Pharmacophore Based Comparative Molecular Field Analysis of CRTh2 Antagonists

Sathya Babu and Thirumurthy Madhavan†

Abstract

Chemoattractant receptor homologous molecule expressed on Th2 cells (CRTh2) is a G-protein coupled receptor targeted for inflammatory diseases such as asthma, allergic rhinitis and atopic dermatitis. In this study, pharmacophore modeling and comparative molecular field analysis (CoMFA) were performed on the series of 2-(2-(benzylthio)-1H-benzo[d]imidazol-1-yl) acetic acids derivatives. Five highly active compounds were used for generation of pharmacophore models using GASP module. The best pharmacophore model was selected and used as template for the alignment of compounds which was used for CoMFA analysis. The best predictions obtained for CoMFA was q² = 0.545, r² = 0.756. The predictive ability of the model was investigated using 15 test set compounds. Contour maps suggested that presence of bulky substituents at 5ᵗʰ position of benzene ring connected to sulphur atoms attached to imidazol ring will increase the activity of the compounds. The results obtained from this study will be useful to design more potent CRTh2 antagonists.

Keywords: CRTh2, Pharmacophore, CoMFA

1. Introduction

Chemoattractant receptor-homologous molecule expressed on Th2 cells (CRTh2) is a chemoattractant receptor with seven transmembrane helices and it belongs to the G-Protein Coupled receptor (GPCR) family[1-5] which plays an important role in inflammatory diseases such as asthma, allergic rhinitis and atopic dermatitis[6-8]. In humans CRTh2 is highly expressed on the cells involved in the pathology of asthma and other allergic disorders including eosinophils, basophils and subset of Th2 lymphocytes[9,10]. CRTh2 mediates the chemoattractant effect of PGD2 in leukocytes where it is present and can also mediate intracellular Ca²⁺ mobilization in response to a factor released from activated mast cells, suggesting that CRTh2 may be closely involved in mast cell–mediated allergic inflammation. Further data in animal models have highlighted the role of CRTh2 in mast cell dependent activation of Th2 cells in chronic allergic skin inflammation and eosinophilic airway inflammation[12,13]. Hence, CRTh2 is as attractive target and antagonizing it could be useful for the treatment of inflammatory diseases.

The rapid increase in three-dimensional structural information has led to the development of 3D QSAR methods. The aim of this study was to identify the important features required for improving the activity of CRTh2 antagonist by 3D-QSAR approach CoMFA. Pharmacophore models were generated using GASP module and best model was used as template for alignment. The series of 65 benzimidazol derivatives was used for CoMFA where 50 compounds was used as training set and 15 compounds was used as test set. Training set was used to build model whereas test set was used to find the predictive ability of the model. The predictive ability of CoMFA (r²(pred)=0.693) were obtained which shows good correlation between experimental and predicted pIC₅₀ values. We hope that our models would be useful for future development of CRTh2 inhibitor molecules.

2. Experimental Section

2.1. Dataset

The structure of 2-(2-(benzylthio)-1H-benzo[d]imidazol-1-yl) acetic acids derivatives and their biological
Table 1. Structures and biological activities (pIC$_{50}$) of CRTh2 inhibitors

