Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis of Thalidomide Analogues as Angiogenesis Inhibitors

Erin R. Lepper,[†] Sylvia S. W. Ng,[†] Michael Gütschow,[‡] Michael Weiss,[§] Sunna Hauschildt,[||] Thomas K. Hecker,[⊥] Frederick A. Luzzio,[⊗] Kurt Eger,[⊥] and William D. Figg*,[†]

Molecular Pharmacology Section, Cancer Therapeutics Branch and Clinical Pharmacology Research Core, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Pharmaceutical Institute, Poppelsdorf, University of Bonn, Bonn, Germany, Center for Cancer Research, Children's Research Institute, and Department of Pediatrics and Biochemistry/Molecular Biology, George Washington University School of Medicine, Washington, DC 22037, Institute of Pharmacy, Pharmaceutical Chemistry and Institute of Zoology, Department of Immunobiology, University of Leipzig, Germany, and Department of Chemistry, University of Louisville, Louisville, Kentucky 40292

Thalidomide, 2-(2,6-dioxo-3-piperidinyl)-1*H*-isoindole-1,3(2*H*)-dione, has been shown to inhibit angiogenesis, the formation of new blood vessels from existing vasculature. As a result, there is renewed interest in this drug as a potential therapy for solid tumors. Thalidomide forms a number of metabolites and has been shown to require metabolic activation for antiangiogenic activity. A series of 39 compounds, based upon the structure of some of these metabolites, was synthesized and tested for their ability to inhibit microvessel growth in the rat aortic ring assay. The results of this testing have been used as the basis for a three-dimensional quantitative structure−activity relationship (3D-QSAR) study, utilizing comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) procedures. The best resulting CoMFA and CoMSIA models have conventional r^2 values of 0.924 and 0.996, respectively. The cross-validated q^2 values are 0.666 and 0.635, respectively. These models offer insight into the structural requirements for activity of thalidomide analogues as angiogenesis inhibitors, since there is only speculative knowledge of the target. Additionally, it appears as though there is more than one active site or mechanism of action.

Introduction

Thalidomide, 2-(2,6-dioxo-3-piperidinyl)-1*H*-isoindole-1,3(2*H*)-dione, **1**, was originally marketed as a sedative during the 1950s. Several years later, thalidomide was seen to cause severe congenital malformations in children whose mothers had taken the drug while pregnant, and it was removed from sale, due to its teratogenic effects.[1] Thalidomide has since seen renewed interest, due to its ability to inhibit angiogenesis, the formation of new vasculature from preexisting blood vessels. This is particularly relevant to solid tumors, since a tumor cannot grow beyond several millimeters without a network of microvessels to supply oxygen and nutrients.[1] Inhibition of angiogenesis has been shown to suppress tumor growth,[2] so design of small molecule inhibitors is currently of great interest.

Thalidomide and several closely related analogues are currently being tested for a variety of cancers in clinical trials.[3,4] It has been shown to be active in multiple myeloma[5] and also shows promise in Kaposi's sarcoma,[6] glioblastoma multiforme,[7] renal cell carcinoma,[8] and prostate cancer.[9] Thalidomide is a prodrug, requiring metabolic activation to exert an antiangiogenic effect.[10] Thalidomide produces several metabolized species as a result of cytochrome P450 biotransformation (specifically, by isoenzyme CYP2C19), including 5′-hydroxythalidomide,[11] which has also been identified in patients taking thalidomide.[12] Additionally, the drug spontaneously hydrolyzes at physiological pH to form a cascade of hydrolysis products.[13]

One of the major side effects of thalidomide in cancer treatment is sedation, the original indication. Though there is fast chiral interconversion of thalidomide,[14] the (*R*)-enantiomer has been shown to be responsible for these sedative effects.[15] Conversely, the (*S*)-enantiomer is a more active inhibitor of angiogenesis.[16]

A series of analogues was synthesized to help elucidate the functional groups required for inhibition of angiogenesis without requiring metabolic activation. Thirty-nine of these compounds were tested in the rat aortic ring assay. We have performed a three-dimensional quantitative structure−activity relationship (3D-QSAR) study, using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Since the target and mode of action of thalidomide and its analogues is still unknown, the resulting CoMFA model will allow for design of more efficacious analogues, with fewer side effects.

Methods

Rat Aortic Ring Angiogenesis Assay. Twelve-well tissue culture-grade plates were covered with 250 μL Matrigel and allowed to gel for 30 to 45 min at 37 °C, 5% CO$_2$. Thoracic

* Corresponding author. Molecular Pharmacology Section, National Cancer Institute, Building 10, Rm 5A01, 9000 Rockville Pike, Bethesda, MD 20892. Tel: +1 301 402 3622. Fax: +1 301 402 8606. E-mail: wdfigg@helix.nih.gov.
† National Cancer Institute.
‡ University of Bonn.
§ George Washington University School of Medicine.
|| Institute of Pharmacy, Pharmaceutical Chemistry, University of Leipzig.
⊥ Institute of Zoology, Department of Immunobiology, University of Leipzig.
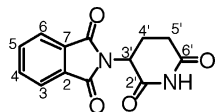⊗ University of Louisville.

Figure 1. Thalidomide (**1**).

aortas were excised from 6- to 8-week-old male Sprague–Dawley rats, and the fibroadipose tissue was removed. The aortas were cut into 1 mm-long cross-sections and placed on the Matrigel coated wells. They were then covered with an additional 250 μL Matrigel and allowed to gel for 30 to 45 min at 37 °C, 5% CO_2. The rings were cultured for 24 h in 1 mL EGM-II. EGM-II consists of endothelial cell basal medium (EBM-II; Clonetics, San Diego, CA) plus endothelial cell growth factors provided as the EGM-II Bulletkit (Clonetics). After 24 h, the medium was removed and replaced with 1 mL EBM-II, supplemented with fetal bovine serum (2%), 0.25 μg/mL amphotericin B, and 10 μg/mL gentamicin. Each compound was dissolved in DMSO (Sigma Aldrich, St. Louis, MO) and added to the EBM-II, before it was added to the well. Drug was administered daily for 4 days. Photos were taken on Day 5. One well was treated with vehicle (0.5% DMSO) only to act as a control and carboxyamidotriazole (NCI, Bethesda, MD) was used as a positive control. The vascular outgrowth was quantified using Adobe Photoshop (Adobe Systems, Inc., San Jose, CA). The percentage inhibition (growth in pixels of treated ring/growth in pixels of control ring) was used as the target property.

