

Theoretical studies on pyrimidine substituent derivatives as dual inhibitors of AP-1 and NF- κ B

Li Qian · Si-Yan Liao · Zu-Liang Huang · Yong Shen · Kang-Cheng Zheng

Received: 5 March 2009 / Accepted: 16 October 2009 / Published online: 27 November 2009
© Springer-Verlag 2009

Abstract Theoretical studies on the three-dimensional (3D) quantitative structure-activity relationship (QSAR) and mechanisms of action of a series of pyrimidine substituent derivatives as dual inhibitors of AP-1 and NF- κ B were carried out using comparative molecular field analysis (CoMFA) and docking methods. The established 3D-QSAR model exhibits a satisfying statistical quality and prediction ability. Docking results show somewhat lower average values of the flexible and rigid energy scores in the chosen binding sites. The docking analysis offers appropriate orientations and conformations of these compounds at the binding sites to both AP-1 and NF- κ B in good agreement with the 3D-QSAR model from CoMFA. The combined CoMFA and docking study suggests the following substituent selections: substituent R₂ should be a kind of H–N–thienyl or CH₃–N–thienyl group; substituent R₅ should be a kind of COO–tBu or COOEt group; and substituent R₄ should be a CH₂CH₃ or 2-thienyl group. The docking analysis also shows that the binding sites fall just at the joint regions between AP-1 (or NF- κ B) and DNA, where these compounds can effectively prevent free AP-1 and NF- κ B from binding to DNA, and this may be the reason that derivatives

with pyrimidine substituents have an inhibition function. In addition, a very interesting finding was that the binding sites of both AP-1 and NF- κ B have a common structural characteristic, thereby providing a reasonable explanation for the dual inhibition functions of these compounds towards both AP-1 and NF- κ B. These theoretical results help to deepen our understanding of the inhibition mechanism of these pyrimidine substituent derivatives, and will aid in directing further drug-molecular design.

Keywords Pyrimidine derivative · 3D-QSAR · Docking analysis · DNA · Activator protein-1 · Nuclear factor kappa B

Introduction

It is now well-accepted that T-lymphocytes (T-cells) play an important role in both the initiation and propagation of immune responses against a number of inflammatory diseases [1, 2]. For example, in allergies and autoimmune diseases such as asthma, psoriasis, rheumatoid arthritis and transplant rejection, T-cell driven immune responses appear to overreact [3, 4]. In activated T cells, transcription factors such as the activator protein-1 (AP-1) and nuclear factor- κ B (NF- κ B) are DNA-binding proteins that act as regulators of inducible gene expression; the former regulates interleukin-2 (IL-2), IL-3 and granulocyte-macrophage colony stimulating factor (GM-CSF) while the latter regulates the proinflammatory cytokines IL-1, IL-6, IL-8 and tumor necrosis factor- α (TNF α) [5–9]. Therefore, potential regulation of a number of inflammatory diseases could converge on inhibition of inhibitors of AP-1 and NF- κ B transcriptional activation in T cells. At present, inhibition of AP-1 and NF- κ B has become an attractive target in the development of novel anti-inflammatory drugs.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-009-0609-8) contains supplementary material, which is available to authorized users.

L. Qian · Z.-L. Huang
Department of Chemistry,
Youjiang Medical College for Nationalities,
Guangxi, Baise 533000, People's Republic of China

L. Qian · S.-Y. Liao · Y. Shen (✉) · K.-C. Zheng
School of Chemistry and Chemical Engineering,
Sun Yat-Sen University,
Guangzhou 510275, People's Republic of China
e-mail: cessay@mail.sysu.edu.cn

Recently, a series of new pyrimidine substituent derivatives that clearly inhibit NF- κ B and AP-1 activation have been designed and synthesized. However, to our knowledge, the mechanism whereby these compounds inhibit AP-1 and NF- κ B activation remains quite unclear. An urgent need for novel dual functional inhibitors towards AP-1 and NF- κ B has provided the impetus for understanding the structural characteristic of inhibitors as well as the inhibition mechanism at a molecular level.

Quantitative structure-activity relationship (QSAR), which quantitatively correlates variations in biological activity with molecular structures and properties, has been used efficiently for the study of biological mechanisms of various reactive chemicals for many years [10–16]. In particular, comparative molecular field analysis (CoMFA) has become well established for ligand-based three dimensional (3D)-QSAR studies. This method offers visual images of established 3D models and thus facilitates understanding of the properties of the ligand at active sites [17–23]. Docking analysis is another useful methodology based on 3D-models, since it can offer a more visual image of the interaction between a drug molecule as ligand, and an organism (e.g., protein and DNA, etc.) as acceptor [24–26]. Therefore, a combined 3D-QSAR and docking study could offer greater insight into the interaction mechanism between drug and organism, and further direct functional drug design.

In this study, theoretical studies on the 3D-QSAR and mechanism of action of a series of newly recorded pyrimidine substituent derivatives as dual inhibitors towards AP-1 and NF- κ B were carried out using CoMFA and docking methods. The main focus of this article is to establish an optimal 3D-QSAR model for this series of compounds and reveal the binding action between the inhibitor and AP-1 (and NF- κ B), and to further explore the dual inhibition mechanism of these compounds towards both AP-1 and NF- κ B. We expect the results obtained to offer some useful theoretical references for experimenters.

Computational methods

Studied compounds and biological activity data

All derivatives studied in this article (Table 1) were selected from publications reported by the same laboratory [4]. Since similar experimental procedures were applied to yield the biological activity data in these publications, biological activity data (represented as IC_{50} values) were considered comparable. A set of 52 pyrimidine substituent derivatives were selected to perform this study. These 52 compounds were divided into a training set (35) and a test set (17) based on such the considerations that the test set adequately

covers the activity range, and compounds of varied but limited range were divided randomly. As a result, the test set comprises compounds 3, 6, 9, 11, 14, 19, 20–23, 29, 35, 36, 42, 46, 47 and 49, shown in Fig. 1. The biological activities of the compounds were expressed as IC_{50} , which is defined as the concentration of compound required to reduce luciferase activity to 50% of the control value. These values are generally transformed to pIC_{50} ($-\log IC_{50}$) in QSAR studies. All pIC_{50} values of selected compounds are distributed equally within the range 4.52–7.46 [4]. The structures and biological activity data of these compounds are shown in Table 1, in which the studied compounds were renumbered according to the structural characteristics determined in this work.

