In silico Modeling and Toxicity Profiling of a Set of Quinoline Derivatives as c-MET Inhibitors in the treatment of Human Tumors

İnsan Tümörlerinin Tedavisinde c-MET İhibitorü Olarak Kullanılan Bir Dizi Kinolin Türevinin In silico Modellemesi ve Toksisite Profili

Gülçin TUĞCU1*, Filiz Esra ÖNEN BAYRAM2, Hande SİPAHIL

1Yeditepe University Faculty of Pharmacy, Department of Toxicology, İstanbul, Turkey
2Yeditepe University Faculty of Pharmacy, Department of Pharmaceutical Chemistry, İstanbul, Turkey

ABSTRACT

Objectives: 4-(2-fluorophenoxy) quinoline derivatives constitute one of the chemical classes of hepatocyte growth factor receptor (c-MET) inhibitors, a promising treatment against various human tumors. There are three aims of the present study: (1) To develop a robust and validated quantitative structure-activity relationship model to predict the c-Met kinase inhibition; (2) to examine the toxicity profiles of these compounds; (3) to design new quinoline derivatives and apply the developed model on these compounds to observe its pertinence.

Materials and Methods: A multiple linear regression method was used to develop the model with calculated descriptors. State-of-the-art internal and external validation parameters were calculated. The in silico toxicity profiles including structural alerts and the lowest observed adverse effect level (LOAEL) values were evaluated using online tools. New derivatives were designed and tested on the developed model.

Results: A series of 4-(2-fluorophenoxy) quinoline derivatives was linearly modeled and vigorously validated to predict the molecules' c-MET kinase inhibition potential. Statistical metrics of the developed model showed that it was robust and able to make successful predictions for this chemical class. The mass, electronegativity, partial charges, and the structure of the molecules had an effect on the activity. Moreover, the toxicity profiles of the studied compounds were found to be adequate.

Conclusion: Five of the synthesized compounds were observed to be auspicious for the toxicity/activity ratio. The developed model is useful in the virtual screening and in the design of new anti-tumor compounds.

Key words: Toxicity, LOAEL, anti-tumor, c-MET, QSAR

ÖZ

Amaç: 4-(2-florofenoksi) kinolin türevleri, insan tümörlerinin tedavisinde kullanılan hepatosit büyümeye nötral reseptörü (c-MET) inhibitörleri kimyasal sınıfından biridir. Bu çalışmanın üç amacı vardır: (1) c-MET kinaz inhibisyonunu etmek için sağlıklı ve doyurulmuş bir kuantitatif yapısı-aktüvite ilişkisi modeli geliştirmek; (2) üzerinde çalışılan bileşiklerin toksisite profillerini incelemek; (3) yeni kinolin türevleri tasarlamak ve geliştirilen modeli bu bileşikler üzerine uygulayarak uygulanışı incelemek.


Anahtar kelimeler: Toksisite, LOAEL, anti-tümör, c-MET, QSAR
INTRODUCTION

Cancer is a worldwide health problem. The Hepatocyte growth factor receptor (c-MET) kinase inhibition has been a novel treatment for various human cancers. In addition to being effective in treating diseases, the approved drugs are expected to be safe at an acceptable level. In other words, the risk-benefit profile of these drugs was considered to be greater than their risks. Among these compounds, crizotinib and cabozantinib are two c-MET inhibitors that meet these criteria and are used in the treatment of various cancers. There are other molecules that have been added to the list of approved drugs. However, studies on promising c-MET tyrosine kinase inhibitors such as JNJ-38877605 were suspended in phase 1 due to the high risk of toxicity. Therefore, before starting clinical trials, it is greatly important to explore the safety profile of a drug candidate in the preclinical phase as early as possible so that studies for safer drugs continue expeditiously. Quantitative structure-activity-relationship (QSAR) modeling can be utilized to reduce costs, save time, and understand mechanisms of active substances even before a chemical is synthesized.