Molecular Modeling. All molecular modeling, including CoMFA studies, was performed on a Silicon Graphics workstation, using Sybyl 6.8 (Tripos, Ltd, St. Louis, MI). Structures were built using the Sketch Molecule function and minimized with Gasteiger-Hückel charges and the Tripos force field, using the MAXIMIN function, with the Powell method and an energy convergence cutoff of 0.05 kcal/mol.

Determination of Active Conformation. Thalidomide is administered as a racemic mixture, since fast racemization of the single enantiomers occurs under physiological conditions, due to the strongly acidic hydrogen atom.[17] With the exception of **2**(CPS3), which is a stable diastereomer, the analogues were also synthesized as racemic mixtures. Where applicable, the *S*-enantiomer of each compound was modeled, since the (*S*)-enantiomer of thalidomide has previously been shown to be more active in the inhibition of angiogenesis.[16] Since the target of thalidomide in this assay is unknown, there is no knowledge of the active conformation. Additionally, due to the wide range of structures it is unreasonable to assume that the minimum energy conformation of the most active is the true conformation of all molecules. Therefore, each compound was first minimized, as described above, then submitted to a GRID search of the rotatable bonds. The lowest energy structures were finally minimized, and the lowest energy conformation of each molecule was used in the model.

Alignment of Molecules. Two different alignments of the molecules were used to form the models. In Alignment 1, molecules were aligned by atoms 2–7 of the phthalimido moiety (see Figure 1). This alignment was used for Models A–D (see Tables 1 and 2). For Alignment 2, which was used for Models E and F, molecules with an intact glutarimide ring were aligned by atoms 1′–6′. Both alignments were manually carried out using the AtomFit tool in Sybyl 6.8. All applicable molecules were originally used in the alignment, based on the assumption that all were interacting with a possible receptor in the same manner.

CoMFA Analysis. The aligned molecules were originally placed in a three-dimensional grid with 2 Å spacing. Steric and electrostatic fields were calculated using the default probe, an sp^3 carbon atom with a charge of +1, at each lattice point. CoMFA standard scaling was used. CoMFA calculates steric fields using a Lennard–Jones potential and electrostatic fields using a Coulombic potential.[18] The steric and electrostatic cutoffs were both set to 30.0 kcal/mol and the dielectric was distance dependent. Electrostatics were dropped within the steric cutoff for each row.

CoMSIA Analysis. As for CoMFA, the aligned molecules were placed in a three-dimensional grid with 2 Å spacing. CoMSIA calculates steric and electrostatic fields, in addition to hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields, using a Gaussian function. The attenuation factor, α, was set as the default, 0.3.

PLS Analysis. An initial partial least-squares (PLS) analysis was carried out using the cross-validated Leave-One-Out option to obtain the optimal number of components to be used in the final analysis.[19] The SAMPLS (SAMple distance PLS)[20] program was used, rather than column filtering. After the optimal number of components was determined, a non-cross validated analysis was performed using column filtering of 2.0 kcal/mol.

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 r^2. This is calculated with the following equation:

$$q^2 = \frac{(\text{SSY} - \text{PRESS})}{\text{PRESS}}$$

SSY is the variance of the biological activity and PRESS is the prediction error sum-of-squares, calculated using the following equation:

$$\text{PRESS} = \frac{\Sigma \, (\text{predicted target property} - \text{observed target property})^2}{\Sigma \, (\text{observed target property} - \text{mean target property})^2}$$

Models are generally considered to be of good predictive value when q^2 is greater than 0.5.

Progressive Scrambling. Progressive scrambling allows for evaluation of the sensitivity of a QSAR model to chance correlations.[21] Thirty scramblings were carried out with a maximum of 10 bins and a minimum of two bins. The critical point was 0.85. Progressive scrambling of the biological data produces three statistics. Q^2 and cSDEP are the predictivity of the model and the calculated crossvalidated standard error, respectively. The susceptibility of the model to chance correlation can be gauged by the slope of q^2 (as originally determined using SAMPLS) with respect to the correlation of the original biological activity versus the scrambled biological activity. This is denoted as dq'^2/d$r^2_{yy'}$. In a model that is highly reliant on chance correlation, this value will be greater than one. In a model where the correlation of activity is not due to chance, the slope will be equal or less than one.

