JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Quantitative Structure–Activity Relationship of Rubiscolin Analogues as δ Opioid Peptides Using Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA)

JULIO CABALLERO,*,† MARIO SAAVEDRA,† MICHAEL FERNÁNDEZ,‡ AND FERNANDO D. GONZÁLEZ-NILO†

Centro de Bioinformática y Simulación Molecular, Universidad de Talca, 2 Norte 685, Casilla 721, Talca, Chile, and Molecular Modeling Group, Center for Biotechnological Studies, University of Matanzas, Matanzas, Cuba

Three-dimensional quantitative structure–activity relationship (3D-QSAR) studies were carried out on a series of 38 rubiscolins as δ opioid peptides using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Quantitative information on structure–activity relationships is provided for further rational development and direction of selective synthesis. All models were carried out over a training set including 30 peptides. The best CoMFA model included electrostatic and steric fields and had a moderate $Q^2 = 0.503$. CoMSIA analysis surpassed the CoMFA results: the best CoMSIA model included only the hydrophobic field and had a $Q^2 = 0.661$. In addition, this model predicted adequately the peptides contained in the test set. Our model identified that the potency of δ opioid activity of rubiscolin analogues essentially exhibited a significant relationship with local hydrophobic and hydrophilic characteristics of amino acids at positions 3, 4, 5, and 6.

KEYWORDS: Rubiscolin analogues; quantitative structure–activity relationships; CoMFA; CoMSIA

INTRODUCTION

Opioid peptides are defined as short sequences of amino acids that mimic the effect of opiates in the brain and play a critical role in a variety of biological processes, including analgesia, constipation, respiration, euphoria, sedation, and meiosis (*1*). Opioid peptides may be produced by the body itself, for example, endorphins (*2*), or be absorbed from partially digested food (casomorphins (*3*), exorphins (*4*), and rubiscolins (*5*)).

Recently, rubiscolin-6 (YPLDLF) and rubiscolin-5 (YPLDL) have been identified as bioactive peptides formed during digestion of D-ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) from spinach leaves (*5*). Because of its abundance in green vegetables, Rubisco serves as a food resource. Rubiscolins exhibited antinociceptive effect in mice even after oral administration (*5*), and orally administered rubiscolin-6 enhanced memory consolidation in a passive avoidance task (*6*). Rubiscolins have an aliphatic amino acid in the third position and are δ selective; in these aspects they differ from most YP type opioid peptides which contain aromatic amino acids in the third position and are mostly μ selective. More recently, Yang et al. investigated the structure–activity relationship (SAR) of rubiscolin analogues (*7*). The authors left the Tyr–Pro sequence at the N-terminus unchanged and made amino acid substitutions at the third, fourth, fifth, and sixth positions. They found sequences giving more potent δ opioid activity than rubiscolin-6.

To understand the rubiscolin–receptor interactions in detail, knowledge of the δ opioid receptor structure at the atomic level is desirable. Unfortunately, because of their size and the inherent obstacles in crystallizing complex membrane proteins, no experimental three-dimensional (3D) structure is currently available. Several 3D models for the human δ opioid receptor have been generated based on the rhodopsin projection maps or the bovine rhodopsin crystal structure (*8–11*). However, the exact site of YP type δ opioid peptides binding to the receptor is unknown.

In this paper, we investigated the structural requirements of rubiscolin analogues for having a high δ opioid activity using 3D quantitative structure–activity relationship (QSAR) analysis. QSAR is a chemoinformatic methodology that can extract information from a data set to produce knowledge about the requirements of the ligands without considering the structure of the biological target. We attempted to obtain a 3D quantitative description of the SAR of rubiscolin analogues using compara-

* Corresponding author: Tel: (56) (71) 201 662. Fax: (56) (71) 201 561. E-mail: jcaballero@utalca.cl and jmcr77@yahoo.com.
† Universidad de Talca.
‡ University of Matanzas.

