ₃ -C ₆ H ₃)	7.74
11	CH ₂ -(4-CH ₃ -C ₆ H ₄)	7.72
12	CH ₂ -C ₆ H ₅	7.72
13	CH ₂ -(3-NH ₂ -C ₆ H ₄)	7.70
14	CH ₂ -(3-CH ₃ -C ₆ H ₄)	7.65
15	CH ₂ -(4-Cl-C ₆ H ₄)	7.57
16	CH ₂ -(3,4-diCl-C ₆ H ₃)	7.54
17	CH ₂ -(2-Cl-C ₆ H ₄)	7.48
18	CH ₂ -(3-Cl-C ₆ H ₄)	7.46
19	CH ₂ -(3,4-diF-C ₆ H ₃)	7.37
20	CH ₂ -(3,4-diOCH ₃ -C ₆ H ₃)	7.01
21	CH ₂ -[2-(5-OCH ₃ -benzofuran)]	6.85
22	2-tetrahydronaphthalene	6.82
23	CH ₂ -(4-OCOCH ₃ -C ₆ H ₄)	5.96
24	CH ₂ -(3-NHCOCH ₂ Cl-C ₆ H ₄)	5.80
25	CH ₂ -(9-anthracene)	5.68
26	CH ₂ -(4-CH ₂ NHCO ₂ C(CH ₃) ₃ -C ₆ H ₄)	5.43
27	CH ₂ -(3-NHSO ₂ CH ₃ -C ₆ H ₄)	5.14
28	CH ₂ -(3-NHCOCH ₃ -C ₆ H ₄)	5.13

composed from the 28 NK1 antagonists resulting from SAR on N-terminal group with known affinities (pIC₅₀ ranges between ~5.0 and 9.26).

PLS analyses were performed on the dataset with pIC₅₀ as the dependent variable (Y) and 37 calculated physicochemical properties as the independent variables—descriptors (X). The results of these analyses are shown in Table 3. The number of PLS vectors extracted in each analysis was determined according to the following criteria: (i) stopping when the statistical significance of the current vector goes above a fixed value (1.0 by default),⁴⁰ (ii) stopping at the first decrease in PRESS, and (iii) stopping at the lowest value of PRESS. The vectors represent the total number of cycles of the PLS algorithm used in the analysis. Each iteration improves the regression equation but increases the risk of overfitting the data. These analyses were performed in order to resolve which of the above criteria gives the most robust model with significant predictive potential. The model from the first of these analyses contained only one statistically significant PLS vector,⁴⁰ the second model contained three PLS vectors, and the third model four PLS vectors.

It is evident from Table 3 that none of the criteria used derived an acceptable model. In the first case, the model was not robust enough, $r^2 = 0.540$, and did not possess sufficient predictive power, $r(\text{cv})^2 = 0.366$, and in the other two cases, the model was overfitting the data as indicated by the noticeable difference between r^2 (0.779, 0.831) and the corresponding $r(\text{cv})^2$ (0.508, 0.550).⁵² These results suggest that the classical QSAR descriptors used may not be the optimal set for the dataset under study.²⁵ In addition to that, PLS does

Table 3. Results of Partial Least Squares Analyses on 37 Classical QSAR Descriptors

analysis	vector ^a	r ²	r(cv) ²	PRESS	statistical significance
1 ^b	1	0.540	0.366	17.12	0.796
2 ^c	3	0.779	0.508	13.27	1.227
3 ^d	4	0.831	0.550	12.15	1.428

^a Number of vectors (components) represents the total number of cycles of the PLS algorithm used in analysis. ^b Number of vectors included were up to a statistical significance of 1. ^c Number of vectors included were up to the first decrease in PRESS. ^d Number of vectors included were up to the lowest value of PRESS.

not handle very well data with many descriptors associated with the one class of property (e.g., size, shape) and only a few associated with the others (electronic and hydrophobic property).⁴⁹ In such a case, shape properties would dominate the analysis by being represented with the majority of independent variables,⁴⁰ but they are not necessarily the most relevant properties in describing the affinity data of the training set.

To distinguish between these two possibilities, the same data table was subjected to MR analysis. To reduce the number of descriptors to a level manageable for MR analysis and to determine which of the 37 calculated physicochemical properties (see Table 1) were highly correlated,⁵² PCA was implemented. The inspection of the correlation matrix resulting from PCA shown graphically in Figure 3 allows the extraction of just one representative property from each highly correlated class. Ideally the chosen properties should be reasonably common and easy to understand and interpret.⁵¹ The descriptors selected from each class of properties are colored red in Figure 3 and are as follows: (i) steric related property—surface area (vector 6), (ii) lipophilic related property—log *P* and log *P*² (vectors 2 and 3), and (iii) electronic related property—total dipole (vector 4). Additionally the number of methyl groups indicated in Figure 3 in green was also used in MR analysis. The total lipole (vector 20 in Figure 3) is correlated to log *P* and therefore was omitted from MR analysis. Kier's ChiV6 (ring) connectivity index (vector 34 in Figure 3) has only limited correlation (<50%) with other steric (shape) properties, but it was not used in MR analysis since it is difficult to understand and interpret its meaning.⁵¹ The same argument was implemented in elimination of the Balaban topological index (vector 36 in Figure 3) from any MR analysis. After selecting a representative set of descriptors, several MR analyses were performed with or without stepping procedures, and the results are shown in Table 4.

In basic regression analysis, all selected *X* variables are used to create a regression equation to predict *Y*. In stepwise regression, a selection algorithm is used to choose a subset of the *X* variables, from which a regression equation is calculated. In the analysis with stepwise regression, the derived model has only two relevant descriptors (total dipole and surface area), both of which are negatively correlated with affinity. Inspection of the data in Table 4 shows that both procedures (basic and stepwise) produce models of comparable quality. However the higher regression coefficient *r*² of the stepwise model (0.671 vs 0.615) and the lower *s* value (0.723 vs 0.734) suggest that this model is more robust.

