

Imidazole-containing farnesyltransferase inhibitors: 3D quantitative structure–activity relationships and molecular docking

Aihua Xie · Srinivas Odde · Sivaprakasam Prasanna · Robert J. Doerksen

Received: 10 December 2008 / Accepted: 2 May 2009 / Published online: 29 May 2009
© Springer Science+Business Media B.V. 2009

Abstract One of the most promising anticancer and recent antimalarial targets is the heterodimeric zinc-containing protein farnesyltransferase (FT). In this work, we studied a highly diverse series of 192 Abbott-initiated imidazole-containing compounds and their FT inhibitory activities using 3D-QSAR and docking, in order to gain understanding of the interaction of these inhibitors with FT to aid development of a rational strategy for further lead optimization. We report several highly significant and predictive CoMFA and CoMSIA models. The best model, composed of CoMFA steric and electrostatic fields combined with CoMSIA hydrophobic and H-bond acceptor fields, had $r^2 = 0.878$, $q^2 = 0.630$, and $r^2_{\text{pred}} = 0.614$. Docking studies on the statistical outliers revealed that some of them had a different binding mode in the FT active site based on steric bulk and available active site space, explaining why the predicted activities differed from the experimental activities.

Keywords Farnesyltransferase inhibitors · CoMFA · CoMSIA · Docking · Imidazoles · Outliers

Introduction

Farnesyltransferase (FT) is considered a promising target of anticancer chemotherapy [1–8]. In a piggy-back approach to the search for novel antimalarials, it has been shown that FT inhibitors could cause significant lysis of *Plasmodium falciparum* (the causative agent of critical malaria) in infected cells, and this was associated with a decrease in farnesylated proteins [9–14]. The FT inhibitors showed substantial selectivity for the malarial parasites over mammalian cells [9].

As a zinc metalloenzyme, FT catalyzes the transfer of a farnesyl residue from farnesylpyrophosphate to the thiol of a cysteine side chain of proteins which carry a *CaaX* sequence at the C terminus [15], in which *C* is the cysteine whose side chain is being farnesylated, *a* is an aliphatic amino acid and *X* is an alanine, glutamine, serine or methionine. The first class of FT inhibitors was *CaaX* peptidomimetics which contained a thiol group to coordinate to the FT zinc atom. Then, the oxidatively unstable thiol group was replaced by another zinc binding moiety, an imidazole group. By adding more hydrophobic factors, researchers found that the imidazole group could also be replaced by aryl residues lacking the ability to coordinate to metal atoms without losing too much of their FT inhibitory activity [16]. High-throughput screening played an important role in the initiation of several new non-thiol, non-peptidic, imidazole or non-imidazole containing series of compounds [5, 16, 17]. Potent compounds which have entered clinical trials, such as SCH66336 (lonafarnib),

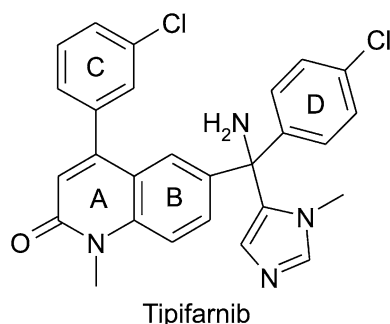
Electronic supplementary material The online version of this article (doi:10.1007/s10822-009-9278-z) contains supplementary material, which is available to authorized users.

A. Xie · S. Odde · S. Prasanna · R. J. Doerksen
Department of Medicinal Chemistry, University of Mississippi,
University, MS 38677-1848, USA

R. J. Doerksen (✉)
Research Institute for Pharmaceutical Sciences, University of
Mississippi, University, MS 38677-1848, USA
e-mail: rjd@olemiss.edu

Present Address:

S. Odde
Department of Medicinal Chemistry, University of Southern
California, Los Angeles, CA 90089, USA



Scheme 1 Tipifarnib, with rings labeled from A–D

R115777 (tipifarnib), BMS-214662, and L-778123, share quite a few common binding characteristics [16, 18, 19].

Mainly based on structural modification of tipifarnib (Scheme 1), one of the most potent and selective non-thiol FT inhibitors in clinical trials, Abbott initiated a highly diverse series of imidazole-containing FT inhibitors through deletion of ring B and transposition of ring D [20–28]. In this work we studied 192 of the Abbott compounds and their FT inhibitory activities (IC_{50}) using CoMFA [29], CoMSIA [30] and docking [31] in order to gain understanding of the interaction between this series of compounds and FT. There have been some reports of 3D-QSAR analysis of FT inhibitors [32–36], but none have treated such a large data set of compounds. There are reports in the literature describing the application of other QSAR methods [37–41] and docking [42–44] to study FT inhibition of other classes of inhibitors. We report on several highly significant and predictive 3D-QSAR models and on the docking binding mode of several outliers. By comparing the CoMFA/CoMSIA contour maps with the ligand-receptor active site interactions, we shed light on spatial and property features important for FT inhibition.

Materials and methods

Data set

Table 1 shows the structures of the 192 compounds used in this study and their observed activities (pIC_{50}) [20–22, 25–28]. The set is composed of 54 pyridones, 55 pyridyl-containing methyl ethers and 83 biphenyl methyl ethers. Figure 1 shows the distribution of pIC_{50} for the whole data set, ranging from 6.47 to 10.44 with a mean of 8.51 and a variance of 0.74.

Putative bioactive conformation and alignment strategy

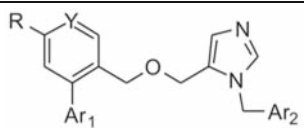
Identification of a putative bioactive conformation template and an effective alignment strategy are key steps for

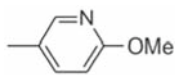
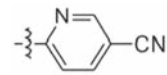
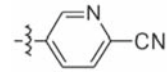
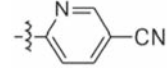
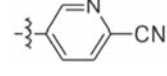
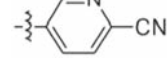
achieving robust models in 3D-QSAR. For this large and highly diverse compound data set, we used two X-ray crystal structures as references: human protein FT complexed with farnesyl diphosphate (FPP) and tipifarnib (R115777) (PDB: 1SA4) [24] and rat FT complexed with a cyano-containing imidazole inhibitor (compound **55**, PDB: 1NI1) [28]. The binding mode conformations of tipifarnib and **55** are shown in Fig. 2, along with the conformation of **2** docked into 1NI1. The main features of the binding mode of the compounds are the spatial arrangement of the imidazole and three other aromatic rings (A, C, D) which are connected by a linker and form a “U” shaped conformation. The imidazole coordinates to the zinc atom, and the three other rings bind into different hydrophobic pockets of the FT active site. Considering also docking studies of ten other representative compounds in the data set, we decided to use the conformation of **55**, extracted from the X-ray co-crystal structure (PDB: 1NI1) and then locally minimized in Sybyl 7.2 [45], using the MMFF94 force field with partial charges also assigned by the MMFF94 force field, as the template (Fig. 3) to construct the whole data set.

We adopted two strategies for alignment in this project. One, which we call centroid alignment, uses the centroids of the imidazole and of the other three rings along with the oxygen or nitrogen in the linker as reference atoms; the other is database alignment based on the maximum structure common to the whole data set, which consists of just the imidazole ring and linker (cf. Fig. 3). All structures were then relaxed to local minima using the MMFF94 force field, as was done for the template **55**.

3D-QSAR methodology

Molecular modeling, CoMFA and CoMSIA analysis were performed using SYBYL 7.2 [45]. CoMFA calculates steric and electrostatic properties for a set of finely aligned compounds according to Lennard-Jones and Coulomb potentials, respectively [29]. CoMSIA calculates similarity indices around the aligned molecules, with similarity expressed in terms of five different physicochemical properties: steric occupancy, electrostatic field, local hydrophobicity, and hydrogen-bond donor and acceptor properties [30]. Visualizations of CoMFA or CoMSIA models provide a set of contour maps indicating modifications favorable for lead optimization, according to the physicochemical properties embedded in the models. For CoMFA and CoMSIA QSAR model development, the methodology was the same as described in our previous work [36]. Each CoMFA or CoMSIA field is a set of descriptors calculated at points in 3D space around the aligned ligands. Using CoMFA and CoMSIA as implemented in Sybyl software, it is possible to select one or two

