Rationalizing the Activities of Diverse Cholecystokinin 2 Receptor Antagonists Using Molecular Field Points

Caroline M. R. Low*,† and J. G. Vinter‡

James Black Foundation, 68 Half Moon Lane, London, SE24 9JE, United Kingdom, and Cresset BioMolecular Discovery Ltd., BioPark Hertfordshire, Broadwater Road, Welwyn Garden City, Hertfordshire, AL7 3AX, United Kingdom

Received July 23, 2007

Cholecystokinin 2 receptor antagonists encompass a wide range of structures. This makes them unsuitable candidates for existing 3D-QSAR methods and has led us to develop an alternative approach to account for their observed biological activities. A diverse set of 21 antagonists was subjected to a novel molecular field-based similarity analysis. The hypothesis is that compounds with similar field patterns will bind at the same target site regardless of their underlying structure. This initial report demonstrates a linear correlation between ligand similarity and biological activity for this challenging data set. A model generated with three molecules was used to predict the activity of 18 test compounds, with different chemotypes, with a root-mean-square error of 0.68 pK_B units. The ability to automatically derive a molecular alignment without knowledge of the protein structure represents an improvement over existing pharmacophore methods and makes the method particularly suitable for scaffold-hopping.

Introduction

Antagonism of the cholecystokinin 2 (CCK_2[a]) G-protein coupled receptor (GPCR) represents an attractive pharmaceutical target with a potential role in treating gastric-acid related conditions,[1] as well as gastrointestinal[2,3] and pancreatic cancers.[4] Receptors for CCK are designated CCK_1 and CCK_2 on the basis of their affinity for the peptide agonists CCK and gastrin. CCK_2 receptors are found throughout the brain with the highest densities in the cerebral cortex, nucleus caudatus, and limbic system. They also regulate stomach acid release at a deeper control level than H/KATPases or histamine. These receptors are activated by the hormone gastrin, a 33-residue peptide that shares the same active C-terminal pentapeptide sequence with cholecystokinin, a related hormone that activates both CCK_1 and CCK_2 receptors. CCK_2 antagonists ideally should be selective against CCK_1 receptors, which are mainly localized in the periphery where they mediate pancreatic enzyme secretion and gallbladder contraction.

A number of diverse compounds have been described as competitive antagonists of CCK_2 receptors (Scheme 1). These include compounds such as the indole **4** (JB93182),[5] peptoid **13** (PD134308),[6] and peptides **16**[7] and **21**,[8] all designed from the structure of the C-terminal sequence of gastrin (Boc-Trp-Met-Asp-Phe-NH_2). The natural product, Asperlicin, has provided an alternative stream of compounds that includes the benzodiazepines **1** (YF476)[9] and **14** (L365260).[10] More recently, imidazoles **7** and **18** were derived by scaffold-hopping[11,12] from indole **4**. In some cases, the structures contain similar elements arranged differently, such that we can envisage defining a relationship between indole **4** and imidazole **7** that superposes the common adamantane, acid, or aromatic groups (Scheme 2). However, the basic benzodiazepine **1** is one of the most potent compounds described to date.[9] This compound contains several aromatic rings and a bulky hydrophobic *t*-butyl group, but also contains two basic nitrogens and no carboxylic acids.

Despite these variations in structure, there is evidence from site-directed mutagenesis that the $CCK_{30–33}$ tetramer and a variety of nonpeptides, including benzodiazepines, quinazolines, and peptoids, all interact with a common residue (Asn353) on the human CCK_2 receptor.[13,14] Taken together, this represents good evidence that the CCK_2 receptor can accommodate a diverse set of structures in the same binding site as the hormone agonist.

Conventional molecular modeling, docking, and virtual screening techniques are based on knowledge of the target binding site or well-characterized antagonist structural cores. CCK_2 receptor antagonists, so far described, represent a wide range of chemotypes, and it is consequently difficult to derive meaningful structure–activity relationships on structural grounds. Furthermore, CCK_2 is a GPCR and, as with most members of this important receptor family, there are no available direct structural data[15] on which to base a molecular model.

Three-dimensional quantitative structure–activity relationship (3D QSAR) methods can be used to derive ligand-based models to estimate the activities of new compounds. Some methods also provide a graphical output indicating regions where increases in affinity might be expected from modifying physical properties such as steric bulk, partial charge, hydrophobicity, or hydrogen-bond donor/acceptor ability. CoMFA[16] (comparative molecular field analysis) and CoMSIA[17] (comparative molecular similarity indices analysis) are well-known examples of these techniques. These methods compare molecules in terms of grid-based field energies or similarity indices and use partial least-squares statistics to generate models that have been widely applied to medicinal chemistry problems. The major disadvantage of most of these methods is that the results are heavily dependent on the choice of bioactive conformation and the way in which compounds are superposed. This has tended to limit their use to congeneric series, where the influence of the operator's choice is minimized.

* To whom correspondence should be addressed: Caroline M. R. Low, Drug Discovery Facility, Imperial College, London, SW7 2AZ, U.K. Phone: +44 20 7594 9776. E-mail: c.low@imperial.ac.uk.

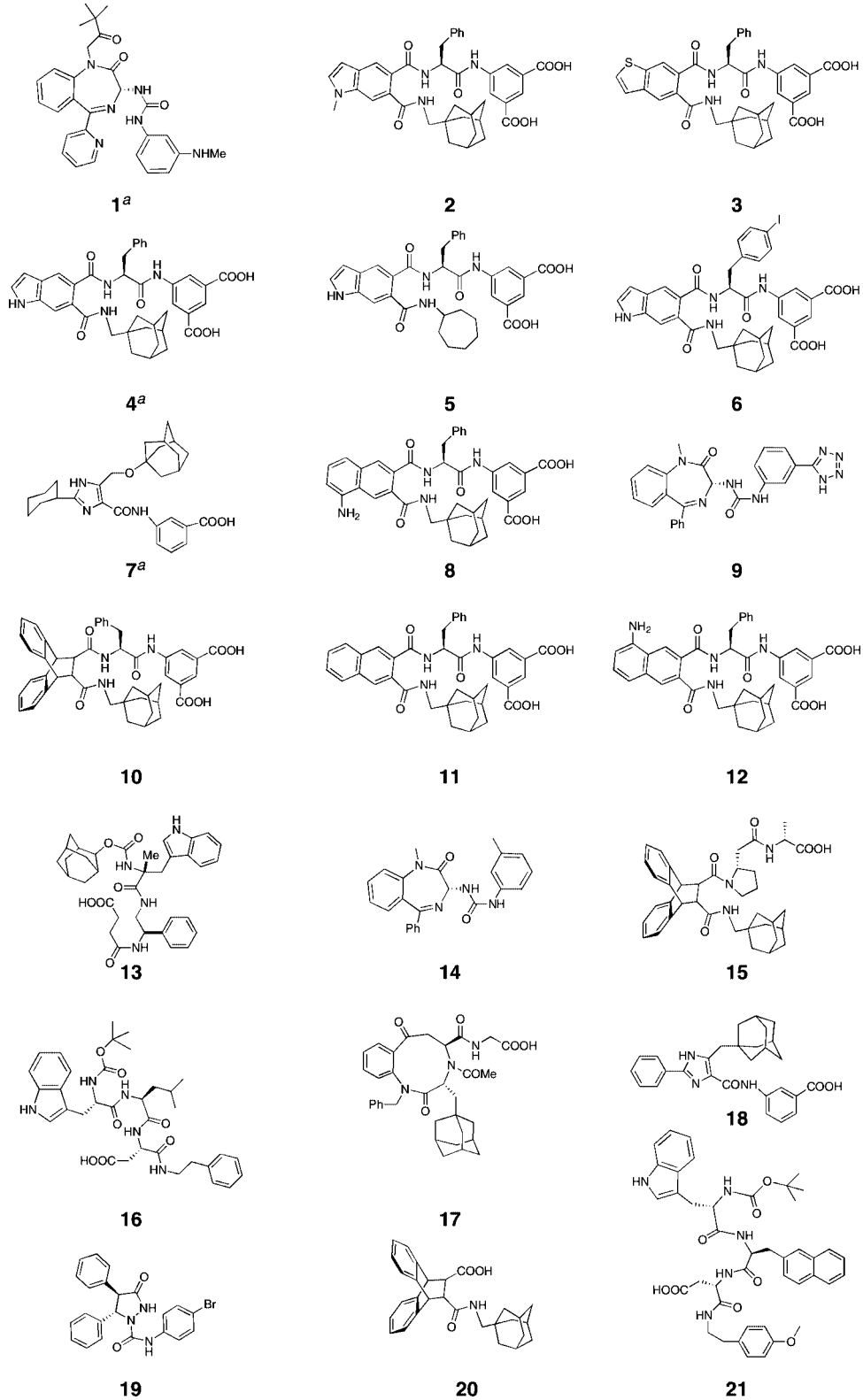
† James Black Foundation.

