Three-Dimensional Quantitative Structure–Activity Relationships of Somatostatin Analogues. 1. Comparative Molecular Field Analysis of Growth Hormone Release-Inhibiting Potencies

Simon J. Hocart,* Vik Reddy,† William A. Murphy, and David H. Coy

Peptide Research Laboratories, Department of Medicine, Tulane University School of Medicine, New Orleans, Louisiana 70112, and Tripos Associates Inc., St Louis, Missouri 63144

Received November 4, 1994®

Somatostatin is a hypothalamic hormone that inhibits the release of growth hormone (GH). It has also been shown to inhibit the release of a broad range of hormones including insulin, glucagon, and gastrin. Presently, five different receptor subtypes of somatostatin have been characterized and cloned. Our previous work on the structure–activity relationship of somatostatin and that of many others has generated a large database of analogues with different biological activities and receptor affinities. This present work is an investigation of the growth hormone release-inhibiting potencies of somatostatin analogues by the three-dimensional quantitative structure–activity paradigm, comparative molecular field analysis (CoMFA). A total of 64 analogues were modeled in SYBYL using structural information from two NMR studies. The molecules were aligned by a root-mean-square fit of atoms and field-fit of the steric and electrostatic molecular fields and the resulting databases analyzed by partial least squares analysis with cross-validation to extract the optimum number of components. The analysis was then repeated without cross-validation to give the final QSAR models. Preliminary investigations with the CoMFA models led to the synthesis of a new somatostatin analogue. This compound together with five other newly synthesized compounds not included in the original training sets were used to test the predictive ability of the CoMFA models. Two models with good predictive powers are presented.

Introduction

Somatostatin, Ala[1]-Gly[2]-c[Cys[3]-Lys[4]-Asn[5]-Phe[6]-Phe[7]-Trp[8]-Lys[9]-Thr[10]-Phe[11]-Thr[12]-Ser[13]-Cys[14]], a hypothalamic factor that inhibits the release of growth hormone (GH) was first isolated from ovine hypothalami in 1972.[1,2] It has since been shown to inhibit the release of a broad range of hormones including insulin, glucagon, and gastrin and has been located in many tissues including the gut, the pancreas, and the nervous system. The broad spectrum of biological activity of somatostatin and its very short half-life in the blood stream reduce the therapeutic value of the native peptide.

Structure–activity studies have demonstrated that the sequence Phe[7]-Trp[8]-Lys[9]-Thr[10] is essential for GH release-inhibiting activity.[3–5] These studies have resulted in the synthesis of a large number of analogues of somatostatin which exhibit a variety of bioactivity profiles. The various actions of these somatostatin analogues are mediated through different membrane-bound receptor subtypes. Recently, the existence of several receptor subtypes has been conclusively demonstrated by molecular cloning. Presently, five receptor subtypes (SSTR1-5) have been characterized.[6–9] Previous research in our laboratory and others has concentrated on the identification of subtype-selective analogues of somatostatin.[10,11] These structure–activity studies have resulted in a large database of noncongeneric somatostatin analogues in the literature.

The comparative molecular field analysis method of three-dimensional structure–activity relationships (3-D QSAR) is based on the assumption that the interactions between a receptor and its ligand, or an enzyme and its substrate or inhibitor, are primarily noncovalent in nature and shape-dependent.[12,13] Therefore, a QSAR may be derived by sampling the steric and electrostatic fields surrounding a set of ligands and correlating the differences in those fields to biological activity. Partial least squares (PLS) analysis is employed to extract a QSAR from the large comparative molecular field analysis (CoMFA) data table produced. In PLS, the "best equation" is obtained by selecting the proper number of components (latent variables) using a cross-validation procedure. In this way the dimensionality of the model is chosen according to its ability to predict the data rather than to fit the data. Several successful CoMFA studies of small, conformationally constrained systems have been reported including polyhalogenated dibenzo-*p*-dioxins, dibenzofurans, and biphenyls,[14] clodronic acid esters,[15] and progestin and androgen receptor binding.[16] Conformationally flexible systems which have been studied include inhibitors of protein-tyrosine kinase,[17] human rhinovirus 14,[17] thermolysin,[18,19] renin,[18] and human immunodeficiency virus protease (I).[20] These studies used the X-ray crystallographic structure of the ligands bound to the enzymes to derive the molecular conformation and alignment rules for the 3-D QSAR models. Angiotensin-converting enzyme (ACE) has been studied using an active analogue approach as the basis for the molecular conformation and alignments of the ACE inhibitors.[19] The results were comparable to those obtained with the closely related thermolysin inhibitors using alignment rules derived from crystallographic data. In this paper we will present a preliminary CoMFA investigation of the GH release-inhibiting

† Current address: National Starch and Chemical Co., Central Analytical Group, Bridgewater, NJ 08807.

® Abstract published in *Advance ACS Abstracts,* May 1, 1995.

0022-2623/95/1838-1974$09.00/0

Table 1. Structure of Somatostatin Analogues and GH Release-Inhibition Potencies

compd no	code name	modeled from	structure	GHI[a]	pGHI[b]
1	BIM-23014	3	DNal-c[Cys-Tyr-DTrp-Lys-Val-Cys]-Thr-NH_2	0.83	0.08
2	BIM-23030	3	c[Mpa-Tyr-DTrp-Lys-Val-Cys]-Phe-NH_2	23	−1.36
3	BIM-23034		DPhe-c[Cys-Tyr-DTrp-Lys-Val-Cys]-Nal-NH_2	0.43	0.37
4	BIM-23042	3	DNal-c[Cys-Tyr-DTrp-Lys-Val-Cys]-Nal-NH_2	11	−1.04
5	BIM-23049	3	DNal-Ala-Tyr-DTrp-Lys-Val-Ala-Thr-NH_2	435	−2.64
6	BIM-23050	3	*N*MeDAla-Tyr-DTrp-Lys-Val-Phe-NH_2	20000	−4.30
7	BIM-23051	5	DPhe-Ala-Phe-DTrp-Lys-Thr-Ala-Thr-NH_2	2500	−3.40
8	BIM-23052	3	DPhe-Phe-Phe-DTrp-Lys-Thr-Phe-Thr-NH_2	77	−1.89
9	BIM-23053	5	DPhe-Ala-Tyr-DTrp-Lys-Val-Ala-Nal-NH_2	15	−1.17
10	BIM-23055	8	DPhe-Phe-Tyr-DTrp-Lys-Val-Phe-DPhe-NH_2	11100	−4.05
11	BIM-23056	8	DPhe-Phe-Tyr-DTrp-Lys-Val-Phe-DNal-NH_2	12500	−4.10
12	BIM-23057	8	DPhe-Cpa-Tyr-DTrp-Lys-Val-Phe-Thr-NH_2	17	−1.23
13	BIM-23058	8	DPhe-Phe-Tyr-DTrp-Lys-Val-Phe-Thr-NH_2	111	−2.04
14	BIM-23059	3	DNal-c[Cys-Tyr-DTrp-Lys-Thr-Cys]-Thr-NH_2	0.19	0.72
15	BIM-23060	3	DPhe-c[Cys-Tyr-DTrp-Lys-Thr-Cys]-Nal-NH_2	0.05	1.30
16	BIM-23063	8	DPhe-Cpa-Tyr-DTrp-Lys-Thr-Phe-Nal-NH_2	100	−2.00
17	BIM-23064	8	DPhe-Cpa-Tyr-DTrp-Lys-Val-Phe-DAla-NH_2	4000	−3.60
18	BIM-23065	3	DNal-Cpa-Tyr-DTrp-Lys-Thr-Phe-Thr-NH_2	17	−1.23
19	BIM-23066	8	DPhe-Nif-Tyr-DTrp-Lys-Val-Phe-Thr-NH_2	159	−2.20
20	BIM-23067	5	DCpa-Ala-Tyr-DTrp-Lys-Val-Ala-DPhe-NH_2	417	−2.62
21	BIM-23068	8	DPhe-Cpa-Tyr-DTrp-Lys-Thr-Phe-Thr-NH_2	3.4	−0.53
22	BIM-23069	5	DCpa-Ala-Tyr-DTrp-Lys-Val-Ala-Nal-NH_2	45	−1.65
23	BIM-23070	5	DPhe-Ala-Tyr-DTrp-Lys-Thr-Ala-Nal-NH_2	16	−1.20
24	L-362-823	3	c[Aha-c[Cys-Phe-DTrp-Lys-Thr-Cys]]	0.81	0.09
25	L-362-855	24	c[Aha-Phe-Trp-DTrp-Lys-Thr-Phe]	16.7	−1.22
26	L-362-862	24	c[Aha-Phe-Cpa-DTrp-Lys-Thr-Phe]	1.4	−0.15
27	L-363-301	28	c[Pro-Phe-DTrp-Lys-Thr-Phe]	0.59	0.23
28	L-363-376	3	c[Pro-Ala-DTrp-Lys-Thr-Phe]	5	−0.70
29	MK-678	28	c[*N*MeAla-Tyr-DTrp-Lys-Val-Phe]	0.02	1.70
30	NC4-28B	3	DPhe-c[Cys-Tyr-DTrp-Lys-Ser-Cys]-Nal-NH_2	0.03	1.52
31	NC8-12	3	DPhe-c[Cys-Tyr-DTrp-Lys-Abu-Cys]-Nal-NH_2	0.03	1.52
32	SMS-201-995	3	DPhe-c[Cys-Phe-DTrp-Lys-Thr-Cys]-Thr-OL	0.93	0.03
33	DC23-60	3	DNal-c[Cys-Tyr-DTrp-Lys-Val-Cys]-Thr-OH	1.07	−0.03
34	EC5-21	3	DPhe-c[Cys-Phe-DTrp-Lys-Thr-Cys]-Nal-NH_2	0.22	−0.66
35		28	c[*N*MeAla-Phe-DTrp-Lys-Thr-Phe][c]	0.29	0.54
36		28	c[Pro-Tyr-DTrp-Lys-Thr-Phe][c]	0.07	1.16
37		28	c[Pro-Phe-DTrp-Lys-Val-Phe][c]	0.29	0.54
38		28	c[Pro-Phe-DTrp-Lys-Ser-Phe][c]	0.71	0.15
39		28	c[Pro-Phe-DTrp-Lys-Abu-Phe][c]	1.25	−0.10
40		24	c[Aha-Phe-Phe-DTrp-Lys-Thr-Phe][d]	1.08	−0.03
41		24	c[Aha-Phe-Tyr-DTrp-Lys-Thr-Phe][d]	1.04	−0.02
42		24	c[Aha-Phe-Fpa-DTrp-Lys-Thr-Phe][d]	20	−1.30
43		24	c[Aha-Phe-Nif-DTrp-Lys-Thr-Phe][d]	3.5	−0.54
44		24	c[Aha-Phe-Amf-DTrp-Lys-Thr-Phe][d]	35	−1.54
45		24	c[Aha-Phe-Thz-DTrp-Lys-Thr-Phe][d]	5	−0.70
46		24	c[Aha-Phe-Leu-DTrp-Lys-Thr-Phe][d]	1.05	−0.02
47		24	c[Aha-Phe-Ala-DTrp-Lys-Thr-Phe][d]	6.25	−0.80
48		24	c[Aha-Phe-Phe-1-MeTrp-Amf-Thr-Phe][d,e]	1.8	−0.26
49		24	c[Aha-Phe-Phe-5-MeTrp-Lys-Thr-Phe][d,e]	14	−1.15
50		24	c[Aha-Phe-Phe-5-FTrp-Lys-Thr-Phe][d,e]	6.7	−0.83
51		24	c[Aha-Phe-Phe-6-FTrp-Lys-Thr-Phe][d,e]	2.4	−0.38
52		24	c[Aha-Phe-Phe-5-BrTrp-Lys-Thr-Phe][d,e]	50	−1.70
53		24	c[Aha-Phe-Phe-5-OMeTrp-Lys-Thr-Phe][d,e]	50	−1.70
54		24	c[Aha-Phe-Phe-DTrp-Orn-Thr-Phe][d]	20	−1.30
55		24	c[Aha-Phe-Phe-DTrp-Arg-Thr-Phe][d]	14	−1.15
56		24	c[Aha-Phe-Phe-DTrp-Lys-MeAla-Phe][d]	50	−1.70
57		24	c[Aha-Phe-Phe-DTrp-Lys-Acp-Phe][d]	6.7	−0.83
58		28	c[Pro-Phe-DTrp-Lys-Thr-Ala][f]	16.7	−1.22
59		28	c[Pro-Phe-DTrp-Lys-Thr-Pro][f]	125	−2.10
60		28	c[DPro-Phe-DTrp-Lys-Thr-DAla][f]	166	−2.22
61		28	c[Pro-Phe-DTrp-Lys-Thr-DAla][f]	500	−2.70
62		28	c[Phe-Phe-DTrp-Lys-Thr-Phe][f]	3.71	−0.57
63		3	c[c[Cys-Phe-DTrp-Lys-Thr-Cys]][g]	0.4	0.40
64		63	c[c[Cys-Tyr-DTrp-Lys-Val-Cys]][g]	0.06	1.20

[a] GH release-inhibiting potency relative to somatostatin (=1.0). [b] Expressed as the negative logarithm of the GH release-inhibiting potency. [c] Reference 35. [d] Reference 36. [e] Tryptophan derivative in this analogue was not resolved so peptide was modeled in the *R* (D) configuration. [f] Reference 37. [g] Reference 38.

activity of somatostatin analogues based on structural data from solution phase NMR studies.[21,22]

Methods

Abbreviations: Abbreviations of the common amino acids are in accordance with the recommendations of IUPAC−IUB.[23] Additional abbreviations include: Abu, 2-aminobutanoic acid; Acp, 1-amino-1-cyclopentanecarboxylic acid; Aha, 7-aminoheptanoic acid; Ahx, 6-aminohexanoic acid; Amf, *p*-aminophenylalanine; Cpa, *p*-chlorophenylalanine; Dip, 3,3-diphenylalanine; Fpa, *p*-fluorophenylalanine; Har, homoarginine; Nal, 3-(2-naphthyl)alanine; Mpa, mercaptoproprionic acid; Nif, *p*-nitrophenylalanine; Thz, 3-(4-thiazolyl)alanine; Tic, 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid. Following the suggestion of Clementi,[24] the abbreviation q^2 is used in favor of

Table 2. Conformation of BIM-23034 from Each CoMFA Model

	model A				model B				model C			
	ϕ	ψ	χ_1	χ_2	ϕ	ψ	χ_1	χ_2	ϕ	ψ	χ_1	χ_2
DPhe[5]		−39	50	76		61	41	74		56	39	71
Cys[6]	−64	161	−34		28	50	−71		32	61	−61	
Tyr[7]	−61	174	64	81	−69	81	−52	97	−73	71	−44	120
DTrp[8]	76	−72	−179	113	68	−139	176	−87	80	−144	177	−94
Lys[9]	−139	41	−54	−173	−63	−33	−63	−179	−60	−31	−68	176
Val[10]	58	16	−170		−82	77	−59		−67	74	−65	
Cys[11]	−60	−179	93		−80	173	−58		−77	163	−67	
Nal[12]	−73	67	−52	105	−141	154	−67	−104	−146	158	−66	86

the SYBYL standard "cross-validated r^2" to distinguish it from the analogous conventional r^2.

