

Predicting Binding Affinity of Some Imatinib Derivatives as BCR-ABL Tyrosine Kinase Inhibitors Based on Monte Carlo Optimization

Shahram Lotfi¹ , Shahin Ahmadi^{2*} , Sharare Vardast Baghmisheh³ , Ali Almasirad² 

¹ Dept of Chemistry, Payame Noor University (PNU), Tehran, Iran

² Dept of Pure and Pharmaceutical Chemistry, Faculty of Pharmaceutical Chemistry, Tehran medical sciences, Islamic Azad University, Tehran, Iran

³ Faculty of Pharmacy, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran

Article Info

Article type:
Research article

Article History:

Received: Dec. 30, 2023

Received in Revised Form:

Mar. 14, 2024

Accepted: May. 19, 2024

Published Online: Oct. 05, 2024

* Correspondence to:

Shahin Ahmadi

Dept of Pure and Pharmaceutical Chemistry, Faculty of Pharmaceutical Chemistry, Tehran medical sciences, Islamic Azad University, Tehran, Iran

Email:

ahmadi.chemometrics@gmail.com

ABSTRACT

Introduction: The Imatinib drug is used to treat blood cancer by inhibiting the BCR-ABL tyrosine kinase enzyme, which prevents the proliferation of cancer cells.

Materials & Methods: In order to predict the binding affinity of 555 compounds of imatinib derivatives as ABL-BCR tyrosine kinase inhibitors, quantitative structure-activity relationship (QSAR) modeling was performed using the Monte Carlo method. The data were randomly divided into four series, including training, invisible training, calibration, and validation sets, as well as they were randomly repeated three times.

Results: The results of three random divisions indicated reliable models for predicting the set of external tests with correlation coefficient (R²) and cross-validation correlation coefficient (Q²) in the range of 0.8575-0.8775 and 0.7620-0.7793. Consequently, the obtained models help identify hybrid descriptors for increasing or decreasing binding affinity (K_i) as BCR-ABL tyrosine kinase inhibitors. The mechanical interpretation of the model is given in the form of a report of descriptors that decrease and increase pK_i, as well as examples of these descriptors.

Conclusion: The results reveal that the designed models can be considerably effective in estimating the biological effect of imatinib derivatives proposed by researchers and medicinal chemists. Therefore, it is possible to predict its possible biological effects by spending less time and money before conducting in vitro or in vivo experiments.

Keywords: Quantitative structure-activity relationship (QSAR), Chronic myeloid leukemia, Imatinib derivatives, Tyrosine kinase inhibitor, Binding affinity

How to cite this paper: Lotfi Sh, Ahmadi Sh, Vardast Baghmisheh Sh, Almasirad A. Predicting Binding Affinity of Some Imatinib Derivatives as BCR-ABL Tyrosine Kinase Inhibitors Based on Monte Carlo Optimization. *Journal of Ilam University of Medical Sciences*. 2024;32(4):66-86.



پیشگویی تمایل اتصال تعدادی از مشتقات ایماتینیب به عنوان مهارکننده‌های تیروزین کیناز بر اساس بهینه‌سازی مونت کارلو

شهرام لطفی^۱، شهین احمدی^{۲*}، شراره وردست باغمیشه^۳، علی الماسی راد^۳

^۱ گروه شیمی، دانشگاه پیام نور، تهران، ایران

^۲ گروه شیمی (محض-دارویی)، دانشکده شیمی دارویی، علوم پزشکی تهران، دانشگاه آزاد اسلامی، تهران، ایران

^۳ گروه شیمی دارویی، دانشکده داروسازی، علوم پزشکی تهران، دانشگاه آزاد اسلامی، تهران، ایران

چکیده

اطلاعات مقاله

مقدمه: بیماری لوسمی میلوئید مزمن نوعی از سرطان خون است که به تولید ژن جدیدی به نام BCR-ABL منجر می‌شود و حاوی دستورالعمل تولید پروتئین تیروزین کیناز توسط سلول‌های غیرطبیعی خون است. داروی ایماتینیب برای درمان این سرطان استفاده می‌گردد که با مهار آنزیم تیروزین کیناز BCR-ABL، از تکثیر سلول‌های سرطانی جلوگیری می‌کند.

مواد و روش‌ها: به منظور پیشگویی تمایل اتصال ۵۵۵ ترکیب از مشتقات ایماتینیب به عنوان مهارکننده ABL-BCR تیروزین کیناز، مدل‌سازی رابطه کمی ساختار-فعالیت (QSAR) با استفاده از روش مونت کارلو انجام شد. داده‌ها به صورت تصادفی به چهار سری آموزش، آموزش منفعل، کالیبراسیون و آزمون تقسیم و به صورت تصادفی ۳ بار تکرار گردید.