‡ Cresset BioMolecular Discovery Ltd.

[a] Abbreviations: CCK, cholecystokinin; 3D QSAR, three-dimensional quantitative structure–activity relationship; GPCR, G-protein coupled receptor; CoMFA, comparative molecular field analysis; CoMSIA, comparative molecular similarity indices analysis; G*sim*, global similarity score.

10.1021/jm070880t CCC: $40.75
Published on Web 01/18/2008

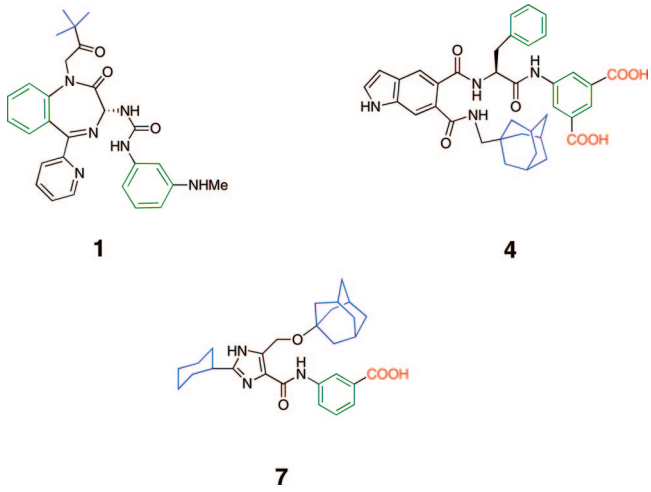
Scheme 1. Structures of CCK_2 Receptor Antagonists Used in this Study[a]



[a] The data set contains at least seven distinct chemotypes. Compounds **1**, **4**, and **7** were used to construct the receptor field point template.

An alternative approach is to derive a pharmacophore hypothesis that attempts to encompass common features of these molecules. One such study has used the Catalyst/Hypogen software (v4.7)[18] to develop a hypothesis for six structurally diverse series of CCK_2 receptor antagonists.[19] Construction of the training set required input from 33 compounds. The authors concluded that their best model consisted of four features: two hydrogen bond donors, one hydrophobic aliphatic, and one hydrophobic aromatic. The linear regression coefficient for actual versus predicted activity (R^2) was 0.78 for the training set. So it seems that the diversity of these structures does not preclude generation of a pharmacophore hypothesis with good predictive powers. Selection of the training set is recognized as a crucial aspect in creating good models using this approach

Scheme 2. Constituents of CCK_2 Receptor Field Point Template[a]



1 4 7

[a] Common bulky hydrophobic (blue), carboxylic acid (red), and aromatic (green) groups that could be superimposed as the basis of a structure-based pharmacophore hypothesis are highlighted.

and a minimum of 16 structurally diverse compounds are claimed to be required to avoid any chance correlation.[20] In addition, there should be no redundancy in structural features, and the biological activities of the set should span 4–5 orders of magnitude. These requirements suggest that this method can only be used in a mature field and is less suited to early drug development.

In reality, molecules interact via their electronic properties: electrostatic and van der Waals forces. If two molecules with diverse structures interact with an enzyme or receptor in a similar way, their bound conformations will have similar properties, although this might not be immediately apparent from a consideration of their structures alone. The idea of a field pattern around a ligand is intuitively appealing as the main criterion for binding recognition and has been acknowledged for many years. Previously we have described in silico methods of defining molecular fields in a form that enables similarity comparisons across molecules in three dimensions and demonstrated how molecular fields can be used as nonstructural templates for defining similar biological behavior.[21] We have so far shown that field patterns can be used to align molecules that act at the same site by their common field pattern and derive the biologically active conformation of a ligand without access to any protein structural data ("Field Templating"). We have also established virtual screening tools using field patterns to search through compound databases looking for potential hits[24] ("Field Screening"). Field Templating and Field Screening rely on the assumption that those molecules whose field patterns are most similar to those of an active search molecule will be the ones most likely to show the same patterns of biological activity and should be chosen for further investigation. The results obtained with these methods suggest that our field points encapsulate the molecular properties of the ligands, as seen by the protein.

In this study we demonstrate that field-based similarities correlate linearly with biological activity even when considering multiple chemotypes. The field patterns of three potent and selective CCK_2 antagonists were amalgamated to give a ligand-based view of the active site of the receptor in field point terms. A test set of compounds was then selected from a very diverse collection of CCK_2 receptor–ligands and each compared to the "receptor template". The field overlay scores for the model system were compared to experimentally determined affinity estimates (pK_B values) for the compounds in a functional in vitro CCK_2 bioassay. This approach provides a novel 3D QSAR method with similar predictive abilities to Catalyst.

Methods

(a) Hardware and Software. All molecular modeling calculations were carried out on a linux cluster of 24 nodes under Mosix distribution control, a Silicon Graphics Octane 2 was used for graphical work running the in-house XEDRAW modeling package, and the PC-based program XEDVIEW, written specifically to handle field point representations, was used for routine visualization. Conformational analyses were carried out using the Xedex software, incorporating the XED force field.[22,23] The FieldTemplater program constructed the putative bioactive field template that was used by FieldAlign to compare across all test-set compounds (http://www.cresset-bmd.com/).[21,24,25]

(b) Fields.[21] If two diverse structures are known to act at the same protein active site, they will be making a similar set of interactions with the protein. We define a molecular field as positive and negative electrostatic regions, hydrophobic regions, and areas of maximum van der Waals attraction and calculate these in terms of the interaction of appropriate charged and neutral probes at and beyond the molecular surface.