Table 1 Structure and activity (IC_{50} in nM) of the data set


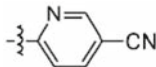
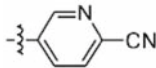
Comp	Ar ₁	R	Y	Ar ₂	IC ₅₀	pIC ₅₀
1	3-Cl-Ph	Cl	CH	4-CN-Ph	0.62	9.21
2	3-Cl-Ph	CN	CH	4-CN-Ph	0.37	9.43
3	3-Cl-Ph	CN	CH	4-CH ₃ SO ₂ -Ph	96	7.02
4	3-OCH ₃ -Ph	CN	CH	4-Cl-Ph	2.2	8.66
5	3-OEt-Ph	CN	CH	4-Cl-Ph	1.4	8.85
6	4-OEt-Ph	CN	CH	4-CN-Ph	1.2	8.92
7	3-OCF ₃ -Ph	CN	CH	4-CN-Ph	0.49	9.31
8	4-OCF ₃ -Ph	CN	CH	4-CN-Ph	0.62	9.21
9	3-AcNH-Ph	CN	CH	4-CN-Ph	13	7.89
10	4-COCH ₃ -Ph	CN	CH	4-CN-Ph	1.3	8.89
11	4-t-Bu-Ph	CN	CH	4-CN-Ph	0.44	9.36
12	4-CF ₃ -Ph	CN	CH	4-CN-Ph	4.3	8.37
13		CN	CH	4-CN-Ph	7.6	8.12
14	3,4-di-Cl-Ph	CN	CH	4-CN-Ph	0.37	9.43
15	3,4-di-Me-Ph	CN	CH	4-CN-Ph	0.60	9.22
16	3,4-OCH ₂ O-Ph	CN	CH	4-Cl-Ph	8.3	8.08
17	2,3-OCH ₂ O-Ph	CN	CH	4-CN-Ph	0.81	9.09
18	2-Naphthyl	CN	CH	4-CN-Ph	0.92	9.04
19	1-Naphthyl	CN	CH	4-CN-Ph	1.3	8.89
20	3-OCH ₃ -Ph	CN	CH	4-CN-Ph	0.46	9.34
21	3-OEt-Ph	CN	CH	4-CN-Ph	0.27	9.57
22	3,4-OCH ₂ O-Ph	CN	CH	4-CN-Ph	0.1	10.00
23	3-Cl-Ph	CN	N	4-CN-Ph	4.3	8.37
24	3-Cl-Ph	CN	CH		1.7	8.77
25	3-Cl-Ph	CN	CH		3.0	8.52
26	3-OCH ₃ -Ph	CN	CH		4.4	8.36
27	3-OCH ₃ -Ph	CN	CH		1.5	8.82
28	3-OEt-Ph	CN	CH		0.58	9.24

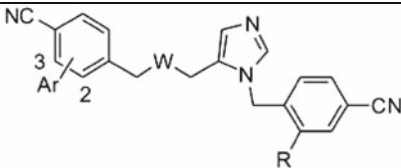
fields from CoMFA and combine them with fields from CoMSIA in a single model. The PLS method to generate a model is then applied in the same way as for CoMFA or CoMSIA separately. We attempted this, selecting steric and electrostatic fields from CoMFA and other fields, such

as hydrophobic, hydrogen bond acceptor and/or hydrogen bond donor from CoMSIA and then formed combined CoMFA/CoMSIA models.

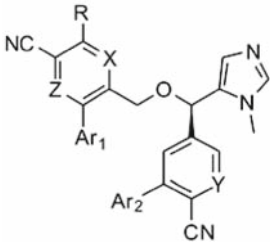
To evaluate the robustness of the QSAR models, we used the following definitions for the squared correlation


Table 1 continued

29	3,4-OCH ₂ O-Ph	CN	CH		4.7	8.32
30	3,4-OCH ₂ O-Ph	CN	CH		3.0	8.52



Comp	Ar	W	R	IC ₅₀	pIC ₅₀
31	2-(3-Cl-Ph)	O	3-Cl-Ph	10	8.0
32	2-(3,4-OCH ₂ O-Ph)	O	3,4-OCH ₂ O-Ph	340	6.47
33	2-(3,4-OCH ₂ O-Ph)	O	1-Naphthyl	6.8	8.17
34	H	O	1-Naphthyl	0.96	9.02
35	3-(1-Naphthyl)	O	H	4.8	8.32
36	3-(1-Naphthyl)	NH	H	6.5	8.19



Comp	Ar ₁	X	Y	Z	Ar ₂	R	IC ₅₀	pIC ₅₀
37	–	CH	CH	CH		–	2.0	8.70
38	3-OCH ₃ -Ph	CH	CH	CH	–	–	0.62	9.21
39	–	N	CH	CH	1-Naphthyl	–	0.43	9.37
40	3-Cl-Ph	CH	N	CH	–	–	0.75	9.12
41	3-OMe-Ph	CH	N	CH	–	–	0.88	9.06
42	3-OCF ₃ -Ph	CH	N	CH	–	–	0.40	9.40
43	3,4-OCH ₂ O-Ph	CH	N	CH	–	–	0.84	9.08
44	3-Cl-4-F-Ph	CH	N	CH	–	–	1.0	9.00
45	3,5-di-F-Ph	CH	N	CH	–	–	2.0	8.70
46	3,5-di-Cl-Ph	CH	N	CH	–	–	0.64	9.19
47	1-Naphthyl	CH	N	CH	–	–	1.7	8.77
48	–	N	CH	CH	–	3-Cl-Ph	12	7.92
49	–	N	CH	CH	–	4-Methyl-piperazin-1-yl	58	7.24
50	–	N	CH	CH	–	2-Naphthyl	8.2	8.09
51	–	N	CH	CH	–	3,5-di-Cl-Ph	8.5	8.07
52	–	N	CH	CH	–	4-Cl-Ph	25	7.60

coefficient, r^2 ; standard error of estimate, SEE; F statistic; predictive residual sum of squares, S_{PRESS} ; and cross-validated r^2 , q^2 (1–6):

$$r^2 = 1 - \frac{\sum (y_{\text{calc}} - y_{\text{act}})^2}{\sum (y_{\text{act}} - y_{\text{mean}})^2} \quad (1)$$

Table 1 continued


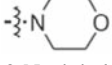
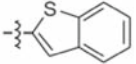
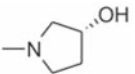

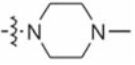
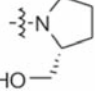
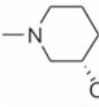
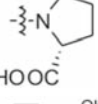
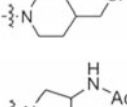
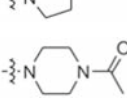
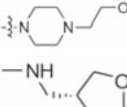
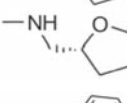
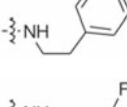
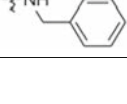

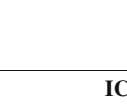
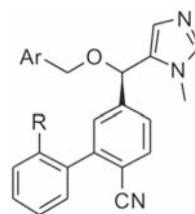
53	–	N	CH	CH	–	3-CN-Ph	68	7.17
54	3-Cl-Ph	N	CH	CH	–	H	0.92	9.04
55	3-Cl-Ph	N	CH	CH	–	Cl	0.90	9.05
56	3-Cl-Ph	N	CH	CH	–	4-Boc-piperazin-1-yl	0.99	9.0
57	3-Cl-Ph	N	CH	CH	–	4-Acyl-piperazin-1-yl	0.98	9.01
58	3-Cl-Ph	N	CH	CH	–	4-Me-piperazin-1-yl	0.57	9.24
59	3-Cl-Ph	N	CH	CH	–	4-OH-piperidin-1-yl	0.51	9.29
60	3-Cl-Ph	N	CH	CH	–	4-Carboxamoyl-piperidin-1-yl	0.73	9.14
61	3-Cl-Ph	N	CH	CH	–	Morpholin-4-yl	0.25	9.60
62	Ph	CH	CH	N	–	–	3.9	8.41
63	3-Cl-Ph	CH	CH	N	–	–	0.82	9.09
64	3-OCF ₃ -Ph	CH	CH	N	–	–	0.18	9.74
65	3,4-OCF ₂ O-Ph	CH	CH	N	–	–	0.18	9.74
66	4-OCF ₃ -Ph	CH	CH	N	–	–	0.51	9.29
67	3,5-di-F-Ph	CH	CH	N	–	–	0.71	9.15
68	3,5-di-Cl-Ph	CH	CH	N	–	–	0.45	9.35
69		CH	CH	N	–	–	44	7.36
70		CH	CH	N	–	–	1.9	8.72
71	2-Naphthyl	CH	CH	N	–	–	0.16	9.80
72	–	CH	CH	CH	o-Tolyl	–	4.8	8.32
73	–	CH	CH	CH	2-CF ₃ -Ph	–	0.8	9.10
74	–	CH	CH	CH	2-CHO-Ph	–	1.9	8.72
75	–	CH	CH	CH	2-HOCH ₂ -Ph	–	5.5	8.26
76	–	CH	CH	CH	2-COOH-Ph	–	14	7.85
77	–	CH	CH	CH	2,4-di-Cl-Ph	–	47	7.33
78	–	CH	CH	CH	2,3-di-Cl-Ph	–	1.5	8.82
79	–	CH	CH	CH	2,3-di-Me-Ph	–	2.1	8.68
80	–	CH	CH	CH		–	160	6.80
81	–	CH	CH	CH	1-Naphthyl	–	0.2	9.70
82	–	N	CH	CH	2-CF ₃ -Ph		3.5	8.46
83	–	N	CH	CH	2-CF ₃ -Ph		1.3	8.89
84	–	N	CH	CH	2-CF ₃ -Ph		1.2	8.92
85	–	N	CH	CH	2-CF ₃ -Ph		2.0	8.70

Table 1 continued

86	—	N	CH	CH	2-CF ₃ -Ph		1.3	8.89
87	—	N	CH	CH	2-CF ₃ -Ph		3.6	8.44
88	—	N	CH	CH	2-CF ₃ -Ph		1.8	8.74
89	—	N	CH	CH	2-CF ₃ -Ph		1.7	8.77
90	—	N	CH	CH	2-CF ₃ -Ph		2.6	8.59
91	—	N	CH	CH	2-CF ₃ -Ph		1.3	8.89
92	—	N	CH	CH	2-CF ₃ -Ph		7.7	8.11
93	—	N	CH	CH	2-CF ₃ -Ph		7.5	8.12
94	—	N	CH	CH	2-CF ₃ -Ph		9.1	8.04
95	—	N	CH	CH	2-CF ₃ -Ph		10	8.00