یافته‌های پژوهش: نتایج حاصل از سه تقسیم تصادفی، مدل‌های قابل اعتمادی را برای پیش‌بینی مجموعه آزمون‌های خارجی با ضریب همبستگی (R2) و ضریب همبستگی اعتبارسنجی متقاطع (Q2) در محدوده ۰/۸۷۷۵-۰/۸۷۷۵ و ۰/۷۷۹۳-۰/۷۶۲۰ نشان داد؛ در نتیجه، مدل‌های به دست آمده به شناسایی توصیفگرهای هیبریدی برای افزایش و کاهش تمایل اتصال (Ki) به عنوان مهارکننده BCR-ABL تیروزین کیناز کمک می‌کنند. تفسیر مکانیکی مدل به صورت گزارشی از توصیفگرهای کاهنده و افزایشنده pKi و همچنین مثال‌هایی از این توصیفگرها آمده است.

بحث و نتیجه‌گیری: نتایج نشان می‌دهد که مدل‌های طراحی شده می‌تواند در تخمین اثر بیولوژیکی مشتقات ایماتینیب پیشنهادی توسط محققان و متخصصان شیمی دارویی بسیار مؤثر باشد؛ بنابراین، با صرف هزینه و زمان کمتر و پیش از انجام آزمایش‌های برون‌تن یا درون‌تن می‌توان آثار بیولوژیک احتمالی آن را پیش‌بینی کرد.

واژه‌های کلیدی: QSAR، لوسمی مزمن میلوئیدی، مشتقات ایماتینیب، مهارکننده‌های تیروزین کیناز، تمایل اتصال

نوع مقاله: پژوهشی

تاریخ دریافت: ۱۴۰۲/۱۰/۰۹

تاریخ ویرایش: ۱۴۰۲/۱۲/۲۴

تاریخ پذیرش: ۱۴۰۳/۰۲/۳۰

تاریخ انتشار: ۱۴۰۳/۰۷/۱۴

نویسنده مسئول:

شهین احمدی

گروه شیمی (محض-دارویی)،
دانشکده شیمی دارویی، علوم
پزشکی تهران، دانشگاه آزاد
اسلامی، تهران، ایران

Email:

ahmadi.chemometrics@gmail.com

استاد: لطفی شهرام، احمدی شهین، وردست باغمیشه شراره، الماسی راد علی. پیشگویی تمایل اتصال تعدادی از مشتقات ایماتینیب به عنوان مهارکننده‌های تیروزین کیناز بر اساس بهینه‌سازی مونت کارلو. مجله دانشگاه علوم پزشکی ایلام، مهر ۱۴۰۳؛ ۳۲(۴): ۸۶-۶۶.



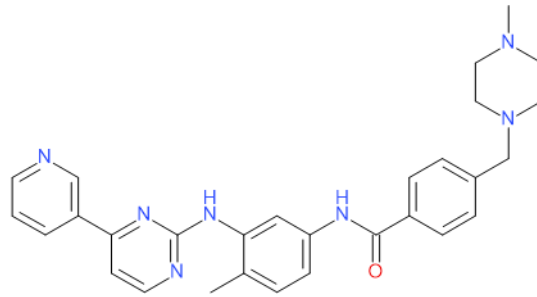
مقدمه

بر اساس گزارش سازمان بهداشت جهانی (WHO)، سرطان دومین عامل مرگ و میر در جهان پس از بیماری‌های قلبی عروقی است و تخمین زده شده است که ۱۵ میلیون نفر تا سال ۲۰۳۰ بر اثر سرطان در سراسر جهان جان خود را از دست بدهند (۱). بیماری لوسمی میلوئید مزمن (CML) یکی از انواع سرطان خون است که از اختلال کروموزومی در سلول‌های بنیادین خون‌ساز مغز استخوان منشأ می‌گیرد. نتیجه این اختلال کروموزومی تولید انکوژن کایمریک BCR-ABL است که تولید یک پروتئین تیروزین کیناز (TK) با فعالیت بالا را کد می‌کند (۲). ایماتینیب (Imatinib) اولین دارویی است که به‌عنوان یک مهارکننده BCR-ABL TK در درمان بیماران CML استفاده شده است. این دارو با مهار رقابتی اتصال ATP به آنزیم BCR-ABL، مانع فسفریلاسیون سوبسترا می‌گردد و فعالیت آن را مسدود می‌کند و از انتقال سیگنال‌های ضروری برای عملکرد سلولی جلوگیری می‌نماید (۳). با این حال، برخی از بیماران به‌ویژه در مراحل پیشرفته CML، به علت جهش‌های نقطه‌ای در بخش ABL کیناز یا با بیان بیش از حد BCR-ABL، نسبت به داروی ایماتینیب مقاومت نشان می‌دهند؛ بنابراین، برای غلبه بر این مقاومت، به توسعه مهارکننده‌های دیگر از مشتقات ایماتینیب نیاز است (۴). مشتقات جدیدی از ایماتینیب مانند پوناتینیب، نیلوتینیب، داساتینیب، بوسوتینیب و غیره به‌عنوان TKI طراحی و در بیماران مبتلا به سرطان خون آزمایش شده است. همسو با آن، نشان داده شده است که مهارکننده‌های BCR-ABL، به‌ویژه مشتقات ایماتینیب، پتانسیل فراوانی در هدایت و توسعه درمان‌های جدیدتر و کارآمدتر در مقابله با CML دارند (۲، ۵).