(c) Single Molecule Field Overlay Principle. Because the computational load to overlay the complete fields of two molecules is unacceptable, we distill the field of the first molecule down to its extrema (its maxima and minima in each region) and use these "field points" to sample the field on the second molecule. The process is repeated for the second molecule on the first, and the average of the two scores is taken as the overlay score of the two molecules. To optimize the field alignment, the overlay score is used to drive a multivariate simplex minimizer through molecular rotation and translation.[21]

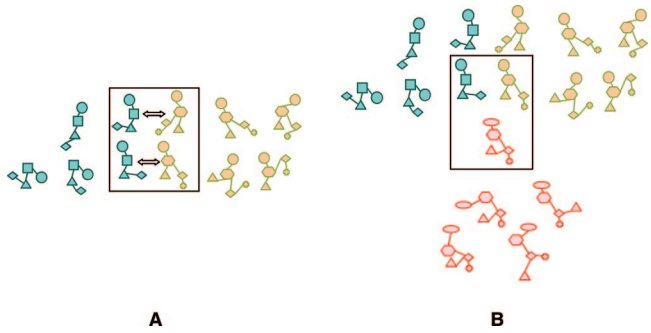
(d) Multiconformation Field Overlay. The bioactive conformations of molecules that are active at the same biological binding site are expected to have similar fields. If relevant structural data are not available from NMR, X-ray, or other experimental sources, many conformations must be investigated to find common field patterns from which to infer the bioactive shape. The most interesting targets (e.g., GPCRs and ion-channels) fall into this category.

We have investigated the field relationships of a set of CCK_2 receptor antagonists whose bioactive conformations are not known. To find the optimal field overlay of two molecules, the field of every conformer of each molecule was compared pairwise until a close field match was found. The conformations from each pair having the most similar fields are assumed to represent the bioactive conformations. Pairs of molecules found to have a "common" field are termed "duos". In practice, a representative collection of 50 molecular conformers was used and assumed to contain the bioactive conformation. This generally covered an energy range of 6 kcal/mol from the global minimum, as found by the Xedex conformation hunter.

(e) Field Template Generation. The conformations from two molecules often generate many duos with high field similarity (Scheme 3A). However, the chance that the "common" field is also generated by a third unrelated active molecule affords a useful refinement step. Therefore, by cross correlating all possible duos from three or more molecules acting at the same site, the conformers with the "common" field are more reliably identified as the bioactive conformers. The principle of cross-correlating duos to create a "trio" of the active site is shown in Scheme 3B.

Combining the "common" fields of progressively more diverse active molecules should increase the definition of the binding site requirements, forming a "field template" (the search molecule field pattern) for that binding site. This process is performed by our "FieldTemplater" software package.[24,25] The result is a set of "trio" templates from three active ligands, "quartets" from four active ligands, and so on, each having a higher probability of containing

Scheme 3. Schematic Representation of the Formation of (A) "Duos" of Common Fields from Pairwise Comparisons of Many Conformations of One Ligand (cyan) with Another (yellow): This Process Usually Identifies Many Potential Solutions; (B) Comparing a Further Set of Fields from a Third Ligand (red) Gives a "Trio" of Common Fields that Form the Basis of a "Field Template" of the Active Site[a]



[a] This cross-comparison reduces the number of common alignments and improves the chances that any trio templates found will include the bioactive conformations of the ligands in their correctly bound pose. As the fields of additional molecules are included, forming quartets, quintets, etc., the template will become better defined, assuming that the molecules are diverse in structure.

the experimentally correct alignment and, hence, the correct bound conformations from the ligand data alone. In this case, FieldTemplater was used to create a multistructural CCK_2 field template from three highly active antagonists **1**, **4**, and **7** (Scheme 2).

(e) Validation of Field Template. Once a multistructural field template had been generated from FieldTemplater, it was used as input to the FieldAlign software package. FieldAlign generates 3D conformers from a 2D molecule and then calculates a field similarity score for each conformer to the 3D field template. In this way it identifies, scores, and displays those conformers of a test ligand that best match the template. This software is capable of working with just one bioactive conformer as a field template or with a template made up from the fields of several individual structures. For the final predictive stage in this CCK_2 example (see Results and Discussion section), the FieldAlign protocol was used to score the field similarity of 50 representative conformations of 21 active ligands against the CCK_2 template.

(f) Field Overlay Scoring Metric *Gsim*. Three overlay scores are computed from which a single overall score (*Gsim*) is derived; raw overlay energy (E_0), field similarity (*Fsim*), and a volume overall similarity (*Vsim*). Full details of the derivation of these terms have been published.[22] E_0 is, by definition, specific to the system being examined and scores the optimized field overlay of a template with an entry conformation in energy units. E_0 is asymmetric and is symmetrized and normalized via a Dice field similarity metric between zero and unity to give *Fsim*. Maximizing this metric between two conformations gives both the best conformational overlay (in terms of field similarity) and a single field similarity score for aligning two conformations. *Fsim* helps to smooth out large molecular weight differences between molecules being compared. *Vsim* is a molecular volume overlay metric that tests how closely the two molecules being compared fit into a common space. This ensures that the field overlays are contained within a single putative active site. The average of *Fsim* and *Vsim* gives *Gsim*, a single universal similarity score used throughout this work.

Although the process of overlaying any two fields will result in many possible orientations, the overlay that scores highest may not correspond to the experimental overlay. However, in this study, no experimental data were available. In consequence, we considered only the highest scoring field overlays on the assumption that if the best overlay is considerably removed from the experimental overlay pose, bioactive correlations across the test set would be expected to be lost and lower lying overlays would then have to be considered.

Table 1. Biological Activities of CCK_2 Receptor Antagonists Used in this Study[a]

cmpd	CCK_2 RS $pK_B \pm$ sem	CCK_1 $pK_i \pm$ sem	ref
1 (YF476)	9.9 ± 0.3	6.4 ± 0.1	*9*
2	9.8 ± 0.3	5.4 ± 0.1	*28*
3	9.4 ± 0.3	6.1 ± 0.1	*5*
4 (JB93182)	9.3 ± 0.2	5.4 ± 0.1	*5*
5	9.3 ± 0.2	5.5 ± 0.1	*5*
6	9.1 ± 0.2	5.7 ± 0.1	*35*
7	9.1 ± 0.3	6.3 ± 0.1	*11*
8	9.0 ± 0.2	5.6 ± 0.1	
9	8.4 ± 0.2	6.5 ± 0.1	*32*
10	8.3 ± 0.2	5.7 ± 0.1	*36*
11	8.2 ± 0.1	6.1 ± 0.1	*5*
12	8.2 ± 0.2	5.6 ± 0.1	
13 (PD134308)	7.9 ± 0.2	6.2 ± 0.1	*6*
14 (L365260)	7.6 ± 0.2	6.5 ± 0.1	*10*
15	7.3 ± 0.2	4.6 ± 0.1	
16	6.9 ± 0.2	5.1 ± 0.1	*7*
17	6.8 ± 0.3	5.1 ± 0.1	*33*
18	6.6 ± 0.3	5.9 ± 0.1	*11*
19 (LY288513)	5.9 ± 1.4	4.7 ± 0.1	*5*
20	5.6 ± 0.3	[b]	*37*
21	5.4 ± 0.2	7.2 ± 0.1	*8*

[a] CCK_2 activities (pK_B values) were determined in an in vitro isolated lumen-perfused immature rat stomach bioassay (see text for details). $pK_B \pm$ sem values were estimated from single shifts of pentagastrin (*t*-Boc-CCK_{29-33}).[26] CCK_1 activity (pK_i values) were obtained from a radioligand binding assay using guinea pig pancreas cell homgenates, in competition with 20pM [^{125}I]BH-CCK-8S.[31] [b] Value not determined.

(g) Biology. The CCK_2 receptor activities of all compounds were examined in the isolated, lumen-perfused immature rat stomach.[26] $pK_B \pm$ sem values were estimated from single shifts of pentagastrin (*t*-Boc-$CCK_{29-33}NH_2$) concentration−effect curves, calculated assuming an underlying Schild slope of unity and fitted using the Gaddum−Schild equation. All compounds behaved as simple competitive antagonists in this assay, although we are aware that some examples behave as partial agonists in other bioassays.[14,27–30] This allowed us to obtain a consistent set of pure affinity estimates for the selected data set.