| Comp No | Structure | pIC$_{50}$ value | Comp No | Structure | pIC$_{50}$ value |
|---------|-----------|------------------|---------|-----------|------------------|
| 1       | ![Structure 1](image1) | 6.229            | 34      | ![Structure 34](image34) | 6.328            |
| 2       | ![Structure 2](image2) | 5.602            | 35      | ![Structure 35](image35) | 5.921            |
| 3       | ![Structure 3](image3) | 5.699            | 36      | ![Structure 36](image36) | 6.131            |
| 4       | ![Structure 4](image4) | 6.105            | 37      | ![Structure 37](image37) | 5.796            |
| 5       | ![Structure 5](image5) | 5.854            | 38      | ![Structure 38](image38) | 7.538            |
| 6       | ![Structure 6](image6) | 6.323            | 39      | ![Structure 39](image39) | 5.538            |
| 7       | ![Structure 7](image7) | 6.055            | 40      | ![Structure 40](image40) | 6.009            |
| 8       | ![Structure 8](image8) | 6.411            | 41      | ![Structure 41](image41) | 6.482            |
| 9       | ![Structure 9](image9) | 5.570            | 42      | ![Structure 42](image42) | 7.347            |
| Comp No | Structure | pIC<sub>50</sub> value | Comp No | Structure | pIC<sub>50</sub> value |
|---------|-----------|-----------------------|---------|-----------|-----------------------|
| 10      | ![Structure Image] | 5.114                 | 43      | ![Structure Image] | 6.824                 |
| 11      | ![Structure Image] | 5.867                 | 44      | ![Structure Image] | 8.097                 |
| 12      | ![Structure Image] | 5.907                 | 45      | ![Structure Image] | 6.638                 |
| 13      | ![Structure Image] | 5.625                 | 46      | ![Structure Image] | 5.657                 |
| 14      | ![Structure Image] | 4.943                 | 47      | ![Structure Image] | 8.301                 |
| 15      | ![Structure Image] | 5.356                 | 48      | ![Structure Image] | 8.699                 |
| 16      | ![Structure Image] | 5.627                 | 49      | ![Structure Image] | 7.347                 |
| 17      | ![Structure Image] | 5.996                 | 50      | ![Structure Image] | 8.699                 |
| 18      | ![Structure Image] | 6.796                 | 51      | ![Structure Image] | 8.523                 |
Table 1. Continued

| Comp No | Structure | pIC<sub>50</sub> value | Comp No | Structure | pIC<sub>50</sub> value |
|---------|-----------|-----------------------|---------|-----------|-----------------------|
| 19      | ![Structure Image](image1) | 6.731 | 52      | ![Structure Image](image2) | 7.387 |
| 20      | ![Structure Image](image3) | 6.432 | 53      | ![Structure Image](image4) | 7.482 |
| 21      | ![Structure Image](image5) | 6.678 | 54      | ![Structure Image](image6) | 8.398 |
| 22      | ![Structure Image](image7) | 7.143 | 55      | ![Structure Image](image8) | 8.398 |
| 23      | ![Structure Image](image9) | 5.959 | 56      | ![Structure Image](image10) | 8.301 |
| 24      | ![Structure Image](image11) | 6.658 | 57      | ![Structure Image](image12) | 8.046 |
| 25      | ![Structure Image](image13) | 6.979 | 58      | ![Structure Image](image14) | 8.770 |
| 26      | ![Structure Image](image15) | 6.092 | 59      | ![Structure Image](image16) | 7.796 |
| 27      | ![Structure Image](image17) | 6.658 | 60      | ![Structure Image](image18) | 7.796 |
activities of 65 compounds were taken from the literature\textsuperscript{[1]}. The IC\textsubscript{50} values of the compounds were converted into pIC\textsubscript{50} (-log IC\textsubscript{50}) in order to use the data as a dependent variable in CoMFA. The dataset is randomly divided into training set of 50 molecules and test set of 15 molecules. The test set molecules comprises of low, middle and high biological activities from the dataset. The structure and biological activities of all compounds are represented in Table 1.

2.2. Pharmacophore Generation
Pharmacophore models were generated using the GASP (Genetic Algorithm Similarity Program)\textsuperscript{[10]} module of SYBYL-X2.0\textsuperscript{[11]}. GASP uses a genetic algorithm developed for the superimposition of set of flexible molecules to align the set of input molecules and to identify the common pharmacophore between them. GASP does not optimize bond length or bond angle, but it can adjust the torsion angles while searching for alignment. GASP uses two unique features to find pharmacophore alignment. Initially, a genetic algorithm is implemented to derive better models. Then, a unique fitness function is implemented that accounts for the protein–ligand interactions of different donors and

