Original article

Receptor-based 3D-QSAR studies of checkpoint Wee1 kinase inhibitors

Kanin Wichapong [a,b], Marc Lindner [b], Somsak Pianwanit [a], Sirirat Kokpol [a], Wolfgang Sippl [b,*]

[a] Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
[b] Institute of Pharmaceutical Chemistry, Martin-Luther-University Halle-Wittenberg, 06120 Halle (Saale), Germany

ABSTRACT

One hundred and seventy-four pyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione derivatives reported as inhibitors of the kinase Wee1 were used for a molecular docking and three-dimensional quantitative structure–activity relationship (3D-QSAR) study. Due to the availability of the three-dimensional structure of the Wee1 kinase a receptor-based alignment strategy was applied. Six available Wee1-inhibitor crystal structures were analyzed using the docking program GOLD resulting in a good reproduction of the experimentally derived position and interaction of the cocrystallized inhibitors. Since only a low correlation between docking scores and inhibitory activities was obtained for the series of 174 inhibitors a receptor-based 3D-QSAR study was performed, dividing the data set into 144 training set molecules and an external test set of 30 compounds. Besides the ligand alignment derived from the docking study we tested several other alignment procedures as basis for the 3D-QSAR analysis. The most predictive model was obtained using the alignment from the GOLD docking study. The CoMFA model was found to be robust ($q^2_{LOO} = 0.764$ and $r^2 = 0.870$). The predictive ability of the model was further examined by carrying out leave-20%-out and leave-50%-out cross-validation ($q^2 = 0.747$ for leave-20%-out and 0.737 for leave-50%-out) and predicting the activities of 30 inhibitors used as external test set ($r^2_{pred} = 0.790$). The graphical analysis of the CoMFA contour plot together with the key residues of the binding pocket provided important insight into the relevant interactions of the inhibitors. The results not only provide information about the essential features of potent Wee1 inhibitors but also show the advantage of using receptor-based alignment for 3D-QSAR analysis.

1. Introduction

Transition one and two of mitosis is determined by the activity of cyclin dependent kinase Cdc2 complexed with cyclins A and B [1]. During the preceding S and G2 phases these complexes were inactivated after their formation by inhibitory phosphorylation of Thr14 and Tyr15 and by missing activation of phosphorylation of Ser161 [2,3]. Two members of Wee1 kinase subfamily mediate these inhibitory phosphorylations. Wee1A mediates phosphorylation of Tyr15 [4], and the dual-specific Myt1 phosphorylates both Thr14 and Tyr15 [5]. Many cancer cells lack a functional p53 signalling pathway, which means that the other significant DNA damage-sensitive checkpoint, G1/S, is not controlled. Wee1 inhibitors have been shown to abrogate the G2/M checkpoint and should preferentially enhance the cytotoxic effects of DNA damaging agents on p53-negative cells [6].

Although Wee1 is functionally a tyrosine kinase, in sequence and structure it most closely resembles serine/threonine kinases

such as the checkpoint kinase Chk1 [7,8]. The first crystal structure of Wee1 was solved in 2005 [9] and showed that although the ATP-binding pocket closely resembles that of other protein kinases, the activation segment contains Wee1-specific features that maintain it in an active conformation. In the last few years, a variety of Weel kinase inhibitors have been described [10,11]. Among them, Palmer et al. developed a series of phenylpyrrolocarbazoles as potent inhibitors of Weel. Using several of the synthesized compounds, further crystal structures of Wee1-inhibitor complexes have been solved recently [12]. The complexes provide important information about the interaction with the residues of the binding site and showed that the protein adopts the same conformation in all six crystal structures (Figs. 1 and 2).

Several studies have been published in which the combination of receptor-based methods and 3D-QSAR was successfully applied for the design and prediction of bioactive compounds [13–15]. The three-dimensional structure of a target protein, along with a docking protocol is used to guide alignment selection for comparative molecular field analysis. It is quite appealing to combine the accuracy of a receptor-based alignment with the computational efficiency of a ligand-based method. Receptor structures, either experimentally resolved or obtained by homology modeling, can provide important
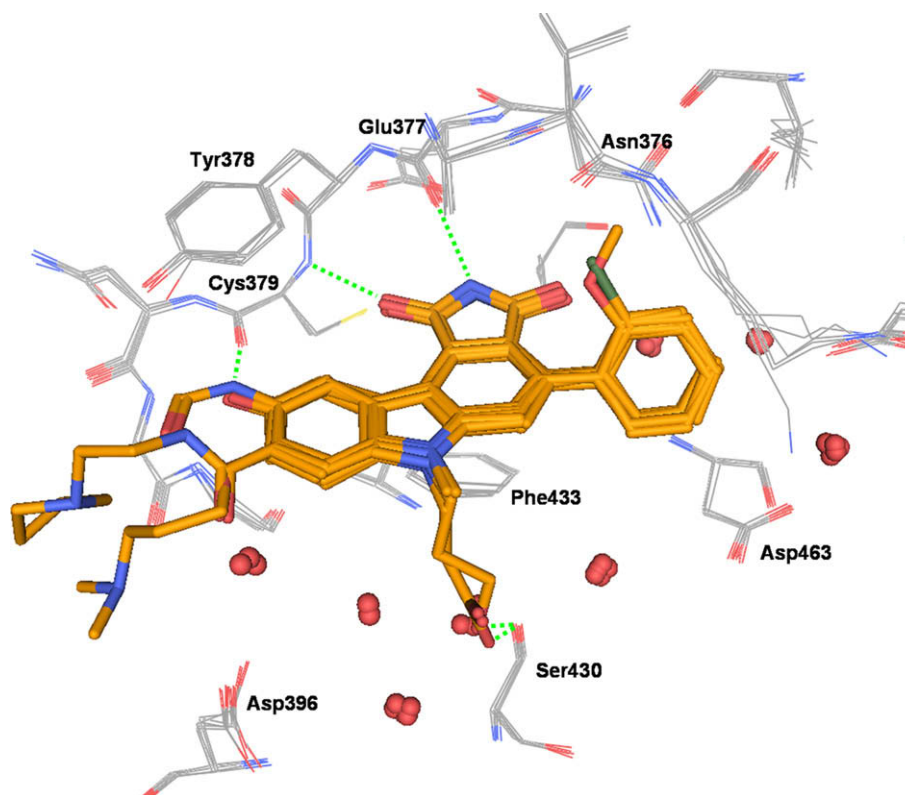
Fig. 1. Superimposition of the six solved crystal structures of Wee1-inhibitor complexes. The pyrrolocarbazole core of the inhibitors (colored orange) shows the same interaction at the ATP-binding site. Clusters of cocrystallized waters are displayed as balls. Hydrogen bonds are shown as dashed green line. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

information that is critical for an alignment in CoMFA [16], while QSAR can provide better prediction of binding energies [13].

In the present work we applied this receptor-based 3D-QSAR technique to a set of 174 Wee1 kinase inhibitors which has been recently developed [10–12]. The crystal structure of the catalytic domain of Wee1 together with an automatic docking program was used to determine the molecular alignment of the ligands. The 3D-QSAR model, which was obtained from the receptor-based alignment, yielded a high correlation between the experimentally determined binding affinity and the calculated molecular interaction fields. We were able to show that the receptor-based 3D-QSAR yields a better prediction of the binding affinity than using an interaction energy-based model or a ligand-based 3D-QSAR analysis.

2. Computational methods

2.1. Inhibitor data set

A set of pyrrolo[3,4-*c*]carbazole-1,3(2*H*,6*H*)-dione derivatives representing inhibitors of the kinase Wee1 [10–12] was used to generate a 3D-QSAR model applying the CoMFA methodology. All inhibitors were developed and tested in the same laboratory using the same assay condition, a prerequisite for generating reliable QSAR models. Inhibitors with no clear IC_{50} value (indicated with >50 μM) and inhibitors with undefined stereochemistry, were omitted. The 174 inhibitors were divided into training and test sets (Chart 1). In this way 144 ligands were randomly selected to generate the CoMFA model. The IC_{50} value of each inhibitor was converted into pIC_{50} ($-\log IC_{50}$) in order to use the data as a dependent variable in the CoMFA model. The structures and activity values (IC_{50} and pIC_{50}) of the studied inhibitors are displayed in Chart 1.