Comp	Ar	R	IC ₅₀	pIC ₅₀
96	Ph	CH ₃	84	7.08
97	o-Tolyl	CH ₃	54	7.27
98	m-Tolyl	CH ₃	61	7.21
99	p-Tolyl	CH ₃	54	7.27
100	p- ⁱ Pr-Ph	CH ₃	51	7.29
101	p- ^t Bu-Ph	CH ₃	87	7.06
102	p-Ph-Ph	CH ₃	13	7.89
103	1-Naphthyl	CH ₃	17	7.77
104	8-Quinolyl	CH ₃	59	7.23
105	p-MeO-Ph	CH ₃	59	7.23
106	m-F ₃ CO-Ph	CH ₃	8.3	8.08
107	p-F ₃ CO-Ph	CH ₃	16	7.80

Table 1 continued

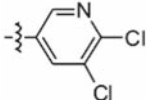
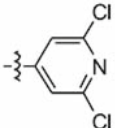
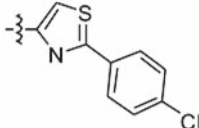
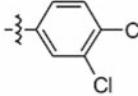
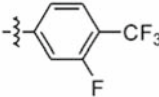
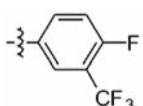
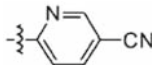
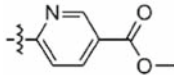
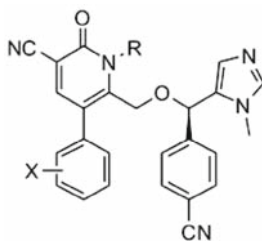
108	m-F ₃ C-Ph	CH ₃	8.1	8.09
109	p-F ₃ C-Ph	CH ₃	12	7.92
110	o-CN-Ph	CH ₃	62	7.21
111	m-CN-Ph	CH ₃	18	7.74
112	p-CN-Ph	CH ₃	4	8.40
113	o-F-Ph	CH ₃	70	7.15
114	m-F-Ph	CH ₃	38	7.42
115	p-F-Ph	CH ₃	32	7.49
116	o-Cl-Ph	CH ₃	48	7.32
117	m-Cl-Ph	CH ₃	19	7.72
118	p-Cl-Ph	CH ₃	11	7.96
119	m-Br-Ph	CH ₃	13	7.89
120	p-Br-Ph	CH ₃	12	7.92
121	m-I-Ph	CH ₃	20	7.70
122	p-I-Ph	CH ₃	13	7.89
123	o-NO ₂ -Ph	CH ₃	33	7.48
124	p-NO ₂ -Ph	CH ₃	7.5	8.12
125	m-CO ₂ CH ₃ -Ph	CH ₃	61	7.21
126	p-CO ₂ CH ₃ -Ph	CH ₃	6.1	8.21
127	p-CH ₃ SO ₂ -Ph	CH ₃	38	7.42
128	3,4-di-Cl-Ph	CH ₃	27	7.57
129	3,5-di-Cl-Ph	CH ₃	19	7.72
130	2,6-di-Cl-Ph	CH ₃	81	7.09
131		CH ₃	15	7.82
132		CH ₃	13	7.89
133		CH ₃	29	7.54
134		CF ₃	1.7	8.77
135		CF ₃	6.1	8.21
136		CF ₃	4.1	8.39


Table 1 continued

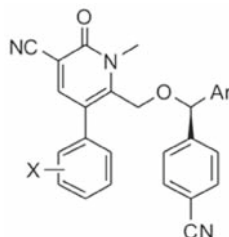
137		CF ₃	2.7	8.57
138		CF ₃	9	8.05







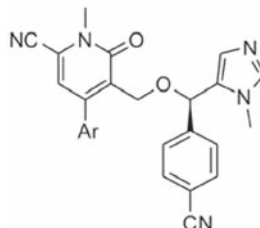
Comp	X	R	Y	IC ₅₀	pIC ₅₀
139	3-Cl	CH ₃	H	0.44	9.36
140	3,4-di-Cl	CH ₃	H	0.49	9.31
141	2,5-di-Cl	CH ₃	H	1.6	8.80
142	2,3-di-Cl	CH ₃	H	6.1	8.21
143	3-F	CH ₃	H	2.1	8.68
144	3,5-di-F	CH ₃	H	0.84	9.08
145	3,4-di-F	CH ₃	H	1.5	8.82
146	2,4-di-F	CH ₃	H	5.5	8.57
147	2,5-di-F	CH ₃	H	2.7	8.57
148	2,3-di-F	CH ₃	H	6.3	8.20
149	3-Cl,5-F	CH ₃	H	0.7	9.15
150	3-Cl,4-F	CH ₃	H	1.1	8.96
151	2-F,3-Cl	CH ₃	H	1.8	8.74
152	3-Br	CH ₃	H	0.51	9.29
153	3-OCH ₃	CH ₃	H	0.41	9.39
154	3-OCH ₂ CH ₃	CH ₃	H	0.67	9.17
155	3-OCF ₃	CH ₃	H	0.42	9.38
156	4-OCF ₃	CH ₃	H	0.73	9.14
157	3,4-OCH ₂ O-	CH ₃	H	0.42	9.38
158	3,4-OCF ₂ O-	CH ₃	H	0.69	9.16
159	3-Cl	CH ₃	Cl	0.97	9.01
160	3-Cl	CH ₃	Br	1.3	8.89
161	3-Cl	CH ₃	I	0.75	9.12
162	3-Cl	CH ₃	3-Cl-Ph	1.0	9.00
163	3,5-di-Cl	CH ₃	H	0.46	9.34
164	3,5-di-Cl	CH ₃	Cl	0.77	9.11
165	3,5-di-Cl	CH ₃	Br	0.42	9.38
166	3,5-di-Cl	CH ₃	I	0.81	9.09
167	3,5-di-Cl	CH ₃	3-Cl-Ph	2.0	8.70
168	4-OCF ₃	CH ₃	Cl	0.36	9.44
169	4-OCF ₃	CH ₃	Br	0.89	9.05

Table 1 continued

170	4-OCF ₃	CH ₃	2-Cl-Ph	3.9	8.41
171	4-OCF ₃	CH ₃	3-Cl-Ph	1.3	8.89
172	4-OCF ₃	CH ₃	4-Cl-Ph	6.5	8.19
173	3-Cl	H	H	0.85	9.07
174	3-Cl	CH ₃ CH ₂	H	0.73	9.14
175	3-Cl	CH ₃ CH ₂ CH ₂	H	0.036	10.44
176	3-Cl		H	0.5	9.30



Comp	Ar	X	IC ₅₀	pIC ₅₀
177		3-Cl	71	7.15
178		3-Cl	5.3	8.28
179		3,5-di-Cl	1.9	8.72
180		4-OCF ₃	9.9	8.00



Comp	Ar	IC ₅₀	pIC ₅₀
181	3-Cl-Ph	1.2	8.92
182	3,5-di-Cl-Ph	1.1	8.96
183	3-CF ₃ -Ph	7.5	8.12

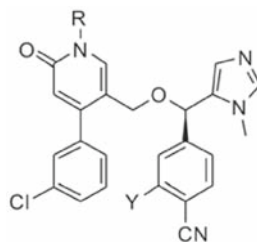
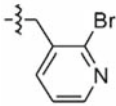
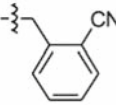
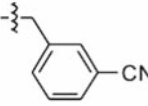
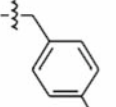
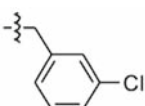
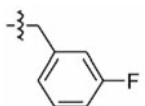
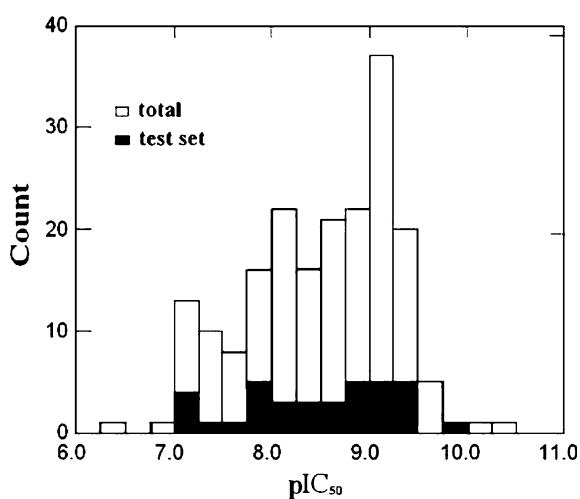


Table 1 continued

Comp	R	IC ₅₀	pIC ₅₀
184	H	11	7.96
185	CH ₃	12	7.92
186	CH ₂ CN	19	7.72
187		3.0	8.52
188		2.3	8.64
189		1.3	8.89
190		5.0	8.30
191		2.1	8.68
192		2.1	8.68

Test set in italic bold

**Fig. 1** Distribution of pIC₅₀ for the data set

$$SEE = \sqrt{\frac{\sum (y_{\text{calc}} - y_{\text{act}})^2}{n - N - 1}} \quad (2)$$

$$F = \frac{\sum (y_{\text{calc}} - y_{\text{mean}})^2 / N}{\sum (y_{\text{calc}} - y_{\text{act}})^2 / (n - N - 1)} \quad (3)$$

$$\text{PRESS} = \sum (y_{\text{pred}} - y_{\text{act}})^2 \quad (4)$$

$$S_{\text{PRESS}} = \sqrt{\frac{\text{PRESS}}{n - N - 1}} \quad (5)$$

$$q^2 = 1 - \frac{\text{PRESS}}{\sum (y_{\text{act}} - y_{\text{mean}})^2} \quad (6)$$

for n , the number of compounds; N , the number of principal components; y_{calc} , calculated activity; y_{act} , observed activity; y_{mean} , mean value of observed activities; y_{pred} , predicted activities (from cross-validation).