Results and Discussion

Thirty-nine compounds with diverse structures (see Figure 2) were tested in the rat aortic ring assay. Of these, nine compounds showed statistically significant inhibition of angiogenesis. Interestingly, some traditional structure–activity relationships can be gained from these results. None of the compounds without intact glutarimide rings exhibited statistically significant inhibition of angiogenesis. Three of these compounds, **3** (CPS29), **4** (CPS30), and **5** (CPS31) are hydrolysis products of thalidomide, all of which were shown to be inactive in this model. These results suggest that the glutarimide ring is required for inhibition of angiogenesis, without metabolic activation. The compounds that were formulated as salts, **6** (CPS16), **7** (CPS17), and **8** (CPS26), all showed good activity, based upon their inhibition of microvessel outgrowth in the rat aortic ring assay (see Figure 3). This could be due to improved solubility or additional interaction of the positively charged nitrogen with the active site. Compounds **5** and **6**, which both possess primary cationic amines, exhibited much greater activity than **7**, with a
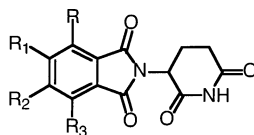
Compound	R	R1	R2	R3	% Inhibition	SEM
1	H	H	H	H		
18	H	OH	OH	H	38.75	13.28
22(CPS-2)	OH	OH	H	H	10.25	4.97
23(CPS-4)	H	OCH₃	OCH₃	H	23.25	7.609
24(CPS-32)	H	H	H	NH₂	5.50	11.81
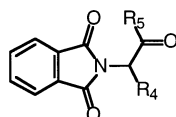
Compound	R4	R5	% Inhibition	SEM
25(CPS-6)	COOH	OH	-2.00	6.99
26(CPS-9)	COOC(CH₃)₃	NHCH(COOH)₂	13.25	6.75
27(CPS-27)		OH	11.67	11.89
28(CPS-28)	NHCH(COOH)CH₂CH₂COOH	OH	25.75	10.47
3	CH₂CH₂COOH	NH₂	15.25	6.87
4	CH₂CH₂CONH₂	OH	1.75	5.648
5	CH₂CH₂COOH	OH	10.75	10.57
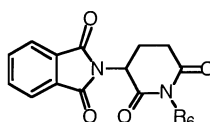
Compound	R₆	% Inhibition	SEM
19	CH₂OH	21.00	4.53
29(CPS13)	CH₂OCOCH(NH₂)Bzl	4.00	6.42
30(CPS14)		25.25	6.80
31(CPS15)		50.00	6.66
6	CH₂OCOCH(NH₃Cl)CH(CH₃)₂	88.00	2.00
7	CH₂OCOCH₂NH₃Cl	71.33	17.29
32(CPS19)	CH₂OCOCH(CH₃)N(CH₃)COOC(CH₃)₃	56.50	11.33
33(CPS20)	CH₂OCOCH(Bzl)NHCOOC(CH₃)₃	33.75	23.25
34(CPS21)	CH₂OCOCH₂NHCOOC(CH₃)₃	40.25	14.24
35(CPS22)	CH₂COOH	39.75	16.46
20	CH₂OCH₂CH₃	32.00	5.72
21	CH₂OCH₃	13.00	7.94
36(CPS25)	CH₂N₃	24.67	15.19
8		29.67	6.12
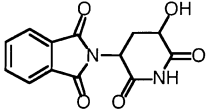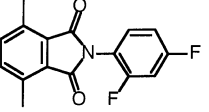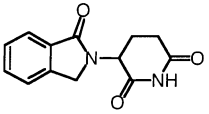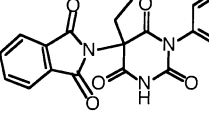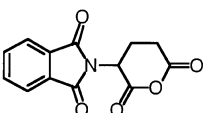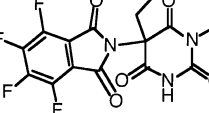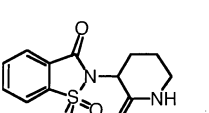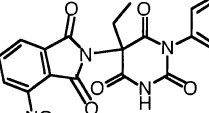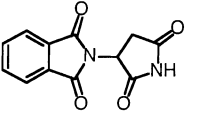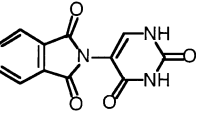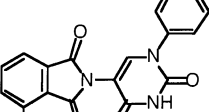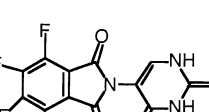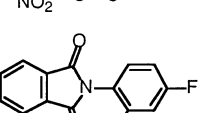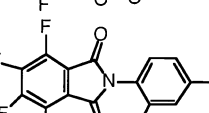
Figure 2. Structures of compounds tested and used in 3D-QSAR analysis
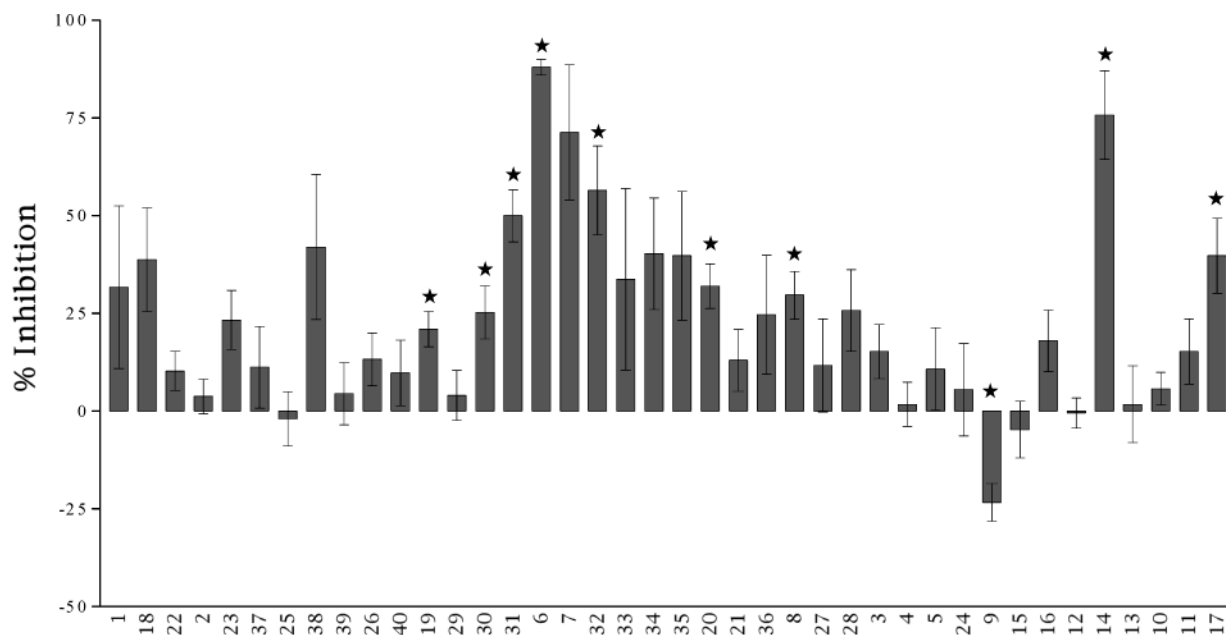
Figure 3. Inhibition of microvessel outgrowth in the rat aortic ring assay at 50 μM daily treatment. ★ denotes that inhibition is statistically significant, as compared to the control ($p < 0.05$). Error bars represent SEM.