2.2. Ligand preparation

The 3D structures of the inhibitors were built based on the X-ray structure of 9-hydroxy-4-phenyl-6*H*-pyrrolo[3,4-*c*]carbazole-1,3-dione (ligand number **1**), which was cocrystallized with human
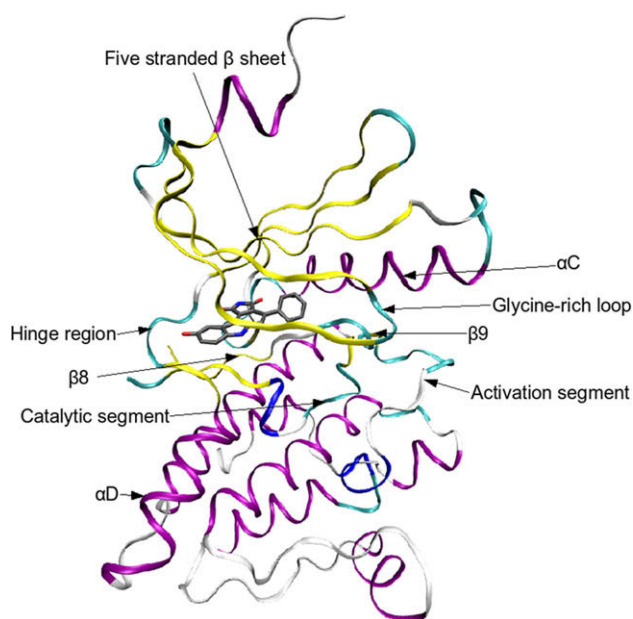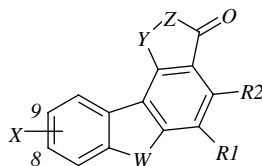

Fig. 2. Overall structure of the Wee1 kinase (X-ray structure of the Wee1/PD0407824 complex, pdb code 1X8B). α-Helices are colored magenta, β-sheets are colored yellow and loops are shown in cyan. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

Training Set (144 compounds)

Cpd	W	X	Y	Z	R1	R2	Activity	
							IC$_{50}$ (µM)	pIC$_{50}$
3	NH	9-OH	CO	NH	H	I	2.3	5.63
4	NH	8-OH	CO	NH	H	Ph	0.31	6.51
5	O	9-OH	CO	NH	H	Ph	0.43	6.37
6	S	9-OH	CO	NH	H	Ph	0.078	7.11
8	NH	9-OH	CO	NH	Me	Ph	0.13	6.89
9	NH	9-OH	CO	NH	Et	Ph	1.6	5.79
10	NH	9-OH	CO	NH	Ph	Me	9.7	5.01
11	NH	9-OH	CO	NH	Ph	Ph	2.3	5.63
12	NH	9-OH	CO	NH	Ph	H	4.0	5.39
13	NH	9-OH	CH$_2$	NH	H	Ph	37	4.43
14	NH	9-OH	CO	N-NH$_2$	H	Ph	3.9	5.40
15	NH	9-OH	CO	NH	H	2-ClPh	0.011	7.96
16	NH	9-OMe	CO	NH	H	2-ClPh	0.64	6.19
17	NMe	9-OH	CO	NH	H	2-ClPh	0.057	7.24
19	NH	9-OH	CO	NH	H	2-FPh	0.33	6.48
21	NH	9-OH	CO	NH	H	2-IPh	0.013	7.89
22	NH	9-OH	CO	NH	H	2-MePh	0.15	6.82
23	NH	9-OH	CO	NH	H	2-EtPh	0.51	6.29
24	NH	9-OH	CO	NH	H	2-CF$_3$Ph	0.58	6.24
25	NH	9-OH	CO	NH	H	2-CH$_2$OHPh	0.45	6.35
26	NH	9-OH	CO	NH	H	2-CNPh	0.19	6.72
27	NH	9-OH	CO	NH	H	2-COMePh	0.83	6.08
28	NH	9-OH	CO	NH	H	2-Ph-Ph	0.57	6.24
31	NH	9-OH	CO	NH	H	2-OEtPh	0.26	6.59
32	NH	9-OH	CO	NH	H	2-SMePh	0.033	7.48
33	NH	9-OH	CO	NH	H	2-SOMePh	0.22	6.66
34	NH	9-OH	CO	NH	H	2-NO$_2$Ph	0.047	7.33
35	NH	9-OH	CO	NH	H	2-NH$_2$Ph	0.21	6.68
36	NH	9-OH	CO	NH	H	3-FPh	0.22	6.66
37	NH	9-OH	CO	NH	H	3-ClPh	0.055	7.26
38	NH	9-OH	CO	NH	H	3-MePh	0.23	6.64
39	NH	9-OH	CO	NH	H	3-CH$_2$OHPh	0.87	6.06
40	NH	9-OH	CO	NH	H	3-CH$_2$NH$_2$Ph	4.4	5.35
41	NH	9-OH	CO	NH	H	3-CNPh	0.18	6.74
42	NH	9-OH	CO	NH	H	3-COMePh	4.3	5.36
45	NH	9-OH	CO	NH	H	3-OMePh	0.62	6.21
46	NH	9-OH	CO	NH	H	3-NO$_2$Ph	0.30	6.52
47	NH	9-OH	CO	NH	H	3-NH$_2$Ph	0.070	7.15
48	NH	9-OH	CO	NH	H	4-FPh	16	4.80
49	NH	9-OH	CO	NH	H	4-ClPh	0.73	6.14
52	NH	9-OH	CO	NH	H	4-CNPh	1.8	5.74
53	NH	9-OH	CO	NH	H	4-COMePh	3.6	5.44
54	NH	9-OH	CO	NH	H	4-OHPh	0.067	7.17
55	NH	9-OH	CO	NH	H	4-OMePh	12	4.90
57	NH	9-OH	CO	NH	H	4-SO$_2$MePh	1.1	5.95
58	NH	9-OH	CO	NH	H	4-NH$_2$Ph	0.15	6.82
59	NH	9-OH	CO	NH	H	2-Cl, 3-ClPh	0.028	7.55
60	NH	9-OH	CO	NH	H	2-Cl, 3-OHPh	0.012	7.55
61	NH	9-OH	CO	NH	H	2-Cl, 3-NH$_2$Ph	0.021	7.92
62	NH	9-OH	CO	NH	H	2-Cl, 4-OHPh	0.023	7.64
63	NH	9-OH	CO	NH	H	4-Cl, 3-NH$_2$Ph	0.024	7.62
64	NH	9-OH	CO	NH	H	2-Cl, 5-ClPh	0.49	6.31
66	NH	9-OH	CO	NH	H	2-Cl, 5-NH$_2$Ph	0.020	7.70
67	NH	9-OH	CO	NH	H	2-Cl, 6-ClPh	0.028	7.55

Cpd	W	X	Y	Z	R1	R2	Activity	
							IC_{50} (µM)	pIC_{50}
68	NH	9-OH	CO	NH	H	2-Cl, 6-OHPh	0.045	7.35
69	NH	9-OH	CO	NH	H	2-Cl, 6-OMePh	0.015	7.82
70	NH	9-OH	CO	NH	H	2-Br, 4-NH_2Ph	0.020	7.70
71	NH	9-OH	CO	NH	H	2-Br, 6-BrPh	0.035	7.46
72	NH	9-OH	CO	NH	H	2-Me, 3-MePh	0.27	6.57
73	NH	9-OH	CO	NH	H	2-Me, 5-MePh	0.96	6.02
74	NH	9-OH	CO	NH	H	2-Me, 6-MePh	0.075	7.12
75	NH	9-OH	CO	NH	H	2-OMe, 4-NH_2Ph	0.019	7.72
76	NH	9-OH	CO	NH	H	2-OMe, 5-NH_2Ph	0.11	6.96
77	NH	9-OH	CO	NH	H	2-OMe, 6-OMePh	0.027	7.57
78	NH	9-OH	CO	NH	H	2-OMe, 6-FPh	0.029	7.54
79	NH	9-OH	CO	NH	H	2,6-diCl, 3-OHPh	0.018	7.74
81	NH	9-OH	CO	NH	H	2-thienyl	0.14	6.85
82	NH	9-OH	CO	NH	H	3-thienyl	0.042	7.38
83	NH	9-OH	CO	NH	H	2-pyrrolyl	0.18	6.74
84	NH	9-OH	CO	NH	H	3-pyrrolyl	0.038	7.42
87	NEt	9-OH	CO	NH	H	2-ClPh	0.050	7.30
88	N-n-Pr	9-OH	CO	NH	H	2-ClPh	0.063	7.20
89	N-i-Pr	9-OH	CO	NH	H	2-ClPh	0.053	7.28
90	N-n-Bu	9-OH	CO	NH	H	2-ClPh	0.059	7.23
91	N(CH_2)$_2$$i$-Pr	9-OH	CO	NH	H	2-ClPh	0.15	6.82
92	N-n-pentyl	9-OH	CO	NH	H	2-ClPh	0.17	6.77
93	NH	9-OH	CO	NH	H	H	0.097	7.01
94	NH	9-OH	CO	NH	H	2,6-diClPh	0.028	7.55
95	NMe	9-OH	CO	NH	H	H	0.14	6.85
96	NMe	9-OH	CO	NH	H	2-ClPh	0.057	7.24
97	Nn-Bu	9-OH	CO	NH	H	2-ClPh	0.059	7.23
98	N(CH_2)$_2$OH	9-OH	CO	NH	H	H	0.025	7.60
99	N(CH_2)$_2$OH	9-OH	CO	NH	H	2-ClPh	0.045	7.35
101	N(CH_2)$_3$OH	9-OH	CO	NH	H	H	0.2	6.70
102	N(CH_2)$_3$OH	9-OH	CO	NH	H	2-ClPh	0.009	8.05
103	N(CH_2)$_3$OH	9-OH	CO	NH	H	2,6-diClPh	0.007	8.15
104	N(CH_2)$_3$OH	9-OH	CO	NH	H	2-OMePh	0.03	7.52
105	N(CH_2)$_2$$CONH_2$	9-OH	CO	NH	H	H	0.021	7.68
106	N(CH_2)$_2$$CONH_2$	9-OH	CO	NH	H	2-ClPh	0.006	8.22
107	N(CH_2)$_2$$CONH_2$	9-OH	CO	NH	H	2,6-diClPh	0.33	6.48
108	N(CH_2)$_2$CN	9-OH	CO	NH	H	2-ClPh	0.015	7.82
109	N(CH_2)$_2$COOMe	9-OH	CO	NH	H	2-ClPh	0.03	7.52
110	N(CH_2)$_3$CN	9-OH	CO	NH	H	2-ClPh	0.033	7.48
112	N(CH_2)$_2$$NMe_2$	9-OH	CO	NH	H	H	0.2	6.70
113	N(CH_2)$_2$$NMe_2$	9-OH	CO	NH	H	2-ClPh	0.096	7.02
114	N(CH_2)$_2$$NMe_2$	9-OH	CO	NH	H	2,6-diClPh	0.17	6.77
116	N(CH_2)$_2$Nmorpholide	9-OH	CO	NH	H	2-ClPh	0.064	7.19
117	N(CH_2)$_2$Nmorpholide	9-OH	CO	NH	H	2,6-diClPh	0.11	6.96
118	N(CH_2)$_2$Nimidazolide	9-OH	CO	NH	H	H	0.23	6.64
119	N(CH_2)$_2$Nimidazolide	9-OH	CO	NH	H	2-ClPh	0.092	7.04
120	N(CH_2)$_2$Nimidazolide	9-OH	CO	NH	H	2,6-diClPh	0.12	6.92
121	N(CH_2)$_3$NHMe	9-OH	CO	NH	H	H	0.28	6.55
122	N(CH_2)$_3$NHMe	9-OH	CO	NH	H	2-ClPh	0.069	7.16
123	N(CH_2)$_3$NHMe	9-OH	CO	NH	H	2,6-diClPh	0.11	6.96
125	N(CH_2)$_3$$NMe_2$	9-OH	CO	NH	H	2-ClPh	0.1	7.00
126	N(CH_2)$_3$$NMe_2$	9-OH	CO	NH	H	2,6-diClPh	0.14	6.85
127	N(CH_2)$_3$Nmorpholide	9-OH	CO	NH	H	H	0.29	6.54
128	N(CH_2)$_3$Nmorpholide	9-OH	CO	NH	H	2-ClPh	0.071	7.15
129	N(CH_2)$_3$Nmorpholide	9-OH	CO	NH	H	2,6-diClPh	0.064	7.19
132	N(CH_2)$_3$Nimidazolide	9-OH	CO	NH	H	2,6-diClPh	0.059	7.23