The cross-validated *r*² values however vary more significantly, from 0.478 in the stepwise model to 0.576

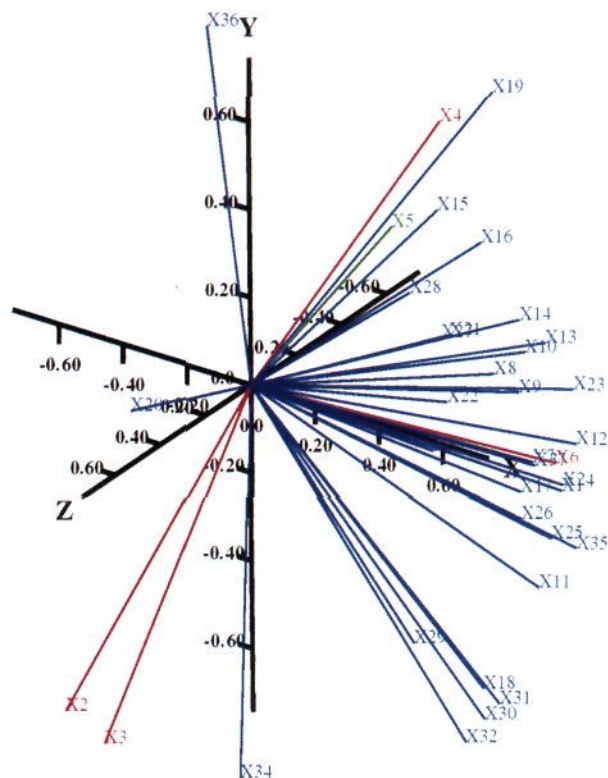


Figure 3. Three-dimensional representation of the correlation matrix derived from PCA of the 37 physicochemical properties calculated by TSAR. Selected descriptors are colored red and are as follows: (i) surface area (vector 6), (ii) log *P* and log *P*² (vectors 2 and 3), (iii) total dipole (vector 4), and (iv) number of methyl groups colored green (vector 5). *x*-Axis: principal component 1. *y*-Axis: principal component 2. *z*-Axis: principal component 3.

Table 4. Results of Multiple Regression Analyses Derived from Classical QSAR Descriptors with and without Stepping Procedure

analysis ^a	r ²	r(cv) ²	PRESS	<i>F</i> value	<i>s</i> value
1 ^b	0.671	0.576	14.83f	8.97	0.723
2 ^c	0.615	0.478	18.25	19.95	0.734

^a Regression was obtained using classical QSAR descriptors: surface area, log *P*, log *P*², total dipole, and number of methyl groups. ^b Regression was derived without stepping procedure implemented. ^c Regression was derived with stepping procedure (*F*_{in enter} = 2.0; *F*_{to leave} = 2.0).

in the basic regression model. Consequently the PRESS value for the basic model is lower than the PRESS value for the stepwise model (14.83 vs 18.25), suggesting better predictive ability of the former model. Although the higher *F* value (19.95 vs 8.97) of the stepwise model seems to contradict this conclusion, *F* values from stepwise procedures are often overestimated by the inherent bias of the variable selection process.⁵⁰ According to the general consensus⁵¹ that the model with higher *r*² and *r*(cv)² values that are reasonably close to each other should be the most robust one with predictive potential, the best regression equations are as follows.

Regression equation (original data):

$$Y = 2.191X1 - 0.428X2 - 0.361X3 - 0.083X4 - 0.016X5 + 8.055 \quad (1a)$$

Regression equation (standardized data):

Table 5. Results of Partial Least Square Analyses

a. Using Different Sets of Carbo Potential Similarity Indices with Gaussian Function Approximation									
Carbo potential similarity indices				vector ^a	r ²	r(cv) ²	st sign ^b	RSS	PSS
shape	electrostatic	lipophilic	refractivity						
28	28	28	28	1	0.647	0.603	0.630	9.54	10.71
28	28	28		1	0.671	0.618	0.618	8.89	10.32
28	28		28	2	0.713	0.590	0.951	7.76	11.08
28		28	28	1	0.595	0.547	0.673	10.94	12.22
	28	28	28	1	0.608	0.559	0.664	10.59	11.91
28	28			2	0.792	0.670	0.946	5.63	8.91
28		28		1	0.613	0.562	0.662	10.45	11.82
28			28	3	0.640	0.574	0.986	9.72	11.49
	28	28		1	0.533	0.451	0.741	12.61	14.81
	28		28	2	0.713	0.556	0.961	7.76	12.00
		28	28	1	0.543	0.496	0.710	12.34	13.61

b. Using Different Sets of Carbo Atomic Field Similarity Indices and/or Carbo Shape Potential Similarity Indices									
Carbo shape and field similarity indices				vector ^a	r ²	r(cv) ²	st sign ^b	RSS	PSS
shape	electrostatic	lipophilic	refractivity						
28	28	28	28	1	0.692	0.638	0.602	8.31	9.77
28	28	28		1	0.710	0.644	0.598	7.84	9.61
28	28		28	2	0.779	0.712	0.891	5.98	7.79
28		28	28	1	0.630	0.572	0.654	9.99	11.56
	28	28	28	1	0.713	0.646	0.595	7.74	9.57
28	28			2	0.815	0.727	0.901	5.00	7.38
28		28		1	0.635	0.571	0.655	9.78	11.58
28			28	2	0.620	0.563	0.943	10.29	11.80
	28	28		1	0.655	0.561	0.662	9.31	11.85
	28		28	2	0.846	0.737	0.987	4.17	7.10
		28	28	1	0.647	0.586	0.644	9.59	11.19

^a Number of vectors calculated represents the total number of cycles of the PLS algorithm used in analysis. ^b Statistical significance.

$$Y = 1.658(\pm 0.489)S1 - 1.767(\pm 0.069)S2 - 0.460(\pm 0.056)S3 - 0.070(\pm 0.267)S4 - 0.447(\pm 0.008)S5 + 7.355(\pm 1.281) \quad (1b)$$

$$n = 28, r = 0.819, F(5,22) = 8.9, s = 0.723$$

where $X1 = \log P$, $X2 = \log P^2$, $X3 = \text{total dipole}$, $X4 = \text{number of methyl groups}$, and $X5 = \text{surface area}$.

The regression equation suggests that the five chosen properties are not equally relevant. Examination of the absolute values of the coefficients in the regression equation with standardized data (eq 1b) leads to the conclusion that the total dipole and the surface area of the N-terminus part of NK1 antagonists under study are dominant and equally important factors in describing the affinity of the training set of compounds. Conversely the lipophilic properties $S1$ and $S2$ (which tend to cancel each other out for positive $\log P$ values) and the number of methyl groups ($S4$) make little or no contribution. It is interesting to note that both regression models, especially the better one with $r^2 = 0.671$ and $r(cv)^2 = 0.576$, indicate a considerable improvement in comparison with the results obtained by PLS analysis with one PLS vector ($r^2 = 0.540$ and $r(cv)^2 = 0.366$), suggesting that MR is the better analysis method for the chosen dataset. However, in spite of this, the best derived model is still not robust enough and lacks high predictive power.