Activity at CCK_1 receptors was established by testing all compounds in a radioligand binding assay in guinea pig pancreas cell homogenates, in competition with 20 pM [^{125}I]BH-CCK-8S, as previously described.[31] Results were taken from at least three separate experiments.

(h) Data Set. The members of the data set were chosen on the basis of their structural diversity (Scheme 1), their wide activity profile (4.5 log units), and selectivity for CCK_2 over CCK_1 receptors (Table 1). A total of 21 compounds was selected from a large number of molecules made and tested in a functional CCK_2 receptor bioassay at the James Black Foundation over several years. It is difficult to obtain good correlations to data derived from multiple sources and so all compounds were tested in the same bioassay and the results for the complete set are reported here for the first time. All the members acted as simple competitive antagonists in the rat stomach (RS) functional in vitro assay[26] and so their pK_B values represent their affinities for the receptor (Table 1).

We are aware that a number of these compounds behave as partial agonists in other bioassays. These include the *N*-methyl indole derivative of **4**, compound **2**;[28] the benzodiazepine **14** (L365,260),[14] the first nonpeptide CCK_2 ligand to be described,[10] and close relatives of the peptoid **13**.[29,30] This behavior is not uncommon with weak partial agonists where differences in levels of receptor expression or cross-species differences can change the degree of agonist response observed. In one regard, the ability of these compounds to elicit the same biological response as the hormone in related CCK_2 bioassays supports the evidence that these compounds bind to the same site as Boc-$CCK_{30-33}NH_2$.[13,14]

A number of the selected compounds feature a (hetero)aromatic bearing identical adjacent carboxamide substituents, **2**–**4**, **8**, and

10–12. The activities of this congeneric subset range from pK_B 8.2 to 9.8, a 40-fold difference. A small set of 3(*R*)-phenylurea benzodiazepines, **1**,[9] **9**,[32] and **14**,[10] was also included. The affinities of this subset cover a range from pK_B 7.6 to 9.9. These two subsets were included to challenge our field technique to discriminate between compounds of the same class in addition to our aim of predicting the activities of compounds of different chemical classes.

We have previously described the scaffold-hopping process from indole **4** that led to imidazoles **7** and **18**,[11,12] and the design of the novel benzodiazonine **17**[34] from our laboratories. We also chose to include representative examples of compounds described by other groups. These include peptoid **13**,[6] as well as the lower affinity diphenylpyrazolidinone **19** (LY288513)[34] and peptide **16**,[7] a close relative of Boc-CCK$_{30-33}$NH$_2$. Finally, peptide derivative **21** was included as an example of a compound that has low affinity for the CCK_2 receptor (pK_B 5.4) and is selective for CCK_1 (pK_i 7.2).[8]

Because no X-ray structural data exist for the CCK_2 GPCR, the ligand bioactive conformations are not known. Most of the ligands are highly flexible, which required that we consider multiple conformations of each. Hence, individual compounds were processed as described in the Methods section and stored as a set of up to 50 conformations, each with its own unique field pattern.

Results and Discussion

We know from the high enrichment factors of our hit rates in field-based virtual screening projects that there is a qualitative relationship between field similarity and biological activity. We wished to discover whether a quantitative relationship between biological activity and field overlay score could be achieved, using the reliable and consistent biological results available for this data set. We also took the opportunity to test whether our field-based approach could distinguish CCK_2 from CCK_1 activity, as well as testing the quality of field-activity relationships on a GPCR target that binds complex and structurally diverse antagonists.

(a) Choosing the Field Template Members. Using any single antagonist to define a template would generate only the field information for the binding of that antagonist into its receptor site. Furthermore, selecting an appropriate conformation for these flexible compounds would be somewhat arbitrary. Our own NMR studies[12] and published X-ray data on some literature compounds[38] gave us limited knowledge of the isolated conformations of some of these compounds, but in general, we could only infer the nature of the bioactive conformation from structure–activity relationships. For example, our conformation hunter (Xedex) detected just one conformer for **13** within the 6 kcal/mol conformation energy search window imposed on the search. This was surprising because visual inspection suggested that the molecule would be expected to be very flexible. The cause of the single conformer stability was found to be a strong intramolecular hydrogen bond across the peptoid backbone and this was confirmed from comparison with an X-ray (refcode: VUWGOJ) of **13** found on the Cambridge Structural Database that showed a good match with the calculated conformer over all groups except the indole. Having confirmed that the match was exactly maintained after overlaying the fields of the calculated and X-ray structures (i.e., the indole field did not influence the field correspondence overall), we felt confident in asserting that this conformation of **13** is either the bioactive conformer or closely related to it in field terms. An X-ray structure of the 3-bromophenylurea analogue of **14** (L365260) and **9** is also available (refcode: PIHFOB).

Amalgamating information from other diverse CCK_2 antagonists might allow us to identify common pharmacophoric features responsible for their biological activity. For the CCK_2 set, templates were generated as "trios" of structures made up from three diverse chemotypes.

Table 2. Three Highest-Ranked CCK_2 Trio Templates Created by FieldTemplater from Structures **1**, **4**, and **7**[a]

		conformation number (energy-ranked)		
template	template rank	benzodiazepine **1**	indole **4**	imidazole **7**
T1	1	4	1	18
T2	2	1	1	18
T3	3	3	2	18

[a] Templates constructed by FieldTemplater are ranked in order of their field overlay similarity. Conformations are numbered from the calculated minimum energy structure (1) upward to the highest (50).

Table 3. Predicted CCK_2 Affinities (RS pK_B) from the Alignments of Trio T1 with Compounds **1–21**

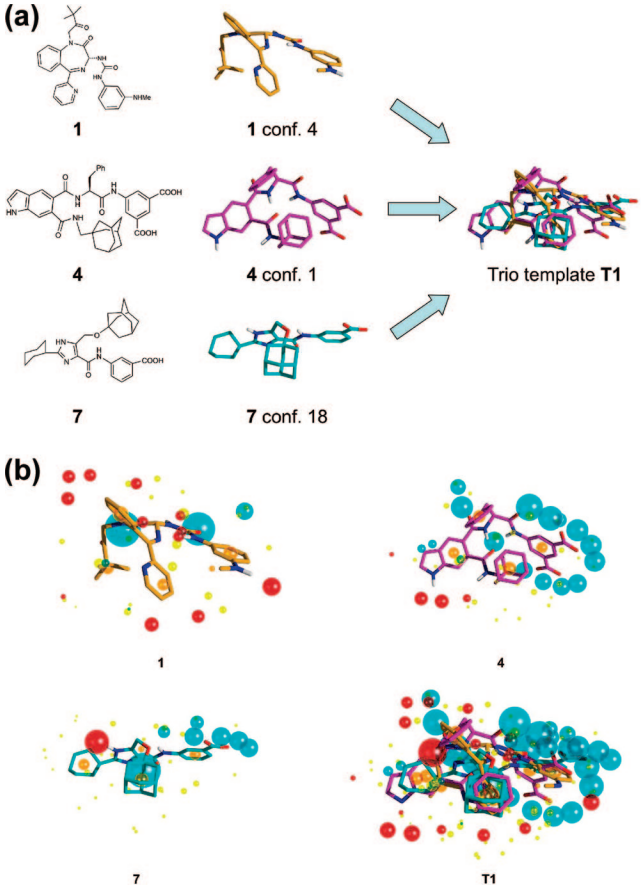
cmpd	conf[a]	*Gsim*	predicted pK_B	measured RS pK_B	measured−predicted
1	4	0.668	9.1	9.9	+0.8
2	1	0.668	9.1	9.8	+0.7
3	1	0.686	9.4	9.4	+0.0
4	1	0.697	9.5	9.3	−0.2
5	8	0.658	9.0	9.3	+0.3
6	3	0.602	8.3	9.1	+0.8
7	18	0.697	9.5	9.1	−0.4
8	1	0.668	9.1	9.0	−0.1
9	1	0.575	7.9	8.4	+0.5
10	1	0.609	8.3	8.3	0.0
11	1	0.641	8.8	8.2	−0.6
12	1	0.668	9.1	8.2	−0.9
13	1	0.488	6.7	7.9	+1.2
14	3	0.480	6.6	7.6	+1.0
15	7	0.487	6.7	7.3	+0.6
16	7	0.495	6.8	6.9	+0.1
17	9	0.494	6.8	6.8	0.0
18	13	0.545	7.5	6.6	−0.9
19	2	0.447	6.2	5.9	−0.3
20	1	0.495	6.8	5.6	−1.2
21	6	0.465	6.4	5.4	−1.0