طراحی، توسعه و تجاری کردن یک دارو فرایندی پیچیده، وقت‌گیر و هزینه‌بر است. به علل یادشده، هر وسیله یا روشی که سرعت و بازده هر مرحله از فرایند کشف دارو را

افزایش دهد، مطلوب است. در علم شیمی دارویی، فعالیت بیولوژیکی مانند ویژگی‌های فیزیکی-شیمیایی و خواص شیمیایی ترکیبات آلی به ساختار مولکولی آن‌ها وابسته است. طراحی دارو به کمک کامپیوتر بر اساس روش‌های کمی ساختار-فعالیت (QSAR) یکی از ابزارهایی است که می‌تواند سرعت و بازده فرایند کشف دارو را افزایش دهد. مدل‌سازی QSAR یک پیشگویی و تشخیص است که برای یافتن ارتباط میان ساختارهای شیمیایی و فعالیت بیولوژیکی بر اساس معادلات ریاضی و روابط آماری استفاده می‌شود و در تلاش برای برآوردن نیاز و تمایل شیمیدانان دارویی برای پیش‌بینی پاسخ بیولوژیکی ترکیبات تازه مهندسی شده یا توسعه‌نیافته تکامل یافته است.

نرم‌افزار CORAL (<http://www.insilico.eu/coral>) به‌عنوان ابزاری مؤثر برای پیش‌بینی نقاط پایانی و مؤلفه‌های ارزیابی عملکرد بیولوژیکی خاص (مانند IC50 [متوسط غلظت کشنده]، Ki [تمایل اتصال]، LD50 [متوسط دوز کشنده]، درصد بازدارنده و غیره) مواد شیمیایی مختلف با استفاده از آنالیز QSAR پیاده‌سازی شده است. این نرم‌افزار با استفاده از یک نماد سیستم ورودی خطی مولکولی ساده شده (SMILES)، مدل‌هایی را برای نقاط پایانی مختلف ایجاد می‌کند تا ساختار شیمیایی را منعکس کند. توصیفگرهای بهینه از SMILES و با استفاده از روش بهینه‌سازی مونت کارلو محاسبه می‌شوند که در آن از ضریب همبستگی بین نقطه پایانی و توصیفگر بهینه برای محاسبه تابع هدف استفاده می‌گردد (۶-۹). هدف از این تحقیق مطالعه رابطه کمی ساختار-فعالیت ۵۵۵ ترکیب از مشتقات ایماتینیب (شکل شماره ۱) بر اساس الگوریتم مونت کارلو و با استفاده از توصیفگرهای SMILES توسط نرم‌افزار CORAL، به‌منظور پیش‌بینی تمایل اتصال (Ki) آن‌ها در برابر آنزیم تیروزین کیناز BCR-ABL (TK) است.



شکل شماره ۱. ساختار مولکولی ایماتینیب

25 set) درصد، آموزش منفعل (passive training set) 20 درصد، کالیبراسیون ۲۰ درصد و آزمون ۳۵ درصد از داده‌ها اختصاص داده شد. وظایف هر دسته از داده‌ها در روش توازن همبستگی عبارت است از: دسته آموزش: برای تهیه مدل QSAR با محاسبه DCW (وزن‌های همبستگی) که بیشترین ضریب همبستگی بین DCW(T, N) و pKi دارد؛ دسته آموزش منفعل: تأیید اینکه ضرایب همبستگی برای ترکیباتی که در دسته آموزش نیستند، خوب است؛ دسته کالیبراسیون: آشکار کردن شروع بیش برازش؛ دسته آزمون: برای تخمین نهایی از قدرت پیشگویی مدل (آشکارساز یک مدل خوب) (۱۱-۱۳). این کار سه بار به صورت تصادفی انجام شده است. برای توسعه مدل‌های QSAR از تابع هدف جدیدی به نام شاخص ایدئال همبستگی استفاده شده که نتایج بهتری نشان داده است. در نهایت، گروه‌های عاملی یا ساختارهای متصل شده به پیکره اصلی ایماتینیب که مسئول افزایش یا کاهش میزان تمایل اتصال به تیروزین کیناز BCR-ABL (TK) هستند، استخراج گردید و می‌تواند برای درک سازوکار تمایل اتصال به کار برده شود.