Training and test set

Initially the whole data set was submitted to CoMFA analysis after alignment, but compounds **3**, **22**, **32**, **34**, **77**, **80**, **175**, and **177** were identified as outliers (with residual

Fig. 2 Putative bioactive conformations of different analogues: **a** tipifarnib (R115777) from 1SA4; **b** **55** from 1NI1; **c** the docked conformation of **2** (into the protein structure from 1NI1)

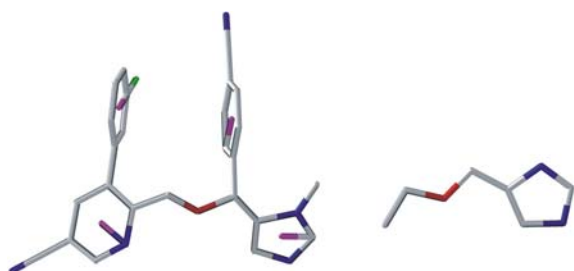
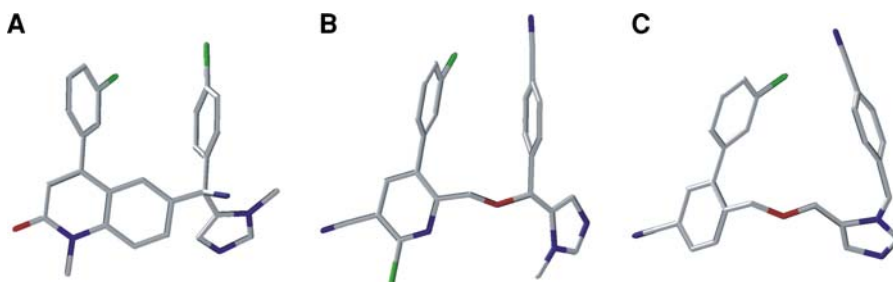


Fig. 3 Templates and reference atoms for alignment. *Left*: centroid alignment; *Right*: database alignment

between experimental and predicted pIC_{50} values >1 log unit). Omitting the eight outliers, we divided the remaining data set, choosing 148 compounds to be in the QSAR training set and 36 compounds in the test set. The test set compounds are indicated in Table 1 using bold italics for their compound number and activities. Care was taken to ensure a uniform distribution of structurally different compounds from each structural class, with a similarly wide range in pIC_{50} for training and test sets. Figure 4 shows the structural diversity of the aligned compounds of the training and test sets.

Docking

Intensive docking studies [31] were performed on the eight outliers identified from initial CoMFA studies in order to understand the cause of the significant difference between their experimental and predicted activity. We used the genetic algorithm-based docking program GOLD (Genetic Optimisation for Ligand Docking) [46] to generate poses and the dimensionless fitness scoring function GOLDScore [47] to rank the poses: the higher the score, the better the docking result. The GOLDScore fitness function has been optimized for the prediction of ligand binding positions, and takes into account hydrogen-bonding energy, van der Waals energy and ligand torsion strain [47].

For docking we used the rat FT protein structure (PDB: 1NI1) [28], because it contained **55** in the active site and hence the side-chain conformations should be suitable for docking structurally related compounds. The protein input file was prepared in Sybyl7.2 [45]. All residues having atoms

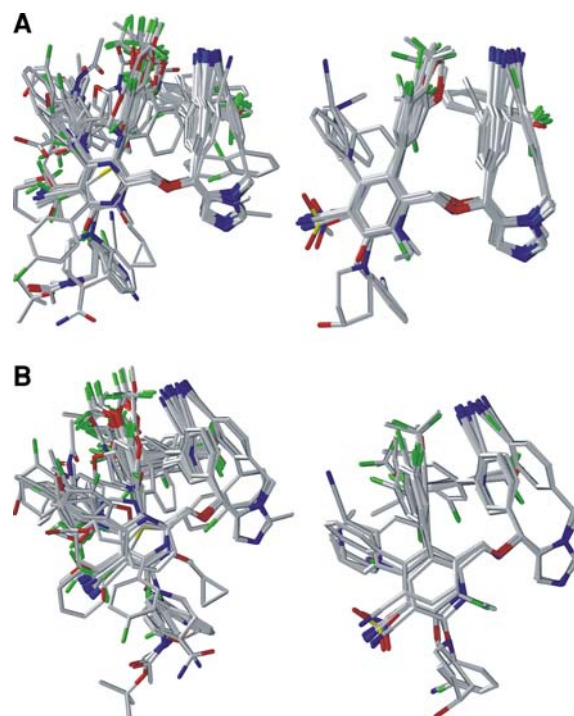


Fig. 4 The alignment of the 148 compounds in the training set (*left*) and 36 compounds in the test set (*right*): **a** centroid alignment; **b** database alignment

within a 10 Å radius of the co-crystallized ligand (**55**) were included in the active site definition. After removal of the bound ligand, hydrogens were added to the protein file. Then minimization was commenced with all heavy atoms constrained and finally the whole protein was minimized, using the Tripos force field, to a gradient convergence of $0.01 \text{ kcal mol}^{-1}$. All water molecules in 1NI1 were deleted from our model because no water molecules have been found to be involved in important interactions in the active site of existing protein-ligand X-ray structures for FT.

Results and discussion

CoMFA and CoMSIA analysis showed that steric, electrostatic, hydrophobic, and hydrogen bond acceptor features have significant influence on the bioactivity of this

series of compounds; however, the hydrogen bond donor fields did not correlate well with FT inhibitory activity.

3D-QSAR models

Supplementary Table S1 and Table 2 summarize the full investigation based on the training set of 148 compounds and test set of 36 compounds. For each of the 15 models, the fields which contribute to the model, the number of PLS components (N), and various statistical measures are listed in the tables. For the single- and multi-field models, r^2 , q^2 and F are generally better for the centroid alignment than for the database alignment. Also, higher r^2_{pred} showed clearly that the centroid-aligned models possess more predictive power than the database-aligned models for the particular test set used. Our centroid alignment scheme takes into consideration more thoroughly the different regions relevant for the ligand-protein interaction, and hence it is a theoretically more reasonable strategy. Thus

we rely on it in our further interpretation of the model results.

From single field models (Table S1), the steric field was found to be the most descriptive field for the SAR (structure–activity relationship) of this series of compounds, while electrostatic, hydrophobic, and hydrogen bond accepting features were also important. We then proceeded to investigate models which contained different combinations of the fields found to be significant in the single field investigations. Table 2 shows that nearly all of the multi-field models had good predictive power. Interestingly, the CoMFA steric plus electrostatic (SE) model was much better than the corresponding CoMSIA one. In order to take full advantage of the better SE model of CoMFA and the hydrophobic and hydrogen bond acceptor fields available in CoMSIA, we prepared combined CoMFA and CoMSIA models (models 13–15) which showed significant improvement in all statistical parameters including r^2_{pred} compared to the corresponding CoMSIA-only models

Table 2 Statistical parameters of CoMFA and CoMSIA models (multi-field), with training set $n = 148$ and test set $n = 36$

PLS	CoMFA ^a	CoMSIA ^b				CoMFA + CoMSIA ^c		
Parameter	8	9	10	11	12	13	14	15
q^2	0.613 ^d	0.530 ^d	0.611 ^d	0.543 ^d	0.610 ^d	0.635 ^d	0.595 ^d	0.630 ^d
	0.574 ^e	0.584 ^e	0.589 ^e	0.525 ^e	0.611 ^e	0.625 ^e	0.578 ^e	0.626 ^e
S_{PRESS}	0.424 ^d	0.468 ^d	0.428 ^d	0.461 ^d	0.429 ^d	0.413 ^d	0.433 ^d	0.418 ^d
	0.441 ^e	0.408 ^e	0.431 ^e	0.475 ^e	0.425 ^e	0.413 ^e	0.446 ^e	0.417 ^e
r^2	0.826 ^d	0.758 ^d	0.846 ^d	0.770 ^d	0.861 ^d	0.841 ^d	0.739 ^d	0.878 ^d
	0.732 ^e	0.748 ^e	0.835 ^e	0.771 ^e	0.904 ^e	0.839 ^e	0.728 ^e	0.873 ^e
SEE	0.285 ^d	0.336 ^d	0.269 ^d	0.328 ^d	0.256 ^d	0.273 ^d	0.347 ^d	0.240 ^d
	0.353 ^e	0.315 ^e	0.273 ^e	0.327 ^e	0.218 ^e	0.268 ^e	0.356 ^e	0.244 ^e
F	169.188 ^d	111.759 ^d	129.480 ^d	119.829 ^d	146.617 ^d	149.832 ^d	136.175 ^d	168.746 ^d
	130.176 ^e	115.89 ^e	122.151 ^e	118.435 ^e	165.787 ^e	145.546 ^e	121.427 ^e	161.435 ^e
N^f	4 ^d	4 ^d	6 ^d	4 ^d	6 ^d	5 ^d	3 ^d	6 ^d
	3 ^e	4 ^e	6 ^e	4 ^e	8 ^e	5 ^e	3 ^e	6 ^e
Field ^g	Contribution							
S	0.526 ^d	0.288 ^d	0.135 ^d	0.195 ^d	0.112 ^d	0.312 ^d	0.375 ^d	0.277 ^d
	0.519 ^e	0.290 ^e	0.133 ^e	0.197 ^e	0.118 ^e	0.316 ^e	0.377 ^e	0.276 ^e
E	0.474 ^d	0.712 ^d	0.472 ^d	0.513 ^d	0.374 ^d	0.332 ^d	0.345 ^d	0.254 ^d
	0.481 ^e	0.710 ^e	0.455 ^e	0.490 ^e	0.340 ^e	0.324 ^e	0.343 ^e	0.257 ^e
H	–	–	0.393 ^d	–	0.347 ^d	0.355 ^d	–	0.307 ^d
	–	–	0.402 ^d	–	0.312 ^d	0.360 ^d	–	0.302 ^d
A	–	–	–	0.292 ^d	0.167 ^d	–	0.279 ^d	0.162 ^d
	–	–	–	0.313 ^e	0.230 ^e	–	0.280 ^e	0.165 ^e
r^2_{pred}	0.616 ^d	0.463 ^d	0.478 ^d	0.470 ^d	0.518 ^d	0.609 ^d	0.509 ^d	0.614 ^d
	0.565 ^e	0.451 ^e	0.455 ^e	0.458 ^e	0.472 ^e	0.583 ^e	0.498 ^e	0.601 ^e