The safety evaluation of drugs is determined as the ratio of the safe amount of the drug exposed to the amount of drug that is effective. All preclinical data on the efficacy and safety endpoints need to be used as early as possible to understand the preliminary therapeutic index (TI). This index is calculated by dividing the highest amount of the drug that does not cause any toxicity to the amount that yields the aimed effect. The derivation of TI ratios from translational studies helps in proper decision-making for a drug in the next stage of drug development.

This study developed a QSAR model to estimate the c-MET kinase inhibition of a set of compounds. The compounds’ toxicity profiles are then explored in silico, and several compounds were determined to be both effective and safe. The model, which revealed satisfying internal and external validation statistics, was then applied to a set of newly designed compounds.

MATERIALS AND METHODS

Data set
A data set of 32 4-(2-fluorophenoxy) quinoline derivatives as c-MET kinase inhibitors was obtained from the literature. The negative logarithm of the activity value (p c-MET) was used as the dependent variable in the model equation. Table 1 lists the structures and biological activities (IC_{50}) of the studied compounds.

Molecular descriptors and descriptor selection

Compounds were drawn in Spartan v.18 (Wavefunction Inc., Irvine, CA) and corresponding geometries were optimized with the semi-empirical PM7 using MOPAC 6. Chemopy v.3.2, PaDEL v.2.21, and alvadesc 1.0.20 (www.alvascience.com/alvadesc) descriptors were calculated using geometry-optimized molecule files.

Descriptor selection was performed via genetic algorithm (GA) using QSARINS v.2.2.4 following the removal of the constant and near-constant value descriptors. A search with GA within the software follows the tournament selection method.

Model development and validation

The data set was split into training and test sets for an external validation procedure by systematically selecting compounds for the test set from the data set sorted according to activity. In this way, the activities were distributed homogenously and the training set was set as large as possible. Additionally, training and test sets are congruent so that the test set resides in the applicability domain. The training set consisted of 80% of the whole data set.

Selected significant descriptors were used as independent variables in MLR models to predict the biological activity. The delta K limit was set at 0.05 to eliminate models with collinear descriptors. The maximum number of descriptors in the model has been limited to at least five compounds per descriptor (Topliss Ratio) to avoid overfitting in the model. Linear models were developed using the ordinary least square method, which is carried out in QSARINS v.2.2.4 with the selected descriptors as the independent variables.

Internal and external validation procedures were performed vigorously using widely known parameters in the literature. The coefficient of determination (R^2), leave-one-out cross-validation (Q_{LOO}^2), standard error (SE), root mean squared error (RMSE), and Fischer statistics (F) as model fitting and internal validation criteria were listed. The reliability of the developed model was tested by the randomization procedure (Y-scrambling). Activity values in the training set were rearranged and new correlation coefficients were calculated for the randomized models. The average correlation coefficient of the new models was expected to be significantly low corresponding to the model correlation coefficient, indicating the absence of a chance correlation. External validation parameters, representing the predictive ability of the model, were listed as R^2, RMSE, Q_{FI}^2, Q_{F2}^2 (0.70), Q_{F3}^2 (0.70), CCC (0.85), and r^2 (0.65) of the test set with the recommended lower limits given in parentheses. Furthermore, conditions described by Golbraikh and Tropsha were applied to the test set predictions:

I. \( (R^2 - R^2_{\text{LOO}})/R^2 < 0.1 \) and \( 0.85 \leq k \leq 1.15 \)

or

\( (R^2 - R^2_{\text{LOO}})/R^2 < 0.1 \) and \( 0.85 \leq k' \leq 1.15 \)

II. \( |R^2_{\text{LOO}} - R^2| < 0.3 \)

where \( R^2 \) corresponds to the predicted vs. observed, \( R^2_{\text{LOO}} \) refers to the observed vs. predicted, k and k’ are slopes, and \( R^2 \) and \( R^2_{\text{LOO}} \) are coefficients of determination passing through the origin.

