Integration of pharmacophore mapping and molecular docking in sequential virtual screening: towards the discovery of novel JAK2 inhibitors†

Ting-Ting Yao,‡ab Jiang-Feng Xie,‡a Xing-Guo Liu,ab Jing-Li Cheng,b Cheng-Yuan Zhu,ab Jin-Hao Zhao,*ab and Xiao-Wu Dong*a

An integrated virtual screening protocol by combining molecular docking and pharmacophore mapping was established to identify novel inhibitors of JAK2 from a commercial compound database. Twelve novel and structurally diverse hits were selected and subjected to in vitro biological tests, and three compounds (A5, A6 and A9) with remarkable JAK2 inhibitory activity were identified. Then, the obtained structures were further used as the template for a subsequent similarity search, leading to the identification of another two promising compounds (B2 and B4). Selectivity profiles of JAK subtype and in vitro anti-cancer activity of the promising compounds were studied, revealing the promising compound B2 was of interest for further study because of its JAK2 selective profile, novelty of skeleton and significantly anti-proliferative effect against cancer cells. Finally, binding patterns of the compounds A5 and B2 were explored to provide a deeper insight for further structural optimization.

Introduction

The Janus kinase family (JAK1, JAK2, JAK3 and TYK2) plays a crucial role in the cytokine and growth factor mediated JAK-STAT signal transduction.1 JAK autophosphorylation leads to dimerization of phosphorylated STAT proteins and their translocation into the nucleus, thereby increasing cellular proliferation and resistance to apoptosis.2 Among the Janus kinase family, JAK2 is proved to be associated with the pathogenesis of myeloproliferative disorders (MPDs), primarily including polycythemia vera (PV), essential thrombocytemia (ET), and primary myelofibrosis (PMF).3–5 In particular, the essential role in cell growth and survival makes JAK2 an important target for cancer treatment.6 Until now, numerous JAK2 inhibitors with different scaffolds have been discovered, including AZD1480, Ruxolitinib, Tofacitinib, Lestaurtinib, CYT387, Pacritinib, etc.7–10 But unfortunately, only Ruxolitinib and Tofacitinib have been pushed into clinical use (Fig. 1). Therefore, the JAK2 inhibitors with better drug-like properties and novel structural skeleton are still highly in demand.

As an excellent computational tool for structurally skeleton hoping in drug design, pharmacophore model has been rapidly developed by medicinal chemists in searching for novel hit/lead compounds in various disease areas.11–14 Recently, a combined screening strategy of pharmacophore mapping and molecular docking show synergetic effect in virtual screening.15,16 As such, the integrally globally and partially essential pharmacophoric features for both of ligand and receptor can be simultaneously characterized, thereby showing more preferable performance (e.g. efficiency and accuracy) than the individually used method. Thus, we envisioned that an integration of ligand-based pharmacophore model and structure-based molecular docking can be conducted with respect to JAK2 inhibitors, and the development of JAK2 inhibitors with new scaffolds can be assisted and accelerated by such strategy. In connection with our previous work on computational medicinal chemistry,17 herein, we reported the establishment of pharmacophore and molecular docking models of structurally diverse JAK2 inhibitors, and

*ZJU-ENS Joint Laboratory of Medicinal Chemistry, Zhejiang Province Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, 310058, P. R. China. E-mail: jinhuazhao@zju.edu.cn; dongxw@zju.edu.cn
†Institute of Pesticide and Environmental Toxicology, Ministry of Agriculture Key Laboratory of Agricultural Entomology, Zhejiang University, Hangzhou 310029, P. R. China
‡School of Chemical Engineering and Light Industry, Guangdong University of Technology, Guangzhou, 510006, P. R. China
†‡ Electronic supplementary information (ESI) available: Fig. S1, S2 and Tables S1–S3. See DOI: 10.1039/c6ra24959k
‡ These authors contributed equally to this work.

Fig. 1 JAK2 inhibitors currently in clinical use or trials.
further sequential virtual screening on commercial SPECS database. To our delight, multiple-substituted aminothiazole derivative B2 bearing the novel scaffolds, the skeleton of which is remarkably different from the existing JAK2 inhibitors. Furthermore, selectivity profiles of JAK subtype and in vitro anti-cancer activity of the promising compounds were studied to demonstrate the further application value. Finally, molecular docking and dynamics were implemented to explore the potential binding mode of the promising compounds with JAK2 protein.