3D-QSAR

CoMFA studies were performed using the molecular modeling software SYBYL v6.9 (Tripos, St. Louis, MO) running on a SGI R2400 workstation. All parameters used in SYBYL v6.9 were default unless stated otherwise [27].

Alignment generation

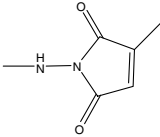
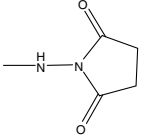
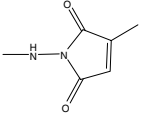
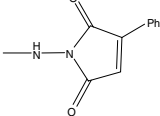
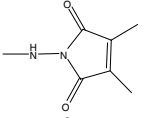
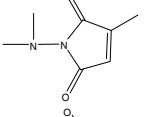
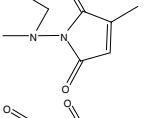
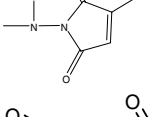
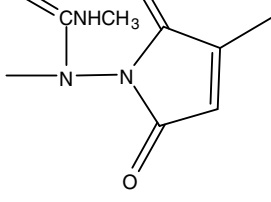
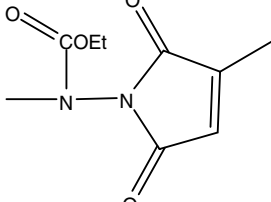
Generation of the correct spatial alignment of the investigated compounds for 3D-QSAR analysis is of vital importance, since the correctness of the analysis is dependent on the quality of the alignment. This challenging step is often impeded by the lack of biologically active conformations of the compounds binding to their target protein in complex. In the absence of any experimental knowledge of the bioactive conformations, determining the low-energy conformation as the biologically active conformation is the first step. Therefore we obtained the global energy-minimum conformation of these compounds by using Chem3D software for the compounds with rotatable bonds.

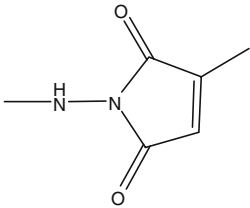
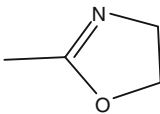
Here, all structures of the compounds in the training set were first obtained in ChemDraw, and optimized by the molecular mechanics method in Chem3D [28]. Then, fully geometric optimizations of 35 compounds were respectively performed using the standard Tripos force field including the electrostatic term calculated from Gasteiger and Hückel atomic charges. The steepest descent and conjugate gradient methods were used to perform energy minimization, and a cut-off value of 0.21 kJ mol^{-1} was adopted. Finally, each molecule was aligned on the common framework of the pyrimidine group (see Fig. 1).

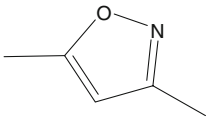
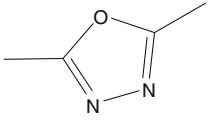
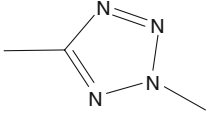
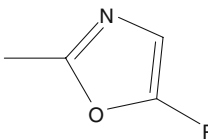
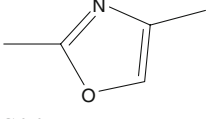
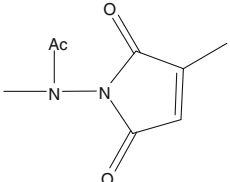
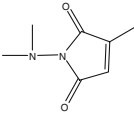
3D-QSAR method

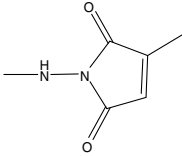
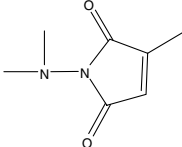
The CoMFA study was performed with the QSAR module of SYBYL v6.9 [27]. The steric (Lennard-Jones) and electrostatic (Coulombic) fields were calculated at each

Table 1 Three-dimensional quantitative structure-activity relationship (3D-QSAR) results of the studied compounds

No.	R ₂	R ₄	R ₅	<i>pIC</i> ₅₀	CoMFA
1.		CF ₃	COOEt	5.70	5.57
2.		CF ₃	COOEt	5.80	5.83
3.		2-(5-Methylthienyl)	COOEt	7.35	6.26
4.		CF ₃	COOEt	5.43	5.56
5.		CF ₃	COOEt	5.41	5.63
6.		CF ₃	COOEt	6.52	5.86
7.		CF ₃	COOEt	6.35	6.40
8.		CF ₃	COOEt	6.08	6.09
9.		CF ₃	COOEt	6.23	6.14
10.		CF ₃	COOEt	6.42	6.43

11.		H	COOEt		
				4.89	5.63
12.	"	CH ₂ CH ₃	COOEt	6.40	6.56
13.	"	CH ₂ Ph	COOEt	5.42	5.33
14.	"	CH ₂ OCH ₃	COOEt	5.22	5.90
15.	"	2-Thienyl	COOEt	6.85	6.67
16.	"	3-Thienyl	COOEt	6.70	6.77
17.	"	2-(5-Chlorothieryl)	COOEt	6.28	6.44
18.	"	2-Benzo[b]thienyl	COOEt	6.10	6.26
19.	"	2-Thiazolyl	COOEt	5.66	6.50
20.	"	Cyclopropyl	COOEt	5.82	6.72
21.	"	CF ₃	COO- <i>t</i> Bu	6.68	6.19
22.	"	CF ₃	COOH	4.52	5.14
23.	"	CF ₃	CONH ₂	4.52	5.35
24.	"	CF ₃	CONMe ₂	4.52	4.65
25.	"	CF ₃	COMe	5.36	5.26
26.	"	CF ₃		5.19	5.03
27.	"	CH ₂ CH ₃	COPh	6.19	6.26
28.	"	CH ₂ CH ₃	COCH ₂ Et	6.35	6.42
29.	"	CH ₂ CH ₃	CH ₂ OH	5.31	5.23
30.	"	CH ₂ CH ₃	CN	5.96	5.70

