Original article

Docking, 3D-QSAR studies and in silico ADME prediction on c-Src tyrosine kinase inhibitors

Cristina Tintori [a], Matteo Magnani [a], Silvia Schenone [b], Maurizio Botta [a],*

[a] Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Alcide de Gasperi, 2, I-53100 Siena, Italy
[b] Dipartimento di Scienze Farmaceutiche, Università degli Studi di Genova, Viale Benedetto XV, I-16132, Genova, Italy

ARTICLE INFO

ABSTRACT

Docking simulations and three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis were performed on a wide set of c-Src inhibitors. The study was conducted using a structure-based alignment and by applying the GRID/GOLPE approach. The present 3D-QSAR investigation proved to be of good statistical value, displaying r^2, q^2 and cross-validation SDEP values of 0.94, 0.84 and 0.42, respectively. Moreover, such a model also proved to be capable of predicting the activities of an external test set of compounds. The availability of the 3D structure of the target made possible the interpretation of steric and electrostatic maps within the binding site environment and provided useful insight into the structural requirements for inhibitory activity against c-Src. Two regions whose occupation by hydrophobic portions of ligands would favourably affect the activity were clearly identified. Moreover, hydrogen bond interactions involving residues Met343, Asp406 and Ser347 emerged as playing a key role in determining the affinity of the active inhibitors toward c-Src. Furthermore, the inhibitors bearing a basic nitrogen provided enhanced potency through protonation and salt bridge formation with Asp350. A preliminary pharmacokinetic profile of the molecules under analysis was also drawn on the basis of Volsurf predictions.

© 2008 Elsevier Masson SAS. All rights reserved.

1. Introduction

Protein tyrosine kinases (PTKs) are enzymes that catalyze the transfer of the terminal ATP phosphate to specific tyrosine residues present on a target substrate [1], thus regulating growth, differentiation, migration and apoptosis in mammalian cells.

PTKs are conventionally partitioned into two families: the transmembrane receptor family (receptor tyrosine kinases, RTKs), activated by the binding of ligands to the receptor extracellular domain [2], and the family of cytoplasmic and nuclear non-receptor tyrosine kinases (TKs), which can be activated through various mechanisms [3].

The Src (acronym of sarcoma) family is the oldest and most studied class of non-receptor TKs; in humans, it consists of 11 members (Blk, Brk, Fgr, Frk, Fyn, Hck, Lck, Lyn, c-Src, Srm, c-Yes). Overexpression, deregulation or mutation of these enzymes has been observed and studied in many human malignancies, including colon, breast and pancreatic cancers [4–6]. Src-family has been shown to be also involved in other pathologic situations, such as osteoporosis [7,8], cardiovascular diseases [9,10] and, more

recently, in prion diseases [11] and angiogenesis [12,13]. Therefore, the Src-family constitutes an attractive target for the design of new therapeutic agents against different pathologies, mainly against cancer and bone diseases. In this regard, in recent years much attention has been paid to the research of novel inhibitors of the c-Src tyrosine kinase.

Although no c-Src kinase inhibitor has been approved for cancer therapy so far, new potent and selective agents targeting this enzyme are continuously being synthesized, as demonstrated by the high number of publications and patents in this field. Several c-Src kinase inhibitors have been identified over the years; they span a variety of structural classes, including dihydropyrimido-quinolinones [14], 4-anilinoquinazolines [15], pyrazolo-pyrimidines [16,17], pyrrolo-pyrimidines [18–21], benzotriazines [22] and others [23,24]. In this context, a series of variously substituted pyrazolo[3,4-d]pyrimidines, possessing an interesting biological profile in terms of both c-Src activity and cellular proliferation inhibition, have been recently synthesized by our group (Chart 1) [17]. Such compounds were shown to block the growth of cancer cells by interfering with the phosphorylation of Src, and act as proapoptotic agents through the inhibition of the anti-apoptotic gene BCL2. Moreover, preliminary molecular docking simulations have been performed to get a better understanding of the peculiar chemical features that influenced their activity. Simulations led to

* Corresponding author. Tel.: +39 0577 234306; fax: +39 0577 234333.
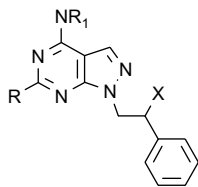 E-mail address: botta@unisi.it (M. Botta).

Chart 1. General formula of in house pyrazolo-pyrimidines.

the identification of two different binding modes, depending on the presence or on the absence of an alkylthio substituent at position 6 of the pyrazolo-pyrimidine nucleus.

In recent years, quantitative structure–activity relationship (QSAR) methods as well as 3D-QSAR methods, such as comparative molecular field analysis (CoMFA) [25] and the GRID/GOLPE combination [26,27], have been widely and successfully applied to guide the design of new PTK inhibitors [28–32]. Nevertheless, the application of 3D-QSAR methodology for the design of c-Src inhibitors has received little attention and, to the best of our knowledge, only few examples have been reported so far in this field [28,33].

In the present work, we report the development of a predictive 3D-QSAR model for a large set of pyrrolo-pyrimidines collected from literature. The model demonstrated to be capable of well predicting/estimating the affinity of the c-Src inhibitors constituting our dataset, thus allowing a better understanding of their quantitative structure–activity relationships and suggesting structural modifications for the design of new derivatives with hopefully improved c-Src inhibitory activity. The newly generated model was also applied to our pyrazolo-pyrimidine inhibitors, with the aim of identifying relevant interactions that are still not exploited by the ligands, and that could therefore guide the rational design of novel and more efficacious derivatives.

As a final step, a preliminary pharmacokinetic profile of all the molecules taken into account was traced by projecting their structures on precalculated ADME models implemented in the Volsurf program [34], and the results obtained for the literature compounds and our c-Src inhibitors were compared.

2. Methods

2.1. Dataset collection

A set of 80 variously functionalized pyrrolo-pyrimidine c-Src inhibitors were collected from the literature along with their activity data [18–21] (Table 1). It should be emphasized that particular attention was paid during the dataset collection process to ensure the homogeneity of biological data. Accordingly, the model was built using only the compounds for which the affinity values toward c-Src kinase were measured by the same group (namely, the Novartis group) and through the same experimental protocol. This allowed us to minimize the errors that could arise from the comparison of data reported by different research groups. The IC_{50} values were converted to pIC_{50} ($-\log IC_{50}$) values, and used as dependent variables in the 3D-QSAR analysis. Next, 65 compounds, denoted by footnote "a" of Table 1, were selected as training set. Training set compounds were selected following the usual guidelines (compounds belonging to the training set should be representative of the molecular diversity of all the compounds under study and uniformly span over the whole range of activity), and were characterized by affinity values spanning more than 4 orders of magnitude. The remaining 15 compounds were used as external test.