Peptide Synthesis. Peptides were synthesized by a solid-phase synthesis methodology as described previously.[25] Briefly, the peptide amides were synthesized on a 4-methylbenzhydrylamine substituted, 1% cross-linked polystyrene resin (Advanced ChemTech, Louisville, KY) using butyloxycarbonyl N^{α} protection. Peptide acids were elongated on a conventional Merrifield resin (Advanced ChemTech). The crude peptides were cleaved and deprotected with anhydrous hydrogen fluoride at 0 °C. The cysteine containing analogues were cyclized in 90% acetic acid with a slight excess of I_2. DC-S-10-96 was synthesized using BocLys(Fmoc). The protected peptide was cleaved from the Merrifield resin by HF, cyclized with PyBOP, and deprotected with piperidine. All the analogues were purified to homogeneity by reversed-phase liquid chromatography and gave satisfactory molecular weights by matrix-assisted laser desorption mass spectrometry (LaserMat, Finnegan MAT, San Jose, CA) and amino acid analyses. The analytical data will be published elsewhere.[26]

GH Release Inhibiting Potency. Assays to determine the *in vitro* GH release inhibiting potency (IC_{50}) were performed as described previously.[27] Anterior pituitaries from adult male rats (200−250 g) were dispersed aseptically by a trypsin/DNase method. The dispersed cells were diluted with sterile Dulbecco's modified Eagles's medium (GIBCO, Grand Island, NY) supplemented with 2.5% fetal calf serum (GIBCO), 3% horse serum (GIBCO), 10% fresh rat serum from the pituitary donors, 1% minimum essential medium nonessential amino acids (GIBCO), 10 ng/ml gentamycin (Sigma Chemical Co., St Louis, MO), and 10 000 units/mL nystatin (GIBCO). The cells were counted with a haemocytometer and randomly plated at a density of 2×10^5 cells/well (Costar Cluster 24, Rochester Scientific, Rochester, NY). The plated cells were maintained in the Dulbecco's medium described above in a humidified atmosphere of 95% air and 5% CO_2 at 37 °C for 96 h. In preparation for hormone challenge, the cells were washed three times with medium 199 (GIBCO). Each dose of analogue (diluted in siliconized test tubes) was tested in quadruplicate wells, in a total volume of 1 mL of medium 199 containing 1% bovine serum albumin. Cells were pulsed in the presence of 1 nM GRF(1-29)NH_2, in the presence or absence of various concentrations of somatostatin analogues. After 3 h at 37 °C in a 95% air/5% CO_2 atmosphere, the medium was removed and stored at −20 °C until the time of the GH radioimmunoassay. GH in plasma and media was measured by a standard double-antibody radioimmunoassay using components generously supplied through NHPP, NIDDK, NICHHD, and USDA. Potencies relative to somatostatin (=1.0) were calculated by four-point assay[28] from the mean of a least three experiments or by comparison of IC_{50}'s from dose−response curves. Additional data and analogues were obtained from the literature as indicated by the references in Table 1.

Molecular Modeling. All molecular modeling and CoMFA studies were performed on a Silicon Graphics Personal Iris 4D/35TG+ computer. Molecular databases containing 64 somatostatin analogues (see Table 1) were modeled in SYBYL 6.03[29] using the standard TRIPOS force field.[30] The first 34 compounds were taken from an earlier publication from this group,[10] and the remainder were obtained from the literature as indicated in Table 1.

The first CoMFA model (model A) was based on the conformational inferences of Van Binst and Tourwé.[21] Compound **3** (BIM-23034; DPhe[5]-c[Cys[6]-Tyr[7]-DTrp[8]-Lys[9]-Val[10]-Cys[11]]-Nal[12]-NH_2) was built from the predefined amino acids and its conformation set to β sheet. The type II′ bend was introduced around DTrp[8]-Lys[9] and the cysteine bridge formed. Partial atomic charges were calculated for the molecule with a charged amino terminus using the Pullman method.[31,32] The geometries of the extra cyclic residues, DPhe[5] and Nal[12], were adjusted visually to facilitate hydrogen bond formation between DPhe[5]CO and Nal[12]NH.[21] The structure was then optimized by energy minimization using the Broyden, Fletcher, Goldfarb, and Shannon (BFGS) algorithm to a final root mean square (rms) gradient of 0.1 kcal mol $Å^{-1}$. A distance-dependent dielectric function[33] was employed together with the default settings for all the other minimization options. This resulted in a conformation for compound **3** in which the N and C termini were hydrogen bonded (see Table 2 and Figure 1a). The remaining compounds were modeled by mutating the residues of compound **3** where possible or by replacing the residues for the cases in which a perfect backbone fit was not possible. The actual analogue used to form each compound is indicated in Table 10. The unchanged residues of the new analogue were defined as an aggregate to constrain their conformation during the initial minimization of the modified residues. The constraints were then removed and the minimization repeated.

The second CoMFA model was based on the folded conformation of He and Huang *et al.*[22,34] Since the minimization of model A to a final gradient of 0.1 kcal mol $Å^{-1}$ had been time consuming, this CoMFA database was minimized in two stages. Initially the analogues were minimized to an intermediate gradient and then the database reminimized to a final gradient using a simple SYBYL Programming Language (SPL) script. The cyclic hexapeptide c[Pro[6]-Phe[7]-(2R,3S)-β-MeTrp[8]-Lys[9]-Thr[10]-Phe[11]] was modeled from the published ϕ, ψ, and χ_1 values,[22] cyclized and minimized with Pullman partial atomic charges to an rms gradient of 1 kcal mol $Å^{-1}$. This structure was used as the basis for the conformation of the template compound **3**. The cyclic hexapeptide was mutated to H-Cys[6]-Phe[7]-DTrp[8]-Lys[9]-Thr[10]-Cys[11]-OH. The cysteine bridge was formed and minimized with the remainder of the sequence (Phe[7]:Thr[10]) defined as an aggregate. The peptide was elongated to the structure of compound **3** and the pendant amino acids and cystine bridge minimized to an rms gradient of 1 kcal mol $Å^{-1}$ with the Phe[7]:Thr[10] aggregate present to remove any bad steric contacts between the extracyclic amino acids while maintaining the conformation of the ring. Finally, the aggregate constraint was removed and the analogue reminimized to a final rms gradient of <10 kcal mol $Å^{-1}$. All the other molecules were modeled, as in model A, by mutating these starting compounds. This database was then minimized to a final rms gradient of 0.1 kcal mol $Å^{-1}$ without constraints using a simple SPL script. Many changes occurred during the final minimization including the distortion of Tyr[7] by the formation of an additional hydrogen bond. This distortion is evident in compound **3** shown in Figure 1c. Since preliminary CoMFA analyses of these two databases gave similar q^2 values, it was decided to produce two CoMFA models from this one conformation, models B and C, differing only in the final minimization gradient, to examine the effect of the minimization. The conformation of compound **3** from model B, minimized to an intermediate gradient of <10 kcal mol $Å^{-1}$ is shown in Figure 1b and Table 2. The conformation of compound **3**

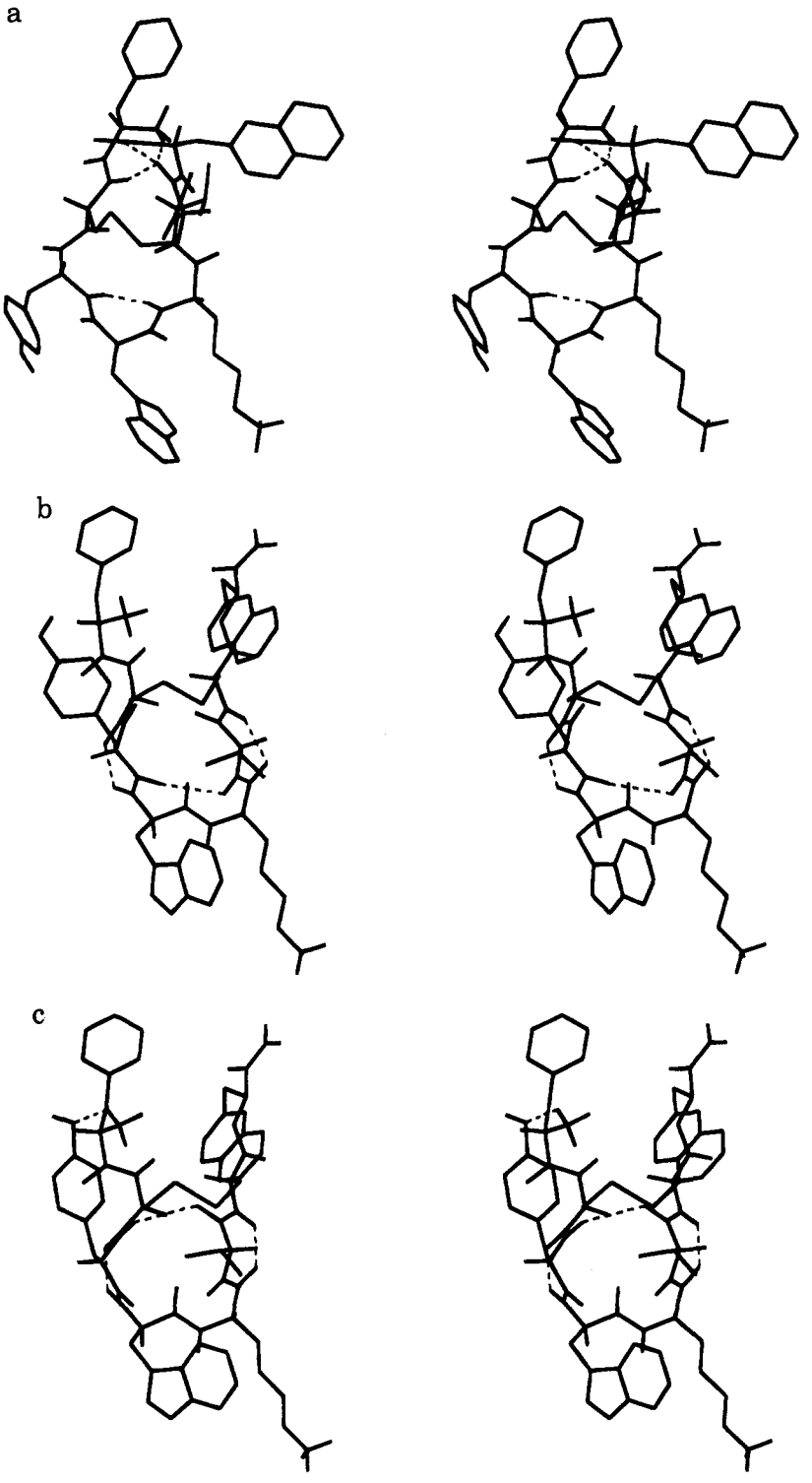

Figure 1. Stereoviews of the conformation of compound **3** (BIM-23034) from (a) model A, (b) model B, and (c) model C. The hydrogen bonds are shown by dotted lines.

from the fully minimized database, model C, is shown in Figure 1c and Table 2.

CoMFA Analysis. The SYBYL CoMFA module was used to define a CoMFA region automatically. This resulted in a regularly spaced (2 Å) three-dimensional lattice (grid), which extended past the van der Waals volumes of all the molecules. The actual dimensions of the region were dependent on the molecular conformation and the applied alignment rule (see Table 4). The steric (Lennard-Jones) and electrostatic (Coulombic) field energies were calculated at all intersections of the grid using an sp^3-hybridized carbon probe atom with a charge of +1 and a distance-dependent dielectric constant. The steric and electrostatic contributions were truncated to ±30 kcal mol^{-1} and the electrostatic contributions were ignored at lattice intersections with maximal steric interactions. QSAR tables were generated from the training sets with the compounds as rows and the target biological data as a column.