10.1021/jf071031h CCC: $37.00
Published on Web 09/06/2007

Table 1. Experimental and Predicted δ Opioid Activities of Rubiscolin Analogues ($\log(10^6/IC_{50})$) using Model CoMSIA-3

peptide	sequence	experimental $\log(10^6/IC_{50})$	predicted $\log(10^6/IC_{50})$
		Training Set	
1	YPLDLF	4.613	4.758
2	YPIDLF	5.217	4.967
3	YPMDLF	5.248	5.302
4	YPVDLF	3.213	3.179
5	YPFDLF	3.611	3.637
6	YPLALF	3.134	3.167
7	YPLELF	2.593	2.530
8	YPLHLF	3.317	3.373
9	YPLDIF	4.257	4.607
10	YPLDMF	3.842	4.003
11	YPLDVF	3.233	3.041
12	YPLDAF	3.000	2.698
13	YPLDLL	4.458	4.530
14	YPLDLI	5.203	5.417
15	YPLDLM	4.580	4.540
16	YPLDLV	5.693	5.215
17	YPLDLA	4.745	4.697
18	YPIDIV	5.071	5.273
19	YPMDIV	5.710	5.608
20	YPIDMV	4.870	4.668
21	YPIDVV	3.446	3.706
22	YPIDAV	3.025	3.364
23	YPIDII	5.799	5.475
24	YPMDII	5.714	5.810
25	YPLDL	4.292	4.426
26	YPMDL	4.824	4.970
27	YPFDL	4.573	4.547
28	YPMDI	4.896	4.812
29	YPMDV	3.910	3.757
30	YPMDF	4.293	4.301
		Test Set	
31	YPADLF	3.234	4.041
32	YPWDLF	3.022	3.619
33	YPIDLV	5.305	5.420
34	YPMDLV	5.917	5.762
35	YPLDIV	4.917	5.060
36	YPFDII	3.873	5.388
37	YPIDL	4.668	4.642
38	YPMDM	4.395	3.871

tive molecular field analysis (CoMFA) (*12*) and comparative molecular similarity indices analysis (CoMSIA) (*13*). The outcome of the present work is a comprehensive qualitative and quantitative description of the molecular features relevant for a high δ opioid activity, which completes the picture of the SAR of this class of compounds.

MATERIALS AND METHODS

The primary structures and activities of 38 Rubiscolin analogues were taken from the literature (*7*). The activities tested by a mouse vas deferens (MVD) assay (*4*), which is sensitive for δ opioids, were collected and transformed into $\log(10^6/IC_{50})$ values. IC_{50} values represent the compound concentrations that inhibit the electrically stimulated muscle contractions by 50% in MVD assays (*4*). The peptide sequences and biological activities used in this study are summarized in **Table 1**.

Molecular modeling was performed using the Sybyl 7.2 software of Tripos (*14*). All the molecules were sketched using the Biopolymer module in Sybyl. Each structure was fully geometry-optimized using a conjugate gradient procedure based on the TRIPOS force field (*15*) and Gasteiger-Marsili charges (*16*) and then aligned by an atom-by-atom least-square fit. We used the backbone, including only the first five N-terminal residues, of the most active compound **34** in its optimized conformation as a template. For a stronger evaluation of model applicability for prediction on new chemicals, the external validation of the models is also recommended (*17*). Therefore, the data set was divided into two subdata sets. Eight compounds were chosen randomly as a test set and were used for external validation of the 3D-QSAR models; the training sets included all the remaining 30 compounds.