2.2. Alignment of compounds

All the dataset molecules were aligned according to the binding models resulting from docking simulations. Each compound was docked within the ATP binding pocket of the active c-Src kinase domain using the program Autodock (v. 3.0) [35]. The recently reported X-ray crystallographic structure (2.3 Å resolution; PDB accession code: 1YOL; chain B) [36] of c-Src in complex with a pyrrolo-pyrimidine derivative was chosen as template, based on the resemblance between the co-crystallized ligand and the compounds under analysis. The X-ray structure was manipulated to prepare the input for Autodock calculations by removing both non-polar hydrogens and water molecules, while Kollman united-atom partial charges and solvent parameters were added. The binding pocket was inserted into a grid box centred on the bound ligand and enclosing residues lying within about 8 Å from the ligand itself. Subsequently, grid maps were generated for several atom probes (those characterized by the same stereoelectronic properties as the atoms constituting the compounds) using the Autogrid 3.0 package, as implemented in Autodock 3.0.
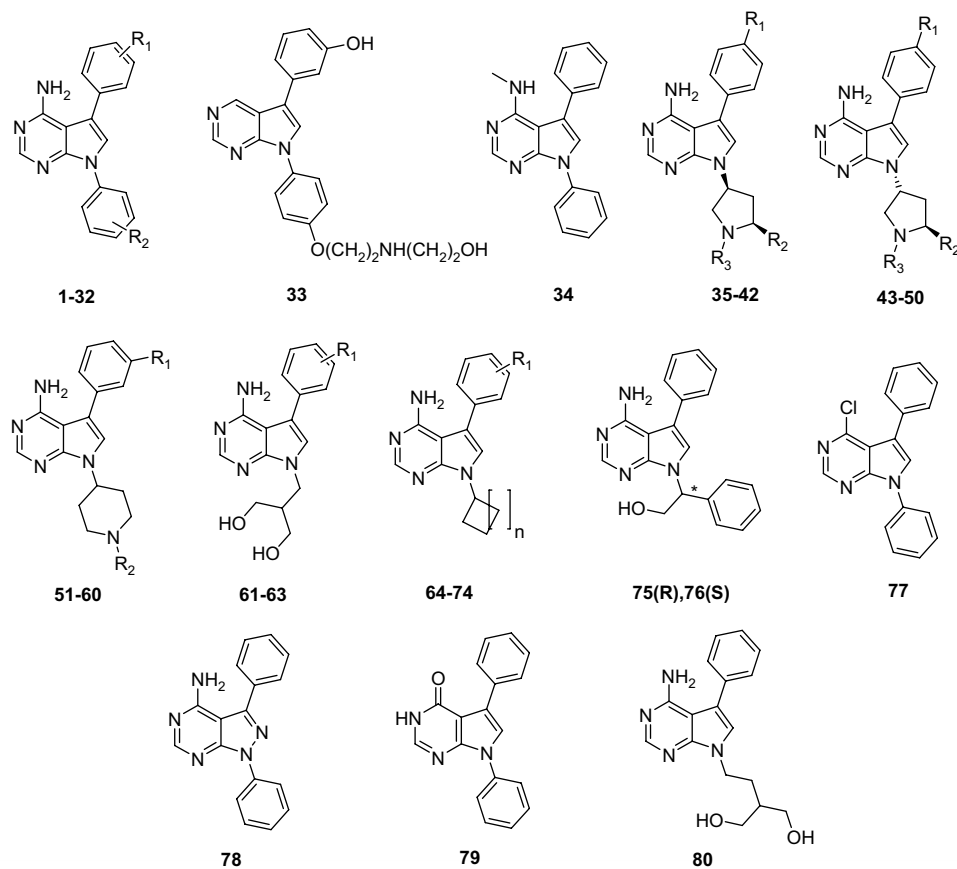
The Open Babel 1.100.2 software [37] was used to add the hydrogens and establish the protonation state at the physiological pH for all the inhibitors, before adding the Gasteiger atomic charges. Finally, the rigid root and the rotatable bonds for each compound were defined using the autotors module of the Auto-dock package. The Lamarckian genetic algorithm (LGA) was employed to explore the possible orientations/conformations of c-Src inhibitors in the binding site. For each of the 100 independent runs, a maximum number of 1 500 000 operations were performed with a population size of 200 individuals. Finally, the best ranked conformation of each molecule was selected for subsequent 3D-QSAR studies. The docking-based alignment protocol described above was also applied to a representative set of our pyrazolo-pyrimidine compounds (reported in Table 2), extracted from our in house collection.

2.3. Grid calculations

The aligned compounds (Fig. 1) were processed through the GRID software [38]. The interaction energies with the C3, O and N1 probes were calculated for each molecule of the dataset, using a grid sufficiently large to accommodate all the aligned ligands in all directions (along X, Y, and Z axes), and a grid spacing of 1 Å. The C3 probe, corresponding to a methyl group, was used to account for hydrophobic contacts and steric hindrance of compounds. The O probe (an sp^2 carbonyl oxygen) and the N1 probe (a neutral flat NH_2, e.g. amide) allowed to evaluate the hydrogen bond donor and hydrogen bond acceptor capabilities, respectively.

2.4. Variable selection and 3D-QSAR model generation

The interaction energy values at each grid point (independent variables) and pIC_{50} values (dependent variables) were subsequently imported into the GOLPE program [39]. It is well known that most of the variables (54 375 in our case) derived from the GRID analysis do not contribute to the correlation between chemical structure and biological activity, but rather they could be considered as noise, thus decreasing the quality of the model. The GOLPE program allowed the removal of non-significant variables before the generation of QSAR models. In detail, a common preliminary pretreatment (consisting in the removal of variables with only two values or with absolute values lower than 0.02 kcal/mol) was first applied to the original variables. Subsequently, the D optimal preselection procedure was used to retain only the most informative variables that correlated with the biological activity. This procedure was repeated two times, using 50% as the reduction

Table 1
Structures, experimental and estimated/predicted affinity of compounds **1**–**80** toward c-Src

Compound	R1	R2	R3	n	−log IC$_{50}$	
					Exp.	Pre.
1[a]	H	H			7.0	6.6
2[a]	m-OCH$_3$	p-CH$_2$CH$_2$OH			7.0	7.0
3[a]	m-OCH$_3$	p-CH$_2$CH$_2$-(4-OH-piperidin-1-yl)			7.7	7.8
4[a]	m-OCH$_3$	H			7.3	7.0
5[a]	m-Cl	H			6.7	6.9
6[a]	p-OCH$_3$	H			6.8	6.6
7[a]	o-CH$_3$	H			5.2	5.7
8[a]	p-COOCH$_2$CH$_3$	H			5.0	5.1
9[a]	H	o-CH$_2$OH			6.5	6.3
10[a]	H	m-CH$_2$OH			7.2	6.7
11[a]	H	p-CH$_2$OH			7.7	7.4
12[a]	m-OCH$_3$	m-CH$_2$OH			7.2	7.3
13[a]	H	o-CH$_2$NHCH$_2$CH$_2$OH			5.3	5.1
14[a]	H	m-CH$_2$NHCH$_2$CH$_2$OH			8.0	7.5
15[a]	H	p-CH$_2$NHCH$_2$CH$_2$OH			7.5	7.4
16[a]	m-OCH$_3$	p-CH$_2$NHCH$_2$CH$_2$OH			7.5	7.4
17[a]	m-OCH$_3$	p-CH$_2$CONHCH$_2$CH$_2$OH			7.7	8.1
18[a]	m-OCH$_3$	p-CH$_2$CH$_2$N(CH$_3$)CH$_2$CH$_2$OH			7.5	7.5
19[a]	m-OCH$_3$	p-CH$_2$COOH			7.3	7.2
20[a]	m-OCH$_3$	p-CH$_2$CH$_2$N(CH$_3$)$_2$			7.4	7.8
21[a]	m-OCH$_3$	p-CH$_2$CH$_2$NHCH$_2$CH$_2$OH			7.7	7.7
22[a]	m-OCH$_3$	p-CH$_2$CH$_2$N(CH$_3$)CH$_2$CH$_2$OCH$_3$			8.0	8.1
23[a]	m-OH	p-OCH$_2$CH$_2$NHCH$_2$CH$_2$OH			9.5	9.6
24[a]	p-OH	p-OCH$_2$CH$_2$NHCH$_2$CH$_2$OH			8.5	8.3
25[a]	m-OH	o-CH$_2$CH$_2$N(CH$_3$)$_2$			9.5	9.4
26	m-OCH$_3$	p-OCH$_2$CH$_2$NHCH$_2$CH$_2$OH			7.7	8.0
27	m-F	p-OCH$_2$CH$_2$NHCH$_2$CH$_2$OH			7.7	8.1
28	m-Cl	p-OCH$_2$CH$_2$NHCH$_2$CH$_2$OH			7.7	8.1
29	m-OH	m-OCH$_2$CH$_2$NHCH$_2$CH$_2$OCH$_3$			8.1	8.0
30	m-OCH$_3$	m-OCH$_2$CH$_2$NHCH$_2$CH$_2$OCH$_3$			6.8	7.3
31	m-OCH$_3$	p-OCH$_2$CH$_2$NHCH$_2$CONHCH$_3$			7.4	7.7
32	m-OCH$_3$	p-OCH$_2$CH$_2$-imidazol-1-yl			6.9	6.4
33					7.4	7.9
34					5.0	5.5
35[a]	H	CH$_2$CO$_2$CH$_2$CH$_3$	CO$_2$t-Bu		5.3	5.1
36[a]	OH	CH$_2$CO$_2$CH$_2$CH$_3$	CO$_2$t-Bu		6.5	6.3