cationic pyridine derivative. In compounds **9** (CPS40), **10** (CPS47), and **11** (CPS48), the glutarimide was replaced with a pyrimidine derivative, in an attempt to

improve the hydrolytic stability. The lack of inhibitory activity suggests that prevention of spontaneous hydrolysis is not sufficient for activity. The fluorinated

compounds that were tested exhibit an unusual pattern of activity. For example, **12** (CPS43) and **13** (CPS46) did not show any anti-angiogenic activity, but **14** (CPS45), the tetrafluorinated derivative, was very active, suggesting that tetrafluorination of the phthalimido ring promotes activity. However, when the activity of **15** (CPS41), and **16** (CPS42), **11**, and **17** (CPS49) is compared, tetrafluorination alone is seen not to be sufficient. For example, neither **15**, nor **16**, both of which contain phenyl groups in place of the glutarimide, shows significant activity. Compound **17**, the tetrafluorinated analogue of **15** exhibits excellent inhibitory action, but **11**, which is also tetrafluorinated, has poor inhibitory activity in the rat aortic ring assay. These results suggest that tetrafluorinated thalidomide analogues may be acting via a different active site or mechanism than the other analogues.

This inhibition data was also used as the basis for CoMFA and CoMSIA analysis. Two different alignments and several different subsets of structures were tested to determine the best predictive model.

Comparative Molecular Field Analysis. All compounds were initially aligned by atoms 2−7 of the phthalimido moiety. The first CoMFA analysis, A1, produced a poorly predictive model, with a cross-validated q^2 of 0.342. Removal of **14**, **11**, and **17**, the compounds that are tetrafluorinated (at atoms 3−6), resulted in a vastly improved model, B1, with a cross-validated q^2 of 0.620. This dramatic improvement suggests that the tetrafluorinated compounds may be interacting with the receptor in a different way or may be interacting with a different receptor entirely. Models C1 and D1 were both aligned by atoms 2−7, as before. It is also possible that the interactions of the fluorine atoms cannot be sufficiently modeled using CoMFA methods. Model C1 excluded only those structures without an intact glutarimide. Since all of the structures were optimized individually, rather than based upon one active conformation, it is possible that those without the glutarimide function, which have a larger number of rotatable bonds, could introduce larger variation into the model, resulting in less predicivity. However, this model produced a cross-validated q^2 of 0.321, a lower value than that originally seen with model A1, which included all compounds. Removal of the three tetrafluorinated compounds, **14**, **11.** and **17**, in model D1, improved the CoMFA model, giving a cross-validated q^2 value of 0.620. Of the models aligned by the phthalimide ring, B1 provides a much more comprehensive sampling of the varied chemistry, suggesting that it is a more representative model.

In the second alignment, used for models E and F, compounds with intact glutarimide rings were aligned by atoms 1′-6′. Model E1 resulted in a q^2 of 0.303, suggesting little predictive value. The additional exclusion of the tetrafluorinated compounds, in model F1, greatly improved the predictive value, with q^2 equal to 0.537. However, the higher q^2 value of model B1, with a larger variety of structures suggests that it is a better model, since a larger variety of substitutions can be included. Therefore, CoMFA model B1 was optimized. This was accomplished by altering the steric and electrostatic field cutoffs, from 20 kcal/mol to 100 kcal/mol, to ensure that the structure−activity correlation
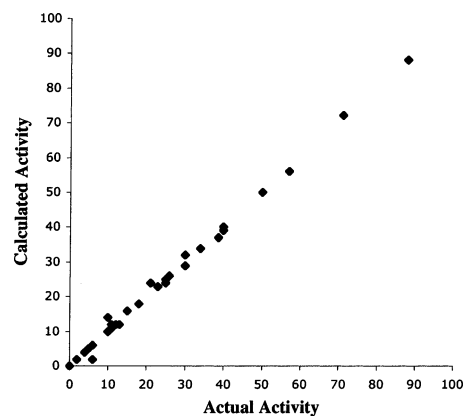

Figure 4. Actual versus calculated activity of final CoMFA model.
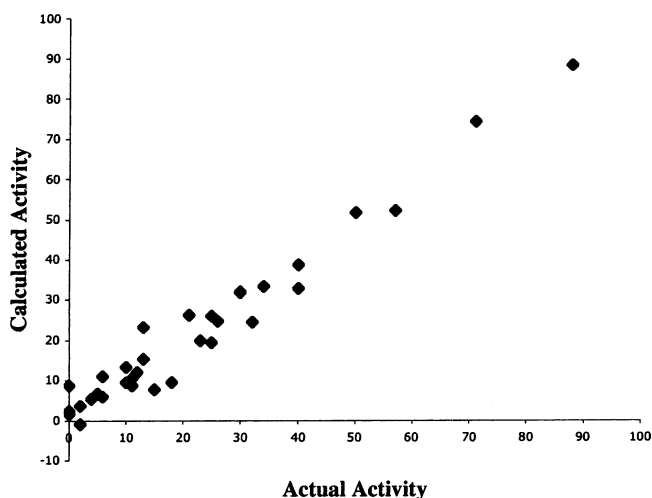
Figure 5. Actual versus calculated activity for best CoMSIA model.

