Receptor-Guided Alignment-Based Comparative 3D-QSAR Studies of Benzylidene Malonitrile Tyrphostins as EGFR and HER-2 Kinase Inhibitors

Shantaram Kamath and John K. Buolamwini*

Department of Pharmaceutical Sciences, College of Pharmacy, University of Tennessee Health Sciences Center, 847 Monroe Avenue Suite 327, Memphis, Tennessee 38163

Received February 11, 2003

The overexpression and/or mutation of the epidermal growth factor receptor family of tyrosine kinases, namely EGFR and HER-2, have been implicated in poor prognosis of human solid tumors and are under investigation as molecular targets for cancer therapy. To gain insights into selectivity in the interaction of inhibitors at the ATP site of the two kinases, we have carried out docking, comparative molecular field analysis (CoMFA), and comparative molecular similarity analysis (CoMSIA) 3D-QSAR studies on 50 benzylidene malonitrile derivatives. Docking studies indicate different binding modes (**A**, **B**, **C**, **D**, and **E**) that are dependent on the R_1 substituent of the compounds. Binding modes **A** and **B** are favored by compounds having a hydroxyl substituent at R_1 whereas a methoxy substituent at R_1 results in compounds occupying binding modes, **C**, **D**, and **E**. The compounds preferred modes **A**, **B**, and **C** in the apo-enzyme whereas modes **B**, **D**, and **E** were preferred in the enzyme in complex with erlotinib. For 3D QSAR studies, based on the multiple binding modes obtained from the docking, four composite alignment strategies (**I**, **II**, **III**, **IV**) were employed and compared with alignment **V**, which is based on pairwise atom alignment of the common structural elements. Alignments **I** and **II** produced models with better predictive ability than those from alignments **III** and **IV** against an external test set. In the EGFR kinase results, alignments **I** and **II** produced comparable 3D-QSAR models with alignment **II** being slightly better than **I**, whereas in the HER-2 results, alignment **I** was better than alignment **II** in its predictive ability. It appears that differences in binding mode preferences at the ATP site might constitute a reason for the selectivity of the dihydroxy compounds as inhibitors of EGFR relative to HER-2. They are more likely to have multiple binding modes at the ATP site of EGFR, i.e., either modes **A** or **B**, than in the ATP site of HER-2 where they are possibly limited to only binding mode, **A**. Selectivity of the methoxy compounds on the other hand appears to depend on hydrogen bonding interactions involving the cyano group and residue 751 in the ATP site.

Introduction

Cancer chemotherapy has entered a new era of molecularly targeted therapeutics, which are highly selective and not associated with the serious toxicities of conventional cytotoxic drugs.[1,2] The first group of these novel anticancer drugs are those targeting mutant or aberrantly expressed oncogenic growth factor receptor and nonreceptor tyrosine kinases involved in mitogenic or proliferative signal transduction pathways in cancer cells.[3,4] The epidermal growth factor receptor (EGFR) and the homologous HER-2 kinases, which are members of the type 1 or erbB/HER receptor tyrosine kinase (RTK) family,[5] have emerged as the most viable anticancer molecular targets in this family of four members, the others being HER-3 and HER-4. The overexpression or activating mutation of EGFR and/or HER-2 occurs frequently in breast, ovarian, and non small cell lung cancers[6−8] making them attractive therapeutic targets in these cancers.[9] The EGFR and HER-2 like other RTKs are single polypeptide chain transmembrane proteins composed of an extracellular ligand-binding domain, a transmembrane lipophilic segment, and an intracellular tyrosine-kinase (TK) domain.[10] Binding of extracellular ligands such as epidermal growth factor (EGF) and transforming growth factor α (TGFα) to the extracellular ligand-binding domain of EGFR results in receptor homo- or heterodimerization, activation of intrinsic TK activity, and autophosphorylation of the receptors, which initiates a mitogenic signaling cascade.[6,11] While a specific ligand for HER-2 has not been found, it undergoes heterodimerization with EGFR or other members of the family when they bind their cognate ligands and get activated.

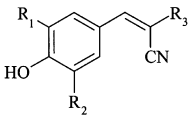
Antibodies and small molecules are being pursued as therapeutics targeting EGFR or HER-2. This has resulted in for example, trastzumamab (Herceptin, Genentech, South San Franscisco, CA), a humanized monoclonal antibody already being used in the clinic for the treatment of patients with HER-2 overexpressing metastatic breast cancer,[12] while an orally active anilinoquinazoline, ZD1839 (Iressa),[9] ATP-site inhibitor of EGFR has recently been approved for clinical use against refractory lung cancers.

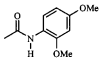
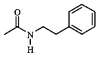
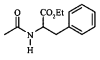
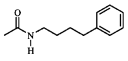
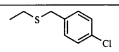
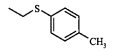
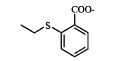
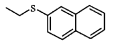
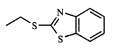
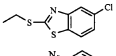
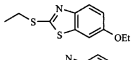
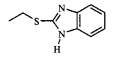
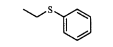
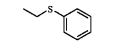
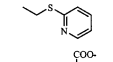
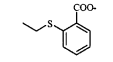
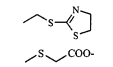
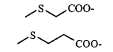
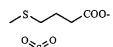
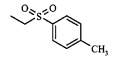
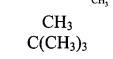
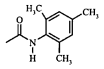
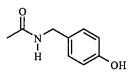
To date, there are no reports on 3D QSAR studies[13] of EGFR kinase inhibitors belonging to the benzylidene malonitrile class and only one report of a limited 3D QSAR study on HER-2 kinase inhibitors.[14] Here we report the application of docking, comparative molecular

* To whom correspondence should be addressed. Phone (901) 448−7533, fax (901) 448−6828, e-mail: jbuolamwini@utmem.edu.

10.1021/jm030065n CCC: $25.00
Published on Web 09/25/2003

Table 1. Structures and Inhibitory Activities of Benzylidene Malonitrile Tyrphostins



Compd	R_1	R_2	R_3	Actual pIC_{50} EGFR	Actual pIC_{50} HER-2
1	OH	H		5.30	4.77
2	OH	H		5.29	4.89
3	OH	H		4.85	4.74
4	OH	H		5.18	4.92
5	OH	H		4.51	4.46
6	OH	H		4.72	4.30
7	OH	H		4.92	5.31
8	OH	H		4.99	4.47
9	OH	H		4.96	5.52
10	OH	H		5.90	4.38
11	OH	H		4.60	3.78
12	OH	H		5.23	4.77
13	OH	H		6.30	4.87
14[a]	OH	H		5.60	4.43
15	OH	H		6.21	4.64
16[b]	OH	H		4.31	3.48
17	OH	H		5.30	3.30
18	OH	CH_3	CN	4.21	4.23
19	OH	H	$CONH_2$	5.35	5.19
20	OH		$CONH_2$	6.19	5.99
21	OCH_3		$CONH_2$	6.05	5.00
22	OCH_3		$CONH_2$	6.00	6.60
23	OCH_3		$CONH_2$	4.80	5.70
24	OCH_3		$CONH_2$	5.76	5.78
25	OCH_3		CN	5.10	5.00
26	OCH_3		$CONH_2$	5.77	5.34
27	OCH_3		$CONH_2$	4.72	6.46
28	OCH_3		$CONH_2$	4.08	4.82
29	OCH_3		$CONH_2$	3.78	-
30	OCH_3		$CONH_2$	4.84	5.21
31	OCH_3		$CONH_2$	6.40	6.89
32	OCH_3		CN	5.17	5.60
33	OCH_3		$CONH_2$	5.47	6.15
34	OCH_3		$CONH_2$	6.15	6.35
35	OCH_3		$CONH_2$	5.06	5.06
36	OCH_3		CN	4.27	5.34
37	OCH_3		$CONH_2$	-	4.04
38	OCH_3		$CONH_2$	4.32	4.80
39	OCH_3		CN	4.64	4.76
40	OCH_3	CH_3	CN	4.06	3.95
41	$C(CH_3)_3$	$C(CH_3)_3$	$CSNH_2$	3.30	6.00
42	OH	H		4.72	4.68
43	OH	H		5.52	5.54
44	OH	H		6.15	4.46
45	OH	H		5.68	4.24
46	OH	H		4.82	5.00
47	OCH_3		$CONH_2$	4.66	5.48
48	OCH_3		$CONH_2$	5.28	5.82

[a] (+) Isomer. [b] Racemate.

field analysis (CoMFA),[15] and comparative molecular similarity analysis (CoMSIA)[16] 3D QSAR methods to a series of benzylidene malonitrile tyrphostins that were shown to be potent inhibitors of EGFR and/or HER-2 kinase activities.[17−20] The study was undertaken to gain insights into the structural determinants and the electrostatic and steric factors that underlie the selectivity of these inhibitors for one kinase or the other, information that will be valuable for designing inhibitors of EGFR and/or HER-2 kinases.