Table 1 (*continued*)

| | | | | | $-\log IC_{50}$ | |
| | | | | | Exp. | Pre. |
Compound	R1	R2	R3	n		
37[a]	H	$CH_2CO_2CH_2CH_3$	H		7.8	8.0
38[a]	OH	$CH_2CO_2CH_2CH_3$	H		8.3	8.2
39[a]	H	CH_2OH	CO_2t-Bu		5.7	5.7
40[a]	OH	CH_2OH	CO_2t-Bu		6.5	6.5
41[a]	H	CH_2OH	H		6.9	7.00
42[a]	OH	CH_2OH	H		7.3	7.6
43[a]	H	$CH_2CO_2CH_2CH_3$	CO_2t-Bu		6.4	6.4
44[a]	OH	$CH_2CO_2CH_2CH_3$	CO_2t-Bu		7.5	7.4
45	OH	$CH_2CO_2CH_2CH_3$	H		7.0	7.2
46[a]	H	$CH_2CO_2CH_2CH_3$	H		6.2	6.5
47[a]	H	CH_2OH	CO_2t-Bu		6.9	6.9
48[a]	OH	CH_2OH	CO_2t-Bu		7.6	7.5
49[a]	H	CH_2OH	H		6.5	6.8
50[a]	OH	CH_2OH	H		7.0	7.0
51[a]	OH	CH_2CONH_2			9.0	9.0
52	OCH_3	$CH_2CH_2N(CH_3)CH_2CH_2OH$			8.2	7.5
53	OCH_3	$CH_2CH_2N(CH_3)CH_2CH_2OCH_3$			7.3	7.2
54	OH	$CH_2CO_2CH_3$			9.0	8.9
55[a]	OH	$CH_2CON(CH_3)_2$			8.5	8.5
56[a]	OH	CH_2CH_2OH			9.00	9.0
57[a]	OCH_3	$CH_2CO_2CH_3$			7.1	7.5
58[a]	OCH_3	CH_2CONH_2			7.7	7.6
59[a]	OCH_3	$CH_2CON(CH_3)_2$			7.2	7.2
60[a]	OCH_3	CH_2CH_2OH			7.6	7.8
61[a]	H				6.7	6.3
62[a]	p-OH				6.3	6.7
63[a]	m-OCH_3				6.5	6.7
64[a]	H			2	6.9	7.5
65[a]	p-OCH_3			2	7.1	7.2
66[a]	p-OH			2	7.7	7.7
67[a]	m-OCH_3			2	8.7	7.9
68[a]	m-OH			2	9.0	8.6
69[a]	H			1	7.0	7.1
70[a]	p-OCH_3			1	6.7	7.0
71[a]	p-OH			1	8.0	7.5
72[a]	m-OCH_3			1	7.4	7.5
73[a]	m-OH			1	8.4	8.4
74	H			0	6.6	6.1
75[a]					6.3	6.4
76[a]					5.3	5.2
77[a]					5.1	5.1
78[a]					6.3	6.3
79[a]					5.0	5.1
80					5.6	7.1

[a] Compounds belonging to the training set.

level and 5 components. Each selection of variables was preceded by the calculation of a new model, based on the last selected variables. Next, a smart region definition (SRD) algorithm was applied with the aim of selecting and grouping the regions of variables with the highest importance for the model. The groups were then used in the fractional factorial design (FFD) procedure to replace the original variables. FFD selection was applied two times, until no further improvement in the q^2 value was observed.

The 3D-QSAR equations were generated using the partial least-squares (PLS) statistical method [40]. PLS algorithm is a variation of principal component regression in which the original variables are replaced by a small set of linear combination thereof.

The internal predictive power of the final model was evaluated by random group cross-validation and the corresponding r^2 and q^2 values were calculated. On the other hand, the external predictive power was assessed by calculating the affinity values of test set compounds and by using the standard deviation error of prediction (SDEP) value as diagnostic statistical parameter.

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - Y_{fit})^2}$$

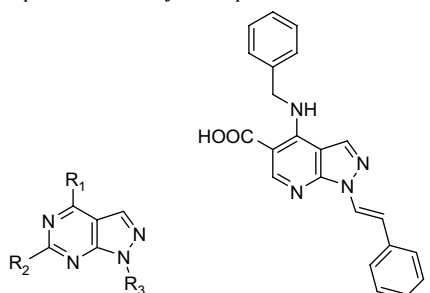$$q^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - Y_{cv})^2}$$

$$SDEP = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - Y_{cv})^2}{N}}$$

where Y_i = experimental value; Y_{fit} = recalculated value; $\overline{Y}$ = mean value; Y_{cv} = predicted value; N = number of objects.

2.5. Volsurf analysis

The pharmacokinetic profile of all the molecules under investigation was predicted by means of four precalculated ADME models provided by the Volsurf program [34]. The interaction energies between the molecules and the water (OH2), hydrophobic (DRY) and H-bonding carbonyl (O) probes were calculated using a 0.5 Å grid space. Caco-2 cell permeability, metabolism stability, blood–brain barrier (BBB) permeation and water solubility of the studied compounds were then predicted by projecting the inhibitors on the corresponding statistical Volsurf models (i.e., CACO2, Metabolic Stability, BBB and SOLY).

Table 2
Structures and experimental affinity of compounds **81–114** toward c-Src

81-113   **114**

Compound	R1	R2	R3	−log (activity)
81	NHcyclopropyl	H	$CH_2CHClC_6H_5$	5.0
82	$NHCH_2CH_2C_6H_5$	H	$CH_2CHClC_6H_5$	5.7
83	NHC_3H_7	H	$CH_2CHClC_6H_5$	5.5
84	$N(C_2H_5)_2$	H	$CH_2CHClC_6H_5$	5.3
85	NHC_4H_9	H	$CH_2CHClC_6H_5$	5.7
86	NHcyclohexyl	H	$CH_2CHClC_6H_5$	5.6
87	1-Piperidinyl	H	$CH_2CHClC_6H_5$	5.2
88	1-Hexahydroazepinyl	H	$CH_2CHClC_6H_5$	5.0
89	NHC_6H_5	H	$CH_2CHClC_6H_5$	5.5
90	NHC_3H_7	H	$CH_2CHBrC_6H_5$	6.1
91	$NHCH(CH_3)_2$	H	$CH_2CHBrC_6H_5$	5.5
92	NHC_4H_9	H	$CH_2CHBrC_6H_5$	5.5
93	$NH(CH_2)_2OC_2H_5$	H	$CH_2CHBrC_6H_5$	5.8
94	1-Pyrrolydinyl	H	$CH_2CHBrC_6H_5$	5.0
95	1-Piperidinyl	H	$CH_2CHBrC_6H_5$	5.8
96	$NHCH_2C_6H_5$	H	$CH_2CHBrC_6H_5$	5.5
97	NHC_3H_7	H	$CH_2CHOHC_6H_5$	5.0
98	$NHCH_2C_6H_5$	H	$CH_2CHOHC_6H_5$	5.6
99	$NHCH_2CH_2C_6H_5$	H	$CH_2CHOHC_6H_5$	5.6
100	NHC_2H_4-pF	H	$CH_2CHOHC_6H_5$	5.0
101	NHcyclopropyl	SCH_3	$CH_2CHClC_6H_5$	5.6
102	NHC_6H_4-mCl	SCH_3	$CH_2CHClC_6H_5$	6.1
103	NHC_3H_7	SCH_3	$CH_2CHClC_6H_5$	6.1
104	NHC_4H_9	SCH_3	$CH_2CHClC_6H_5$	5.7
105	$N(C_2H_5)_2$	SCH_3	$CH_2CHClC_6H_5$	6.3
106	1-Piperidinyl	SCH_3	$CH_2CHClC_6H_5$	5.6
107	4-Morpholinyl	SCH_3	$CH_2CHClC_6H_5$	5.2
108	$NHCH_2C_6H_5$	SCH_3	$CH_2CHClC_6H_5$	5.4
109	NHC_3H_7	SC_2H_5	$CH_2CHClC_6H_5$	6.2
110	NHC_4H_9	SC_2H_5	$CH_2CHClC_6H_5$	6.2
111	1-Piperidinyl	SC_2H_5	$CH_2CHClC_6H_5$	5.1
112	4-Morpholinyl	SC_2H_5	$CH_2CHClC_6H_5$	5.4
113	$NHCH_2C_6H_5$	SC_2H_5	$CH_2CHClC_6H_5$	5.1
114				5.4

Fig. 1. Alignment of the pyrrolo-pyrimidine compounds used for 3D-QSAR model generation.

moiety at C5 was located into the hydrophobic pocket I, in Van der Waals contact with Ala295, Lys297, Met316, Val325, Ile338, Thr340 and Ala405. Finally, the N7 substituents were accommodated within the hydrophobic pocket II, showing favourable lipophilic interactions with Gly346 and Leu275. Interestingly, all pyrrolo-pyrimidines bearing a positively charged nitrogen on N7 side chain pointed such moiety away from the binding pocket, in a solvent accessible area, and were involved in a favourable charge–charge interaction with Asp350. A different binding mode was identified for a restricted number of pyrrolo-pyrimidines (namely compounds **8**–**11**, **34**, **75**–**77**, **79**, Table 1), as exemplified in Fig. 2B. In particular, such molecules still disposed their pyrrolo-pyrimidine nucleus within the adenine region but, unlike the crystal structure bound orientation, substituents at N7 were accommodated within the hydrophobic pocket I, whereas the phenyl moiety at C5 was located into the hydrophobic pocket II. Such compounds were still engaged in three hydrogen bonds with the hinge region, though with a different network; in fact, such hydrogen bonds involved the N1 and the OH side chain of Thr340, the amino substituent and the Met343 carbonyl backbone group, and the N3 and the NH backbone of Met343.