The fact that descriptors commonly used in classical statistical analysis failed to identify a robust model in either MR or PLS analyses led to the idea that similarity indices used successfully in several reported QSAR studies^{36,41,46} may provide a better set of descriptors for the compounds under study. Therefore both potential and field similarity indices were calculated comparing the electrostatic, lipophilic, and size- and shape-related properties of the N-terminus substituents to one an-

other. The resulting data table has significantly more columns than rows, and hence it was analyzed using PLS. The results of several PLS analyses using different sets of similarity indices are summarized in Table 5.

The number of vectors calculated in each analysis was determined according to criteria defined by Stahle and Wold,⁴⁰ stopping when the statistical significance of the current vector goes above a fixed value (1.0 by default). For each variable there is a regression coefficient and a jackknife estimate of the standard error on each coefficient at each step of the analysis.

In the first PLS performed (Table 5a, row 1), the entire matrix of potential and shape similarity indices was submitted to analysis. The following 10 rows (Table 5a, rows 2–11) represent results of PLS analyses where different sets of potential and/or shape similarity indices were used in the calculation. Table 5b is arranged in the same manner as Table 5a with the difference that instead of potential similarity indices, field similarity indices were used in the PLS analyses. The higher r^2 and $r(CV)^2$ (columns 3 and 4 in Table 5b vs columns 3 and 4 in Table 5a) and the lower residual sum of squares (RSS) and predictive sum of squares (PSS) values (columns 6 and 7 in Table 5b vs columns 6 and 7 in Table 5a) indicate that more robust models are obtained by using field descriptors, which give information about dipolar interactions, rather than potential descriptors, which give information about the location and strength of ionic interactions between ligand and receptor. Similar findings have been reported elsewhere.⁴¹

For each derived model in Table 5b, the r^2 and $r(cv)^2$ values are close to each other which suggests that none of the models is overfitting the data. The most robust model (row 10 in Table 5b) was obtained when both electronic and size properties were used in the analysis.

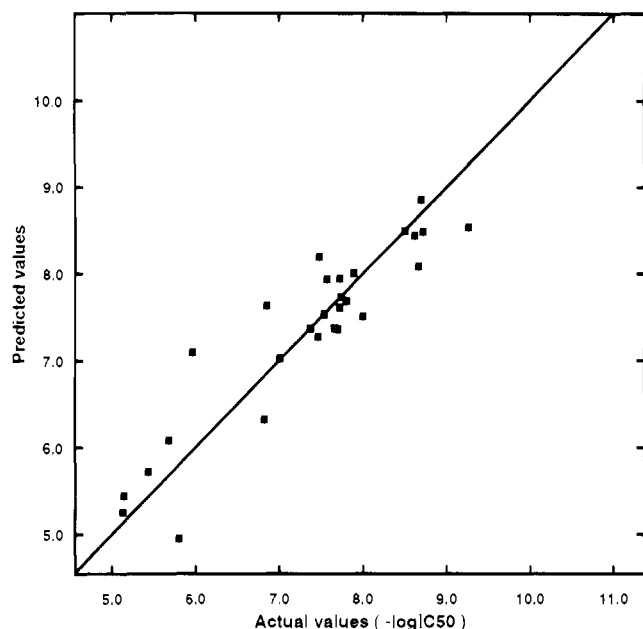


Figure 4. Observed versus predicted pIC_{50} values for the N-terminus fragments in the training data set derived using PLS analysis on similarity matrices; $r^2 = 0.846$, $r(cv)^2 = 0.737$, and $n = 28$.

Table 6. Results of Multiple Regression Analyses Using Different Numbers of PCs Extracted from Carbo Field and Shape Potential Similarity Indices

no. of descriptors	r^2	$r(CV)^2$	F values	s values
15 ^a	0.829	0.326	38.75	0.499
9 ^b	0.825	0.702	37.65	0.505
6 ^c	0.801	0.706	32.23	0.538
3 ^d	0.829	0.756	38.75	0.499

^a Analysis was performed using stepping procedure with default number of steps and all relevant PCs extracted from field and shape potential similarity indices. ^b Analysis was performed using stepping procedure with default number of steps and the first three PCs extracted from field similarity indices. ^c Analysis was performed using stepping procedure with $F_{to\ enter} = 2.0$ and $F_{to\ leave} = 2.0$ and the first three PCs extracted from electrostatic and refractivity field similarity indices. ^d Analysis was performed using basic procedure and one PC extracted from each electrostatic and refractivity field similarity indices and shape potential similarity indices with Gaussian function approximation.

It has $r^2 = 0.846$, $RSS = 4.17$, $PRESS = 7.10$, and the best predictive ability, $r(cv)^2 = 0.737$, and is shown in Figure 4.

To check internally the robustness of the PLS models derived from the similarity indices, several MR analyses were also performed. Firstly, PCA was used on each set of field and shape similarity indices to reduce the large number of descriptors to a much smaller number of components that still contain the same information. The results of MR analyses on selected sets of PCs are summarized in Table 6. The first MR analysis performed using the stepping procedure produced a very robust model ($r^2 = 0.829$) but with insignificant predictive power ($r(cv)^2 = 0.326$). Therefore the predictive accuracy of the model will be much worse than the s value of 0.499 suggests. However the other stepwise regression analyses (Table 6, rows 2 and 3) which used a smaller number of PCs produced models that were both robust ($r^2 = 0.825$ and 0.801) and with high predictive power ($r(cv)^2 = 0.702$ and 0.706). The regression equation of the best derived model ($r^2 = 0.829$, $r(cv)^2 = 0.756$, and $s = 0.499$) was obtained using the backward elimination procedure⁴⁸ followed by basic

