Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 14 (2006) 6874–6885
Bioorganic & Medicinal Chemistry

Development of new CoMFA and CoMSIA 3D-QSAR models for anti-inflammatory phthalimide-containing TNFα modulators

Carolina Martins Avila,[a] Nelilma Correia Romeiro,[a] Gilberto M. Sperandio da Silva,[a] Carlos M. R. Sant'Anna,[a,b] Eliezer J. Barreiro[a] and Carlos A. M. Fraga[a,*]

[a]*Laboratório de Avaliação e Síntese de Substâncias Bioativas (LASSBio), Faculdade de Farmácia, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, PO Box 68023, RJ 21944-970, Brazil*
[b]*Departamento de Química, ICE, Universidade Federal Rural do Rio de Janeiro (UFRRJ), Seropédica, RJ 23851-970, Brazil*

Received 8 May 2006; revised 15 June 2006; accepted 19 June 2006
Available online 14 July 2006

Abstract—In the present study, we describe a new 3D-QSAR analysis of 42 previously reported thalidomide analogues, with the ability to modulate the pro-inflammatory cytokine TNFα, by using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Three statistically significant models were obtained. The best resulting CoMFA and CoMSIA models have conventional r^2 values of 0.996 and 0.983, respectively. The cross-validated q^2 values are 0.869 and 0.868, respectively. The analysis of CoMFA and CoMSIA contour maps provided insight into the possible sites for structural modification of the thalidomide analogues for better activity and reduced toxicity.
© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

Tumor necrosis factor α (TNFα) is a key cytokine produced primarily by monocytes and macrophages, which is involved in the host immune response, contributing to the pathogenesis of both infectious and autoimmune diseases.[1] During normal host defense, low levels of serum TNFα confer protection against infectious agents, tumors, and tissue damage, and have an important role in the development of the humoral immune response in mice. On the other hand, increased concentrations of TNFα have been shown to trigger the lethal effects of septic shock syndrome.[2]

Thalidomide, α-*N*-phthalimideglutarimide (**1**, Fig. 1), approved by the Food and Drug Administration (FDA) agency in July 1998 for the treatment of erytrema nodosum leprosum (ENL), is an anti-inflammatory and immunomodulatory drug that was originally used as a sedative, although it is now widely associated to its teratogenic and neurotoxic properties.[3] Thalidomide (**1**) selectively inhibits TNFα production by lipopolysaccharide (LPS)-stimulated human monocytes.[4] Its inhibitory action on TNFα seems to be exerted by enhancing mRNA degradation.[5,6]

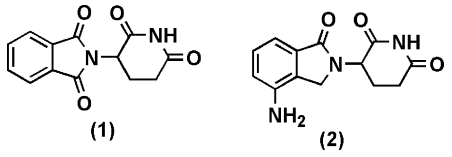


Figure 1. Chemical structures of anti-inflammatory drugs thalidomide (**1**) and lenalidomide (**2**).

The levels of other cytokines, such as IL-1β, IL-6, and granulocyte macrophage-colony stimulating factor (GMCSF), are also inhibited by thalidomide, whereas IL-10 is stimulated.[7] The interest in this drug was renewed due to the potential clinical benefits associated to its selective inhibition of TNFα biosynthesis. Present applications include treatment of ENL,[8] rheumatoid arthritis,[9] HIV-associated aphthous ulceration,[10] chronic tuberculosis,[11] chronic graft-versus-host disease,[12] and a variety of tumors.[13]

In order to obtain new lead compounds with increased anti-TNFα activity and reduced toxicity, useful for the treatment of chronic anti-inflammatory diseases, novel

Keywords: Thalidomide; Lenalidomide; TNFα; Phthalimide derivatives; 3D-QSAR; CoMFA; CoMSIA.
* Corresponding author. Tel.: +55 21 25626503; fax: +55 21 25626644; e-mail: cmfraga@pharma.ufrj.br

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.06.042

TNFα inhibitors have been designed using thalidomide (**1**) as structural template.[1] A good example of that is lenalidomide (**2**, Revlimid™) (Fig. 1), approved by FDA in December 2005 for the treatment of patients with myelodysplastic syndrome. Lenalidomide (**2**) is a potent inhibitor of TNFα production, 2000 times more potent than thalidomide (**1**).[14]

This strategy has demanded larger knowledge of the relationships between the chemical structure and the bioactivity profile related to TNFα modulation. In this context, molecular modeling strategies come as important tools in understanding the mechanism of interaction between various receptors and ligands, with prominence for the CoMFA and CoMSIA methods among the most powerful tools in the 3D-QSAR approach.[15,16] Recently, CoMFA and CoMSIA models for a series of thalidomide analogues as angiogenesis inhibitors have been reported, and shed light on the structural requisites that are important for the antiangiogenic activity.[17]

In this context, as part of a research program aiming at the discovery of new phthalimide-containing anti-inflammatory prototypes able to modulate selectively the cytokine TNFα,[18,19] we have become interested in applying 3D-QSAR methods to construct CoMFA and CoMSIA models for a series of TNFα modulators that have emerged as potential drugs for the treatment of anti-inflammatory diseases. These studies are especially important because the mechanism of action of these substances is not clear and their target bioreceptor has not been identified yet. The resulting models are expected to give insight into the influence of structural characteristics on the anti-inflammatory profile and thus aid in designing new potent TNFα modulators with fewer side effects.

2. Results and discussion

After the generation of CoMFA and CoMSIA models, the statistical validity of the models was judged by high values of cross-validated q^2 (more than 0.8) and non-validated r^2 (more than 0.9), and also the lowest standard errors of estimation (SEE). The models that fulfill these criteria were selected for discussion.

2.1. CoMFA studies

Initially two different alignments were tested (Fig. 3). In the first alignment, which considered only the rigid phenyl ring of the phthalimide moiety, a good q^2 value of 0.624 was obtained. However, high statistical errors, for example, a S_{cv} of 0.776 (data not shown), discouraged us to proceed with this strategy. It is worth mentioning that the grid spacing variation of 1.0, 1.5, and 2.0 Å did not improve the models. These results may be rationalized by the observation that, in this alignment, the adjacent structural moiety of some compounds adopts different orientations. Thus, high variation in the orientation of these moieties, that seem to account for most of differential activity of this class of compounds, was detrimental to the success of the model. The sensitivity of CoMFA to the overall orientation of rigidly aligned compounds has been previously reported.[20]

In an attempt to improve these results, we have next tried a better superimposition through the strategy of alignment 2, including atoms from the structural framework adjacent to phthalimide moiety, which in fact occurred. The results obtained from the CoMFA studies using alignment 2 are summarized in Table 3.

The best CoMFA model was obtained with the combination of cutoffs of 25 and 35 kcal mol^{-1} for the steric and electrostatic field contributions, respectively. The optimal number of components (N) was found to be 6 (data from components 1–5 not shown), and grid spacing was 1.0 Å. This result is in accordance with previously reported data about the effect of decreasing grid spacing on the generation of high q^2 values.[20] Most probably, the reason for this behavior was that the decrease in grid spacing increases the number of interaction points which, in turn, should raise the probability of placing the probe atom in a region where the steric and electrostatic field changes can be best correlated with biological activity. Thus, all further results for steric and electrostatic field contributions alone have been based on these values.[20]

The analysis of the resulting 3D-QSAR models showed that the best CoMFA model was obtained with combined steric/electrostatic fields, that yields a cross-validated q^2 of 0.869, a non-cross-validated r^2 of 0.996, an estimated F value of 952.856, and low standard errors of prediction (S_{cv}) and of estimation (SEE) of 0.422 and 0.078, respectively. These statistical indexes are very good, suggesting that the CoMFA model has a good predictive ability.