^a CoMFA models 8; ^b CoMSIA models 9–12; ^c CoMFA + CoMSIA models 13–15; ^d Centroid aligned models; ^e Database-aligned models;

^f Number of principal components; ^g Field type: S—Steric, E—Electrostatic, H—Hydrophobic, D—Hydrogen bond donor, A—Hydrogen bond acceptor

(models 10–12). This approach was not reported in the literature before our recent report [48], but apparently could be broadly applicable since it is sometimes found, as in this work, that for the SE models CoMFA performs better than CoMSIA, and hence the CoMFA S and E fields are the better descriptors to use in conjunction with other fields of interest to prepare more statistically significant models. Table 3 shows the residuals of the test set

predictions for all multi-field models based on the centroid alignment.

It is difficult to say which of the centroid-aligned multi-field models 8, 13, or 15 is the best. Models 13 and 15 have the highest q^2 and lowest S_{PRESS} . Model 15 has the highest r^2 and lowest SEE. Models 8 and 15 have the highest F and r^2_{pred} . We chose to analyze model 15 as the best model since it is statistically robust and contains the four kinds of fields which gave significant single-field models.

Table 3 Residuals^a of test set compounds for CoMFA and CoMSIA models

Comp	QSAR model							
	8	9	10	11	12	13	14	15
5	−0.06	−0.08	−0.12	−0.22	−0.22	0.05	0.03	−0.19
12	0.42	0.71	1.05	0.73	0.99	0.69	0.43	0.82
26	0.87	0.87	0.65	0.79	0.60	0.75	0.85	0.66
27	0.17	0.13	0.02	0.20	0.02	0.10	0.31	0.06
29	0.59	0.78	0.59	0.68	0.52	0.52	0.63	0.40
35	0.06	−0.21	−0.15	−0.02	−0.07	−0.01	−0.03	0.20
46	−0.07	−0.36	0.03	−0.39	−0.03	0.11	−0.30	−0.09
53	0.37	0.59	0.74	0.49	0.76	0.55	0.30	0.63
59	−0.19	−0.43	−0.01	−0.44	0.10	−0.19	−0.01	0.15
63	0.42	−0.15	0.06	0.05	0.17	0.31	0.08	0.22
71	−0.28	−0.91	−0.72	−0.73	−0.64	−0.26	−0.56	−0.35
73	−0.48	−0.42	−0.54	−0.44	−0.44	−0.43	−0.38	−0.34
76	0.90	1.24	1.02	1.13	1.01	0.84	1.00	0.67
84	−0.38	−0.69	−0.59	−0.60	−0.62	−0.43	−0.37	−0.49
102	−0.05	0.44	0.34	0.11	0.13	0.04	0.00	−0.12
105	0.31	0.18	−0.08	0.25	−0.09	0.22	0.46	0.07
113	0.27	0.40	0.26	0.41	0.22	0.20	0.38	0.14
117	−0.67	−0.68	−0.39	−0.56	−0.40	−0.37	−0.51	−0.36
119	−0.58	−0.42	−0.37	−0.32	−0.34	−0.54	−0.48	−0.40
121	−0.42	−0.24	−0.13	−0.14	−0.08	−0.29	−0.30	−0.09
122	−0.45	−0.59	−0.19	−0.42	−0.05	−0.21	−0.34	−0.02
124	0.31	0.40	−1.30	0.77	−1.05	−0.94	0.76	−1.08
127	0.82	0.48	0.10	0.88	0.48	0.58	1.15	0.65
130	0.30	0.75	0.71	0.68	0.60	0.50	0.60	0.36
139	−0.96	−1.10	−1.16	−1.34	−1.22	−1.00	−1.31	−1.05
140	−0.49	−0.46	−0.50	−0.50	−0.52	−0.38	−0.52	−0.42
142	0.52	0.67	0.58	0.64	0.55	0.55	0.55	0.54
154	−0.12	−0.08	−0.63	−0.12	−0.62	−0.50	−0.18	−0.44
155	−0.43	−0.16	−0.19	−0.20	−0.21	−0.28	−0.37	−0.23
160	0.24	0.27	0.25	0.21	0.26	0.26	0.10	0.30
167	0.12	0.23	0.25	0.20	0.24	0.45	0.27	0.48
174	−0.29	−0.22	−0.35	−0.25	−0.35	−0.34	−0.33	−0.32
179	0.11	0.15	0.14	0.15	0.17	0.17	0.13	0.16
182	0.36	0.23	−0.19	0.24	−0.33	0.05	0.28	−0.26
189	−0.44	−0.37	−0.13	−0.35	−0.27	−0.48	−0.41	−0.44
192	−0.35	−0.24	0.01	−0.19	−0.10	−0.35	−0.24	−0.33

^a Calculated pIC_{50} − experimental pIC_{50}

Graphical interpretation of the QSAR results

The contour maps of CoMFA/CoMSIA 3D-QSAR models indicate which structural modification will enhance the bioactivity. Figure 5 shows the steric (a), electrostatic (b), hydrophobic (c), and hydrogen bond acceptor (d) contours of CoMFA and CoMSIA fields derived from our best model, the centroid-aligned model 15. Also shown in the images are the template molecule (**55**) and the rat FT active pocket which is represented as a Connolly surface of the protein onto which different properties are mapped.

Figure 5a shows the CoMFA steric field contour with calculated hydrophobicity mapped onto the protein active site. From the picture, we can clearly see that the indications of polyhedra are consistent with the active site structural features. For steric field contours, green represents where more bulky groups are favorable for improving bioactivity, while yellow means in those positions bulky groups are unfavorable for improving activity. The active site surface shows that regions surrounding the green polyhedra are deep spacious pockets which could accommodate more bulky ligand substructures. The activity trends of compounds **62–68** and **20–22** are consistent with this assertion. For the yellow polyhedra which indicate sterically unfavorable regions, the adjacent pockets of the binding site are shallow and crowded. Compounds such as **96–132** showed poorer activities ($7.0 < \text{pIC}_{50} < 8.4$) compared to the structurally similar **54–61** and **63–68** ($9.0 < \text{pIC}_{50} < 9.7$), mainly because the “A” ring substituents of this compound group did not occupy the spacious area of the active site that is usually occupied by the “C” ring. Also, for those compounds the substituents on the “D” ring are not in ideal position for active site binding (cf. Fig. 5a).

Figure 5b shows the CoMFA electrostatic field contour with the electrostatic potential mapped onto the active site. A blue contour means an electronegative group there will favor the activity, while red represents regions where electronegative groups will reduce the activity of the molecule. The ligand binding region of the active site shows no obvious variation in protein electrostatic potential, so there is mainly cyan-colored surface near the template molecule and close to the CoMFA electrostatic field