| Comp No | Structure | pIC\textsubscript{50} value | Comp No | Structure | pIC\textsubscript{50} value |
|---------|-----------|----------------------------|---------|-----------|----------------------------|
| 28      | ![Image](https://example.com/image1.png) | 6.432                       | 61      | ![Image](https://example.com/image2.png) | 7.229                       |
| 29      | ![Image](https://example.com/image3.png) | 6.237                       | 62      | ![Image](https://example.com/image4.png) | 7.143                       |
| 30      | ![Image](https://example.com/image5.png) | 4.975                       | 63      | ![Image](https://example.com/image6.png) | 6.305                       |
| 31      | ![Image](https://example.com/image7.png) | 6.347                       | 64      | ![Image](https://example.com/image8.png) | 6.971                       |
| 32      | ![Image](https://example.com/image9.png) | 6.284                       | 65      | ![Image](https://example.com/image10.png) | 8.523                       |
| 33      | ![Image](https://example.com/image11.png) | 6.420                       |         |           |                            |
acceptors. In this study, a set of five potent compounds (comp48, comp50, comp51, comp58, and comp65) was used to develop pharmacophore models using default parameters. A model with the highest fitness score was selected for molecular alignment, and the alignment obtained was subsequently used for CoMFA studies.

2.3. CoMFA Model Calculation

CoMFA has been used widely to relate the structure to their chemical and biological properties. The steric and electrostatic potential fields were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å using Tripos force field. The steric and electrostatic interaction energies were calculated for each molecule at each grid point using a sp$^3$ carbon probe with van der Waals radius of 1.52 Å and +1 charge at each lattice point. The generated steric and electrostatic fields were scaled by CoMFA standard scaling method with default energy cut off value of 30 kcal/mol. With standard option for the scaling of variables, the regression analysis was carried out using cross-validated partial least squares approach (PLS)\cite{12} of LOO (leave-one-out)\cite{13}. After the optimal number of components was determined, a non-cross-validated analysis was carried out without column filtering.

2.4. Partial Least Squares (PLS) Analysis

In 3D-QSAR, the CoMFA descriptors were used as independent variables and pIC$_{50}$ values were used as the dependent variable. The partial least squares (PLS) method was used to explore a linear correlation between the CoMFA fields and the biological activity values\cite{14}. CoMFA cutoff values were set to 30 kcal mol for both the steric and electrostatic fields, and all fields were scaled by the default options in SYBYL. The cross-validation correlation coefficient ($q^2$) that resulted in a minimal number of components and the lowest cross-validated standard error of estimate was considered for further analysis and calculates using the formula:

$$q^2 = 1 - \frac{\sum (\gamma_{\text{pred}} - \gamma_{\text{actual}})^2}{\sum (\gamma_{\text{actual}} - \gamma_{\text{mean}})^2}$$

where $\gamma_{\text{pred}}$, $\gamma_{\text{actual}}$, and $\gamma_{\text{mean}}$ are the predicted, actual, and mean values of the target property (pIC$_{50}$), respectively. CoMFA results were then graphically interpreted by field contribution maps.

The predictive power of 3D-QSAR models were derived from the test set molecules, which were excluded during model development. The activity of the test set was predicted by using model derived from training set. The predictive correlation coefficient $r^2_{\text{pred}}$ for developed model was determined by using following formula:

$$r^2_{\text{pred}} = \frac{(SD - PRESS)}{SD}$$

where, PRESS is the sum of the squared deviation between the predicted and actual activity of the test set molecules, and SD is defined as the sum of the square deviation between the biological activity of the test set compounds and the mean activity of the training set molecules.

3. Results and Discussion

3.1. Pharmacophore Generation

The pharmacophore model was generated using five highly active compounds (compound 48,50,51,58 and 65). A total of 100 models were generated with good fitness value and variable sizes. After inspection of pharmacophore models, an optimal model with high fitness score and good pharmacophore alignment was selected. The statistical values of the selected model was tabulated in Table 2. The model was associated with pharmacophore features consisting of 3 hydrophobes (HY1-HY3), 11 H-bond acceptors (AA1-AA11) and 11 Donor sites (DS1-DS11). The selected pharmacophore model represented in Fig. 1 was used to align rest of the compounds and it is subsequently used for ligand based CoMFA analysis.

| GASP statistical terms | Values |
|------------------------|--------|
| Fitness                | 6785.9900 |
| Size                   | 25     |
| Hits                   | 5      |
| Dmean                  | 7.2245 |

Table 2. Statistical values of the best pharmacophore model (Model 10) developed using GASP

Fitness: fitness score of the model, Size: number of features in the model, Hits: number of molecules that matched during the search, Dmean: average interpoint distance.
3.2. CoMFA Statistical Analysis

The pharmacophore model was used as a template and the entire molecules in the dataset were aligned over the selected pharmacophore and is shown in Fig. 2 and it was used for generation of various CoMFA models. The best CoMFA model was selected based on its statistical results. The pharmacophore based alignment has produced reasonable results for CoMFA with a $q^2$ of 0.545 and $r^2$ of 0.756. The statistical results of CoMFA model are shown in Table 3.