Methods

Docking-based virtual screening by LigandFit

The crystal structure of JAK2 (PDB entry: 4AQC) was employed as the template for molecular docking with the LigandFit protocol. All crystallographic water molecules were removed, and the missing hydrogen atoms were added using CHARMM forcefield using "Prepare Protein module". The resulting target protein structure was subsequently utilized to define the docking site by the bound inhibitor using "Find Sites as Volume of Selected Ligand" tool. For the conformationer generation of different ligands in the virtual library, variable numbers of Monte Carlo simulations was implemented. All the ligand–receptor interaction energies during the docking steps were calculated based on the Piecewise Linear Potential 1 (PLP1) forcefield. A short rigid body minimization was then performed and the top one pose for each ligand was saved. Scoring was performed with a set of scoring functions employed in LigandFit module, including LigScore1_Dreiding, LigScore2_Dreiding, -PLP1, -PLP2, Jain, -PMF, DOCK_SCORE.

Pharmacophore hypothesis generation

The training set, which consisted of 15 compounds, was selected by considering the structural diversity and the wide variety of activity ranges presented in Fig. S1 in the ESI. Inhibitory activities against JAK2 in the training set were reported as IC50 values ranging from 0.0004 μM to 7.218 μM, covering 5 orders of magnitude (see Table 1). All structures were built and minimized to the closest local minimum by using a modified CHARMM forcefield in the Catalyst program implemented in Discovery Studio 2.5 (Accelrys Software, Inc., San Diego, CA, USA), and the CAESAR algorithm was used for conformational analysis of each molecule. In this study, the maximum value of the setting number of the conformer was limited to 250 by using the “best conformers generation” method with a 20 kcal mol⁻¹ energy cutoff, and the default setting was used for the rest of the parameters. Four features (hydrogen bond acceptor, hydrogen bond donor, excluded volumes, and hydrophobic features) were selected to generate the pharmacophore hypothesis; the 3D-QSAR Pharmacophore Generation (HypoGen) module in Discovery Studio 2.5 was used. In addition, the HypoGen program with the default uncertainty value of 2 was used to find pharmacophore models that contain a maximum of five features.

Table 1. Actual and estimated IC50 (μM) for the training set based on the best pharmacophore hypothesis Hypo 1

| No. | Estimated IC50 (μM) | Actual IC50 (μM) | Error ratio | Ref. |
|-----|--------------------|-----------------|-------------|------|
| 1   | 0.0007             | 0.0004          | −0.7500     | 19   |
| 2   | 0.0097             | 0.0590          | 0.8356      | 19   |
| 3   | 0.0103             | 0.0120          | 0.1417      | 20   |
| 4   | 0.0105             | 0.0030          | −2.5000     | 21   |
| 5   | 0.0280             | 0.0510          | 0.4510      | 20   |
| 6   | 0.0711             | 0.0580          | −0.2259     | 22   |
| 7   | 0.2967             | 0.1550          | −0.9142     | 22   |
| 8   | 0.4060             | 0.1060          | −2.8302     | 23   |
| 9   | 0.5145             | 0.2310          | −1.2273     | 23   |
| 10  | 0.5257             | 0.2020          | −1.6025     | 21   |
| 11  | 0.8173             | 0.5330          | −0.5334     | 21   |
| 12  | 0.8281             | 1.4200          | 0.4168      | 23   |
| 13  | 1.5023             | 2.8900          | 0.4802      | 22   |
| 14  | 1.9519             | 7.2180          | 0.7296      | 20   |
| 15  | 2.1109             | 2.7000          | 0.2182      | 19   |

Hit-to-lead (H2L) screening

A second in silico screening of the constructed drug-like database was performed with “Find Similar Molecules by Fingerprints” protocol using the core structures of potent hits (A5, A6 and A9) as queries. 2D molecular fingerprint FCFc_6 method was applied to virtually screen the drug-like database. The minimum similarity was set as 0.3, and other parameters were maintained as default configuration. Following the similarity search, the obtained compounds were further screened using molecular docking and pharmacophore mapping, as described above.