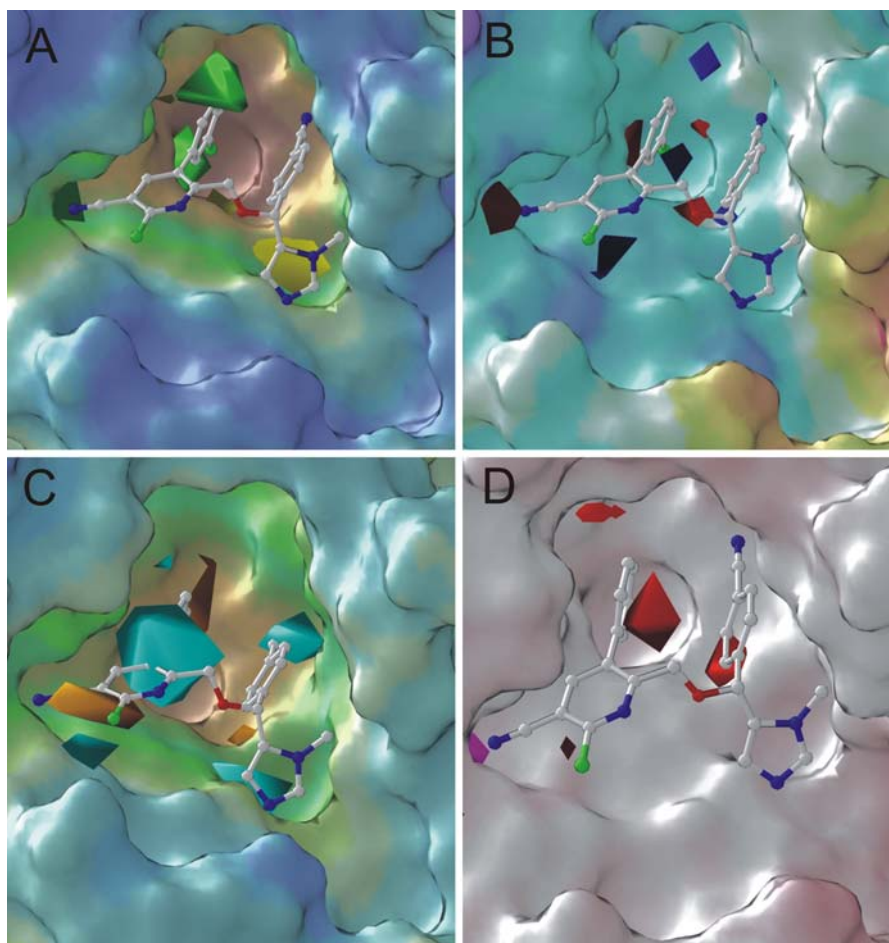


Fig. 5 CoMFA/CoMSIA contour plots, from model 15, inside the rat FT (1NI1) active site mapped with different properties [**a** and **c** hydrophobic potential (from *brown* to *green* to *blue*, for most hydrophobic to most hydrophilic); **b** electrostatic potential (from *red* to *yellow* to *cyan* to *purple*, for electropositive to electronegative); **d**, hydrogen bond donor density (from *pink* to *white*, for H-bond donor density from high to low)] mapped onto the Connolly surface of the protein active site. **a** CoMFA steric map with *green* (*yellow*) polyhedra indicating regions where bulky groups should increase (decrease) FT binding affinity; **b** CoMFA electrostatic map with *blue* (*red*) polyhedra indicating regions where electronegative groups should increase (decrease) FT binding affinity; **c** CoMSIA

hydrophobic map with *orange* (*cyan*) polyhedra indicating regions where hydrophobic groups should increase (decrease) FT binding affinity. **d** CoMSIA hydrogen bond acceptor field map with *magenta* (*red*) polyhedra indicating regions where hydrogen bond acceptor groups should increase (decrease) FT binding affinity. The template compound (**55**) is also shown in each image. The images were prepared in Sybyl as follows: prepare X-ray structure, including co-crystallized **55**; prepare 3D-QSAR contour map, including template **55**; superimpose the two images, carefully aligning the images of **55**; delete the co-crystallized **55**, to avoid having two images of it; add in protein surface property representation

contours. However, the CoMFA contour map is more useful since it shows the different requirements for the electrostatic nature of substituents in different locations. For example, the big red block on the left side of the figure indicates that less electronegative groups are best for substitutions in the para position of the “A” ring. The activity trend of compounds **112**, **115**, **118**, **120**, **126**, and **127** supports this.

Figure 5c shows the CoMSIA hydrophobic field contour with hydrophobic potential mapped onto the active site. The colors from brown to green to cyan indicate the areas of the protein surface ranging from most hydrophobic to

most hydrophilic. The CoMSIA hydrophobic contours are consistent with the hydrophobicity mapped surface. Orange polyhedra indicate that hydrophobic groups in those locations will increase the activity, while cyan polyhedra are located where hydrophobic groups will decrease the activity. All the orange blocks matched with the deep brown pockets of the active site which are composed of a group of hydrophobic amino acids including Tyr93, Leu96, Trp102, Leu103, Trp106, Phe360, and Tyr361, while all the cyan blocks are in green or cyan areas and the biggest cyan block is actually exposed to solvent. Compounds **54** and **62** are perfect examples for interpreting the orange and

cyan blocks at the very left corner of the binding site. Compound **61** is particularly active and has a 3-Cl-Ph group matching the orange block indicating a hydrophobic group on the “A” ring and a morpholin-4-yl group matching the cyan block indicating a hydrophilic group on the “D” ring.

Figure 5d shows the CoMSIA H-bond acceptor field contour with the H-bond donor density mapped onto the active site. For the protein surface, pink means high H-bond donor density. For the CoMSIA H-bond acceptor field contour, magenta polyhedra indicate regions where hydrogen bond acceptor groups should increase the compound binding affinity, while red polyhedra indicate regions where hydrogen bond acceptor groups will decrease the compound binding affinity. The magenta polyhedron on the very left part of Fig. 5d is the only spot showing that a hydrogen bond acceptor group will increase the compound activity. Appropriate residues available for hydrogen bond donation can be seen in the PDB:1NI1 X-ray structure of FT, in which the cyano group of the “A” ring of **55** accepts a hydrogen bond from the side chain OH of Tyr93 and from Asp359 at this position (hydrogen bond distances between 2.7 and 3.4 Å). Comparing compounds **1** and **14**, with a cyano group (H-bond acceptor) at R, **14** has a $\text{pIC}_{50} = 9.43$, while with Cl, **1** has a slightly lower $\text{pIC}_{50} = 9.21$. More significant examples are found in compounds **96–133**, in which only compounds having an H-bond acceptor group at the *para* position of ring “A”, such as **112**, **124**, **126**, exhibited high activity ($\text{pIC}_{50} > 8.00$).

Outliers

When first submitting the whole 195 compound dataset to CoMFA analysis, we identified eight outliers (**3**, **22**, **32**, **34**, **77**, **80**, **175**, and **177**) and so proceeded to analyze each of them using docking to rat FT (1NI1). Tables S3 and 4 provide the residuals for the outliers for all the QSAR models, and Fig. 6 shows the docked binding mode of each

outlier. The residuals revealed that the compounds were outliers in most of the 15 models, with only **22**, **34** and **77** having a residual with absolute value <1 for four, one and two models, respectively. For every model, the pIC_{50} 's were underestimated compared to experiment for **22**, **34** and **175**, and overestimated for the other outliers.

There are several possible reasons why a compound may be an outlier in QSAR analysis based on binding affinity to a receptor. Variability in structure can cause a compound to adopt a different binding mode with a conformation that does not match the one selected as the alignment template. If an unusual functional group is present in only one compound, a QSAR model may not be able to account for the group properly. Also, experimental errors may be the source of extremely high or low activity for a structurally normal compound for which the activity falls outside the SAR trend of other analogues.

As shown in Fig. 6, compounds **77**, **175** and **177** exhibited the same docking binding mode as the template molecule shown in Fig. 5, and their structures are not much different compared to other analogues with respect to the four key properties (steric, electrostatic, hydrophobic, and H-bond accepting) used in the QSAR analysis. Their residuals (Tables S3 and 4) are large and fairly constant in different QSAR models. It is unclear why these three compounds behaved as outliers. The experimental pIC_{50} for **77** is 1.5 less than that of **78**, although a shift from 2,3-dichloro to 2,4-dichloro on the Ar_2 ring is the only structural difference. Similarly an extra methylene enhances the pIC_{50} of **175** compared to **174** by 1.3 and two extra methyls increase the pIC_{50} of **178** versus **177** by 1.1. Perhaps these subtle differences allow specific interactions with the FT binding site which could be modeled better using docking calculations allowing protein flexibility.

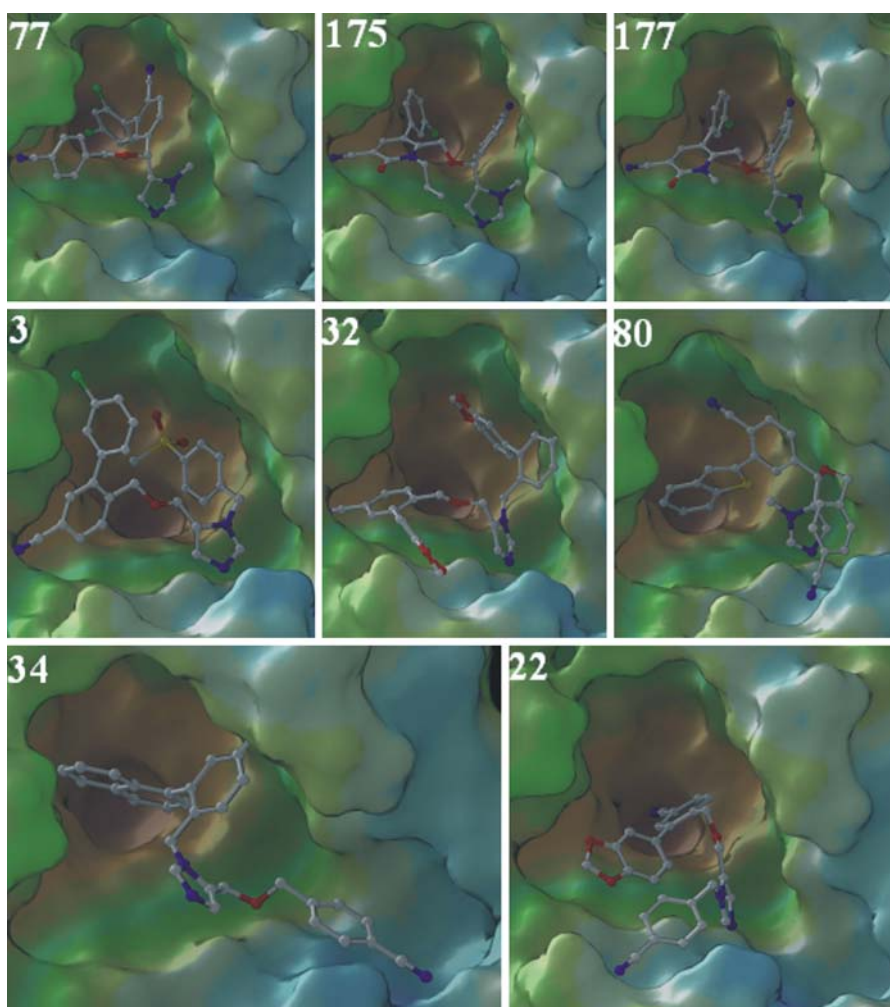
The other five outliers all had different docked binding modes than that of the template molecule; however, they can be further divided into two categories. Compounds **3**, **32**, and **80** had positive residuals, with their predicted activities (pIC_{50}) much higher than their actual activities. From their docked binding modes, shown in Fig. 6, none of them showed good affinity with the FT active site. For **3**, the steric hindrance of the relatively bulky methylsulfonyl group in the *para* position of ring “D” forced ring “C” to come out of the active site. Compound **32** is the most rigid in the whole dataset, and its large benzodioxole group in the *ortho* position of ring “D” forced ring “C”, which contains a fused dioxolane ring, out of the deep pocket and to be flipped to the opposite orientation compared to other compounds. These two compounds obviously cannot make use of the deepest lipophilic pocket in the active site to enhance binding to the protein, which greatly reduces their binding affinity. In the docked pose of compound **80**, the bicyclic ring at the *meta* position of the “D” ring does