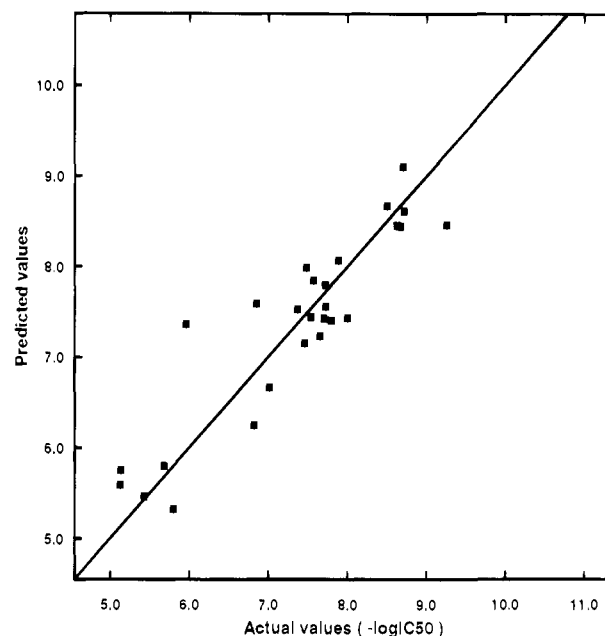


Figure 5. Observed versus predicted pIC_{50} values for the N-terminus fragments in the training data set derived using MR analysis on PCs extracted from similarity matrices; $r^2 = 0.829$, $r(cv)^2 = 0.756$, and $n = 28$.

regression analysis. It is displayed graphically in Figure 5 and numerically below.

Regression equation (original data):

$$Y = 0.387X1 + 0.572X2 + 0.445X3 + 7.335 \quad (2a)$$

Regression equation (standardized data):

$$Y = 0.394(\pm 0.140)S1 + 0.582(\pm 0.043)S2 + 0.454(\pm 0.056)S3 + 7.355(\pm 0.107) \quad (2b)$$

$$n = 28, r = 0.910, F(3,24) = 38.75, s = 0.499$$

where $X1$ is a principal component extracted from the shape similarity indices, $X2$ is a principal component representing the electrostatic field similarity indices, $X3$ is a principal component derived from the refractivity field similarity indices, and $S1-S3$ are their respective standardized values.

Inspection of the coefficients in eq 2b suggests that the dominant descriptor in explaining the affinity of the training set of compounds is related to the electronic distribution around the N-terminus substituent and that the shape and size of the substituent also play significant although not so dominant roles. This is in agreement with the observations obtained from the results of the PLS analyses where the best model extracted also suggested the relevance of the same descriptors.

The observed pIC_{50} values and their predicted pIC_{50} values together with their corresponding residuals derived using the best models from both the PLS and the MR analyses shown in Figures 4 and 5 are listed in Table 7. The residuals vary from 0.00 to 1.13 in the PLS model and from 0.06 to 1.42 in the MR model. The biggest differences between the observed and predicted pIC_{50} values in both models were obtained for compound **23** whose affinity was overpredicted (observed $IC_{50} = 1100$ nM vs predicted $IC_{50}(PLS) = 81$ nM and $IC_{50}(MR) = 44$ nM) and compound 1 whose affinity was under-

Table 7. Observed pIC₅₀ Values vs Predicted pIC₅₀ Values and Corresponding Residuals for the Training Set of Compounds

compd	pIC ₅₀	PLS analysis (<i>r</i> ² = 0.846; <i>r</i> (cv) ² = 0.737)		MR analysis (<i>r</i> ² = 0.829; <i>r</i> (cv) ² = 0.756)	
		pIC ₅₀ (predicted)	residual	pIC ₅₀ (predicted)	residual
1	9.26	8.54	0.72	8.46	0.80
2	8.72	8.48	0.24	8.61	0.11
3	8.70	8.86	-0.16	9.10	-0.40
4	8.67	8.08	0.59	8.44	0.23
5	8.62	8.44	0.18	8.45	0.17
6	8.50	8.49	0.01	8.66	-0.16
7	8.00	7.50	0.50	7.43	0.57
8	7.89	8.01	-0.12	8.07	-0.18
9	7.80	7.69	0.11	7.40	0.40
10	7.74	7.73	0.01	7.55	0.19
11	7.72	7.95	-0.23	7.80	-0.08
12	7.72	7.61	0.11	7.56	0.16
13	7.70	7.36	0.34	7.43	0.27
14	7.65	7.37	0.28	7.23	0.42
15	7.57	7.94	-0.37	7.85	0.28
16	7.54	7.53	0.01	7.44	0.10
17	7.48	8.19	-0.71	7.99	-0.51
18	7.46	7.27	0.19	7.16	0.30
19	7.37	7.37	0.00	7.52	-0.15
20	7.01	7.02	-0.01	6.66	0.35
21	6.85	7.63	-0.78	7.59	-0.74
22	6.82	6.32	0.50	6.25	0.57
23	5.96	7.09	-1.13	7.36	-1.40
24	5.80	4.95	0.85	5.32	0.48
25	5.68	6.08	-0.40	5.80	-0.12
26	5.43	5.72	-0.29	5.47	-0.04
27	5.14	5.45	-0.31	5.76	-0.62
28	5.13	5.26	-0.13	5.59	-0.46
rms residuals			0.44		0.46

predicted (observed IC₅₀ = 0.55 nM vs predicted IC₅₀(PLS) = 2.9 nM and IC₅₀(MR) = 3.5 nM).

The overprediction of compound **23** is due to the fact that this compound shares over 90% of overall shape and size similarities with the five most potent compounds (IC₅₀s between 0.55 and 2.4 nM) although its electronic properties are different (between 10% and 40% similarity). Appropriate shape and size properties are one of the major reasons for the high affinity of the leading compounds according to both derived models, and therefore compound **23** is predicted to have a binding affinity of ~50 nM, although its overall electronic properties are not optimal.

The predicted IC₅₀ values for compound **1** are 5–6-fold higher than its true affinity. The reason for this could lie with the fact that the training set contains only two compounds, **5** (IC₅₀ = 2.4 nM) and **8** (IC₅₀ = 13 nM), with physicochemical properties very similar to compound **1** (<96% shape and size similarity and ~75% electronic similarity). Therefore compound **1** was predicted to have an affinity closer to the affinity of these compounds than to its actual value.