Materials and Methods

Dataset. The data set comprises a series of 50 benzylidene malonitrile derivatives (Table 1),[17−20] which were demonstrated to be ATP competitive inhibitors of EGFR and HER-2.[21] For compounds **14** and **16**, both stereoisomers were considered. The test set used in validating the QSAR models consisted of seven compounds, **42**−**48**. The IC_{50} values for inhibition of phosphorylation by EGFR and HER-2[17−20] were converted to pIC_{50} (−log IC_{50}, Table 1) values and used as dependent variable in the CoMFA and CoMSIA 3D-QSAR analyses.

Molecular Modeling and Alignment. CoMFA and CoMSIA model development require that the 3D structures of the molecules be aligned according to a reasonable conformation, possibly the bioactive conformation. There are no reports of crystal structures of any of these compounds either alone or in complex with any of the RTKs. Therefore, to obtain reasonable molecular alignments, the inhibitors were docked into the ATP site of each kinase since they have been shown by Osherov et al.[21] to be ATP competitive inhibitors. Docking studies were carried out on the ATP site of the apo-enzyme as well as the ATP site of the enzyme bound to erlotinib to identify any differences in binding that may result from perturbation caused by complexation with the ligand. After the binding modes were identified, the molecular alignments were obtained for CoMFA and CoMSIA studies by superimposing the compounds in their major docking modes.

Homology Modeling. A homology model of the tyrosine kinase domain of HER-2 was built using HOMOLOGY module of the InsightII modeling software (version 2000) from Accelrys Inc. (San Diego, CA) running on an SGI Octane 2 (R12000) workstation. The crystal structures of apo-form (PDB code 1M14) and the complexed form (PDB code 1M17) of EGFR kinase domain were used as the templates to build homology models of the corresponding forms of HER-2. A single template was used as both enzymes belong to the same family and share greater than 80% sequence identity in the tyrosine kinase domain. The residues that were mutated in HER-2 were subjected to minimizations with 500 cycles of steepest descents while the rest of the structure was fixed. The structural statistics of the generated homology model complied with the PROSTAT checks in Homology.

Docking. The program GOLD (version 2.0)[22] from Cambridge Crystallographic Data Center, UK, was used to dock the inhibitors into the ATP site of the enzymes. GOLD is an automated ligand docking program that uses a genetic algorithm to explore ligand conformational flexibility. It also incorporates receptor flexibility to a limited extent by optimizing the torsion angles of serine and threonine hydroxyl groups, and lysine NH_3^+ groups. A built-in cavity detection algorithm is used to restrict the region of interest to concave solvent-accessible areas. Chemscore as implemented in the GOLD program was used as the scoring function. The ATP site was defined as residues lying within 15 Å of Thr766 in both enzymes. Up to 20 different docking solutions were obtained for each molecule, the docking being terminated when the top 10 solutions were within a root-mean-square deviation (RMSD) of ≤1.5. Preliminary docking using the default torsion distribution file indicated the inability of the program to retain the planarity of the styryl-conjugated system in the docked solutions. Therefore, the torsion distribution file was modified so as to maintain a nearly planar styryl-conjugated system in the docked solutions. We also carried out pilot docking runs of erlotinib into EGFR using the same definition of the ATP site, as mentioned above, to confirm the ability of the program to reproduce the crystal structure. We were able to reproduce the crystal structure of erlotinib in EGFR (Figure 1a) only after introduction of the critical water molecule, which is present in the crystal structure that forms a bridge between the N3 of the quinazoline ring and the hydroxyl group of Thr766.

CoMFA and CoMSIA 3D QSAR Studies. CoMFA and CoMSIA models were generated using SYBYL 6.8, Tripos Associates (St. Louis, MO). A 3D cubic lattice with a grid spacing of 2.0 Å and extending 4.0 Å beyond the aligned molecules in all directions was used for deriving the descriptor fields. CoMFA descriptor fields were calculated using an sp^3 carbon probe with a van der Waal radius of 1.52 Å and a charge of +1.0 to generate steric and electrostatic fields using a distance dependent dielectric at each lattice point. An energy cutoff of 30 kcal/mol was applied. The CoMFA steric and electrostatic fields were scaled by the CoMFA-STD method in SYBYL. CoMSIA similarity index descriptors were derived according to Klebe *et al.*[16] using a similar lattice box as used in CoMFA calculations. Five properties, namely steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor, were evaluated using the probe atom of 1.4 Å radius and +1.0 charge, applying the default attenuation factor of 0.3. The CoMSIA steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from atomic partial charges, the hydrophobic fields are derived from atom-based parameters developed by Viswanadhan et al.,[23] and the hydrogen bond donor and acceptor indices are obtained from a rule-based method derived from experimental values.

The CoMFA and CoMSIA descriptors were used as independent variables and pIC_{50} values for kinase inhibitions were used as dependent variables in partial least squares (PLS) regression analyses to derive 3D QSAR models using the standard implementation in the SYBYL package running the SGI workstation mentioned above. The internal consistency of the models was evaluated by leave-one-out (LOO) cross-validation. The cross-validated coefficient, q^2, was calculated using eq 1.

$$q^2 = 1 - \frac{\Sigma(Y_{\text{predicted}} - Y_{\text{actual}})^2}{\Sigma (Y_{\text{actual}} - Y_{\text{mean}})^2} \qquad (1)$$

where $Y_{\text{predicted}}$, Y_{actual}, and Y_{mean} are predicted, actual and mean values of the target property (pIC_{50}), respectively. The optimum number of PLS components for deriving the final regression models was the one that corresponds to the lowest value of $\Sigma(Y_{\text{predicted}} - Y_{\text{actual}})^2$, the predictive sum of squares (PRESS). In addition to the q^2, the corresponding PRESS, the conventional correlation coefficient (r^2) and its standard error (s) were also computed. CoMFA and CoMSIA coefficient maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding CoMFA or CoMSIA descriptor values.

Results and Discussion

Receptor-Guided Alignments. To obtain alignments for the 3D QSAR studies, we performed a flexible docking of the molecules into the ATP-binding sites of each kinase. The 3D coordinates of EGFR kinase domain available in the PDB (code 1M14 and 1M17) were used to define its ATP-binding site. Since the 3D structure of HER-2 is not available, homology models in the apo-form and the complex form of its kinase domain were constructed using the corresponding EGFR kinase domain as template. In the ATP site definition, only one amino acid difference exists between the two receptors, i.e., a change of Cys 751 in EGFR to Ser in HER-2. This is the only instance wherein the residue that is changed exposes its side chain to the ATP site. The other changes close to the ATP site, namely Phe 688 to Leu, Ser 744 to Gly, Asp 746 to Gly, Asn 747 to Ser, His 749 to Tyr and Ile 765 to Val, that do not expose their side chains to the ATP could be possibly contributing to the differences in selectivity of these enzymes toward various inhibitors. The effects of these residues in terms of their effects on dynamics of the ATP site is difficult to model using static docking on the receptor models.

There were differences in the docking modes of the compounds in the apo-enzyme and the complexed enzyme receptor sites. In examining the first ranked docking solutions five major binding modes were identified, which have been designated modes **A**, **B**, **C**, **D**, and **E** (see Figure 1). Modes **A**, **B**, and **C** were the major docking modes adopted in the apo-enzyme model while modes **B**, **D**, and **E** were adopted in the enzyme model derived from the complex. Mode **A** does not occur in the model derived from the complex. Mode **C** happens to