2.2. Visual inspection of 3D contour maps generated by CoMFA

In order to visualize the information content in the resulting 3D-QSAR models, CoMFA contour maps were generated by interpolating the products between the 3D-QSAR coefficients and their associated standard deviations. The three-dimensional representations of the CoMFA field contributions as contour maps reveal where variability in molecules' fields is able to explain experimental binding differences.

The CoMFA contour maps have been generated from the analysis based on alignment 2 for the best model derived by steric/electrostatic field combination. One of the most active compounds of the training set (**2**, $pIC_{50} = 7.00$), that is the commercially available TNFα modulator lenalidomide, and the least active derivative (**3**, $pIC_{50} = 3.58$) were superimposed on both maps (Fig. 4) in order to understand the different activity profiles of these compounds.

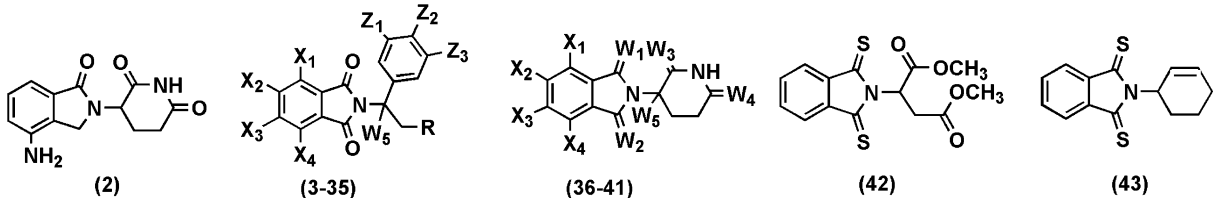
The representation illustrated in Figure 4 shows the electrostatic maps contribution of the best model generated by CoMFA, in which the blue and red contours

correspond to regions where an increase in positive or negative charge, respectively, will enhance the bioactivity profile. The blue and red contour maps are mainly localized around the adjacent moiety attached to the phthalimide ring, suggesting their importance for the electrostatic interactions with the putative receptor.

When we analyze the electrostatic contour maps around compound **3** and compare them to those around compound **2**, it is clear that the carbonyl of the amide group of **3** falls in a region where an increase in positive charge is activity enhancing. Since the electron density of the carbonyl group is polarized toward the more electronegative oxygen atom, it is probably one of the structural reasons for the decreased activity of this derivative. On the other hand, there are no electron-rich groups in the molecular structure of **2** pointing toward this region.

Additionally, the non-substituted phenyl ring of **3** is surrounded by red contour maps in Figure 4, while, in

Table 1. Molecular structures of phthalimide-containing derivatives of the training and test sets (bold italics) along with their respective pIC_{50} values



Compound	X_1	X_2	X_3	X_4	Z_1	Z_2	Z_3	R	W_1	W_2	W_3	W_4	W_5	pIC_{50}[a]
2	—	—	—	—	—	—	—	—	—	—	—	—	—	7.00
3	H	H	H	H	H	H	H	$CONH_2$	O	O	—	—	H	3.58
4	H	H	H	H	H	H	CN	$CONH_2$	O	O	—	—	H	4.05
5	H	H	H	H	H	CN	H	$CONH_2$	O	O	—	—	H	3.82
6	H	H	H	H	H	H	OMe	$CONH_2$	O	O	—	—	H	3.92
7	H	H	H	H	H	OMe	H	$CONH_2$	O	O	—	—	H	4.20
8	H	H	H	H	H	OMe	OMe	$CONH_2$	O	O	—	—	H	4.92
9	H	H	H	H	H	OEt	OEt	$CONH_2$	O	O	—	—	H	5.25
10	H	H	H	H	H	OPr	O-*n*-Pr	$CONH_2$	O	O	—	—	H	4.25
11	H	H	H	H	H	Cl	Cl	$CONH_2$	O	O	—	—	H	4.15
12	H	H	H	H	H	Ph		$CONH_2$	O	O	—		—	4.29
13	H	H	H	H	OMe	H	OMe	$CONH_2$	O	O	—	—	H	4.40
14	H	H	H	H	H	OMe	OMe	CONHMe	O	O	—	—	H	4.92
15	H	H	H	H	H	OMe	OMe	CONHEt	O	O	—	—	H	4.27
16	H	H	H	H	H	OMe	OMe	COCNBn	O	O	—	—	H	4.07
17	***H***	***H***	***H***	***H***	***H***	***OMe***	***OMe***	***CO_2H***	***O***	***O***	—	—	***H***	***4.22***
18	***H***	***H***	***H***	***H***	***H***	***OMe***	***OMe***	***CH_2OH***	***O***	***O***	—	—	***H***	***5.12***
19	***H***	***H***	***NO_2***	***H***	***H***	***OMe***	***OMe***	***CO_2Me***	***O***	***O***	—	—	***H***	***4.46***
20	***H***	***H***	***NH_2***	***H***	***H***	***OMe***	***OMe***	***CO_2Me***	***O***	***O***	—	—	***H***	***6.42***
21	H	H	H	OH	H	OMe	OMe	CO_2Me	O	O	—	—	H	4.82
22	***H***	***Ph***		***H***	***H***	***OMe***	***OMe***	***CO_2Me***	***O***	***O***	—	—	***H***	***5.32***
23	H	Cl	Cl	H	H	OMe	OMe	CO_2Me	O	O	—	—	H	4.88
24	Cl	H	H	Cl	H	OMe	OMe	CO_2Me	O	O	—	—	H	5.10
25	H	H	*t*-Bu	H	H	OMe	OMe	CO_2Me	O	O	—	—	H	5.37
26	H	H	H	H	H	OMe	OMe	CO_2Me	O	O	—	—	H	5.53
27	H	H	H	H	H	OMe	OEt	CO_2Me	O	O	—	—	H	6.15
28	H	H	H	H	H	OMe	O-Cyclopentyl	CO_2Me	O	O	—	—	H	5.79
29	H	H	H	H	H	OMe	O-Norbornyl	CO_2Me	O	O	—	—	H	5.61
30	H	H	H	H	H	OMe	OMe	$CONH_2$	O	O	—	—	H	4.88
31	H	H	H	H	H	OMe	OEt	$CONH_2$	O	O	—	—	H	5.56
32	H	H	H	H	H	OMe	O-Cyclopentyl	$CONH_2$	O	O	—	—	H	5.60
33	***H***	***H***	***H***	***H***	***H***	***OMe***	***OMe***	***CN***	***O***	***O***	—	—	***H***	***5.76***
34	H	H	H	H	H	OMe	O-Cyclopentyl	CN	O	O	—	—	H	5.79
35	F	F	F	F	H	OMe	OEt	CO_2Me	O	O	—	—	H	6.42
36	H	H	H	NH_2	—	—	—	—	O	O	O	O	Me	7.35
37	H	H	H	NH_2	—	—	—	—	O	O	O	O	H	7.88
38	H	H	H	NH_2	—	—	—	—	O	O	O	O	F	6.63
39	***H***	***H***	***H***	***H***	—	—	—	—	***O***	***O***	***S***	***S***	***H***	***4.69***
40	***H***	***H***	***H***	***H***	—	—	—	—	***S***	***O***	***O***	***S***	***H***	***5.22***
41	***H***	***H***	***H***	***H***	—	—	—	—	***S***	***O***	***S***	***S***	***H***	***5.09***
42	—	—	—	—	—	—	—	—	—	—	—	—	—	***5.00***
43	—	—	—	—	—	—	—	—	—	—	—	—	—	***4.82***

[a] Data obtained from Refs. 24–28.