The best models derived from both the PLS and MR analyses were also used to predict the binding affinities of a test set of 11 compounds (Table 8). Both PLS and MR analyses predict the same test compounds to have high binding affinity (~10 nM), namely, **test6** and **test9**. Compound **test6** has a methyl substituent in position 3 of the benzofuran ring and belongs to the same family of compounds as the high-affinity compounds **1** (0.55 nM) and **5** (2.4 nM). It has 94% and 76% electronic similarity, 99% and 95% shape similarity, and 97% and 94% size similarity to those compounds. Its size is somewhat larger (molecular volume = 119.6 Å³ in comparison with 106.54 Å³ for

Table 8. Compounds Belonging to the QSAR Test Set

compd	R
test1	CH ₂ -(2-quinoline)
test2	CH ₂ -(3-benzofuran)
test3	CH ₂ -(2-(3-F-benzofuran))
test4	CH ₂ -(2-(3-Cl-benzofuran))
test5	CH ₂ -(2-(3-OCH ₃ -benzofuran))
test6	CH ₂ -(2-(3-CH ₃ -benzofuran))
test7	CH ₂ -(2-OH-C ₆ H ₄)
test8	CH ₂ -(3-OH-C ₆ H ₄)
test9	CH ₂ -(2-F-5-CH ₃ -C ₆ H ₄)
test10	CH ₂ -(2-F-5-OH-C ₆ H ₄)
test11	CH ₂ -(2-F-5-CN-C ₆ H ₄)

Table 9. Observed pIC₅₀ Values vs Predicted pIC₅₀ Values and Corresponding Residuals for the Test Set of Compounds

compd	PLS analysis (<i>r</i> ² = 0.846; <i>r</i> (cv) ² = 0.737)		MR analysis (<i>r</i> ² = 0.829; <i>r</i> (cv) ² = 0.756)	
	pIC ₅₀ (predicted)	residual	pIC ₅₀ (predicted)	residual
test1	6.99		6.94	
test2	5.51		5.24	
test3	7.40		7.58	
test4	7.41		7.56	
test5	7.06		7.57	
test6	7.94	0.05	7.99	-0.05
test7	6.92		7.07	
test8	7.89	0.72	7.40	0.49
test9	8.17	0.34	8.14	0.03
test10	7.86		7.92	
test11	7.03		7.51	
rms residuals		0.46		0.29

compound **1** and 118.2 Å³ for compound **5**) which both models suggested to be beneficial for overall binding affinity.

Similarly compound **test9** is a member of the same congeneric series as compounds **2–4** and **6**. They share more than 95% shape and size similarity and between 43% (compound **2**) and 88% (compound **6**) electronic similarity. The overall high similarity in relevant properties to the high-affinity compounds **2–4** and **6** suggested that **test9** had the best potential to bind to the receptor with high affinity.

As a consequence of these findings, compounds **test6** and **test9** were synthesized and tested for their binding affinity. Additionally, compound **test8** was also synthesized and tested. This compound was chosen because it has >90% similarity in shape and size with compounds **2–4** and **6** but considerably different electronic properties (between 1% dissimilarity and 39% similarity). Therefore it was predicted to have moderate binding affinity (IC₅₀(predicted) ~50 nM), but it was considered to be an interesting test of the overall relevance of the direction and magnitude of the N-terminus dipole to binding affinity.

The predicted pIC₅₀ values for the selected compounds and their subsequently determined pIC₅₀ values together with their residuals are shown in Table 9. Inspection of the data in Table 9 suggests that the experimentally observed values and QSAR-derived values are in good agreement (rms residuals = 0.46 (PLS) and 0.29 (MR)). Moreover the predicted IC₅₀ value for compounds **test6** (13 nM (PLS) and 10 nM (MR)), **test8** (68 nM (PLS) and 40 nM (MR)), and **test9** (15 nM (PLS) and 7.2 nM (MR)) vs the observed IC₅₀ values (12, 13, and 6.8 nM, respectively) are within the experimental error of the biological tests.

Conclusion

A comparison was made between the models derived from both PLS and MR analyses using conventional QSAR descriptors and similarity indices on the identical training and test datasets. The best correlations derived using PLS and MR analyses on similarity matrices with their r^2 values of 0.846 and 0.829 and corresponding $r(\text{cv})^2$ values of 0.737 and 0.756 are superior to the best model obtained on conventional QSAR parameters from MR analysis ($r^2 = 0.671$ and $r(\text{cv})^2 = 0.576$). These results are in contrast to previously reported results²⁶ which suggest that little or no advantage is obtained by using CoMFA methodology over the classical Hansch approach. Both potential and field similarity indices proved to be useful descriptors for deriving robust 3D-QSAR models. Relevance of size, shape, and electrostatic field surrounding the N-terminus in defining binding affinity of the training set of compounds was emphasized. All the results discussed above indicated that by using either PLS or MR analysis with field and/or shape potential similarity indices as descriptors, very robust models can be derived that also possess very powerful predictive abilities. The predicted affinities of the three compounds in the test set that were synthesized and experimentally tested subsequently were in the range of the average experimental error of the biological tests. The weakness of similarity matrices is that they are clearly dependent on the 3D alignment of the substituents and therefore more influenced by subjectiveness. In contrast, the classical QSAR approach is procedurally less difficult and time consuming but did not prove to be useful in deriving an equally robust model. Finally, the distinguished advantage of QSAR based on similarity indices is its ability to supply insight into the topographical properties essential for binding.

Experimental Section

Biological Assay. NK1 receptor-binding assay was performed on human IM9 cells using the method given in ref 21.

Chemistry. All compounds were prepared using methods analogous to those given in ref 21. Melting points were determined with a Mettler FP80 or a Reichert Thermovar hot-stage apparatus. Proton NMR spectra were recorded on a Varian Unity +400 spectrometer; chemical shifts were recorded in parts per million (ppm) downfield from tetramethylsilane. IR spectra were recorded with the compound either neat (oils and liquids) or as a Nujol mull on a sodium chloride disk on a Perkin-Elmer System 2000 Fourier transform spectrophotometer. Optical rotations were determined with a Perkin-Elmer 241 polarimeter. Mass spectra were recorded with a Fisons VG Trio-2A or Finnigan MAT TSQ70 triple quadrupole mass spectrometer. Element analyses were determined by Medac Ltd., Uxbridge, U.K. Normal-phase silica gel used for chromatography was Merck No. 9385 (230–400 mesh); reverse-phase silica gel used was Lichroprep RP-18 (230–400 mesh). For further details, see ref 21.