این پژوهش مطابق با نظر کمیته ی اخلاق دانشگاه علوم پزشکی آزاد اسلامی تهران به شماره IR.IAU.PS.REC.1400.260 انجام شده است.

انتخاب توصیفگرهای مولکولی مناسب:

محاسبه و تولید توصیفگرهای مولکولی یکی از مراحل مهم در مدل‌سازی QSAR است. سه گزینه برای انتخاب توصیفگرهای بهینه در نرم‌افزار مدل‌سازی CORAL وجود دارد: الف. توصیفگرهای مبتنی بر SMILES؛ ب. توصیفگرهای مولکولی بر اساس گراف مولکولی؛ ج.

مواد و روش‌ها

تهیه و آماده‌سازی داده‌ها یکی از مسائلی است که در حوزه طراحی دارو از اهمیت ویژه‌ای برخوردار است. معمولاً در مدل‌سازی QSAR، از فعالیت بیولوژیکی ترکیبات که پیش‌تر به صورت تجربی در آزمایشگاه اندازه‌گیری شده باشد، استفاده می‌شود. به این منظور، مشتقات ایماتینیب استفاده‌شده به همراه داده‌های میزان تمایل اتصال (Ki) آن‌ها برای ایجاد مدل QSAR از مقاله زین و همکارانش استخراج شده است (۱۰). ساختار مولکولی ترکیبات انتخاب‌شده توسط نرم‌افزار BIOVIA 2019 رسم گردید و به‌عنوان داده‌های ورودی نرم‌افزار CORAL به صورت نمادهای اسملی تبدیل شد. مدل‌سازی به کمک نرم‌افزار کورال، به بهینه‌سازی مولکول که فرایندی زمان‌بر است (به‌ویژه برای داده‌های بسیار فراوان این تحقیق)، نیاز ندارد و می‌توان مدل‌های جهانی را برای داده‌های بزرگ توسعه داد. برخلاف بسیاری از نرم‌افزارهای مدل‌سازی دیگر، همه مراحل مدل‌سازی در محیط نرم‌افزار انجام می‌شود و به نرم‌افزارهای تولید توصیفگر و انتخاب متغیر جداگانه نیاز ندارد.

نماد اسملی ترکیبات و میزان عددی تمایل اتصال آن‌ها در جدول S1 به‌عنوان داده‌های همراه مقاله ارائه گردیده است. داده‌های تجربی مربوط به این ترکیبات در برابر تیروزین کیناز BCR-ABL (TK) به صورت منفی لگاریتم (pKi) به کار برده شد. برای مدل‌سازی رابطه کمی ساختار-فعالیت از تابع pKi به‌عنوان متغیر وابسته استفاده گردید که مقدار آن از ۴/۲۴ تا ۱۰/۷۳ متغیر است. داده‌ها به صورت تصادفی به چهار گروه، دسته آموزش، آموزش منفعل، کالیبراسیون و آزمون تقسیم شده است. به دسته آموزش فعال (Active Training)

توصیفگرهای هیبریدی که از ترکیب هر دو گراف مولکولی و SMILES محاسبه می‌شوند؛ د. توصیفگرهای مولکولی که شرایط انجام آزمایش در آن تعریف می‌شود (Quasi-SMILES) (۱۴، ۱۵)؛ همچنین در نرم‌افزار CORAL، سه نوع گراف مولکولی ارائه داده می‌شود: ۱. گراف مولکولی بدون هیدروژن (HSG)؛ ۲. گراف مولکولی با هیدروژن (HFG) و

۳. گراف اوربیتال‌های اتمی (GAO) (۱۶، ۱۷).

توصیفگر بهینه ترکیبی با استفاده از ترکیبی از خصوصیات مولکولی استخراج شده از SMILES و ویژگی‌های مولکولی استخراج شده از گراف مولکولی محاسبه می‌شود. وزن همبستگی حاصل از توصیفگر بهینه ترکیبی به کمک معادله زیر محاسبه می‌گردد:

$$\text{HybridDCW}(T, N_{\text{epoch}}) = \text{SMILESDCW}(T, N_{\text{epoch}}) + \text{GraphDCW}(T, N_{\text{epoch}}) \quad (1)$$