On the other hand, as previously described [17], two different binding modes emerged from docking studies on our pyrazolo-pyrimidine inhibitors, depending on the presence or on the absence of an alkylthio substituent at position 6 of the pyrazolo-pyrimidine ring (Fig. 2C and D). However, differently from the results previously reported, not only the 6-unsubstituted compounds, but also the 6-substituted derivatives formed two hydrogen bonds, involving the amino group at C4 and the side chain of Thr340, and the N6 and the NH backbone of Met343. The reason for such variation is likely to lie in the fact that in previous calculations [17] we used the X-ray structure of c-Src kinase complexed with an Imatinib analog, completely different from our pyrazolo-pyrimidines from a structural point of view. Conversely, the results reported herein were derived from studies on a more recently deposited crystal structure, which was not available when we performed the first docking simulations. As already mentioned, in this new structure the active c-Src kinase domain was complexed with a pyrrolo-pyrimidine inhibitor and it was therefore judged a more suitable template to analyse the binding mode of the molecules constituting our dataset. It should be pointed out that the orientation of the studied molecules within the two active sites was essentially the same; only a few differences, which could be ascribed to a different ligand induced fit, were found in the side chain position of some residues; nevertheless, such variations

3. Results

3.1. Docking analysis

Docking studies were performed on all the compounds under investigation (**1**–**114** in Tables 1 and 2) and the results are summarized in Fig. 2 and Table 3. The reliability of the docking protocol was first checked by comparing the best docking pose obtained for the crystallized inhibitor with its bound conformation. As a result, a root mean square deviation (rmsd) of 0.8 Å was found suggesting that the docking procedure could be relied on to predict the binding mode of our compounds.

Most of the pyrrolo-pyrimidines were found to bind into the ATP binding site of c-Src similar to the crystallized inhibitor (Fig. 3), and showed an interaction pattern analogous to the canonical adenine binding of protein kinases, characterized by three hydrogen bonds involving the pyrrolo-pyrimidine nucleus and the hinge region (Fig. 2A): one between the N3 and the NH backbone of Met343, and two between the amino substituent and the OH side chain of Thr340 and the carbonyl backbone of Glu341. Moreover, the phenyl
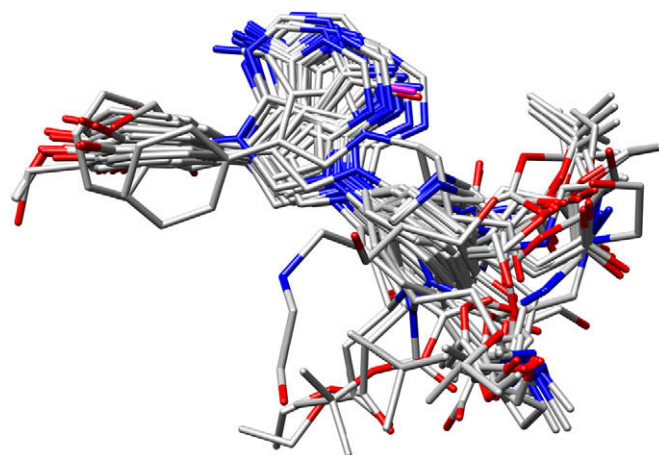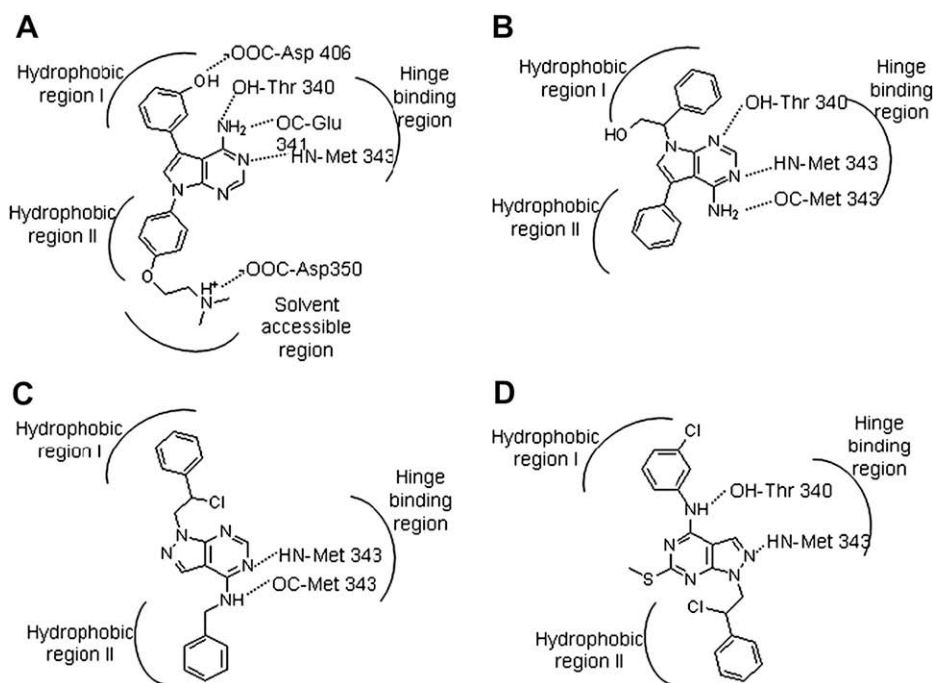
Fig. 2. Schematic representation of the predicted binding modes for the docked compounds. As an example, compounds **25**, **75**, **89** and **102** were shown in parts A, B, C and D, respectively.

were responsible for the different described pattern of hydrogen bonds.

3.2. 3D-QSAR model

The 3D-QSAR analysis was carried out on the 65 pyrrolo-pyrimidines of the training set, using the GRID/GOLPE procedure (for details, see Section 2). As a result, a final model consisting of 4 components was generated on a reduced set of 1554 variables, which were derived from the variable selection process. The model showed a good correlation coefficient ($r^2 = 0.94$) as well as

a satisfactory internal and external predictive powers. The internal validation procedure (a cross-validation routine with 5 random sets of compounds) resulted in a cross-validated correlation coefficient (q^2) of 0.84 and a cross-validation SDEP value of 0.42. Moreover, the external predictive power of the model was assessed by calculating the affinity values of the test set compounds, as reported in Table 1. The data showed a good agreement between the experimental and predicted affinity values for test set compounds (Fig. 4), as an SDEP value of 0.57 was found and all residuals fell within a statistically tolerable error range.