Table 4 Residuals of eight outliers for CoMFA and CoMSIA models

Comp	QSAR model							
	8	9	10	11	12	13	14	15
3	1.90	1.57	2.12	1.73	2.14	2.01	1.87	2.11
22	−1.25	−1.00	−1.03	−0.93	−0.98	−1.22	−1.11	−1.10
32	1.62	1.92	1.84	2.11	1.86	1.72	2.12	1.83
34	−1.91	−1.95	−2.34	−1.96	−2.45	−1.88	−1.54	−1.96
77	1.42	1.61	1.61	1.50	1.63	1.43	1.48	1.44
80	1.93	2.13	1.83	2.05	1.91	1.84	1.94	1.80
175	−2.01	−2.09	−2.10	−2.30	−2.22	−2.01	−2.27	−2.05
177	1.28	1.24	1.19	1.12	1.12	1.27	1.14	1.22

Calculated pIC_{50} − experimental pIC_{50}

Fig. 6 Docking binding modes of eight outliers into the rat FT (INI1) active pocket with the hydrophobic potential (from brown to green to blue, for most hydrophobic to most hydrophilic) mapped onto the Connolly surface of the protein active site. Compound numbers are listed in each image



make use of the deepest pocket. Because there is no “C” ring in this molecule, ring “A” faces less steric blocking than it does in **32**, and therefore it flipped towards the imidazole binding spot to make the benzothiophene group fit better in the deepest pocket; but the structure of **32** also prevents ring “A” from fitting into the other possible aryl group binding spot shown at the lower right corner of the pictures. Thus, compound **80** did not show a good binding conformation, and its actual activity was much lower than that predicted using our QSAR models.

Compounds **34** and **22** both had negative residuals, with their predicted activities (pIC_{50}) lower than their experimental activities. There is no ring “C” in **34**, and ring “D” has a big α -naphthyl group at the *ortho* position. Upon binding, the ligand could be accommodated appropriately in the active site, and as shown in Fig. 6, every part of **34** was placed in an energetically favorable position, especially ring “A”. Unlike in compound **158**, in **34**, ring “A” had its cyano group pointed into the aryl group binding site and made good use of the high H-bond donor density region, so it is understandable that the experimental

activity of **34** was higher than the QSAR-predicted activity, considering that the lack of ring “C” and the presence of the bulky naphthyl group on ring “D” were both unfavorable factors in the QSAR models. It is a little confusing that **22** was defined as an outlier, because four of its residuals were less than one log unit (Tables S3 and 4) and **22** fits the CoMSIA models (steric, electrostatic, and H-bond accepting field) quite well. However, we identified outliers based on the CoMFA steric and electrostatic field model (model 8). Compared to the template binding mode, **22** bound to the active site in a totally upside-down pose. However, this pose also fit the active site very well which matches with the reasonably high experimental activity of **22**.

Conclusion

CoMFA and CoMSIA 3D-QSAR models with statistical significance and high r^2_{pred} have been built based on 192 highly diverse imidazole containing FT inhibitors. We used

two alignment strategies, and compared the differences between CoMFA and CoMSIA models. In this study, the centroid alignment, which took different regions of the ligand-protein interaction into consideration, gave statistically more significant models than the common-fragment database alignment. While the CoMFA model with steric and electrostatic fields (model 8) was statistically robust, in order to take advantage of CoMFA and CoMSIA fields, we built combined CoMFA/CoMSIA models 13–15 which were better in some statistical measures than model 8 and were significantly better than the corresponding CoMSIA-alone models 10–12. The field contour maps of the best model, model 15, with $r^2 = 0.878$, $q^2 = 0.630$, and $r^2_{\text{pred}} = 0.614$, provide explicit indication for lead optimization regarding steric, electrostatic, hydrophobic, and H-bond accepting properties. Docking studies on the statistical outliers revealed that some of them had a different binding mode in the FT active site, explaining why the predicted activities differed from the experimental activities based on steric bulk and available active site space.

Acknowledgments Funding from National Center for Zoonotic, Vector-borne, and Enteric Diseases (CK) of the Centers for Disease Control and Prevention (CDC) (U01/CI000211); National Science Foundation (EPS-0556308); University of Mississippi, including from its Office of Research and Sponsored Programs; as well as Laboratory for Applied Drug Design and Synthesis and MCSR computing facilities are greatly appreciated. SP is a Natural Products Neuroscience Fellow [P20 RR021929 from National Center for Research Resources (NCRR), National Institutes of Health (NIH)]. This investigation was conducted in a facility constructed with support from research facilities improvement program C06 RR-14503-01 from the NIH National Center for Research Resources.

