Comparative Molecular Field Analysis of Pyrrolopyrimidines as LRRK2 Kinase Inhibitors

Anand Balupuri[1†], Pavithra K. Balasubramanian[1†], and Seung Joo Cho[1,2†]

Abstract

Leucine rich repeat kinase 2 (LRRK2) is a highly promising target for Parkinson's disease (PD) that affects millions of people worldwide. A three-dimensional quantitative structure-activity relationship (3D-QSAR) analysis was performed on a series of pyrrolopyrimidine-based selective LRRK2 kinase inhibitors. This study was performed to rationalize the structural requirements responsible for the inhibitory activity of these compounds. A reliable 3D-QSAR model was developed using comparative molecular field analysis (CoMFA) technique. The model produced statistically acceptable results with a cross-validated correlation coefficient (q^2) of 0.539 and a non-cross-validated correlation coefficient (r^2) of 0.871. Robustness of the model was further evaluated by bootstrapping and progressive scrambling analysis. This work could assist in designing more potent LRRK2 inhibitors.

Keywords: Parkinson's Disease, LRRK2 Kinase, Pyrrolopyrimidines, 3D-QSAR, CoMFA.

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects millions of people worldwide. Symptoms include muscle rigidity, tremors, and changes in speech and gait. The cause and underlying disease mechanisms are not well understood[1]. Recent genome-wide association studies (GWAS) studies have identified leucine rich repeat kinase 2 (LRRK2) as a highly promising target for PD[2,3]. Several genetic variants in LRRK2 have been identified as having an increased PD risk, indicating that it is important in the cause and pathogenesis of PD. LRRK2 dysfunction/ dysregulation is involved in the development of PD[4,5]. Due to potential of disease modification, LRRK2 has attracted the attention of pharmaceutical industry.

LRRK2 is a serine/threonine kinase and shares sequence homology with leucine rich repeat kinase 1 (LRRK1) and receptor-interacting protein (RIP) kinases[4,6]. The chronic nature of PD and aging patient population, require highly selective inhibitors for excellent safety profile and successful therapy. Recently, a series of pyrrolopyrimidines has been reported as highly selective LRRK2 kinase inhibitors[7]. However, a three-dimensional quantitative structure-activity relationship (3D-QSAR) analysis was not performed on these inhibitors to determine the relation between chemical structures and the inhibitory values. Our research group is involved in molecular modeling studies[8-12]. Here, we have carried out comparative molecular field analysis (CoMFA) to identify the key structural elements that are required in the rational design of novel LRRK2 kinase inhibitors.

2. Methodology

2.1. Data Set

A data set of 37 pyrrolopyrimidines possessing LRRK2 kinase inhibitory activity was collected[7]. Activity values were reported as IC_{50} values. These inhibitory values were converted into pIC_{50} values for 3D-QSAR analysis. Activity (pIC_{50}) values were employed as dependent variables for deriving CoMFA model. The extracted co-crystallized ligand (compound **8**) was used as template to construct and align the 3D structures of other data set compounds. All structures were sketched using SKETCH function of SYBYL-X2.0[13].

[1]Department of Biomedical Sciences, College of Medicine, Chosun University, Gwangju 501-759, Republic of Korea
[2]Department of Cellular and Molecular Medicine, College of Medicine, Chosun University, Gwangju 501-759, Republic of Korea

†Corresponding author : anandbalupuri.niper@gmail.com, pavithrabioinfo@gmail.com, chosj@chosun.ac.kr

Table 1. Chemical structures and biological activities of pyrrolopyrimidine-based LRRK2 kinase inhibitors
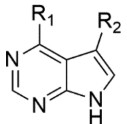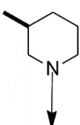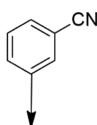
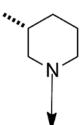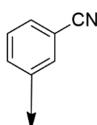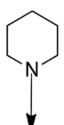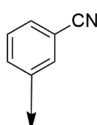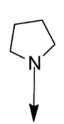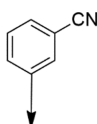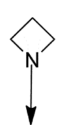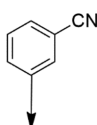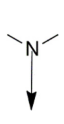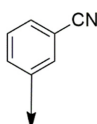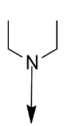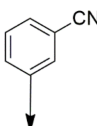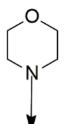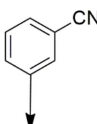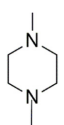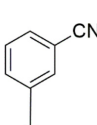| Compound	R_1	R_2	IC$_{50}$ (nM)	pIC$_{50}$
1			16	7.796
2			69	7.161
3			6	8.222
4			8	8.097
5*			115	6.939
6			5	8.301
7			20	7.699
8			3	8.523
9			1830	5.738