Table 3
Summary of the interactions between the main residues constituting the c-Src binding site and the studied inhibitors, according to the binding models derived from docking studies

| Residue	Residue location	Molecules	Kind of interaction
Thr340	Hinge region	**1–32, 34–78, 80**	Hydrogen bond
Glu341	Hinge region	**1–7, 12–32, 35–74, 78, 80**	Hydrogen bond
Met343	Hinge region	**1–83, 85–86, 89–93, 96–104, 108–110, 113, 114**	Hydrogen bond
Asp406	Hinge region	**23, 25, 29, 45, 51, 54, 68, 73, 75**	Hydrogen bond
Tyr342	Hinge region	**1–114**	Hydrophobic interaction
Tyr342	Hinge region	**75, 76**	Hydrogen bond
Met316	Hydrophobic region I	**1–114**	Hydrophobic interaction
Val325	Hydrophobic region I	**1–114**	Hydrophobic interaction
Ile338	Hydrophobic region I	**10, 50**	Hydrogen bond
Ile338	Hydrophobic region I	**1–114**	Hydrophobic interaction
Ala405	Hydrophobic region I	**1–114**	Hydrophobic interaction
Ala295	Hydrophobic region I	**10**	Hydrogen bond
Ala295	Hydrophobic region I	**1–114**	Hydrophobic interaction
Glu312	Hydrophobic region I	**11**	Hydrogen bond
Leu275	Hydrophobic region II	**2, 37, 38, 41, 42, 47, 48**	Hydrogen bond
Leu275	Hydrophobic region II	**1–114**	Hydrophobic interaction
Gly346	Hydrophobic region II	**1–114**	Hydrophobic interaction
Leu349	Hydrophobic region II	**22, 29**	Hydrophobic interaction
Asp350	Solvent accessible region	**3, 14–16, 18, 20–29, 31, 33, 52**	Salt bridge
Asp350	Solvent accessible region	**3, 12, 14–18, 20–29, 31, 33, 39–42, 45, 49–52, 56, 58, 60, 80**	Hydrogen bond
Ser347	Solvent accessible region	**8, 12, 14–18, 21, 23–28, 31, 33, 37, 38, 41–50, 52, 56, 60, 80**	Hydrogen bond
Gln277	Solvent accessible region	**19, 35, 36**	Hydrogen bond
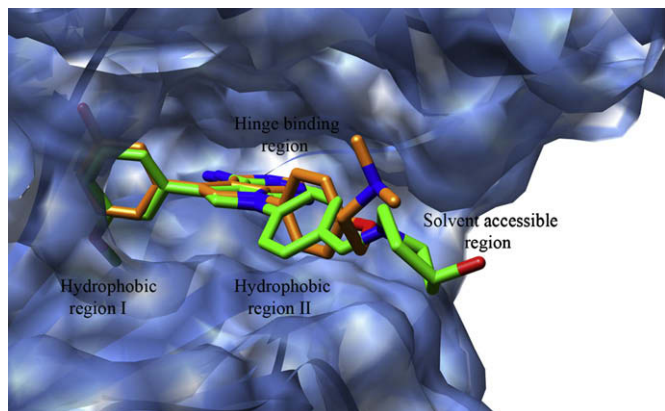
Fig. 3. Binding mode of docked compound **25** (orange) within the ATP binding site of c-Src. The crystallized ligand is represented in green.

Concerning our pyrazolo-pyrimidines (Table 2), the computation of statistical parameters such as SDEP or r^2 seemed inappropriate, since the biological data came from a different source with respect to training set and were expressed as K_i instead of IC_{50}. Notwithstanding, the analysis of our compounds within the derived PLS coefficient plots could be useful to design novel inhibitors, thus providing a rational basis for the synthesis of new derivatives suggested by the 3D maps of the model itself (actually, the synthesis of some of such derivatives is already in progress and results will be published in the due time).

3.3. 3D-QSAR model interpretation

A fundamental role of a 3D-QSAR model, besides predicting the biological activity of new untested molecules, is to allow a more exhaustive interpretation of biological data and to provide hints
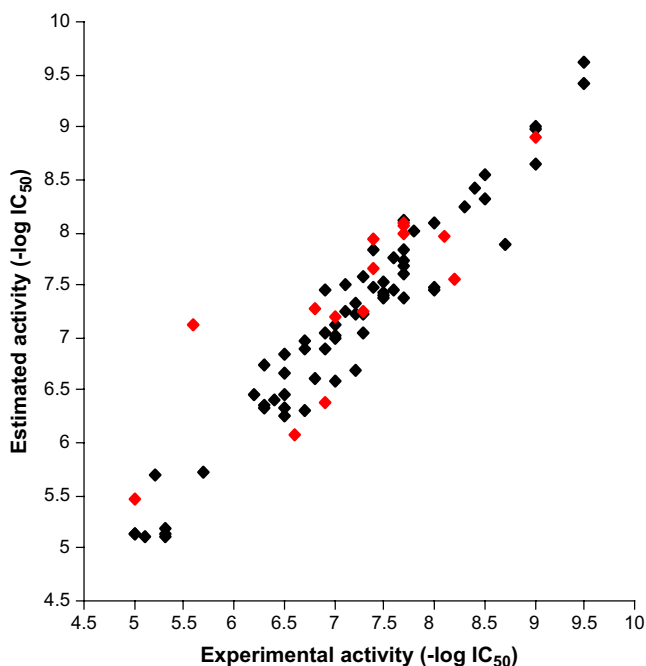
Fig. 4. Experimental activity versus predicted activity in the final 3D-QSAR model. (Black diamonds represent predictions for the training set while red diamonds represent predictions for test set.)

about the molecular regions that are directly related to the measured activity. In this context, the analysis of GOLPE PLS pseudo-coefficient plots could be very useful for SAR studies and for the design of novel molecules with improved activities. In fact, such plots highlight the regions of space in which a positive interaction of ligands with a given probe results in an increased activity, and vice versa. Moreover, since in this case the model was developed following a structure-based approach, a further contribution to the comprehension of the structure–activity relationships could be given by the analysis of the PLS pseudo-coefficient plots within the protein environment. In such a way, the key residues in determining the activities of the studied compounds could be easily detected.

Contours maps for the C3 probe indicate the regions of space that are sterically accessible and inaccessible (yellow and cyan, respectively) to alkyl (or, more generally, hydrophobic) groups of ligands. Fig. 5a shows the contour maps generated by the model for the C3 probe, superimposed to the X-ray structure of the ATP binding site of c-Src. Yellow contour maps A and B indicated two regions whose occupation by hydrophobic portions of ligands would favourably affect the activity; such regions corresponded to the hydrophobic region I, bounded by residues Val325, Ile338, Thr340, Ala405, Met316 and Leu327, and to the hydrophobic region II, including residues Gly346, Leu275 and Leu349. Interestingly, in a previous study we had already pointed out the importance of the full occupancy of these two regions for the activity of our c-Src inhibitors [17]. On the contrary, cyan volumes represented regions in which the presence of alkyl (or hydrophobic) groups would result in a decreased activity. Accordingly, the C region partially superimposed on protein residues, confirming a steric hindrance in the receptor, whereas the cyan region D was in close proximity of residues Asp350 and Ser347 and corresponded to a solvent accessible area. Consistent with these findings, the inhibitors containing a basic nitrogen in D region provided enhanced potency through protonation and interaction with Asp350 and Ser347.

Coefficient plots generated for the N1 and O probes (Fig. 5b and c, respectively) have been used to investigate the hydrogen bond donor and hydrogen bond acceptor capabilities relevant for the biological activity of compounds. Cyan polyhedra for the N1 (O) probe indicates regions in which the presence of a hydrogen bond donor (acceptor) group positively interacting with the ligands results in increased activity. An opposite situation is described by yellow polyhedra. It is worth noting that some yellow and cyan polyhedra of both N1 and O probes occupied the same regions of space of C3 polyhedra, suggesting that their contribution in such regions could be mainly steric in nature. However, two additional significant regions in both N1 and O coefficient plots were identified. Cyan volumes E and F for N1 probe (Fig. 5b) lay near residues Met343 and Asp406, respectively, revealing that such amino acids should be important in determining the affinity toward c-Src acting as hydrogen bond donors. Similarly, cyan volumes G and H in the coefficient plot corresponding to the O probe, being located near residues Glu341 and Asp350, pointed out that such residues may act as hydrogen bond acceptors, thus favouring the activity of those inhibitors which are able to locate a hydrogen donor group in their proximity.