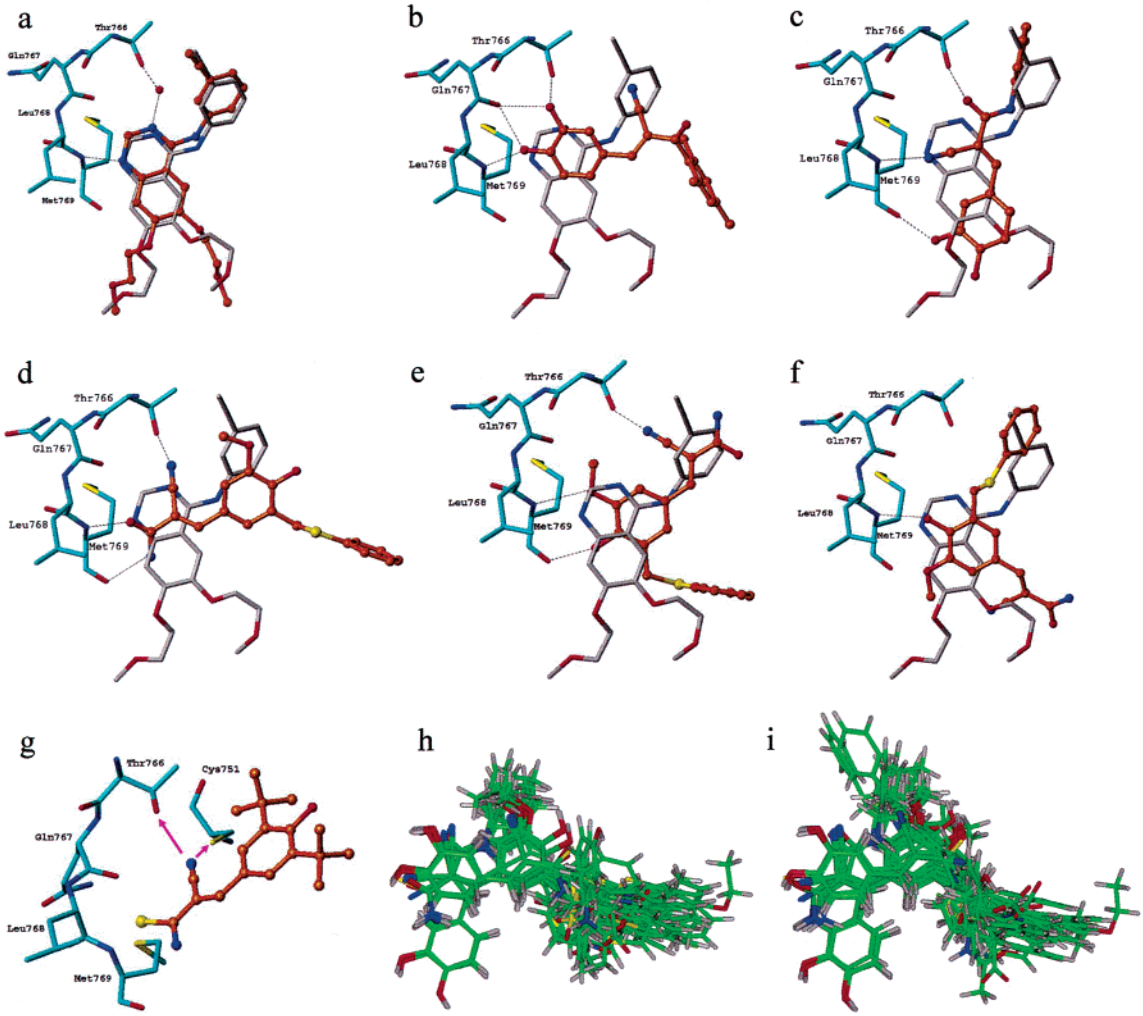


Figure 1. Docking modes of the compounds overlayed on the bound conformation of erlotinib. The reproduction of the crystal bound conformation and orientation by GOLD is shown in **a**. Binding Modes **A**, **B**, **C**, **D**, and **E** are shown in **b**, **c**, **d**, **e**, and **f**, respectively. The docked molecules are rendered as orange colored ball-and-stick while the residues of the hinge region of the enzyme are shown in cyan. Erlotinib in its binding orientation as it is in complex with EGFR (PDB code: 1M17) is colored gray. Hydrogen bonds are shown as dotted lines. The juxtaposition of the cyano group of mode **C** docking compounds between Thr766 and Cys751 in EGFR is shown in **g**. Hydrogens are not shown.. Alignments **I** and **II** are shown in **h** and **i**, respectively.

be a minor mode in the complex structure model, whereas modes **D** and **E** are minor modes in the apo-enzyme model. The occurrence of a particular mode was structure related (Table 1), in that compounds that have hydroxyl substitution at R_1 (also referred to as the dihydroxy compounds) preferred either mode **A** or **B** whereas molecules that have a methoxy substitution at R_1 adopted mode **C**, **D** and **E**. It is worth noting that the latter group of compounds is likely to have high inhibitory activity in both kinases while the former group of compounds is generally more selective toward EGFR. A comparison of the interactions of erlotinib with the hinge region of EGFR kinase (PDB code: 1M17) and the interactions of the benzylidene malonitriles in the five docking modes (Figure 1b−f) brings out the possible commonalities in the interactions shared by benzylidene malonitriles and erlotinib (an ATP competitive site inhibitor). The interactions of the inhibitors with the hinge region of the kinases, especially residues Thr766, Gln767, and Met 769, have been shown to be important for inhibitory activity.[24] In mode **A**, the oxygen atoms of the hydroxyl groups occupy a positions equivalent to the N1 and N3 atoms of the quinazoline moiety of erlotinib while the hydrogens attached to these oxygen atoms hydrogen bond with Gln767 backbone carbonyl as seen for the H-2 of erlortinib.[25] Thus the hydroxyl groups interact with residues Thr766 and Met769. In the case of mode **B**, the cyano group and the carbonyl of the amide share the same interactions as N1 and N3 of erlotinib, respectively, while the amide nitrogen and its substituent occupy a similar location as the substituted anilino group of erlotinib. Here the interactions appear more like those of erlotinib than in mode **A**, as acceptor-only groups, cyano and the carbonyl, interact with Thr766 and Met 769. In mode **C**, the carbonyl of the amide and the cyano group occupy positions similar to the N1 and N3, respectively, while the methoxy substituent at R_1 occupies the location of the phenyl group of the anilino moiety of erlotinib. In mode **C**, the interactions are similar to those in mode **B** but the acceptors are interchanged. In mode **D**, the oxygen atom of the methoxy group interacts with the NH of the Met769 and the hydroxyl group forms a hydrogen bond with the CO of Met769, and the cyano group shares the same interaction as in Mode **C**. In mode **E**, the hydroxyl oxygen forms only one hydrogen bond with the hinge

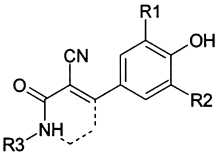


Figure 2. A structural insight from docking conformations of the benzylidene malonitriles. The dotted line indicates a possible linking strategy that would place the two acceptors on the same side allowing favorable interactions with the hinge regions of the kinase.

region residues, i.e., NH of Met769. Recently, it has been shown by mutagenesis studies that Thr766 is an important interaction residue on EGFR that contributes to kinase inhibitory activity.[26] In light of the reduced number of hydrogen bond interactions with important residues and the apparent lack of involvement of Thr766 in the interactions of this binding mode (**E**), it is less compelling as a mode for the methoxy compounds.

Based on these docking modes, five alignments were generated for comparative 3D QSAR studies, as follows. Alignment **I** (Figure 1h) consisted of binding conformations of compounds that could occupy mode **A** (irrespective of the rank in docking), compounds that occupied only mode **B**, and compounds that occupied mode **C**. Alignment **II** (Figure 1i) had binding conformations of compounds that occupied only mode **A**, all compounds that could occupy mode **B** (irrespective of the rank in docking), and compounds that occupied mode **C**. Alignment **III** (Figure not shown) had conformations of compounds in mode **B** and compounds in mode **D** whereas alignment **IV** (Figure not shown) had conformations of compounds in mode **B** and compounds in mode **E**. Conformations for alignments **I** and **II** were obtained from docking onto the apo-enzyme whereas alignments **III** and **IV** were obtained from docking onto the complex structure's ATP site. Alignment **V** (Figure not shown), a classical alignment strategy in 3D-QSAR was obtained by using the MULTIFIT routine in SYBYL, which uses a pairwise atom-forcing using a spring constant of 20.0 kcal/Å/mol to the homologous portion of the compounds on to a selected template. The most active compound, **31**, was selected as a template.

By performing QSAR studies using these composite alignments (**I**, **II**, **III**, and **IV)**, we envisaged that the influence of multiple binding modes would be included in the 3D-QSAR models and provide a means to distinguish between binding modes preferred for inhibitory activity. This issue arises because of the inadequacy of current scoring functions to accurately calculate binding energies of docked structures, which often leaves an element of uncertainty in identifying correct binding mode(s).

Another insight that was obtained from the docking simulations is that these molecules possibly bind with the cyano and carbonyl groups on the same side, thereby suggesting that incorporating these into a ring system, as shown in Figure 2, might enhance binding affinity.

3D QSAR Studies. As already mentioned, 3D QSAR studies were undertaken with the above five receptor-based composite alignments, **I**, **II**, **III**, and **IV** and the MULTIFIT alignment **V** to investigate the probable binding mode(s) at the ATP sites, to determine the effects of steric, electrostatic, hydrophobic, and hydrogen bonding on selective kinase inhibitory activity toward either EGFR or HER-2 kinase, and to obtain models for predicting inhibitory activity. PLS analysis of all the compounds in the training set resulted in 3D QSAR models with low q^2 values (approximately 0.1 in many instances). Improvements in q^2 values up 0.702 were attained by the omission of two to five possible outliers depending on the particular model. The PLS statistics of the various 3D-QSAR models for EGFR and HER-2 inhibition are summarized in Tables 2 and 3, respectively. The fitted curves for activity prediction of the training and test sets for EGFR and HER-2 inhibition are shown in Figures 3 and 4, respectively. The various QSAR models that were derived are discussed in the following sections.