31.	"	CH ₃	COCH ₃	5.11	5.13
32.	"	CH ₂ CH ₃		5.00	4.95
33.	"	CH ₂ CH ₃		5.00	5.11
34.	"	CH ₂ CH ₃		5.00	4.97
35.	"	CH ₂ CH ₃		5.55	4.89
36.	"	2-Thienyl		6.12	5.49
37.	"	2-Thienyl	COO- <i>t</i> Bu	7.30	7.14
38.		CH ₂ CH ₂ CH ₃	COOEt	6.37	6.27
39.		CF ₂ CF ₃	COOEt	6.35	6.52
40.	"	CH ₂ CH ₃	COOEt	7.46	7.14
41.	"	2-Benzo[b]thienyl	COOEt	6.89	6.84
42.	"	CH ₂ CH ₃	COOPh	6.39	6.88
43.	"	CF ₂ CH ₃	COOEt	6.46	6.25
44.	"	2-Thienyl	COOEt	7.03	7.29
45.	"	3-Thienyl	COOEt	6.92	6.80
46.	"	2-(5-Methylthienyl)	COOEt	6.46	7.19

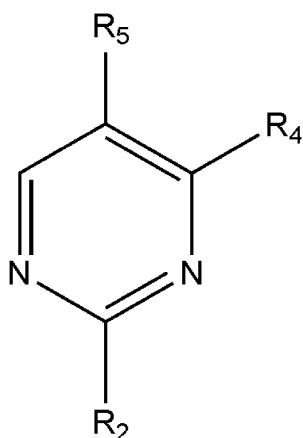
47.	"	CF ₂ CH ₃	CH ₂ OCH ₃	6.20	5.79
48.		CH ₃	COOEt	5.72	5.67
49.	"	CF ₂ CH ₃	COOEt	6.70	5.96
50.	"	Ph	COOEt	5.92	5.96
51.	"	2-Furanyl	COOEt	6.00	6.19
52.		2-Thiazolyl	COOEt	5.77	5.73

grid point using a sp^3 carbon probe with +1 charge, 0.2 nm grid spacing and 0.152 nm van der Waals radius, and both electrostatic and steric cutoffs were set at $125.46 \text{ kJ mol}^{-1}$. Meanwhile, the default value of 0.3 was used for the attenuation factor (α). The partial least squares (PLS) method was used to construct and validate the CoMFA model, in which the column filtering value was set at $8.364 \text{ kJ mol}^{-1}$ [29]. Cross-validation was performed with the leave-one-out procedure [30, 31]. The optional number of components N for final PLS analysis was defined as that yielding the highest cross-validation coefficient (q^2). Finally, the correlation of biological activity with the diversified field was described by 3D contour maps of CoMFA.

Molecular docking

The 3D structures of AP-1 and NF- κ B, obtained from the Protein Data Bank (pdb ids: 1NFK and 2H7H, respectively)

Fig. 1 Structural schematic diagram of the pyrimidine substituent derivatives. $pIC_{50} - \log IC_{50}$ of the concentration of compound required to reduce luciferase activity to 50% of the control value, CoMFA Comparative molecular field analysis



were used for docking analysis. All water molecules were removed. Docking studies were performed with the DOCK 6.0 program [32].

Beginning with docking, the program-package SYBYL 6.9 was used to add hydrogen atoms and charges to the protein. The ligand charge was obtained from Sybyl 6.9 by adopting the Gasteiger Hückel charge, and the surface of the protein was calculated using the default parameters of the Dms program [27–33]. To obtain binding sites, some spheres were generated and selected by the SPHGEN module of the DOCK 6.0 program [32].

In docking analysis, there is rigid docking and flexible docking, which represent two quite different docking approaches. In rigid docking, which is based on receptor spheres generated by the SPHGEN module as well as the heavy atom centers of the ligand, the ligand is docked rigidly to the receptor, whereas in flexible docking, which is based on the anchor-and-growing algorithm, the ligand is docked flexibly to the receptor. Obviously, these two approaches differ in principle. In order to obtain reliable binding sites from these different approaches, we adopted both rigid and flexible docking to perform this research. The convergent cycle, computational path and computational time are also different due to the different computational principles.

After docking to the active site, the structures were further optimized and quite changed, thus the docked structures were quite different and hard to align. Although we tried to align them, the results are quite poor. 3D-QSAR analysis is basically a statistical method. Structural alignment is based on the assumption that all ligands interact with the active site of the acceptor by taking the same or a similar way, i.e., the distribution of a pharmacophore

should be the same or similar. Although there are some differences in principle between docking analysis and 3D-QSAR study, combining the complementary advantages of each of these two approaches can offer more insight into understanding the interaction mechanism between drug and organism.

The 52 compounds were each flexibly and rigidly docked into the binding sites using the DOCK 6.0 program. The box size, grid space and energy cutoff distance were set at 12 Å, 0.3 Å and 9,999 Å, respectively. The maximum orientation in flexible docking was set as 500 and that in rigid docking was set as 120,000 and 250,000 for AP-1 and NF-κB, respectively. All parameters used in docking were default unless stated otherwise. The interaction between each of the docked ligands and the binding pocket of the protein model was also studied using the DOCK 6.0 program.

Results and discussion

CoMFA analysis

CoMFA analyses were performed with steric and electrostatic fields. The contributions of the steric and electrostatic fields to biological activity in the CoMFA model are 59.8% and 40.2%, respectively. This indicates that both the steric and electrostatic fields make a significant contribution to biological activity, and that the steric field is more predominant than the electrostatic field. So, changing either the volume or the polarity of the compound can be used to improve its inhibitory activity towards AP-1 and NF-κB transcriptional activation.

The CoMFA-predicted biological activity values as well as experimental biological activity values of the 35 compounds in training set are listed in Table 1; the good correlation between them is shown in Fig. 2. The correlation coefficient of non-cross validation R^2 is 0.922 (or R of 0.960), cross-validation coefficient q^2 is 0.519. The F value and standard error of estimation are 112.7 and 0.156, respectively. All data points are close to the line, confirming the good statistical quality and predictive ability of the model.

The results of CoMFA can be shown as 3D contour maps (see Fig. 3). The various colored contours are composed of points having a larger influence on biological activity and whose equivalence spaces in molecule exhibit different steric and electrostatic character.

