Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
ELSEVIER
Bioorganic & Medicinal Chemistry Letters 16 (2006) 4467–4474

Comparative QSAR modeling of CCR5 receptor binding affinity of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas

J. Thomas Leonard and Kunal Roy*

Drug Theoretics and Cheminformatics Lab, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

Received 6 April 2006; revised 6 June 2006; accepted 12 June 2006
Available online 27 June 2006

Abstract—The present QSAR study attempts to explore the structural and physicochemical requirements of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas for CCR5 binding affinity using linear free energy-related (LFER) model of Hansch. QSAR models have been developed using electronic (Hammett σ), hydrophobicity (π), and steric (molar refractivity and STERIMOL L, $B1$, and $B5$) parameters of phenyl ring substituents of the compounds along with appropriate dummy variables. Whole molecular descriptor like partition coefficient ($\log P_{\text{calcd}}$) was also tried as an additional descriptor. Statistical techniques like stepwise regression, multiple linear regression with factor analysis as the data preprocessing step (FA-MLR), partial least squares with factor analysis as the preprocessing step (FA-PLS), principal component regression analysis (PCRA), multiple linear regression with genetic function approximation (GFA-MLR), and genetic partial least squares (G/PLS) were applied to identify the structural and physicochemical requirements for the CCR5 binding affinity. The generated equations were statistically validated using leave-one-out technique. The quality of equations obtained from stepwise regression, FA-MLR, FA-PLS, and PCRA is of acceptable statistical range (explained variance ranging from 71.9% to 80.4%, while predicted variance ranging from 67.4% to 77.0%). The GFA-derived models show high intercorrelation among predictor variables used in the equations while the G/PLS model shows lowest statistical quality among all types of models. The best models were also subjected to leave-25%-out crossvalidation.

Acquired immunodeficiency syndrome (AIDS) is a fatal disorder for which no complete and successful chemotherapy has been developed so far. Human immunodeficiency virus subtype 1 (HIV-1), a retrovirus of the lentivirus family, has been found to be prevalent in causing this disease. HIV-1 produces a progressive immunosuppression by destruction of $CD4^+$ T lymphocytes ('helper' cells, which lead attack against infections), and results in opportunistic infections and death.[1]

The replicative cycle of HIV can be divided into entry and postentry steps.[2,3] Entry of the HIV into a target cell consists of three vital steps: (1) the trimeric HIV-1 envelope glycoprotein complex-mediated viral entry into susceptible target cells: the surface subunit (gp120) attaches to the receptor (CD4); (2) gp120-co-receptor (CXCR4 or CCR5) interaction, which results in the exposure of a co-receptor-binding domain in gp120 on the cell surface; (3) and subsequent conformational changes within the Env complex which lead to membrane fusion mediated by the transmembrane subunit (gp41). Each of the stages can serve as a target for the HIV entry.

Postentry steps[4] require the viral reverse transcriptase (RT), integrase (IN), and protease (PR) enzymes to complete the viral replication cycle. The virally encoded RT enzyme mediates reverse transcription. RT is a heterodimeric (p51 and p66 subunits) and multifunctional enzyme presenting both RNA and DNA polymerase and RNaseH activities, being responsible for the conversion of the single-stranded viral RNA into the double-stranded proviral DNA.[1] The viral integrase enzyme is required for the integration of proviral DNA into the host genome before replication. When the infected cell synthesizes new protein, integrated proviral DNA is also translated into the protein building blocks of new viral progeny. Subsequent expression of the virus by the host cells produces

Abbreviations: QSAR, quantitative structure–activity relationships; AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus; LFER, linear free energy related; G/PLS, genetic partial least squares; GFA, genetic function approximation; FA, factor analysis.

Keywords: QSAR; Hansch analysis; LFER; CCR5; Piperidinyl amide; Urea.

* Corresponding author. Fax: +91 33 2837 1078; e-mail: kunalroy_in@yahoo.com
URL: http://www.geocities.com/kunalroy_in.

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.06.031

the gag and gag-pol proteins Pr44 and Pr160 of HIV-DNA that are processed by the HIV-encoded PR into functional proteins and enzymes. The viral components then assemble on the cell surface and bud out as immature viral particles. The final maturation of newly formed viruses requires the HIV-1 protease to make up an infectious virion. The inhibition of the key enzymes, HIV-1 reverse transcriptase and HIV-1 protease, provides the most attractive target for the anti-HIV drug development.[5–7]

Among various methods of anti-HIV activity screening, some important methods are cytoprotection assay, integration enzyme assay, RT inhibition assay, HIV attachment assay, fusion assay, etc.[8,9]

The present group of authors has developed a few quantitative structure–activity relationship (QSAR) models for anti-HIV activities of different group of compounds, for example, 2-amino-6-arylsulfonylbenzonitriles,[10] benzylpyrazoles,[11] imidazoles,[12] phenylpropylamines,[13] and mannitol[14] derivatives. In continuation of such efforts, the present paper deals with QSAR modeling of CCR5 binding affinity data of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas.[15,16]