Table 1. Continued
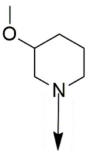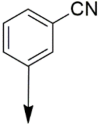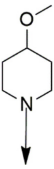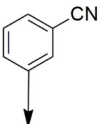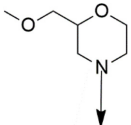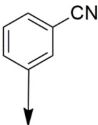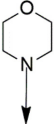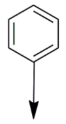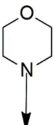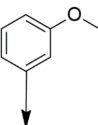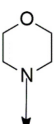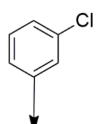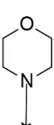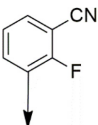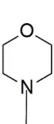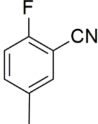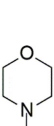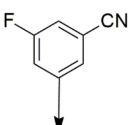
Compound	R_1	R_2	IC_{50} (nM)	pIC_{50}
10			79	7.102
11			393	6.406
12			95	7.022
13			9	8.046
14			28	7.553
15			8	8.097
16			3	8.523
17*			117	6.932
18			80	7.097

Table 1. Continued
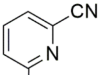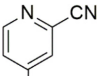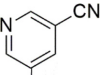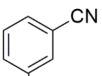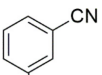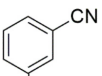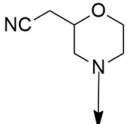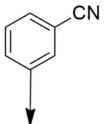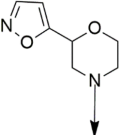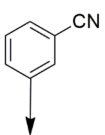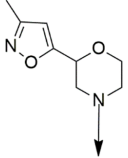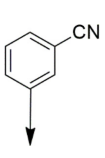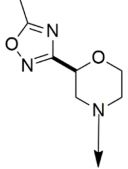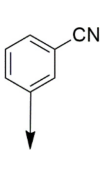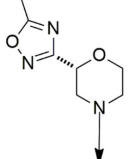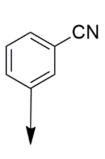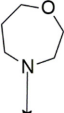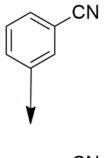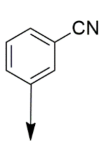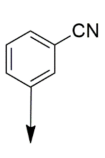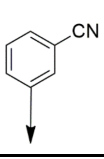
Compound	R_1	R_2	IC_{50} (nM)	pIC_{50}
19			9	8.046
20*			686	6.164
21*			2243	5.649
22			12	7.921
23			53	7.276
24			18	7.745
25*			914	6.039
26			18	7.745
27			7	8.155
28			42	7.377

Table 1. Continued

Compound	R$_1$	R$_2$	IC$_{50}$ (nM)	pIC$_{50}$
29			5	8.301
30			37	7.432
31			20	7.699
32			8	8.097
33			204	6.690
34			111	6.955
35			2331	5.633
36			139	6.857
37			222	6.654

*Compounds are considered as outliers.

Partial atomic charges were calculated by the Gasteiger-Hückel method. Energy minimizations were performed using the Tripos force field with a distance-dependent dielectric and the Powell conjugate gradient algorithm. The minimized structures were aligned to the template compound using common substructure-based alignment method. Structures and activities of the data set compounds are listed in Table 1.