is not due to chance and that the predictive value is not due to noise. Grid sizes of 1.0, 2.0 and 3.0 Å were tested and column filtering was varied from 1.0 to 4.0 kcal/mol for noise reduction. The best CoMFA model used a 2.0 Å grid size, with 20 kcal/mol steric and electrostatic field cutoffs and 2.0 kcal/mol column filtering to give a final q^2 of 0.663 at 3 components. Compound **18** (CPS1) was also excluded from the model, as it was a dramatic outlier, based upon residual values and due to the high SEM of the experimental data. These statistics suggest that the experimentally determined value was inflated. The non-cross-validated r^2 is 0.924 with a standard estimated error of 5.986 and an F-value of 125.838. The steric and electrostatic fields account for 55.4 and 44.6% of the model, respectively. Figure 4 shows the actual versus predicted values for the final model. Figure 6 provides a graphical representation of the CoMFA model. Figure 6a shows the CoMFA steric field contour map. Green (80% contribution) and yellow (10% contribution) contours represent regions where steric bulk increases and decreases biological activity, respectively. Compound **6**, the most active compound, can be seen to overlap with much of the green areas, where bulk is desired, and to avoid the yellow areas, where it is not. The shape of the contours suggests that compounds with a U-shaped low energy conformation, rather than a long straight chain, are more active. The large green region, where steric bulk improves activity, may allow for

Figure 6. (a) CoMFA steric field contour map. Yellow contours (10%) indicate regions where steric bulk decreases biological activity. Gr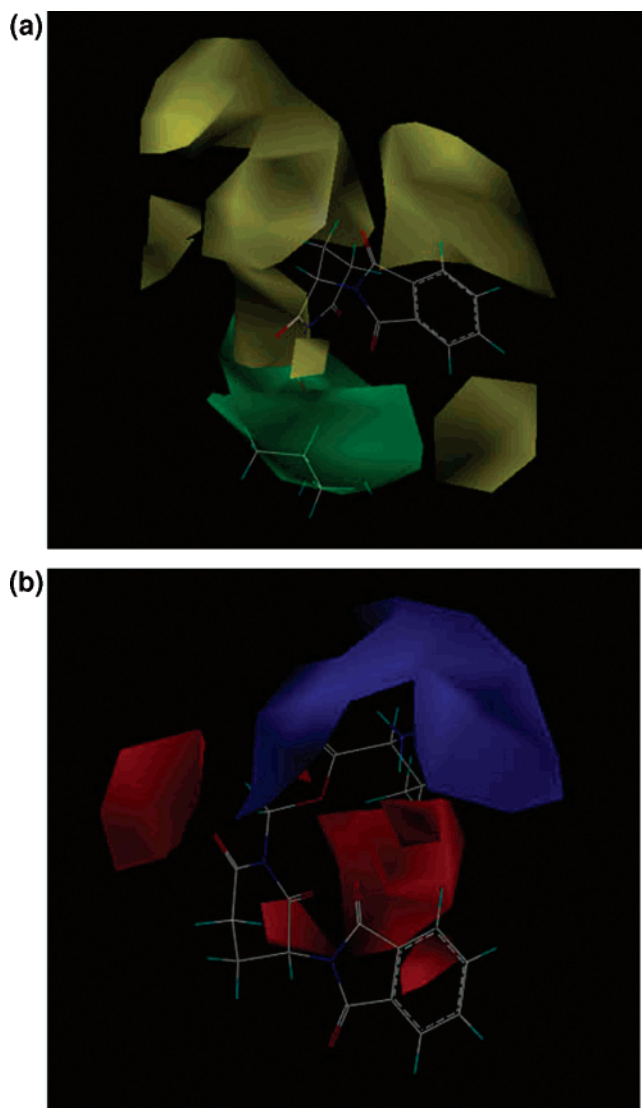een contours (80%) indicate regions where steric bulk increases activity. (b) CoMFA electrostatic field contour map. Blue (20%) and red regions (20%) indicate where biological activity is improved by positive and negative charge, respectively. Compound **6**, the most active ligand, is superimposed on both maps.

optimization of the pharmacology, since large bulky groups could potentially be added, improving the hydrophobic character. Figure 6b shows the CoMFA electrostatic field contour map. Blue (80% contribution) and red (20% contribution) contours represent regions where positive and negative charge increase activity, respectively. For example, the cationic amine on **6** is surrounded by a blue region, suggesting that this functionality greatly improves activity. The lack of colored contours in some areas does not necessarily indicate that these regions have no effect on activity. This can result if all the compounds tested have similar functionality located there.

Comparative Molecular Similarity Indices Analysis. Comparative molecular similarity indices analysis calculates steric and electrostatic fields, as in CoMFA, but additionally uses hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields. The resulting field contour maps are easier to interpret than in CoMFA,

since a Gaussian function is used to determine the distance dependence. Therefore, the similarity indices can also be calculated at grid points inside the molecule, not just outside, as with CoMFA.[22] The same alignments and models as used in the CoMFA analysis were used in the CoMSIA analysis. The first analysis, using the phthalimide ring for alignment and including all structures, was completed using steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields. It resulted in a poor model with $q^2 = 0.251$. Model B2, which excluded the three tetrafluorinated compounds, **14**, **11** and **17**, produced a dramatically improved cross-validated q^2 of 0.552 at six components. This is very similar to the CoMFA analysis. Again, the tetrafluorinated compounds do not seem to be interacting with the same active site as the remainder of the compounds. Compounds **15**, **16**, and **17** all have fluorine atoms on an aromatic ring, yet do not appear to be outliers, suggesting that the discordance of the tetrafluorinated compounds is not due to the calculation of the fields surrounding fluorine atoms, but due to these compounds interacting in another way. Additionally, the presence of fluorine atoms on the pthalimide ring may prevent metabolism, since they could block oxidation by cytochrome P450 isozymes that normally form the 5-hydroxy metabolite.[11] Preliminary microassay screening of mice treated with **19** (CPS11), **14**, or **17** (unpublished data) suggest that treatment with **19** produces a different effect than **14** and **17**, providing further evidence that the tetrafluorinated compounds are, in fact, operating via a different mode of action.[23]

Models C2 and D2 were also aligned by atoms 2−7, as before. Model C2 excluded only those compounds that do not possess an intact glutarimide ring. This resulted in a poor model, with a validated q^2 of 0.183 at one component. Further exclusion of the three tetrafluorinated compounds, in model D2, again dramatically improved the model, giving a cross-validated q^2 of 0.564 at five components. As can be seen in Table 2, the hydrogen bond donor field is the major contributor to the model (37.4%).