The CCR5 binding affinity data reported by Burrows et al.[15,16] have been used as the model dataset for the present QSAR study: the affinity (50% inhibitory concentration) data [IC_{50} (μM) and IC_{50} (nM)] of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas (Table 1) for ^{125}I-labeled RANTES (regulated on activation normal T-cell expressed and secreted) to Chinese hamster ovary (CHO) cells expressing human CCR5 have been converted to the logarithmic scale [pIC_{50} (mM)] and then used for subsequent QSAR analyses as the response variable. There are five regions of structural variations in the compounds: one is the R^1-position (showing limited substitution pattern), second one is the X position (showing limited structural variations), and the remaining are R^2, R^3, and R^4 positions of the phenyl rings (showing diverse substitution pattern) (Table 1). This paper uses classical LFER approach using substituent constants;[17–19] thus, compounds containing the common scaffold of [1-(3,3-diphenylpropyl)piperidin-4-yl]-2-phenylacetamides and [1-(3,3-diphenylpropyl)piperidin-4-yl]-*N*-benzylureas were only considered for the present analysis. The objective of the work was to find out the contribution pattern of the phenyl ring substituents. The binding affinity data were subjected to classical QSAR analysis using linear free energy-related (LFER) model of Hansch[17–19] with lipophilicity (π), electronic (Hammett σ), and steric (molar refractivity mr and STERIMOL *L*, *B*1, and *B*5) parameters of the phenyl ring substituents along with appropriate dummy parameters as descriptors. Various indicator variables used in the study have been defined in Table 2. The values of the physiochemical substituent constants (Table S1 in Supplementary material) were taken from the literature.[17] Hydrophobic whole molecular descriptor (partition coefficient $\log P_{calcd}$) was also tried as predictor variable. SMILES were generated from the structures using the JME molecular editor (http://www.molinspiration.com/jme/) and then $\log P$ values were calculated using the ALOGPS 2.1 software [Virtual Computational Chemistry Laboratory (VCC-LAB); http://vcclab.org/lab/alogps]. The calculated $\log P$ ($\log P_{calcd}$) values for all the compounds are given in Table 1.

For the development of equations, six methods were used: (1) stepwise regression,[20] (2) multiple linear regression with factor analysis[21,22] as the data pre-processing step for variable selection (FA-MLR), (3) partial least squares[23,24] with factor analysis as preprocessing step (FA-PLS), (4) principal component regression analysis (PCRA),[22] (5) multiple linear regression with genetic function approximation (GFA-MLR),[24,25] and (6) genetic partial least squares (G/PLS).[26] The details of the methods are given Supplementary material.

The stepwise regression, factor analysis (FA), and principal component regression analysis were performed using the statistical software SPSS.[27] PLS was performed using statistical software MINITAB.[28] Model extraction from the data using genetic function approximation (GFA) and genetic partial least squares G/PLS was done using QSAR+ environment of Cerius^2 software.[29]

The statistical qualities of the MLR equations[30] were judged by the parameters like explained variance (R_a^2), correlation coefficient (R), standard error of estimate (s), and variance ratio (F) at specified degrees of freedom (df). All accepted MLR equations have regression coefficients and F ratios significant at 95% and 99% levels, respectively, if not stated otherwise. The generated QSAR equations were validated by predicted residual sum of squares (PRESS) (leave-one-out or LOO),[31,32] crossvalidation R^2 (Q^2), standard deviation based on PRESS (S_{PRESS}) and standard deviation of error of prediction (SDEP). Finally, leave-25%-out crossvalidation was applied on selected equations.

Stepwise regression. Using stepping criterion based on F value ($F = 3$ for inclusion; $F = 2.9$ for exclusion), the following best equation was derived with eight variables.

$$\begin{aligned} pIC_{50} = & -0.863(\pm0.820)\sigma_{R2\_m} + 1.230(\pm0.355)\sigma_{R2\_p} \\ & + 0.737(\pm0.614)\sigma_{R3\_p} + 1.179(\pm0.469)mr_{R3\_p} \\ & - 0.356(\pm0.142)mr^2_{R3\_p} + 0.791(\pm0.371)I_{NHCH2} \\ & - 1.279(\pm0.670)I_{BRANCH} - 1.206(\pm0.666)I_{R4\_4S} \\ & + 2.986(\pm0.185) \\ n = 79,\ & R_a^2 = 0.719,\ R^2 = 0.747,\ R = 0.865, \\ F = 25.9(df8, 70),\ & s = 0.455,\ SDEP = 0.487, \\ S_{PRESS} = 0.517,\ & Q^2 = 0.674,\ PRESS = 18.706 \end{aligned} \quad (1)$$

The 95% confidence intervals of the regression coefficients are mentioned within parentheses. Eq. 1 can explain and predict 71.9% and 67.4%, respectively, of the variance of the binding affinity data. The calculated binding affinity values according to Eq. 1 are given in Table 1. Various indicator variables used in this study have been defined in Table 2. The regression