Table 2. Different cutoff combinations for the steric and electrostatic fields applied in the development of the 3D-QSAR models for phthalimide-containing TNFα modulators

Steric (kcal mol^{-1})	Electrostatic (kcal mol^{-1})
25.00	25.00
30.00	30.00
30.00	35.00
35.00	30.00
25.00	30.00
30.00	25.00
25.00	35.00
35.00	25.00
35.00	35.00

Table 3. Statistical results for the best CoMFA models obtained for phthalimide-containing TNFα modulators

	Steric/electrostatic	Steric	Electrostatic
q^{2}[a]	0.869	0.830	0.535
N[b]	6	6	6
S_{cv}[c]	0.422	0.492	0.535
r^{2}[d]	0.996	0.986	0.986
SEE[e]	0.078	0.152	0.142
F[f]	952.856	261.169	301.638

[a] Cross-validation correlation coefficient.
[b] Number of components.
[c] Standard error of prediction.
[d] Correlation coefficient.
[e] Standard error of estimation.
[f] F-ratio.

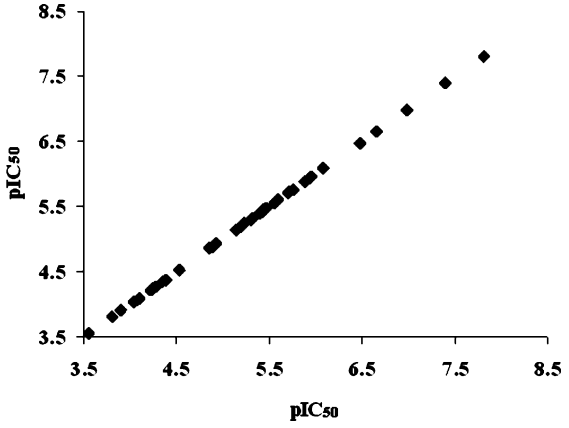


Figure 2. Distribution of pIC_{50} values of the phthalimide compounds elected for this 3D-QSAR study.

compound **2**, the two carbonyl groups of the glutarimide ring fall in this region. Since an increase in negative charge in this region will enhance activity, this observation supports the assumption of a better interaction of this compound with an electron-deficient counterpart in the target bioreceptor.

One of the limitations of CoMFA methodology concerns its insensitivity to molecular subunits that have similar substituents. It results in the lack of contour maps in the areas occupied by these groups. However, the absence of contour maps does not necessarily mean that the occupation of that region in space will not result in steric or electrostatic terms significantly correlated to activity. Besides, in this case study, the presence of substituents or new frameworks replacing the original glutarimide ring is more important because it has been reported as a toxicoforic subunit of thalidomide and some analogues.[21]

It is worth noting that our best CoMFA model was successful in discriminating the carbonyl groups and their importance for the modulatory activity on TNFα, as can be seen in Figure 4C, which depicts the CoMFA plots of compound **37** (CC-4047, $pIC_{50} = 7.88$), the most active compound in this study.

It has been experimentally shown that the withdrawal of one carbonyl group from the phthalimide ring (e.g., in compound **2**) does not cause a high decrease in the ability to modulate the TNFα action, and a small red contour map can be observed around the carbonyl group of the phthalimide moiety of these derivatives, which is indicative of the importance of at least one electron-rich group in this region for maximal activity. Interestingly, in the whole set there is only one compound, that is, phthalimide derivative (**2**), that does not bear two carbonyl groups, but it was successfully evaluated by CoMFA, in opposition to the above observation about its insensitivity to similar functionalities.

Figure 5 shows the steric maps contribution of the best 3D-QSAR model derived by steric/electrostatic fields combination in CoMFA. Green contours indicate regions where an increase in steric bulk will enhance activity, and yellow contours indicate regions where an increase in steric bulk will reduce activity. The steric contour maps are localized around the moiety attached to the phthalimide nitrogen atom, suggesting their importance for eventual steric interactions with the putative receptor, being more significative around the phenyl ring of some derivatives of the database, that is, compounds **3**–**35** (Table 1). As it was done in CoMFA studies of the electrostatic contour maps, compounds **3** and **2** were superimposed on both maps (Fig. 5) in order to try to establish SAR and understand the different activities of the derivatives.

The analysis of the steric contour maps around compound **1** permitted us to visualize a sterically unfavored region around the non-substituted phenyl ring, which should contribute detrimentally for the interaction with the target bioreceptor. Since compound **2** does not bear this structural feature, we may consider this aspect as a particular feature of the functionalized *N*-benzyl-phthalimide derivatives and take another analogue with reasonable activity, like compound **32** ($pIC_{50} = 5.60$), as a reference compound for comparison with the least active compound **3** (Fig. 5).

We may infer that the cyclopentyl group of compound **32** lies in a green region of the steric contour map, indicating that some steric bulk interaction is necessary for a good bioactivity in this series of derivatives. This profile

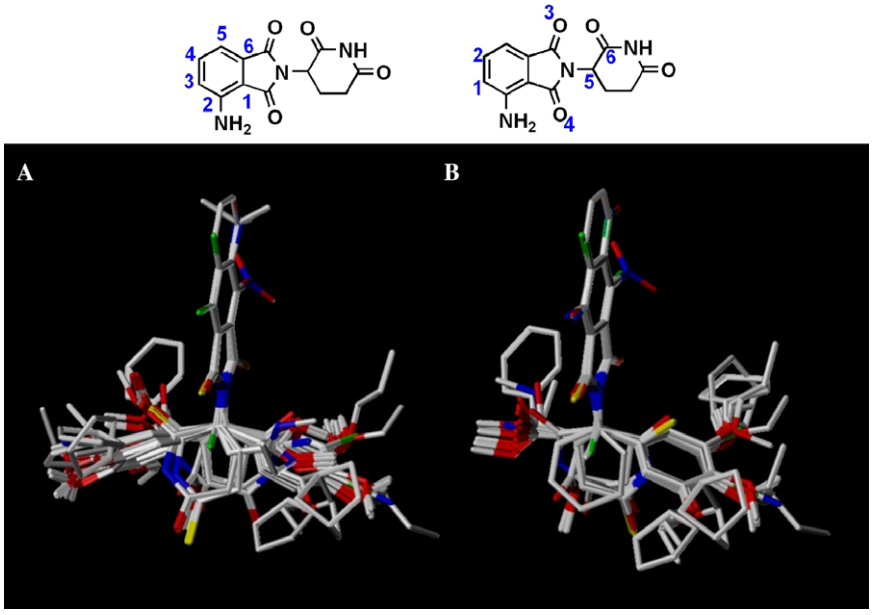


Figure 3. Atom definition and superimposition of the phthalimide derivatives (**2**–**43**) following the alignments 1 (A) or 2 (B) selected for this study. CC-4047 (**37**) was used as the reference compound for atom numbering.