The second alignment used for CoMFA analysis was also subjected to CoMSIA analysis. Model E2 is based upon those compounds with intact glutarimide rings, aligned by atoms 1′-6′. It has little predictive value, with a cross-validated q^2 of 0.145 at 1 component. Again, exclusion of the tetrafluorinated compounds leads to a much improved model F2, with q^2 equal to 0.45 at 2 components.

Though model D2 has a slightly greater cross-validated q^2 value than model B2, 0.564 and 0.552 respectively, model B2 was optimized, since it had a higher r^2 value and lower standard error values. Additionally, inclusion of a wider variety of structures should lead to a more complete model, even though it results in a reduced q^2 value. The model was recreated excluding different fields, to see if the correlation between these structures and their activity is better represented with a different combination of fields. Models B3, B4, B5, B6, and B7 explored the different possible field combinations, as seen in Table 2. The exclusion of hydrogen-bond donor and acceptor fields improved the q^2 value, but decreased the r^2 value. Therefore, a final CoMSIA model was developed, using all five fields. It was based

Table 1. Results of CoMFA Analysis[a]

CoMFA model	results of PLS analysis					field		compounds excluded
	q^2	components	r^2	standard error	F	steric	electrostatic	
A1	0.342	1	0.541	15.005	43.684	0.394	0.606	none
B1	0.620	2	0.848	8.297	92.064	0.529	0.471	**45**, **48**, **49**
C1	0.321	1	0.528	16.097	33.584	0.396	0.604	**6**, **9**, **27**, **28**, **29**, **30**, **31**
D1	0.620	2	0.860	8.548	79.803	0.560	0.440	**6**, **9**, **27**, **28**, **29**, **30**, **31**, **45**, **48**, **49**
E1	0.303	1	0.500	16.695	29.050	0.365	0.635	**6**, **9**, **27**, **28**, **29**, **30**, **31**
F1	0.537	2	0.806	10.219	51.810	0.545	0.455	**6**, **9**, **27**, **28**, **29**, **30**, **31**, **45**, **48**, **49**
final	0.663	3	0.924	5.986	125.838	0.554	0.446	**1**, **45**, **48**, **49**

[a] q^2 = leave-one-out cross-validated r^2 value; r^2 = non-cross-validated regression coefficient; F = F-statistic for analysis.

Table 2. Results of CoMSIA Analysis[a]

CoMSIA model	results of PLS analysis					field				
	q^2	components	r^2	standard error	F	steric	electrostatic	hydrophobic	donor	acceptor
A2	0.251	1	0.616	13.735	59.293	0.14	0.211	0.176	0.334	0.139
B2	0.552	6	0.975	3.587	188.736	0.144	0.219	0.158	0.354	0.126
C2	0.183	1	0.588	15.308	42.86	0.125	0.196	0.172	0.38	0.128
D2	0.564	5	0.969	4.262	144.691	0.145	0.218	0.139	0.374	0.124
E2	0.145	1	0.538	16.06	33.732	0.118	0.19	0.169	0.411	0.111
F2	0.45	2	0.881	8.008	92.241	0.128	0.237	0.153	0.362	0.121
B3	0.603	2	0.87	7.681	110.104	0.312	0.427	0.261	–	–
B4	0.489	6	0.972	3.788	168.669	–	–	0.254	0.529	0.217
B5	0.534	2	0.882	7.302	123.559	0.203	0.255	–	0.389	0.153
B6	0.425	6	0.969	4.012	149.839	–	–	–	0.61	0.39
B7	0.618	2	0.85	8.23	93.758	0.447	0.553	–	–	–
final	0.662	3	0.958	4.45	236.086	0.165	0.226	0.147	0.333	0.129

[a] q^2 = leave-one-out cross-validated r^2 value; r^2 = non-cross-validated regression coefficient; F = F-statistic for analysis.

on all molecules, excluding **14**, **11,** and **17**, due to the tetrafluorination and **18** again because it was an outlier and had a high SEM. The final q^2 was 0.662 with three components and the r^2 value is 0.958. The standard error is only 4.45 and the F-value is 236.086. A plot of calculated versus actual activity for this model is shown in Figure 5. Figure 7a displays the steric and electrostatic field maps for this CoMSIA model. Compound **6** is overlaid. The yellow regions, indicating where steric bulk decreases activity, along with the red regions, where negative charge is desired, suggest that for optimal activity, the glutarimide ring should be perpendicular to the phthalimido ring. The hydrophobic field contour map, shown in Figure 7b, shows that hydrophobicity around the phthalimido ring promotes activity, while hydrophilic regions improve activity in the side chain. Figure 7c effectively shows where hydrogen-bond donors and acceptors on the receptor enhance or decrease binding. The cyan contours represent locations where a hydrogen bond acceptor group on the receptor would increase activity. For example, there are two cyan regions in close proximity to the cationic amine on **6**. These regions are present because the active compounds analyzed in forming of the model have hydrogen bond donors within 1.8 Å of these points. Conversely, the purple contours indicate regions where hydrogen bond acceptor groups on the receptor decrease activity. This arises because inactive compounds have hydrogen bond donor groups close to this region. The magenta contours map areas where hydrogen bond donor groups on the receptor increase activity. Therefore, for optimization of activity, the ligand should have hydrogen bond donors and acceptors within 1.8 Å of the cyan and magenta contours, respectively. Interestingly, the large band of red surrounding the glutarimide ring of **6**, and the red contour near it, suggest that any substitution of hydrogen bond acceptors onto this ring will only decrease activity. This could partially

explain the lack of activity seen for **9**, **12**, **13**, and **10**, all of which have either oxygen or nitrogen lone pairs in the vicinity.