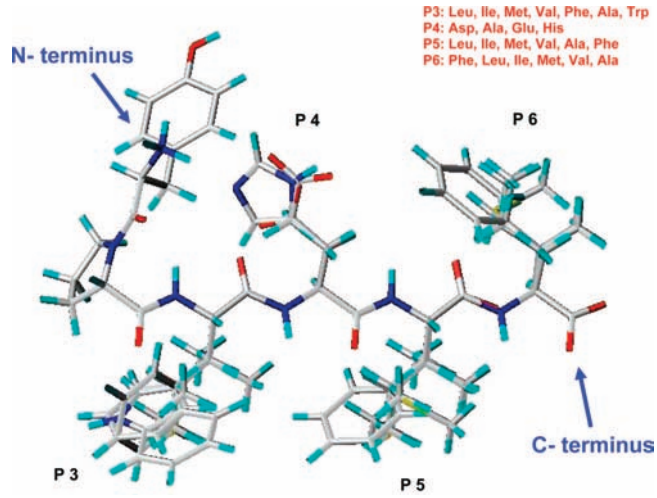

Figure 1. Atom-by-atom superposition used for 3D-QSAR analysis. Amino acids at positions 3, 4, 5, and 6 of rubiscolin's sequence are indicated.

For the QSAR calculations, molecules were placed in a rectangular grid and the interaction energies between a probe atom and all compounds were computed at the surrounding points, using a volume-dependent lattice with 2.0 Å grid spacing. Then, standard Sybyl parameters were used for a partial least squares (PLS) analysis. The number of components in the PLS models were optimized by using a Q^2 value, obtained from the leave-one-out (LOO) cross validation procedure, with the SAMPLS (*18*) sampling method. The number of components was increased until additional components did not increase Q^2 by at least 5% per added component. The CoMFA models were generated by using steric and electrostatic probes with standard 30 kcal/mol cutoffs. In the CoMSIA analyses, similarity is expressed in terms of steric occupancy, electrostatic interactions, local hydrophobicity, and H-bond donor and acceptor properties, using a 0.3 attenuation factor.

The modeling capability (goodness of fit) was judged by the correlation coefficient squared, R^2. The prediction capability (goodness of prediction) was indicated by the cross-validated R^2 (Q^2) and by the percent relative error (PRE) of the predictions of the test set.

RESULTS AND DISCUSSION

Figure 1 shows the aligned molecules within the grid box (grid spacing 2.0 Å) used to generate the CoMFA and CoMSIA columns. In the structure of rubiscolins, the protonated Tyr at the N-terminus and the second residue Pro are obligatory for opioid activity, so they are unchanged in all structures (*19, 20*). The alignment ensures a straightforward determination of the relevant effects at positions 3, 4, 5, and 6 of the rubiscolin structure.

The stepwise development of CoMFA and CoMSIA models using different fields is presented in **Table 2**. The predictability of the models is the most important criterion for assessment of both methods. CoMFA models using steric or electrostatic fields are statistically unacceptable ($Q^2 < 0.5$). After both fields were considered, a CoMFA model with a better statistical significance ($Q^2 = 0.503$) was obtained (CoMFA-3). This model reveals that steric field has a major influence on the opioid activity.

In comparison to CoMFA, CoMSIA methodology has the advantage of exploring more fields. The best model (CoMSIA-3) has a Q^2 value of 0.661 using nine components and only includes the hydrophobic field; meanwhile, models that only include steric, electrostatic, H-bond-acceptor and H-bond-donor fields had Q^2 values below 0.5. We evaluated whether the

Table 2. Results of the CoMFA and CoMSIA Analyses using Several Different Field Combinations[a]