به‌عنوان یک قانون کلی در نرم‌افزار کورال، به‌منظور انتخاب توصیفگرهای بهینه، با توجه به اتم‌های تشکیل‌دهنده مولکول و حضور باندهای دو گانه، سه گانه و باندهای فضایی می‌توان تعدادی از توصیفگرهای مبتنی بر SMILES را انتخاب کرد؛ مثلاً اگر مولکولی اتم هالوژن داشته باشد، «HALO» انتخاب می‌گردد؛ در غیر این صورت، این توصیفگر برای مدل‌سازی در نظر گرفته نمی‌شود. برای حضور حلقه‌های چندعضوی و سایر توصیفگرهای مبتنی بر SMILES به همین صورت انتخاب می‌گردند. در صورتی که توصیفگرهای مبتنی بر SMILES مدلی با معیارهای پیشگویی مناسب ارائه ندهد، از توصیفگرهای گراف مولکولی هم کمک گرفته می‌شود. از آنجا که توصیفگرهای گراف اوربیتال‌های اتمی به‌خوبی قابل تفسیر نیستند، به‌ندرت در مدل‌سازی استفاده می‌گردند. برای هر مجموعه تصادفی از مجموعه داده‌ها، چندین مدل مختلف با دسته‌های مختلف توصیفگرها ایجاد می‌شود و پس از تجزیه و تحلیل نتایج آماری مدل‌سازی با این توصیفگرها، بهترین مجموعه توصیفگرها انتخاب می‌گردد. در این تحقیق از توصیفگر

هیبریدی بهینه که ترکیبی از SMILES و HSG است، برای تولید مدل‌های QSAR استفاده شده است. درباره نسبت‌های جفت اتم‌ها، هنگامی از این دسته از توصیفگرها استفاده می‌شود که بخواهیم حضور دوتایی از توصیفگرهای Br، Cl، N، O، S، P، باند دوگانه (=) و باند سه گانه (#) را در مدل‌سازی بررسی کنیم. در صورتی که این اتم‌ها در ساختارها مشاهده نگردد، وضعیت انتخاب را در وضعیت «۰» قرار می‌دهیم. همان‌طور که در مدل‌سازی این دسته از داده‌ها، برای اتم فسفر (P) این وضعیت را «۰» قرار دادیم (شکل شماره ۲)، در صورتی که اتم‌ها به‌صورت تکی در ساختارها دیده شوند، در وضعیت «۱» قرار می‌دهیم و اگر به‌صورت چندتایی اتم‌ها در ساختار مولکول‌ها مشاهده گردند، در وضعیت «۲» قرار می‌دهیم. همان‌طور که در شکل شماره ۲ مشخص است، همه اتم‌های نام‌برده، به‌جز P، در وضعیت «۲» قرار داده شده است؛ همچنین توصیفگرهای مبتنی بر SMILES با رنگ سبز و توصیفگرهای مبتنی بر گراف مولکولی با رنگ صورتی مشخص گردیده است و توصیفگرهایی که کنار آن‌ها تیک خورده است، در مدل‌سازی وارد شده‌اند.

شکل شماره ۲. نمایش محیط نرم افزار کورال ۲۰۲۰ برای مدل سازی QSAR

انتخاب شده نشان می دهد:

معادله زیر توصیفگرهای بهینه مولکولی را برای مدل ترکیبات BCR-ABL تیروزین کیناز علیه سرطان

$$DCW(T, N) = \sum CW(SSS_k) + CW(BOND) + CW(NOSP) + CW(HALO) + CW(HARD) + CW(C5) + CW(C6) + \sum CW(pt2_k) + \sum CW(pt3_k) + \sum CW(S2_k) + \sum CW(nn_k) + CW(APP) \quad (2)$$

توصیفگرهای مبتنی بر SMILES است که نشان دهنده حضور چندتایی از جفت اتم های N, Br, Cl, O, S است. $CW(X)$ وزن همبستگی یک توصیفگر خاص را نشان می دهد. مؤلفه T ضریب آستانه است و از آن برای تشخیص توصیفگرهای مختلف مولکولی به دست آمده از SMILES به دو دسته فعال و نادر استفاده می شود و N تعداد دوره های بهینه سازی مونت کارلو را نشان می دهد (۱۸). توضیحات کاملی از توصیفگرها همراه با رسم شکل در فایل راهنمای نرم افزار کورال وجود دارد (<http://www.insilico.eu/coral>).