2.2. CoMFA

CoMFA technique is based on the hypothesis that changes in the inhibitory activity of compounds are related to the variations in the steric and electrostatic fields[14]. The Lennard-Jones potential terms and Coulombic terms represent steric and electrostatic fields respectively. SYBYL-X2.0 was used for CoMFA calculations. Potential fields for all compounds were determined at each lattice intersection of a regularly spaced grid of 2.0 Å. An sp^3 hybridized carbon probe atom carrying +1 charge and van der Waals radius of 1.52 Å was used for the calculation of interaction fields. Energy cutoff value of 30 kcal mol^{-1} was selected for both steric and electrostatic fields.

Partial least squares (PLS) regression algorithm was employed for structural parameters and inhibitory activity values[15]. Activity (pIC_{50}) values were used as dependent variables whereas CoMFA descriptors were used as independent variables in the PLS analysis. The leave-one-out (LOO) cross-validation was performed to obtain cross-validated correlation coefficient (q^2), optimal number of components (NOC) and standard error of prediction (SEP). Non-cross-validated correlation coefficient (r^2), standard error of estimate (SEE) and F-test value (F) were obtained by non-cross-validated analysis. CoMFA model was further validated by bootstrapping analysis[16] and progressive scrambling. Bootstrapping analysis was carried out for 1000 runs while 100 independent scramblings were performed with a maximum of 10 bins and a minimum of 2 bins.

3. Results and Discussion

3.1. CoMFA Model

A series of pyrrolopyrimidine-based LRRK2 kinase inhibitors was used to develop a 3D-QSAR model. Whole data set was employed for developing CoMFA model. All compounds were aligned over the template
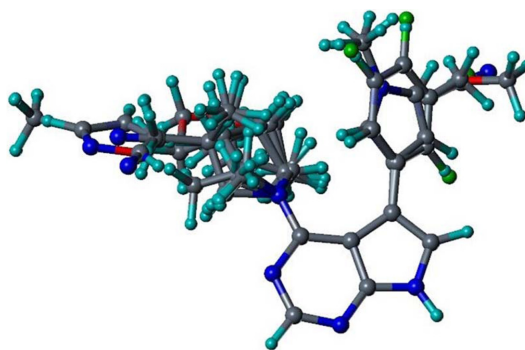
Fig. 1. Common substructure-based alignment of data set compounds using compound **8** as a template.

Table 2. Statistical parameters of the CoMFA model

Parameters	CoMFA
q^2	0.539
NOC	4
SEP	0.561
r^2	0.871
SEE	0.297
F	40.371
BS-r^2	0.919
BS-sd	0.039
Q^2	0.403
Steric contribution	58.4
Electrostatic contribution	41.6

Note: q^2 is cross-validated correlation coefficient, NOC is number of components, SEP is standard error of prediction, r^2 is non-cross-validated correlation coefficient, SEE is standard error of estimation; F is F-test value, BS-r^2 is bootstrapping r^2 mean, BS-SD is bootstrapping standard deviation, Q^2 is corrected q^2 dependency.

(compound **8**) using common substructure alignment method. Alignment of data set compounds are shown in Fig. 1. Data set was not split into training and test sets during model generation due to less number of compounds. During the development of CoMFA model, five compounds (**5**, **17**, **20**, **21**, and **25**) were removed from data set as outliers based on the high residual values.

A statistically acceptable CoMFA model was developed. Statistical values for the model are listed in Table 2. Model showed a q^2 value of 0.539 with 4 components. The non-cross-validated analysis generated r^2, SEE and F values of 0.871, 0.297 and 40.371, respec-

Table 3. Actual and predicted activity values with residuals of the data set compounds

Compound	Actual pIC$_{50}$	CoMFA	
		Predicted pIC$_{50}$	Residual
1	7.796	7.245	0.551
2	7.161	7.106	0.055
3	8.222	7.371	0.851
4	8.097	8.091	0.006
6	8.301	8.482	-0.181
7	7.699	8.211	-0.513
8	8.523	7.995	0.528
9	5.738	6.169	-0.432
10	7.102	6.973	0.129
11	6.406	6.144	0.262
12	7.022	6.984	0.038
13	8.046	7.927	0.119
14	7.553	7.592	-0.040
15	8.097	7.918	0.179
16	8.523	8.152	0.371
18	7.097	7.774	-0.677
19	8.046	8.280	-0.234
22	7.921	8.081	-0.161
23	7.276	7.939	-0.663
24	7.745	7.506	0.239
26	7.745	7.870	-0.125
27	8.155	7.708	0.447
28	7.377	7.622	-0.246
29	8.301	8.067	0.234
30	7.432	7.526	-0.094
31	7.699	7.438	0.261
32	8.097	8.096	0.001
33	6.690	6.739	-0.048
34	6.955	7.167	-0.213
35	5.633	5.743	-0.111
36	6.857	6.732	0.125
37	6.654	6.574	0.080