In vitro JAK inhibition assays

The inhibitory activity of compounds against JAK1/2/3 enzyme was evaluated according to published procedure. The incubation mixtures contained the following components: 1.1 nM JAK1/2/3, 1.5 μM peptide substrate (5-FAM-KKKKEEIYFFFG-OH for JAK2) and 30 μM ATP. After the incubation for 180 min, the reaction mixture was analyzed on a Caliper LabChip 3000 (Caliper LifeSciences, Hopinkton, MA, USA) by electrophoretic separation of the fluorescent substrate and phosphorylated product. The virtual screening hits were firstly evaluated at the concentration of 20 μg mL⁻¹, and further tested for seven concentrations prepared from 3- or 4-fold serial dilution if the preliminary activity was more than 60%. The IC50 values were generated by the analytical software Prism 5.0 (GraphPad Software Pte Ltd.).

Cell viability assay

Cancer cells were counted in logarithmic phase and 5000 cells were placed in 96-well plates. After treatment with compound B2 for 48 hours, cells were incubated for an additional 2 h with CCK-8 reagent (100 μL mL⁻¹ medium) and the absorbance was read at 450 nm using a microplate reader (Sunnyvale, CA, USA). Cell proliferation inhibition rates were calculated according to the following formula: the proliferation inhibition ratio (%) = 1
undesirable moieties, compounds in the Specs database were subjected to “Lipinski’s rule of five” filter. After the above preparation, the final drug-like virtual library includes 153 691 compounds.

Evaluation of the scoring functions in LigandFit

Before the chemical libraries were screened, the accuracy of scoring functions in LigandFit for identifying the known JAK2 inhibitors was evaluated using distribution analysis and receiver operating characteristic (ROC) curve study. Therefore, 239 known JAK2 inhibitors were retrieved from the BindingDB database, while 10 000 inactive compounds were randomly chosen from the Maybridge database using the “Find Diverse Molecules” protocol in Discovery Studio 2.5; and they had the maximum diversity evaluated by the 2D similarities (Tanimoto coefficient) based on the FCPF_6 fingerprints. Then the JAK2 inhibitors and inactive compounds were processed with the “Prepare Ligands” protocol, and subsequently docked into the active site of JAK2 and evaluated by different scoring functions of LigandFit module. As shown in Fig. 2(a)–(g), all of the distributions analysis of seven docking scores (LigScore1_Dreiding, LigScore2_Dreiding, -PLP1, -PLP2, Jain, -PMF, DOCK_SCORE) for the JAK2 inhibitors and inactive compounds can distinguish them in different extent. Accordingly, the differentiating capacity of Ligscore1_Dreiding is the most promising one, owing to the minimal overlaying area of the JAK2 inhibitors and inactive compounds. Furthermore, discrimination ability of these docking scores (LigScore1_Dreiding, LigScore2_Dreiding, -PLP1, -PLP2, Jain, -PMF, DOCK_SCORE) were examined by ROC curve studies (Fig. 2(h)), in similar, the Ligscore1_Dreiding is the best, with a preferable AUC (area under ROC curve) value of 0.964, when comparing with that of Ligscore2_Dreiding (0.859), -PLP1 (0.924), -PLP2 (0.901), Jain (0.924), -PMF (0.940), DOCK_SCORE (0.924).