The preparation methods and analytical data of the three novel compounds only are reported here. Compounds **test6**, **test8**, and **test9** were prepared using methods analogous to those given in ref 21. The required starting alcohols 2-fluoro-5-methylbenzenemethanol and 3-methylbenzofuran-2-methanol were prepared by the reduction of commercially available carboxylic acids using NaBH_4/I_2 ⁵³ and a LiBH_4 reduction of mixed anhydride, respectively. The acetic acid 3-(hydroxymethyl)phenyl ester was prepared using the selective acylating agent 1-acetyl-1*H*-1,2,3-triazolo[4,5-*b*]pyridine⁵⁴ with 3-hydroxybenzyl alcohol.

[R(R*,S*)]-[2-(1*H*-Indol-3-yl)-1-methyl-1-[(1-phenylethyl)carbamoyl]ethyl]carbamic acid 3-methylbenzofuran-2-yl methyl ester (test6): mp 75–78 °C; $[\alpha]^{22}_{\text{D}} = +14^\circ$ ($c =$

0.15, MeOH); IR (film) 3322, 1720, 1657, 1494, 1455, 1246, 1069, 742 cm^{-1} ; NMR (CDCl_3) δ 1.28 (3H, d, $J = 6.8$ Hz), 1.58 (3H, s), 2.26 (3H, s), 3.22 (1H, d, $J = 14.7$ Hz), 3.44 (1H, d, $J = 14.7$ Hz), 4.95–5.02 (1H, m), 5.11 (1H, d, $J = 13.2$ Hz), 5.21 (1H, d, $J = 13.2$ Hz), 5.30 (1H, s), 6.34 (1H, d, $J = 7.3$ Hz), 6.70 (1H, m), 7.02–7.41 (11H, m), 6.48–7.49 (2H, m), 7.75 (1H, s); MS m/e (FAB) 510.4 ($\text{M}^- + \text{H}$, 35), 466.3 (25), 447.4 (15), 433.4 (17), 419.5 (21), 392.3 (21), 371.1 (28), 355.1 (21), 341 (16), 325 (20), 281 (44), 236.9 (42). Anal. ($\text{C}_{31}\text{H}_{31}\text{N}_3\text{O}_4$) C, H, N.

[R(R*,S*)]-[2-(1*H*-Indol-3-yl)-1-methyl-1-[(1-phenylethyl)carbamoyl]ethyl]carbamic Acid 3-Hydroxybenzyl Ester (test8). A saturated aqueous solution of K_2CO_3 (10 mL) was added to a solution of [R(R*,S*)]-acetic acid 3-[[[2-(1*H*-indol-3-yl)-1-methyl-1-[(1-phenylethyl)carbamoyl]ethyl]carbamoyl]oxy]methyl]phenyl ester (400 mg, 7.8 mmol) in methanol (10 mL) and stirred vigorously for 1 h. The reaction mixture was acidified using 2 N HCl, and the product was extracted using EtOAc (3 × 30 mL). The combined organic layers were dried (MgSO_4) and evaporated under reduced pressure. The residue was purified using RP silica gel with MeOH/ H_2O followed by crystallization from MeOH/ H_2O to give **test8** (320 mg, 87%) as white diamond plates: mp 91–93 °C; $[\alpha]^{19}_{\text{D}} = +10.6^\circ$ ($c = 0.5$, MeOH); IR (film) 3341, 3058, 2979, 1705, 1651, 1602, 1591, 1495, 1456, 1376, 1341, 1250, 1157, 1072, 780, 741 cm^{-1} ; NMR (CDCl_3) δ 1.31 (3H, d, $J = 6.8$ Hz), 1.59 (3H, s), 3.27 (1H, d, $J = 14.6$ Hz), 3.44 (1H, d, $J = 14.4$ Hz), 4.95–5.07 (3H, m), 5.36 (1H, s), 5.90–6.10 (1H, br s), 6.42 (1H, d, $J = 7.8$ Hz), 6.75–6.81 (3H, m), 6.84 (1H, d, $J = 7.3$ Hz), 7.05–7.35 (9H, m), 7.57 (1H, d, $J = 7.8$ Hz), 8.09 (1H, s); MS m/e (CI^-) 472 ($\text{M}^+ + \text{H}$, 2), 471 (M^+ , 0.2), 348 (23), 322 (8), 244 (27), 131 (11), 130 (60), 107 (100), 95 (32). Anal. ($\text{C}_{28}\text{H}_{29}\text{N}_3\text{O}_4$) C, H, N.

[R(R*,S*)]-[2-(1*H*-Indol-3-yl)-1-methyl-1-[(1-phenylethyl)carbamoyl]ethyl]carbamic acid 2-fluoro-5-methylbenzyl ester (test9): mp 104–106 °C; $[\alpha]^{23}_{\text{D}} = +7.7^\circ$ ($c = 0.16$, MeOH); IR (film) 3334, 2927, 1715, 1652, 1505, 1456, 1251, 1072, 742 cm^{-1} ; NMR (CDCl_3) δ 1.28 (3H, d, $J = 7.2$ Hz), 1.60 (3H, s), 2.26 (3H, s), 3.24 (1H, d, $J = 14.8$ Hz), 3.45 (1H, d, $J = 14.8$ Hz), 4.95–5.03 (1H, m), 5.06 (1H, d, $J = 12.5$ Hz), 5.12 (1H, d, $J = 12.0$ Hz), 5.30–5.40 (1H, m), 6.34 (1H, d, $J = 7.8$ Hz), 6.79–7.32 (12H, m), 7.56 (1H, d, $J = 8.1$ Hz), 7.90–7.95 (1H, br s); MS m/e (FAB) 489.6 (17), 488.3 ($\text{M}^+ + \text{H}$, 100), 444.3 (9), 358 (12), 322.6 (12), 304.3 (18), 295.3 (22). Anal. ($\text{C}_{29}\text{H}_{30}\text{FN}_3\text{O}_3$) C, H, N.

Acknowledgment. The authors wish to thank Mrs. Nirmala Suman-Chauhan for generation of NK1 binding data.