Cpd	W	X	Y	Z	R1	R2	Activity	
							IC$_{50}$ (µM)	pIC$_{50}$
133	N(CH$_2$)$_3$N(4-Mepiperazine)	9-OH	CO	NH	H	H	0.3	6.52
134	N(CH$_2$)$_3$N(4-Mepiperazine)	9-OH	CO	NH	H	2-ClPh	0.082	7.09
135	N(CH$_2$)$_3$N(4-Mepiperazine)	9-OH	CO	NH	H	2,6-diClPh	0.062	7.21
136	N(CH$_2$)$_3$NHPh	9-OH	CO	NH	H	H	0.093	7.03
139	N(CH$_2$)$_2$CONH(CH$_2$)$_2$NMe$_2$	9-OH	CO	NH	H	H	0.17	6.77
140	N(CH$_2$)$_2$CONH(CH$_2$)$_2$NMe$_2$	9-OH	CO	NH	H	2-ClPh	0.035	7.46
141	N(CH$_2$)$_2$CONH(CH$_2$)$_2$NMe$_2$	9-OH	CO	NH	H	2,6-diClPh	0.014	7.85
142	N(CH$_2$)$_2$COOH	9-OH	CO	NH	H	H	0.023	7.64
143	N(CH$_2$)$_2$COOH	9-OH	CO	NH	H	2-ClPh	0.009	8.05
144	N(CH$_2$)$_2$COOH	9-OH	CO	NH	H	2,6-diClPh	0.39	6.41
145	N(CH$_2$)$_2$CONHSO$_2$Me	9-OH	CO	NH	H	2-ClPh	0.012	7.92
146	N(CH$_2$)$_2$CONHSO$_2$Ph	9-OH	CO	NH	H	2-ClPh	0.007	8.15
147	N(CH$_2$)$_2$C-tetrazole	9-OH	CO	NH	H	2-ClPh	0.021	7.68
148	N(CH$_2$)$_2$S-triazole	9-OH	CO	NH	H	2-ClPh	0.024	7.62
149	N(CH$_2$)$_2$SO-triazole	9-OH	CO	NH	H	2-ClPh	0.009	8.05
152	N(CH$_2$)$_3$C-tetrazole	9-OH	CO	NH	H	2-ClPh	0.016	7.80
153	NH	8-(CH$_2$)$_4$-NMe$_2$, 9-OH	CO	NH	H	2-ClPh	0.049	7.31
154	NH	8-(CH$_2$)$_4$-Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.05	7.30
155	NH	8-(CH$_2$)$_4$-Nmorph, 9-OH	CO	NH	H	2-ClPh	0.037	7.43
156	NMe	8-(CH$_2$)$_4$-NMe$_2$, 9-OH	CO	NH	H	2-ClPh	0.034	7.47
157	NMe	8-(CH$_2$)$_4$-Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.036	7.44
159	N(CH$_2$)$_2$OH	8-(CH$_2$)$_4$-Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.024	7.62
160	N(CH$_2$)$_2$OH	8-(CH$_2$)$_4$-Nmorph, 9-OH	CO	NH	H	2-ClPh	0.019	7.72
161	NH	8-O(CH$_2$)$_3$-NMe$_2$, 9-OH	CO	NH	H	2-ClPh	0.026	7.59
163	NH	8-O(CH$_2$)$_3$-Nmorph, 9-OH	CO	NH	H	2-ClPh	0.026	7.59
164	NMe	8-O(CH$_2$)$_3$-NMe$_2$, 9-OH	CO	NH	H	2-ClPh	0.058	7.24
165	NMe	8-O(CH$_2$)$_3$-Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.075	7.12
167	N(CH$_2$)$_2$OH	8-O(CH$_2$)$_3$-NMe, 9-OH$_2$	CO	NH	H	2-ClPh	0.018	7.74
168	N(CH$_2$)$_2$OH	8-O(CH$_2$)$_3$-Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.024	7.62
169	N(CH$_2$)$_2$OH	8-O(CH$_2$)$_3$-Nmorph, 9-OH	CO	NH	H	2-ClPh	0.015	7.82
170	NMe	8-S(CH$_2$)$_3$-Npyrro, 9-OH	CO	NH	H	2-ClPh	0.02	7.70
171	NMe	8-SO(CH$_2$)$_3$-Npyrro, 9-OH	CO	NH	H	2-ClPh	0.033	7.48
173	NMe	8-SO$_2$NH(CH$_2$)$_2$-Npyrro, 9-OH	CO	NH	H	2-ClPh	0.046	7.34
174	NMe	8-CONH(CH$_2$)$_2$-Npyrro, 9-OH	CO	NH	H	2-ClPh	0.015	7.82

Test Set (30 compounds)

Cpd	W	X	Y	Z	R1	R2	Activity	
							IC$_{50}$ (µM)	pIC$_{50}$
1	NH	9-OH	CO	NH	H	Ph	0.097	7.01
2	NH	9-OH	CO	NH	H	H	4.0	5.39
7	NMe	9-OH	CO	NH	H	Ph	0.26	6.59
18	O	9-OH	CO	NH	H	2-ClPh	0.033	7.48
20	NH	9-OH	CO	NH	H	2-BrPh	0.023	7.64
29	NH	9-OH	CO	NH	H	2-OHPh	0.060	7.22
30	NH	9-OH	CO	NH	H	2-OMePh	0.024	7.62

Cpd	W	X	Y	Z	R1	R2	Activity IC$_{50}$ (µM)	pIC$_{50}$
43	NH	9-OH	CO	NH	H	4-Biphenyl	40	4.30
44	NH	9-OH	CO	NH	H	3-OHPh	0.089	7.05
50	NH	9-OH	CO	NH	H	4-MePh	3.3	5.48
51	NH	9-OH	CO	NH	H	4-CH$_2$OHPh	1.2	5.92
56	NH	9-OH	CO	NH	H	4-SMe	29	4.50
65	NH	9-OH	CO	NH	H	2-Cl, 5-OHPh	0.042	7.38
80	NH	9-OH	CO	NH	H	2,6-diCl, 4-OHPh	0.049	7.31
85	NH	9-OH	CO	NH	H	4-pyridyl	0.82	6.09
86	NH	9-OH	CO	NH	H	3-pyridyl	0.58	6.24
101	N(CH2)$_2$OH	9-OH	CO	NH	H	2,6-diClPh	0.008	8.10
112	N(CH2)$_3$OMe	9-OH	CO	NH	H	2-ClPh	0.027	7.57
116	N(CH2)$_2$Nmorpholide	9-OH	CO	NH	H	H	0.14	6.85
125	N(CH2)$_3$NMe2	9-OH	CO	NH	H	H	0.36	6.44
131	N(CH2)$_3$Nimidazolide	9-OH	CO	NH	H	H	0.11	6.96
132	N(CH2)$_3$Nimidazolide	9-OH	CO	NH	H	2-ClPh	0.054	7.27
138	N(CH2)$_3$NHPh	9-OH	CO	NH	H	2-ClPh	0.074	7.13
139	N(CH2)$_3$NHPh	9-OH	CO	NH	H	2,6-diClPh	0.067	7.17
151	N(CH2)$_2$SO$_2$triazole	9-OH	CO	NH	H	2-ClPh	0.019	7.72
152	N(CH2)$_2$COOH	9-OH	CO	NH	H	2-ClPh	0.013	7.89
159	NMe	8-(CH2)$_4$-N-morph, 9-OH	CO	NH	H	2-ClPh	0.03	7.52
163	NH	8-O(CH2)$_3$-N-pyrrol, 9-OH	CO	NH	H	2-ClPh	0.036	7.44
167	NMe	8-O(CH2)$_3$-N-morph, 9-OH	CO	NH	H	2-ClPh	0.057	7.24
173	NMe	8-SO$_2$(CH2)$_3$-N-pyrrol, 9-OH	CO	NH	H	2-ClPh	0.16	6.80

Wee1 kinase (PDB code 1X8B) [11,17]. The molecular structures of the inhibitors were generated using Sybyl7.2 [18]. The structures were energy minimized using the MMFF94s force field [19] and the BFGS method [20–23], until the default derivative convergence criterion of 0.05 kcal/mol was reached. All compounds were generated in the protonation state at pH 7.1 using the MOE protonate 3D method [24].