Pharmacophore model generation and validation

The top 10 hypotheses describing the activity of JAK2 inhibitors in training set were generated by the HypoGen algorithm. The pharmacophore composition and parameters of significance, such as cost value, correlation coefficient (r), and root mean square (RMS) deviations, can be found in the ESI Table S1.† The fixed cost (59.69 bits) of the ten best hypotheses was well separated from the null hypothesis cost (146.29 bits), indicating an over 90% statistical significance of the resultant model. The configuration cost of 15.02 bits represents the acceptable complexity of the hypotheses space. The top-ranked pharmacophore model (Hypo 1) consists of two hydrogen-bond donor (HBD), two hydrophobic groups (HY) and four excluded volumes (EV), and the geometric parameters is shown in Fig. 3(a). The developed Hypo 1 had the optimal predictive capability and statistical significance, indicated by the high correlation coefficient (r = 0.94), low RMS deviation (1.25) and cost deviation of 73.08. As shown in Fig. 3(b), the most active compound 1 (IC50 = 0.0007 μM, see Table 2) can perfectly mapped to Hypo 1.
To validate the predictive quality of Hypo 1, we used it to estimate the activities of training set molecules. The experimental IC50 values along with predicted ones of the training set compounds are reported in Table 1. The error ratio represents the ratio of the difference value between actual and predicted activity and the actual activity. The developed Hypo 1 hypothesis mapped very well to the training set molecules, indicating the high capability of Hypo 1 to correctly predict activities. To further verify the accuracy of Hypo 1, a test set of 9 structurally diverse JAK2 inhibitors (see Fig. S2 in the ESI†) were predicted by Hypo 1 pharmacophore. According to Table 2, all the test set molecules were predicted correctly. Thus, the validated Hypo 1 pharmacophore could be applied for virtual screening to discriminate the JAK2 inhibitors from inactive compounds.

**Sequential virtual screening using molecular docking and pharmacophore mapping**

All the pre-treated structures in chemical SPECS database were docked and scored using LigScore1_Dreiding, the top ranked 3000 compounds were selected for further study. Then, Hypo 1 was employed to further map the key pharmacophoric features onto the structures of initially screened compounds, compounds were picked up via globally and partially pharmacophoric features for both of ligand and receptor. Then, further...
visual inspection was performed with respect to chemical stability, toxicity and molecular diversity. For example, the compounds with a hydrolyzable group (e.g. ester moiety) as well as electrophilic warheads (e.g. hydrazones) were excluded due to the hepatotoxicity. Finally, twelve compounds A1–A12 (see Table S2 in the ESI†) with preferable docking scores, pharmacophore mapping fit value and well drug-like properties were purchased for further studies.

In vitro biological assay for JAK2 inhibitory activities

The screened compounds A1–A12 were further evaluated for their JAK2 inhibitory activities, to our delight, three compounds showed significantly JAK2 inhibitory activities (inhibitory rate of A5: 67.67%, A6: 28.00%, A9: 31.26%) at a concentration of 20 μg mL⁻¹ (Fig. 4). Among them, compound A5 bearing diaminoypyrimidine moiety showed the best JAK2 inhibitory activities with IC₅₀ value of 4.64 μg mL⁻¹. As a result, the hit rate in our sequential virtual screening process for identification of JAK2 inhibitors is 33.33%, showing significantly superior performance in the identification of novel JAK2 inhibitors when comparing with that of conventional high throughput screening (HTS) strategy. Moreover, A6 bearing aminothiazole moiety is

Table 2  Actual and estimated IC₅₀ (μM) for the test set based on the best pharmacophore hypothesis Hypo 1

| No. | Estimated IC₅₀ (μM) | Actual IC₅₀ (μM) | Error ratio | Ref. |
|-----|---------------------|------------------|-------------|-----|
| 16  | 0.0027              | 0.0030           | 0.1000      | 19  |
| 17  | 0.0195              | 0.0230           | 0.1522      | 19  |
| 18  | 0.0427              | 0.1500           | 0.7153      | 25  |
| 19  | 0.0400              | 0.0440           | 0.0909      | 24  |
| 20  | 0.1074              | 0.0720           | 0.4917      | 22  |
| 21  | 4.0691              | 3.9400           | 0.0328      | 22  |
| 22  | 0.1123              | 0.3220           | 0.6512      | 21  |
| 23  | 11.4614             | 16.6000          | 0.3096      | 26  |
| 24  | 22.0345             | 16.9000          | −0.3038     | 27  |

Fig. 3  The best pharmacophore model (Hypo 1) generated by the 3D-QSAR pharmacophore module: (a) features at a specific distance. (Features are color-coded as follows: blue, hydrophobic; magenta, hydrogen-bond donor; and dark gray, excluded volume.); (b) mapping of the most active inhibitor compound 1 (ball and stick) onto the Hypo 1 hypothesis.