ضریب همبستگی میان pK_i و $(\wedge_)$ $DCW(T^*, N^*)$ یک تابع ریاضی از وزن های همبستگی و با تابع هدف شاخص ایدئال همبستگی برای یافتن وزن های همبستگی توصیفگرهای SMILES است (۱۹). وزن هایی که بیشترین ضریب همبستگی را دارند، با روش مونت کارلو به دست می آیند. وزن همبستگی از نظر داده های

در رابطه شماره ۲، SSSK نشان دهنده حضور قطعات سه عضوی در نمایش SMILES است؛ به عنوان مثال، برای مولکول بنزن با اسمایلز (c1ccccc1)، توصیفگرهای SSSK شامل «c1c»، «lcc»، «ccc» و... است. BOND نشان دهنده حضور داشتن یا نداشتن باندهای مختلف مانند پیوند دو گانه (=)، سه گانه (#) و استریشیمی (@, @, @) است. NOSP نشان دهنده حضور داشتن یا نداشتن اتم های نیتروژن، اکسیژن، گوگرد و فسفر در ساختار SMILES ترکیبات است. HALO نشان دهنده حضور داشتن یا نداشتن هالوژن در ترکیبات است. HARD مربوط به BOND, NOSP و HALO در SMILES ها است. در اینجا C5 و C6 نشان دهنده وجود حلقه های پنج تایی و شش تایی است. pt2k و pt3k نشان دهنده تعداد مسیرهای طول ۲ و ۳ است که از رأس k شروع می شود و S2K نشان دهنده لایه های ظرفیت درجه ۲ است که از رأس K شروع می گردند. در معادله بالا، توصیفگر APP از دسته

بالا ذکر شد، برای معادله شماره ۳، با استفاده از روش بهینه‌سازی مونت کارلو محاسبه می‌شود. برای ساخت مدل QSAR در نرم‌افزار CORAL، سه نوع تابع هدف یعنی ۱. طرح کلاسیک، ۲. تعادل همبستگی و ۳. شاخص ایدئال همبستگی (IIC) طراحی شده است. شاخص ایدئال همبستگی به‌عنوان یک معیار جدید برای قضاوت درباره پتانسیل پیش‌بینی مدل‌های توسعه یافته QSAR به کار می‌رود.

در این مطالعه از دو تابع هدف برای بهینه‌سازی مونت کارلو استفاده شده است: تعادل همبستگی بدون IIC، (TF1) و تعادل همبستگی با IIC (TF2). وزن‌های همبستگی هر توصیفگر با استفاده از روش مونت کارلو، به‌منظور به‌دست آوردن حداکثر مقدار تابع هدف (TF) بهینه می‌گردند:

$$TF_1 = R_{TRN} + R_{iTRN} - |R_{TRN} - R_{iTRN}| \times Const \quad (4)$$

$$TF_2 = TF_1 + IIC_{CAL} \times Const \quad (5)$$

Const نیز ثابت تجربی است.

به‌منظور محاسبه IIC برای مجموعه کالیبراسیون، از معادله ریاضی زیر استفاده می‌شود:

$$IIC = R_{CAL} \times \frac{\min(-MAE_{CAL}, +MAE_{CAL})}{\max(-MAE_{CAL}, +MAE_{CAL})} \quad (6)$$

کالیبراسیون است. MAE میانگین خطای مطلق است که با روابط زیر محاسبه می‌گردد:

$$-MAE_{CLB} = -\frac{1}{N} \sum_{y=1}^{N^-} |\Delta_k| \quad \Delta_k < 0, \quad N^- \text{ is the number of } \Delta_k < 0 \quad (7)$$

$$+MAE_{CLB} = +\frac{1}{N} \sum_{y=1}^{N^+} |\Delta_k| \quad \Delta_k \geq 0, \quad N^+ \text{ is the number of } \Delta_k \geq 0 \quad (8)$$

$$\Delta_k = Obs_k - Calc_k \quad (9)$$

توسعه یافته QSAR توصیه می‌گردد.

تجزیه و تحلیل دامنه کاربرد به ما اجازه می‌دهد تا بررسی کنیم که آیا مدل QSAR ساخته شده می‌تواند برای هر مجموعه‌ای از مولکول‌ها اعمال شود یا خیر. پیش‌بینی یک نقطه پایانی مدل شده با استفاده از QSAR تنها زمانی مناسب است که مولکول پیش‌بینی شده در محدوده دامنه کاربرد مدل توسعه یافته QSAR قرار گیرد؛ بنابراین، دامنه کاربرد را می‌توان به‌عنوان اطلاعات فضای فیزیکوشیمیایی، بیولوژیکی

عددی آمار بهتری را برای مجموعه‌های کالیبراسیون ارائه می‌دهند و از مجموعه آموزش و آموزش منفعل محاسبه می‌شوند. در نرم‌افزار CORAL معادله نهایی مدل به‌صورت زیر است:

$$pK_i = C_0 + C_1 \times DCW(T, N) \quad (3)$$

مؤلفه‌های C_0 و C_1 اعداد ثابت‌اند و برای هر مدل، پس از انجام بهینه‌سازی ضرایب و ایجاد رابطه خطی میان متغیر وابسته و DCW مربوط به مولکول‌ها به‌دست می‌آید. DCW مجموع ضرایب همبستگی همه توصیفگرهای مولکولی مدل برای هر ترکیب است. درواقع، به کمک معادله بالا می‌توان با داشتن DCW هر ساختار، فعالیت (pKi) آن را پیشگویی کرد.