EGFR QSAR Models. Alignments **I**, **II**, **III**, and **V** gave good CoMFA QSAR models with reasonable PLS statistics. The improvement of q^2 for alignment **IV** required the removal of eight possible outliers. The large number of outliers for alignment **IV** indicates a poor alignment, ruling out mode **E** as a favorable docking mode in EGFR. The PLS statistics are shown in Tables 2, and the corresponding fitted curves of the CoMFA models are shown in Figure 3a−d. Based on the scatter of the data points about the fitted lines (see Figure 3a−d), the CoMFA models from alignments **I**, **II**, and **III** outperformed that from alignment **V**, regardless of the q^2 values. It is worth noting that even though the q^2 is important for internal consistency in a QSAR model it may not always be the best measure of predictive ability as has been shown recently.[27] Models from alignments **I** and **II** also performed better at predicting the activities of the compounds in the test set than did the model from alignments **III** and **V**. Alignment **III** was the worst among the various alignments in predicting the test set. This suggests that the receptor-based alignments are advantageous over the MULTIFIT alignment in deriving predictive QSAR models for this series of compounds provided the docking mode is correctly identified. Between models from alignments **I** and **II**, the latter model performed slightly better than the former in predicting the test set (compare Figure 3a with Figure 3b).

Paralleling the CoMFA QSAR models, alignments **I**, **II**, **III**, and **V** gave good CoMSIA QSAR models. Alignment **IV** was eliminated for reasons mentioned above. The fitted curves of the CoMSIA models are shown in Figure 3e−h and the corresponding PLS statistics are reported in Table 2. CoMSIA models from alignments **I**, **II**, and **III** performed better than the model derived from alignment **V** in predicting the activities of the training set. However, models from alignments **I** and **II** predicted the test set compounds better than that from alignment **V**. The model from alignment **III** performed poorly on the test set, regardless of the q^2 value. As was the case of the CoMFA analysis, the CoMSIA model from alignment **II** is slightly better than that from alignment **I** in predicting the activities of the test set.

Therefore, for EGFR kinase inhibitory activity, the results from both the CoMFA and CoMSIA models agree and suggest that alignment **II** (Figure 1i), which had a greater number of dihydroxy compounds in binding mode **B** (Figure 1c), provided the best model. However, the fact that alignment **I** (Figure 1h) also provided a good QSAR model means that perhaps both binding

Table 2. PLS Statistics of 3D-QSAR Models for EGFR Inhibitory Activities from Various Alignments for Benzylidene Malonitrile Tyrphostins

	alignments			
PLS statistics	**I**	**II**	**III**	**V**
		CoMFA		
q^2	0.522	0.540	0.702	0.651
components	6	6	5	6
outliers	**19**, **28**, **29**, **38**	**18**, **21**, **29**, **40**, **41**	**15**, **23**, **28**, **34**	**15**, **23**
PRESS	0.477	0.452	0.439	0.418
r^2	0.959	0.971	0.977	0.921
s	0.140	0.113	0.122	0.192
F	120.183	175.420	218.744	69.507
contribution				
steric	0.445	0.533	0.476	0.473
electrostatics	0.555	0.469	0.524	0.527
		CoMSIA		
q^2	0.498	0.519	0.667	0.614
components	6	6	4	6
outliers	**21**, **28**, **29**, **38**, **41**	**21**, **29**, **38**, **41**	**15**, **23**, **28**, **38**	**15**, **23**, **38**
PRESS	0.476	0.478	0.460	0.436
r^2	0.947	0.958	0.955	0.921
s	0.154	0.141	0.170	0.197
F	92.705	122.769	142.309	62.639
contribution				
steric	0.119	0.107	0.151	0.172
electrostatics	0.247	0.243	0.215	0.289
hydrophobic	0.260	0.213	0.183	0.265
donor	0.209	0.242	0.298	0.131
acceptor	0.165	0.194	0.153	0.143

Table 3. PLS Statistics of 3D-QSAR Models for HER-2 Inhibitory Activities from Various Alignments for Benzylidene Malonitrile Tyrphostins

	alignments				
PLS statistics	**I**	**II**	**III**	**IV**	**V**
			CoMFA		
q^2	0.610	0.630	0.676	0.643	0.614
components	5	3	4	5	6
outliers	**28**, **37**, **40**	**17**, **28**, **30**, **37**, **38**	**39**	**25**	**17**, **27**, **28**, **36**
PRESS	0.562	0.528	0.523	0.554	0.541
r^2	0.953	0.855	0.939	0.955	0.945
s	0.194	0.330	0.227	0.198	0.205
F	139.126	65.906	130.996	117.532	87.969
contribution					
steric	0.449	0.492	0.546	0.492	0.470
electrostatics	0.551	0.508	0.454	0.508	0.530
			CoMSIA		
q^2	0.646	0.642	0.672	0.653	0.594
components	6	5	5	6	6
outliers	**19**, **28**, **37**	**17**, **28**, **37**	**37**, **38**	**18**, **21**	**7**, **17**, **27**, **28**, **36**
PRESS	0.555	0.520	0.533	0.560	0.563
r^2	0.974	0.964	0.969	0.967	0.920
s	0.152	0.164	0.169	0.172	0.250
F	202.369	183.54	200.896	127.663	57.569
contribution					
steric	0.131	0.114	0.151	0.189	0.148
electrostatics	0.297	0.330	0.287	0.210	0.331
hydrophobic	0.239	0.229	0.119	0.206	0.253
donor	0.216	0.219	0.231	0.210	0.125
acceptor	0.117	0.107	0.131	0.184	0.143

modes **A** and **B** are possible for these compounds in the ATP site of EGFR, although mode **B** is preferred. Thus the possibility of multiple binding modes for these dihydroxy compounds exists at the ATP site of EGFR.

HER-2 QSAR Models. All the alignments produced CoMFA QSAR models with reasonable PLS statistics. The PLS statistics are reported in Table 3 and the fitted curves of the CoMFA models are shown in Figure 4a–e. The CoMFA models from alignments **I** and **IV** are better than those from alignments **II**, **III**, and **V** in predicting the activities of compounds in the training set. The CoMFA model from alignment **I** performs the best in its ability to predict the activities of the test set. These CoMFA results have similarities as well as differences when compared to the CoMFA results from the EGFR data sets. In both the EGFR and HER-2 CoMFA analyses, alignments **III** and **IV** did not perform well. On the other hand, alignment **I** performed best in the HER-2 CoMFA models, in contrast to the EGFR CoMFA models where alignment **II** performed best. In this instance, the results suggest that the dihydroxy compounds prefer binding in mode **A** (Figure 1b) at the