Some main characteristics of the steric field contours can be seen in Fig. 3a: (1) three rather larger yellow-colored contours surround substituent R_2 (mainly surrounding the CH_3 sites of the R_2 termini); (2) a very large green-colored contour also surrounds the terminus of substituent R_4 ; (3) three rather larger yellow-colored contours along with a

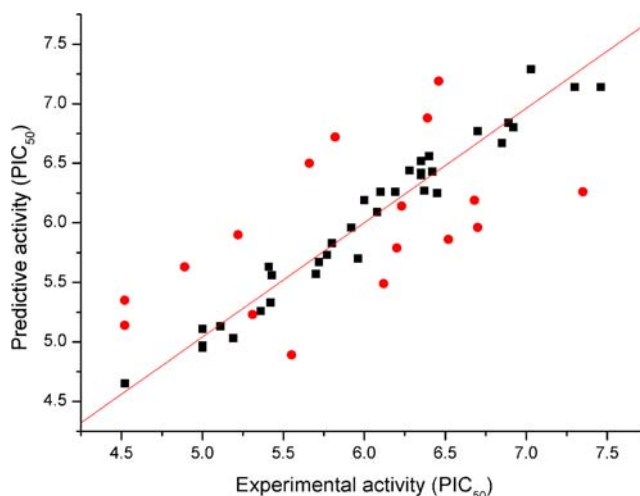
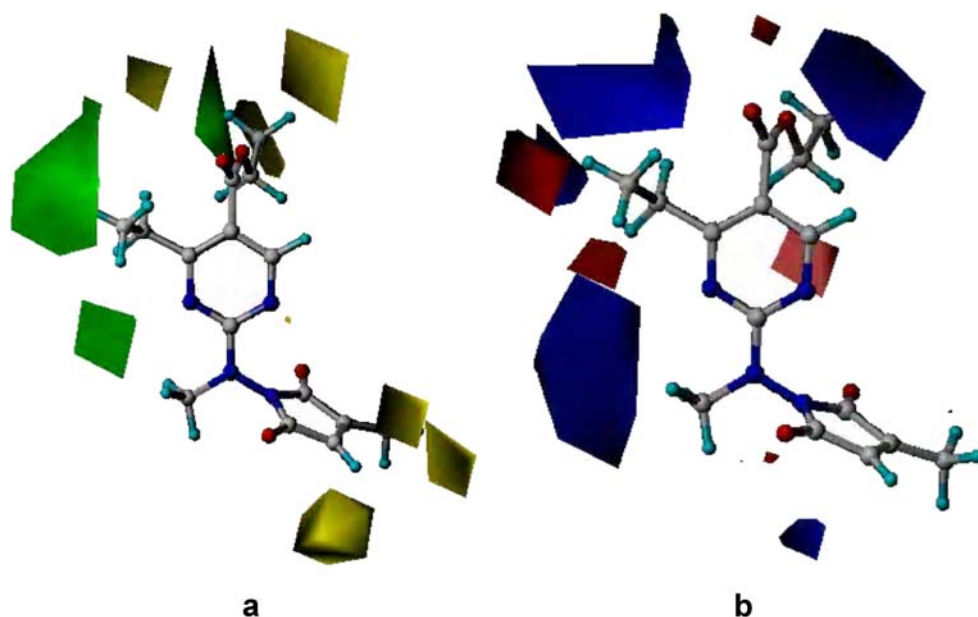


Fig. 2 Comparison between CoMFA-predicted and experimental biological activity values of training set (*black squares*) and test set (*red circles*) compounds

green-colored contour surround substituent R_5 . These steric characteristics suggest that selecting substituents with a suitable volume is very important for improving biological activity. Of the compounds differing only in R_2 , the activity of compound **2** ($\text{pIC}_{50}=5.80$) is higher than those of **1**, **4**, and **5** (5.70, 5.43 and 5.41, respectively), because the terminus of R_2 of compound **2** is an H atom, which does not enter the yellow area, whereas that of the latter three compounds is CH_3 or Ph, which can enter the yellow area. Of the compounds differing only in R_4 , compound **3**, where R_4 is 2-(5-Methylthienyl), has the highest activity ($\text{pIC}_{50}=7.35$), whereas compound **11** ($R_4 = \text{H atom}$) has the lowest activity (4.89) of compounds **11–20** and **3** (4.89–7.35) because the volume of R_4 of compound **3** is very large and thus can enter the large green area, whereas the situation is just the opposite in compound **11**. Of the compounds differing only in R_5 , compound **21** ($R_5 = \text{COO-tBu}$) has the highest activity (6.68) of compounds **21–26** (<5.40, except **21**), because its R_5 volume fits the steric field. In addition, the activity of compound **40** (7.46) is higher than that of compound **42** (6.39), because the R_5 volume of the former ($R_5 = \text{COOEt}$) is well suited to the steric field whereas that of the latter ($R_5 = \text{COOPh}$) may be too large to enter the yellow area.

On the other hand, Fig. 3b illustrates some main characteristics of the electrostatic field contours: (1) a very large blue-colored contour appears around CH_3 , binding to N of the substituent R_2 , especially around the terminal H atoms (with electropositive charges) of CH_3 ; (2) a large blue-colored contour and a medium-sized blue-colored contour as well as two medium-sized red-colored contours surround the terminus of R_4 ; (3) a very large blue-colored contour surrounds the terminus of R_5 . These characteristics of the electrostatic field contours can be used to explain the

Fig. 3 **a** Steric field contour map of CoMFA. **b** Electrostatic field contour map of CoMFA (the selected molecule is compound **40**, which has the highest inhibition activity)



experimental results shown in Table 1. In the series of compounds studied, compound **40** has the highest activity ($\text{pIC}_{50}=7.46$). This may be due to the fact that its R_2 , R_4 and R_5 substituents are well suited not only to the steric field but also to the electrostatic field, in which all of the terminal H atoms (with positive charges) of substituents R_2 , R_4 and R_5 fall into the blue areas. Similarly, compound **3** with $R_4 = 2$ -(5-Methylthienyl) has a higher activity (7.35) because it suits not only the steric field but also the electrostatic field. Meanwhile, compounds with $R_5 = \text{COO-tBu}$ have relatively higher activities, e.g., compound **21** has the highest activity (6.68) of compounds **21–26** (<5.40 except for **21**) because the terminal H atoms of COO-tBu fall just within the large blue contour, besides which its volume is suitable for the steric field; the activity (7.30) of compound **37**, with R_5 of COO-tBu, is also somewhat higher than that of compound **36** (6.12), possibly also for this reason. Compounds **44** (7.03), **45** (6.92), **49** (6.70), **15** (6.85), **16** (6.70) and so on also have higher activities because their R_2 and R_5 substituents have similar structural characteristics to those mentioned above. In addition, the activity (7.03) of compound **44** with $R_4 = 2$ -thienyl is also higher than that of compound **36** (6.12) with $R_4 = 3$ -thienyl, and this may also be the reason that the S atom of 2-thienyl with its negative charges more effectively falls into the red contour, and the H atoms of 2-thienyl with their positive charges more effectively fall into the blue contour relative to R_4 of 3-thienyl.