Applicability domain

The applicability domain of the model was defined via the Williams plot. Leverage (hat) values lie on the x-axis, while standardized residuals are on the y-axis in the graph. Response outliers were diagnosed by the limit of ±2.5 standardized
### Table 1. Molecular structures and activities of studied compounds

<table>
<thead>
<tr>
<th>Compound</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>IC&lt;sub&gt;50&lt;/sub&gt; (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10a</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Cyclopentyl</td>
<td></td>
<td>52.86</td>
</tr>
<tr>
<td>10b</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>2-thiophenyl</td>
<td></td>
<td>37.46</td>
</tr>
<tr>
<td>10c</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Naphtyl-</td>
<td></td>
<td>48.36</td>
</tr>
<tr>
<td>10d</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Phenyl-</td>
<td></td>
<td>22.74</td>
</tr>
<tr>
<td>10e</td>
<td>propyl</td>
<td>Phenyl-</td>
<td></td>
<td>19.45</td>
</tr>
<tr>
<td>10f</td>
<td>cyclohexyl</td>
<td>Phenyl-</td>
<td></td>
<td>30.38</td>
</tr>
<tr>
<td>10g</td>
<td>phenyl</td>
<td>Phenyl-</td>
<td></td>
<td>54.42</td>
</tr>
<tr>
<td>10h</td>
<td>'But</td>
<td>Phenyl-</td>
<td></td>
<td>8.35</td>
</tr>
<tr>
<td>10i</td>
<td>'But</td>
<td>4-methylphenyl-</td>
<td></td>
<td>12.35</td>
</tr>
<tr>
<td>10j</td>
<td>'But</td>
<td>4-methoxyphenyl-</td>
<td></td>
<td>21.48</td>
</tr>
<tr>
<td>10k</td>
<td>'But</td>
<td>3,4,5-trimethoxyphenyl-</td>
<td></td>
<td>35.64</td>
</tr>
<tr>
<td>10l</td>
<td>'But</td>
<td>'But-</td>
<td></td>
<td>18.45</td>
</tr>
<tr>
<td>10m</td>
<td>'But</td>
<td>4-fluorophenyl-</td>
<td></td>
<td>2.49</td>
</tr>
<tr>
<td>10n</td>
<td>'But</td>
<td>3-fluorophenyl-</td>
<td></td>
<td>4.72</td>
</tr>
<tr>
<td>10o</td>
<td>'But</td>
<td>2-fluorophenyl-</td>
<td></td>
<td>3.13</td>
</tr>
<tr>
<td>10p</td>
<td>'But</td>
<td>4-chlorophenyl-</td>
<td></td>
<td>7.43</td>
</tr>
<tr>
<td>10q</td>
<td>'But</td>
<td>3-chlorophenyl-</td>
<td></td>
<td>7.86</td>
</tr>
<tr>
<td>10r</td>
<td>'But</td>
<td>2-chlorophenyl-</td>
<td></td>
<td>8.12</td>
</tr>
<tr>
<td>10s</td>
<td>'But</td>
<td>4-bromophenyl-</td>
<td></td>
<td>18.44</td>
</tr>
<tr>
<td>10t</td>
<td>'But</td>
<td>4-CF&lt;sub&gt;3&lt;/sub&gt;-phenyl-</td>
<td></td>
<td>36.62</td>
</tr>
<tr>
<td>10u</td>
<td>'But</td>
<td>2,3-dichlorophenyl-</td>
<td></td>
<td>18.66</td>
</tr>
<tr>
<td>10v</td>
<td>'But</td>
<td>2-thiophenyl</td>
<td></td>
<td>32.74</td>
</tr>
<tr>
<td>10w</td>
<td>'But</td>
<td>2-furanyl-</td>
<td></td>
<td>42.68</td>
</tr>
<tr>
<td>10x</td>
<td>'But</td>
<td>Cyclopentyl</td>
<td></td>
<td>66.54</td>
</tr>
<tr>
<td>10y</td>
<td>'But</td>
<td>Naphtyl-</td>
<td></td>
<td>46.87</td>
</tr>
</tbody>
</table>

6,7-dimethoxy-4-(2-fluorophenoxy)quinoline derivatives
residuals, while the structural outlier threshold was set at the critical hat \( h^* \). The critical hat value was calculated as \( 3p/n \), where \( p \) is the number of descriptors plus one and \( n \) is the number of molecules in the training set.