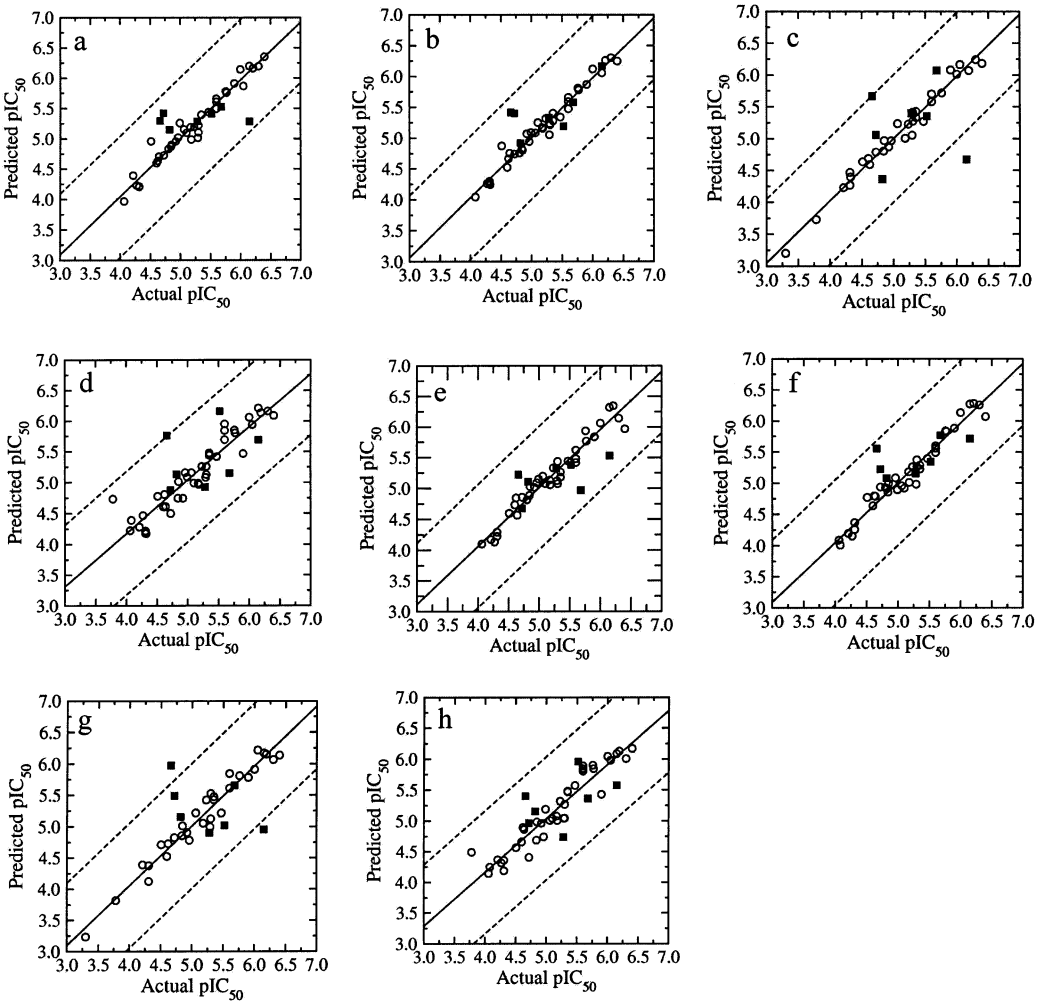


Figure 3. Curves for the prediction of the activities of the training set and test set compounds by QSAR models for inhibition of EGFR kinase by benzylidene malonitriles. Plots **a**, **b**, **c**, and **d** show predictions by CoMFA models using alignments **I**, **II**, **III** and **V**, respectively, while plots **e**, **f**, **g**, and **h** show predictions by CoMSIA models using alignments **I**, **II**, **III**, and **V**, respectively. The activity predictions for the training set are shown as open circles whereas filled squares indicate predictions for the test set. The solid line represents the regression fitted line of the training set whereas the dotted line shows the error limit of ±1 log unit about the corresponding fitted curve.

ATP site in HER-2. This is interesting, in that these dihydroxy compounds are selective inhibitors of EGFR kinase relative to HER-2 kinase (see Table 1). It appears that this selectivity may have to do, at least in part, with different mode preferences in binding of these compounds to the two kinases. Mode **B** appears to be the slightly preferred mode in EGFR, although mode **A** is suggested to be good as well. In HER-2, the situation is more clear-cut, where mode **A** is preferred over mode **B**.

All the alignments fared well in generation of CoMSIA QSAR models. The PLS statistics are shown in Table 3, and the corresponding fitted curves of the CoMSIA models are shown in Figure 4f−j. CoMSIA models from alignments **I** and **II** performed better than that from alignment **V** in the predictive ability toward both the training set and the test set compounds. Models from alignments **IV** and **V** fared the worst in their predictive ability of the test set even though their ability to predict the training set was comparable to models from alignment **I** and **II**. In agreement with the CoMFA models, the CoMSIA models show that alignment **III** and **IV** give the worst model in terms of predictions of the activities of the test set, regardless of the q^2 values. The CoMSIA models from alignments **I** and **II** are comparable in predicting the test set, but the fitted curve was better for the model from alignment **I**. Therefore, overall, the CoMSIA results agree with the CoMFA results in suggesting that most of the dihydroxy compounds prefer binding mode **A** at the ATP site of HER-2.

3D QSAR Contour Maps. The coefficients from CoMFA and CoMSIA QSAR models are usually plotted to generate contours maps that indicate regions in 3D space around the molecules where changes in the respective physicochemical properties are predicted to increase or decrease potency. The steric and electrostatic contours for the CoMFA models, along with hydrophobic and hydrogen bond contours of the CoMSIA models for EGFR inhibitory activity using alignments **I** and **II**, and some of interacting residues close to the contours in the ATP site are presented in Figure 5, and the corresponding contours for the 3D-QSAR models for HER-2 inhibitory activity are presented in Figure 6. The CoMFA and CoMSIA contours maps for models from a classical 3D-QSAR alignment strategy, **V**, are shown in Figure 7.

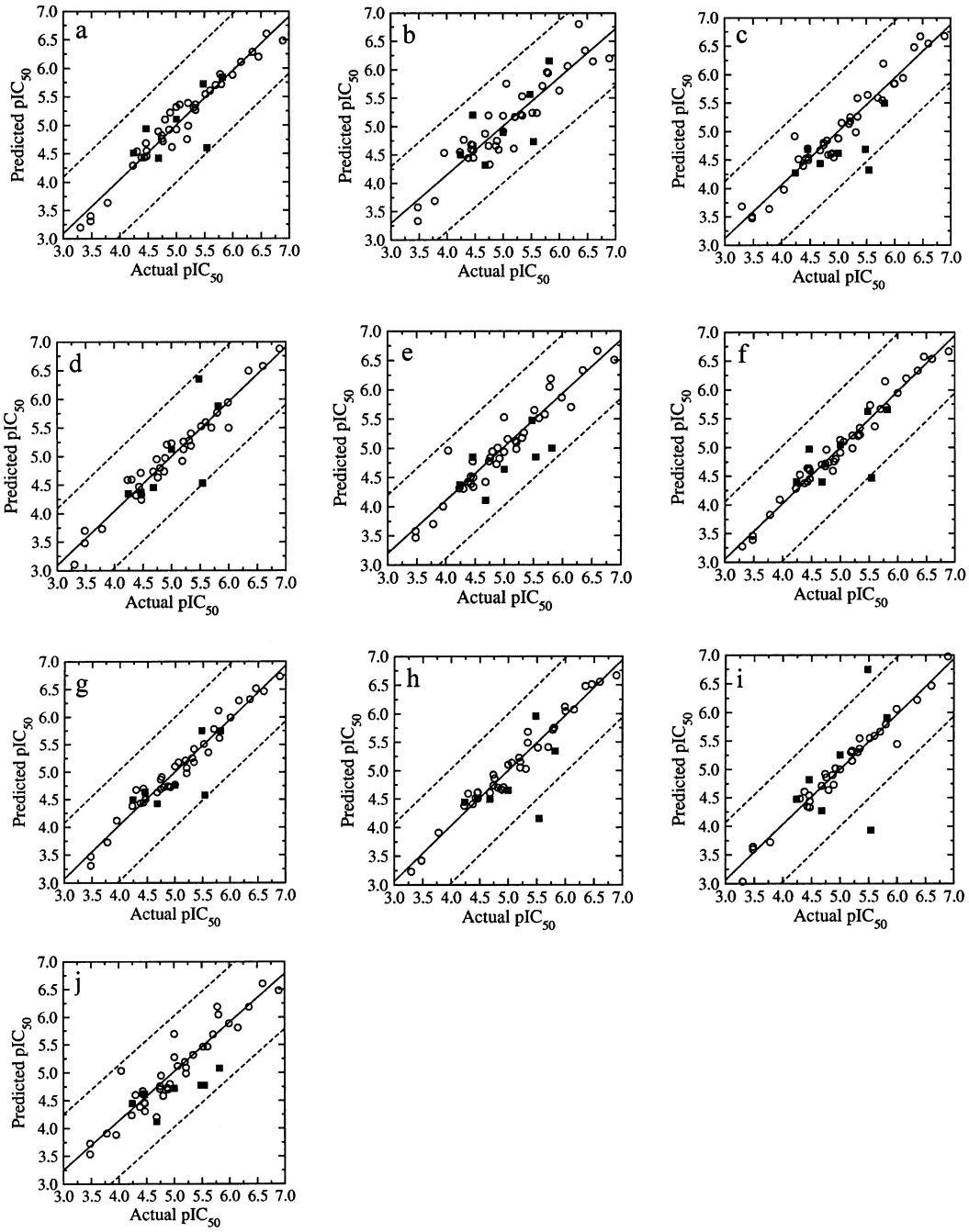


Figure 4. Curves for training set and test set activity predictions by QSAR models for inhibition of HER-2 kinase by benzylidene malonitriles. Plots **a**, **b**, **c**, **d**, and **e** show predictions by CoMFA models employing alignments **I**, **II**, **III**, **IV**, and **V**, respectively, while plots **f**, **g**, **h**, **i**, and **j** show predictions by CoMSIA models using alignments **I**, **II**, **III**, **IV**, and **V**, respectively. The activity predictions for the training set are shown as open circles whereas filled squares indicate predictions for the test set. The solid line represents the regression fitted line of the training set where as the dotted line shows the error limit of ±1 log unit about the corresponding fitted curve.