Fig. 6a–c displays one of the most potent compounds of the dataset (compound **25** in Table 1, $pIC_{50} = 9.50$) embedded into the contour maps of C3, N1 and O probes, respectively. The *m*-hydroxy-phenyl substituent at position 4, as well as the aromatic ring of the side chain at position 5, fully occupied both the hydrophobic pockets I and II (yellow polyhedra in Fig. 6a), consistent with the key role played by these regions in influencing the activity of compounds. In addition, compound **25**
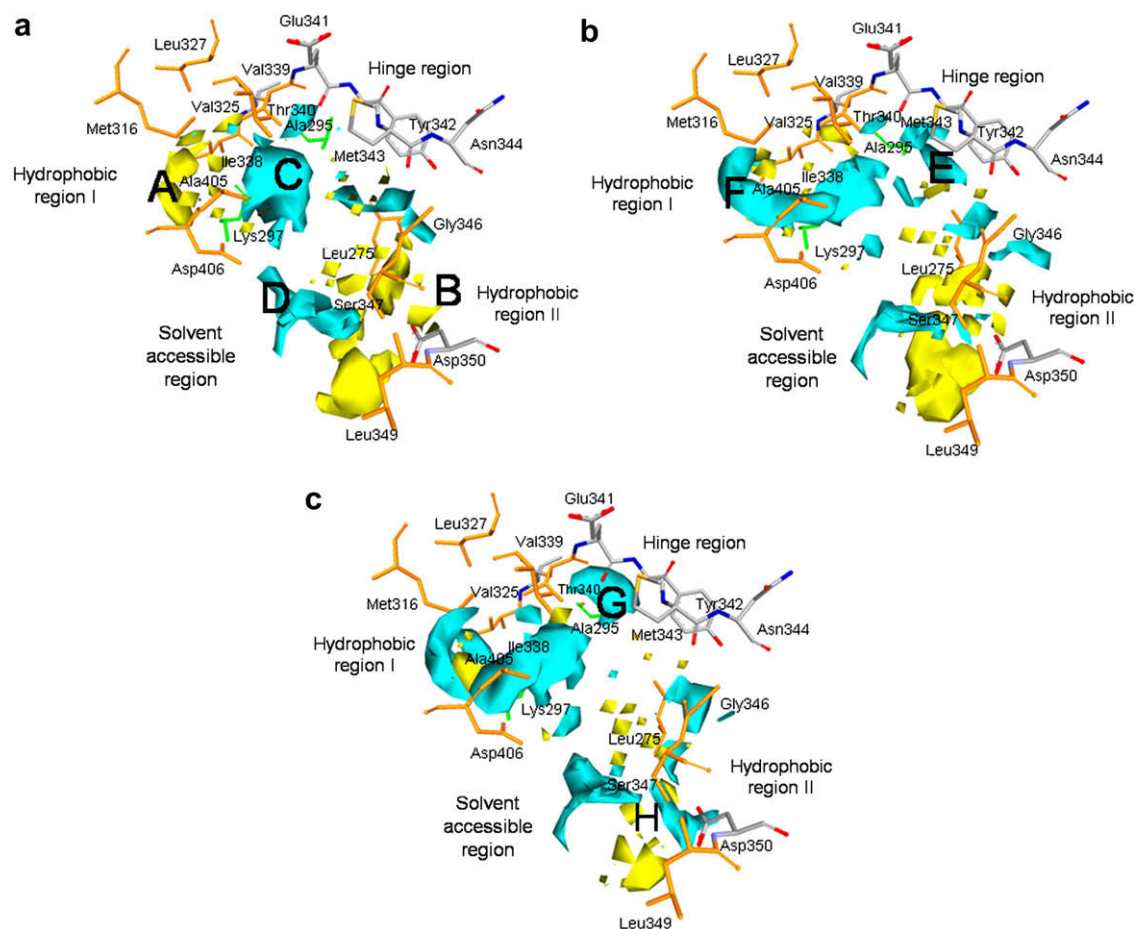
Fig. 5. PLS coefficient plots obtained for C3 (a), N1 (b) and O: (c) probes superimposed to the ATP binding site on c-Src. The residues belonging to hydrophobic region I or II are orange colored, Ala295 and Lys297 are green colored while the amino acids of the hinge region as well as the residue Asp350 are colored by atom.

showed favourable interactions within the region D, being able to locate the hydrophilic *N*-dimethylamino moiety in proximity of this solvent exposed area. As regards hydrogen bond interactions, these were analysed through the contour maps obtained with N1 and O probes. The N3 of the ligand was in hydrogen bond distance from the NH amide of Met343, consistent with the close proximity of such residue to the cyan contour map E (Fig. 6b). Both data contributed to point out the importance of this interaction in modulating the activity against c-Src. Very similarly, the *m*-hydroxy group formed a hydrogen bond with the NH backbone of Asp406; this hydrogen bond donor group was located near the cyan region F of probe N1, highlighting the relevance of hydrogen bond acceptor capabilities of ligands in this region of the binding pocket. Finally, the amino group at C4 was involved in two hydrogen bonds with the carbonyl backbone of Glu341 and with the OH side chain of Thr340; these hydrogen bond acceptor groups fell within one of the cyan regions of probe O (namely, region G in Fig. 6c), indicating that also these interactions were expected to be crucial for the activity of c-Src inhibitors.

Concerning pyrazolo-pyrimidines **81**–**114**, the most relevant aspect that emerged from the analysis of their binding modes through the above-described contour plots was the lack of favourable interactions with residues Asp406 and Asp350. As a consequence, we can speculate that the activity of such compounds should be significantly improved by structural modifications aimed at establishing strong interactions with such residues.

3.4. Prediction of ADME properties

The pharmacokinetic features of all c-Src inhibitors taken into account in this work were preliminarily investigated by projecting the compounds on predefined Volsurf models. In fact, it would be extremely advantageous if information about the ADME properties of the studied molecules could be produced in the early stages of the drug discovery process. Once obtained, this information is expected to help chemists to ameliorate the pharmacokinetic profile of the compounds. The use of in silico methods to predict ADME properties is intended as a first step in this direction, and the results of such analysis are herein reported and discussed. The main goal was to predict the ADME properties of our synthesized pyrazolo-pyrimidine compounds using some models provided by Volsurf, and compare the results with those obtained for the Novartis pyrrolo-pyrimidines. Projections of the object molecules on the Caco-2 permeability, water solubility, BBB penetration and metabolic stability models are shown in Fig. 7A–D, respectively. The projected pyrrolo-pyrimidine and pyrazolo-pyrimidine are colored in green and yellow, respectively, whereas blue, cyan and red points represent the compounds used to generate the model. The plots indicated for our compounds an efficacious permeation of intestinal epithelium and a good crossing of the blood–brain barrier together with low metabolic stability and aqueous solubility. As a whole, the results of the in silico study pointed out a non-optimal pharmacokinetic profile, since the low solubility suggested that the per os administration of these molecules could not be effective. In this regard, work is in progress to ameliorate the solubility of
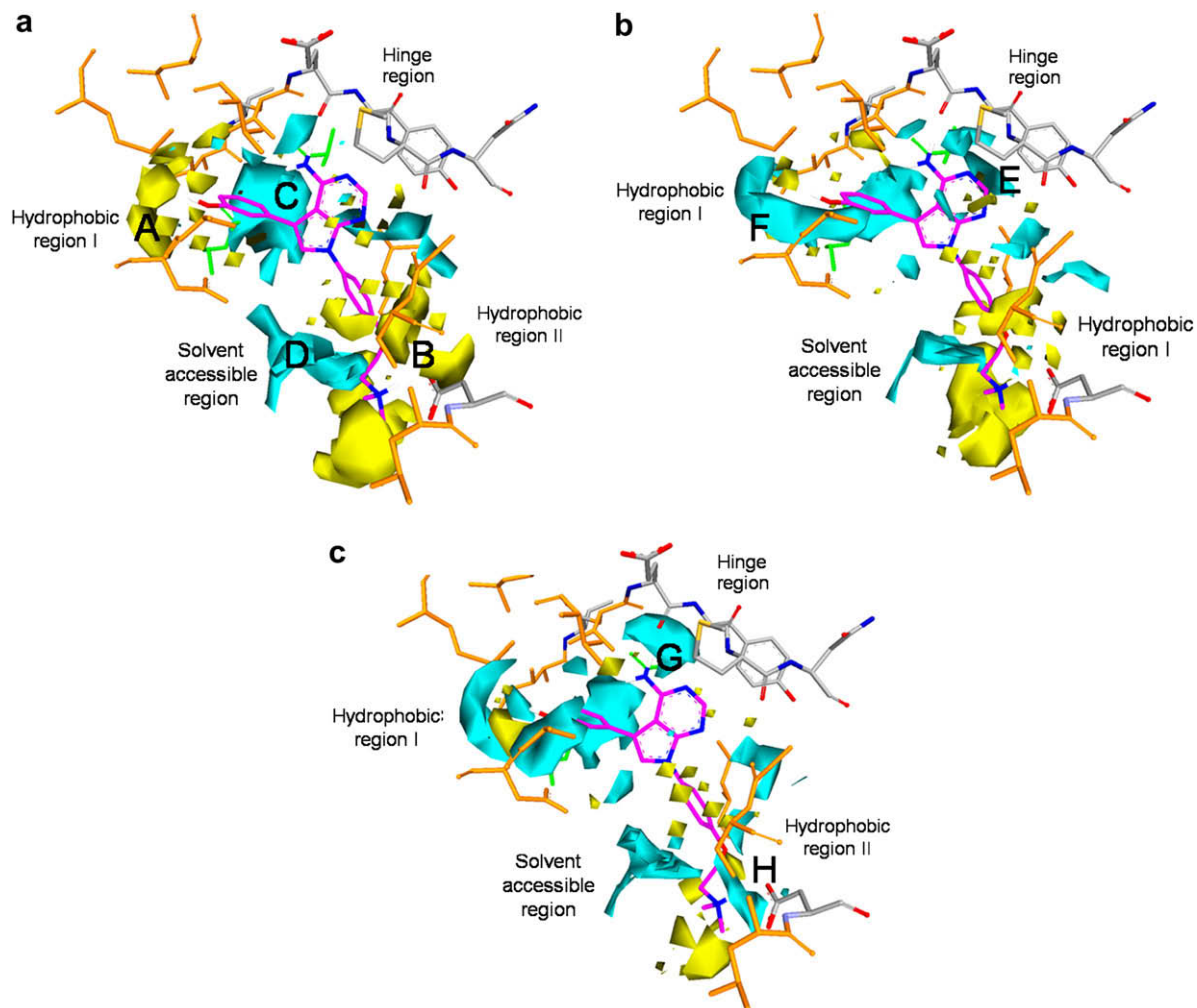
Fig. 6. Compound **25** embedded into the PLS coefficient plots obtained for C3 (a), N1 (b) and O: (c) probes. The residues belonging to hydrophobic region I or II are orange colored, Ala295 and Lys297 are green colored while the amino acids of hinge region as well as the residue Asp350 are colored by atom.