							fraction			
	NC	R^2	s	F	Q^2	s_{CV}	steric	electrostatic	hydrophobic	H-bond acceptor	H-bond donor
CoMFA-1	3	0.721	0.519	22.427	0.458	0.724	1				
CoMFA-2	9	0.920	0.316	25.692	0.404	0.865		1			
CoMFA-3	3	0.770	0.472	28.968	0.503	0.693	0.834	0.166			
CoMSIA-1	3	0.691	0.546	19.400	0.438	0.737	1				
CoMSIA-2	9	0.906	0.344	21.401	0.465	0.820		1			
CoMSIA-3	**9**	**0.956**	**0.234**	**48.793**	**0.661**	**0.551**			**1**		
CoMSIA-4	2	0.046	0.943	0.646	−0.09	1.071				1	
CoMSIA-5	4	0.621	0.617	10.228	0.218	0.887					1
CoMSIA-6	4	0.845	0.395	34.007	0.575	0.653	0.637	0.363			
CoMSIA-7	4	0.609	0.627	9.751	0.178	0.909				0.070	0.930
CoMSIA-8	9	0.957	0.233	49.211	0.612	0.698	0.306		0.694		
CoMSIA-9	10	0.957	0.238	42.624	0.544	0.777		0.261	0.739		
CoMSIA-10	8	0.954	0.236	53.876	0.627	0.689			0.986	0.014	
CoMSIA-11	8	0.955	0.233	55.235	0.403	0.846			0.659		0.341
CoMSIA-12	3	0.816	0.421	38.532	0.567	0.647	0.247	0.269	0.484		
CoMSIA-13	8	0.956	0.230	56.531	0.552	0.733	0.297		0.685	0.018	
CoMSIA-14	3	0.764	0.478	28.064	0.420	0.749	0.235		0.457		0.308
CoMSIA-15	11	0.957	0.244	36.844	0.539	0.803		0.256	0.736	0.008	
CoMSIA-16	10	0.958	0.236	43.213	0.494	0.818		0.173	0.626		0.202
CoMSIA-17	3	0.749	0.493	25.828	0.374	0.778			0.582	0.051	0.367
CoMSIA-18	3	0.812	0.426	37.428	0.540	0.667	0.243	0.251	0.485	0.021	
CoMSIA-19	3	0.794	0.446	33.391	0.495	0.699	0.195	0.208	0.388		0.209
CoMSIA-20	3	0.758	0.484	27.091	0.406	0.758	0.225		0.449	0.044	0.282
CoMSIA-21	8	0.952	0.240	51.744	0.485	0.785		0.178	0.612	0.006	0.204
CoMSIA-22	4	0.842	0.398	33.382	0.491	0.715	0.193	0.220	0.375	0.016	0.196

[a] NC is the number of components from PLS analysis, R^2 is the square of the correlation coefficient, s is the standard deviation of the regression, F is the Fischer ratio, and Q^2 and s_{CV} are the correlation coefficient and standard deviation of the leave-one-out (LOO) cross-validation, respectively.

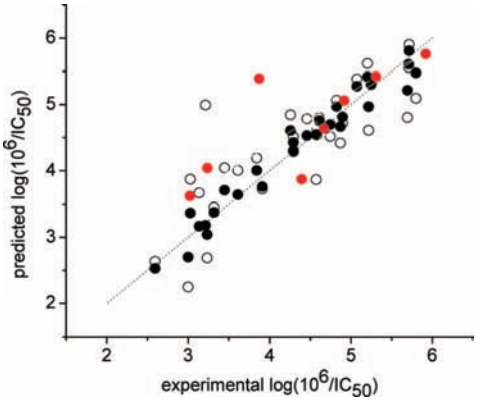

Figure 2. Scatter plot of the experimental activities versus predicted activities for model CoMSIA-3: (●) training set predictions, (○) LOO cross-validated predictions, (red circles) test set predictions.

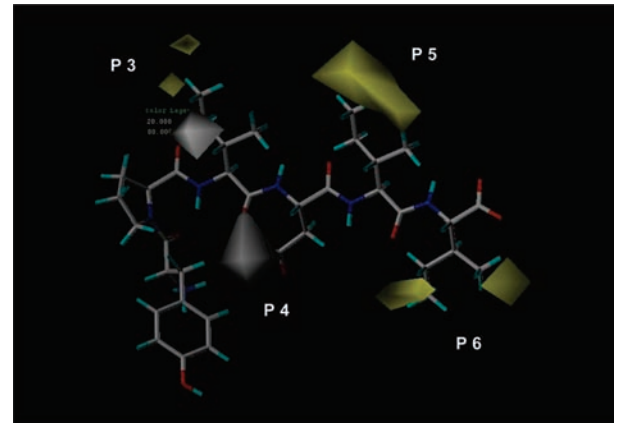

Figure 3. Hydrophobic CoMSIA stdev*coeff contour map. Peptide **23** is shown inside the field. Yellow and gray polyhedra indicate regions where hydrophobic or hydrophilic groups, respectively, will enhance the affinity. Amino acids at positions 3, 4, 5, and 6 of rubiscolin's sequence are indicated.