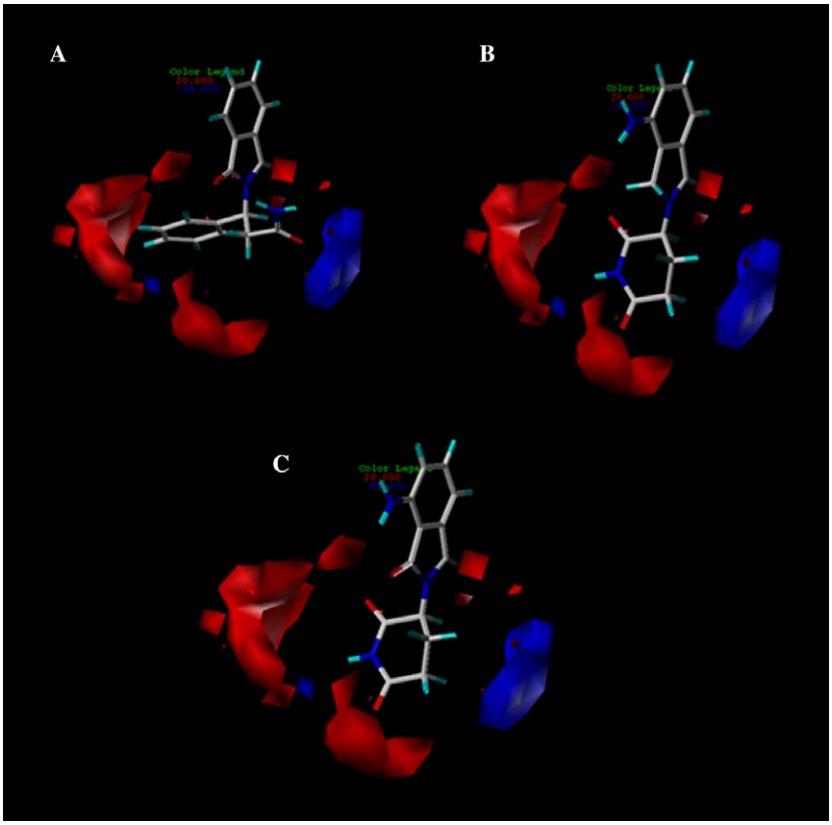


Figure 4. Electrostatic contour map from the best CoMFA model. Compound **3** (A), lenalidomide (**2**) (B), and CC-4047 (**37**) (C) are shown inside fields. Electrostatic contour plots: blue contours (80% contribution) and red contours (20% contribution) correspond to regions where an increase in positive or negative charge, respectively, is favorable for binding properties.

helps us to understand why the non-substituted compound **3** has a lower activity ($pIC_{50} = 3.58$) in comparison to compounds **30** ($pIC_{50} = 4.88$), **31** ($pIC_{50} = 5.56$) and **32** ($pIC_{50} = 5.60$), that bear a methyl, an ethyl, and a cyclopentyl group, respectively. Finally, these data are indicative of the specific steric requirements of the binding pocket of the hypothetical target bioreceptor.

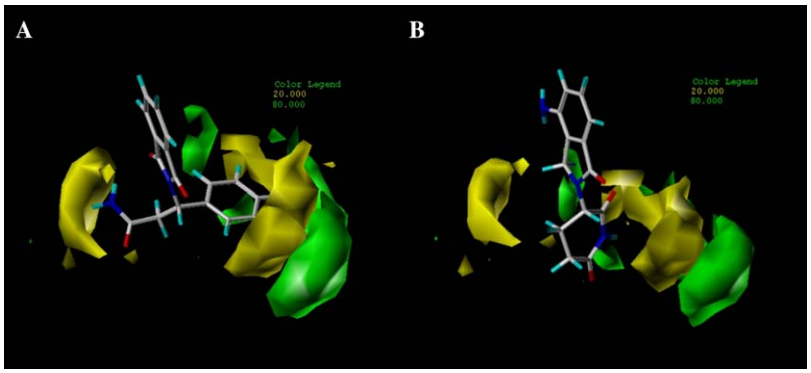


Figure 5. Steric contour map from the best CoMFA model. Compound **3** (A) and lenalidomide (**2**) (B) are shown inside fields. Steric contour plots: green contours (80% contribution) indicate regions where an increase in steric bulk will enhance affinity, whereas yellow contours (20% contribution) indicate sterically disfavored regions.

2.3. CoMSIA studies

As outlined in the methods section, CoMSIA results have been obtained applying the most successful alignment from CoMFA studies, that is, alignment 2 (Fig. 2). The results obtained from the CoMSIA studies are summarized in Table 4.

The analysis of the resulting 3D-QSAR models shows that the best CoMSIA model was obtained with combined steric/electrostatic fields, as it was previously observed during CoMFA studies. This model yields a cross-validated q^2 of 0.868, a non-cross-validated r^2 of 0.983, and an estimated F value of 234.426; low standard errors of prediction (S_{cv}) and of estimation (SEE), 0.417 and 0.157, respectively, were obtained. These statistical indexes suggest that this model has a good predictive ability. For the hydrogen bonding donor/acceptor field, the q^2 value was 0.665, while the hydrogen bonding donor and hydrogen bonding acceptor fields gave, separately, q^2 values of 0.629 and 0.545 (Table 4) and high standard errors of prediction ($S_{cv} > 0.6$), showing that these fields do not discriminate adequately the structural features associated to the molecular activities of these phthalimide-containing derivatives.

Finally, the analysis of the hydrophobic field shows a cross-validated q^2 of 0.826, a non-cross-validated r^2 of 0.976, an estimated F value of 165.431, and low standard errors of prediction (S_{cv}) and estimation (SEE), 0.487 and 0.187, respectively. These statistical indexes suggest that this model has a reasonable predictive ability and seems to be able to discriminate among the different hydrophobic contributions of the fragments present in the compounds of the database exploited in this study.

2.4. Visual inspection of 3D contour maps generated by CoMSIA

Figure 6 shows the electrostatic maps contribution to the best CoMSIA model. In accordance with CoMFA model, the electrostatic contour maps are localized around the adjacent subunit attached to the phthalimide ring of both compounds, confirming its importance for the electrostatic interaction involved in the molecular recognition by the hypothetical bioreceptor.

Figure 7 shows the steric maps contribution to the best model derived by steric/electrostatic field combination in CoMSIA. The steric contour maps are localized around the structural subunit attached to the phthalimide ring of both compounds **3** and **2**. However, the yellow contour, that accounts for a region of the receptor where steric bulk contribution is not desired, is localized very close to the peripheral aromatic ring. This result suggests that CoMSIA is putting more emphasis on the vol-

Table 4. Statistical results for the best CoMSIA models obtained for phthalimide-containing TNFα modulators

	Steric/electrostatic	Steric	Electrostatic	Donor/acceptor	Donor	Acceptor	Hydrophobic
q^{2a}	0.868	0.783	0.808	0.665	0.629	0.545	0.826
N^b	6	6	6	6	6	6	6
$S_{cv}{}^c$	0.417	0.533	0.519	0.663	0.687	0.934	0.487
r^{2d}	0.983	0.937	0.970	0.841	0.800	0.822	0.976
SEE[e]	0.157	0.304	0.210	0.485	0.544	0.513	0.187
F^f	234.426	59.596	130.013	21.105	16.634	18.424	165.431

[a] Cross-validation correlation coefficient.
[b] Number of components.
[c] Standard error of prediction.
[d] Correlation coefficient.
[e] Standard error of estimation.
[f] F-ratio.