pyrazolo-pyrimidines entrapping them in liposomes [41]. On the other hand, the high penetration through the BBB led us to speculate about the possibility to use such compounds as potential therapeutic agents for the treatment of acute ischemic stroke, as already reported for similar molecules with a pyrazolo-pyrimidine nucleus [42].

As shown in Fig. 7A–D, the pyrrolo-pyrimidines were predicted to have good Caco-2 cell permeability as well, but they seemed to be characterized by a better water solubility and metabolic stability compared to our compounds, while being endowed with a lower permeability across the BBB.

4. Conclusion

In this paper we reported the first 3D-QSAR study on a wide set of pyrrolo-pyrimidines acting as c-Src kinase inhibitors. After using the crystallographic structure of c-Src as template for a structure-based alignment, a robust 3D-QSAR model was generated by applying the GRID/GOLPE approach. An internal validation based on the cross-validation test, together with the good correlation between experimental and predicted activity values for an external test set of compounds, proved the predictive power of the model. Remarkably, insertion of contour maps provided by the model within the protein binding site made the interpretation of results more straightforward and allowed the identification of not only the

chemical features responsible for the inhibitory activity of compounds toward the c-Src kinase, but also of the most important c-Src residues involved in the binding process. Analysis of results led to the identification of both hydrophobic and hydrogen bond interactions that should be established by c-Src inhibitors to improve their affinity to their target. Two regions whose occupation by hydrophobic portions of ligands would favourably affect the activity were clearly identified. Moreover, hydrogen bond interactions involving residues Met343, Asp406 and Ser347 emerged as playing a key role in determining the affinity of the active inhibitors toward c-Src. Furthermore, the inhibitors bearing a basic nitrogen provided enhanced potency through protonation and salt bridge formation with Asp350.

Remarkably, the 3D-QSAR model was used to derive indications about structural modifications for a set of in house c-Src inhibitors bearing the pyrazolo-pyrimidine nucleus.

The pharmacokinetic profile of both pyrrolo-pyrimidine and pyrazolo-pyrimidine compounds was preliminarily investigated by means of the Volsurf approach, allowing for some important considerations and suggestions to direct the development of our inhibitors.

Taken together, the 3D-QSAR model and the computational analysis of the pharmacokinetic properties described herein will constitute a valuable tool for the design of novel structurally related pyrazolo-pyrimidine inhibitors endowed with increased affinities toward c-Src and improved ADME profiles.
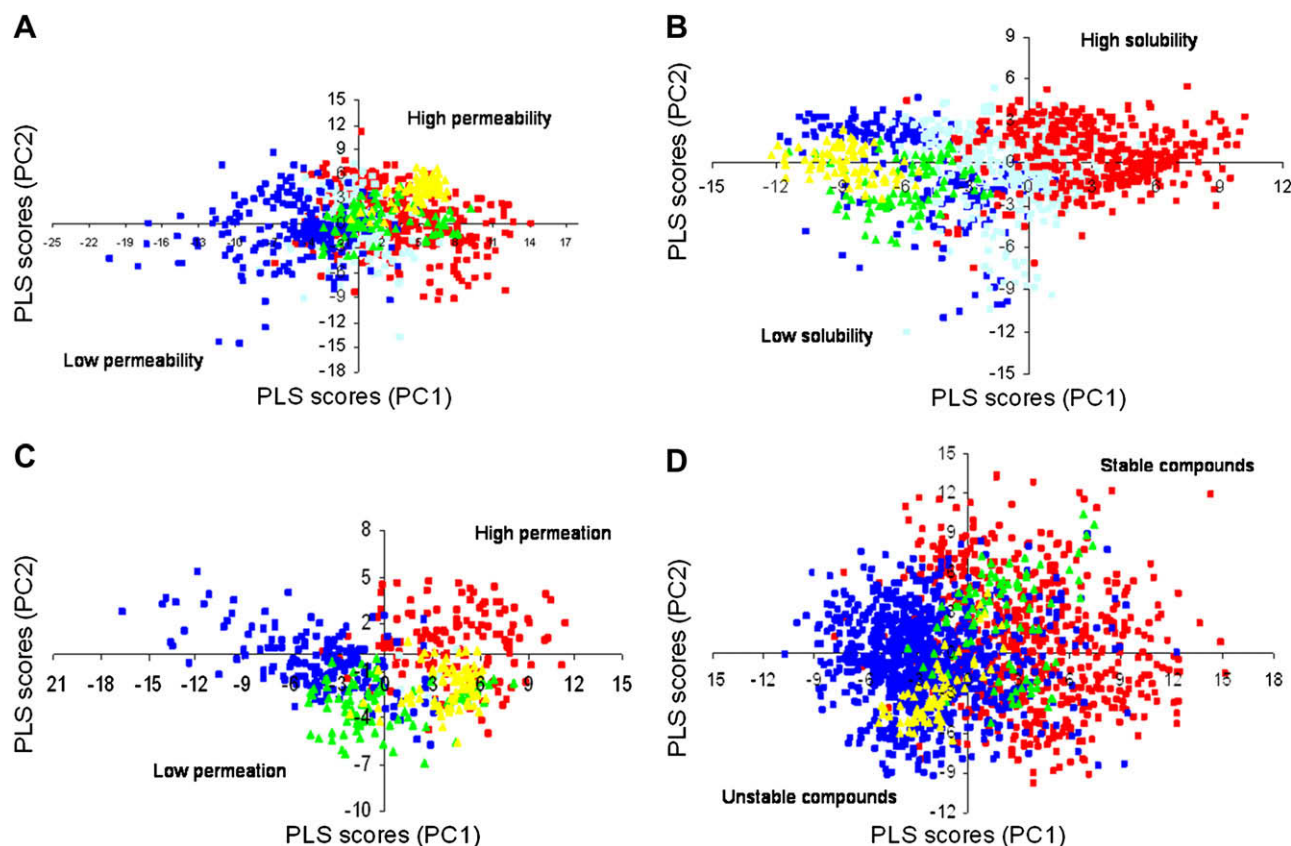
Fig. 7. Projection of the studied compounds (compounds **1–80**: green, compounds **81–114**: yellow), on the precalculated Volsurf models: Caco-2 (A); solubility (B); blood–brain barrier (C); metabolic stability (D).