3.3. Validation of CoMFA Model

The generated model were validated using the test set of 15 molecules that were excluded during model development.
The predictive power of CoMFA model was 
$r^2_{\text{pred}} = 0.693$. The statistical results indicates the robustness of the developed model and is in good agreement between experimental and predicted pIC$_{50}$ values. The experimental and predicted pIC$_{50}$ values of the training and test set molecules was shown in Table 4. The plot of actual and predicted pIC$_{50}$ of training and test set is represented in Fig. 3a and 3b for CoMFA model.

Table 3. Statistical results of CoMFA model obtained from pharmacophore based alignment

| PLS statistics | CoMFA model |
|----------------|-------------|
| $q^2$          | 0.545       |
| $N$            | 3           |
| $r^2$          | 0.756       |
| SEE            | 0.545       |
| F-value        | 47.583      |
| $r^2_{\text{pred}}$ | 0.693       |

Field contribution

- Steric: 0.458
- Electrostatic: 0.542

$q^2$ = cross-validated correlation coefficient; $N$ = number of statistical components; $r^2$ = non-cross validated correlation coefficient; SEE = standard estimated error; $F$ = Fisher value; $r^2_{\text{pred}}$ = predictive correlation coefficient for test set.

Table 4. Predicted activities of CoMFA model compared with the experimental pIC$_{50}$ values

| Compound | Actual pIC$_{50}$ | Predicted pIC$_{50}$ | Residual |
|----------|-------------------|----------------------|----------|
| 1        | 6.229             | 5.930                | 0.299    |
| 2*       | 5.602             | 5.923                | -0.321   |
| 3        | 5.699             | 5.888                | -0.189   |
| 4        | 6.105             | 5.843                | 0.262    |
| 5        | 5.854             | 5.988                | -0.134   |
| 6        | 6.323             | 6.051                | 0.273    |
| 7        | 6.055             | 6.083                | -0.028   |
| 8*       | 6.411             | 6.562                | -0.151   |
| 9        | 5.570             | 5.942                | -0.372   |
| 10       | 5.114             | 5.785                | -0.671   |
| 11       | 5.867             | 5.967                | -0.100   |
| 12       | 5.907             | 5.957                | -0.051   |
| 13*      | 5.625             | 6.181                | -0.556   |
| 14       | 4.943             | 5.736                | -0.793   |
| 15       | 5.356             | 5.707                | -0.351   |
| 16*      | 5.627             | 5.771                | -0.144   |
| 17       | 5.996             | 6.030                | -0.034   |
| 18       | 6.796             | 6.500                | 0.295    |

* Test set compounds
Pharmacophore Based Comparative Molecular Field Analysis of CRTh2 Antagonists

3.4. CoMFA Contour Map Analysis

The results of CoMFA analysis shows that steric and electrostatic contribution were responsible for inhibitory activity. In CoMFA, steric interaction are represented by green and yellow contour where green represents the bulky group increases activity and yellow indicates bulky group decreases activity whereas in electrostatic interaction, blue contour represents region where electropositive group enhances activity and red contour indicates electronegative decreases activity. The CoMFA contour map for highly active molecule 58 was represented in Fig. 4a and 4b for steric and electrostatic contribution respectively. In steric, the green contour over the phenyl ring attached to 5th position of benzene ring indicates the bulky substituents at this position increases activity, and presence of bulky substituents near the sulphur atom attached to imidazoyl ring also increase activity. The yellow contour indicates that the presence of bulky group to oxygen atom attached at 2nd position of benzene ring decreases activity. In electrostatic contribution, the blue contour over the substituents attached to the imidazol ring suggest that the more electropositive group at that position is responsible for increased activity of the compound.

4. Conclusion

In this study, pharmacophore model was generated using highly active compounds in the dataset. Ligand based CoMFA model was generated from the alignment obtained using pharmacophore. The presence of bulky substituents at 5th position of benzene ring will increases the activity of compounds. The good correlation between experimental and predicted pIC_{50} values was obtained which indicates the developed models were robust. The results from this study can be used to design...
more CRTh2 antagonist with high inhibitory activities.