In summary, according to the established 3D-QSAR model, both the steric and electrostatic fields make a significant contribution to biological activity, with the steric field being more important than the electrostatic field. The R_2 substituent selected should be an H–N–Thienyl or CH_3 –N–Thienyl-type group, and the R_5 substituent should be a

COO-tBu or COOEt group because such a selection is well suited not only to the steric field but also to the electrostatic field. The R_4 substituent selected should be a CH_2CH_3 or 2-thienyl group because their volume suits the steric field, which plays the main role in this position.

Docking study

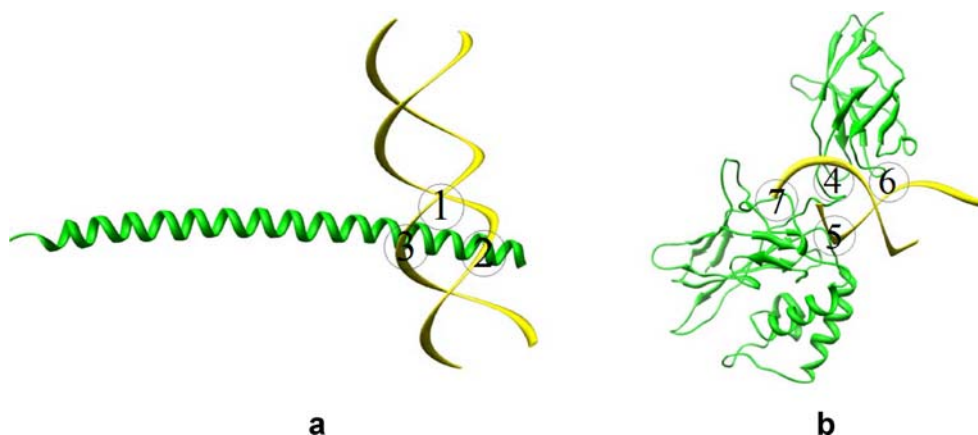
Figure 4 shows the 3D structures of AP-1 and NF- κ B, as well as the possible DNA binding sites (1–7) of the analogues tested here according to the docking study. The energy scoring values of DOCK are approximate molecular mechanics interaction energies, consisting of van der Waals and electrostatic components. The lower the energy scoring value, the more effective the binding of ligand to protein.

The compounds in training set were each flexibly and rigidly docked into the binding sites. The energy scores in these different binding sites are summarized in Tables S1 and S2. The energetic differences suggest a preference for sites neglecting entropic contributions.

From Tables S1 and S2, the average values of the flexible and rigid energy scores were -26.298 , -24.839 , -26.352 and -24.240 , -23.906 , -24.251 $\text{kcal}\cdot\text{mol}^{-1}$ from site 1 to 3 for AP-1, respectively. Binding to site 2 not only has a high average value of the flexible energy scores but also has a high average value of the rigid energy scores, compared with binding to sites 1 and 3. We can also see that the average energy score (either flexible or rigid) on binding site 3 is lower than that on binding site 1. So, site 3 is the most probable binding site from among sites 1–3, and will be used to further discuss the inhibition of the title complexes towards AP-1.

On the other hand, the average values of the flexible and rigid energy scores were -37.278 , -33.699 , -31.533 , -33.379

Fig. 4 Possible sites of action (1–7) of the set of pyrimidine substituent compounds investigated here with activator protein-1 (AP-1) binding to DNA (a) and nuclear factor- κ B (NF- κ B) binding to DNA (b). Green AP-1 and NF- κ B protein, yellow DNA



and -35.701 , -28.773 , -30.4887 , -33.379 kcal·mol $^{-1}$ from site 4 to 7 for NF- κ B, respectively. Site 4 obviously has lower average values in both flexible and rigid energy scores than sites 5–7. So, site 4 is the most probable binding site from among sites 5–7. In addition, it is interesting to find that site 4 lies at the junction between P50 and the DNA sequence 5'-GGGACTCCTC-3', which is part of a long-terminal repeat (LTR) sequence of DNA. Recent studies have shown that the P50 subunit of the NF- κ B complex is the one that mainly interacts with the LTR sequence [31, 34, 35]. Site 4 will be also used to further discuss the inhibition of the title complexes towards NF- κ B.

Figure 5 shows the binding information obtained for compound **40** to the surfaces of site 3 and site 4, respectively. It was very interesting to find that these two pictures of compound **40** binding to site 3 and site 4 have key common characteristics in good agreement with the

3D-QSAR model at first glance they appear to be different. We can see clearly that there are some pockets suited to the conformation of compound **40** for site 3 (AP-1) and site 4 (NF- κ B), respectively. Moreover, the binding orientations of the compound are also in line, on the whole, with the 3D-QSAR model. For example, there are several large spaces for pyrimidine in the main part of the molecule in sites 3 and 4; for R₂ (CH₃-N-thienyl), its branch group CH₃ is blocked and the H atoms of the CH₃ group fall just within the blue surface (N atom) of the protein, while its other branch group thienyl is not blocked and there is an extensive space suitable to house it; for an R₅ group of the COO-tBu or COOEt type, there is a very suitable long groove to accommodate it, and its terminal H atoms fall just within the red surface (O atom) of the protein; there is also a suitable space for R₄ of CH₂CH₃ or the thienyl group, and so on.