References

[1] A. Ullrich, J. Schlessinger, Cell 61 (1990) 203–212.
[2] L.N. Johnson, M.E. Noble, D.J. Owen, Cell 85 (1996) 149–158.
[3] K. Neet, T. Hunter, Genes Cells 1 (1996) 147–169.
[4] H. Aligayer, D.D. Boyd, M.M. Heiss, E.K. Abdalla, S.A. Curley, G.E. Gallick, Cancer 94 (2002) 344–351.
[5] A.P. Belsches-Jablonski, J.S. Biscardi, D.R. Peavy, D.A. Tice, D.A. Romney, S.J. Parsons, Oncogene 20 (2001) 1465–1475.
[6] L.A. MacMillan-Crow, J.S. Greendorfer, S.M. Vickers, J.A. Thompson, Arch. Biochem. Biophys. 377 (2000) 350–356.
[7] B.F. Boyce, T. Yoneda, C. Lowe, P. Soriano, G.R. Mundy, J. Clin. Invest. 90 (1992) 1622–1627.
[8] H. Umemori, S. Sato, T. Yagi, S. Aizawa, T. Yamamoto, Nature 367 (1994) 572–576.
[9] S. Weis, S. Shintani, A. Weber, R. Kirchmair, M. Wood, A. Cravens, H. McSharry, A. Iwakura, Y.S. Yoon, N. Himes, D. Burstein, J. Doukas, R. Soll, D. Losordo, D. Cheresh, J. Clin. Invest. 113 (2004) 885–894.
[10] R. Paul, Z.G. Zhang, B.P. Eliceiri, Q. Jiang, A.D. Boccia, R.L. Zhang, M. Chopp, D.A. Cheresh, Nat. Med. 7 (2001) 222–227.
[11] R.R. Nixon, J. Biol. Chem. 280 (2005) 2455–2462.
[12] S. Schenone, F. Manetti, M. Botta, Curr. Pharm. Des. 13 (2007) 2118–2128.
[13] S. Donnini, M. Monti, C. Castagnini, R. Solito, M. Botta, S. Schenone, A. Giochetti, M. Ziche, Int. J. Cancer 120 (2007) 995–1004.
[14] R.L. Dow, B.M. Bechle, T.T. Chou, C. Goddard, E.R. Larson, Bioorg. Med. Chem. Lett. 5 (1995) 1007–1010.
[15] M.R. Myers, N.N. Setzer, A.P. Spada, A.L. Zulli, C.-Y.J. Hsu, A. Zilberstein, S.E. Johnson, L.E. Hook, M.V. Jacoski, Bioorg. Med. Chem. Lett. 7 (1997) 417–420.

[16] J.H. Hanke, J.P. Gardner, R.L. Dow, P.S. Changelian, W.H. Brissette, E.J. Weringer, B.A. Pollok, P.A. Connelly, J. Biol. Chem. 271 (1996) 695–701.
[17] F. Carraro, A. Naldini, A. Pucci, G.A. Locatelli, G. Maga, S. Schenone, O. Bruno, A. Ranise, F. Bondavalli, C. Brullo, P. Fossa, G. Menozzi, L. Mosti, M. Modugno, C. Tintori, F. Manetti, M. Botta, J. Med. Chem. 49 (2006) 1549–1561.
[18] M. Missbach, E. Altmann, L. Widler, M. Susa, E. Buchdunger, H. Mett, T. Meyer, J. Green, Bioorg. Med. Chem. Lett. 10 (2000) 945–949.
[19] L. Widler, J. Green, M. Missbach, M. Susa, E. Altmann, Bioorg. Med. Chem. Lett. 11 (2001) 849–852.
[20] E. Altmann, M. Missbach, J. Green, M. Susa, H.-A. Wagenknecht, L. Widler, Bioorg. Med. Chem. Lett. 11 (2001) 853–856.
[21] E. Altmann, L. Widler, M. Missbach, Mini Rev. Med. Chem. 2 (2002) 201–208.
[22] G. Noronha, K. Barrett, A. Boccia, T. Brodhag, J. Cao, C.P. Chow, E. Dneprovskaia, J. Doukas, R. Fine, X. Gong, C. Gritzen, H. Gu, E. Hanna, J.D. Hood, S. Hu, X. Kang, J. Key, B. Klebansky, A. Kousba, G. Li, D. Lohse, C.C. Mak, A. McPherson, M.S. Palanki, V.P. Pathak, J. Renick, F. Shi, R. Soll, U. Splittgerber, S. Stoughton, S. Tang, S. Yee, B. Zeng, N. Zhao, H. Zhu, Bioorg. Med. Chem. Lett. 17 (2007) 602–608.
[23] S.R. Klutchko, J.M. Hamby, D.H. Boschelli, Z. Wu, A.J. Kraker, A.M. Amar, B.G. Hartl, C. Shen, W.D. Klohs, R.W. Steinkampf, D.L. Driscoll, J.M. Nelson, W.L. Elliott, B.J. Roberts, C.L. Stoner, P.W. Vincent, D.J. Dykes, R.L. Panek, G.H. Lu, T.C. Major, T.K. Dahring, H. Hallak, L.A. Bradford, H.D. Showalter, A.M. Doherty, J. Med. Chem. 41 (1998) 3276–3292.
[24] A.M. Thompson, G.W. Rewcastle, S.L. Biushelle, B.G. Hartl, A.J. Kraker, G.H. Lu, B.L. Batley, R.L. Panek, H.D.H. Showalter, W.A. Denny, J. Med. Chem. 43 (2000) 3134–3147.
[25] R.D. Cramer III, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5939–5967.
[26] G. Cruciani, K.A. Watson, J. Med. Chem. 37 (1994) 2589–2601.
[27] R. Ragno, S. Simeoni, S. Valente, S. Massa, A. Mai, J. Chem. Inf. Model. 46 (2006) 1420–1430.
[28] J. Caballero, M. Fernández, M. Saavedra, F.D. Gonzáles-Nilo, Bioorg. Med. Chem. 16 (2008) 810–821.
[29] M. Oblak, M. Randic, T. Solmajer, J. Chem. Inf. Comput. Sci. 4 (2000) 994–1001.
[30] Q. Shen, Q.-Z. Lu, J.-H. Jiang, G.-L. Shen, R.-Q. Yu, Eur. J. Pharm. Sci. 20 (2003) 63–71.
[31] R. Guha, P.C. Jurs, J. Chem. Inf. Comput. Sci. 44 (2004) 2179–2189.
[32] G. Chen, X. Luo, W. Zhu, C. Luo, H. Liu, C.M. Phua, K. Chena, H. Jiang, Bioorg. Med. Chem. 12 (2004) 2409–2417.

[33] R. Thaimattam, P.R. Daga, R. Banerjee, J. Iqbal, Bioorg. Med. Chem. 13 (2005) 4704–4712.
[34] G. Cruciani, M. Pastor, W. Guba, Eur. J. Pharm. Sci. 11 (Suppl. 2) (2000) S29–S39.
[35] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639–1662.
[36] C.B. Breitenlechner, A.K. Norman, K.H.S. Scheiblich, H. Koll, E. Greiter, S. Koch, W. Schafer, R. Huber, R.A. Engh, J. Mol. Biol. 353 (2005) 222–231.
[37] Further information at the web site: <http://openbabel.sourceforge.net>.
[38] GRID 22, Molecular Discovery Ltd, 215 Marsh Road, Pinner, Middlesex, UK.

[39] GOLPE 4.5, Multivariate Infometric Analyses Srl, Viale dei Castagni 16, Perugia, Italy, 1999.
[40] S. Wold, E. Johansson, M. Cocchi, in: H. Kubinyi (Ed.), 3D QSAR in Drug Design: Theory, Methods and Applications, Escom, Leiden, 1993, pp. 523–550.
[41] M. Celano, S. Schenone, D. Cosco, M. Navarra, E. Puxeddu, L. Racanicchi, C. Brullo, E. Varano, S. Alcaro, E. Ferretti, G. Botta, S. Filetti, M. Fresta, M. Botta, D. Russo, Endocr. Relat. Cancer, 15 (2008) 499–510.
[42] H. Mukaiyama, T. Nishimura, H. Shiohara, S. Kobayashi, Y. Komatsu, S. Kikuchi, E. Tsuji, N. Kamada, H. Ohnota, H. Kusama, Chem. Pharm. Bull. 55 (2007) 881–889.