Only contours from the best two models from the receptor guided alignments (**I** and **II**) are shown and discussed for brevity.

Although in models from each kinase there are similarities in the appearance of the steric and electrostatic contour maps, there occur significant differences in comparing maps from the different kinases (compare Figure 5a,b for EGFR, and Figure 6a,b for HER-2). For example in alignment **I**, the CoMFA contours show more steric hindrance regions in the EGFR models than the HER-2 models (compare Figure 5a and 6a) whereas for alignment **II** the opposite is true (compare Figure 5b and 6b). This implies that alignment **I** will be preferred in HER-2 and alignment **II** in EGFR. In terms of electrostatics in the CoMFA models, there are electropositive favored regions at the R_2 substituent in the EGFR models that are absent in the HER-2 models (compare Figure 5a and 5b with Figure 6a and 6b). Around the R_3 substituent, the HER-2 models tend to have prominent electropositive favored contours, which is not the case in the EGFR models. This implies that substitution of electropositive groups at this region will enhance selectivity toward HER-2 inhibitory activity. The contours show general complimentarity with the receptor surface physicochemical properties (data not shown).

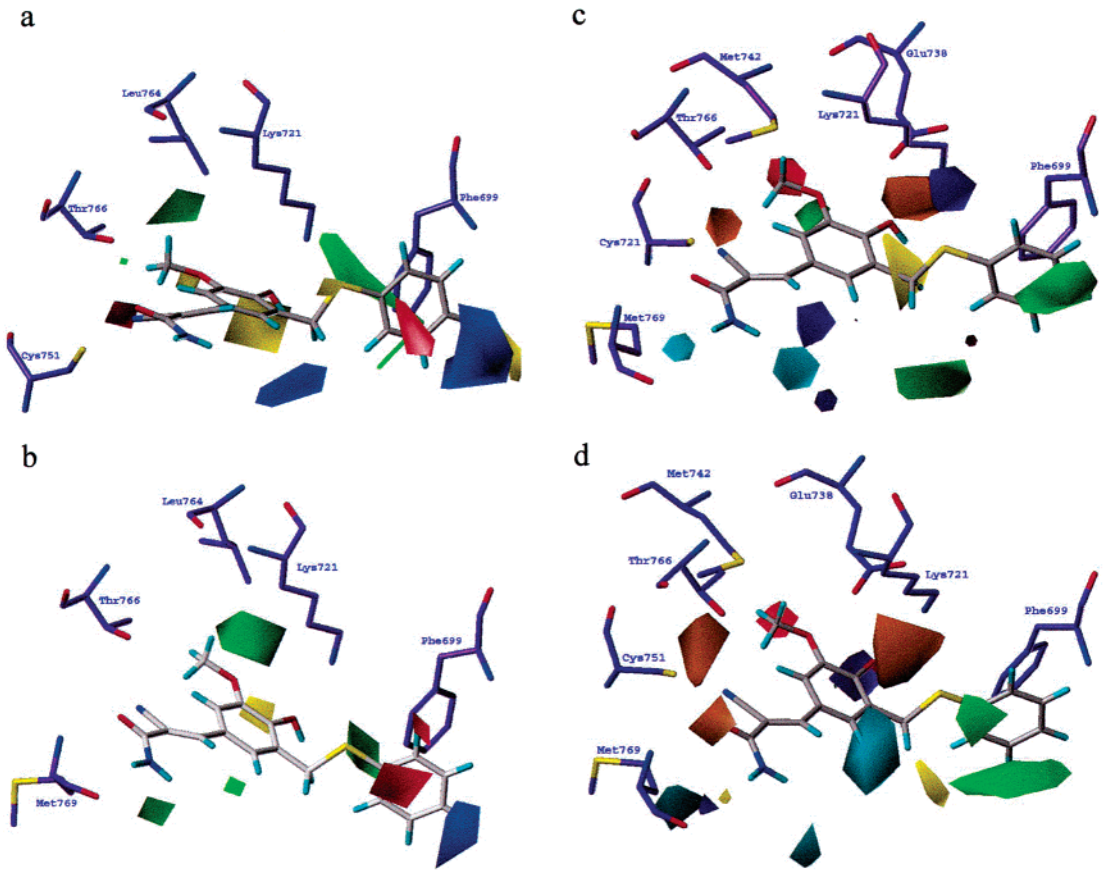


Figure 5. PLS coefficient*stdev contour maps of CoMFA and CoMSIA QSAR models of benzylidene malonitriles along with the interacting residues in the ATP site of EGFR. CoMFA contours of the models from alignment **I** and **II** are shown in **a** and **b**, respectively, while the corresponding CoMSIA contours are shown in **c** and **d**, respectively. Only the hydrophobic and the hydrogen bond contours are shown for the CoMSIA models. The green contours indicate sterically favored regions in the CoMFA models and indicate hydrophobic favored areas in the CoMSIA models. The yellow contours in the CoMFA models denote regions that are sterically unfavorable while in the CoMSIA models they show areas where hydrophobic groups decrease activity. The blue-colored areas favor electropositive groups and red-colored regions favor electronegative substituents. The cyan-colored contours show areas where a hydrogen bond donor is favored, and purple-colored contours indicate the opposite. Hydrogen bond acceptors are favored in areas indicated by orange contours whereas red contours are areas where hydrogen bond acceptors are disfavored.

The CoMSIA hydrophobic contours for alignments **I** and **II** for both EGFR (Figure 5c,d) and HER-2 (Figure 6c,d) are similar, with hydrophobic groups favored on the sulfur substituent. They also complement the receptor site surface physicochemical properties (data not shown). An additional hydrophobic-favored contour is seen near the R_1 substituent in the models derived from alignment **I** for EGFR (Figure 5c) and alignment **II** in the case of HER-2 (Figure 6d). This again suggests that different hydrophobicity requirements might contribute to the selectivity of the compounds in inhibiting the two kinases.

The hydrogen bond donor and acceptor contours from alignments **I** and **II** for EGFR and HER-2 bear an overall similarity (Figure 5c,d and Figure 6c,d). The hydrogen bond acceptor contours on the cyano and the carbonyl of the amide, and donor contours on the amide, indicate the importance of these sites in contributing to the activity. The model derived from alignment **II** in EGFR (Figure 5d) shows an additional hydrogen bond donor-favored contour close to the R_2 substituent, which is not seen in any other model from either the EGFR or HER-2. Since alignment **II** appears preferred over alignment **I** in EGFR CoMFA models, placing strong hydrogen bond donor groups in this location might enhance selectivity of compounds for EGFR.

The following describes the contours obtained from models based on alignment **V**. The steric contours for the CoMFA model of EGFR (Figure 7a) differ from those for HER-2 (Figure 7b). Steric bulk is favored around the amide region in the EGFR model whereas it is disfavored in the HER-2 model. In the HER-2 CoMFA model, steric bulk is favored around the sulfur (R_2) region. In the CoMSIA model, steric bulk is predicted to be favored around both the amide and the sulfur substituent in the EGFR model (Figure 7c) whereas it is favored only around the latter in the HER-2 model (Figure 7d). The electrostatic CoMFA contours for both EGFR and HER-2 (Figure 7a and 7b) are similar, with electropositive contours near the R_3 substituent and electronegative contours around the R_2 substituent. The CoMSIA electrostatic contours for EGFR show an electropositive area around R_3 in addition to that seen in the CoMFA model (compare Figures 7a and 7c) whereas in the model for HER-2 (Figure 7d), the electropositive contour near R_3 is absent. The CoMSIA model suggests that hydrophobic groups are favored around both the amide (R_3) and the sulfur (R_2) substituents in EGFR (Figure 7e) whereas it is only favored near the sulfur substituent in HER-2 (Figure 7f). Hydrogen bond donor favored regions are at similar locations as in the other alignments, but the acceptor-favored contours are predicted to be important