بهینه‌سازی وزن‌ها به روش مونت کارلو:

مقادیر عددی اوزان همبستگی توصیفگر بهینه که در

RTRN و RiTRN ضرایب همبستگی میان مقادیر مشاهده شده و محاسبه شده توصیفگر بهینه از فعالیت یا خاصیت برای مجموعه آموزش و آموزش منفعل است.

در این معادله، RCAL ضریب همبستگی میان مقادیر آزمایشی و پیش‌بینی شده نقطه انتهایی برای مجموعه

متغیر K از ۱ تا N تغییر می‌کند و Obs_k و Calc_k

مقادیر آزمایشی و محاسباتی نقطه پایانی (pIC50) است.

دامنه کاربرد مدل:

مؤلفه‌ها یا توصیفگرهای استفاده شده در توسعه مدل‌های QSAR باید تجزیه و تحلیل شوند. تفسیر مکانیکی مدل‌های QSAR به درک تأثیر توصیفگرها در پیش‌بینی فعالیت یا نقطه پایانی کمک می‌کند. بر اساس اصل سوم OECD، دامنه کاربرد (AD) برای اعتبارسنجی مدل‌های

اگر ترکیبی خارج از محدوده دامنه کاربرد قرار گیرد، به عنوان یک داده پرت کدگذاری می‌گردد. در مدل ساخته شده توسط CORAL از نقض آماری برای تعریف دامنه کاربرد استفاده می‌شود. «نقص آماری» Defect_{AK} بر اساس معادله ریاضی زیر محاسبه می‌گردد:

$$\begin{aligned} \text{Defect}_{AK} &= \frac{|P_T(A_K) - P_{CAL}(A_K)|}{N_T(A_K) + N_{CAL}(A_K)} & \text{If } A_K > 0 \\ \text{Defect}_{AK} &= 1 & \text{If } A_K = 0 \end{aligned} \quad (10)$$

اعتبارسنجی مدل:

در این تحقیق برای ارزیابی قابلیت اطمینان و پیشگویی مدل‌های ایجادشده، از سه روش استفاده شده است:

- اعتبارسنجی داخلی یا متقابل توسط معیارهای R², IIC، CCC، Q² و F با استفاده از داده‌های مجموعه آموزشی؛
- اعتبارسنجی خارجی که توسط معیارهای Q²_{F1}، Q²_{F2}، Q²_{F3}، MAE، s، Cr_p²، \overline{r}_m^2 و Δr_m^2 استفاده از داده‌های دسته آزمون صورت می‌گیرد و تصادفی سازی داده‌ها یا درهم آمیختگی Y که با استفاده از داده‌های دسته آزمون انجام می‌شود. روابط ریاضی این شاخص‌های آماری در مقالات پیشین ارائه شده است (۲۰-۲۲) و همچنین در جدول شماره ۱ آمده است.

یا ساختاری بیان کرد که بر اساس آن، مجموعه آموزشی مدل ساخته می‌شود و مدل برای پیش‌بینی مولکول‌های جدید در حوزه‌های خاص اعمال می‌گردد.

توزیع توصیفگرهای SMILES در مجموعه‌های آموزش و کالیبراسیون برای تعیین دامنه کاربرد در مدل‌های ساخته شده QSAR توسط نرم‌افزار CORAL استفاده می‌شود.

مؤلفه‌های P_{CAL}(A_K) و P_T(A_K) به ترتیب احتمال یک توصیفگر در دسته آموزش و کالیبراسیون هستند. نقص آماری را می‌توان به عنوان مجموع نقص‌های آماری همه ویژگی‌های موجود در نماد SMILES محاسبه کرد.

$$\text{Defect}_{\text{compound}} = \sum_{k=1}^{N_A} \text{Defect}_{A_K} \quad (11)$$

NA تعداد توصیفگرهای فعال برای یک ترکیب است. در این محاسبات آماری، یک ترکیب در محدوده دامنه کاربرد قرار می‌گیرد، اگر نامساوی زیر برقرار گردد:

$$\text{Defect}_{\text{compound}} < 2 \times \overline{\text{Defect}}_T \quad (12)$$

که $\overline{\text{Defect}}_T$ میانگین نقص آماری برای دسته آموزش است.