CoMFA Alignment and Analysis. A series of alignment rules were developed to minimize the predictive residual sum of squares (PRESS) and maximize q^2 by cross-validated partial least squares (PLS) analysis (see Table 3). The superimpositions of molecules in the alignments are shown in Figures 2, 3 and 4 for models A, B and C, respectively. The figures illustrate 11 representative compounds: five linear octapeptides (compounds **5**, **8**, 11, **17**, and **21**), four cyclic octapeptides (compounds 1, **15**, **30**, and **34**), a cyclic heptapeptide (compound

Table 3. Alignment Rules and Associated Atomic Expressions Used in the CoMFA Models

alignment	SYBYL type	reference compound	SPL atomic expression
1	match	**3**	({*7}+{*8}+{*9}+{*10})&({BACKBONE}-<H>)
2	CoMFA field-fit	**30**	–
3	fit[a]	**29**	{*7.CG}+{*8.CG}+{*9.CG}+{*11.CG}

[a] SYBYL requires an atom selection rather than an expression to perform a fit, so this expression was converted to a selection for each compound *via* an SPL program. The analogues which did not have a Cγ carbon atom were fitted using the Cβ atom of the reference and compound for that residue.

Table 4. Dimensions of the Regions Used in the CoMFA Models

	model A: alignment			model B: alignment			model C: alignment		
	1	2	3	1	2	3	1	2	3
X, Å	30	28	30	24	24	24	24	24	24
Y, Å	24	24	26	24	22	22	24	22	22
Z, Å	24	26	26	24	26	28	28	28	28
points	2704	2730	3136	2366	2184	2340	2535	2340	2340

26), and a cyclic hexapeptide (compound **29**). Cross-validated PLS analyses gave the optimum number of components for each model/alignment combination which was used to generate the final models without cross-validation. q^2 was calculated from the equations

$$q^2 = 1 - \frac{\text{PRESS}}{\sum(y - y_{\text{mean}})^2}$$

$$\text{where} \quad \text{PRESS} = \sum(y - y_{\text{pred}})^2$$

Values of q^2 can range between 1, indicating a perfect model, to less than 0, where the errors of prediction are greater than the error from assigning each compound the mean activity of the model.

Cross-validated PLS analysis runs employed 10 components and the leave-one-out procedure in which one compound was dropped in turn and a model generated from the remaining compounds. The model was used to predict the activity of the dropped compound. This procedure was repeated until all the compounds had been predicted. The q^2 and number of principal components were tabulated. The optimum number of principal components was taken as the number required to increase q^2 by ~5% from the model with one fewer component rather than the default SYBYL estimate which is that which gave the highest q^2 value. All cross-validated PLS analyses were performed with a minimum σ (column filter) value of 2.0 kcal mol^{-1} which minimized the influence of column noise and reduced computation time. The final CoMFA analyses were produced by repeating the PLS analyses with the optimum numbers of components but without cross-validation, yielding conventional r^2's. The fitted predicted activities of each training compound from the 34 and 64 compound non-cross-validated analyses are given in Tables 8 and 9, respectively. The q^2, r^2, and associated values are shown in Tables 5, 6, and 7 for models A, B, and C, respectively.

CoMFA Model Validation. The predictive ability of each analysis was determined from a set of six new compounds which were not included in the training set (see Table 10). These compounds were aligned and their activities predicted by each PLS analysis. The "predictive r^2" (r^2_{pred}, calculated in the same way as q^2), PRESS, and $r^2_{\text{correlation}}$ from a plot of the actual *versus* predicted activities of the test compounds were calculated for each analysis from these values (see Tables 11, 12, and 13 for models A, B, and C respectively).

Results and Discussion

Molecular Models. The first 34 somatostatin analogues used in this study were reported in a previous paper from this group and collaborators.[10] These analogues included both linear (16 compounds) and cyclic octapeptides (10 compounds) and cyclic hepta- (4 compounds) and hexapeptides (3 compounds) and a linear hexapeptide (compounds **1**–**34**, see Table 1). These analogues were supplemented with 30 additional analogues from the literature (12 cyclic hexapeptides and 18 cyclic heptapeptides) giving a database of 64 compounds with a wide range of growth hormone release-inhibiting activities (compounds **35**–**64**, see Table 1).[35–38] Since no X-ray crystallographic data on the conformation of a somatostatin analogue bound to a receptor were available, NMR data from somatostatin analogues were employed for the molecular modeling. The intention was to generate self-consistent databases of reasonable molecular structures which had common features in similar orientations. These databases are available for downloading *via* FTP from the world wide web.[48]

To model the compounds in Table 1, the octapeptide, compound **3** (BIM-23034), was used as a template from which the other analogues were produced by mutation or replacement of amino acids. This octapeptide was chosen as it has a cyclic hexapeptide core with two pendant, bulky amino acids and could be easily mutated to produce the various cyclic and linear analogues with similar conformations. Many (44 of 64) of the peptides in this study contained unusual amino acids not present in the standard protein dictionaries of SYBYL, so the use of *ab initio* (STO-3G)-based Kollman charges was precluded. Consequently, Pullman charges were employed as they are computationally simple and yet give good estimates for the dipole moments of proteins.[31,32]

Model A. CoMFA model (model A) was based on NMR structural inferences from our cyclic octapeptides.[21] This work indicated the presence of a type II′ bend and likely intramolecular hydrogen bonds but included no information on the population of side chain rotamers. Thus the template molecule, compound **3** (BIM-23034) was built from the dictionary definitions of the component amino acids with no regard to the side chain conformation. Extensive energy minimization to a final rms gradient of 0.1 kcal mol Å^{-1} produced the conformation shown in Figure 1a. The corresponding φ, ψ, and χ_1 angles of this conformation are given in Table 2. This structure was then used as a template to model the remaining compounds of the database as indicated in Table 1. When the resulting molecules were superimposed by aligning the backbone heavy atoms (alignment 1), the analogues were markedly similar in appearance (see Figure 2a), with the Tyr^7 hydroxyl groups in the same orientation and a similar hydrogen bonding pattern for the set: DPhe^{5}NH, Cys^{11}-CO; Cys^{6}NH, Cys^{11}CO; Nal^{11}CONH$_2$, Cys^{11}CO; and Lys^{8}NH, Tyr^{7}CO. These hydrogen bonds maintained the conformation of the two pendant amino acids in the octapeptide analogues (see Figure 1a).

Model B. The second model was based on the more complete NMR data from a constrained cyclic hexapeptide, c[Pro^6-Phe^7-(2R,3S)-β-MeTrp^8-Lys^9-Thr^{10}-Phe^{11}].[22,34]

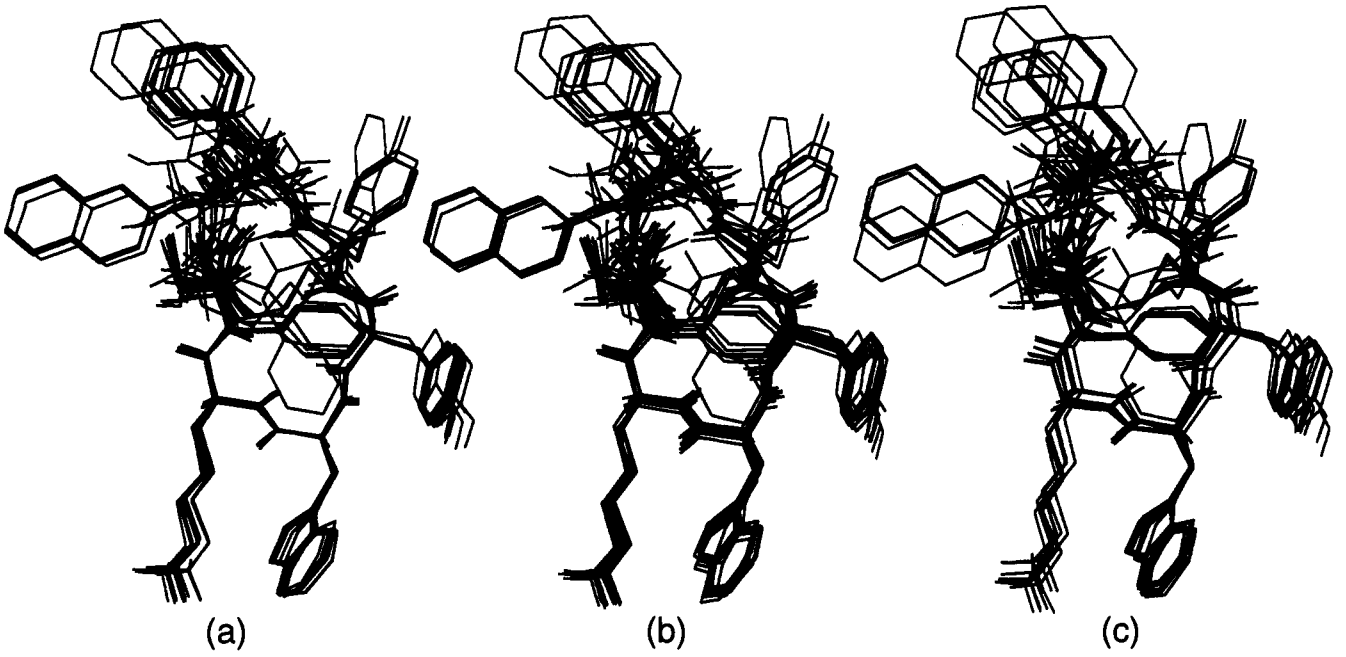

Figure 2. Superimposition of 11 representative analogues from model A in (a) alignment 1, (b) alignment 2, and (c) alignment 3.

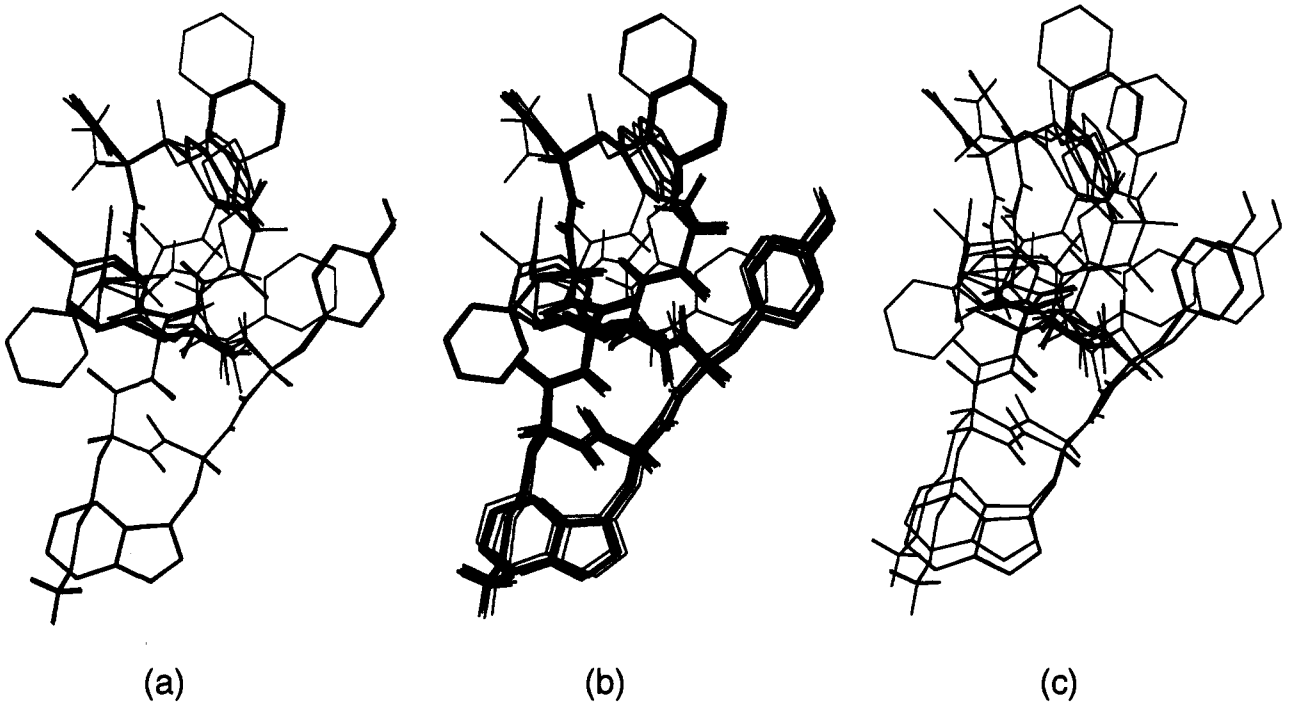

Figure 3. Superimposition of 11 representative analogues from model B in (a) alignment 1, (b) alignment 2, and (c) alignment 3.

Since the minimization of model A to a final gradient of 0.1 kcal mol $Å^{-1}$ had been time consuming, this model was produced in two stages. Initially the analogues were minimized to an intermediate gradient (<10 kcal mol $Å^{-1}$) and then the database reminimized to a final rms gradient (0.1 kcal mol $Å^{-1}$) using a simple SYBYL Programming Language (SPL) script. Preliminary CoMFA analyses of these two databases gave similar q^2 values so it was decided to produce two CoMFA models from this one conformation, models B and C differing only in the final minimization gradient, to examine the effect of the minimization.