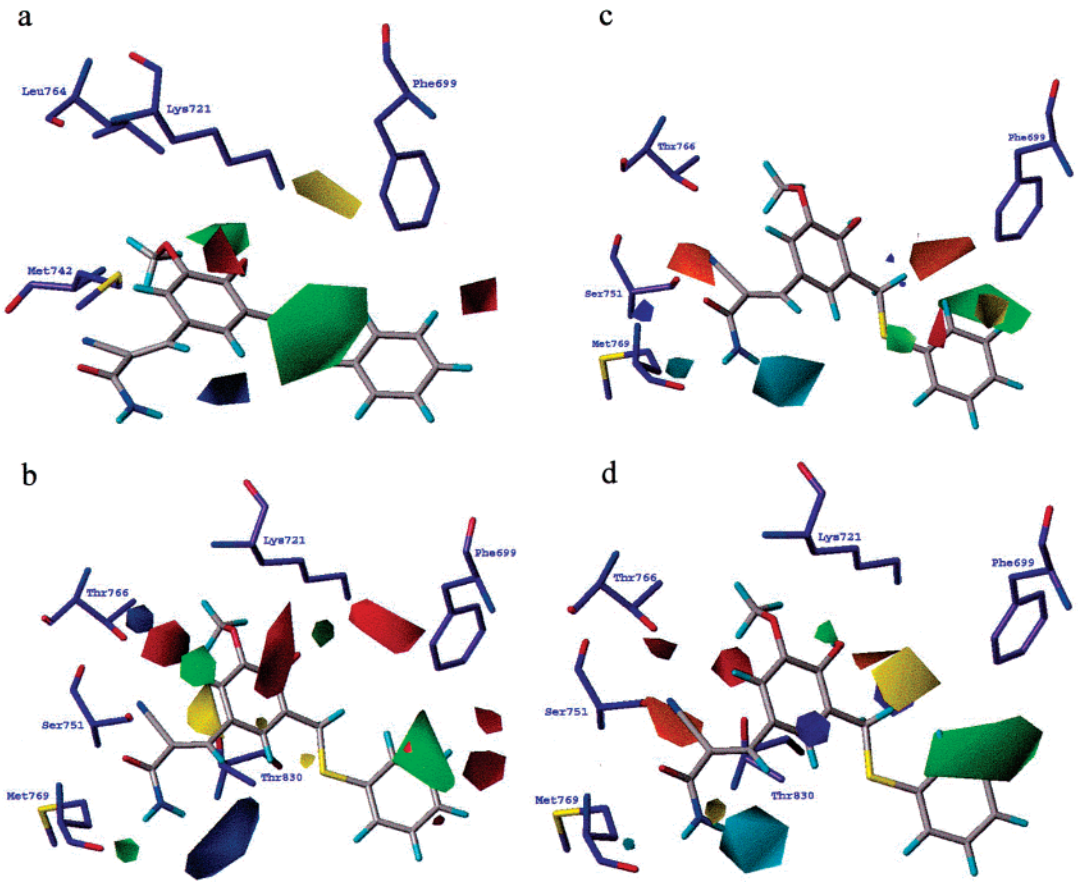


Figure 6. PLS coefficient*stdev contour maps of CoMFA and CoMSIA models of benzylidene malonitriles along with the interacting residues in the ATP site of HER-2. CoMFA contours of the models from alignments **I** and **II** are shown in **a** and **b**, respectively, while the corresponding CoMSIA contours are shown in **c** and **d**, respectively. Only the hydrophobic and the hydrogen bond contours are shown for the CoMSIA models. The green contours indicate sterically favored regions in the CoMFA models and indicate hydrophobic favored areas in CoMSIA models. The yellow contours in the CoMFA models denote regions that are sterically unfavorable whereas in the CoMSIA models they show areas where hydrophobic groups decrease activity. The blue-colored areas favor electropositive groups and red-colored regions favor electronegative substituents. The cyan-colored contours show areas where a hydrogen bond donor is favored and purple-colored contours denote areas where a hydrogen bond donor is disfavored. Hydrogen bond acceptors are favored in areas indicated by orange contours whereas red contours indicate the opposite.

only on the carbonyl of the amide. The QSAR predictive performance of models from this alignment was one of the worst and therefore much weight cannot be given to these suggestions unless they are in agreement with contours from alignments **I** and **II**.

Conclusion

In this study, both docking and 3D QSAR analyses have been employed to explore the binding modes of benzylidene malonitrile tyrphostins at ATP sites of EGFR and HER-2 kinases and to gain insights into possible factors contributing to the selectivity of benzylidene malonitriles as inhibitors of EGFR or HER-2 kinase activity. In summary, multiple binding modes have been suggested for the dihydroxy compounds, which differ from the binding mode preference of the methoxy compounds, which are suggested to prefer a single binding mode **C**, as suggested by the best QSAR models (from alignments **I** and **II**). The results also indicate that the ATP site of EGFR is more accommodating of the multiple binding modes of the dihydroxy compounds than does that of HER-2. This difference in binding mode preferences for the dihydroxy compounds might contribute in part to the selectivity of these compounds as better inhibitors of EGFR than HER-2. This implies that even a conservative amino acid change, i.e., from Cys751 in EGFR to Ser in HER-2 in the ATP-binding site that would not be considered as a dramatic change can have a marked effect on inhibitor binding at this site. The selectivity of the methoxy compounds toward HER-2 does not appear to be related to binding mode differences, but rather may have to do with steric, electrostatic, hydrophobicity, and hydrogen bond donor effects. In binding mode **C**, the cyano moiety is juxtaposed between Thr766 and Cys751 in EGFR or Ser751 in HER-2 (Figure 1g). This creates the possibility for hydrogen bonding between the cyano and Thr766 and/or Cys/Ser751 in the methoxy compounds. This might be explained by the increase in hydrogen bonding strength that could result from changing SH in Cys751 to OH in Ser751.

The selectivity of specific compounds in the dihydroxy set, which bind in mode **A** or mode **B**, such as compound **13**, and the methoxy set and compound **41**, which bind only in mode **C**, may be partly attributed to hydrogen bond strength differences relating to the Cys751 SH and Ser751 OH in EGFR and HER-2, respectively. In the event that during binding, water molecules hydrogen bonding to the side chain of residue 751 have to be displaced, it will be easier in EGFR than HER-2. This