2.3. Molecular docking

To predict the appropriate binding orientation of Weel kinase inhibitors the molecular docking program GOLD (version 3.2

Fig. 3. Comparison of the docking solution for **1** (colored cyan) compared with its position in the crystal structure (colored grey, hydrogen bonds are shown as dashed line, cocrystallized water molecules as red balls and the magnesium ion as green ball). For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

[25,26]) was employed to generate an ensemble of docking conformations. To test whether GOLD is able to reproduce the experimental data we selected the six available Wee1-inhibitor complexes from the PDB databank (entries 1X8B, 2ZZW, 2IN6, 2IO6, 3BI6, and 3BIZ). The protein structures were prepared for docking using the Sybyl7.2 software [18]. The original ligand and all ions were removed from the Weel protein complexes. The default GOLD parameters were used except that the 'flood fill radius' was set to 20 Å around Cys379.

In the Wee1 crystal structures a variety of cocrystallized water molecules are found in close proximity to the bound inhibitors (Fig. 1). Since the docking of the six inhibitors into the corresponding protein structures resulted in low accuracy (high RMSD values for the top-ranked docking pose) when no water molecules or constraints were used, we tested a variety of further docking settings. First, we used the water toggling mode within the GOLD program. Eight clusters of cocrystallized water molecules (shown as red balls in Fig. 1) which were observed in close proximity to the inhibitors in the Wee1 X-ray structures were considered as potential interaction sites for the docked inhibitors. For interpretation of the references to color in this paragraph and subsequent paragraphs, the reader is referred to the web version of this article.

As a second setting, we considered two hydrogen bonds to the Wee1 hinge backbone region to be relevant for inhibitor binding. Therefore two protein hydrogen bond constraints were defined in GOLD to the backbone oxygen of Glu377 and to the backbone NH of Cys379. The ligands were scored based on the fitness function 'GoldScore'. GOLD was run to save up to 10 top-ranked docking solutions for the ligands. The results were visually analyzed using MOE2006.8 [24].

Both docking settings resulted in one favourable cluster of docking poses in case of the reference inhibitor **1** (RMSD values between 0.347 and 0.520 for the 10 top-ranked docking poses), whereas for the more flexible inhibitors several individual clusters
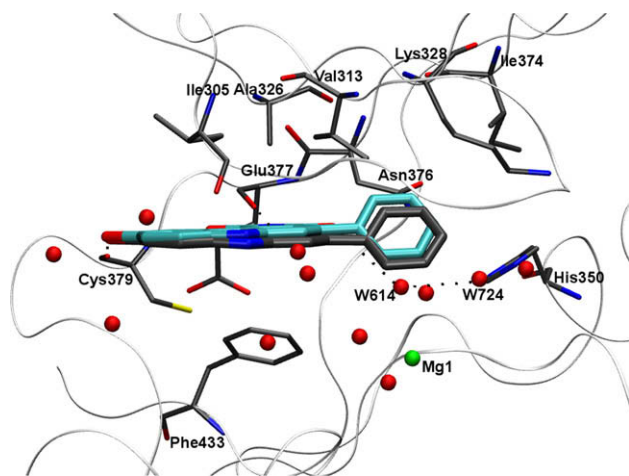
were detected. In the case of several clusters we selected the ligand conformation which showed the highest GoldScore.

2.4. Ligand alignment method

The result of a CoMFA model strongly depends on the quality of the ligand alignment. However, when studying a large number of ligands there is no certain preferable alignment method. Therefore, we tested three different alignment procedures in order to identify the most efficient alignment approach for this data set. The first alignment (Alignment 1) was derived using the top-ranked GOLD docking pose for each inhibitor.

For the ligand-based Alignment 2, a flexible alignment was carried out using the structure of compound **1** derived from the crystal structure 1X8B.pdb and using the flexible alignment tool within MOE2006.8. In this procedure, a ligand-based alignment was derived using the similarity function as a criteria to align each molecule onto the conformation of the potent inhibitor **1** as template structure.

Alignment 3 was derived by refining the GOLD docking solutions using the Amber force field 99 [27]. The minimization of the Wee1-inhibitor complexes was performed using the MOE2006.08 program. PEOE charges [28] were assigned for the inhibitors whereas the Amber force field 99 charges were applied for the kinase. During the energy minimization process, the kinase was constrained by tethering heavy atoms (force constant 100 kcal/mol) whereas the inhibitor structures were relaxed.

2.5. CoMFA model

CoMFA calculations were carried out by applying the default settings. The standard CoMFA fields performing the Lennard–Jones potential for the steric and Coulomb potential for the electrostatic were used. A cut-off value for the fields was set at 30 kcal/mol. Both steric and electrostatic interactions at each grid point were calculated by applying the C-sp3 probe atom with $+1$ charge. Three different grid spacing values, 1 Å, 1.5 Å and 2 Å, were employed in order to investigate the influence of the grid spacing. The Sybyl7.2 program [18] was used for generating the CoMFA models.

To extend the scope of CoMFA the region focusing approach was applied. Region focusing is a procedure to refine CoMFA models by contributing the weight to the lattice points. Several region focusing procedures such as StDev × Coefficients (StDev.), Discriminant power (Disc.), Sigma field (Sigma) and PLS coefficient (PLS) as implemented in Sybyl7.2 were tested.

2.6. Partial least-square (PLS) analysis

To form the basis for a statistical significant model, the method of partial least-squares (PLS) regression was used to analyze the inhibitors by correlating variations in their biological activities with variations in their interaction fields. The optimum number of PLS components corresponding to the smallest standard error of prediction was determined by the leave-one-out cross-validation procedure. Using the optimal number of components, the final PLS analysis was carried out without cross-validation to generate a predictive model with a conventional correlation coefficient. The leave-one-out cross-validation method might lead to high q^2 values, which do not necessarily reflect a general predictive ability of a model. Therefore further cross-validation, using five and two groups of approximately the same size in which the objects were assigned randomly, was performed. In this method 80% or 50% of the compounds were randomly selected and a model is generated, which is then used to predict the remaining compounds (leave-20%-out, leave-50%-out). This cross-validation technique, especially

the leave-50%-out procedure, has been found to be stricter and to give a better estimate of the robustness of a model than the normal leave-one-out procedure [13–15]. However, it must be stated that there is no qualitative difference between the different cross-validation procedures; all methods have the problem that they are not able to estimate the external predictivity of a QSAR model. Therefore, an external test set containing 30 randomly selected inhibitors was used in addition to test the predictivity of the final CoMFA model.

3. Results and discussion

3.1. Examination of the ATP and inhibitor binding site

In a first step the X-ray structures of Wee1 (Figs. 1 and 2) in complex with six pyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione derivatives were analyzed. All cocrystallized inhibitors form several hydrogen bonds between its maleimido ring and the backbone atoms of Glu377 and Cys379 and the sidechain amide of Asn376 (Fig. 3). These interactions are well-known from other kinase inhibitors, like staurosporine. Two cocrystallized water molecules mediate a hydrogen bond to His350. The phenyl ring of the inhibitors occupies a hydrophobic pocket between the sidechains of Val313, Lys328 and Ile374. Stabilization of the planar aromatic system is provided by Ile305, Val313, Ala326 and Phe433 [9].

3.2. Ligand docking

We first analyzed, whether GOLD is able to reproduce the position of the cocrystallized inhibitors observed in the six Wee1 X-ray structures. Docking of the inhibitors into the Wee1 ATP-binding site, using two hydrogen bonding constraints and without considering the cocrystallized water molecules, yielded slightly lower root-mean square deviation (RMSD) values compared to the docking when ordered water molecules were considered to be part of the protein (toggeling water mode in GOLD) (Table 1). Therefore we used these two hydrogen bond constraints for all 174 pyrrolocarbazole derivatives. Due to the structural similarity of the analyzed data set, it is likely that all active compounds possessing the pyrrolocarbazole moiety show a similar interaction at the Wee1 ATP-binding pocket. The superimposition of the 174 compounds (Fig. 4) reveals a well conserved overall conformation, particularly for the pyrrolocarbazole ring, with slight variations in the orientation of the attached sidechains. Besides the four hydrogen bonds of the pyrrolocarbazole to Ans376 (sidechain amide), Glu377 (backbone CO), and Cys379 (backbone NH and CO), the most potent inhibitor **106** forms an additional hydrogen bond between the N6-substituent and the backbone CO of Ser430 (Fig. 5(A) and (B)).
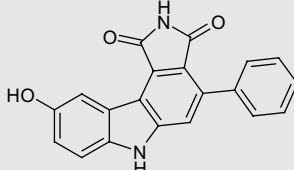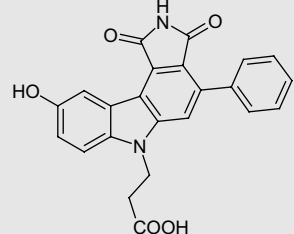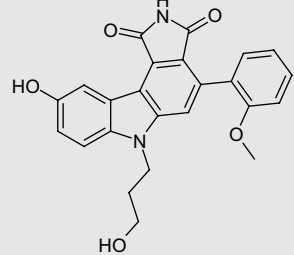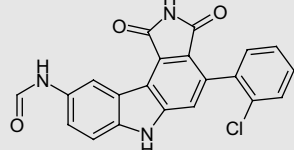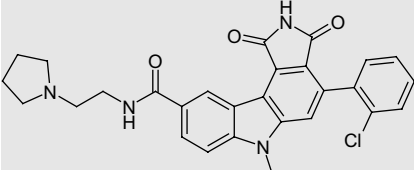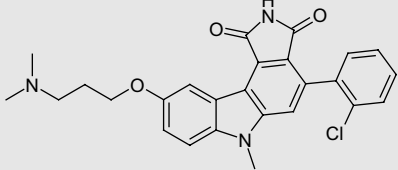
Subsequently, we analyzed whether the docking scores (top-score and average score) can be correlated to the biological data of the inhibitors. Only a low correlation was observed for the studied data set of 174 inhibitors ($r^2 = 0.31$). To counteract the tendency of larger molecules to produce better docking scores, the calculated scores were divided by the square root of the number of heavy atoms [29,30]. However, this resulted in no significant improvement of the correlation coefficient ($r^2 = 0.34$).