addition of other fields produces an improvement in the internal validation of the model CoMSIA-3. Models including hydrophobic field with other fields had lower Q^2 values (**Table 2**). Furthermore, the hydrophobic field always had the major contribution in these models. This means that the hydrophobic field is enough for describing the relationship between the structure of rubiscolin analogues and their δ opioid activity, and the inclusion of other fields does not contribute to the QSAR analysis.

The best model CoMSIA-3 also surpassed the results achieved by CoMFA analysis (the statistical parameters are highlighted in **Table 2**). This model explains 95.6% of the variance and has a low standard deviation ($s = 0.234$) and a high Fischer ratio ($F = 48.793$). The predictions of $\log(10^6/IC_{50})$ values for the 30 rubiscolin analogues using model CoMSIA-3 is shown in **Table 1**. The correlation between the calculated and the experimental values of $\log(10^6/IC_{50})$ (from training and LOO cross-validation) is shown in **Figure 2**. We also used model CoMSIA-3 to predict the opioid activities of the test set compounds. The values are given in **Table 1**, and correlation between the calculated and the experimental values are represented in **Figure 2**. This analysis reveals that the proposed model fails in the prediction of compound **36** [$\log(10^6/IC_{50}) = 5.388$ instead of 3.873; PRE = 39.2]. However, the remaining compounds were predicted adequately.

The contour plot of the CoMSIA hydrophobic fields (stedv*coeff) is presented in **Figure 3**. Favored and disfavored levels fixed at 80 and 20%, respectively, were used. Peptide **23** (YPIDII) is shown inside the hydrophobic field. The peptide is positioned with the N-terminus to the left and the C-terminus to the right. The well-tolerated hydrophobic groups are shown as yellow contours. Such areas exist at positions P3, P5, and P6. The favored hydrophilic groups are shown as gray contours. They are at positions P3 and P4.

The hydrophobic map indicates that an amino acid bearing a hydrophilic atom at position δ of the side chain and a hydrophobic atom in the outer is preferred at position 3. According to the map, the most suitable amino acid for this position is Met, whereas aromatic residues or short hydrophobic residues such as Val and Ala decreased the δ opioid activity. Yang et al. identified that charged residues are required at the fourth position in rubiscolin analogues (*7*). The hydrophobic map states that a hydrophilic atom must be contained at position δ of the side chain of the amino acid at position 4. This requirement is only accomplished by Asp. Substitutions of Asp with larger residues such as Glu and His or shorter residues such as Ala greatly weakened the opioid activity.

According to the CoMSIA hydrophobic map, and previous SAR analysis, rubiscolin analogues in the fifth and sixth positions must contain hydrophobic residues for increasing their δ opioid activity. The hydrophobic map at position 5 indicates that large hydrophobic side chains are preferred for this position. Large hydrophobic amino acids such as Leu, Ile, Met, and Phe contribute positively to the δ opioid activity. When these residues are replaced by shorter residues such as Val and Ala, the activity decreases considerably. Otherwise, replacements at position 6 have a minor influence on the δ opioid activity. Since yellow contours are at this position, the absence of residues decreases the activity; in general, rubiscolin-6 analogues are more potent than rubiscolin-5 analogues. The contours show that β-branched residues such as Ile and Val might be the better hydrophobic residues at position 6.