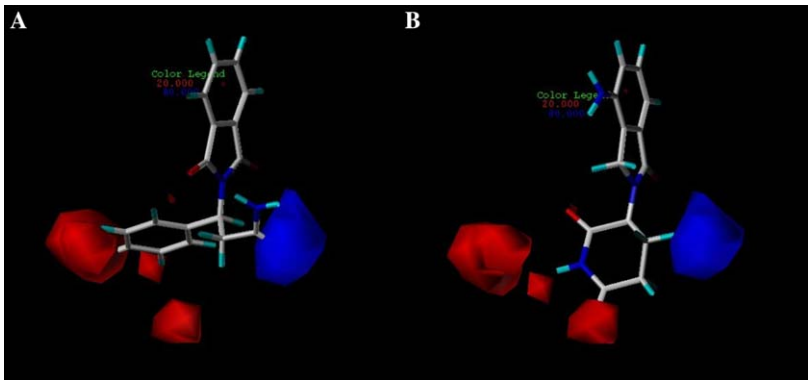


Figure 6. Electrostatic contour map from the best CoMSIA model. Compound **3** (A) and lenalidomide (**2**) (B), the least and the most active derivatives, respectively, are shown inside fields. Blue (80% contribution) and red (20% contribution) contours encompass areas where an increase in positive or negative charge, respectively, is favorable for binding properties.

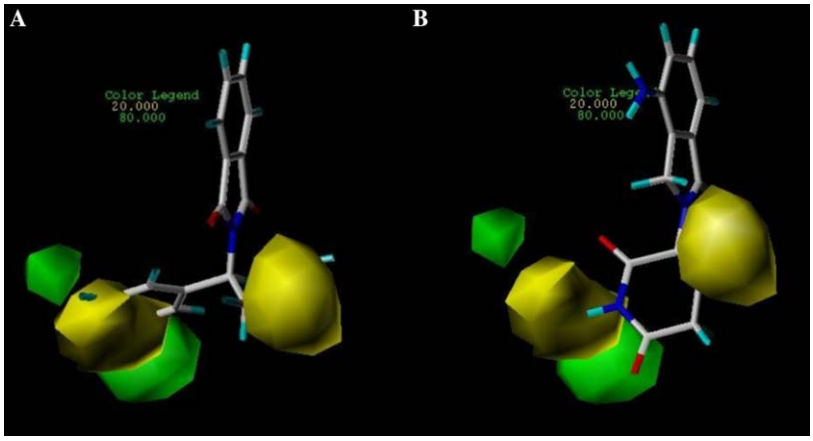


Figure 7. Contour plots of steric fields from the best CoMSIA model. Compound **3** (A) and lenalidomide (**2**) (B), the least and one of the most active derivatives, respectively, are shown inside fields. Regions where increasing the molecular volume increases bioactivity are in green (80% contribution) and regions where increasing the molecular volume decreases the activity are in yellow (20% contribution).

ume occupied by the aromatic carbons of the phenyl group. The same result was observed for the green contour maps, which account for sterically allowed regions of the receptor.

Figure 8 shows the hydrophobic maps contribution to the best model derived by CoMSIA. Yellow regions indicate areas where hydrophobic groups increase activity and white regions indicate areas where hydrophobic groups decrease activity.

The analysis of the hydrophobic contour maps around compound **3** led us to identify a white region around the phenyl ring, which seems to be indicative that the receptor demands the presence of a hydrophilic group for maximal interaction. This assumption could be confirmed when we analyzed the hydrophobic contour map around compound **2**, one of the most active derivatives of this study, since it had the hydrophilic glutarimide ring occupying that region.

Closer inspection of the hydrophobic contour maps around compound **3** showed that its polar amide group lies in a yellow contour region, that represents a binding cavity of the hypothetical receptor that demands a hydrophobic group for better interaction. This observation justifies the lower activity of compound **3**.

Comparisons between the obtained models derived from CoMFA and CoMSIA methods were made to assess their predictive abilities, to gain insight into the observed variance in the bioactivity profile and to suggest structural modifications to improve their modulatory action on TNFα. Therefore, we performed internal and external validation of the best models generated with CoMFA and CoMSIA methods, which consists of the calculations of the standard deviations of the residuals of fit (SD_{res}, data not shown) for the training and test sets and comparison of the residuals (*s*) obtained (Tables 5 and 6).

The analysis of the values of *s* obtained for the combined steric/electrostatic fields model, showed that best statistical indexes in CoMFA methodology were achieved. The value of SD_{res} obtained in the analysis of the predicted activities of the training set was 0.07, which may be considered as very low. The only

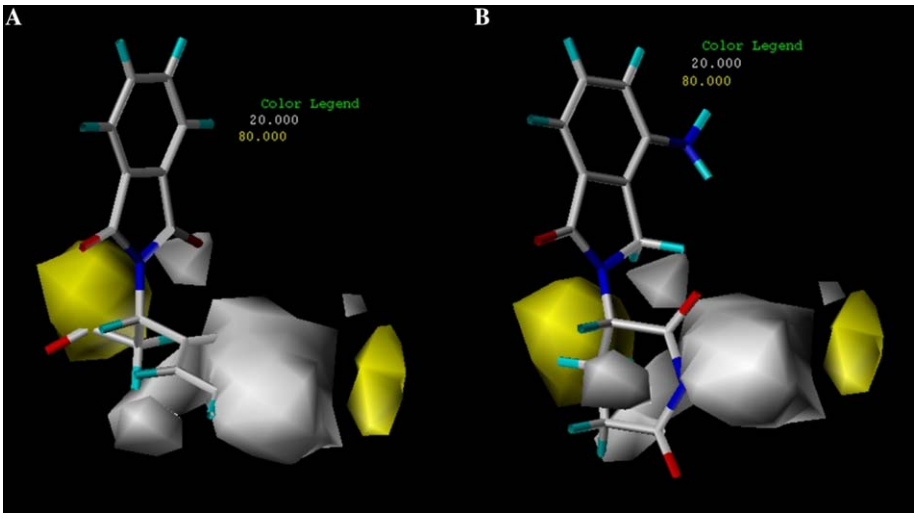


Figure 8. CoMSIA hydrophobic contour maps. Compound **3** (A) and lenalidomide (**2**) (B) are shown inside fields. Yellow regions (80% contribution) indicate areas where hydrophobic groups increase activity and white regions (20% contribution) indicate areas where hydrophobic groups decrease activity.

Table 5. Experimental (pIC_{50}) versus CoMFA- and CoMSIA-predicted activities (PA), with residuals (*s*), for the phthalimide derivatives of the training set