The hexapeptide c[Pro^6-Phe^7-(2R,3S)-β-MeTrp^8-Lys^9-Thr^{10}-Phe^{11}] was built from the published ϕ, ψ, and χ_1 values. Although the conformation of this peptide was fully described, it lacked the amino acids in positions 5 and 12 of compound **3** (BIM-23034). Two cysteine residues were introduced in place of the bridging Pro^6 and Phe^{11} residues and the conformation of the remainder of the ring constrained as an aggregate. Minimization of this intermediate structure followed by addition of the pendant amino acids, DPhe^5 and Nal^{12}, and reminimization produced the structure shown in Figure 1b. This structure contained hydrogen bonds only in the cyclic portion of the molecule: DTrp^{8}NH, Cys^{6}CO; Val^{10}NH, Lys^{9}CO; Lys^{9}NH, Tyr^{7}CO. The analogues were modeled from the hexapeptide and derived octapeptide by mutation and minimization to a final

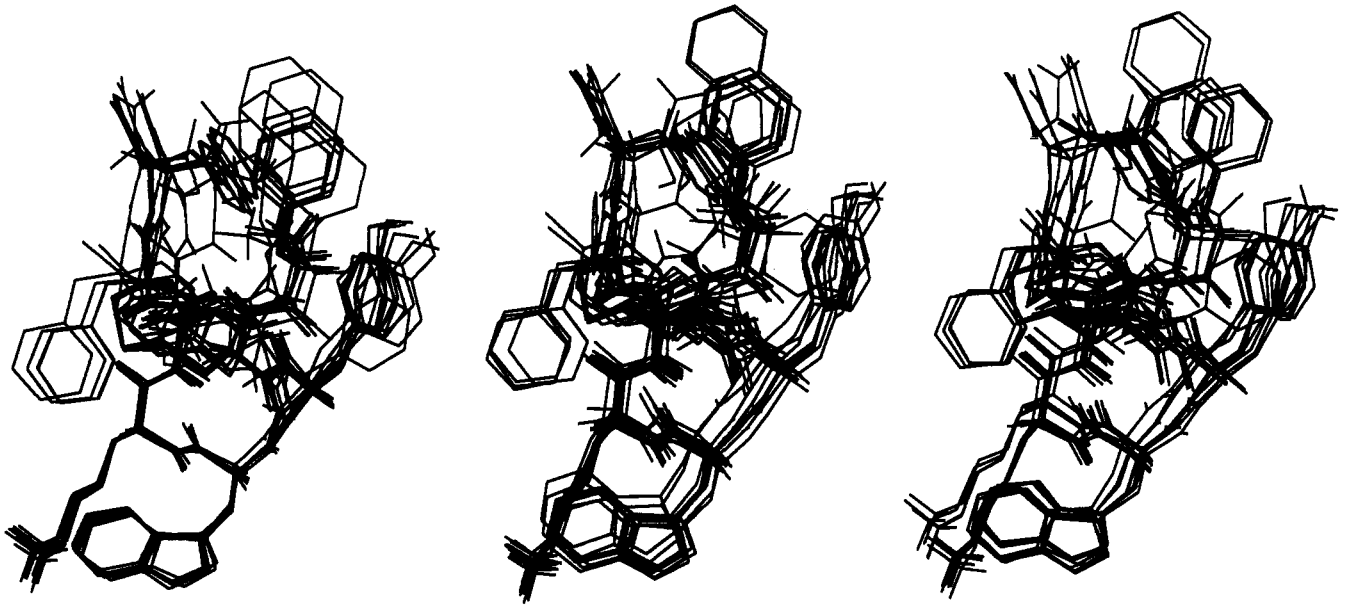

Figure 4. Superimposition of 11 representative analogues from model C in (a) alignment 1, (b) alignment 2, and (c) alignment 3.

Table 5. Effect of Alignment Rule on the PLS Analysis of Model A

	alignment 1			alignment 2			alignment 3		
	64[b]	64	34	64	64	34	64	64	34
q^2	0.507	0.560	0.536	0.574	0.597	0.631	0.546	0.586	0.546
PC[a]	5	8[c]	4	5	7[c]	5	5	7[c]	5
SEP[b]	1.019	0.988	1.216	0.946	0.937	1.104	0.978	0.950	1.203
r^2	0.861	0.965	0.923	0.922	0.972	0.956	0.880	0.948	0.921
s	0.542	0.279	0.496	0.404	0.246	0.380	0.503	0.338	0.500
F test	71.594	189.156	86.526	137.874	280.334	75.501	84.775	144.400	85.066
p value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
elect contrib	0.392	0.393	0.383	0.408	0.400	0.405	0.390	0.381	0.349
steric contrib	0.608	0.607	0.617	0.592	0.600	0.595	0.610	0.619	0.651

[a] Optimum number of components in PLS analysis; number of components which increase q^2 by ~5%. [b] Standard error of prediction. [c] Number of components corresponding to the minimum value of SEP.

Table 6. Effect of Alignment Rule on the PLS Analysis of Model B

	alignment 1		alignment 2		alignment 3	
	64[b]	34	64	34	64	34
q^2	0.566	0.489	0.522	0.410	0.613	0.464
PC[a]	8	8	5	5	6	5
SEP[b]	0.981	1.374	1.102	1.395	0.910	1.330
r^2	0.922	0.961	0.836	0.897	0.884	0.895
s	0.415	0.381	0.587	0.584	0.497	0.588
F test	81.629	76.428	59.291	48.647	72.740	47.843
p value	0.000	0.000	0.000	0.000	0.000	0.000
elect contrib	0.336	0.356	0.389	0.368	0.343	0.315
steric contrib	0.664	0.644	0.611	0.632	0.657	0.685

[a] Optimum number of components in PLS analysis; number of components which increase q^2 by ~5%. [b] Standard error of prediction.

gradient of ~5–10 Kcal mol Å^{-1} to give model B as indicated in Table 1. This resulted in a database of molecules which superimposed well by visual inspection, with common features (e.g., Tyr hydroxyl) aligned similarly (see Figure 3a).

Model C. The database of molecules from model B was duplicated and reminimized without constraints to a final gradient of 0.1 kcal mol Å^{-1} to give model C. The conformation of compound **3** (BIM-23034) from model C is shown in Figure 1c.

Table 7. Effect of Alignment Rule on the PLS Analysis of Model C

	alignment 1		alignment 2		alignment 3		
	64[b]	34	64	34	64	34	34
q^2	0.617	0.506	0.517	0.481	0.700	0.581	0.659
PC[a]	5	3	3	3	4	4	8[c]
SEP[b]	0.897	1.233	0.991	1.264	0.788	1.155	1.123
r^2	0.892	0.828	0.789	0.846	0.900	0.921	0.992
s	0.476	0.727	0.655	0.688	0.454	0.501	0.177
F test	96.261	48.298	74.914	55.123	133.375	84.750	364.950
p value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
electr contrib	0.355	0.339	0.387	0.368	0.323	0.328	0.340
steric contrib	0.645	0.661	0.613	0.632	0.677	0.672	0.660

[a] Optimum number of components in PLS analysis; number of components which increase the q^2 by ~5%. [b] Standard error of prediction. [c] Number of components corresponding to the minimum value of SEP.

Reminimization caused several changes in conformation; the transannular hydrogen bond disappeared and was replaced by one between Tyr^{7}NH and Val^{10}CO, and an additional hydrogen bond formed between DPhe^5-NH and Tyr^{7}OH. This resulted in a marked displacement of the Tyr^7 side chain and a distortion of its previously planar aromatic ring. As can be seen from Figure 1c, the conformational space available to the

Table 8. Fitted Predictions of GH Release-Inhibition Potencies from the 34 Compound CoMFA PLS Analyses

		model A			model B			model C			
compd no.	pGHI[a]	1[b] 4[c]	2 5	3 5	1 8	2 5	3 5	1 3	2 3	3 4	3 8
1	0.08	0.18	−0.28	0.35	−0.35	−0.32	−0.43	−0.39	−0.44	−0.08	−0.08
2	−1.36	−2.18	−1.84	−1.87	−1.44	−1.70	−1.57	0.53	0.24	−1.24	−1.24
3	0.37	0.56	0.59	0.36	0.64	0.69	0.70	0.47	0.64	0.18	0.18
4	−1.04	−0.26	−0.67	−0.47	0.14	0.37	0.23	0.41	0.48	−0.71	−0.71
5	−2.64	−2.82	−3.04	−2.81	−2.58	−2.38	−2.17	−2.59	−2.31	−2.58	−2.58
6	−4.30	−4.15	−4.21	−4.37	−4.36	−3.74	−3.49	−4.63	−4.53	−4.16	−4.16
7	−3.40	−2.66	−3.16	−2.52	−2.60	−1.99	−1.87	−2.76	−2.31	−3.55	−3.55
8	−1.89	−1.33	−1.86	−1.34	−2.31	−2.11	−1.52	−2.27	−2.19	−1.81	−1.81
9	−1.17	−1.57	−1.40	−1.63	−1.59	−1.71	−1.63	−1.94	−1.71	−1.20	−1.20
10	−4.05	−4.06	−4.10	−4.04	−4.11	−4.22	−4.92	−3.82	−4.21	−3.98	−3.98
11	−4.10	−3.84	−3.68	−3.96	−4.32	−3.58	−4.30	−3.52	−3.73	−4.18	−4.18
12	−1.23	−2.17	−2.30	−2.22	−1.14	−1.45	−1.66	−1.71	−1.73	−1.49	−1.49
13	−2.04	−2.15	−2.08	−2.22	−1.98	−1.93	−1.92	−2.13	−1.98	−2.13	−2.13
14	0.72	0.84	0.92	1.11	0.14	0.17	−0.10	−0.01	−0.06	0.85	0.85
15	1.30	1.02	1.08	0.84	1.21	1.04	1.04	0.92	0.74	1.18	1.18
16	−2.00	−2.11	−1.68	−1.99	−1.87	−2.10	−1.78	−2.44	−2.32	−1.95	−1.95
17	−3.60	−2.89	−3.26	−2.91	−3.58	−3.07	−3.24	−2.53	−3.07	−3.70	−3.70
18	−1.23	−1.26	−1.41	−0.96	−1.03	−1.58	−1.13	−0.58	−0.57	−1.21	−1.21
19	−2.20	−1.99	−1.64	−1.76	−2.00	−1.83	−1.89	−1.39	−1.54	−1.93	−1.93
20	−2.62	−3.26	−2.91	−3.29	−2.36	−3.50	−3.26	−2.88	−3.30	−2.65	−2.65
21	−0.53	−1.26	−0.81	−1.05	−0.56	−1.31	−1.05	−0.53	−0.52	−0.60	−0.60
22	−1.65	−1.53	−1.36	−1.71	−1.84	−1.63	−1.51	−1.90	−1.64	−1.39	−1.39
23	−1.20	−1.05	−1.13	−0.99	−1.32	−1.43	−1.29	−1.64	−1.56	−1.56	−1.56
24	0.09	0.51	0.42	0.24	0.05	0.13	0.07	−0.05	0.02	−0.01	−0.01
25	−1.22	−1.05	−1.25	−1.14	−1.14	−1.40	−1.45	−0.98	−0.85	−1.11	−1.11
26	−0.15	−0.15	0.04	−0.67	0.12	0.16	−0.42	−0.87	−0.79	−0.08	−0.08
27	0.23	0.66	0.55	0.59	0.35	1.01	0.80	0.36	0.18	0.23	0.23
28	−0.70	−0.40	−0.70	−0.06	−0.92	−1.44	−0.71	0.21	−0.48	−0.80	−0.80
29	1.70	0.87	1.02	0.82	1.69	1.29	1.11	0.15	0.28	1.70	1.70
30	1.52	1.03	1.12	1.07	1.47	1.45	1.42	1.05	1.17	1.47	1.47
31	1.52	1.00	1.31	1.00	1.21	1.04	1.04	0.90	0.73	1.60	1.60
32	0.03	0.70	0.52	0.52	−0.12	−0.18	−0.11	−0.38	−0.35	0.29	0.29
33	−0.03	0.00	0.16	0.26	−0.11	0.36	0.42	0.02	0.56	−0.18	−0.18
34	−0.66	0.63	0.71	0.68	0.24	0.55	0.57	0.78	0.82	0.63	0.63

[a] Expressed as the negative logarithm of the GH release-inhibiting potency. [b] Alignment rule. [c] Number of principal components.

pendant amino acids was severely constrained by the Tyr[7] side chain. This database had the poorest visual superimposition of molecules of the three databases (see Figure 4). The close proximity of the Tyr[7] side chain and the N terminus permitted electrostatic interactions between them causing the position of the Tyr[7] ring and hydroxyl orientation to vary greatly between molecules. This distortion is also illustrated in Figure 1c.

Alignments. To perform a CoMFA analysis, the molecules of interest have to be aligned to maximize the interaction of the ligands with the active site. In the case of previous studies of peptide inhibitors, X-ray crystallographic data of inhibitors cocrystallized with the enzyme were used to derive experimental alignment rules.[17-20] However, as shown in the study of inhibitors of HIV(I) protease, different classes of inhibitors may interact in different ways with the active site of the enzyme, giving different relative positions of each class of inhibitor with respect to one another in the active site.[20] This required specific alignment rules for each class of inhibitor. In this study, solid state X-ray data was not available, so arbitrary rules had to be derived. Most of the analogues contained the common sequence DTrp[8]-Lys[9] so the simple superimposition of common features was though to be expedient. Three different alignment rules were defined in an attempt to maximize q^2 (see Table 3). The alignments are illustrated in Figures 2, 3, and 4 with 11 representative molecules for models A, B, and C, respectively.

Alignment 1. The initial alignment was based on the SYBYL rms match command using the backbone heavy atoms from the common residue sequence 7:10. The active analogue, compound **3** (BIM-23034), from each model database was used as the reference molecule for this match as it had been used as a template for modeling the other structures in the database. This alignment was designed to maximize the overlap of the backbone atoms in the interior of the analogues in the biologically important and least variable DTrp[8]-Lys[9] portion of the analogues. The SYBYL expression used to select the atoms involved in the match for this alignment is given in Table 3.

Alignment 2. The second alignment was based on the SYBYL QSAR rigid-body field fit command. Compound **30** was used as the reference compound since it was one of the most potent octapeptides. Its size and bulky pendant groups ensured more extensive fields than the other molecules so facilitating the rigid-body field fit. To diminish the problem of multiple local minima,[40] the molecules were first prealigned by backbone atoms (alignment 1) before being fitted to the fields of the reference compound in the same orientation. Field fit uses a Simplex algorithm in SYBYL which minimizes the rms difference of the steric and electrostatic fields averaged over all the lattice intersections to find the best fit. This fit is dependent on the similarity and initial orientations of the molecules, and thus, it is possible to generate slightly different orientations with the same compounds. This alignment maximized the similarity in the molecular fields of the molecules with those of the most active octapeptide analogue.