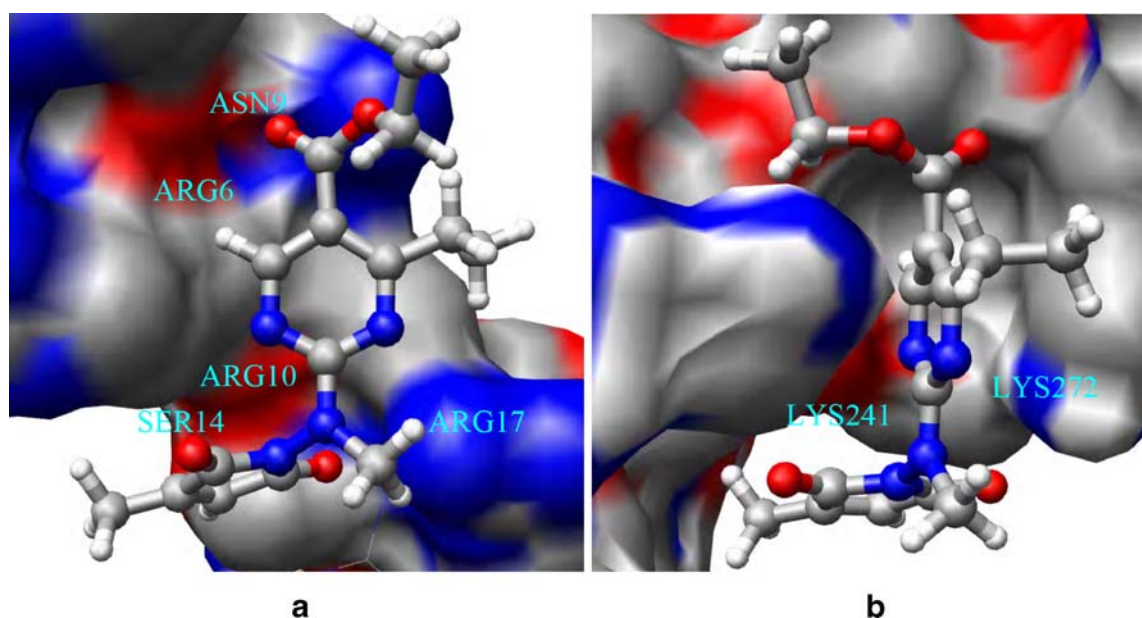


Fig. 5 a Possible binding mode picture of inhibitory compound to the surface of AP-1. b Possible binding mode picture of inhibitory compound to the surface of NF- κ B. The selected molecule is

compound **40** with the highest inhibition towards AP-1 and NF- κ B transcriptional activation; red surface of oxygen atoms, blue surface of nitrogen atoms

Figure 6 further shows the interaction between the ligand and the proteins of AP-1 and NF- κ B. For AP-1, the residues nearest to the ligand are ASN9, ARG10 and ARG17. For NF- κ B, the residues nearest to the ligand are TYR19, SER208, LYS241 and LYS272. From these figures, we can see that some hydrogen bonds clearly exist between the O (or N) atoms of the ligand and H atoms of AP-1 (or NF- κ B).

In short, the docking simulation and 3D-QSAR analysis were found to be in satisfactory agreement with each other, suggesting that site 3 (AP-1) and site 4 (NF- κ B) share a common structural characteristic suitable for the conformation of this kind of compound. This may explain why these compounds have dual inhibition functions towards AP-1 and NF- κ B. Moreover, these results further show that sites 3 and 4 are the most probable binding sites for AP-1 and NF- κ B, respectively. Thus, sites 3 and 4 are selected to further discuss the inhibition mechanism.

3D-QSAR and docking analysis in the test set

The biological activity values predicted by CoMFA and the experimental values for the 17 compounds in the test set are also listed in Table 1, and the considerable correlation between them is also shown in Fig. 2 (dots). The correlation coefficient of non-cross validation R_{test} is 0.556 (or R_{ext}^2 is 0.308) and the standard deviation SD_{test} is 0.547. All these data points are distributed symmetrically about the line. Such a correlation coefficient for an external test set is indeed on the low side. However, considering the correlation coefficient of non-cross validation R of 0.960 (or R^2 of 0.922) and

cross-validation coefficient q^2 of 0.519 for the training set as well as the consistency between the 3D-QSAR model and docking analysis, the established model should be acceptable for such structurally variable compounds.

The 17 compounds in the test set were also each flexibly and rigidly docked into the binding sites. From Tables S1 and S2, the average values of the flexible and rigid energy scores were -25.580 , -24.472 , -25.850 and -23.395 , -23.516 , -23.925 kcal·mol $^{-1}$ from site 1 to 3 for AP-1, respectively. The data in the test set can be regarded as being essentially in line with those in training set.

Similarly, the average values of the flexible and rigid energy scores were -35.518 , -32.731 , -30.838 , -32.318 and -34.415 , -27.772 , -29.837 , -32.126 kcal·mol $^{-1}$ from site 4 to 7 for NF- κ B, respectively. We found that the flexible and rigid energy scores in site 4 have clearly lower average values than those in sites 5–7, respectively. Thus, site 4 can be regarded as the most probable from among sites 4–7.

To summarize, the results from the test set also indicate that sites 3 and 4 are the most probable binding sites for AP-1 and NF- κ B, respectively, in agreement with the results from the training set.

Elucidating the inhibition mechanism of pyrimidine derivatives towards AP-1 and NF- κ B

Transcription factors such as AP-1 and NF- κ B have recently been recognised as DNA-binding proteins that are regulators of inducible gene expression. Therefore, regulation of a number of inflammatory diseases appears to