Compound	pIC_{50}	CoMFA		CoMSIA			
		PA[a]	*s*	PA[a]	*s*	PA[b]	*s*
2	7.00	6.98	0.02	6.99	−0.01	7.12	0.12
3	3.58	3.56	−0.02	3.52	−0.06	3.68	0.1
4	4.05	4.10	0.05	4.14	0.09	3.72	−0.33
5	3.82	3.81	−0.01	3.76	−0.06	3.75	−0.07
6	3.92	3.90	−0.02	4.06	0.14	4.16	0.24
7	4.2	4.24	0.04	4.29	0.09	4.31	0.11
8	4.92	4.86	−0.06	4.91	−0.01	4.8	−0.12
9	5.25	5.32	0.07	5.18	−0.07	5.2	−0.05
10	4.25	4.35	0.10	4.55	0.30	4.46	0.21
11	4.15	4.04	−0.11	4.04	−0.11	4.07	−0.08
12	4.29	4.23	−0.06	4.18	−0.11	4.39	0.1
13	4.4	4.38	−0.02	4.37	−0.03	4.30	−0.1
14	4.92	4.89	−0.03	4.8	−0.12	5.02	0.1
15	4.27	4.27	0	4.42	0.15	4.28	0.01
16	4.07	4.07	0	3.89	−0.18	4.06	−0.01
21	4.82	4.93	0.11	4.95	0.13	5.19	0.37
23	4.88	4.89	0.01	4.83	−0.05	4.82	−0.06
24	5.1	5.14	0.04	5.28	0.18	5.14	0.04
25	5.37	5.40	0.03	5.44	0.07	5.42	0.05
26	5.53	5.43	−0.1	5.45	−0.08	5.42	−0.11
27	6.15	5.96	−0.19	5.86	−0.29	5.85	−0.3
28	5.79	5.88	0.09	5.86	0.07	5.70	−0.09
29	5.61	5.56	−0.05	5.47	−0.14	5.54	−0.07
30	4.88	4.86	−0.02	4.91	0.03	4.80	−0.08
31	5.56	5.47	−0.09	5.32	−0.24	5.18	−0.38
32	5.6	5.71	0.11	5.78	0.18	5.88	0.28
34	5.79	5.72	−0.07	5.61	−0.18	5.76	−0.03
35	6.42	6.48	0.06	6.54	0.12	6.34	−0.08
36	7.35	7.40	0.05	7.49	0.14	7.32	−0.03
37	7.88	7.82	−0.06	7.69	−0.19	7.75	−0.13
38	6.63	6.66	0.03	6.71	0.08	6.77	0.14

[a] Steric/electrostatic fields.
[b] Hydrophobic field.

compound that exceeded two times this value was compound **27** (Table 5) but, still, it seems to be nonsignificant.

Table 6. Experimental (pIC_{50}) versus CoMFA- and CoMSIA-predicted activities (PA), with residuals (*s*), for the phthalimide derivatives of the test set

Compound	pIC_{50}	CoMFA		CoMSIA			
		PA[a]	*s*	PA[a]	*s*	PA[b]	*s*
17	4.22	5.39	1.17	5.92	1.70	5.95	1.73
18	5.12	5.19	0.07	6.12	1.00	4.87	−0.25
19	4.46	4.52	0.06	8.57	4.11	5.70	1.24
20	6.42	5.58	−0.84	5.63	−0.79	5.71	−0.71
22	5.32	6.09	0.77	6.68	1.36	6.25	0.93
33	5.76	5.60	−0.16	5.67	−0.09	6.26	0.50
39	4.69	5.44	0.75	7.35	2.66	5.49	0.80
40	5.22	5.24	0.02	7.55	2.33	7.67	2.45
41	5.09	5.30	0.21	7.38	2.29	7.13	2.04
42	5.00	5.94	0.94	6.12	1.12	5.77	0.77
43	4.82	5.76	0.94	5.96	1.14	5.67	0.85

[a] Steric/electrostatic fields.
[b] Hydrophobic field.

When we analyze the predictivity of the steric/electrostatic CoMFA model for the test set (external validation), we get a value of SD_{res} of 0.61. The observation of the residuals of fit for the test set (Table 6) and the plot of predicted versus experimental pIC_{50} values derived from the CoMFA model for the training and test sets (Fig. 9) shows that all the 11 compounds were very well predicted by the model.

In accordance to CoMFA results, the combined steric/electrostatic fields model obtained with CoMSIA showed the best statistical indexes. The value of SD_{res} obtained in the analysis of the predicted activities of the training set was 0.14 and two compounds exceeded two times this value: **27** and **31**. These data suggest that there are limitations in relation to the orientation of atoms that have not been assigned in the alignment,[20] since both compounds bear ethoxy groups in the phenyl ring.

The results of the CoMSIA model for the test set (external validation) showed a value of SD_{res} of 1.33 but the

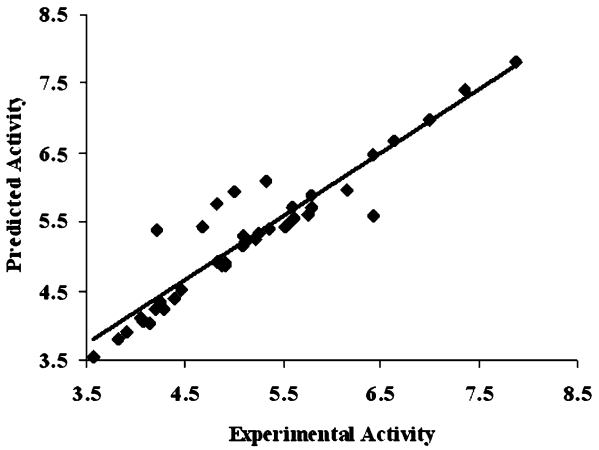


Figure 9. Predicted versus experimental pIC_{50} values derived from the steric/electrostatic CoMFA model of the training and test sets of phthalimide-containing TNFα modulators.

values of *s* (Table 6) reveal that just two compounds, that is, **19** and **39** had outlier behavior (*s* values of 4.11 and 2.66, respectively), were not well predicted by the model. The plot of predicted versus experimental pIC_{50} values derived from the CoMSIA model for the training and test sets (Fig. 10) shows the overall deviation of the model.

The outlier behavior of compound **19** may be explained by the fact that there are only a few molecules bearing substituents in position 5 of the phthalimide ring in the training set. Also, compound **39** bears thiocarbonyl groups, which are not present in the molecules of the training set. This fact is not relevant for the application of the model because of its ability to evaluate the role of structural modifications in the toxicoforic glutarimide ring,[21] very important for the design of new drug candidates, as already mentioned.

The hydrophobic field model showed the best statistical indexes in CoMSIA methodology (Table 4). The value of SD_{res} obtained in the analysis of the predicted activities of the training set was 0.17. Only compounds **21** and **31** exceeded two times this value.

Taking into consideration the method by which CoMSIA hydrophobic fields are based on,[22] we can suggest an explanation for the outlier behavior of compounds **21** and **31**. For example, the method calculates the hydrophobic fields of molecules based on the atomic values for physicochemical properties, such as the octanol–water partition coefficient and atomic refractivity.

The main reason in using molar refractivity parameters in the calculation of hydrophobic fields is the possibility of representing the dispersive interaction between the ligand and the receptor as a linear function of the property of the ligand alone, as long as attention is confined to a small local region of the ligand. However, although this work has been based on the atomic contributions to hydrophobicity taking into consideration the hybridization state of the bonded atoms based on a large data set, there are always limitations due to variations in the atomic environment.

Concerning the outlier behavior of compound **21**, that has a hydroxyl group attached in the phthalimide ring, it is possible that, in function of the fact that this pattern of substitution in the heterocyclic ring is not frequent in the data set, it was difficult to be discriminated and correlated to biological activity. However, as to compound **31**, the fragmentation method used to calculate the hydrophobic field does not discriminate between alkyl chain sizes, which may be the reason for underestimation of the activity of this compound.[23]

When we analyze the predictivity of the hydrophobic CoMSIA model for the test set (external validation), we get a value of SD_{res} of 0.93. From this result, two outliers, that is, compounds **40** and **41**, were identified. The plot of predicted versus experimental pIC_{50} values derived from the hydrophobic CoMSIA model for the training and test sets (Fig. 11) shows the overall deviation of the model.