Table 9. Fitted Predictions of GH Release-Inhibition Potencies from the 64 Compound CoMFA PLS Analyses

		model A						model B			model C		
compd		1[b]	1	2	2	3	3	1	2	3	1	2	3
no.	pGHI[a]	5[c]	8	5	7	5	7	8	5	6	5	3	4
1	0.08	0.60	−0.14	−0.04	−0.13	0.37	0.15	−0.25	−0.63	−0.44	−0.01	−0.40	−0.66
2	−1.36	−0.32	−1.12	−0.96	−1.49	−1.00	−1.66	−2.00	−0.71	−1.10	−0.92	0.25	−0.80
3	0.37	0.50	0.78	0.78	0.82	0.60	0.63	0.75	0.89	0.74	0.46	0.78	0.62
4	−1.04	0.08	−0.75	−0.28	−0.92	−0.08	−0.59	0.26	0.39	0.15	0.12	0.73	0.18
5	−2.64	−2.60	−2.93	−2.75	−2.91	−2.51	−3.11	−2.45	−2.26	−2.06	−1.88	−2.09	−2.94
6	−4.30	−4.61	−4.21	−4.50	−4.12	−4.39	−4.36	−4.37	−3.94	−3.82	−4.52	−5.26	−4.20
7	−3.40	−2.65	−3.34	−2.87	−3.35	−2.64	−3.03	−2.22	−1.78	−1.79	−2.64	−2.26	−3.19
8	−1.89	−1.23	−2.14	−1.42	−1.87	−1.26	−1.79	−2.04	−2.17	−1.50	−2.06	−1.97	−1.88
9	−1.17	−1.93	−1.15	−1.49	−1.20	−2.06	−1.24	−1.68	−1.10	−1.54	−1.58	−1.86	−1.72
10	−4.05	−3.90	−4.01	−3.96	−4.07	−3.86	−4.02	−4.06	−4.44	−4.98	−4.31	−3.92	−3.85
11	−4.10	−3.72	−4.15	−3.71	−3.84	−3.96	−3.94	−4.08	−3.50	−4.35	−3.89	−3.42	−3.49
12	−1.23	−2.01	−1.67	−1.46	−1.95	−2.20	−2.22	−1.39	−1.51	−1.59	−1.88	−1.75	−1.82
13	−2.04	−1.99	−1.81	−2.32	−1.98	−2.10	−2.24	−1.93	−1.87	−1.89	−2.22	−1.94	−2.30
14	0.72	1.25	0.57	0.51	0.64	1.07	0.86	0.10	−0.27	−0.11	0.85	0.00	0.48
15	1.30	0.84	1.17	1.17	1.28	1.03	1.02	1.28	1.06	1.07	0.95	0.85	1.40
16	−2.00	−2.19	−2.22	−1.56	−2.11	−2.07	−2.05	−2.31	−1.73	−1.77	−2.32	−2.06	−2.25
17	−3.60	−2.64	−3.46	−3.15	−3.49	−2.85	−3.20	−3.44	−2.65	−3.15	−2.79	−3.09	−2.96
18	−1.23	−1.12	−1.02	−1.42	−1.51	−0.91	−1.01	−1.37	−1.69	1.14	−1.23	−0.73	−1.10
19	−2.20	−1.78	−1.57	−2.21	−1.49	−1.58	−1.63	−1.97	−1.91	−1.84	−1.79	−1.74	−1.71
20	−2.62	−3.63	−2.52	−3.69	−2.60	−3.70	−2.73	−2.40	−3.80	−3.27	−3.03	−3.49	−3.02
21	−0.53	−1.22	−0.84	−0.80	−0.63	−0.95	−0.94	−0.78	−1.45	−1.03	−1.04	−0.74	−0.81
22	−1.65	−1.82	−1.64	−1.59	−1.33	−2.11	−1.45	−1.67	−1.17	−1.36	−1.54	−1.75	−1.67
23	−1.20	−1.50	−1.25	−1.39	−1.37	−1.54	−0.82	−1.43	−1.13	−1.22	−1.46	−1.77	−1.83
24	0.09	0.12	0.28	0.01	0.14	−0.08	0.23	0.32	0.00	0.22	−0.32	−0.29	−0.33
25	−1.22	−1.28	−1.07	−1.26	−1.00	−1.16	−1.19	−1.06	−1.48	−0.93	−0.99	−0.64	−0.91
26	−0.15	−0.69	−0.51	−0.65	−0.10	−0.70	−0.48	−0.42	−0.24	−0.40	−0.75	−0.89	−0.75
27	0.23	0.22	0.31	0.40	0.42	0.30	0.43	0.24	0.84	0.29	0.17	−0.09	0.31
28	−0.70	0.10	−0.55	−0.61	−0.63	0.40	−0.52	−0.27	−0.70	−0.38	0.12	−0.34	−0.12
29	1.70	0.64	1.46	1.01	1.90	0.57	1.23	1.46	1.50	1.03	1.16	0.43	1.31
30	1.52	0.85	1.58	0.84	1.48	0.97	1.22	1.14	1.18	1.15	1.06	0.93	1.28
31	1.52	0.76	1.26	1.17	1.32	1.02	1.08	1.12	0.93	0.81	1.03	0.73	1.58
32	0.03	0.51	0.34	0.44	0.02	0.37	0.32	0.24	−0.43	−0.17	−0.03	−0.38	−0.14
33	−0.03	0.49	0.13	−0.04	−0.16	0.65	0.04	−0.11	−1.10	0.22	−0.29	0.70	−0.29
34	−0.66	0.49	0.63	0.70	0.66	0.70	0.47	0.48	0.66	0.46	1.06	0.94	1.12
35	0.54	0.29	0.62	0.61	0.66	0.38	0.69	0.69	0.58	0.55	0.33	−0.30	0.31
36	1.16	0.61	1.19	0.78	1.02	0.70	1.10	1.15	0.59	0.63	0.72	0.60	0.93
37	0.54	0.23	0.41	0.65	0.58	0.13	0.34	0.23	0.48	−0.05	1.06	−0.19	1.03
38	0.15	0.04	0.08	0.23	0.09	0.23	0.31	0.24	0.37	0.27	0.31	0.47	0.47
39	−0.10	0.11	0.23	0.29	0.02	0.08	0.23	0.27	0.29	−0.06	0.17	−0.08	0.60
40	−0.03	−0.79	−0.61	−0.61	−0.34	−0.94	−0.79	−0.79	−0.39	−0.73	−0.77	−0.80	−0.88
41	−0.02	−0.52	0.05	−0.19	0.20	−0.51	−0.07	−0.11	−0.39	−0.36	−0.60	−0.59	−0.76
42	−1.30	−0.75	−0.79	−0.58	−1.10	−0.90	−0.85	−0.67	−0.83	−0.71	−0.71	−0.95	−0.87
43	−0.54	−0.64	−0.42	−0.62	−0.55	−0.64	−0.31	−0.36	−0.58	−0.45	−0.71	−0.77	−0.80
44	−1.54	−0.95	−1.18	−1.05	−1.52	−0.92	−0.88	−0.41	−0.61	−0.71	−1.05	−0.90	−0.91
45	−0.70	−0.75	−0.58	−0.85	−0.75	−0.71	−0.63	−0.63	−1.31	−0.86	−1.06	−0.80	−0.81
46	−0.02	−0.61	−0.27	0.05	−0.04	−0.40	−0.44	−0.24	−0.72	−0.47	−0.28	−0.38	−0.55
47	−0.80	−0.87	−0.73	−0.96	−0.83	−0.77	−1.01	−1.14	−0.81	−1.26	−0.77	−0.68	−0.81
48	−0.26	−0.59	−0.48	−0.92	−0.40	−0.45	−0.14	−0.52	−0.54	−0.19	−0.88	−0.14	−0.73
49	−1.15	−0.91	−1.51	−1.01	−1.30	−0.94	−1.41	−1.53	−1.56	−1.48	−0.92	−0.84	−1.16
50	−0.83	−0.84	−0.88	−0.76	−0.73	−1.00	−0.90	−0.71	−0.36	−0.65	−0.88	−1.01	−0.93
51	−0.38	−0.85	−0.62	−0.59	−0.65	−0.93	−0.74	−0.74	−0.50	−0.72	−0.66	−0.72	−0.89
52	−1.70	−0.97	−1.67	−1.17	−1.96	−1.02	−1.51	−1.40	−1.05	−1.28	−0.89	−0.97	−1.10
53	−1.70	−1.00	−1.45	−1.11	−1.33	−1.16	−1.30	−1.63	−1.68	−1.55	−1.41	−1.63	−1.26
54	−1.30	−0.89	−1.24	−0.89	−1.11	−0.95	−1.29	−1.31	−0.80	−1.28	−0.84	−0.78	−1.13
55	−1.15	−1.08	−1.24	−0.90	−1.32	−0.99	−0.94	−1.18	−0.64	−0.99	−0.84	−0.74	−0.90
56	−1.70	−1.78	−1.25	−2.16	−1.37	−1.65	−1.29	−1.45	−1.08	−1.14	−0.88	−1.29	−1.08
57	−0.83	−1.70	−0.94	−1.48	−0.79	−1.54	−0.94	−1.10	−1.07	−1.11	−1.11	−1.25	−1.17
58	−1.22	−1.61	−1.84	−1.65	−1.89	−1.45	−1.54	−1.75	−1.54	−1.49	−1.69	−0.91	−1.40
59	−2.10	−2.11	−2.20	−2.49	−2.32	−2.08	−2.61	−2.06	−2.82	−2.10	−2.74	−1.41	−2.38
60	−2.22	−1.65	−2.45	−1.92	−2.13	−1.87	−2.13	−2.47	−2.09	−2.34	−2.44	−2.41	−2.01
61	−2.70	−1.83	−2.17	−2.19	−2.39	−2.11	−2.73	2.35	−2.37	−2.96	−2.08	−1.27	−2.08
62	−0.57	−0.11	−0.66	−0.23	−0.53	−0.04	−0.29	−0.47	−0.22	−0.31	−0.57	−0.17	−0.51
63	0.40	0.94	0.49	0.59	0.42	0.74	0.44	0.60	0.40	1.02	0.65	0.11	0.49
64	1.20	1.13	1.06	1.24	0.90	1.15	1.17	1.13	0.48	1.45	1.01	0.31	0.73

[a] Expressed as the negative logarithm of the GH release-inhibiting potency. [b] Alignment rule. [c] Number of principal components.

Alignment 3. The third alignment was based on a postulated pharmacophore for somatostatin of Huang et al.[34] Huang described a folded active conformation for somatostatin and defined a pharmacophore based on the distances between four γ-carbon atoms of a cyclic hexapeptide; the aromatic residues at positions 7, 8, and 11 and Lys[9]. This conformation was also used as the basis for models B and C. Alignment 3 was based on the SYBYL fit command and defined as a four-atom rms fit of the γ-carbon atoms of residues 7, 8, 9, and 11 using compound **29** (MK-678) as the reference molecule. This potent analogue closely matched the structure of the hexapeptides used in the original definition of the pharmacophore. The atomic expression for this fit (see,

Table 10. Structure of Somatostatin Test Analogues and GH Release-Inhibition Potency

compd no.	code	structure	GHI[a]	pGHI[b]
65	DC35-58	DNal-c[Cys-Tyr-DTrp-Lys-Val-Cys]-Nal-OH	1.12	-0.05
66	NC9-28	DPhe-c[Cys-Tyr-DTrp-Lys-Ser-Cys]-Nal-OH	0.06	1.22
67	DC33-35	DDip-c[Cys-Tyr-DTrp-Lys-Val-Cys]-Nal-NH_2	0.91	0.04
68	NC8-59	DNal-c[Cys-Tyr-DTrp-Orn-Val-Cys]-Nal-NH_2	4.17	-0.62
69	DC35-53	DPhe-c[Cys-Trp-DTrp-Lys-Thr-Cys]-Nal-NH_2	0.20	0.69
70	DC-S-10-96	c[Tic-Tyr-DTrp-Lys-Abu-Phe]	0.26	0.58

[a] GH release-inhibiting potency relative to somatostatin (=1.0). [b] Expressed as the negative logarithm of the GH release-inhibiting potency.

Table 11. Predicted Activity of Somatostatin Test Analogues from Model A

		alignment 1			alignment 2			alignment 3		
code	pGHI[a]	64[b]	64	34	64	64	34	64	64	34
DC35-58	−0.05	0.18	0.03	−0.67	1.75	1.23	1.28	−0.09	−0.60	−0.71
NC9-28	1.22	0.87	1.29	0.57	2.03	1.75	1.84	1.16	0.87	0.96
DC33-35	0.04	0.45	0.94	0.63	0.67	0.75	0.63	0.66	0.82	0.52
NC8-59	−0.62	−0.11	−0.76	−0.51	−0.27	−0.74	−0.49	−0.11	−0.72	−0.43
DC35-53	0.69	0.64	1.13	0.67	0.64	0.85	0.85	0.64	0.63	0.62
DC-S-10-96	0.58	0.36	0.68	0.86	0.80	0.95	1.04	0.36	0.71	0.50
r^2_{pred}		0.929	0.885	0.880	0.521	0.705	0.700	0.924	0.886	0.925
$r^2_{correlation}$		0.880	0.774	0.513	0.414	0.666	0.687	0.724	0.619	0.679
t test		0.562	0.173	0.809	0.059	0.056	0.028	0.413	0.899	0.692
intercept		0.245	0.225	0.031	0.680	0.473	0.550	0.238	0.008	−0.015
slope		0.496	1.055	0.734	0.829	1.048	0.995	0.640	0.893	0.833
PRESS		0.654	1.044	1.246	4.476	2.600	2.755	0.701	1.064	0.781
P.C.		5	8	4	5	7	5	5	7	4

[a] Expressed as the negative logarithm of the GH release-inhibiting potency. [b] Number of training compounds in analysis.