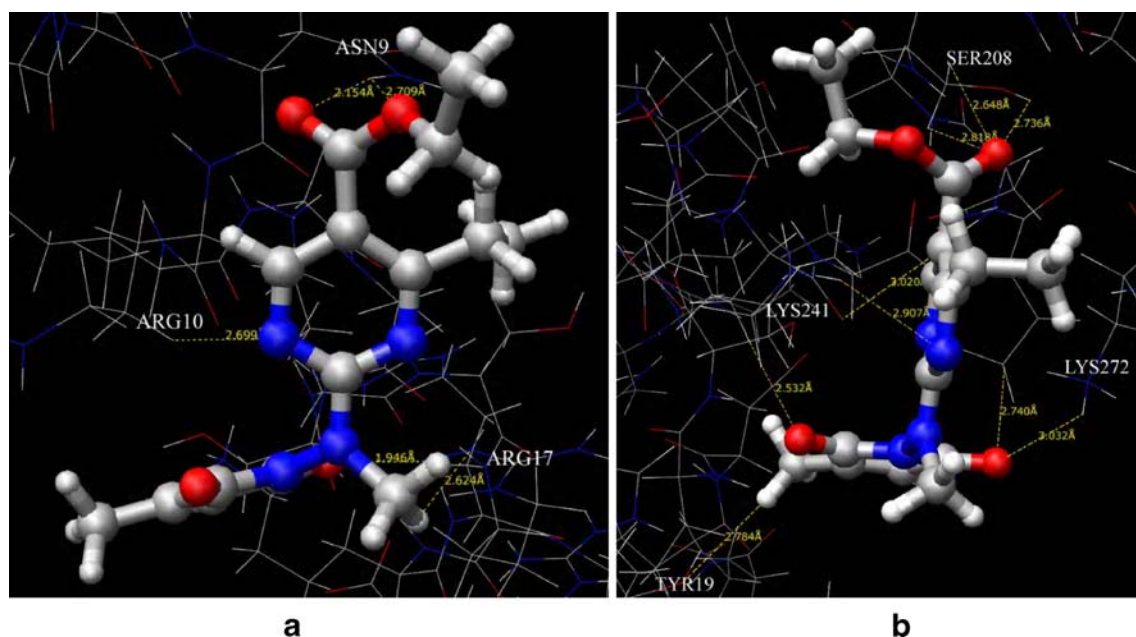


Fig. 6 **a** Interaction picture of inhibitory compound with the surface of AP-1. **b** Interaction picture of the compound with the surface of NF- κ B. The selected molecule is compound 40 with the highest inhibition towards AP-1 and NF- κ B transcriptional activation

come down to the action of an inhibitor of AP-1 and NF- κ B transcriptional activation in T cells [2, 3, 36, 37].

From Fig. 4, we can clearly see that, in the absence of an inhibitor, the interaction between AP-1 (Fig. 4a) or NF- κ B (Fig. 4b) and DNA at the point where they come into contact should be very strong because the distance is very short. That is to say, the inducible gene expression produced via the action of these DNA-binding proteins in a number of inflammatory diseases should be very strong. However, when an inhibitor exists at sites 1–7, in particular, at site 3 for AP-1 and site 4 for NF- κ B, there is a stronger interaction between AP-1 (and NF- κ B) and the inhibitor, meaning that the inhibitor can easily remain at these sites. Since these binding sites fall just within the region between AP-1 (and NF- κ B) and DNA, they must greatly weaken the interaction between AP-1 (and NF- κ B) and DNA, and thus induce a strong inhibition of the gene expression that causes these inflammatory diseases. This may explain why the pyrimidine derivatives investigated here have an inhibitory function. On the other hand, we also find that the distribution of the steric and electrostatic fields in the established 3D-QSAR model are in good agreement with the structural characteristics of the corresponding binding sites. Therefore, combining the docking results with 3D-model information could effectively direct the molecular design of this kind of inhibitor in addition to helping understand their mechanism of action.

Conclusions

Comparative molecular field analysis (CoMFA) and docking methods were applied synergistically to study novel pyrimidine derivatives as dual inhibitors of AP-1 and NF- κ B transcriptional activation. A 3D-QSAR model with good statistical quality and considerable predictive ability was established using CoMFA, allowing the main factors affecting activity to be discerned. Docking results show that there are somewhat lower average values for the flexible and rigid energy scores on the binding sites found, and that sites 3 and 4 seem to be the most likely binding sites for AP-1 and NF- κ B, respectively. The structural characteristics of the most probable binding sites from docking analysis were found to be in good agreement with CoMFA field distributions. Thus, our combined CoMFA and docking study suggests the following: the R₂ substituent selected should be an H–N–thienyl- or CH₃–N–thienyl-type group and the R₅ substituent selected should be a COO–tBu or COOEt group because such selections are well suited not only to the steric field but also the electrostatic field. Meanwhile, the R₄ substituent selected should be a CH₂CH₃ or 2-thienyl group because their volume suits the steric field playing the main role in this position. The

docking results also show that the most probable binding sites fall just within the region where AP-1 (or NF- κ B) and DNA overlap. So it may be reasonable to suggest that this kind of compound could act as NF- κ B-DNA and AP-1-DNA binding inhibitors, which can prevent free AP-1 and NF- κ B from binding to DNA. In addition, the dual inhibition functions of these compounds towards AP-1 and NF- κ B may be due to the common structural characteristics of site 3 (AP-1) and site 4 (NF- κ B), which suit the conformation of this kind of compound.

Acknowledgments We are pleased to thank the National Natural Science Foundation of China for financial support (No. 90608012). We heartily thank Molecular Discovery Ltd. for providing the Dock 6.0 program as freeware. We are also pleased to thank the College of Life Sciences, Sun Yat-Sen University for SYBYL 6.9 computation environment support.

References

1. Palanki MSS, Manning AM (1999) Interleukin-2 inhibitors in autoimmune disease. *Exp Opin Ther Patents* 9:27–39
2. Palanki MSS, Erdman PE, Ren MH, Suto AM, Bennett BL, Manning A, Ransone L, Spooner C, Ow A, Totsuka R, Tsaob P, Toriumib W (2003) The design and synthesis of novel orally active inhibitors of AP-1 and NF- κ B mediated transcriptional activation SAR of In vitro and In vivo studies. *Bioorg Med Chem Lett* 13:4077–4080
3. Palanki MSS, Erdman PE, Manning A, Ow A, Ransone L, Spooner C, Suto C, Suto M (2000) Novel inhibitors of AP-1 and NF-kappa B mediated gene expression: Structure-activity relationship studies of ethyl-4-[(3-methyl-2, 5-dioxo(3-pyrrolinyl)) amino]-2-(trifluoro-methyl)pyrimidine-5-carboxylate. *Bioorg Med Chem Lett* 10:1645–1648
4. Palanki MSS, Gayo-Fung LM, Shevlin GI, Erdman P, Sato M, Goldman M, Ransone L, Spooner C (2002) Structure-activity relationship studies of ethyl 2-[(3-methyl-2, 5-dioxo(3-pyrrolinyl)) amino]-4-(trifluoromethyl)pyrimidine-5-carboxylate: an inhibitor of AP-1 and NF-kappa B mediated gene expression. *Bioorg Med Chem Lett* 12:2573–2577
5. Manning AM, Lewis AJ (1999) New targets for anti-inflammatory drugs. *Curr Opin Chem Biol* 3:489–494
6. Pande V, Ramos MJ (2005) NF-kappaB in human disease: current inhibitors and prospects for de novo structure based design of inhibitors. *Curr Med Chem* 12:357–374
7. Li Q, Verma IM (2002) NF-kappaB regulation in the immune system. *Nat Rev Immunol* 2:725–734
8. Tobe M, Isobe Y, Tomizawa H, Nagasaki T, Takahashi H, Fukazawa T, Hayashi H (2003) Discovery of quinazolines as a novel structural class of potent inhibitors of NF-kappaB activation. *Bioorg Med Chem* 11:383–391
9. Palanki MSS (2002) Inhibitors of AP-1 and NF-kappa B mediated transcriptional activation: Therapeutic potential in autoimmune diseases and structural diversity. *Curr Med Chem* 9:219–227
10. Isayev O, Rasulev B, Gorb L, Leszczynski J (2006) Structure-toxicity relationships of nitroaromatic compounds. *Mol Divers* 10:233–245
11. Li RL (2004) Drug structure-activity relationship, 1st edn. Chinese Medicine Science and Technology Press, Beijing
12. Liao SY, Qain L, Chen JC, Lu HL, Zheng KC (2008) 2D and 3D-QSAR studies on antiproliferative thiazolidine analogs. *Int J Quantum Chem* 108:1380–1390