**Exploring toxicity profile of c-MET inhibitors**

The toxicity profile of c-MET inhibitors was predicted using online tools and software. As a first step, structural alerts were sought for pan-assay interference compounds (PAINS)\(^{20} \) and Brenk et al.\(^{21} \) using SwissADME.\(^{22} \) These alerts filter unwanted substructural features and screen compounds for lead-likeness. One of the parameters used in the safety evaluation is the lowest observed adverse effect level (LOAEL), which is defined as the lowest dose in a study at which adverse effects were observed, although it can lead to a significant overestimation of the TI.\(^{3} \) LOAEL values (obtained from oral rat chronic toxicity tests) were predicted via pkCSM.\(^{23} \)

The TI can be used as the ratio of the highest amount of the drug that does not cause any toxicity to the amount that yields creates the desired effect in a translational research environment. This study used in vitro efficacy and in silico safety endpoint to compare safety margin interpretations. The greater the difference between the effective dose of a drug candidate molecule and its toxic dose, the safer that drug is. Consequently, a ratio was calculated as LOAEL to IC\(_{50}\) to evaluate the toxicity over the activity. This study defined a threshold level of 1 for this ratio.

**Design of new c-MET inhibitors**

New compounds were designed to explore potential anti-tumor agents. Inspired from papers of Parikh and Ghate\(^{24} \) and Liu et al.\(^{25} \) useful molecular fragments were selected. For instance, pyrazol and oxolane functional groups were introduced to the designed molecules. Additionally, the single methoxy group on the quinoline backbone was proved on some compounds. Designed compounds were tested using the developed model, and predicted activities and leverages were plotted on a graph.

**RESULTS AND DISCUSSION**

The training and test set selection was performed using systematic division. Activity values were sorted in decreasing order. The most and the least active compounds were assigned into the training set to provide a broader applicability domain. Finally, 80% of the data set was allocated into the training set. A total of 5182 descriptors were calculated using Chemopy (633), Padel (1444), and alvaDesc (3105) programs for all compounds. Selected descriptors were used as independent variables in the model developed on the training set. A GA-based search with parameters of 500 iterations, a population of 50 models, and a mutation rate of 20 was employed to find significant descriptors for the model. The four-descriptor MLR model (Eqn. 1) below was found to be predictive with an acceptable goodness-of-fit value. SEs of the coefficients were given in the parentheses (p<0.05).

\[
\begin{align*}
\text{c-MET} & = 30.428 \pm 8.794 + 0.087 \pm 0.018 \text{PEOEVSA2} + 0.034 \pm 0.016 \text{AATSC4m} - 5.751 \pm 1.094 \text{SpMin8_Bh(e)} - 25.005 \pm 8.584 \text{VR2_RG} \\
& = (26, R^2 = 0.812, R^2_{\text{adj}} = 0.776, \text{RMSE}_{\text{tr}} = 0.163, \text{MAE}_{\text{tr}} = 0.138) \\
& = (6, R^2_{\text{test}} = 0.782, \text{RMSE}_{\text{test}} = 0.130, \text{MAE}_{\text{test}} = 0.102) \\
& = (0.800, Q^2_{F1} = 0.756, Q^2_{F2} = 0.881, \text{CCC} = 0.877) \\
& = (0.692, |R^2_{o} - R^2_{r}| = 0.018) \\
& = (0.756), k = 1.019, (R^2 - R^2_{o})/R^2 = 0.030 \\
& = (0.776), k = 0.974, (R^2 - R^2_{o})/R^2 = 0.007
\end{align*}
\]

The reliability of the model was investigated using a randomization test, which was run for 2000 iterations for the developed model. The average \( R^2 \) was 0.159. Additionally, the distinct distribution of \( R^2 \) and \( Q^2 \) values belong to randomized models, and the developed model (Figure SI) revealed that the generated model is robust and the descriptors are not selected by chance.