3.3. CoMFA models

In this work, statistical models linking the biological activities to the different descriptors were built by means of partial least-square (PLS) regression, the degree of correlation of experimental vs. predicted values was expressed in terms of the square of the correlation coefficient (r^2), indicating the fraction of explained variance, and the internal predictability was measured in terms of cross-validated r^2, hereafter referred to as q_{LOO}^2, after cross-

Table 1
RMSD values between crystal structure and the top-ranked docking pose using different docking settings.

Cocrystallized inhibitor	PDB code	RMSD water (Å)	RMSD constr. (Å)
	1X8B	0.437	0.434
	2IN6	1.210	0.793
	2IB6	1.257	0.585
	2Z2W	0.373	0.431
	3BI6	0.637	0.504
	3BIZ	0.445	0.491

RMSD water: RMSD observed for the docking using the toggeling water mode in GOLD. RMSD constr.: RMSD observed for the docking using two hydrogen bond constraints.
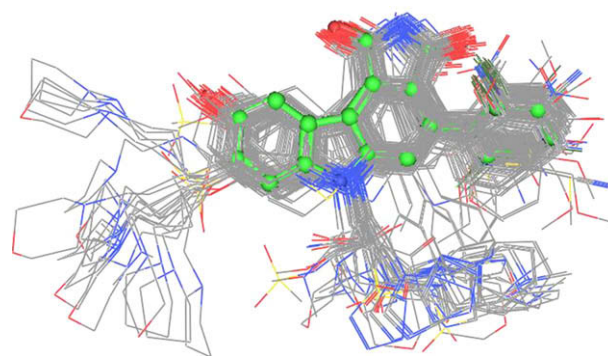

Fig. 4. Docking-based Alignment 1 derived from the GOLD ligand docking. Inhibitor **1** is shown in green. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

validation using the leave-one-out method. A first CoMFA model was generated by using 144 randomly selected compounds as training set (Chart 1). Analyzing the CoMFA model ($q_{LOO}^2 = 0.474$, SDEP = 0.558, $r^2 = 0.706$ and SDEE = 0.417) we found that ligands
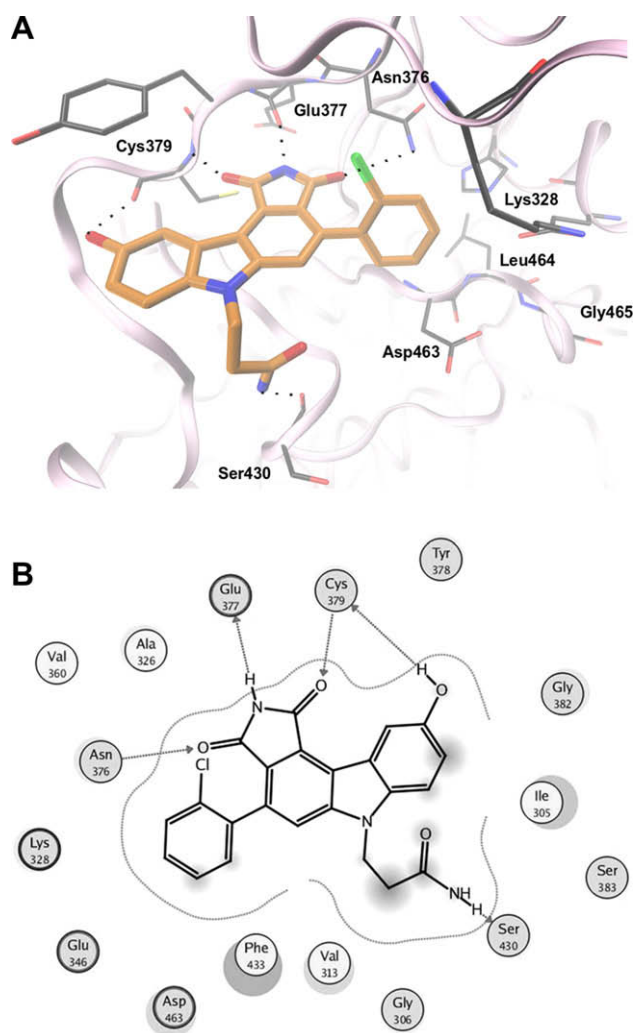

Fig. 5. (A) GOLD docking solution for the most potent inhibitor **107** (orange). Hydrogen bonds are shown as dashed line. (B) Schematic representation of the interaction of the most potent inhibitor **106** with the residues of the Wee1 binding site. Hydrogen bonds are indicated as arrows. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

Table 2
Statistics of generated CoMFA models using different alignment methods.

Model	Grid spacing (Å)	Leave-one-out cross-validation			Non-cross-validation			Field contribution	
		q^2	Comp.	SDEP	r^2	SDEE	F-value	Steric	Elec.
Alignment 1 (docking-based alignment)									
1	1.0	0.549	6	0.475	0.773	0.337	75.590	0.546	0.454
2	**1.5**	**0.685**	**6**	**0.397**	**0.836**	**0.286**	**113.017**	**0.570**	**0.430**
3	2.0	0.624	6	0.433	0.801	0.316	89.120	0.546	0.454
Alignment 2 (ligand-based flexible alignment)									
4	1.0	0.506	4	0.488	0.625	0.430	56.319	0.565	0.435
5	1.5	0.502	4	0.495	0.614	0.436	53.769	0.569	0.431
6	2.0	0.502	5	0.497	0.656	0.413	51.118	0.561	0.439
Alignment 3 (alignment from protein-inhibitor minimization)									
7	1.0	0.470	5	0.513	0.682	0.397	57.509	0.486	0.514
8	1.5	0.462	5	0.517	0.672	0.403	54.991	0.485	0.515
9	2.0	0.462	5	0.517	0.676	0.401	55.890	0.482	0.518

13, **40**, **48**, **107** and **144** have high residual values (actual pIC_{50} vs. predicted pIC_{50}). Their residual values are 2.0 times higher as the SDEP of the model indicating that these five ligands are being truly outliers. We visually analyzed the docking solutions of the outliers and tried to find an explanation for that. In the case of compound **13**, which is the only compound without a second carbonyl group at the pyrrolidine ring, the 3D-QSAR model was not able to cover this difference. In addition, a variety of different possible binding orientations were observed for compound **13** in the docking study. This variability was only observed for compound **13**, whereas all other inhibitors show the conserved binding mode due to the further hydrogen bonding possibilities. Compound **40** is the only inhibitor which possesses an aliphatic amino group ($-CH_2NH_2$) attached to the phenyl ring, and it is not clear whether the amino group is protonated or not when interacting at the buried hydrophobic part of the binding pocket. To test the influence of the protonation state we generated a variety of different CoMFA models (all aliphatic amines protonated, all amines neutral, only the amino group of **40** neutral) but compound **40** was an outlier in all the models. In the case of compound **48** it is not clear why the fluoro substituent in position 4 results in a complete loss of activity, also the QSAR model is not able to explain this discrepancy. Similar observations were made for compounds **107** and **144**. It was interesting to see that the five compounds were also outliers in the CoMFA model derived from the ligand-based alignment (data not shown).

Therefore, we removed the five outliers and generated a CoMFA model for the remaining 139 inhibitors. A variety of CoMFA models were generated based on three different alignments and testing different CoMFA settings. Table 2 summarizes the statistics derived for the different models. The best CoMFA model was obtained from the docking-based Alignment 1 (model 2, $q^2_{LOO} = 0.685$, SDEP = 0.397, $r^2 = 0.836$ and SDEE = 0.286), whereas the other two alignment procedures resulted in lower q^2 values. In addition, we applied a region focusing strategy to analyze whether reducing the number of variables is able to increase the predictive ability of the CoMFA model 2. The resulting CoMFA model of the 139 ligands with PLS region focusing method (model 10 exhibited in Table 3) gave the model with the highest $q^2_{LOO} = 0.764$.