tively. The steric and electrostatic contributions were 58.4% and 41.6%, respectively. Actual and predicted activity values along with the residual values for data set compounds are given in Table 3. The scatter plot for actual versus predicted activity values is displayed in Fig. 2. Predicted activities are in accordance with the experimental values indicating that a reliable CoMFA model was developed.

3.2. CoMFA Contour Maps
One of the attractive features of 3D-QSAR modeling

is the visualization of information content of the derived models by contour maps. These maps indicate the regions in 3D space around the compounds where variations in the fields are predicted to either enhance or reduce the activity. The contour maps of different fields are shown with the template molecule (compound **8**) in Fig. 3 and 4.

Steric contour map is exhibited in Fig. 3. Green contours represent favorable regions while yellow contours represent unfavorable regions for the substitution of bulky groups. Green contours observed near R$_1$ substi-

Fig. 2. Scatter plot of the actual versus predicted activities based on the CoMFA model.
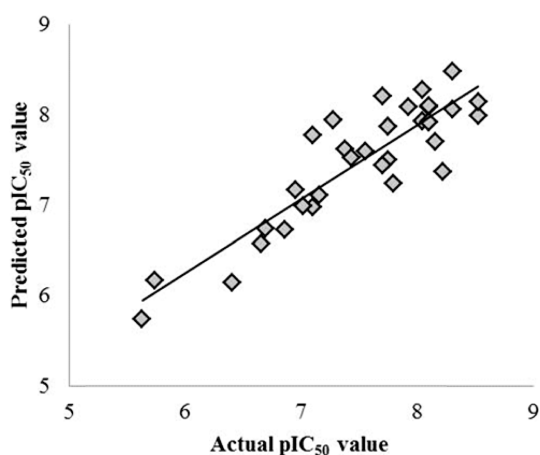

Fig. 4. CoMFA electrostatic contour map with template (compound **8**) as a reference. Blue contours represent favorable regions for electropositive substituents.
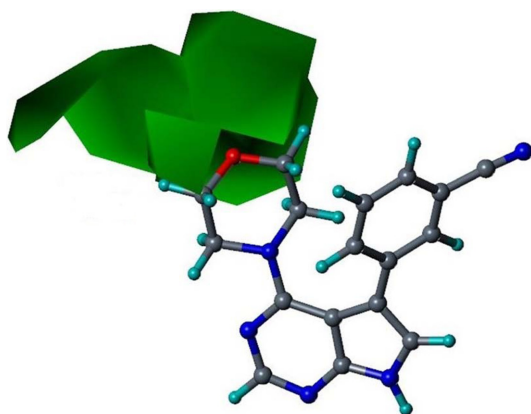

Fig. 3. CoMFA steric contour map with template (compound **8**) as a reference. Green contours indicate sterically favored regions.

tution indicated that bulky groups in this region are favorable for improving the inhibitory activity. This could be the possible reason for better inhibitory activity of compound **8** as compared to compound **5**. Due to the same possible reason, compound **32** possess higher activity than compound **26**.

Electrostatic contour map is exhibited in Fig. 4. Blue contours represent regions where electropositive substitutions are favored while red contours represent favorable regions for electronegative substitutions to improve the activity. A large blue contour observed near R_1 substitution indicated that electropositive groups at that position could enhance the activity. This might be the

reason for higher activity of compound **27** than compound **28**. Also, compound **1** demonstrates better inhibitory activity as compared to compounds **11** and **12** because of the same possible reason.