Fig. 4  Comparison of JAK2 inhibitory abilities of hits (A5, A6, A7, A9, A12, B2 and B4) and staurosporine at the concentration of 20 μg mL⁻¹.
significantly different from the reported JAK2 inhibitors in literatures, which is very interesting for us in further development of novel JAK2 inhibitors.

**Similarity search for other structurally analogical JAK2 inhibitors**

To detect analogues of the sequential VS hits, similarity search using the core structures of potent hits (A5, A6 and A9) as queries through the drug-like dataset was performed. As a result, 158 analogues were enriched and further evaluated in combination of molecular docking and pharmacophore mapping, finally, five compounds B1–B5 (see Table S2 in the ESI†) were identified and subjected to the JAK2 inhibition test. To be of interest, two compounds showed moderate JAK2 inhibitory activities (inhibitory rate of B2: 63.18%, B4: 26.40%) at a concentration of 20 μg mL⁻¹. To be of interest, we identified another compound B2 with the promising skeleton (aminothiazole), which exhibited potent JAK2 inhibitory activity (IC₅₀: 22.86 μg mL⁻¹) (Fig. 4).

**Selectivity profiles of JAK kinase inhibitory activities**

Then, two compounds A5 and B2 were sent to analyze their selectivity profiles of JAK kinase inhibitory activities, including

| Compd. | IC₅₀ (μg mL⁻¹) | JAK1 | JAK2 | JAK3 |
|--------|----------------|------|------|------|
| A5     | 1.18           | 4.64 | 0.293|      |
| B2     | >100           | 22.86| >100 | >100 |

Fig. 5 Anti-proliferative activities of compound B2 against cancer cell H1975, PANC-1 and PC-9.

Fig. 6 MD simulations and interaction mechanism analysis for the most potent inhibitors: (a) and (b) plot the distances of H-bond between Leu932 and A5 and B2 during 200 ps of MD simulation; (c) and (d) plot the binding pattern of A5 and B2; (e) and (f) plot the inhibitor–residue interaction spectrum for A5 and B2, respectively.
JAK1, JAK2, and JAK3 (Table 3). To be of interest, compound A5 also shows good inhibitory activities against JAK1 and JAK3 with IC\textsubscript{50} value of 1.18 and 0.293 \mu M\textsuperscript{-1}, respectively, which were even better than that of JAK2 (4.64 \mu M\textsuperscript{-1}). Thus, compound A5 would be a promising lead compound in structural optimization as a novel selective JAK3 inhibitor. In addition, compound B2 showed more selective inhibitory activity against JAK2, which showed no activity against JAK1 and JAK3 at even very high concentration (100 \mu M\textsuperscript{-1}).

In vitro antiproliferative assay against cancer cells

Considering the novelty of the skeleton and the selective profiles (JAK2) of compound B2, it was further biologically evaluated for its further anticancer activities, involving with antiproliferative effect against acquired gefitinib-resistant human lung cancer cell (H1975), human pancreatic carcinoma (PANC-1) and human lung cancer cell line (PC-9). As shown in Fig. 5, compound B2 exhibited significantly antiproliferative activity against all of these cancer cell lines, with IC\textsubscript{50} value of 18.1, 58.3, 40.6 \mu M\textsuperscript{-1} against PC-9, H1975 and PANC-1, respectively.

Insight into the binding of A5 and B2 to JAK2

In order to investigate the interaction mode between the preferable compounds A5 and B2 and JAK2 protein, molecular dynamics (MD) simulations and interaction decomposition analysis were carried out to explore the quantitative energy contributions per-residue to the binding affinity of two promising inhibitors (A5 and B2). After 200 ps production simulation, distance between inhibitors (A5 and B2) and key amino acid Leu932 of JAK2 trends to converge, indicating that the system reached equilibrium conditions (see Fig. 6(a) and (b)). Based on the stable conformation obtained from MD simulation, interactions involved in the protein/ligand complex were analyzed. As shown in Fig. 6(c) and (d), the =NH and =N moieties of aminopyrimidine (A5) aminothiazole (B2) core could form two H-bonds with hinge residue Leu932 of ATP-binding pocket, which accounts for a significant proportion of their potency. In accordance with the interaction energy decomposition, the contribution of Leu932 for A5 and B2 is -8.33 and -6.89 kcal mol\textsuperscript{-1}, respectively (see Fig. 6(e) and (f)).