[a] Column two records the conformer chosen by FieldAlign as the best overlay with T1. Conformer numbers are in ascending order of gas-phase energy from the global minimum, as defined by the Xedex conformation searcher.

Our previous studies had shown that the best models are generated using templates created from highly active compounds, which would be expected to contain the most relevant binding information. We chose three compounds as a training set. These were the benzodiazepine **1**,[9] the most potent compound described to date, whose structure was based on that of the natural product, asperlicin; the indole derivative **4**,[36] which was originally based on the structure of Boc-CCK$_{30-33}$NH$_2$, by way of an intermediate active set of bicyclooctanes (e.g., **10**)[5] and imidazole **7**, the product of a scaffold-hopping exercise that used an early version of the field comparison technique to improve bioavailability problems inherent in compounds related to indole **4**.[11,12]

(b) Field Templating. The templating protocol (see Methods) has been consolidated into a single software package called FieldTemplater. FieldTemplater takes three or more structures, optimally aligns the fields of their conformation sets, and outputs a series of templates. Each template in the series was ranked according to how well its constituents overlaid in field and volume space using the *Gsim* score. The three highest ranked templates from FieldTemplater for the trio combination are shown in Table 2. The top-ranking template (T1) was chosen as the master template on which to base the calculation of field similarities across the whole data set. Scheme 4a exemplifies the relationship of conformers of the constituent compounds to the trio field template T1. All of the conformations identified fall into the lowest 20% of those calculated. The field patterns

Scheme 4. (a) Structural and Conformational Makeup of Trio T1; Field Templater Overlaid the Field Patterns of all Conformers of **1**, **4**, and **7** To Find a Single Common Field Pattern Assumed To Reflect the Binding Requirements of the CCK_2 Receptor: The Three Conformers With the Most Similar Individual Field Patterns (Conformer 4 of **1**; conformer 1 of **4**, and Conformer 18 of **7**) were Returned as the T1 Trio Template; (b) Field Point Patterns of the Three Conformers of **1**, **4**, and **7** that Showed the Most Mutual Similarity and their Final Alignment in Trio Template T1[a]



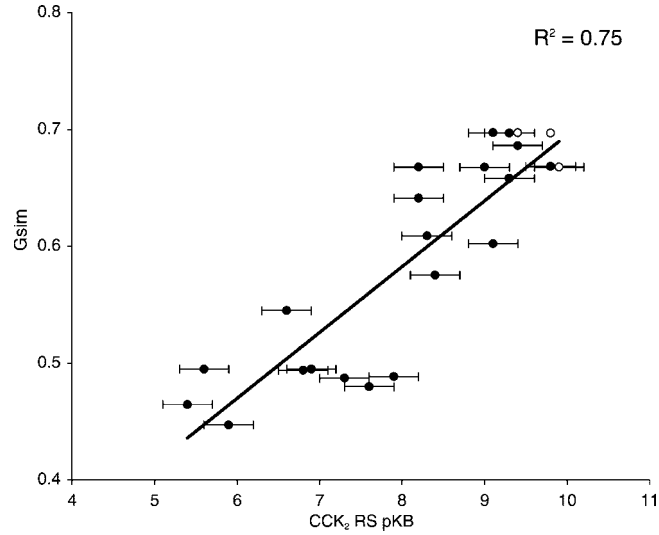
[a] The structural arrangement that underpin each field pattern are included for interpretative clarity. Field points are color coded as follows: negative charge, blue; positive charge, red; van der Waal's surface, yellow; hydrophobes, orange.

associated with each individual component of the template and the combined field of template T1 are illustrated in Scheme 4b.

(c) Correlation of Overlay Scores for Field Template T1 with Experimental Biological Activity. The field of each conformation of every member of the data set (**1**–**21**, Scheme 2) was optimally overlaid with that of template T1 to yield a combined field/volume similarity score (*Gsim*) using FieldAlign. In each case, only the top scoring alignment, the best fit between the field points of the template and ligand, was used to plot a linear regression of the experimental RS pK_B against *Gsim*. Scheme 5 plots the regression for template T1 over the complete active data set (including those compounds that made up the template). There is clearly a strong linear correlation between the score calculated for the field point model and the measured affinity for the CCK_2 receptor that extends over the full 4.5 log unit range. The R^2 value (0.75) and *p*-value ($<10^{-6}$) indicate that this is a good correlation.

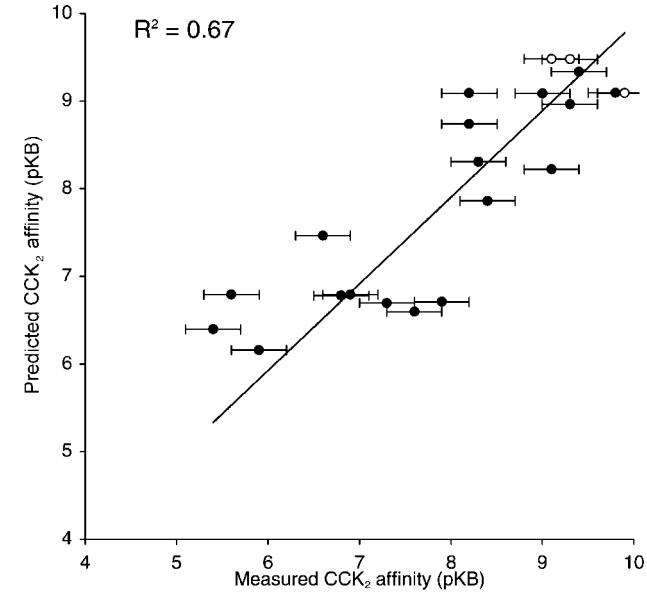
If we convert this to a plot of predicted affinity versus measured pK_B (Scheme 6), we can compare our results to those from the Catalyst study (Scheme 7).[19] The constrained (through

Scheme 5. Linear Least-Squares Fit of CCK_2 RS pK_B (Biological Affinity) vs the *Gsim* Score for Structures **1**–**21** of the Dataset Listed in Table *3* with the Trio T1 Template[a]



[a] Data for the three template compounds, **1**. **4**, and **7**, are shown as open circles (○) and those for the 18 test compounds as filled circles (●).