4. Conclusions

In this work, a CoMFA model was developed for a series of pyrrolopyrimidine-based LRRK2 kinase inhibitors. Model produced statistically reliable results in terms of q^2 and r^2 values. Robustness of model was further validated by bootstrapping and progressive sampling analyses. Analysis of contour maps suggested regions for structural modification to enhance the activity of compounds. Bulky groups with electropositive properties are desirable at R_1 substitution for improving the inhibitory potency. The information provided by the contour maps could be utilized to design more potent LRRK2 kinase inhibitors.

Acknowledgements

References

[1] K.-L. Lim and C.-W. Zhang, "Molecular events underlying Parkinson's disease - an interwoven tap-

estry", Frontiers in Neurology, Vol. 4, 2013.

[2] C. Labbé and O. A. Ross, "Association studies of sporadic Parkinson's disease in the genomic era", Curr. Genomics, Vol. 15, pp. 2-10, 2014.

[3] S. Lubbe and H. R. Morris, "Recent advances in Parkinson's disease genetics", J. Neurol., Vol. 261, pp. 259-266, 2014.

[4] I. Marin, "The Parkinson disease gene LRRK2: evolutionary and structural insights", Mol. Biol. Evol., Vol. 23, pp 2423-2433, 2006.

[5] A. R. Esteves, R. H. Swedlow, and S. M. Cardoso, "LRRK2, a puzzling protein: insights into Parkinson's disease pathogenesis", Exp. Neurol., Vol. 261, pp. 206-216, 2014.

[6] T. T. Wager, X. Hou, P. R. Verhoest, and A. Villalobos, "Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties", ACS Chem. Neurosci., Vol. 1, pp. 435-449, 2010.

[7] J. L. Henderson, B. L. Kormos, M. M. Hayward, K. J. Coffman, J. Jasti, R. G. Kurumbail, T. T. Wager, P. R. Verhoest, G. S. Noell, Y. Chen, E. Needle, Z. Berger, S. J. Steyn, C. Houle, W. D. Hirst, and P. Galatsis, "Discovery and preclinical profiling of 3-[4-(morpholin-4-yl)-7H-pyrrolo[2,3-d]pyrimidin-5-yl]benzonitrile (PF-06447475), a highly potent, selective, brain penetrant, and in vivo active LRRK2 kinase inhibitor", J. Med. Chem., Vol. 58, pp 419-432, 2015.

[8] A. Balupuri and S. J. Cho, "Exploration of the binding mode of indole derivatives as potent HIV-1 inhibitors using molecular docking simulations", J. Chosun Natural Sci., Vol. 6, pp. 138-142, 2013.

[9] A. Balupuri, P. K. Balasubramanian, and S. J. Cho, "A CoMFA study of glycogen synthase kinase 3 inhibitors", J. Chosun Natural Sci., Vol. 8, pp. 40-47, 2015.

[10] A. Balupuri, P. K. Balasubramanian, and S. J. Cho, "A CoMFA study of quinazoline-based anticancer agents", J. Chosun Natural Sci., Vol. 8, pp. 214-220, 2015.

[11] P. K. Balasubramanian, A. Balupuri, and S. J. Cho, "A CoMFA study of phenoxypyridine-based JNK3 inhibitors using various partial charge schemes", J. Chosun Natural Sci., Vol. 7, pp. 45-49, 2014.

[12] P. K. Balasubramanian, A. Balupuri, and S. J. Cho, "Ligand-based CoMFA study on pyridylpyrazolopyridine derivatives as PKCθ kinase inhibitors", J. Chosun Natural Sci., Vol. 7, pp. 253-259, 2014.

[13] SYBYLx2.1, Tripos International, 1699 South Hanley Road, St. Louis, Missouri, 63144, USA.

[14] R. D. Cramer, D. E. Patterson, and J. D. Bunce, "Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins", J. Am. Chem. Soc., Vol. 110, pp. 5959-5967, 1988.

[15] S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, III, "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses", SIAM Journal on Scientific and Statistical Computing, Vol. 5, pp 735-743, 1984.

[16] R. D. Cramer, J. D. Bunce, D. E. Patterson, and I. E. Frank, "Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies", Quantitative Structure-Activity Relationships, Vol. 7, pp. 18-25, 1988.