The actual pIC_{50}, the predicted pIC_{50} and the estimated pIC_{50} of each inhibitor, which was calculated from the CoMFA model 10, are compared and listed in Table 4. Actual pIC_{50} vs. predicted pIC_{50} and estimated pIC_{50} values are plotted on the graph in Fig. 6(A) and (B), respectively. The cross-validation demonstrates that the predicted pIC_{50} values are in good agreement with the actual pIC_{50}. The leave-one-out cross-validation method might lead to high q^2 values, which do not necessarily reflect a predictivity of a model. Therefore, further cross-validation, using five and two groups of approximately the same size in which the objects were assigned randomly, was performed. These models have been shown in the literature to provide a better quantitative estimate of the robustness of a QSAR model. Thus, q^2 values for leave-20%-out (0.747) and leave-50%-out (0.737) were determined. The high q^2 values of both cross-validation procedures, which are in the same range as the leave-one-out value, indicate that even with a limited number of molecules robust QSAR models could be derived. In addition, CoMFA model 10 was also applied to predict the activity of the external test set containing 30 inhibitors which were not used for model generation (Table 5). The prediction shows a good correlation between the actual and the predicted pIC_{50} values with $r^2_{pred} = 0.790$ (Fig. 6(C)). In addition, CoMFA models 4 and 7 were also used to predict the activities of the test set but resulted in much lower predictive r^2 values ($r^2_{pred} = 0.551$ and 0.501, respectively).

3.4. CoMFA graphical contour plot

The subsequent PLS analysis, using the energy fields as descriptors and the biological activity as the dependent variable, can highlight the relative importance for affinity of certain types of interaction and certain regions around the compounds. Since the structure of the Wee1 kinase is known, the results obtained by the 3D-QSAR analysis were compared with the geometry and properties of the binding pocket. It is necessary to note that, in general, such comparison should be attempted carefully. In receptor-based 3D-QSAR models, the variance observed in the field values reflects both the structural diversity of the ligands and their diverse location within the binding site. The PLS coefficient contour maps can by no means be regarded as a set of low resolution picture of the

Table 3
Statistics of the resulting CoMFA models.

Model	RF	Leave-one-out cross-validated			Non-cross-validated			Field contribution	
		q^2	Comp.	SDEP	r^2	SDEE	F-value	Steric	Elec.
Receptor-based CoMFA model									
2	–	**0.685**	**6**	**0.397**	**0.836**	**0.286**	**113.017**	**0.570**	**0.430**
CoMFA model, region focusing method									
10	**PLS**	**0.764**	**6**	**0.342**	**0.870**	**0.254**	**148.973**	**0.478**	**0.522**

Table 4
List of actual pIC$_{50}$, predicted pIC$_{50}$ (derived from leave-one-out cross-validation), estimated pIC$_{50}$ values and residual for the 139 training set compounds as derived from the CoMFA model 10.

Cpd	Actual pIC$_{50}$	Pred. pIC$_{50}$	Res.	Est. pIC$_{50}$	Res.	Cpd	Actual pIC$_{50}$	Pred. pIC$_{50}$	Res.	Est. pIC$_{50}$	Res.
3	5.63	5.81	−0.18	5.71	−0.08	90	7.23	6.94	0.29	7.02	0.21
4	6.51	6.87	−0.36	6.80	−0.29	91	6.82	7.12	−0.30	7.01	−0.19
5	6.37	6.77	−0.40	6.62	−0.25	92	6.77	7.33	−0.56	7.16	−0.39
6	7.11	6.72	0.39	6.83	0.28	93	7.01	7.01	0.00	7.01	0.00
8	6.89	6.32	0.57	6.42	0.47	94	7.55	7.35	0.20	7.38	0.17
9	5.79	6.36	−0.57	6.17	−0.38	95	6.85	7.15	−0.30	7.10	−0.25
10	5.01	5.39	−0.38	5.12	−0.11	96	7.24	7.57	−0.33	7.55	−0.31
11	5.63	5.64	−0.01	5.60	0.03	97	7.23	7.55	−0.32	7.50	−0.27
12	5.39	5.17	0.22	5.17	0.22	98	7.60	7.33	0.27	7.39	0.21
14	5.40	5.08	0.32	5.08	0.32	99	7.35	7.72	−0.37	7.63	−0.28
15	7.96	7.29	0.67	7.36	0.60	101	6.70	7.37	−0.67	7.22	−0.52
16	6.19	6.79	−0.60	6.52	−0.33	102	8.05	7.52	0.53	7.63	0.42
17	7.24	7.60	−0.36	7.53	−0.29	103	8.15	7.51	0.64	7.68	0.47
19	6.48	6.58	−0.10	6.60	−0.12	104	7.52	7.47	0.05	7.47	0.05
21	7.89	7.35	0.54	7.47	0.42	105	7.68	7.30	0.38	7.47	0.22
22	6.82	7.39	−0.57	7.29	−0.47	106	8.22	7.57	0.65	7.73	0.49
23	6.29	6.45	−0.16	6.40	−0.11	108	7.82	7.63	0.19	7.64	0.18
24	6.24	6.58	−0.34	6.41	−0.17	109	7.52	7.55	−0.03	7.53	−0.01
25	6.35	6.85	−0.50	6.64	−0.29	110	7.48	7.40	0.08	7.42	0.07
26	6.72	7.34	−0.62	7.28	−0.56	111	6.70	6.53	0.18	6.61	0.09
27	6.08	6.41	−0.33	6.23	−0.15	113	7.02	7.01	0.01	6.97	0.05
28	6.24	6.02	0.22	6.08	0.16	114	6.77	6.97	−0.20	6.86	−0.09
31	6.59	6.35	0.24	6.44	0.15	116	7.19	7.32	−0.13	7.12	0.07
32	7.48	7.36	0.12	7.41	0.07	117	6.96	6.85	0.12	6.82	0.14
33	6.66	6.75	−0.09	6.59	0.07	118	6.64	6.88	−0.24	6.80	−0.16
34	7.33	7.77	−0.44	7.60	−0.27	119	7.04	7.24	−0.20	7.16	−0.12
35	6.68	7.35	−0.67	7.26	−0.58	120	6.92	7.11	−0.19	7.01	−0.09
36	6.66	6.86	−0.20	6.87	−0.21	121	6.55	6.56	−0.01	6.55	0.00
37	7.26	6.81	0.45	6.89	0.37	122	7.16	6.58	0.58	6.86	0.30
38	6.64	6.52	0.12	6.57	0.07	123	6.96	7.19	−0.23	6.78	0.18
39	6.06	6.50	−0.44	6.35	−0.29	125	7.00	6.72	0.28	6.78	0.22
41	6.74	6.37	0.37	6.44	0.30	126	6.85	6.98	−0.13	6.83	0.02
42	5.36	6.25	−0.89	5.66	−0.30	127	6.54	6.79	−0.25	6.72	−0.18
45	6.21	6.38	−0.17	6.31	−0.10	128	7.15	7.10	0.05	7.18	−0.03
46	6.52	7.05	−0.53	6.97	−0.45	129	7.19	7.09	0.11	7.19	0.00
47	7.15	7.14	0.01	7.17	−0.02	132	7.23	6.99	0.24	7.07	0.16
49	6.14	6.75	−0.61	6.65	−0.51	133	6.52	6.76	−0.24	6.65	−0.13
52	5.74	5.78	−0.04	5.77	−0.03	134	7.09	6.90	0.19	6.99	0.10
53	5.44	5.30	0.14	5.27	0.17	135	7.21	7.00	0.21	7.12	0.09
54	7.17	6.94	0.23	6.98	0.19	136	7.03	7.12	−0.09	7.22	−0.19
55	4.90	4.97	−0.07	4.79	0.11	139	6.77	7.24	−0.47	7.02	−0.25
57	5.95	6.27	−0.32	6.10	−0.15	140	7.46	7.41	0.05	7.68	−0.22
58	6.82	6.97	−0.15	6.99	−0.17	141	7.85	8.01	−0.16	8.05	−0.20
59	7.55	7.16	0.39	7.25	0.30	142	7.64	7.89	−0.25	7.75	−0.11
60	7.55	7.16	0.39	7.37	0.18	143	8.05	8.16	−0.11	8.08	−0.03
61	7.92	7.18	0.74	7.28	0.64	145	7.92	7.38	0.54	7.63	0.30
62	7.64	7.14	0.50	7.25	0.39	146	8.15	8.04	0.11	8.31	−0.16
63	7.62	7.21	0.41	7.30	0.32	147	7.68	7.45	0.23	7.52	0.16
64	6.31	6.86	−0.55	6.77	−0.46	148	7.62	7.75	−0.13	7.74	−0.12
66	7.70	7.00	0.70	7.14	0.56	149	8.05	7.71	0.34	7.94	0.11
67	7.55	7.46	0.09	7.49	0.06	152	7.80	7.47	0.33	7.57	0.24
68	7.35	7.52	−0.17	7.50	−0.15	153	7.31	7.39	−0.08	7.31	0.00
69	7.82	7.42	0.40	7.55	0.27	154	7.30	7.39	−0.09	7.31	−0.01
70	7.70	7.44	0.26	7.51	0.19	155	7.43	7.28	0.15	7.27	0.16
71	7.46	7.25	0.21	7.29	0.17	156	7.47	7.66	−0.19	7.44	0.03
72	6.57	7.24	−0.67	7.06	−0.49	157	7.44	7.21	0.23	7.34	0.11
73	6.02	6.23	−0.21	6.02	0.00	159	7.62	7.74	−0.12	7.71	−0.09
74	7.12	6.84	0.29	6.85	0.27	160	7.72	7.57	0.15	7.58	0.14
75	7.72	7.93	−0.21	7.92	−0.20	161	7.59	7.55	0.04	7.61	−0.02
76	6.96	6.97	−0.01	6.96	0.00	163	7.59	7.66	−0.07	7.67	−0.08
77	7.57	7.38	0.19	7.51	0.06	164	7.24	7.43	−0.19	7.37	−0.13
78	7.54	7.73	−0.19	7.72	−0.18	165	7.12	7.55	−0.43	7.29	−0.17
79	7.74	7.44	0.30	7.61	0.13	167	7.74	7.50	0.24	7.72	0.02
81	6.85	6.90	−0.05	6.91	−0.06	168	7.62	7.68	−0.06	7.75	−0.13
82	6.32	6.92	−0.60	7.05	−0.73	169	7.82	7.94	−0.12	7.97	−0.15
83	6.74	6.62	0.12	6.68	0.06	170	7.70	7.69	0.01	7.69	0.01
84	7.42	6.67	0.75	6.88	0.54	171	7.48	7.28	0.20	7.72	−0.24
87	7.30	7.56	−0.26	7.54	−0.24	173	7.34	7.53	−0.19	7.31	0.03
88	7.20	7.50	−0.30	7.44	−0.24	174	7.82	7.62	0.20	7.96	−0.14
89	7.28	7.57	−0.29	7.51	−0.23						