Table 12. Predicted Activity of Somatostatin Test Analogues from Model B

		alignment 1		alignment 2		alignment 3	
code	pGHI[a]	64[b]	34	64	34	64	34
DC35-58	−0.05	−0.24	−0.24	−0.18	0.77	0.68	1.06
NC9-28	1.22	0.39	0.76	0.60	1.46	1.56	2.14
DC33-35	0.04	0.82	0.73	0.62	0.34	0.02	−0.03
NC8-59	−0.62	0.08	0.34	0.37	0.54	0.06	0.12
DC35-53	0.69	0.54	−0.05	0.73	0.48	0.21	0.06
DC-S-10-96	0.58	1.02	1.40	0.56	1.32	0.60	0.95
r^2_{pred}		0.778	0.560	0.813	0.702	0.853	0.665
$r^2_{correlation}$		0.171	0.102	0.242	0.433	0.520	0.406
t test		0.644	0.573	0.571	0.052	0.314	0.191
intercept		0.343	0.398	0.373	0.672	0.323	0.461
slope		0.296	0.295	0.250	0.472	0.642	0.824
PRESS		2.040	1.720	2.757	2.757	1.342	3.165
P.C.		8	8	5	5	6	5

[a] Expressed as the negative logarithm of the GH release-inhibiting potency. [b] Number of training compounds in analysis.

Table 3) was converted to an atomic selection for each fit to the reference compound using a simple SPL program. Since each CoMFA model had a different conformation, the molecules in each model database were aligned with the conformation of compound **29** in that database. Molecules which were missing the requisite γ-carbon atoms of the definition were fitted using the corresponding β-carbon atoms for those residues. This alignment was designed to maximize the overlap of atoms near the exterior of the molecules. However, the fact that some analogues lacked requisite γ-carbon atoms caused a poorer visual superimposition of the molecules than expected. This is particularly evident in models Band C (see Figures 2c, 3c, and 4c). Only molecules in models B and C fitted the original definition of the pharmacophore. In the case of model A, most (4 out of 6) of the distances were too short by >2 Å.

CoMFA PLS Analyses. The CoMFA PLS analyses were implemented by generating tables based on each model database containing the prealigned molecules. Each compound corresponded to one row in the table (one conformation). Columns were defined for the growth hormone release-inhibiting activity of the analogues and the data read in from a file. These data were then transformed to pGHI (log(1/GHI)) to reduce the spread of the values. CoMFA columns corresponding to each alignment rule were added using the default options in SYBYL. The CoMFA regions were calculated automatically and the resulting region dimensions and the number of lattice points are given in Table 4.

The first PLS analysis for each alignment rule was cross-validated using the leave-one-out procedure, which gives reproducible estimates of q^2 at the expense of computation time (>50 min for 64 cross-validation groups). The results are summarized in Tables 5, 6, and 7 for models A, B, and C, respectively. The optimum number of components was extracted from a cross-validated PLS analysis by examining the incremental change in q^2 with each additional component. To maximize the predictive power of the analysis rather than its ability to fit the data, the optimum number of components was judged to be that which increased the total q^2 by ~5% from the PLS analysis with one fewer component. This number usually coincided with the minimum of the standard error of prediction (SEP). For those analyses where this was not the case, the number of components corresponding to the minimum SEP was also used for a non-cross-validated PLS analysis. PLS analyses without cross-validation were then run with the optimum number of components for each alignment to derive the final QSAR models and corresponding conventional r^2's. The fitted predictions for the 34 and 64 compound analyses are given in Tables 8 and 9, respectively. Representative data from alignment 1 is also illustrated graphically for the 64 compound analyses in Figure 5. The figures show the non-cross-validated fitted predictions of activity of the training set (open circles) together with the predicted activities of the six test compounds (filled triangles) with the

Table 13. Predicted Activity of Somatostatin Test Analogues from Model C

		alignment 1		alignment 2		alignment 3		
code	pGHI[a]	64[b]	34	64	34	64	34	34
DC35-58	−0.05	0.29	0.14	0.91	0.81	−1.08	−0.68	−1.13
NC9-28	1.22	0.17	0.25	0.29	0.56	−0.79	−0.29	−0.43
DC33-35	0.04	0.27	0.32	0.60	0.33	−0.99	−0.60	−0.82
NC8-59	−0.62	0.05	0.34	0.53	0.31	−1.24	−0.99	−1.58
DC35-53	0.69	−0.65	0.03	0.51	0.84	−0.26	0.17	−0.16
DC-S-10-96	0.58	1.16	0.12	0.74	0.33	−0.12	0.08	0.36
r^2_{pred}		0.592	0.750	0.614	0.775	0.250	0.683	0.489
$r^2_{correlation}$		0.000	0.207	0.223	0.097	0.416	0.605	0.591
t test		0.800	0.722	0.405	0.424	0.004	0.009	0.004
intercept		0.215	0.227	0.644	0.493	−0.887	−0.554	−0.882
slope		−0.001	−0.086	−0.154	0.118	0.454	0.544	0.825
PRESS		3.852	2.624	3.481	2.209	7.938	3.744	6.321
P.C.		5	3	3	3	4	4	8

[a] Expressed as the negative logarithm of the GH release-inhibiting potency. [b] Number of training compounds in analysis.

associated regression lines and 95% confidence limits for the training set. As can be seen from Figure 5 and Tables 8 and 9, all three models gave reasonable estimates of the activities of the training compounds with statistically significant q^2 values.

CoMFA Model Validation. To test the predictive power of the CoMFA models, the activities of six newly synthesized analogues not included in the training sets were predicted (Table 10). Preliminary investigations with model A of the most active training analogues had suggested the synthesis of a carboxylate-terminated octapeptide analogue. This compound (NC9-28) was synthesized and tested. Five other newly synthesized analogues, including two which contained unusual amino acids not present in the original training sets, were synthesized, characterized, and tested for biological activity. These six compounds were modeled and aligned for each of the CoMFA models. The activities of the six test analogues were then predicted by each model and alignment combination. These predictions are given in Tables 11−13. The tabular data includes the square of the correlation coefficient ($r^2_{correlation}$) from a plot of the actual *versus* predicted activities of the test compounds and the associated slopes and intercepts which ideally should be 1, 1, and 0, respectively. The Student's t test of the means of the actual and predicted data, r^2_{pred} and PRESS are also given. The predicted activities of the compounds with alignment 1 for each model are also shown graphically in Figure 5.

All three models were capable of producing significant estimates of q^2 with each alignment rule. As can be seen from Tables 5−7, each model was able to give a q^2 greater than 0.4 with many analyses giving values in the range 0.5−0.6. All PLS analyses had similar contributions from the electrostatic and steric fields. Model A gave estimates of q^2 in the range 0.507 to 0.631 (mean = 0.565) for each alignment/set combination which in turn produced high conventional r^2's for the corresponding non-cross-validated PLS analyses. Each alignment of the 64 molecule training set in model A gave lower estimates for the optimum number of components by the 5% rule than by the minimum SEP rule (see Table 5). Thus, increasing the number of components from the optimum number determined by the 5% rule to that corresponding to the minima in the SEP also increased the q^2 and r^2 and decreased the standard errors of the analyses. However, examination of the predicted activities of the six test compounds from these analyses showed that the r^2_{pred} and the correlation coefficient dropped and the PRESS increased as more components were added. This illustrates the hypothesis that a larger number of components gives a better fit to the training data at the expense of the predictive ability of the analysis; the additional components are fitting noise rather than signal in the analysis and so degrade the predictive power. Surprisingly, the field fit alignment was an exception. Here, the increase in q^2 by the addition of the extra components was half that seen with the atomic alignments 1 and 3. Although the PRESS decreased slightly with the addition of extra components, it was still much larger than the PRESS from the other alignments (see Table 11). The visual superimposition of molecules from alignment 2 showed more scatter than alignment 1 but less than alignment 3 (see Figure 2). However, examination of the alignments in model B (Figure 3) show alignment 2 to have the most random scatter, a product of the local minima problems associated with field fit. This random scatter probably accounts for the higher PRESS and lower r^2_{pred} seen with alignment 2. The 34 compound analyses all gave single estimates of the optimum number of components. Again, the atomic alignments were superior in predictive ability to the field fit alignment even though they had lower q^2's.

Unlike model A, model B gave higher estimates of q^2 with the 64 molecule set than with the smaller set (q^2 range 0.410−0.613; mean 0.511, see Table 6). Visual examination of the superimposition of the molecules in each alignment showed much less scatter than observed with model A. Unlike model A, the highest estimates of q^2 and r^2 were associated with the atomic alignments 1 and 3 rather than with the field fit alignment. Examination of the predicted activity of the test compounds (see Table 12) showed alignment 3 with the 64 compound training set to give the lowest PRESS and highest r^2_{pred}, though the values were inferior to those from model A. Interestingly, only in model B did the highest q^2 correspond with the highest r^2_{pred} and the lowest PRESS.

Model C produced the highest estimate of q^2 of the three models with a range of 0.418−0.700 and mean 0.580 (see Table 7). In this case, the highest q^2 was from the 64 compound set with alignment 3 and the highest r^2 from the same alignment with 34 compounds. However, the predictive power of this model was much inferior to the others. As can be seen from Table 13, the correlation between the actual and predicted activities of the test compounds ($r^2_{correlation}$) was highest for

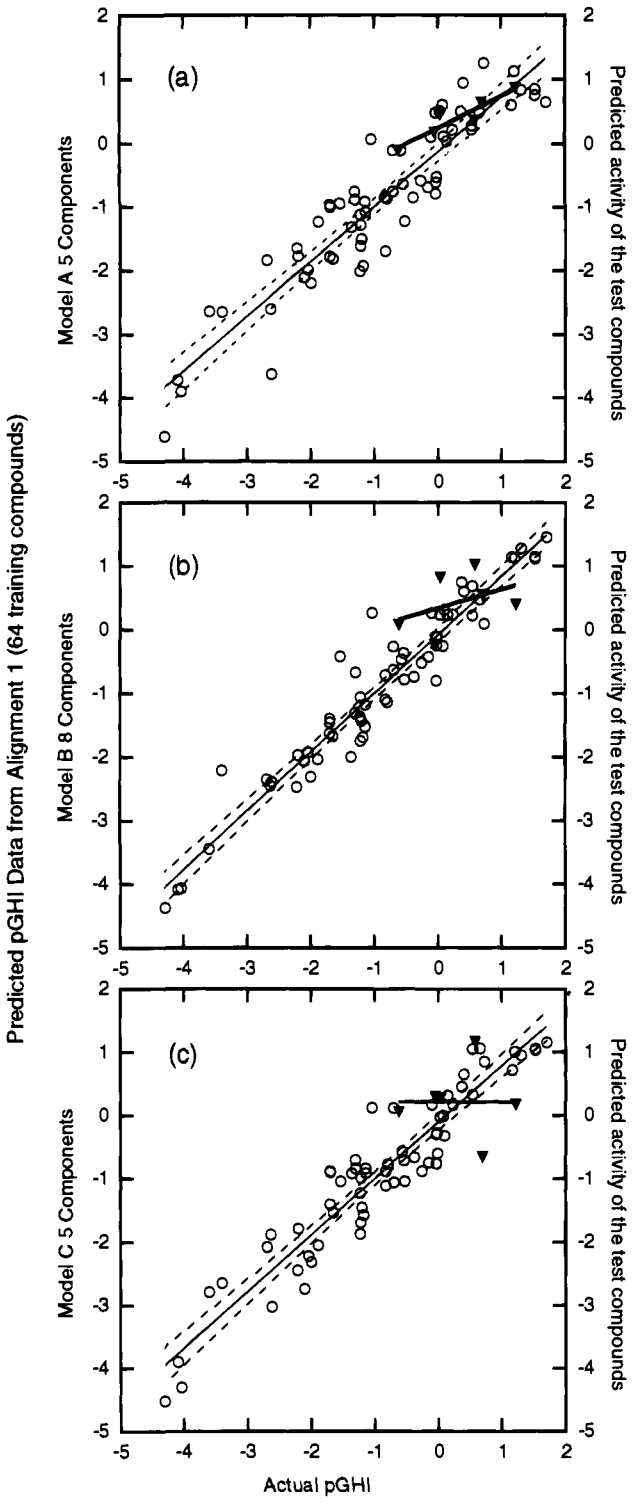

Figure 5. Fitted predictions of activity (pGHI) of the 64 training compounds (open circles) and the six test compounds (filled triangles) with regression lines from (a) model A, (b) model B, and (c) model C, with alignment 1. The 95% confidence limits are shown by dotted lines.

alignment 3, but the predictions were consistently underestimated leading to very large PRESS values. Indeed, a pairwise Student's *t*-test applied to the means of the actual and predicted data showed that they are different at the 95% confidence level. A visual inspection of the alignments of molecules in model C revealed the poorest superimposition of common features of the three models including large variations in the orientation of Tyr^7OH and distortions of the aromatic ring (see Figure 4). Hence, the PLS analysis was swamped by the noise of such artifacts leading to a poor CoMFA model.