13. Liao SY, Chen JC, Qain L, Shen Y, Zheng KC (2008) QSAR studies and molecular design of phenanthrene-based tylophorine derivatives with anticancer activity. *QSAR Comb Sci* 27:280–288
14. Wang XD, Lin ZF, Yin DQ, Liu SS, Wang LS (2005) 2D/3D-QSAR comparative study on mutagenicity of nitroaromatics. *Sci China Ser B-Chem* 48:246–252
15. Wu WJ, Chen JC, Qain L, Zheng KC (2007) QSAR and molecular design of benzo[b] acronycine derivatives as antitumor agents. *J Theor Comput Chem* 6:223–231
16. Basak SC, Mills D, Hawkins DM (2008) Predicting allergic contact dermatitis: a hierarchical structure-activity relationship (SAR) approach to chemical classification using topological and quantum chemical descriptors. *J Comput Aided Mol Des* 22:339–343
17. Kharkar PS, Reith MEA, Dutta AK (2008) Three-dimensional quantitative structure-activity relationship (3D-QSAR) and pharmacophore elucidation of tetrahydropyran derivatives as serotonin and norepinephrine transporter inhibitors. *J Comput Aided Mol Des* 22:1–17
18. Datar PA, Khedkar SA, Malde AK, Coutinho EC (2006) Comparative residue interaction analysis (CoRIA): a 3D-QSAR approach to explore the binding contributions of active site residues with ligands. *J Comput Aided Mol Des* 20:343–360
19. Xiao AJ, Zhang ZY, An LY, Xiang YH (2008) 3D-QSAR and docking studies of 3-aryquinazolinethione derivatives as selective estrogen receptor modulators. *J Mol Model* 14:149–159
20. Cramer RD, Patterson DE, Bunce JD (1988) Comparative Molecular Field Analysis (CoMFA). I. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110:5959–5967
21. Qian L, Shen Y, Chen JC, Zheng KC (2006) 3D-QSAR study on a series of indolo[1, 2-b]quinazoline derivatives with anticancer activity and their molecular design. *Acta Phys-Chim Sin* 22:1372–1376
22. Qian L, Shen Y, Chen JC, Wang YX, Wu XT, Chen TJ, Zheng KC (2008) 3D-QSAR and docking studies of quinazoline derivatives with the inhibitory activity toward NF- κ B. *QSAR Comb Sci* 27:984–995
23. Zou XJ, Lai LH, Jin GY (2005) Study on the three-dimensional quantitative structure-activity relationship of pyridazinonyl-substituted 1, 3, 4-thiadiazoles. *Chin J Chem* 23:1120–1122
24. DaSilva CTP, Carvalho I, Taft CA (2006) Molecular dynamics, docking, density functional and admet studies of HIV-1 reverse transcriptase inhibitors. *J Theor Comput Chem* 5:579–586
25. Pande V, Sharma RK, Inoue JI, Otsuka M, Ramos JR (2003) A molecular modeling study of inhibitors of nuclear factor kappa-B (p50)-DNA binding. *J Comput-Aided Mol Des* 17:825–836
26. Yao Y, Han WW, Zhou YH, Li ZS, Li Q, Zhong DF (2007) Molecular docking study of the affinity of CYP2C9 and CYP2D6 for imrecoxib. *J Theor Comput Chem* 6:541–548
27. SYBYL 6.9 (2001) Tripos Associates, St Louis
28. Chemoffice (2005) CambridgeSoft, Cambridge
29. Cramer RD, Bunce JD, Patterson DE (1988) Crossvalidation, bootstrapping and partial least squares compared with multiple regression in conventional QSAR studies. *Quant Struct Act Relat* 7:18–25
30. Center V, Massart DL, DeNorrd OE, DeJong S, Vandeginste BM, Aterna C (1996) Elimination of uninformative variables for multivariate calibration. *Anal Chem* 68:3851–3858
31. Ghosh G, Van DG, Ghosh S, Sigler PE (1995) Structure of NF- κ B p50 homodimer bound to a κ B site. *Nature* 373:303–310
32. Irwin D, Kuntz DT, Moustakas P, Therese L (2005) DOCK 6.0 @ University of California Last updated October 2007. Web address: <http://dock.compbio.ucsf.edu>
33. Schaftenaar G, Noordik JH (2000) Molden: a pre- and post-processing program for molecular and electronic structures. *J Comput-Aided Mol Des* 14:123–134
34. Sharma RK, Chopra S, Sharma SD, Pande V, Ramos MJ, Meguro K, Inoue JI, Otsuka M (2006) Biological evaluation, chelation, and molecular modeling studies of novel metal-chelating inhibitors of NF-kappa B-DNA binding: Structure activity relationships. *J Med Chem* 49:3595–3601
35. Chen-Park FE, Huang DB, Noro B, Thanos D, Ghosh G (2002) The kappa B DNA sequence from the HIV long terminal repeat functions as an allosteric regulator of HIV transcription. *J Biol Chem* 277:24701–24708
36. Kumar NV, Bernstein LR (2001) Ten ERK-related proteins in three distinct classes associate with AP-1 proteins and/or AP-1 DNA. *J Biol Chem* 276:32362–32372
37. Pande V, Ramos M (2005) NF-kappa B in human disease: current inhibitors and prospects for de novo structure based design of inhibitors. *J Curr Med Chem* 12:357–374