References

[1] J. Pothier, M. A. Riederer, O. Peter, X. Leroy, A. Valdenaire, C. Gnerre, and H. Fretz, “Novel 2-(2-(benzylthio)-1H-benzo[d]imidazol-1-yl) acetic acids: Discovery and hit-to-lead evolution of a selective CRTh2 receptor antagonist chemotype”, Bioorg. Med. Chem. Lett., Vol. 22, pp. 4660-4664, 2012.

[2] R. Pettipher and M. Whittaker, “Update on the development of antagonists of chemoattractant receptor-homologous molecule expressed on Th2 cells (CRTH2). From lead optimization to clinical proof-of-concept in asthma and allergic rhinitis”, J. Med. Chem., Vol. 55, pp. 2915-2931, 2012.

[3] D. Bonafoux, A. Abibi, B. Bettencourt, A. Burchat, A. Ericsson, C. M. Harris, T. Kebede, M. Morytko, M. McPherson, G. Wallace, and X. Wu, “Thienopyrrole acetic acids as antagonists of the CRTH2 receptor”, Bioorg. Med. Chem. Lett., Vol. 21, pp. 1861-1864, 2011.

[4] T. Ulven and E. Kostenis, “Novel CRTH2 antagonists: a review of patents from 2006 to 2009”, Expert Opin. Ther. Pat., Vol. 20, pp. 1505-1530, 2010.

[5] A. N. Hata, T. P. Lybrand, and R. M. Breyer, “Identification of determinants of ligand binding affinity and selectivity in the prostaglandin D2 receptor CRTH2”, J. Biol. Chem., Vol. 280, pp. 32442-32451, 2005.

[6] T. N. Birkinshaw, S. J. Teague, C. Beech, R. V. Bonnett, S. Hill, A. Patel, S. Reakes, H. Sangamee, I. G. Dougall, T. T. Phillips, S. Salter, J. Schmidt, E. C. Arrowsmith, J. J. Carrillo, F. M. Bell, S. W. Paine, and R. Weaver, “Discovery of potent CRTh2 (DP2) receptor antagonists”, Bioorg. Med. Chem. Lett., Vol. 16, pp. 4287-4290, 2006.

[7] R. Pettipher, “The roles of the prostaglandin D2 receptors DP1 and CRTH2 in promoting allergic responses”, Brit. J. Pharmacol., Vol. 153, pp. S191-S199, 2008.

[8] S. Crosignani, P. Page, M. Missotten, V. Colovray, C. Cleva, J.-F. Arrighi, J. Atherall, J. Macritchie, T. Martin, Y. Humbert, M. Gaudent, D. Pupowicz, M. Maio, P.-A. Pittet, L. Golzio, C. Giachetti, C. Rocha, G. Bernardinelli, Y. Filimchuk, A. Scheer, M. K. Schwarz, and A. Chollet, “Discovery of a new class of potent, selective, and orally bioavailable CRTH2 (DP2) receptor antagonists for the treatment of allergic inflammatory diseases”, J. Med. Chem., Vol. 51, pp. 2227-2243, 2008.

[9] A. K. Jain, N. Manocha, V. Ravichandran, V. K. Mourya, and R. K. Agrawal, “Three-dimensional Qsar study of 2,4-disubstituted-phenoxo acetic acid derivatives as a Crth2 receptor antagonist: Using the K-nearest neighbor method”, Dig. J. Nanomater. Bios., Vol. 3, pp. 147-158, 2008.

[10] G. Jones, P. Willett, and R. C. Glen, “A genetic algorithm for flexible molecular overlay and pharmacophore elucidation”, J. Comput. Aid. Mol. Des., Vol.9, pp. 532-549, 1995.

[11] SYBYL Software, Version X2.0., Tripos Associates Inc, St. Louis, USA, 2006.

[12] R. D. Cramer, D. E Patterson, and J. 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.

[13] S. Wold, “Cross-validatory estimation of the number of components in factor and principal component model”, Technometrics, Vol. 20, pp. 397-405, 1978.

[14] S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, “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.