The four-descriptor linear model has fulfilled the internal and external validation criteria listed in the materials and methods section. Figure 1 depicts the predicted versus observed activities, which are scattered around the y = x line. Table S11 gives the predicted values by the model, leverages, and standardized residuals for the studied compounds. Table SII gives the calculated descriptor values, while Table

![Figure 1. Predicted versus experimental activity values of training and test set compounds](Image 337x70 to 547x275)
lists the correlation matrix of the descriptors of the model, showing that the descriptors were not intercorrelated. The Williams plot was depicted to show both the structurally distant compounds to the others (leverages) and the standardized residuals that represent the model’s prediction accuracy. Horizontal reference lines are at ±2.5 σ and the vertical reference line is at the critical hat value (h*: 0.577). The model did not have any outliers in terms of leverages and standardized residuals. Hence, all compounds are located within the AD of the model. Supplementary information Table S11 lists the leverages and standardized residuals for the studied compounds.

The compounds were inspected in silico according to their structural alerts and LOAELs to explore the toxicity profile of the set of c-MET inhibitors as potential anti-tumor agents. No alert was fired for PAINS and Brenk screens in the SwissADME software. LOAEL values that were predicted via pkCSM were used to evaluate the safety of drugs. LOAEL predictions varied between 0.237 and 75.683 mg/kg/day. 10m, 11a, 11b, 11c, and 11g with the highest toxic dose/ effective dose ratio were seen to be safe at the effective dose as expected from a drug (Table S13) (Figure 2).

As described in the materials and methods section in detail, 20 quinoline derivatives were designed. The developed model was deployed on this set of new compounds (Table S14) to explore possible inhibitors. Among 20 compounds, 16 of them were located in the AD and showed promising results as drug candidates (Figure 3). Compounds D06 and D07 were structurally distant from the modeled compounds. D02 and D13 had extreme prediction values, resulting in falling outside the AD of the model.

### CONCLUSION

A QSAR model was developed with four descriptors to predict the c-MET inhibitory activity. The model revealed that the mass, electronegativity, partial charges, and the structure of the molecules have an influence on the c-MET inhibition. The model was validated internally and externally to prove its robustness and predictivity. Compounds were predicted for their availability as drug candidates and toxicity profile. It can be concluded that the compounds are opportune to be drug candidates and five of them ensued promising as both active and relatively safe. Additionally, the developed model was applied to 20 newly designed compounds. The model showed promising results on the new compounds. The model can be used in the early design process of c-MET inhibitors as anti-tumor agents.

**Table 2. Descriptors used in the model along with their standardized coefficients**

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Software</th>
<th>Meaning</th>
<th>Standardized coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEOEVSA2</td>
<td>ChemoPy</td>
<td>MOE-type descriptor using partial charges and surface area contributions</td>
<td>0.492</td>
</tr>
<tr>
<td>AATSC4m</td>
<td>PaDEL</td>
<td>Autocorrelation descriptor</td>
<td>0.297</td>
</tr>
<tr>
<td>SpMin8_Bh(e)</td>
<td>alvaDesc</td>
<td>Burden eigenvalues</td>
<td>-0.610</td>
</tr>
<tr>
<td>VR2_RG</td>
<td>alvaDesc</td>
<td>3D matrix-based descriptor</td>
<td>-0.350</td>
</tr>
</tbody>
</table>

ChemoPy: Chemoinformatics in Python, PaDEL: PaDEL-Descriptor, alvaDesc: Alvascience Descriptor

**Table 3. Correlation matrix of descriptors used in the model**

<table>
<thead>
<tr>
<th></th>
<th>PEOEVSA2</th>
<th>AATSC4m</th>
<th>SpMin8_Bh(e)</th>
<th>VR2_RG</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEOEVSA2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AATSC4m</td>
<td>0.14</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SpMin8_Bh(e)</td>
<td>-0.17</td>
<td>0.52</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>VR2_RG</td>
<td>0.02</td>
<td>-0.61</td>
<td>-0.25</td>
<td>1</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

The authors would like to thank Prof. Paola Gramatica for providing QSARINS program.

Conflict of interest: No conflict of interest was declared by the authors. The authors are solely responsible for the content and writing of this paper.

REFERENCES

1. WHO, 2018. Available from: https://www.who.int/health-topics/cancer#tab=tab_1