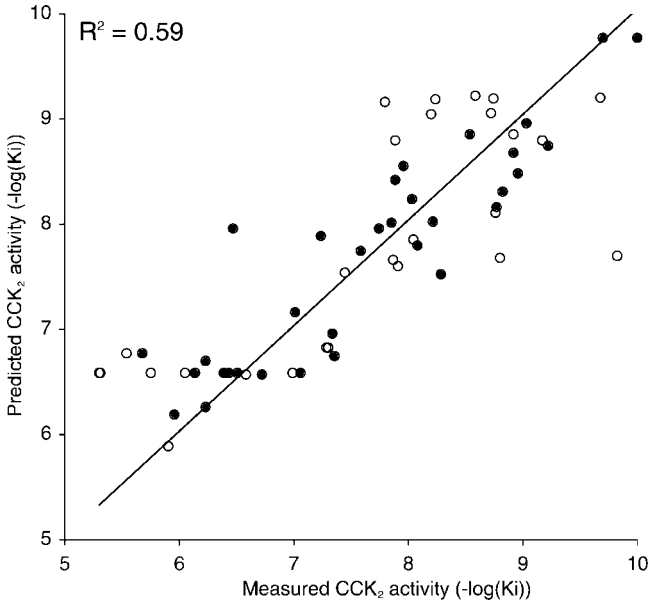
Scheme 6. Linear Least-Squares Fit of Predicted CCK_2 Affinity vs Measured Affinity (pK_B) for Compounds **1**–**21** ($R^2 = 0.67$, *p*-Value $< 10^{-6}$)[a]



[a] Data for the 33 compounds used to derive the pharmacophore hypothesis are shown as open circles (○) and those for the 27 test compounds are shown as filled circles (●).

zero) R^2 value from our study (0.67), which includes data from 18 test compounds, in addition to the three used to generate the field point template, is higher than that reported for the combined training and test sets of the Catalyst/Hypogen model (0.59). At this stage, we cannot comment on whether these differences are significant, as the two data sets are not identical. Nevertheless, it indicates that the field point model is capable of predicting the biological activities of this diverse set of compounds to a comparable, if not better, degree than Catalyst/Hypogen. Analysis of the residuals for the measured and predicted pK_B values in Scheme 6 indicates that there is no significant systematic error in the predictions made using the field template (mean residual is 0.07 ± 0.15). The standard

Scheme 7. Linear Least-Squares Fit of Predicted CCK_2 Activity ($-\log(K_i)$) vs Measured Activity for 60 Structures Used in Catalyst/Hypogen Study (Replotted from Literature[19] $R^2 = 0.59$, p-Value $< 10^{-16}$)[a]



[a] Data for the 33 compounds used to derive the pharmacophore hypothesis are shown as open circles (○) and those for the 27 test compounds are shown as filled circles (●).

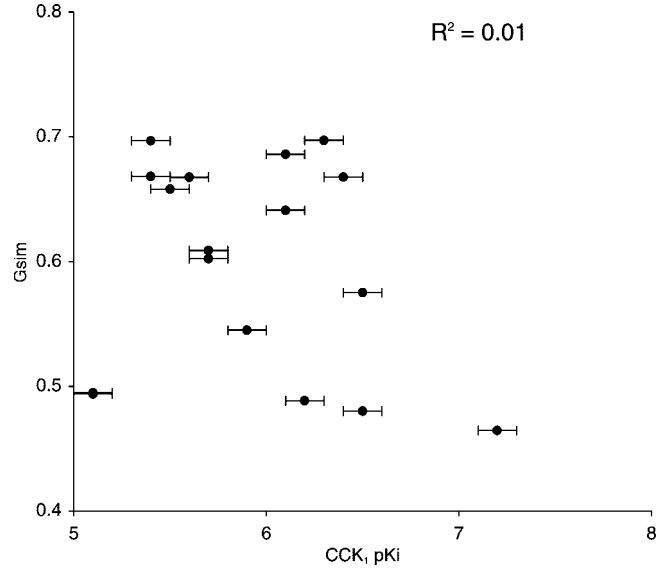
deviation for the predicted values is highly acceptable at 0.68 and is of the same order as that for the measured pK_B values. The predictions reported for the Catalyst/Hypogen model (Scheme 7) show a slight tendency to under read (mean residual is -0.14 ± 0.16) and the standard deviation is higher (0.82).

Values predicted by us for the three 3(*R*)-phenylurea benzodiazepines, **1**,[9] **9**,[32] and **14**,[10] place these three compounds in the correct rank order. This is not true of the subset of compounds **2**–**4**, **8**, and **10**–**12**, whose structures all contain identical adjacent carboxamide substituents attached to a variety of (hetero)aromatics. However, all were correctly predicted to be potent compounds and there are only two cases (compounds **2** and **12**), where the estimated affinities are greater than 1 standard deviation from the measured value. The reasons for these inaccuracies are unclear and have prompted careful examination of the field point overlays. These results lead us to conclude that the small actual energy differences involved in increasing the binding affinity of a compound from pK_B 8.2 to 9.9 are not reflected with sufficient accuracy by our current methodology. We hope to tackle this by extending the number of compounds examined in this system or applying this approach to systems where the nature of the protein binding site has been experimentally determined and can provide additional constraints to the template design.

Nevertheless, a significant difference between this report and the Catalyst study is that good predictions were obtained for chemotypes that are not present in the training set. In contrast, all members of the Catalyst study test set were represented in the training set. These results may suggest that the field point approach is particularly suited to generating new leads by scaffold hopping rather than simple optimization of a known core structure.

(d) Template Selectivity: Correlations with CCK_1 Activity. It is clearly important that any QSAR modeling technique can distinguish activity at the target receptor from that at a closely related subtype. In the case of the cholecystokinin receptors, this is complicated by the observation that enantiomeric pairs of ligands are known to show opposite subtype selectivities. For example, a number of independent reports have described opposite CCK_1/CCK_2 selectivities for enantiomeric pairs of peptoid,[39] benzodiazepine,[10,40] and diphenylpyrazolidinone[34] ligands that are close relatives of compounds in this study. However, given the obvious dependence of fields on structure, substitution, and charge state across the whole molecule and their subtler dependence on absolute configuration and conformation, the template should be able to discriminate activity at the CCK_2 receptor from that at other receptors, assuming the two subsites in question utilize even slightly different conformations. This led us to investigate whether there was any correlation to biological activity at the closely related CCK_1 receptor with the T1 template across the 20 compounds whose CCK_1 activity were measured (Table 1). Regression analysis returned an R^2 value of 0.01 (Scheme 8), verifying no significant correlation between the T1 trio template and the CCK_1 assay data, indicating that the T1 template was selective for CCK_2 over CCK_1 antagonists.

Scheme 8. Plot of CCK_1 pK_i vs the *Gsim* Score for Structures **1**–**19** and **21** of the Dataset with the Trio T1 Template



(e) Graphical Interpretation of Results. 3D-QSAR methods have two aims: predict the activity of new compounds and suggest modifications that might lead to increased activity. Existing methods fulfill this second aim to different extents. Catalyst/Hypogen can indicate ways in which a compound does not fit a pharmacophore hypothesis but is limited to five points of reference that describe the influence of isolated features such as hydrogen-bond donors/acceptors and alkyl and aromatic groups. CoMFA and CoMSIA both produce field contribution maps that allow physicochemical properties responsible for binding to be mapped back onto molecular structures. This facility is a strength of these techniques and has been extensively used in lead optimization programs.