Cpd = compound number, Pred. pIC$_{50}$ = predicted pIC$_{50}$, Est. pIC$_{50}$ = estimated pIC$_{50}$ and Res. = residual value.
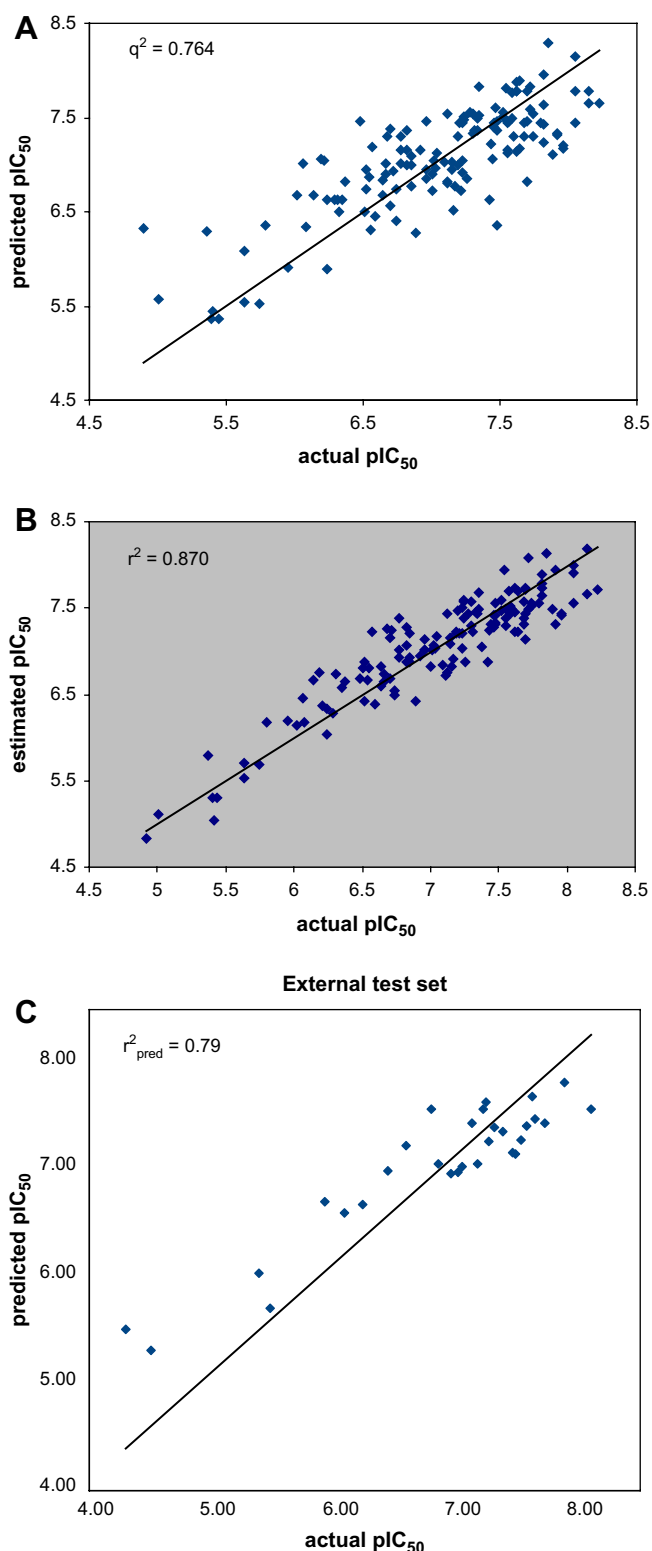
Fig. 6. (A) Correlation between the actual pIC$_{50}$ and the predicted pIC$_{50}$ obtained with the best CoMFA model 10. (B) Correlation between the actual pIC$_{50}$ and the estimated pIC$_{50}$ of the best CoMFA model 10. (C) Prediction of the external test set containing 30 inhibitors.

Table 5
Actual pIC$_{50}$, predicted pIC$_{50}$ and residual values of the 30 test set compounds (derived from the final CoMFA model 10).

Cpd	Actual pIC$_{50}$	Predicted pIC$_{50}$	Residual
1	7.01	6.98	0.03
2	5.39	5.92	−0.53
7	6.59	7.17	−0.58
18	7.48	7.18	0.30
20	7.64	7.47	0.17
29	7.22	7.44	−0.22
30	7.62	7.73	−0.11
43	4.30	5.37	−1.07
44	7.05	7.03	0.02
50	5.48	5.70	−0.22
51	5.92	6.64	−0.72
56	4.50	5.29	−0.79
65	7.38	7.39	−0.01
80	7.31	7.37	−0.06
85	6.09	6.70	−0.61
86	6.24	6.76	−0.53
100	8.10	7.51	0.58
111	7.57	7.57	0.00
115	6.85	6.75	0.10
124	6.44	6.81	−0.37
130	6.96	6.79	0.17
131	7.27	7.15	0.12
137	7.13	7.38	−0.25
138	7.17	7.23	−0.05
150	7.72	7.68	0.04
151	7.89	7.63	0.25
160	7.52	7.47	0.05
162	7.46	7.03	0.42
166	7.24	7.55	−0.31
172	6.80	7.56	−0.77

binding site, since the contour maps reflect only those regions in space, where the ligand–probe interaction energy is correlated with a variance of the biological activity. However, it provides an opportunity to interpret features indicated in the contour maps with respect to the protein environment and to check whether the variance in the field values corresponds to regions known to be important for inhibitor binding. We superimposed the coefficient contour maps and the Wee1 ATP-binding pocket. Fig. 7(A) and (B) shows the plot of the PLS coefficients for the steric and electrostatic fields, respectively.

The steric interaction is represented by green and yellow contours, in which green-colored regions indicate areas where increased steric bulk is associated with enhanced activity, and yellow regions suggest areas where increased steric bulk is unfavourable. Electrostatic interaction is indicated by red and blue contours, among which blue-colored regions show areas where more positively charged groups are favoured, and red regions highlight areas where groups with more negative partial charges are favoured.

A large favourable steric field (green) is observed around Ala326 and Asn376 next to the *ortho* position of the phenyl ring. Potent inhibitors without further activity-increasing substituents at the pyrrolocarbazole core, such as **21**, **29**, **30**, **32**, **34**, and **69** possess substituents at the *ortho* position of the phenyl ring. A second smaller favourable PLS region is found nearby the *para* position of the phenyl ring. However, the size of the hydrophobic pocket is restricted by His350, Glu346 and Leu464. Compounds with too large substituents at positions 3, 4 and 5 of the phenyl ring (**42**, **43**, **55**, **56** and **57**) are able to interact with the hydrophobic pocket but show no optimal position to make the important hydrogen bonds to the hinge backbone region. Based on these results we suggest that adding a smaller electronegative substituent at position 6 of the phenyl ring would be favourable when combined with the 2-Cl substituent. The prediction of the fluoro analog of **106** (2-Cl, 6-FPhe) resulted in a higher pIC$_{50}$ value compared to the original
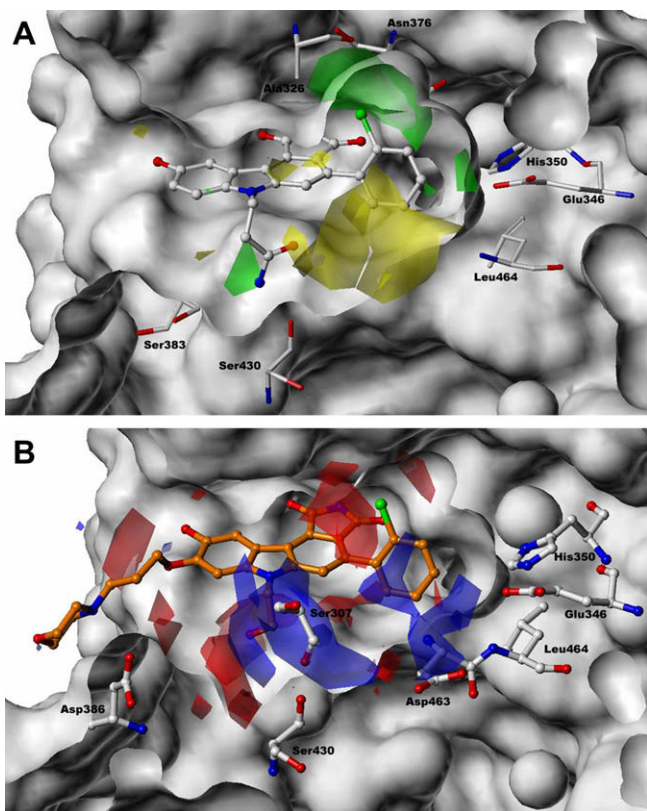
Fig. 7. (A) CoMFA steric field projected on the Wee1 binding pocket. The docked inhibitor **106** is displayed in ball and stick. The Connolly molecular surface of the enzyme is shown in white. (Color code: favoured, green (contribution level 80%); disfavoured, yellow (contribution level 20%).) (B) CoMFA electrostatic field overlaid on the docked inhibitor **169** (colored orange). The Connolly molecular surface of the enzyme is shown in white. (Color code: increase in positive charge favoured, blue (contribution level 80%); increase in negative charge favoured, red (contribution level 20%).) For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

inhibitor. The same positive effect was observed when adding the 6-F group to the phenyl ring of compound **103**.