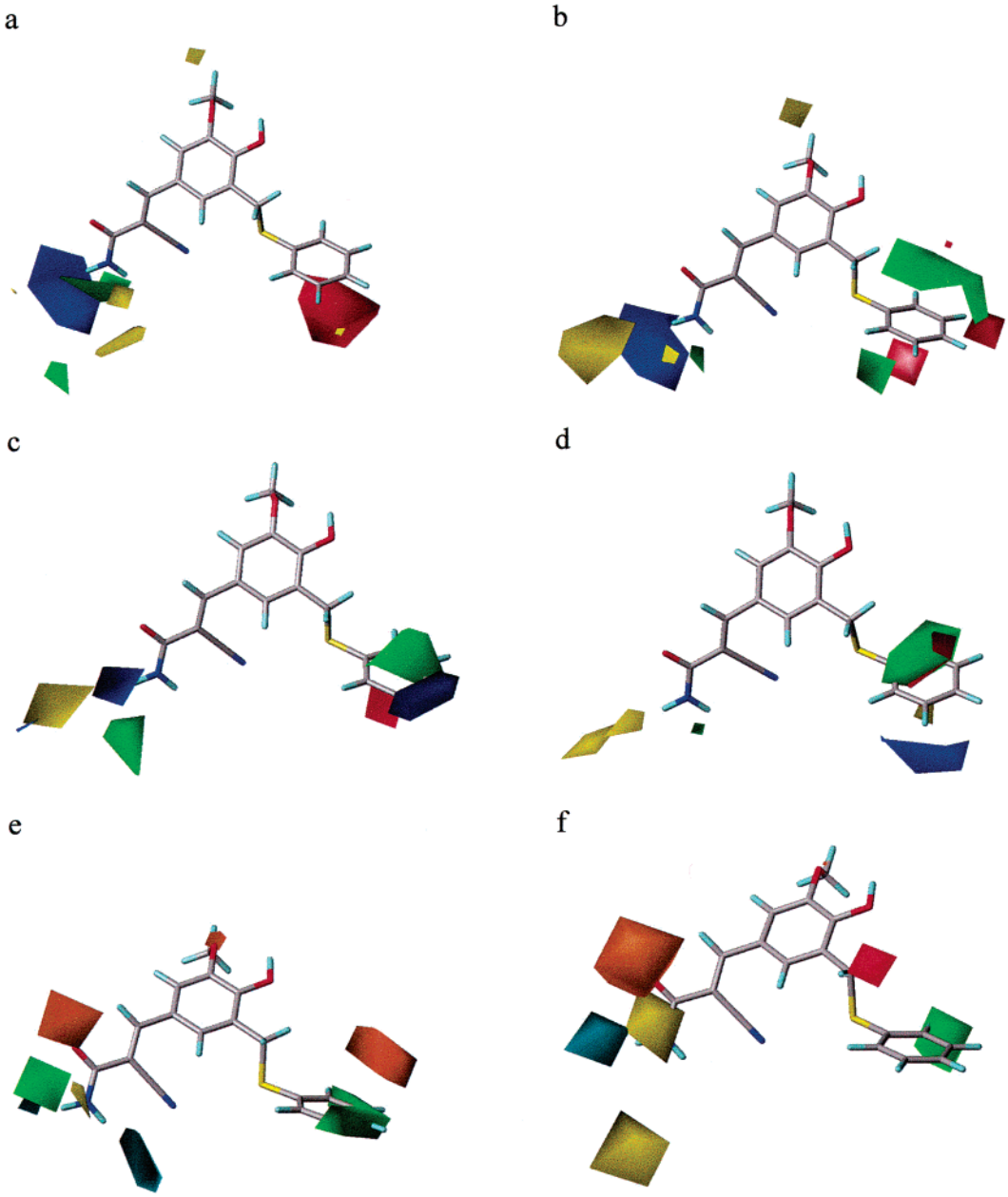


Figure 7. PLS contours from 3D QSAR models for EGFR and HER-2 inhibition using alignment **III**. Steric and electrostatic CoMFA contours from models for EGFR and HER-2 inhibition are shown in **a** and **b**, respectively while corresponding CoMSIA contours from models for EGFR and HER-2 are shown in **c** and **d**, respectively. The green contours indicate sterically favored regions whereas the yellow contours denote regions that are sterically unfavorable. The blue-colored areas favor electropositive groups and red-colored regions favor electronegative substituents. Hydrophobic and hydrogen bond donor/acceptor CoMSIA contours for EGFR and HER-2 are shown in **e** and **f**, respectively. The green contours indicate regions where hydrophobic groups are favored whereas the yellow contours denote regions that disfavor hydrophobicity. The cyan and purple contours indicate favorable and unfavorable regions for hydrogen bond donors, respectively, whereas the orange and red contours are indicative of favorable and unfavorable positions for hydrogen bond acceptors, respectively.

displacement, we speculate, will involve the OH groups of the dihydroxy compounds docked in the mode **A**. Since our models suggest that the dihydroxy compounds prefer to bind in mode **A** in HER-2, this may partly explain the lower inhibitory activity of these compounds against HER-2. For the methoxy compounds such as **27**, as well as compound **41**, which adopt only mode **C**, the situation involving the change of SH to OH in residue 751 is different. In this case, it is the strength of the possible hydrogen bond between the cyano group and the SH or OH that may confer selectivity (see Figure 1g). This would favor HER-2 selectivity since the CN···H−O hydrogen bond should be stronger than the CN···H−S hydrogen bond.

It should be noted, however, that the dynamic effects and other hydration pattern changes that occur at the receptor site upon binding that would also influence potency and selectivity, were not considered in our present models. Further, some of the molecules may be mixed competitive inhibitors as was hinted by Osherov et al.,[21] and this would cause differences that the present model was not able to address as well. What is clear from the present study is that docking in conjunction with 3D QSAR studies has provided some rationale

for the selectivity in the inhibitory effects of this series of benzylidene malonitrile tyrphostins against EGFR and HER-2 kinase activities.

Another point that this study brings out is that composite alignments in which different subalignments are combined can be useful for 3D QSAR studies if they are receptor-guided. The QSAR results also show, as indicated by the spread in both the fitted predictions and test set predictions, that a traditional atom by atom alignment such as the MULTIFIT method of SYBYL will not be appropriate for these molecules, which reflects the fact that they may assume different binding modes at the kinase active site as our docking results suggest. Last, the docking has suggested that incorporating the cyano and carbonyl moieties into a ring system might improve their binding affinity, which is being explored in our laboratory.

References

(1) Seymore, L. Novel Anti-Cancer Agents in Development: Exciting Prospects and New Challenges. *Cancer Treat. Revs*. **1999**, *25*, 301−312.

(2) Buolamwini, J. K. Novel Molecular Targets for Cancer Drug Discovery In*The Molecular Basis of Human Cancer*; Coleman, W. B.; Tsongalis, G. J., Eds.; Humana Press: Totowa, NJ, 2002; pp 521−540.

(3) Adjei, A. A. Blocking Oncogenic Ras Signaling for Cancer Therapy. *J. Natl. Cancer Inst*. **2001**, *93*, 1062−1074.

(4) Druker, B. J. STI571 (Gleevec) as a Paradigm for Cancer Therapy. *Trends Mol. Med*. **2002**, *8*, S14−S18.

(5) Fantl, W. J.; Johnson, D. E.; Williams, L. T. Signaling by Receptor Tyrosine Kinases. *Annu. Rev. Biochem*. **1993**, *62*, 453−481.

(6) Ullrich, A. and Schlessinger, J. Signal Transduction by Receptors with Tyrosine Kinase Activity. *Cell* **1991**, *61*, 203−212.

(7) Kolibaba, K. S.; Druker, B. J. Protein Tyrosine Kinases and Cancer. *Biochim. Biophys. Acta* **1997**, *1333*, F217−F248.

(8) Aaronson, S. A. Growth Factors and Cancer. *Science* **1991**, *254*, 1146−1152.

(9) de Bono, J. S.; Rowinsky, E. K. The ErbB Receptor Family: A Therapeutic Target for Cancer. *Trends Mol. Med*. **2002**, *8*, S19−S26.

(10) Workman, P. The Potential for Molecular Oncology to Define New Drug Targets In *New Molecular Targets for Cancer Chemotherapy*; Kerr, D. J.; Workman, P., Eds., CRC Press: Boca Raton, FL, 1994; pp 1−44.

(11) Sauseville, E. A.; Longo, D. L. Growth Factors and Growth Factor Inhibitors In *Cancer Therapeutics: Experimental and Clinical Agents*; Teicher, B., Eds., Humana Press: Totowa, NJ, 1994; pp 337−370.

(12) Slamon, D. J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; Baselga, J.; Norton, L. Use of Chemotherapy Plus a Monoclonal Antibody against HER-2 for Metastatic Breast Cancer that Overexpresses HER-2. *N. Engl. J. Med*. **2001**, *344*, 783−792.

(13) Kubinyi, H. Eds. *3D QSAR in Drug Design: Theory, Methods and Applications*; ESCOM Science Publishers: Leiden, 1993.

(14) Zhu, L. L.; Hou, T. J.; Chen, L. R.; Xu, X. J. 3D QSAR Analyses of Novel Tyrosine Kinase **I**nhibitors Based on Pharmacophore Alignment. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1032−1040.

(15) Cramer, III, R. D.; 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.

(16) 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.

(17) Gazit, A.; Yaish, P.; Gilon, C.; Levitzki, A. Tyrphostins 1: Synthesis and Biological Activity of Protein Tyrosine Kinase Inhibitors *J. Med. Chem*. **1989**, *32*, 2344−2352.

(18) Gazit, A.; Osherov, N.; Posner, I.; Yaish, P.; Poradosu, E.; Gilon, C.; Levitzki, A. Tyrphostins 2: Heterocyclic and α-Substituted Benzomalonitrile Tyrphostins as Potent Inhibitors of EGF Receptor and *HER-2/neu* Tyrosine Kinases. *J. Med. Chem*. **1991**, *34*, 1896−1907.

(19) Gazit, A.; Osherov, N.; Posner, I.; Bar-Sinai, Gilon, C.; Levitzki, A. Tyrphostins 3: Structure−Activity Relationship Studies of α-Substituted Benzylidene Malonitrile 5-*S*-Aryltyrphostins. *J. Med. Chem*. **1993**, *36*, 3556−3564.

(20) Levitzki, A.; Gazit Aviv. Tyrosine Kinase Inhibition: An Approach to Drug Development. *Science* **1995**, *267*, 1782−1788.

(21) Osherov, N., Gazit, A., Gilon, C. and Levistzki, A. Selective Inhibition of Epidermal Growth Factor and *HER-2/Neu* Receptors by Tyrphostins. *J. Biol. Chem*. **1993**, *268*, 11134−11142.

(22) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol*. **1997**, *267*, 727−748.

(23) Viswanadhan, V. N.; Ghose, A. K.; Revenkar, G. R.; Robins, R. Atomic and Physichochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure−Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and their Application for an Automated Superposition of Certain Naturally Occurring Antibiotics. *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 163−172.

(24) Taxler, P., Green, J., Mett, H., Sequin, U., Furet P. Use of a Pharmacophore Model for the Design of EGFR Tyrosine Kinase Inhibitors: Isoflavones and 3-Phenyl-4(1*H*)-quinolones. *J. Med. Chem*. **1999**, *42*, 1018−1026.

(25) Pierce, A. C., Sandretto, K. L., Bemis G. W. Kinase Inhibitors and the Case for CH···O Hydrogen Bonds in Protein−ligand Binding. *Proteins: Struct. Funct. Genet*. **2001**, *49*, 567−576.

(26) Blencke, S., Ullrich, A., Daub, H. Mutation of Threonine 766 in the Epidermal Growth Factor Receptor Reveals a Hotspot for Resistance Formation against Selective Tyrosine Kinase Inhibitors *J. Biol. Chem*. **2003**, *278*, 15435−15440.

(27) Golbraikh, A., Tropsha, A. Beware of q^2! *J. Mol. Graphics Modeling* **2002**, *20*, 269−276.

JM030065N