Progressive Scrambling. Due to the limited number of compounds demonstrating statistically significant activity, all structures were used for the analyses, rather than dividing them into training and validation sets. Therefore, to validate these models, progressive scrambling was used, which gauges the dependence of the model on chance correlations. SAMPLS can potentially overstate the q^2 value when sets with redundant data are examined. Progressive scrambling of the final CoMFA model gave the best correlation at three components, with $q^2 = 0.357$ and $dq^2/dr^2_{yy} = 1.040$ (Table 3). These values suggest that the original CoMFA PLS analysis overstated the utility of the model.

Progressive scrambling of the CoMSIA PLS demonstrated that the model was not dependent on chance correlation. The best analysis was with two components, producing a q^2 of 0.43 with a dq^2/dr^2_{yy} value of 0.3646. The cSDEP value is 15.6626 which is much more comparable with the actual experimental error seen in the assay. Though the initial PLS analysis overstated the value of the model, we feel that this validation still warrants it very useful for the design of new inhibitors.

Since there are only three tetrafluorinated compounds that were excluded from the main model, it is not possible to develop another CoMFA or CoMSIA model to learn about their structure−activity relationships. Several approaches were taken to confirm that the fluorinated compounds warranted exclusion from the models. LogP values were calculated for all the compounds using ClogP (Biobyte Corp., Claremont, CA). However, there was no significant correlation between logP and inhibitory activity. Second, the charges were calculated and geometry optimized for all compounds, using MOPAC with AM1. This was done to ensure that it was not the Gasteiger−Hückel charges that led to the
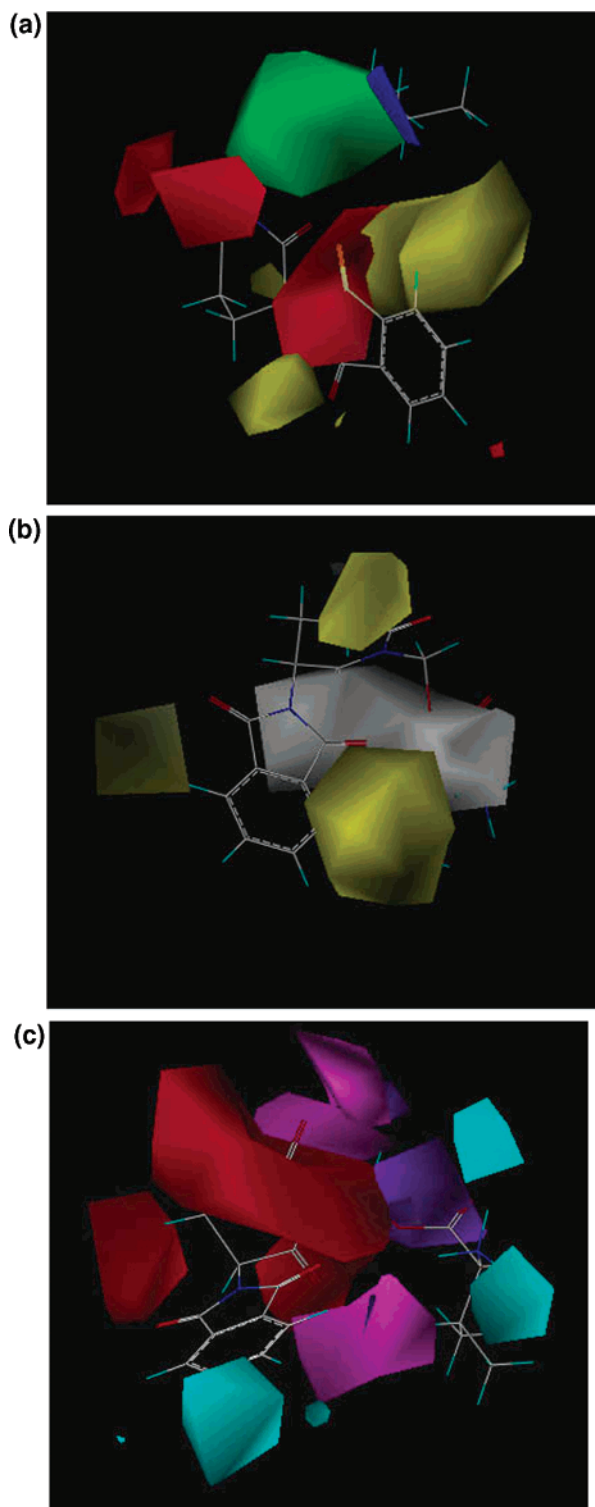
Table 3. Results of Progressive Scrambling of CoMFA Analysis

components	q^2	cSDEP	$dq^{2\prime}/dr^2_{yy\prime}$
1	0.31	17.221	1.006
2	0.34	16.9112	1.123
3	0.36	16.6343	1.040
4	0.26	17.8341	2.007

Table 4. Results of Progressive Scrambling of CoMSIA Analysis

components	q^2	cSDEP	$dq^{2\prime}/dr^2_{yy\prime}$
1	0.42	15.7388	0.249
2	0.43	15.6626	0.365
3	0.42	15.8412	0.665
4	0.40	16.0281	0.937

outlier status of the fluorinated compounds. CoMFA fields were calculated, but there was no improvement using this semiempirical method to calculated charges. A comparison of **11** and **17**, which are very similar in shape, suggests that hydrogen bonding at positions 1′ and 5′ may significantly affect the activity. The **11** glutarimide ring contains two NH-units not present in **17** and shows very reduced activity, though it is a polar isostere of **17**. Additionally, the decreased lipophilicity of **11** may result in the decline in activity.