Field points represent positions of maximum interaction of a molecule with its electrostatic, steric and hydrophobic surroundings. Comparing the field pattern of an individual test molecule with the "active template" pattern can reveal where chemical variation might be made to optimize the field pattern toward higher activity. The origins of field points are generally intuitive, given a basic knowledge of organic chemistry, and are generated from the properties of the whole molecule rather than isolated features, such as hydrogen bonds or π-electron clouds. This leads

to a more subtle and sensitive gauge of molecular potential than from conventional pharmacophores and QSAR descriptors. Correlating the similarity of the field pattern of prospective test molecules against the field point template affords a ready means of predicting a pK_B value for the new compound and establishing priority for synthesis. We have not yet developed software that is specifically directed toward the lead optimization process. Nevertheless, we have already shown that it is possible to scaffold hop from one chemotype to another using a basic understanding of the origins of particular field point patterns[12] and plan to investigate this aspect further.

Conclusion

The structural diversity of existing CCK_2 receptor antagonists represents a unique data set ideally suited to investigating the molecular properties that determine ligand binding. Without the knowledge of the bioactive conformations or a means of molecular alignment, both of which are provided by the field point approach, it would be a challenge for techniques such as CoMFA and CoMSIA to analyze these flexible molecules and provide an appropriate rationale on which to relate one chemotype to another. In this paper, we have validated the proposal that field comparisons correlate with biological activity. By using a small set of three diverse structures, without the advantages of protein X-ray data, a consistent picture of the active site of the receptor was built up, based on structure-free molecular fields. These patterns of field points provide a representation of the whole molecule based on real physical interactions and divorce the concept of chemotype from structure. Useful predictive value was obtained from this model that has significantly contributed to the understanding of ligand binding modes and helped to progress the project. Furthermore, no correlation to activity at the CCK_1 receptor subtype was found, demonstrating that the derived CCK_2 active site field is selective for CCK_2 over CCK_1.

The results obtained using this method compare well with those obtained using Catalyst/Hypogen to create a pharmacophore hypothesis for a similar set of compounds.[19] The advantages of the field point method are that significantly fewer compounds are required to derive the initial model (3 vs 33). It is notable that the field point method gave good results despite the absence of several chemotypes from the training set, in contrast to the Catalyst study where all members of the test set were represented. This last point reiterates the advantages of developing methods for comparing molecules that are not tied to traditional representations of chemical structure and can naturally handle molecular diversity. We hope that these will improve our insight into the factors that influence the protein's view of a ligand and, in turn, our ability to design new drugs.

Acknowledgment. The authors would like to thank all at the James Black Foundation and at Cresset Biomolecular Discovery for their scientific input and support and Drs. James Melville and Sally Rose for their help in preparing this manuscript.

Supporting Information Available: Conformer collections of each structure used in this report. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dockray, G. J. Clinical endocrinology and metabolism. Gastrin. *Best Pract. Res. Clin. Endocrinol. Metab.* **2004**, *18*, 555–568.

(2) Aly, A.; Shulkes, A.; Baldwin, G. S. Gastrins, cholecystokinins and gastrointestinal cancer. *Biochim. Biophys. Acta* **2004**, *1704*, 1–10.

(3) Haigh, C. R.; Attwood, S. E.; Thompson, D. G.; Jankowski, J. A.; Kirton, C. M.; Pritchard, D. M.; Varro, A.; Dimaline, R. Gastrin induces proliferation in Barrett's metaplasia through activation of the CCK2 receptor. *Gastroenterology* **2003**, *124*, 615–625.

(4) Chau, I.; Cunningham, D.; Russell, C.; Norman, A. R.; Kurzawinski, T.; Harper, P.; Harrison, P.; Middleton, G.; Daniels, F.; Hickish, T.; Prendeville, J.; Ross, P. J.; Theis, B.; Hull, R.; Walker, M.; Shankley, N.; Kalindjian, B.; Murray, G.; Gillbanks, A.; Black, J. Gastrazole (JB95008), a novel CCK2/gastrin receptor antagonist, in the treatment of advanced pancreatic cancer: Results from two randomized controlled trials. *Br. J. Cancer* **2006**, *94*, 1107–1115.

(5) Kalindjian, S. B.; Buck, I. M.; Davies, J. M.; Dunstone, D. J.; Hudson, M. L.; Low, C. M.; McDonald, I. M.; Pether, M. J.; Steel, K. I.; Tozer, M. J.; Vinter, J. G. Non-peptide cholecystokinin-B/gastrin receptor antagonists based on bicyclic, heteroaromatic skeletons. *J. Med. Chem.* **1996**, *39*, 1806–1815.

(6) Horwell, D. C.; Hughes, J.; Hunter, J. C.; Pritchard, M. C.; Richardson, R. S.; Roberts, E.; Woodruff, G. N. Rationally designed "dipeptoid" analogues of CCK. α-Methyltryptophan derivatives as highly selective and orally active gastrin and CCK-B antagonists with potent anxiolytic properties. *J. Med. Chem.* **1991**, *34*, 404–414.

(7) Martinez, J.; Rodriguez, M.; Bali, J.-P.; Laur, J. Phenylethylamide derivatives of the C-terminal tetrapeptide of gastrin. *Int. J. Pept. Protein Res.* **1986**, *28*, 529–535.

(8) Low, C. M.; Black, J. W.; Broughton, H. B.; Buck, I. M.; Davies, J. M.; Dunstone, D. J.; Hull, R. A.; Kalindjian, S. B.; McDonald, I. M.; Pether, M. J.; Shankley, N. P.; Steel, K. I. Development of peptide 3D structure mimetics: Rational design of novel peptoid cholecystokinin receptor antagonists. *J. Med. Chem.* **2000**, *43*, 3505–3517.

(9) Semple, G.; Ryder, H.; Rooker, D. P.; Batt, A. R.; Kendrick, D. A.; Szelke, M.; Ohta, M.; Satoh, M.; Nishida, A.; Akuzawa, S.; Miyata, K. (3*R*)-*N*-(1-(*tert*-Butylcarbonylmethyl)-2,3-dihydro-2-oxo-5-(2-pyridyl)-1*H*-1,4-benzodiazepin-3-yl)-*N*′-(3-(methylamino)phenyl)urea (YF476): A potent and orally active gastrin/CCK-B antagonist. *J. Med. Chem.* **1997**, *40*, 331–341.

(10) Bock, M. G.; DiPardo, R. M.; Evans, B. E.; Rittle, K. E.; Whitter, W. L.; Veber, D. E.; Anderson, P. S.; Freidinger, R. M. Benzodiazepine gastrin and brain cholecystokinin receptor ligands: L-365,260. *J. Med. Chem.* **1989**, *32*, 13–16.

(11) Buck, I. M.; Black, J. W.; Cooke, T.; Dunstone, D. J.; Gaffen, J. D.; Griffin, E. P.; Harper, E. A.; Hull, R. A.; Kalindjian, S. B.; Lilley, E. J.; Linney, I. D.; Low, C. M.; McDonald, I. M.; Pether, M. J.; Roberts, S. P.; Shankley, N. P.; Shaxted, M. E.; Steel, K. I.; Sykes, D. A.; Tozer, M. J.; Watt, G. F.; Walker, M. K.; Wright, L.; Wright, P. T. Optimization of the in vitro and in vivo properties of a novel series of 2,4,5-trisubstituted imidazoles as potent cholecystokinin-2 (CCK2) antagonists. *J. Med. Chem.* **2005**, *48*, 6803–6812.

(12) Low, C. M.; Buck, I. M.; Cooke, T.; Cushnir, J. R.; Kalindjian, S. B.; Kotecha, A.; Pether, M. J.; Shankley, N. P.; Vinter, J. G.; Wright, L. Scaffold hopping with molecular field points: identification of a cholecystokinin-2 (CCK2) receptor pharmacophore and its use in the design of a prototypical series of pyrrole- and imidazole-based CCK2 antagonists. *J. Med. Chem.* **2005**, *48*, 6790–6802.