CoMFA Field Plots

The CoMFA steric and electrostatic fields from PLS analyses are usually visualized as contour plots of the product of the standard deviation associated with the CoMFA column and the coefficient (sd × coeff) at each lattice point. The values corresponding to 80% and 20% contribution are plotted as colored, closed polyhedra which are associated with increased and decreased biological activity respectively. Representative stereoview mesh contour plots of the contributions for models A, B, and C with alignment 1, 64 training compounds and 5, 8, and 5 components respectively are shown in Figures 6–11. The numbers in the figures denote polyhedra referenced in the discussion. The steric contributions are shown in Figures 6–8 with an active analogue, compound **34** (red) and an inactive analogue, compound **10** (blue). These two molecules were chosen because of their low residuals in the three models. The steric fields are colored green where an increase in biological activity correlated with increased steric bulk or yellow where a decrease in biological activity correlated with increased steric bulk. Figures 9, 10, and 11 show stereoview mesh contour plots of the electrostatic contributions for models A, B, and C, respectively. These figures show an active analogue, compound **34** (green), and an inactive analogue, compound **10** (cyan). The electrostatic fields are colored blue where an increase in activity was correlated with increased positive electrostatic charge or red where an increase in activity was correlated with increased negative charge.

As expected, most of the CoMFA contour plot regions were associated with the variable terminal and bridging portions of the molecules, with only minor regions associated with the biologically critical but conserved $DTrp^8$-Lys^9 sequence. The most notable features of the steric contour plots for all the models were large, sterically unfavorable regions associated with the amino acids in the bridging region; Phe^6,Phe^{11} in the inactive (blue) analogue and Cys^6,Cys^{11} in the active (red) analogue. Model A had a large disfavored region (Figure 6; yellow polyhedron, area 1) surrounding the Phe^6 side chain of the inactive compound extending to the backbone of Phe^{12} and a smaller favored region (green, area 2) associated with the other bridging residue at position 11. However, in models B and C, this pattern was reversed with the disfavored region (Figures 7 and 8; yellow, 1) surrounding Phe^{11} and the favored region (green, 2) surrounding Phe^6. Model B alone displayed a large disfavored region encompassing the entire C terminal amide with an extension near to Nal^{12} in the inactive compound (Figure 7; yellow, area 3), whereas, in models A and C, this group was linked with sterically favored regions (Figures 6 and 8; green, 3). Models A and C had both favored and disfavored regions associated with Phe^5 (Figures 6 and 8; area 4). In model B the disfavored region was minor and the favorable region (Figure 7; green, 4) was displaced to a location equidistant from positions 5 and 6 and the amide terminus. Models A and B had sterically favorable polyhedra associated with position 7 (Figures 6 and 7; green, area 5) but the same position was unfavorable in model C (Figure 8, yellow, 5). Only model A had any

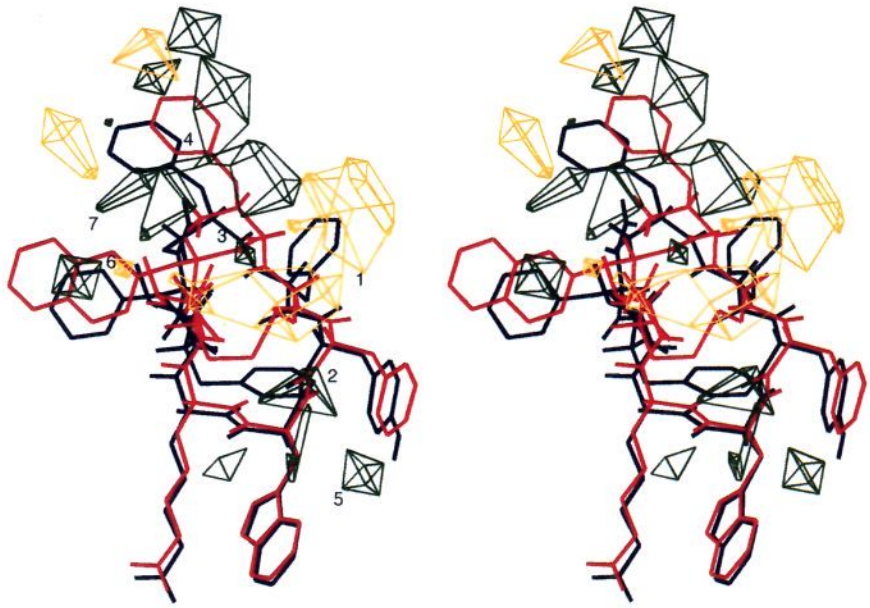

Figure 6. Stereoview of the CoMFA steric sd × coeff contour plot from the PLS analysis based on model A and alignment 1 with 5 components and 64 training compounds. Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. The highly active analogue, compound **34**, is shown in red, and an inactive analogue, compound **10**, is shown in blue. The numbers denote CoMFA polyhedra referenced in the discussion.

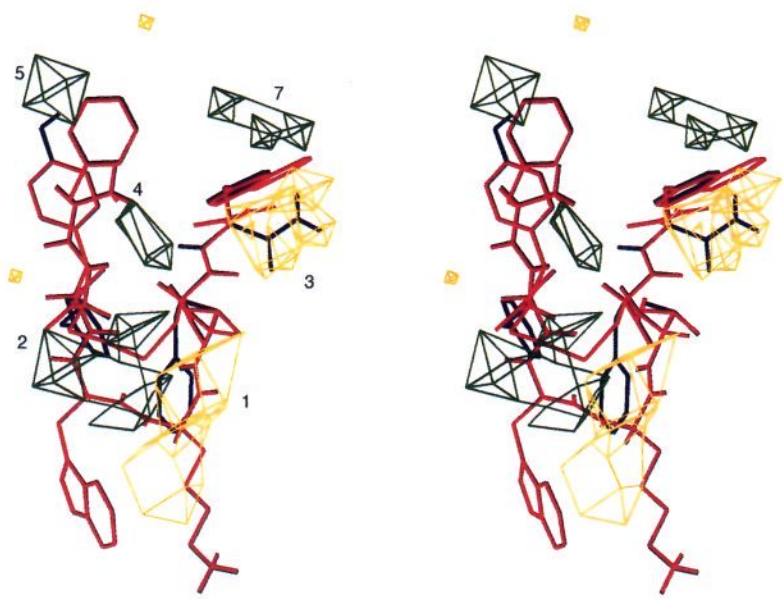

Figure 7. Stereoview of the CoMFA steric sd × coeff contour plot from the PLS analysis based on model B and alignment 1 with 8 components and 64 training compounds. Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. The highly active analogue, compound **34**, is shown in red and an inactive analogue, compound **10**, is shown in blue. The numbers denote CoMFA polyhedra referenced in the discussion.

contours near position 10, which had a large sterically favored region and a smaller disfavored region associated with it (Figure 6; area 6). Models A and B had favored regions distantly associated with position 12 in the active (red) compound (Figures 6 & 7; green, area 7). This residue was also associated with a disfavored region in model B which extended through the amide (Figure 7; yellow, area 3).

For the electrostatic contour plots of models A and B, the major features were large regions favoring increased positive charge (Figures 9 and 10; blue polyhedra, area 1) in vicinity of the termini (Phe[5] and Nal[12] for the green compound depicted in the figures), with a smaller region favoring increased negative charge (red, area 2) beneath it, closer to the end groups. In the case of model C, the region favoring increased positive charge (Figure 11; blue, area 1) was displaced toward position 7 which, in Tyr-containing analogues, was involved in a hydrogen bond with the amino terminus. This region was sandwiched by two regions favoring increased negative charge, one near position 7 (red, 2) and the other near the amide group (red, 2a). Additionally, models B and C had regions favoring increased positive charge near the terminal amide

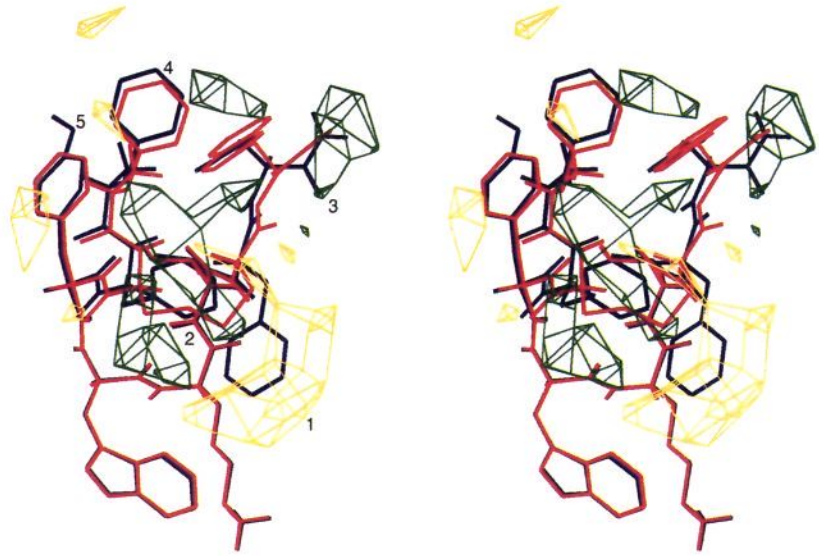

Figure 8. Stereoview of the CoMFA steric sd × coeff contour plot from the PLS analysis based on model C and alignment 1 with 5 components and 64 training compounds. Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. The highly active analogue, compound **34**, is shown in red and an inactive analogue, compound **10**, is shown in blue. The numbers denote CoMFA polyhedra referenced in the discussion.

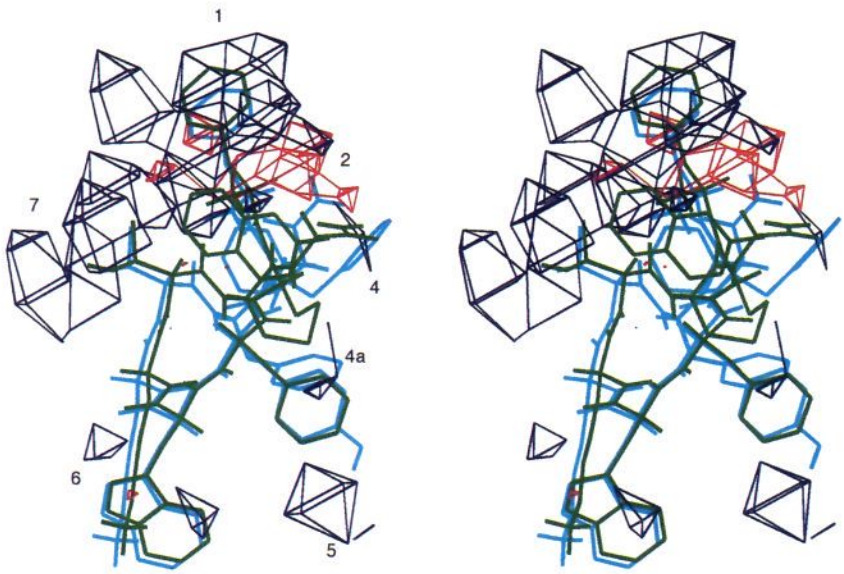

Figure 9. Stereoview of the CoMFA electrostatic sd × coeff contour plot from the PLS analysis based on model A and alignment 1 with 5 components and 64 training compounds. Positive charge favored areas (contribution level of 80%) are represented by blue polyhedra. Negative charge favored areas (contribution level of 20%) are represented by red polyhedra. The highly active analogue, compound **34** (green), is shown with an inactive analogue, compound **10** (cyan). The numbers denote CoMFA polyhedra referenced in the discussion.

groups (Figures 10 and 11; blue, 3). Model C had a large positive region (Figure 11; blue, 4) associated with the side chain of position 6. The analogous region was absent from model B and very minor in model A which was accompanied by a similarly small region associated with the other bridging residue at position 11 (Figure 9; blue, 4a). All the models had positive charge favoring regions near position 7, corresponding with the hydroxyl moiety of Tyr. This was a separate region in model A (Figure 9; blue, 5) but part of the major region associated with the termini in models B and C (Figures 10 and 11; blue, area 1). Model A had several minor electrostatic regions of both types associated with the Lys^9 amino group (Figure 9; area 6). Equivalent regions were absent from model B and only one small negative region was present in model C (Figure 11; red, 6). Model A had a large polyhedron favoring increased positive charge surrounding position 10 (Figure 9; blue, 7). This region was much smaller in model B (Figure 10; blue 7) and absent from model C.

Conclusion

Successful CoMFA models of the growth hormone release-inhibiting activity of somatostatin analogues have been generated using two distinctly different conformations as a basis for molecular modeling in the absence of X-ray crystallographic structures. A convenient objective ranking of the predictive power of a PLS analysis is the PRESS which tends to zero as the predicted activities approach the actual activities. This

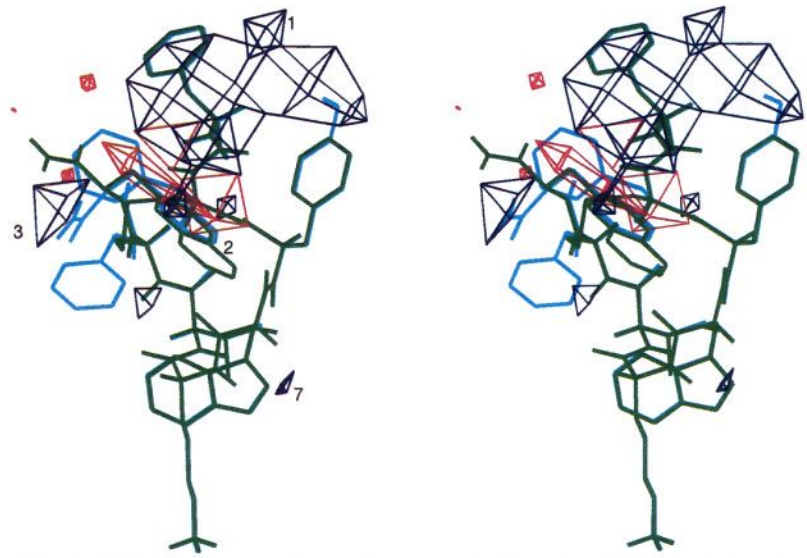

Figure 10. Stereoview of the CoMFA electrostatic sd × coeff contour plot from the PLS analysis based on model B and alignment 1 with 8 components and 64 training compounds. Positive charge favored areas (contribution level of 80%) are represented by blue polyhedra. Negative charge favored areas (contribution level of 20%) are represented by red polyhedra. The highly active analogue, compound **34** (green), is shown with an inactive analogue, compound **10** (cyan). The numbers denote CoMFA polyhedra referenced in the discussion.