Besides, the NH\textsubscript{2} bearing on the right 2-methylquinolin moiety of A5 also forms H-bond with the Tyr931, while the NH\textsubscript{2} group bearing on the thiazole core of B2 forms an additional H-bond with Glu930. The C5-methyl of pyrimidine core (A5) and the C5-amino of thiazole skeleton (B2) were stacked against the gatekeeper residue Met929. For the selective JAK2 inhibitor B2, an intramolecular hydrogen bond was formed, holding the chlorothiophene substituent coplanar with the aminothiazole core. Thus, the chlorothiophene moiety was found to be located in the binding pocket adjacent to Val863 and Leu983, and extended towards the Asp994 of activation loop and the Gly993 of glycine-rich loop. This orientation may confer to the high selectivity of B2 over the other JAK family kinases. Therefore, according to the binding pattern and chemical structure of A5 and B2, we speculate that an aromatic scaffold containing hydrogen-donors and receptors, and a substituent extending towards the activation loop and glycine-rich loop may be an important pharmacophore feature of JAK2 inhibitors.

Conclusions

Herein, a reasonable and reliable sequential VS protocol by integrating the prediction results of structure-based molecular docking and ligand-based pharmacophore model was successfully constructed to identify potent JAK2 inhibitors. With the sequential virtual screen workflow in hand, twelve novel and structurally diverse hits were obtained and subjected to in vitro biological test, and three compounds (A5, A6 and A9) with potent JAK2 inhibitory activity were discovered. In pursuit of more JAK2 inhibitors efficiently, the obtained structures were further used as the template for subsequent similarity search, leading to the identification of another two promising compounds (B2 and B4). Moreover, selectivity profiles of two compounds A5 and B2 against JAK kinase was studied, revealing that these compounds bearing two different skeletons show different JAK subtype inhibitory activities. The compound B2 with a novel aminothiazole skeleton and better JAK2 inhibitory selectivity was further demonstrated for its anti-proliferative effect against cancer cells with micromolar IC\textsubscript{50}, suggesting that it would be a good lead to do further structure-activity research to find more high potent JAK2 selective inhibitors. Finally, binding patterns of the two most potent JAK2 inhibitors (A5 and B2) were explored by molecular docking, MD simulations, and interaction energy decomposition analysis. The novel intramolecular hydrogen bond of B2 and its favourable orientation towards the activation loop and glycine-rich loop of JAK2 may confer to its high specificity over the other JAK family kinases, providing computational clue for structurally optimization of the lead compounds.

Acknowledgements

This study was financially supported by Natural Science Foundation of Zhejiang Province, China (LY15C140002, LY13H300002) and Science and Technology Planning Project of Zhejiang Province (2015C32061).

Notes and references

1 K. Yamaoka, P. Saharinen, M. Pesu, V. E. T. Holt, O. Silvennoinen and J. J. O’Shea, *Genome Biol.*, 2004, 5, 253.
2 L. L. C. Marotta, V. Almendro, A. Marušk, M. Shipitsin, J. Schemme, S. R. Walker, N. Bloushtain-Qimron, J. J. Kim, S. A. Choudhary, R. Maruyma, Z. Wu, M. Gonen, L. A. Mulvey, M. O. Bessarabova, S. J. Huh, S. J. Silver, S. Y. Kim, S. Y. Park, H. E. Lee, K. S. Anderson, A. L. Richardson, T. Nikolskaya, Y. Nikolsky, S. X. Liu, D. E. Root, W. C. Hahn, D. A. Frank and K. Polyak, *J. Clin. Invest.*, 2011, 121, 2723–2735.
3 R. L. Levine, M. Wadleigh, J. Cools, B. L. Ebert, G. Wernig, B. J. P. Huntly, T. J. Boggon, L. Wlodarska, J. J. Clark, S. Moore, J. Adelsperger, S. Koo, J. C. Lee, S. Gabriel,