(13) Foucaud, M.; Tikhonova, I. G.; Langer, I.; Escrieut, C.; Dufresne, M.; Seva, C.; Maigret, B.; Fourmy, D. Partial agonism, neutral antagonism, and inverse agonism at the human wild-type and constitutively active cholecystokinin-2 receptors. *Mol. Pharmacol.* **2006**, *69*, 680–690.

(14) Kopin, A. S.; McBride, E. W.; Chen, C.; Freidinger, R. M.; Chen, D.; Zhao, C. M.; Beinborn, M. Identification of a series of CCK-2 receptor nonpeptide agonists: Sensitivity to stereochemistry and a receptor point mutation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5525–5530.

(15) Giragossian, C.; Mierke, D. F. Intermolecular interactions between cholecystokinin-8 and the third extracellular loop of the cholecystokinin-2 receptor. *Biochemistry* **2002**, *41*, 4560–4566.

(16) Cramer, R.; Patterson, D.; Bunce, J. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(17) Böhm, M.; Stürzebecher, J.; Klebe, G. Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*, 458–477.

(18) *Catalyst 4.7*; Accelrys, Inc.: San Diego, CA, 2000.

(19) Chopra, M.; Mishra, A. K. Ligand-based molecular modeling study on a chemically diverse series of cholecystokinin-B/gastrin receptor antagonists: Generation of predictive model. *J. Chem. Inf. Model.* **2005**, *45*, 1934–1942.

(20) Guner, O.; Clement, O.; Kurogi, Y. Pharmacophore modeling and three dimensional database searching for drug design using Catalyst: Recent advances. *Curr. Med. Chem.* **2004**, *11*, 2991–3005.

(21) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular field extrema as descriptors of biological activity: definition and validation. *J. Chem. Inf. Model.* **2006**, *46*, 665–76.
(22) Chessari, G.; Hunter, C. A.; Low, C. M.; Packer, M. J.; Vinter, J. G.; Zonta, C. An evaluation of force-field treatments of aromatic interactions. *Chem.—Eur. J.* **2002**, *8*, 2860–2867.
(23) Vinter, J. G. Extended electron distributions applied to the molecular mechanics of some intermolecular interactions. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 653–668.
(24) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular field technology applied to virtual screening and finding the bioactive conformation. *Expert Opin. Drug Discovery* **2007**, *2*, 131–144.
(25) Cheeseright, T.; Mackey, M.; Vinter, A. Peptides to non-peptides: Leads from structureless virtual screening. *Drug Discovery Today: BIOSILICO* **2004**, *2*, 57–60.
(26) Welsh, N. J.; Shankley, N. P.; Black, J. W. Comparative analysis of the vagal stimulation of gastric acid secretion in rodent isolated stomach preparations. *Br. J. Pharmacol.* **1994**, *112*, 93–6.
(27) Blaker, M.; Ren, Y.; Seshadri, L.; McBride, E. W.; Beinborn, M.; Kopin, A. S. CCK-B/gastrin receptor transmembrane domain mutations selectively alter synthetic agonist efficacy without affecting the activity of endogenous peptides. *Mol. Pharmacol.* **2000**, *58*, 399–406.
(28) Kalindjian, S. B.; Dunstone, D. J.; Low, C. M.; Pether, M. J.; Roberts, S. P.; Tozer, M. J.; Watt, G. F.; Shankley, N. P. Nonpeptide cholecystokinin-2 receptor agonists. *J. Med. Chem.* **2001**, *44*, 1125–1133.
(29) Schmassmann, A.; Garner, A.; Flogerzi, B.; Hasan, M.; Sanner, M.; Varga, L.; Halter, F. Cholecystokinin type B receptor antagonist PD-136,450 is a partial secretory agonist in the stomach and a full agonist in the pancreas of the rat. *Gut* **1994**, *35*, 270–274.
(30) Schmassmann, A.; Varga, L.; Flogerzi, B.; Halter, F. The potent gastrin receptor antagonist CAM-1189 exerts a partial acid secretory agonist effect in concious chronic gastric fistula rats. Gastrin, Dana Point, CA, 1992; pD3.
(31) Hull, R. A. D.; Shankley, N. P.; Harper, E. A.; Gerskowitch, V. P.; Black, J. W. 2-Naphthalenesulphonyl L-aspartyl-(2-phenethyl)amide (2-Nap)—A selective cholecystokinin CCK(A)-receptor antagonist. *Br. J. Pharmacol.* **1993**, *108*, 734–740.
(32) Bock, M. G.; Dipardo, R. M.; Mellin, E. C.; Newton, R. C.; Veber, D. F.; Freedman, S. B.; Smith, A. J.; Patel, S.; Kemp, J. A.; Marshall, G. R.; Fletcher, A. E.; Chapman, K. L.; Anderson, P. S.; Freidinger, R. M. 2nd-Generation benzodiazepine Cck-B antagonists—Development of subnanomolar analogs with selectivity and water solubility. *J. Med. Chem.* **1994**, *37*, 722–724.
(33) McDonald, I. M.; Dunstone, D. J.; Kalindjian, S. B.; Linney, I. D.; Low, C. M.; Pether, M. J.; Steel, K. I.; Tozer, M. J.; Vinter, J. G. 2,7-Dioxo-2,3,4,5,6,7-hexahydro-1*H*-benzo[*h*][1,4]diazonine as a new template for the design of CCK(2) receptor antagonists. *J. Med. Chem.* **2000**, *43*, 3518–3529.
(34) Howbert, J. J.; Lobb, K. L.; Britton, T. C.; Mason, N. R.; Bruns, R. F. Diphenylpyrazolidinone and benzodiazepine cholecystokinin antagonists—A case of convergent evolution in medicinal chemistry. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 875–880.
(35) Kalindjian, S. B.; Harper, E. A.; Pether, M. J. The synthesis of a radioligand with high potency and selectivity for CCKB/gastrin receptors. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1171–1174.
(36) Kalindjian, S. B.; Buck, I. M.; Cushnir, J. R.; Dunstone, D. J.; Hudson, M. L.; Low, C. M.; McDonald, I. M.; Pether, M. J.; Steel, K. I.; Tozer, M. J. Improving the affinity and selectivity of a nonpeptide series of cholecystokinin-B/gastrin receptor antagonists based on the dibenzobicyclo[2.2.2]octane skeleton. *J. Med. Chem.* **1995**, *38*, 4294–4302.
(37) Kalindjian, S. B.; Bodkin, M. J.; Buck, I. M.; Dunstone, D. J.; Low, C. M. R.; McDonald, I. M.; Pether, M. J.; Steel, K. I. M. A new class of nonpeptidic cholecystokinin-B gastrin receptor antagonists based on dibenzobicyclo[2.2.2]octane. *J. Med. Chem.* **1994**, *37*, 3671–3673.
(38) Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallogr.* **2002**, 380–388.
(39) Higginbottom, M.; Horwell, D. C.; Roberts, E. Selective ligands for cholecystokinin receptor subtypes Cck-a and Cck-B within a single structural class. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 881–884.
(40) Evans, B. E.; Rittle, K. E.; Bock, M. G.; DiPardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S.; Chang, R. S.; Lotti, V. J.; Cerino, D. J.; Chen, T. B.; Kling, P. J.; Kunkel, K. A.; Springer, J. P.; Hirshfield, J. Methods for drug discovery: Development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* **1988**, *31*, 2235–2246.

JM070880T