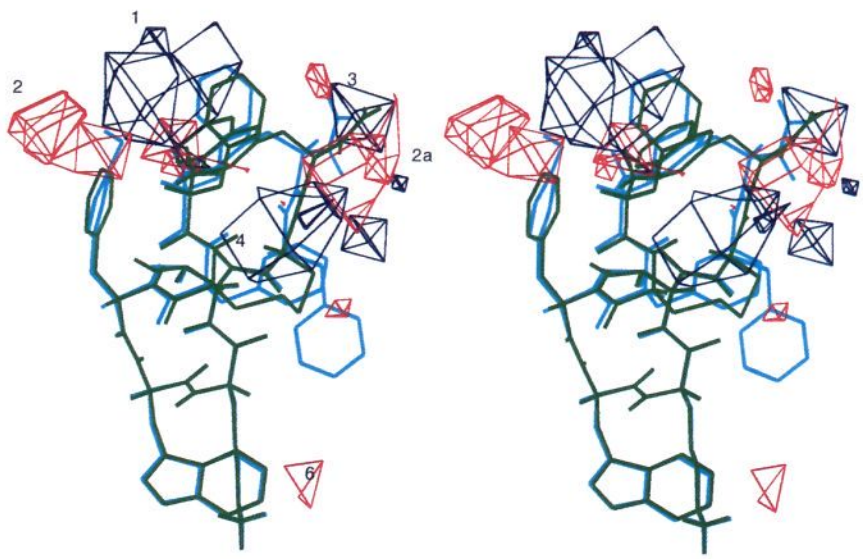

Figure 11. Stereoview of the CoMFA electrostatic sd × coeff contour plot from the PLS analysis based on model C and alignment 1 with 5 components and 64 training compounds. Positive charge favored areas (contribution level of 80%) are represented by blue polyhedra. Negative charge favored areas (contribution level of 20%) are represented by red polyhedra. The highly active analogue, compound **34**, (green), is shown with an inactive analogue, compound **10** (cyan). The numbers denote CoMFA polyhedra referenced in the discussion.

value is a more sensitive measure of a models predictive ability than r^2_{pred} in this work because the test compounds are all more active than the mean of the training set compounds. This was a deliberate choice since the final application of any QSAR work is the discovery of compounds of higher activity, thus it seemed valid to only test the models predictive power in the upper range of the data. The insensitivity of r^2_{pred} to poor predictions in this work is illustrated by model A in Table 11. Alignments 1 and 3 gave reasonable predictions of the activity of test compound DC35-58 but the field-fit alignment (alignment 2) vastly overestimated the activity. Examination of r^2_{pred} and PRESS show that r^2_{pred} dropped only slightly for alignment 2, but PRESS increased 4-fold. This also illustrates the sensitivity of the CoMFA method to slight changes in the orientation of the compounds. Comparison of the CoMFA contour plots of analyses with the three different alignments also demonstrates this sensitivity to orientation. In the case of model A, the steric and electrostatic CoMFA plots for the three alignments contain many of the same features, such as the unfavorable steric region near residue 6 (see Figure 6, area 1), but the size and location of the regions vary with the alignment. This complicates the interpretation of the contour plots with regard to the structural modifications required to enhance biological activity. Indeed, at the standard SYBYL contribution contour levels used in these CoMFA plots, compound **29**, a small highly active analogue, had few features near any of the regions.

Using PRESS as a measure, models A and B gave good estimates of the activity of the test compounds

based on simple rms alignments of the molecules. While model A gave lower PRESS values than model B, the higher minimization gradient end point used for modelling the compounds in model B significantly reduced the computational burden of building the molecular databases. Further work is in progress to extend and refine these models and to investigate the receptor binding affinity of somatostatin to its five known receptors.

Acknowledgment. We would like to thank Dr. Frank Momany for his invaluable suggestions and encouragement, Ms. Etchie Yauger for her excellent support of this work in performing the purification and analysis of the peptides, Ms. Vienna Mackey for her assistance with the biological assays and Ms. Robyn Denenea and Ms. Karyn Van Buren for their administrative support.

References

(1) Vale, W.; Brazeau, P.; Grant, G.; Nussey, A.; Burgus, R.; Rivier, J. E.; Ling, N. Guillemin, R. Premiéres Observations sur le Mode D'action de la Somatostatine, un Facteur Hypothalamique qui Inhibe la Sécrétion de L'hormone de Croissance. *C. R. Seances. Acad. Sci. Paris* **1972**, *275*, 2913–2916.

(2) Brazeau, P.; Vale, W.; Burgus, R.; Ling, L.; Butcher, M.; Rivier, J.; Guillemin, R. Hypothalamic Polypeptide That Inhibits the Secretion of Immunoreactive Pituitary Growth Hormone. *Science* **1973**, *179*, 77–79.

(3) Vale,W.; Rivier, J.; Ling, N.; Brown, M. Biologic and Immunologic Activities and Applications of Somatostatin Analogs. *Metabolism* **1978**, *27*, (Suppl. 1), 1391–1401.

(4) Veber, D. R.; Holly, F. W.; Nutt, R. F.; Bergstrand, S. J.; Brady, S. F.; Hirschmann, R.; Glitzer, M. and Saperstein, R. Highly Active Cyclic and Bicyclic Somatostatin Analogues of Reduced Ring Size. *Nature* **1979**, *280*, 512–514.

(5) Vander Elst, P.; van den Berg, E.; Pepermans, H.; Vander Auwera, L.; Zeeuws, R.; Tourwé, D.; Van Binst, G. Synthesis and Conformational Study of a Cyclic Hexapeptide Analogue of Somatostatin. *Int. J. Pept. Protein Res.* **1987**, *29*, 318–330.

(6) Yamada, Y.; Post. S.; Wand, K.; Tager, H.; Bell, G.; Seino, S. Cloning and Functional Characterization of a Family of Human and Mouse Somatostatin Receptors Expressed in Brain, Gastrointestinal Tract, and Kidney. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 251–255.

(7) Yasuda, K.; Rens-Domiano, S.; Breder, C.; Law, S.; Saper, C.; Reisine, T.; Bell, G. Cloning of a Novel Somatostatin Receptor, SSTR3, Coupled to Adenylyl Cyclase. *J. Biol. Chem.* **1992**, *267*, 20422–20428.

(8) O'Carrol, A.-M.; Lolait, S.; Konig, M.; Mahan, L. Molecular Cloning and Expression of a Pituitary Somatostatin Receptor with Preferential Affinity for Somatostatin-28. *Mol. Pharmacol.* **1992**, *42*, 939–946.

(9) Bruno, J.; Xu, Y.; Song, J.; Berelowitz, M. Molecular Cloning and Functional Expression of a Novel Brain Specific Somatostatin Receptor. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 11151–11155.

(10) Raynor, K.; Murphy, W. A.; Coy, D. H.; Taylor, J. E.; Moreau, J.-P.; Yasuda, K.; Bell, G. I.; Reisine, T. Cloned Somatostatin Receptors: Identification of Subtype-Selective Peptides and Demonstration of High Affinity Binding of Linear Peptides. *Mol. Pharmacol.* **1993**, *43*, 838–844.

(11) Raynor, K.; O'Carroll, A.-M.; Kong, H.; Yasuda, K.; Mahan, L. C.; Bell, G. I.; and Reisine, T. Characterization of Cloned Somatostatin Receptors SSTR4 and SSTR5. *Mol. Pharmacol.* **1993**, *44*, 385–392.

(12) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Bindign of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(13) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. *Tetrahedron Comput. Method* **1990**, *3*, 47–59.

(14) Waller, C. L.; McKinney, J. D. Comparative Molecular Field Analysis of Polyhalogenated Dibenzo-p-dioxins, Dibenzofurans, and Biphenyls. *J. Med. Chem.* **1992**, *6*, 487–504.

(15) Bjorkroth, J.-P.; Pakkanen, T. A.; Lindroos, J. Comparative Molecular Field Analysis of Some Clodronic Acid Esters. *J. Med. Chem.* **1991**, *34*, 2338–2343.

(16) Loughney, S. A.; Schwender, C. F. A Comparison fo Progestin and Androgen Receptor Binding using the CoMFA Technique. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 569–581.

(17) Nicklaus, M. C.; Milne, G. W. A.; Burke, T. R., Jr. QSAR of Conformationally Flexible Molecules: Comparative Molecular Field Analysis of Protein-Tyrosine Kinase Inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 487–504.

(18) Klebe, G.; Abraham, U. On the Prediction of Binding Properties of Drug Molecules by Comparative Molecular Field Analysis. *J. Med. Chem.* **1993**, *36*, 70–80.

(19) 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.

(20) Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-Dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-Determined Alignment Rules. *J. Med. Chem.* **1993**, *36*, 4215–4160.

(21) Van Binst, G.; Tourwé, D. Backbone Modifications in Somatostatin Analogues: Relation Between Conformation and Activity. *Pept. Res.* **1992**, *5*, 1, 8–13.

(22) He, Y.-A.; Huang, Z.; Raynor, K.; Reisine, T.; Goodman, M. Syntheses and Conformations of Somatostatin-Related Cyclic Hexapeptides Incorporating Specific α- and β-Methylated Residues. *J. Am. Chem. Soc.* **1993**, *115*, 8066–8072.

(23) IUPAC–IUB Commission of Biochemical Nomenclature (CBN). Symbols for Amino-Acid Derivatives and Peptides, Recommendations (1971). *Eur. J. Biochem.* **1972**, *27*, 201–207.

(24) Baroni, M.; Constantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20.

(25) Coy, D. H.; Coy, E. J.; Arimura, A.; Schally, A. V. Solid-Phase Synthesis of Growth Hormone Release-Inhibiting Factor. *Biochem. Biophys. Res. Commun.* **1973**, *54*, 1267–1271.

(26) Hocart, S. J.; Murphy, W. A.; Coy, D. H. Proceedings of the Fourteenth American Peptide Symposium, June 18th–23rd, 1995, Columbus, OH, manuscript in preparation.

(27) Murphy, W. A.; Taylor, J.; Moreau, J.-P.; Coy, D. H. Novel Heptapeptide Somatostatin Analog Displays Anti-Tumor Activity Independent of Effects on Growth Hormone Secretion. *Pept. Res.* **1989**, *2*, 128–132.

(28) Pugsley, L. I. The Application of the Principles of Statistical Analysis to the Biological Assay of Hormones. *Endocrinology* **1946**, *39*, 161.

(29) SYBYL molecular modeling software, version 6.03, Tripos Associates Inc.; 1699 Hanley Rd, St. Louis, MO 63144-2913.

(30) Clark, M.; Cramer, R. D., III; Van Opdenbosh, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.

(31) Berthod, H.; Pullman, A. Calculation of the σ Structure of Conjugated Molecules. *J. Chem. Phys.* **1965**, *62*, 942–946.

(32) Berthod, H.; Prettre, C. G.; Pullman, A. Role of σ and π electrons on the Properties of Halogens-Substituted Conjugated Molecules. Application to the Study of Uracil and Fluororacil. *Theor. Chim. Acta* **1967**, *8*, 212–222.

(33) McCammon, J. A.; Wolynes, P. G.; Karplus, M. Picosecond Dynamics of Tyrosine Side Chains in Proteins. *Biochemistry* **1979**, *18*, 927–942.

(34) Huang, Z.; He, Y.-B.; Raynor, K.; Tallent, M.; Reisine, T.; Goodman, M. Main Chain and Side Chain Chiral Methylated Somatostatin Analogs: Syntheses and Conformational Analyses. *J. Am. Chem. Soc.* **1992**, *114*, 9390–9401.

(35) Brady, S. F.; Nutt, R. F.; Holly, F. W.; Paleveda, W. J.; Strachan, R. G.; Bergstrand, S. J.; Veber, D. F.; Saperstein, R. Synthesis and Biological Activity of Somatostatin Analogs of Reduced Ring Size. In *Peptides: Synthesis, structure, function. Proceedings of the Seventh American Peptide Symposium*; Rich, D. H., Gross, E., Eds.; Pierce Chemical Co.: New York, 1981; pp 653–656.

(36) Veber, D. F.; Saperstein, R.; Nutt, R. F.; Freidinger, R. M.; Brady, S. F.; Curley, P.; Perlow, D. S.; Paleveda, W. J.; Colton, C. D.; Zacchei, A. G.; Tocco, D. J.; Hoff, D. R.; Vandlen, R. L.; Gerich, J. E.; Hall, L.; Mandarino, L.; Cordes, E. H.; Anderson, P. S.; Hirschmann, R. A Super Active Cyclic Hexapeptide Analog of Somatostatin. *Life Sci.* **1984**, *34*, 1371–137.

(37) Veber, D. F. Design of a Highly Active Cyclic Hexapaptide Analog of Somatostatin. In *Peptides: Synthesis, structure, function. Proceedings of the Seventh American Peptide Symposium*; Rich, D. H., Gross, E., Eds.; Pierce Chemical Co.: New York, 1981; pp 685–694.

(38) Veber, D. F. Design and Discovery in the Development of Peptide Analogs. In *Peptides: Chemistry and Biology. Proceedings of the Twelfth American Peptide Symposium*; Smith, J. A., Rivier, J. E., Eds.; ESCOM: Leiden, 1992; pp 3–14.

(39) Databases containing the molecules of the three models can be downloaded using the following universal resource locator (URL) http://www.mcl.tulane.edu/

(40) *SYBYL 6.1 Theory Manual*; Tripos Associates: St. Louis, MO, August 1994; p 59.

JM940733S