In conclusion, traditional CoMFA and CoMSIA approaches have been applied to derive quantitative relationships between the structure of rubisculin analogues and their δ opioid activity. The present study indicates that the steric and electrostatic CoMFA fields are not enough to describe fully the rubiscolin analogues' δ opioid activity. When using the hydrophobic CoMSIA field, a statistically meaningful model was derived. Thus, prediction of δ opioid activities with sufficient accuracy should be possible. Moreover, an interpretation of the respective hydrophobic field makes it possible to draw conclusions concerning the most appropriate amino acids for positions 3, 4, 5, and 6 of the rubiscolin analogues.

LITERATURE CITED

(1) Vaccarino, A. L.; Kastin, A. J. Endogenous Opiates: 2000. *Peptides* **2001**, *22*, 2257–2328.

(2) Zadina, J. E.; Hackler, L.; Ge, L. J.; Kastin, A. J. A potent and selective endogenous agonist for the μ-opiate receptor. *Nature* **1997**, *386*, 499–502.

(3) Brantl, V.; Teschemacher, H.; Henschen, A.; Lottspeich, F. Novel opioid peptides derived from casein (β-casomorphins)I. Isolation from bovine casein peptone. *Hoppe-Seyler's Z. Physiol. Chem.* **1979**, *360*, 1211–1216.

(4) Fukudome, S.; Yoshikawa, M. Opioid peptides derived from wheat gluten: their isolation and characterization. *FEBS Lett.* **1992**, *296*, 107–111.

(5) Yang, S.; Yunden, J.; Sonoda, S.; Doyama, N.; Lipkowski, A. W.; Kawamura, Y.; Yoshikawa, M. Rubiscolin, a δ selective opioid peptide derived from plant Rubisco. *FEBS Lett.* **2001**, *509*, 213–217.

(6) Yang, S.; Kawamura, Y.; Yoshikawa, M. Effect of rubiscolin, a δ opioid peptide derived from Rubisco, on memory consolidation. *Peptides* **2003**, *24*, 325–328.

(7) Yang, S.; Sonoda, S.; Chen, L.; Yoshikawa, M. Structure-activity relationship of rubiscolins as δ opioid peptides. *Peptides* **2003**, *24*, 503–508.

(8) Alkorta, I.; Loew, G. H. A 3D model of the δ opioid receptor and ligand-receptor complexes. *Protein Eng.* **1996**, *9*, 573–583.

(9) Strahs, D.; Weinstein, H. Comparative modeling and molecular dynamics studies of the δ, κ and μ opioid receptors. *Protein Eng.* **1997**, *10*, 1019–1038.

(10) Pogozheva, I. D.; Lomize, A. L.; Mosberg, H. I. Opioid receptor three-dimensional structures from distance geometry calculations with hydrogen bonding constraints. *Biophys. J.* **1998**, *75*, 612–634.

(11) Aburi, M.; Smith, P. E. Modeling and simulation of the human δ opioid receptor. *Protein Sci.* **2004**, *13*, 1997–2008.

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

(13) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.

(14) *SYBYL*, version 7.2; Tripos Inc.: 1699 South Hanley Rd., St. Louis, MO 63144, USA.

(15) Clark, M.; Cramer, R. D.; Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012.

(16) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.

(17) Golbraikh, A.; Tropsha, A. Beware of q2. *J. Mol. Graph. Model.* **2002**, *20*, 269–276.

(18) Bush, B. L.; Nachbar, R. B., Jr. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.

(19) Hruby, V. J.; Gehrig, C. A. Recent developments in the design of receptor specific opioid peptides. *Med. Res. Rev.* **1989**, *9*, 343–401.

(20) Mierke, D. F.; Nossner, G.; Schiller, P. W.; Goodman M. Morphiceptin analogs containing 2-aminocyclopentane carboxylic acid as a peptidomimetic for proline. *Int. J. Pept. Protein Res.* **1990**, *35*, 35–45.

Received for review April 9, 2007. Revised manuscript received August 6, 2007. Accepted August 8, 2007. This work was supported by Innova Chile (FDI, CORFO), Project 06CN12PAT-51. We express our thanks to "Red Iberoamericana de Bioinformática" RIB from CYTED, for partial financial support.

JF071031H