Summary and Conclusion

In the absence of any knowledge about the target of these thalidomide analogues, the final CoMFA and CoMSIA models provide a guide for the design of more active and potent analogues. Both models have good cross-validated q^2 values, suggesting good predictive ability. The progressive scrambling gives the CoMSIA model a higher r^2 and suggests that this model is much more resilient to chance correlation. It offers a more interpretable representation of the structure activity correlations, with the inclusion of hydrogen bonding parameters, since this is obviously of great importance in the binding of these compounds. Interestingly, these models have also suggested that the tetrafluorinated analogues are operating via a different mechanism. Further experiments will have to be carried out to confirm this.

Figure 7. (a) CoMSIA steric and electrostatic field contour map. Green and yellow represent sterically favored (80%) and disfavored (3%) regions, respectively. Blue and red represent electrostatically favored (80%) and disfavored (20%) regions, respectively. (b) CoMSIA hydrophobic field contour map. Yellow regions (80%) indicate areas where hydrophobic groups increase activity and white regions (20%) indicate areas where hydrophobic groups decrease activity. (c) CoMSIA H-bond donor and acceptor contour maps. Cyan (80%) and purple (20%) contours indicate regions where hydrogen bond acceptor groups on the receptor increase and decrease activity, respectively. Magenta (80%) and red (20%) contours indicate regions where hydrogen bond donor groups on the receptor increase and decrease activity, respectively. **6**, the most active compound, is overlaid in each plot.

References

(1) Folkman, J. Tumor angiogenesis: a possible control point in tumor growth. *Ann. Intern. Med.* **1975**, *82*, 96−100.
(2) Kim, K. J.; Li, B.; Winer, J.; Armanini, M.; Gillett, N. et al. Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. *Nature* **1993**, *362*, 841−844.
(3) Singhal, S.; Mehta, J. Thalidomide in cancer: potential uses and limitations. *BioDrugs* **2001**, *15*, 163−172.
(4) Dredge, K.; Dalgleish, A. G.; Marriott, J. B. Thalidomide analogues as emerging anti-cancer drugs. *Anticancer Drugs* **2003**, *14*, 331−335.
(5) Singhal, S.; Mehta, J.; Desikan, R.; Ayers, D.; Roberson, P. et al. Antitumor activity of thalidomide in refractory multiple myeloma. *N. Engl. J. Med.* **1999**, *341*, 1565−1571.
(6) Little, R. F.; Wyvill, K. M.; Pluda, J. M.; Welles, L.; Marshall, V. et al. Activity of thalidomide in AIDS-related Kaposi's sarcoma. *J. Clin. Oncol.* **2000**, *18*, 2593−2602.
(7) Fine, H. A.; Figg, W. D.; Jaeckle, K.; Wen, P. Y.; Kyritsis, A. P. et al. Phase II trial of the antiangiogenic agent thalidomide in patients with recurrent high-grade gliomas. *J. Clin. Oncol.* **2000**, *18*, 708−715.
(8) Eisen, T.; Boshoff, C.; Mak, I.; Sapunar, F.; Vaughan, M. M. et al. Continuous low dose Thalidomide: a phase II study in advanced melanoma, renal cell, ovarian and breast cancer. *Br. J. Cancer* **2000**, *82*, 812−817.

(9) Figg, W. D.; Dahut, W.; Duray, P.; Hamilton, M.; Tompkins, A. et al. A randomized phase II trial of thalidomide, an angiogenesis inhibitor, in patients with androgen-independent prostate cancer. *Clin. Cancer Res.* **2001**, *7*, 1888−1893.

(10) Bauer, K. S.; Dixon, S. C.; Figg, W. D. Inhibition of angiogenesis by thalidomide requires metabolic activation, which is species-dependent. *Biochem. Pharmacol.* **1998**, *55*, 1827−1834.

(11) Ando, Y.; Fuse, E.; Figg, W. D. Thalidomide metabolism by the CYP2C subfamily. *Clin. Cancer Res.* **2002**, *8*, 1964−1973.

(12) Eriksson, T.; Bjorkman, S.; Roth, B.; Bjork, H.; Hoglund, P. Hydroxylated metabolites of thalidomide: formation in-vitro and in-vivo in man. *J. Pharm. Pharmacol.* **1998**, *50*, 1409−1416.

(13) Schumacher, H.; Smith, R. L.; Williams, R. T. The metabolism of thalidomide: the spontaneous hydrolysis of thalidomide in solution. *Br. J. Pharmacol.* **1965**, *25*, 324−337.

(14) Eriksson, T.; Bjorkman, S.; Hoglund, P. Clinical pharmacology of thalidomide. *Eur. J. Clin. Pharmacol.* **2001**, *57*, 365−376.

(15) Hoglund, P.; Eriksson, T.; Bjorkman, S. A double-blind study of the sedative effects of the thalidomide enantiomers in humans. *J. Pharmacokinet. Biopharm.* **1998**, *26*, 363−383.

(16) Kenyon, B. M.; Browne, F.; D'Amato, R. J. Effects of thalidomide and related metabolites in a mouse corneal model of neovascularization. *Exp. Eye Res.* **1997**, *64*, 971−978.

(17) Knoche, B.; Blaschke, G. Investigations on the in-vitro racemization of thalidomide by high-performance liquid-chromatography. *J. Chromatog. A* **1994**, *666*, 235−240.

(18) Cramer, R.; Patterson, D.; Bunce, J. Comparative Molecular-Field Analysis (CoMFA) 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(19) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. The collinearity problem in linear regression − The partial least-squares (PLS) approach to generalized inverses. *Siam J. Sci. Stat. Comput.* **1984**, *5*, 735−743.

(20) Bush, B. L.; Nachbar, R. B. Sample-distance Partial Least-Squares PLS Optimized for Many Variables, with Application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587−619.

(21) Clark, R. D.; Sprous, D. G.; Leonard, J. M. Validating models based on large data sets. *Rational Approaches to Drug Design*; Prous Science SA: Barcelona, 2001; pp 475−485.

(22) 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.

(23) Macpherson, G. R. Unpublished results.