References

- (1) Von Euler, U. S.; Gaddum, J. H. An unidentified depressor substance in certain tissue extracts. *J. Physiol. (London)* **1931**, *72*, 577–583.
- (2) Chang, M.; Leeman, S. E. Isolation of a sialogogic peptide from bovine hypothalamic tissue and its characterization as substance P. *J. Biol. Chem.* **1970**, *245*, 4784–4790.
- (3) Erspamer, V. The tachykinin peptide family. *Trends Neurosci.* **1981**, *4*, 267–269.
- (4) Negri, L.; Melichiorri, P. Nonmammalian Tachykinins: their Contribution to the Discovery and Biological Characterization of the Mammalian Neurokinin System. *Regul. Pept.* **1988**, *22*, 13–17.
- (5) Wang, J.-X.; Dipasquale, A. J.; Bray, A. M.; Maeji, N. J.; Spillmeyer, D. C.; Geysen, H. M. Systematic Study of Substance P analogs. *Int. J. Pept. Protein Res.* **1993**, *42*, 392–399.
- (6) Regoli, D.; Drapeau, G.; Dion, S.; Couture, R. New Selective Agonists for Neurokinin Receptors: Pharmacological Tools for Receptor Characterization. *Trends Pharmacol. Sci.* **1988**, *9*, 290–295.
- (7) Guard, S.; Watson, S. P. Tachykinin Receptor Types: Classification and Transmembrane Signalling Mechanisms. *Neurochem. Int.* **1991**, *18*, 149–165.
- (8) Pernow, B. Substance P. *Pharmacol. Rev.* **1983**, *35*, 85–141.
- (9) Maggio, J. E. Tachykinins. *Annu. Rev. Neurosci.* **1988**, *11*, 13–28.
- (10) Leeman, S. E.; Hammerschlag, R. Stimulation of Salivary Secretion by a Factor Extracted from Hypothalamic Tissue. *Endocrinology* **1967**, *81*, 803–810.