An additional favourable region is located nearby Ser430 and Ser383. This result agrees well with the experimental data which show that adding a substituent to the nitrogen atom of the pyrrolo ring results in highly active inhibitors. The unfavourable regions are located above and behind the phenyl substituent, indicating that the conformational preference of the phenyl ring is important for high activity.

A favourable electrostatic field (Fig. 7(B), colored blue) for positively charged groups is found close to the positions 4 and 5 of the phenyl ring, indicating that more positively charged groups can favourably interact with the surrounding residues Glu346 and Asp463. Glu346 and Asp463 mainly contribute to the negative electrostatic potential in that region of the binding pocket. This observation is in agreement with the experimental data; compounds **44**, **47**, **54**, **58**, **65**, **66**, **70**, **75** and **80** bearing a substituent at position 4 or 5 of the phenyl with partial positive charge show higher pIC$_{50}$ values compared to compounds **3**, **6**, **8**, **10**–**15**, **42**, **43**, **59**, **60**, **64** and **75** which possess neutral or more negatively charged substituents. Based on the contour plots, we suggest that inhibitors with 2-Cl, 4-OH, 6-FPhe or 2-F, 4-OH, 6-FPhe ring attached to the pyrrolocarbazole core represent interesting compounds for future synthesis.

A second favourable electrostatic field for positively charged groups can be observed nearby Ser307 and Ser430. The docking of the potent inhibitor **106** showed that the amide group in the

sidechain of **106** is donating a hydrogen bond to Ser430 (Fig. 5(A)). In addition, the highly active inhibitors **147**, **148**, **149**, **150** and **152**, show also a hydrogen bond with Ser307. These results imply that the side chain of the phenylpyrrolocarbazole core plays a key role in the interaction with Wee1. Inhibitors **122**, **130**–**132**, **136**, **137** and **138** which contain partially positive charged groups in this region show high pIC$_{50}$ values, whereas on the other side compounds **117**, **118**, **124**–**126**, **127** and **133** bearing a neutral substituent at this position show lower pIC$_{50}$ values. From the GOLD docking and the established CoMFA model, we suggest that partial positively charged groups at this position are favourable to establish strong interactions with Ser307 and Ser430.

A favourable electrostatic field for electronegative substituents (Fig. 7(B), colored red) is found nearby Tyr378 and Cys379. This is in agreement with the observation that the 9-OH substituent is important for the interaction with these two amino acid residues. Another favourable electrostatic field is observed above and below the planar pyrrolocarbazole ring system. These two fields result from the slightly different position of the aromatic system as derived from the docking study.

4. Conclusion

The study was initially undertaken to explore the structural features needed for high inhibitory activity at the Wee1 kinase. The availability of crystal structures of Wee1 complexed with an inhibitor enabled us to establish and analyze a model of the binding site. The good match of predicted and experimental structures gave confidence that the docking method is able to provide relevant information about the inhibitor interaction. The docking study revealed that the most potent inhibitors (**103**, **106** and **146**) establish hydrogen bonds to the backbone atoms of the hinge region (Asn376, Glu377, Cys379). In addition the sidechain group of the most potent inhibitors interacts with Ser307 or Ser430. Inhibitors bearing a substituted tetrazole or triazole ring in the sidechain are also able to form a hydrogen bond to Ser307, resulting in good inhibitory activity.

Using a combination of receptor-based alignment and 3D-QSAR yielded a significant and predictive model, indicated by the high cross-correlation coefficient and the low SDEP value. Besides the docking-based alignment, two other alignment methods were tested which resulted in models with lower predictive quality. It was shown that the docking scores could not be used to establish a reliable QSAR model, whereas the receptor-based 3D-QSAR model gave a significant correlation and can be used to point out which interaction sites in the binding pocket might be responsible for the variance in biological activities. In this context, it must be considered that a PLS analysis indicates only where a variation in the interaction fields is correlated with a variation in the biological activities. If all molecules of a data set would show a certain important interaction with the receptor, indicated by similar interaction energy at a particular grid point for all compounds, this would not be reflected by the resulting PLS model. Thus the degree of correspondence depends strongly on the structural diversity of the training set. If one considers these circumstances, useful information can be obtained from a comparison of the contour maps and the binding site, which can then be integrated in the further optimization process.

References

[1] J. Pines, C.L. Rieder, Nat. Cell Biol. 3 (2001) E3–E6.

[2] D.O. Morgan, Nature 374 (1995) 131–134.

[3] D. Desai, H.C. Wessling, R.P. Fisher, D.O. Morgan, Mol. Cell Biol. 15 (1995) 345–350.

[4] N. Watanabe, M. Broome, T. Hunter, EMBO J. 14 (1995) 1878–1891.

[5] F. Liu, C. Rothenblum-Oviatt, C.E. Ryan, H. Piwnica-Worms, Mol. Cell Biol. 19 (1999) 5113–5123.

[6] Y. Wang, J. Li, R.N. Booher, A. Kraker, T. Lawrence, W.R. Leopold, Y. Sun, Cancer Res. 61 (2001) 8211–8217.

[7] I. Vastrik, P. D'Eustachio, E. Schmidt, G. Joshi-Tope, G. Gopinath, D. Croft, B. de Bono, M. Gillespie, B. Jassal, S. Lewis, L. Matthews, G. Wu, E. Birney, L. Stein, Genome Biol. 8 (2007) R39.

[8] G. Manning, D.B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, Science 298 (2002) 1912–1934.

[9] C.J. Squire, J.M. Dickson, I. Ivanovic, E.N. Baker, Structure 13 (2005) 541–550.

[10] B.D. Palmer, A.M. Thompson, R.J. Booth, E.M. Dobrusin, A.J. Kraker, H.H. Lee, E.A. Lunney, L.H. Mitchell, D.F. Ortwine, J.B. Smaill, L.M. Swan, W.A. Denny, J. Med. Chem. 49 (2006) 4896–4911.

[11] J.B. Smaill, E.N. Baker, R.J. Booth, A.J. Bridges, J.M. Dickson, E.M. Dobrusin, I. Ivanovic, A.J. Kraker, H.H. Lee, E.A. Lunney, D.F. Ortwine, B.D. Palmer, J. III Quin, C.J. Squire, A.M. Thompson, W.A. Denny, Eur. J. Med. Chem. 43 (2008) 1276–1296.

[12] J.B. Smaill, H.H. Lee, B.D. Palmer, A.M. Thompson, C.J. Squire, E.N. Baker, R.J. Booth, A. Kraker, K. Hook, W.A. Denny, Bioorg. Med. Chem. Lett. 18 (2008) 929–933.

[13] W. Sippl, Methods and principles in medicinal chemistry, T. Langer, R. Hofmann (Eds.), Pharmacophores and Pharmacophore Concepts, VCH Publisher, New York, 2006, pp. 223–249.

[14] G.H. Lushington, J.X. Guo, J.L. Wang, Curr. Med. Chem. 14 (2007) 1863–1877.

[15] R. Frederick, W.A. Denny, J. Chem. Inf. Model. 48 (2008) 629–638.

[16] R.D. III Cramer, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.

[17] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.

[18] Sybyl 7.2, Tripos Inc., South Hanley, St. Louis, MO, 2007.

[19] T.A. Halgren, J. Comput. Chem. 17 (1996) 490–519;
(b) T.A. Halgren, J. Comput. Chem. (1996) 520–552;
(c) T.A. Halgren, J. Comput. Chem. 17 (1996) 553–586;
(d) T.A. Halgen, R.B. Nachbar, J. Comput. Chem. 17 (1996) 587–615;
(e) T.A. Halgren, J. Comput. Chem. 17 (1996) 616–641.

[20] C.G. Broyden, J. Inst. Math. Appl. 6 (1970) 76–90.

[21] R. Fletcher, Comput. J. 13 (1970) 317–322.

[22] D. Goldfarb, Math. Comput. 24 (1970) 23–26.

[23] D. Shanno, Math. Comput. 24 (1970) 647–656.

[24] MOE 2006.08, Chemical Computing Group, Inc, Montreal, Canada, 2006.

[25] G. Jones, P. Willett, R.C. Glen, J. Mol. Biol. 245 (1995) 43–53.

[26] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727–748.

[27] J.W. Ponder, D.A. Case, Adv. Protein Chem. 66 (2003) 27–85.

[28] J. Gasteiger, M. Marsili, Tetrahedron 36 (1980) 3219–3228.

[29] Y. Pan, N. Huang, S. Cho, A.D. MacKerell Jr., J. Chem. Inf. Comput. Sci. 43 (2003) 267–272.

[30] G. Carta, A.J. Knox, D.G. Lloyd, J. Chem. Inf. Model. 47 (2007) 1564–1571.