References

- Caponigro F (2002) *Anticancer Drugs* 13:891. doi:[10.1097/00001813-200209000-00016](https://doi.org/10.1097/00001813-200209000-00016)
- Casey PJ, Solski PA, Der CJ, Buss JE (1989) *Proc Natl Acad Sci USA* 86:8323. doi:[10.1073/pnas.86.21.8323](https://doi.org/10.1073/pnas.86.21.8323)
- Graaf MR, Richel DJ, van Noorden CJF, Guchelaar H-J (2004) *Cancer Treat Rev* 30:609. doi:[10.1016/j.ctrv.2004.06.010](https://doi.org/10.1016/j.ctrv.2004.06.010)
- Johnson BE, Heymach JV (2004) *Clin Cancer Res* 10:4254S. doi:[10.1158/1078-0432.CCR-040016](https://doi.org/10.1158/1078-0432.CCR-040016)
- Mazieres J, Pradines A, Favre G (2004) *Cancer Lett* 206:159. doi:[10.1016/j.canlet.2003.08.033](https://doi.org/10.1016/j.canlet.2003.08.033)
- Powers S, Michaelis S, Broek D, Santa A-AS, Field J, Herskowitz I, Wigler M (1986) *Cell* 47:413
- Rolland D, Camara-Clayette V, Barbarat A, Salles G, Coiffier B, Ribrag V, Thieblemont C (2008) *Cancer Chemother Pharmacol* 61:855. doi:[10.1007/s00280-007-0543-3](https://doi.org/10.1007/s00280-007-0543-3)
- Sousa SF, Fernandes PA, Ramos MJ (2008) *Curr Med Chem* 15:1478. doi:[10.2174/092986708784638825](https://doi.org/10.2174/092986708784638825)
- Eastman RT, Buckner FS, Yokoyama K, Gelb MH, Van Voorhis WC (2006) *J Lipid Res* 47:233. doi:[10.1194/jlr.R500016-JLR200](https://doi.org/10.1194/jlr.R500016-JLR200)
- Gelb MH, Van Voorhis WC, Buckner FS, Yokoyama K, Eastman R, Carpenter EP, Panethymitaki C, Brown KA, Smith DF (2003) *Mol Biochem Parasitol* 126:155. doi:[10.1016/S0166-6851\(02\)00282-7](https://doi.org/10.1016/S0166-6851(02)00282-7)
- Glenn MP, Chang S-Y, Hucke O, Verlinde CLMJ, Rivas K, Hornéy C, Yokoyama K, Buckner FS, Pendyala PR, Chakrabarti D, Gelb M, Van Voorhis WC, Sebt SM, Hamilton AD (2005) *Angew Chem Int Ed* 44:4903. doi:[10.1002/anie.200500674](https://doi.org/10.1002/anie.200500674)
- Nallan L, Bauer KD, Bendale P, Rivas K, Yokoyama K, Horney CP, Pendyala PR, Floyd D, Lombardo LJ, Williams DK, Hamilton A, Sebt S, Windsor WT, Weber PC, Buckner FS, Chakrabarti D, Gelb MH, VanVoorhis WC (2005) *J Med Chem* 48:3704. doi:[10.1021/jm0491039](https://doi.org/10.1021/jm0491039)
- Ohkanda J, Lockman JW, Yokoyama K, Gelb MH, Croft SL, Kendrick H, Harrell MI, Feagin JE, Blaskovich MA, Sebt SM, Hamilton AD (2001) *Bioorg Med Chem Lett* 11:761. doi:[10.1016/S0960-894X\(01\)00055-5](https://doi.org/10.1016/S0960-894X(01)00055-5)
- Wiesner J, Kettler K, Sakowski J, Ortmann R, Katzin AM, Kimura EA, Silber K, Klebe G, Jomaa H, Schlitzer M (2004) *Angew Chem Int Ed* 43:251. doi:[10.1002/anie.200351169](https://doi.org/10.1002/anie.200351169)
- Strickland CL, Windsor WT, Syto R, Wang L, Bond R, Wu Z, Schwartz J, Le HV, Beese LS, Weber PC (1998) *Biochemistry* 37:16601. doi:[10.1021/bi981197z](https://doi.org/10.1021/bi981197z)
- Appels NMGM, Beijnen JH, Schellens JHM (2005) *Oncologist* 10:565. doi:[10.1634/theoncologist.10-8-565](https://doi.org/10.1634/theoncologist.10-8-565)
- Hasvold LA, Wang W, Gwaltney SL, Rockway TW, Nelson LTJ, Mantei RA, Fakhoury SA, Sullivan GM, Li Q, Lin N-H, Wang L, Zhang H, Cohen J, Gu W-Z, Marsh K, Bauch J, Rosenberg S, Sham HL (2003) *Bioorg Med Chem Lett* 13:4001. doi:[10.1016/j.bmcl.2003.08.058](https://doi.org/10.1016/j.bmcl.2003.08.058)
- Brunner TB, Hahn SM, Gupta AK, Muschel RJ, McKenna WG, Bernhard EJ (2003) *Cancer Res* 63:5656
- Heymach JV, Johnson DH, Khuri FR, Safran H, Schlabach LL, Yunus F, DeVore RFIII, De Porre PM, Richards HM, Jia X, Zhang S, Johnson BE (2004) *Ann Oncol* 15:1187. doi:[10.1093/annonc/mdh315](https://doi.org/10.1093/annonc/mdh315)
- Li Q, Li T, Woods KW, Gu W-Z, Cohen J, Stoll VS, Galicia T, Hutchins C, Frost D, Rosenberg SH, Sham HL (2005) *Bioorg Med Chem Lett* 15:2918. doi:[10.1016/j.bmcl.2005.03.049](https://doi.org/10.1016/j.bmcl.2005.03.049)
- Li Q, Wang GT, Li T, Gwaltney SLII, Woods KW, Claiborne A, Wang X, Gu W, Cohen J, Stoll VS, Hutchins C, Frost D, Rosenberg SH, Sham HL (2004) *Bioorg Med Chem Lett* 14:5371. doi:[10.1016/j.bmcl.2004.08.011](https://doi.org/10.1016/j.bmcl.2004.08.011)
- Lin N-H, Wang L, Wang X, Wang GT, Cohen J, Gu W-Z, Zhang H, Rosenberg SH, Sham HL (2004) *Bioorg Med Chem Lett* 14:5057. doi:[10.1016/j.bmcl.2004.07.083](https://doi.org/10.1016/j.bmcl.2004.07.083)
- Ohkanda J, Strickland CL, Blaskovich MA, Carrico D, Lockman JW, Vogt A, Bucher CJ, Sun I, Qian Y, Knowles D, Pusateri EE, Sebt SM, Hamilton AD (2006) *Org Biomol Chem* 4:482. doi:[10.1039/b508184j](https://doi.org/10.1039/b508184j)
- Reid TS, Beese LS (2004) *Biochemistry* 43:6877. doi:[10.1021/bi049723b](https://doi.org/10.1021/bi049723b)
- Tong Y, Lin N-H, Wang L, Hasvold L, Wang W, Leonard N, Li T, Li Q, Cohen J, Gu W-Z, Zhang H, Stoll V, Bauch J, Marsh K, Rosenberg SH, Sham HL (2003) *Bioorg Med Chem Lett* 13:1571. doi:[10.1016/S0960-894X\(03\)00195-1](https://doi.org/10.1016/S0960-894X(03)00195-1)
- Wang GT, Wang X, Wang W, Hasvold LA, Sullivan G, Hutchins CW, O'Conner S, Gentiles R, Sowin T, Cohen J, Gu W-Z, Zhang H, Rosenberg SH, Sham HL (2005) *Bioorg Med Chem Lett* 15:153. doi:[10.1016/j.bmcl.2004.10.014](https://doi.org/10.1016/j.bmcl.2004.10.014)
- Wang L, Lin N-H, Li Q, Henry RF, Zhang H, Cohen J, Gu W-Z, Marsh KC, Bauch JL, Rosenberg SH, Sham HL (2004) *Bioorg Med Chem Lett* 14:4603. doi:[10.1016/j.bmcl.2004.07.004](https://doi.org/10.1016/j.bmcl.2004.07.004)
- Wang L, Wang GT, Wang X, Tong Y, Sullivan G, Park D, Leonard NM, Li Q, Cohen J, Gu WZ, Zhang H, Bauch JL, Jakob CG, Hutchins CW, Stoll VS, Marsh K, Rosenberg SH, Sham HL, Lin NH (2004) *J Med Chem* 47:612. doi:[10.1021/jm030434f](https://doi.org/10.1021/jm030434f)

29. Cramer RD III, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959. doi:[10.1021/ja00226a005](https://doi.org/10.1021/ja00226a005)
30. Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130. doi:[10.1021/jm00050a010](https://doi.org/10.1021/jm00050a010)
31. Halperin I, Ma B, Wolfson H, Nussinov R (2002) *Proteins* 47:409. doi:[10.1002/prot.10115](https://doi.org/10.1002/prot.10115)
32. Equbal T, Silakari O, Ravikumar M (2008) *Eur J Med Chem* 43:204. doi:[10.1016/j.ejmech.2007.02.013](https://doi.org/10.1016/j.ejmech.2007.02.013)
33. Puntambekar DS, Giridhar R, Yadav MR (2008) *Eur J Med Chem* 43:142. doi:[10.1016/j.ejmech.2007.02.003](https://doi.org/10.1016/j.ejmech.2007.02.003)
34. Puntambekar D, Giridhar R, Yadav MR (2006) *Bioorg Med Chem Lett* 16:1821. doi:[10.1016/j.bmcl.2006.01.019](https://doi.org/10.1016/j.bmcl.2006.01.019)
35. Puntambekar DS, Giridhar R, Yadav MR (2006) *Eur J Med Chem* 41:1279. doi:[10.1016/j.ejmech.2006.07.002](https://doi.org/10.1016/j.ejmech.2006.07.002)
36. Xie A, Sivaprakasam P, Doerksen RJ (2006) *Bioorg Med Chem* 14:7311. doi:[10.1016/j.bmc.2006.06.041](https://doi.org/10.1016/j.bmc.2006.06.041)
37. Chaurasia SH, Srivastava AK, Math A, Srivastava MK, Pandey A (2007) *Oxid Commun* 30:778
38. Fernández M, Tundidor-Camba A, Caballero JM (2005) *Mol Simul* 31:575. doi:[10.1080/08927020500134144](https://doi.org/10.1080/08927020500134144)
39. Giraud E, Luttmann C, Lavelle F, Riou JF, Mailliet P, Laoui A (2000) *J Med Chem* 43:1807. doi:[10.1021/jm991166h](https://doi.org/10.1021/jm991166h)
40. González MP, Caballero J, Tundidor-Camba A, Helguera AM, Fernández M (2006) *Bioorg Med Chem* 14:200. doi:[10.1016/j.bmc.2005.08.009](https://doi.org/10.1016/j.bmc.2005.08.009)
41. Polley MJ, Winkler DA, Burden FR (2004) *J Med Chem* 47:6230. doi:[10.1021/jm049621j](https://doi.org/10.1021/jm049621j)
42. Bolchi C, Pallavicini M, Rusconi C, Diomede L, Ferri N, Corsini A, Fumagalli L, Pedretti A, Vistolia G, Valoti E (2007) *Bioorg Med Chem Lett* 17:6192. doi:[10.1016/j.bmcl.2007.09.015](https://doi.org/10.1016/j.bmcl.2007.09.015)
43. Guida WC, Hamilton AD, Crotty JW, Sebt SM (2005) *J Comput Aided Mol Des* 19:871. doi:[10.1007/s10822-005-9030-2](https://doi.org/10.1007/s10822-005-9030-2)
44. Kohring K, Wiesner J, Altenkamper M, Sakowski J, Silber K, Hillebrecht A, Haebel P, Dahse HM, Ortmann R, Jomaa H, Klebe G, Schlitzer M (2008) *ChemMedChem* 3:1217. doi:[10.1002/cmdc.200800043](https://doi.org/10.1002/cmdc.200800043)
45. Tripos Inc. (2005) St Louis, MO 63144, USA. <http://www.tripos.com>
46. CCDC Software Ltd (2005) Cambridge CB2 1EZ, UK. <http://www.ccdc.cam.ac.uk>
47. GOLD User Guide & Tutorials, CCDC Software Ltd (2005) http://www.ccdc.cam.ac.uk/support/documentation/gold/4_0/gold/toc.html
48. Xie A, Clark SR, Prasanna S, Doerksen RJ (2009) *J Enzyme Inhib Med Chem* 24 (in press)