- (11) Hagiwara, D.; Miyake, H.; Morimoto, H.; Murai, M.; Fujii, T.; Matsuo, M. The Discovery of a Tripeptide Substance P Antagonist and its Structure-Activity Relationships. *J. Pharmacobiodyn.* **1991**, *14*, 5104.
- (12) Cascieri, M. A.; Huang, R.-R. C.; Fong, T. M.; Cheung, A. H.; Sadowski, S.; Ber, E.; Strader, C. D. Determination of the Amino Acid Residues in Substance P Conferring Selectivity and Specificity for the Rat Neurokinin Receptors. *Mol. Pharmacol.* **1992**, *41*, 1096-1099.
- (13) Hagiwara, D.; Miyake, H.; Morimoto, H.; Murai, M.; Fujii, T.; Matsuo, M. Studies on Neurokinin Antagonists. 1. The Design of Novel Tripeptides Possessing the Glutamyl-D-tryptophyl-phenylalanine Sequence as Substance P Antagonists. *J. Med. Chem.* **1992**, *35*, 2015-2025.
- (14) Hagiwara, D.; Miyake, H.; Morimoto, H.; Murai, M.; Fujii, T.; Matsuo, M. Studies on Neurokinin Antagonists. 2. Design and Structure-Activity Relationships of Novel Tripeptide Substance P Antagonists, N^α-[N^α-Acetyl-L-threonyl]-N¹-formyl-D-tryptophyl]-N-methyl-N-(phenyl-methyl)-L-phenylalaninamide and Its Related Compounds. *J. Med. Chem.* **1992**, *35*, 3184-3191.
- (15) Hagiwara, D.; Miyake, H.; Murano, K.; Morimoto, H.; Murai, M.; Fujii, T.; Nakanishi, I.; Matsuo, M. Studies on Neurokinin Antagonists. 3. Design and Structure-Activity Relationships of New Branched Tripeptides N^α-(Substituted L-aspartyl, L-ornithyl, or L-lysyl)-N-methyl-N-(phenylmethyl)-L-phenylalaninamides as Substance P Antagonists. *J. Med. Chem.* **1993**, *36*, 2266-2278.
- (16) Snider, R. M.; Constantine, J. W.; Lowe, J. A., III; Longo, K. P.; Lebel, W. S.; Woody, H. A.; Drozda, S. E.; Desai, M. C.; Vinick, F. J.; Spencer, R. W.; Hess, H.-J. A Potent Nonpeptide Antagonist of the Substance P (NK₁) Receptor. *Science* **1991**, *25*, 435-437.
- (17) Garret, C.; Carruette, A.; Fardin, V.; Moussaoui, S.; Peyronel, J.-F.; Blanchard, J.-C.; Laduron, P. M. Pharmacological Properties of a Potent and Selective Nonpeptide Substance P Antagonist. *Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10208-10212.
- (18) Lawrence, K. B.; Venepalli, B. R.; Appell, K. C.; Goswami, R.; Logan, M. E.; Tomczuk, B. E.; Yanni, J. M. Synthesis and Substance P Antagonist Activity of Naphthimidazolium Derivatives. *J. Med. Chem.* **1992**, *35*, 1273-1279.
- (19) MacLeod, A. M.; Merchant, K. J.; Cascieri, M. A.; Sadowski, S.; Ber, E.; Swain, C. J.; Baker, R. J. N-Acyl-L-tryptophan Benzyl Esters: Potent Substance P Receptor Antagonists. *J. Med. Chem.* **1993**, *36*, 2044-2045.
- (20) McLean, S.; Ganong, A.; Seymour, P. A.; Sinder, R. M.; Desai, M. C.; Rosen, T.; Bryce, D. K.; Longo, K. P.; Reynolds, L. S.; Robinson, G.; Schmidt, A. W.; Siok, C.; Heym, J. Pharmacology of CP-99,994: a Nonpeptide Antagonist of the Tachykinin-1 Receptor. *J. Pharmacol. Exp. Ther.* **1993**, *267*, 472-479.
- (21) Boyle, S.; Guard, S.; Higginbottom, M.; Horwell, D. C.; Howson, W.; McKnight, A. T.; Martin, K.; Pritchard, M. C.; O'Toole, J.; Raphy, J.; Rees, D. C.; Roberts, E.; Watling, K. J.; Woodruff, G. N.; Hughes, J. Rational Design of High Affinity Tachykinin NK1 Receptor Antagonists. *Bioorg. Med. Chem.* **1994**, *2*, 357-370.
- (22) Higginbottom, M.; Kneen, C.; Ratcliffe, G. S. Rationally Designed "Dipeptoid" Analogues of CCK. A Free-Wilson/Fujita-Ban Analysis of some α -Methyltryptophan Derivatives as CCK-B Antagonists. *J. Med. Chem.* **1992**, *35*, 1572-1577.
- (23) Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P.; Geiger, E.; Streich, M. Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients. *J. Am. Chem. Soc.* **1963**, *85*, 2817-2824.
- (24) Hansch, C.; Fujita, T. ρ - σ - π Analysis. A Model for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.
- (25) Topliss, J. G. Some Observations on Classical QSAR. *Perspect. Drug Discovery Des.* **1993**, *1*, 253-268.
- (26) Agarwal, A.; Pearson, P. P.; Taylor, E. W.; Li, H. B.; Dahlgren, T.; Herslof, M.; Yang, Y.; Lambert, G.; Nelson, D. L.; Regan, J. W.; Martin, A. R. Three-Dimensional Quantitative Structure-Activity Relationships of 5-HT Receptor Binding Data for Tetrahydropyridinylindole Derivatives: A Comparison of the Hansch and CoMFA Methods. *J. Med. Chem.* **1993**, *36*, 4006-4014.
- (27) Labanovski, J.; Motoc, I.; Naylor, C. B.; Mayer, D.; Dammkoehler, R. A. Three-Dimensional Quantitative Structure-Activity Relationships. 2. Conformational Mimicry and Topographical Similarity of Flexible Molecules. *Quant. Struct.-Act. Relat.* **1986**, *5*, 138-152.
- (28) Carbo, R.; Leyda, L.; Arnan, M. An Electron Density Measure of the Similarity Between Two Compounds. *Int. J. Quant. Chem.* **1980**, *17*, 1185-1189.
- (29) Hodgkin, E. E.; Richards, W. G. Molecular Similarity Based on Electrostatic Potential and Electric Field. *Int. J. Quant. Chem. Quant. Biol. Symp.* **1987**, *14*, 105-110.
- (30) Hopfinger, A. J. Theory and Analysis of Molecular Potential Energy Fields in Molecular Shape Analysis: A QSAR Study of 2,4-diamino-5-benzylpyrimidines as DHFR Inhibitors. *J. Med. Chem.* **1983**, *26*, 990-996.
- (31) Doweiko, A. M. The Hypothetical Active Site Lattice. An Approach to Modeling Active Site from Data on Inhibitor Molecules. *J. Med. Chem.* **1988**, *31*, 1396-1406.
- (32) Ghose, A. K.; Crippen, G. M. Use of Physicochemical Parameters in Distance Geometry and Related Three-Dimensional Quantitative Structure-Activity Relationships: A Demonstration Using Escherichia Coli Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1985**, *28*, 333-346.
- (33) Kato, Y.; Itai, A.; Iitake, Y. A Novel Method for Superimposing Molecules and Receptor Mapping. *Tetrahedron* **1987**, *43*, 5229-5236.
- (34) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3-D 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.
- (35) 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.
- (36) Burt, C.; Huxley, P.; Richards, W. G. The Application of Molecular Similarity Calculations. *J. Comput. Chem.* **1990**, *11*, 1139-1146.
- (37) Good, A. C.; Hodgkin, E. E.; Richards, W. G. Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188-191.
- (38) Good, A. C.; Richards, W. G. Rapid Evaluation of Shape Similarity Using Gaussian Functions. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112-116.
- (39) Wold, S.; Wold, H.; Ruhe, A.; Dunn, W. J., III. The Colinearity Problems in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735-743.
- (40) Stahle, L.; Wold, S. Principal Components Analysis and Partial Least Squares Regression. In *Progress in Medicinal Chemistry*; Ellis, G. P., West, G. B., Eds.; Elsevier: Amsterdam, The Netherlands, 1988; Vol. 25, p 1.
- (41) Good, A. C.; Peterson, S. J.; Richards, W. G. QSAR's from Similarity Matrices. Technique Validation and Application in the Comparison of Different Similarity Evaluation Methods. *J. Med. Chem.* **1993**, *36*, 2929-2937.
- (42) Sybyl 6.00/6.10. Tripos Assoc. Inc., 1699 S. Hanly Rd., St. Louis, MO 63144.
- (43) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982-1012.
- (44) (a) Abraham, R. J.; Smith, P. E. Charge Calculations in Molecular Mechanics IV: A General Method for Conjugated Systems. *J. Comput. Chem.* **1987**, *9*, 288-297. (b) Abraham, R. J.; Smith, P. E. Charge Calculations in Molecular Mechanics 7: Application to Polar π Systems Incorporating Nitro, Cyano, Amino, C=S and Thio Substituents. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 175-187. (c) Abraham, R. J.; Grant, G. H. Charge Calculation in Molecular Mechanics. IX. A General Parametrisation of the Scheme for Saturated Halogen, Oxygen and Nitrogen Compounds. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 273-286.
- (45) Tsar, 2.20. Oxford Molecular Ltd., The Magdalen Centre Oxford Science Park, Standford-on-Thames, Oxford OX4 4GA, U.K.
- (46) Burt, C.; Richards, W. G. Molecular Similarity: The Introduction of Flexible Fitting. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 231-238.
- (47) Meyer, A. Y.; Richards, W. G. Similarity of Molecular Shape. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 427-439.
- (48) Draper, N. R.; Smith, H. *Applied Regression Analysis*, 2nd ed.; John Wiley & Sons: New York, 1981; p 1.
- (49) Cramer, R. D., III. Partial Least Squares (PLS): Its Strengths and Limitations. *Perspect. Drug Discovery Des.* **1993**, *1*, 269-278.
- (50) Topliss, J. G.; Costello, R. J. Chance Correlations in Structure-Activity Studies Using Multiple Regression Analysis. *J. Med. Chem.* **1972**, *15*, 1066-1068.
- (51) Tute, M. S. Quantitative Drug Design. In *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, U.K., 1990; Vol. 4, pp 1-31.
- (52) Craig, P. N. Interdependence between Physical Parameters and Selection of Substituent Groups for Correlation Studies. *J. Med. Chem.* **1971**, *14*, 680-684.
- (53) Bhaskar Kanth, J. V.; Periasamy, M. Selective Reduction of Carboxylic Acids into Alcohols using NaBH₄ and I₂. *J. Org. Chem.* **1991**, *56*, 5964-5965.
- (54) Paradisi, P. M.; Zecchini, P. G.; Torrini, I. Selective Acylations of Aminophenols and Hydroxyalkylphenols with 1-acetyl-*v*-triazolo[4,5-*b*]pyridine. *Tetrahedron Lett.* **1986**, *27*, 5029-5032.