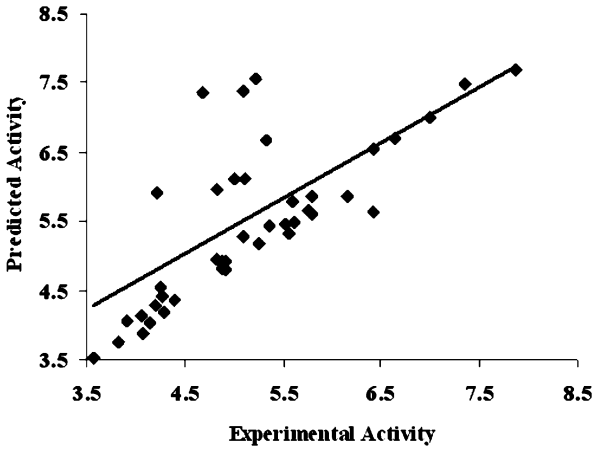


Figure 10. Predicted versus experimental pIC_{50} values derived from the steric/electrostatic CoMSIA model for the phthalimide-containing TNFα modulators of the training and test sets.

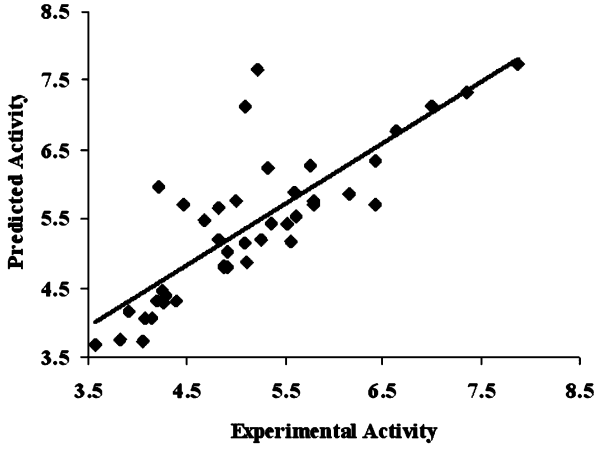


Figure 11. Predicted versus experimental pIC_{50} values derived from the hydrophobic CoMSIA model for the phthalimide-containing TNFα modulators of the training and test sets.

The outlier behavior of compounds **39** and **41** may be rationalized in terms of the presence of a thiocarbonyl group, as discussed earlier.

After analysis of the statistical parameters of the best 3D-QSAR models obtained with CoMFA and CoMSIA methods, and also the analysis of the outliers, we have come to the conclusion that both CoMFA and CoMSIA have shown to be predictive methods. Either CoMFA or CoMSIA steric/electrostatic models have shown a discriminatory character and provided a good understanding of the variation in the biological activity of the series of molecules of this study, being highly predictive for lenalidomide (**2**) ($pIC_{50(exp)} = 7.00$; $pIC_{50(pred)} = 6.98$) and CC-4047 (**37**) ($pIC_{50(exp)} = 7.88$; $pIC_{50(pred)} = 7.82$).

3. Conclusions

In this study, the 3D-QSAR methods, CoMFA and CoMSIA, were applied to predict the TNFα modulatory activity of a series of phthalimide-containing derivatives. The best 3D-QSAR models were generated by CoMFA combined steric/electrostatic fields, providing the most significant correlation with biological activity.

This model had good statistical results in terms of q^2 and r^2 values, and showed a great predictivity of the test set, in the external validation, showing no outliers. Additionally, although CoMSIA models have been less predictive, steric/electrostatic and hydrophobic field models have also shown correlation with the biological actions on TNFα.

In conclusion, the 3D-QSAR CoMFA and CoMSIA models for thalidomide analogue TNFα modulators gave insight into the influence of various structural attributes for the biological activity and may be considered as a powerful tool in designing and forecasting more efficacious analogues, since they point toward the molecular sites that may be explored, like the region represented by the substituted glutarimide ring, in order to maximize the bioprofile and prevent toxicity. Finally, the models developed herein are useful for the rational design of new bioactive prototypes for the treatment of chronic inflammatory diseases.

4. Methods

4.1. Data set

The 31 and 11 compounds of the training and the test sets (Table 1), respectively, were selected from the literature.[24–28] The anti-TNFα activity of all the 42 compounds used in this study was measured as IC_{50} (Table 1). These values have been obtained using the same pharmacological protocol on TNFα inhibitory activity in LPS-stimulated human peripheral blood mononuclear cells (PBMC) and were expressed in negative logarithmic units, $-\log IC_{50}$ or pIC_{50}. The elected inhibitors are phthalimide derivatives[24–28] showing good structural variation and pIC_{50} range (3.58–7.88, Fig. 2), making them suitable for 3D-QSAR studies.

4.2. Molecular modeling

Molecular modeling studies were performed using SYBYL software version 7.1 (Tripos Ltd, St. Louis, MI) running on a Silicon Graphics workstation Onyx3 VPro. Initially, the structures were built using PC *Spartan Pro* 1.0 for Windows (Wavefunction), and the conformer distribution of each one was calculated by molecular mechanics with the Merck molecular force field (MMFF).[22] Subsequently, the lowest-energy conformation found for each structure was submitted to optimization with the semiempirical AM1 method.[29] For further calculations with CoMFA/CoMSIA, Gasteiger–Hückel charges were also assigned using SYBYL software version 7.1.

4.3. Proposition of bioactive conformation

CC-4047 (**37**, Actimid™), the most active phthalimide-containing derivative in the series of compounds used in this study, has been used as template for proposition of the bioactive conformation. It is administered as a racemic mixture, since fast racemization of the single enantiomers occurs under physiological conditions due to the strongly acidic hydrogen atom,[30] but it has been demonstrated that the eutomeric form of **37** corresponds to the *S* enantiomer ($IC_{50} = 3.9$ nM), which is 23-fold more active than the corresponding *R*-enantiomer ($IC_{50} = 93$ nM) as TNFα modulator.[26] Therefore, although the majority of compounds used in this study have been synthesized and biologically evaluated as the corresponding racemic mixture,[24–28] the *S*-enantiomer of each compound was modeled where applicable in this study.[17]

4.4. Alignment of molecules

One of the most important adjustable parameters in CoMFA is the relative alignment of all the compounds to one another so that they have a comparable conformation and a similar orientation of pharmacophoric groups in space. Therefore, this is considered a critical step in CoMFA studies, since experience has shown that the resulting 3D-QSAR model is often sensitive to a particular alignment scheme.[31]

In this study, the most active molecule of the database (compound **37**) was used as a template for superimposition, assuming that its lowest-energy conformation represents the most probable bioactive conformation of the phthalimide-containing thalidomide analogues at the putative receptor. This assumption was made because the molecular mechanism of this class of compounds on the TNFα modulation is not completely elucidated yet. Two atom-based alignments were manually carried out using the Fit Atoms tool in SYBYL 7.1. (a) Alignment 1: All atoms of the phenyl ring of the phthalimide moiety were selected and are numbered 1 to 6 (Fig. 3);[17] (b) Alignment 2: 6 atoms were also selected for the alignment of all compounds. These atoms are numbered from 1 to 6 and involve not only atoms from the phthalimide

moiety, but also from the adjacent structural subunit (Fig. 3).

4.5. CoMFA studies

CoMFA was performed with the QSAR module of SYBYL 7.1 for each combination of steric and electrostatic molecular fields, which were sampled at each point of regularly spaced grids of 1.0, 1.5, and 2.0 Å. The steric and electrostatic fields were calculated using the default probe, a sp^3 carbon atom with a charge of +1. CoMFA calculates steric fields using a Lennard-Jones potential and electrostatic fields using a Coulomb potential.[15] Different cutoff combinations for the steric and electrostatic fields were tested (Table 2). Then, the energies of the steric and electrostatic fields were calculated separately for the best model obtained.

4.6. CoMSIA studies

Similar to the CoMFA approach, a data table has been constructed from similarity indices calculated via the default probe, which is placed at grid points of a regularly spaced lattice.[32] In the CoMSIA methodology, the alignment that generated the most predictive models in CoMFA methodology was used. Also, five physicochemical properties k (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were evaluated using a Gaussian function. The steric and electrostatic, hydrogen-bond donor and acceptor contributions were combined and also calculated separately. The attenuation factor α was set at the default, 0.3.

4.7. PLS analysis

The partial least-squares (PLS) analysis algorithm was used in conjugation with the cross-validation leave-one-out (LOO)[32] option to obtain an optimum number of components, which were used to generate the final CoMFA/CoMSIA models without cross-validation. To avoid overfitted 3D-QSAR, the optimum number of components (N) used in the model derivation was chosen from the analysis with the highest q^2 value. Finally, column filtering was set at 2.0 kcal/mol so that only those steric and electrostatic energies with values greater than 2.0 kcal/mol were considered in the PLS analysis. This procedure speeds up analysis and reduces noise. In the above stage, the predictive quality of the 'best' correlation model was determined.

4.8. Predictive ability of CoMFA and CoMSIA models

The predictive ability of CoMFA and CoMSIA models can be evaluated based on q^2, the cross-validated leave-one-out correlation coefficient, which quantifies the predictive ability of the model. Models are considered to be of good predictivity when q^2 is greater than 0.5. Moreover, the standard deviation of the residuals of fit (SD_{res}) was calculated. Compounds with residuals (s) greater than twice the standard deviation of the residuals have been considered as outliers, that is, compounds whose predicted activity was under- or overestimated by the model, considering the training set (internal validation) and the test set (external validation). This methodology evaluates the robustness of the generated 3D-QSAR models and also gives deep insight into the full SAR of the compounds of this study.

Acknowledgments

The authors thank CNPq (Br, #420.015/05-1), FAPERJ (Br), PRONEX-Rio (Br), and IM-INOFAR (Br) for the financial support and fellowships (to C.M.A., E.J.B., and C.A.M.F.).

References and notes

1. Corral, L. G.; Kaplan, G. *Ann. Rheum. Dis.* **1999**, *58*, 107.
2. Magna, A. M.; Takiya, C. M.; Aruda, L. B.; Pascarelli, B.; Gomes, R. N.; Faria, H. C.; Lima, L. M.; Barreiro, E. J. *Int. Immunopharmacol.* **2005**, *5*, 485.
3. Marriott, J. B.; Westby, M.; Cookson, S.; Guckian, M.; Goodbourn, S.; Muller, G.; Shire, M. G.; Stirling, D.; Dalgleish, A. G. *J. Immunol.* **1998**, *161*, 4236.
4. Sampaio, E. P.; Sarno, E. N.; Galilly, R.; Cohn, Z. A.; Kaplan, G. *J. Exp. Med.* **1991**, *173*, 699.
5. Moreira, A. L.; Sampaio, E. P.; Zmuidizinas, A.; Galilly, R.; Smith, K. A.; Kaplan, G. *J. Exp. Med.* **1993**, *177*, 1675.
6. Lima, L. M.; Fraga, C. A. M.; Koatz, V. L. G.; Barreiro, E. J. *Anti-Inflammatory Anti-Allergy Agents Med. Chem.* **2006**, *5*, 79.
7. Corral, L.; Haslett, P.; Muller, G.; Chen, R.; Wong, L.; Ocampo, C. J.; Patterson, R. T.; Stirling, D. I.; Kaplan, G. *J. Immunol.* **1999**, *163*, 380.
8. Sampaio, E. P.; Kaplan, G.; Miranda, A.; Nery, J. A.; Miguel, C. P.; Viana, S. M.; Sarno, E. N. *J. Infect. Dis.* **1993**, *168*, 408.
9. Schuler, U.; Ehninger, G. *Drug Safety* **1995**, *12*, 364.
10. Youle, M.; Clarbour, J.; Furthing, C.; Connolly, M.; Hawkins, D.; Staughton, R.; Gazzard, B. *Br. Med. J.* **1989**, *298*, 432.
11. Klausner, J. D.; Makonkawkeyoon, S.; Akarasewi, P.; Nakata, K.; Kasinrerk, W.; Corral, L.; Dewar, R. L.; Lane, H. C.; Freedman, V. H.; Kaplan, G. *J. Acquir. Immune Defic. Syndr.* **1996**, *11*, 247.
12. Vogelsang, G. B.; Farmer, E. R.; Hess, A. D.; Altamonte, V.; Beschorner, W. E.; Jabs, D. A.; Corio, R. L.; Levin, L. S.; Colvin, O. M.; Wingard, J. R.; Santos, G. W. *N. Eng. J. Med.* **1992**, *326*, 1055.
13. Pagnini, G.; Dicarlo, R. *Boll. Soc. Ital. Biol. Sper.* **1963**, *39*, 1360.
14. Teo, S. K. *The AAPS J.* **2005**, *7*, E14.
15. Cramer, R.; Patterson, D.; Bunce, J. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
16. Klebe, G. *Perspect. Drug Discovery* **1998**, *12*, 87.
17. Lepper, E. R.; Ng, S. S. W.; Gütschow, M.; Weiss, M.; Hauschildt, S.; Hecker, T. K.; Luzzio, F. A.; Eger, K.; Figg, W. D. *J. Med. Chem.* **2004**, *47*, 2219.
18. Lima, L. M.; Castro, P.; Machado, A. L.; Fraga, C. A. M.; Lugnier, C.; Koatz, V. L. G.; Barreiro, E. J. *Bioorg. Med. Chem.* **2002**, *10*, 3067.
19. Machado, A. L.; Lima, L. M.; Araújo, J. X., Jr.; Fraga, C. A. M.; Koatz, V. L. G.; Barreiro, E. J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1169.
20. Cho, S. J.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 1060.
21. Dewar, M. J. S.; Huang, P. H.; McBride, W. *Teratog. Carcinog. Mutagen.* **1997**, *17*, 1.
22. Halgren, T. A. *J. Am. Chem. Soc.* **1990**, *112*, 4710.

23. Viswanadhan, V.; Chose, A. K.; Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.
24. Muller, G.; Corral, L.; Shire, M. G.; Chen, R.; Wang, H.; Moreira, A.; Kaplan, G.; Stirling, D. I. *J. Med. Chem.* **1996**, *39*, 3238.
25. Muller, G.; Shire, M. G.; Wong, L. M.; Corral, L.; Patterson, R. T.; Chen, Y.; Stirling, D. I. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2669.
26. Muller, G.; Chen, R.; Huang, S.; Corral, L.; Wong, L. M.; Patterson, R. T.; Chen, Y.; Kaplan, G.; Stirling, D. I. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1625.
27. Man, H.; Corral, L.; Stirling, D. I.; Muller, G. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3415.
28. Zhu, X.; Giordano, T.; Yu, Q.; Holloway, H. W.; Perry, T. A.; Lahiri, D. K.; Brossi, A.; Greig, N. H. *J. Med. Chem.* **2003**, *46*, 5222.
29. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
30. Knoche, B.; Blaschke, G. *J. Chromatogr., A.* **1994**, *666*, 235.
31. Zou, X.; Lai, L.; Jin, G.; Zhang, Z. *J. Agric. Food Chem.* **2002**, *50*, 3757.
32. Wold, S. *Quant. Struct.-Act. Relat.* **1991**, *10*, 191.