جدول شماره ۱. معادله ریاضی معیارهای مختلف ارزیابی قابلیت اطمینان و پیشگویی مدل‌های QSAR

معیار پتانسیل پیشگویی	نوع اعتبارسنجی		
$R^2 = 1 - \frac{\sum(Y_{\text{obs}} - Y_{\text{prd}})^2}{\sum(Y_{\text{obs}} - \bar{Y})^2}$	داخلی	(۲۳)	
$Q^2 = 1 - \frac{\sum(Y_{\text{prd}} - Y_{\text{obs}})^2}{\sum(Y_{\text{obs}} - \bar{Y}_{\text{train}})^2}$			
$Q_{F1}^2 = 1 - \frac{\sum(Y_{\text{per}(\text{test})} - Y_{\text{obs}(\text{test})})^2}{\sum(Y_{\text{obs}(\text{test})} - \bar{Y}_{\text{train}})^2}$	خارجی	(۲۴)	
$Q_{F2}^2 = 1 - \frac{\sum(Y_{\text{prd}(\text{test})} - Y_{\text{obs}(\text{test})})^2}{\sum(Y_{\text{obs}(\text{test})} - \bar{Y}_{\text{ext}})^2}$			
$Q_{F3}^2 = 1 - \frac{\sum(Y_{\text{prd}(\text{test})} - Y_{\text{obs}(\text{test})})^2 / n_{\text{ext}}}{\sum(Y_{\text{obs}(\text{test})} - \bar{Y}_{\text{train}})^2 / n_{\text{train}}}$			
$R_m^2 = R^2 \times \left(1 - \sqrt{R^2 - R_0^2}\right)$			(۲۵)
$CCC = \frac{2 \sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2 + n((\bar{X} - \bar{Y})^2)}$			(۲۶)
$MAE = \frac{1}{n} \times \sum Y_{\text{obs}} - Y_{\text{prd}} $		(۲۳)	

$C_{R_p^2} = R\sqrt{(R^2 - R_p^2)}$	آزمون تصادفی سازی	(۲۷)
-------------------------------------	-------------------	------

مدل‌های QSAR برای تمایل اتصال مشتقات ایماتینیب در برابر BCR-ABL تیروزین کیناز بر اساس تابع هدف شاخص ایدئال همبستگی (IIC) با استفاده از تابع هدف TF2 و روش مونت کارلو برای سه توزیع تصادفی داده‌ها به دسته‌های آموزش، آموزش منفعل، کالیبراسیون و آزمون محاسبه‌شده عبارت‌اند از:

توانایی پیشگویی مدل QSAR در صورتی قابل قبول است که مقدار عددی Q2 و R2 بیشتر از ۰/۷ باشد (۲۸) و از نظر اعتبارسنجی خارجی، اگر CCC بزرگ‌تر از ۰/۸۵ باشد، مدل قابلیت پیش‌بینی خوبی دارد (۲۹)؛ همچنین اگر مقادیر $r^2_m > 0.5$ و $\Delta F_m < 0.2$ باشد، مدل را می‌توان به‌عنوان یک مدل قابل اعتماد بیان کرد.

یافته‌های پژوهش

شکاف ۱:

$$pk_i = 4.539 (\pm 0.011) + 0.096 (\pm 0.001) \times DCW(1, 7) \quad (13)$$

شکاف ۲:

$$pk_i = 5.086 (\pm 0.012) + 0.091 (\pm 0.001) \times DCW(1, 7) \quad (14)$$

شکاف ۳:

$$pk_i = 4.539 (\pm 0.012) + 0.0740 (\pm 0.001) \times DCW(1, 7) \quad (15)$$

مختلف موافقت دارند. مقادیر pk_i تجربی و پیش‌بینی شده و همچنین دامنه کاربرد ترکیبات انتخاب شده در جداول S1 آورده شده است. پیش‌بینی‌های یک مدل QSAR زمانی واقعی تر هستند که مولکول‌های پیش‌بینی شده در حوزه دامنه کاربرد باشند. حدود ۸۲ درصد، ۸۰ درصد و ۸۰ درصد از ترکیبات در دامنه کاربرد مدل‌های ۱۳ تا ۱۵ به ترتیب قرار دارند و این موضوع نشان می‌دهد، هر سه مدل توانایی پیش‌بینی بیش از ۸۰ درصد از داده‌های جدید را دارند. در ادامه، آزمون Y برای مجموعه اعتبارسنجی نیز برای تأیید قدرت پیش‌بینی مدل‌های QSAR تولیدشده به کار گرفته شد. برای آزمون تصادفی‌سازی Y اگر $cR2p$ بیشتر از ۰/۵ باشد، مدل‌های ایجادشده ارتباط شانس ندارند (۳۲). مقادیر $cR2p$ برای مدل‌های ایجادشده بیشتر از ۰/۷۲ است.

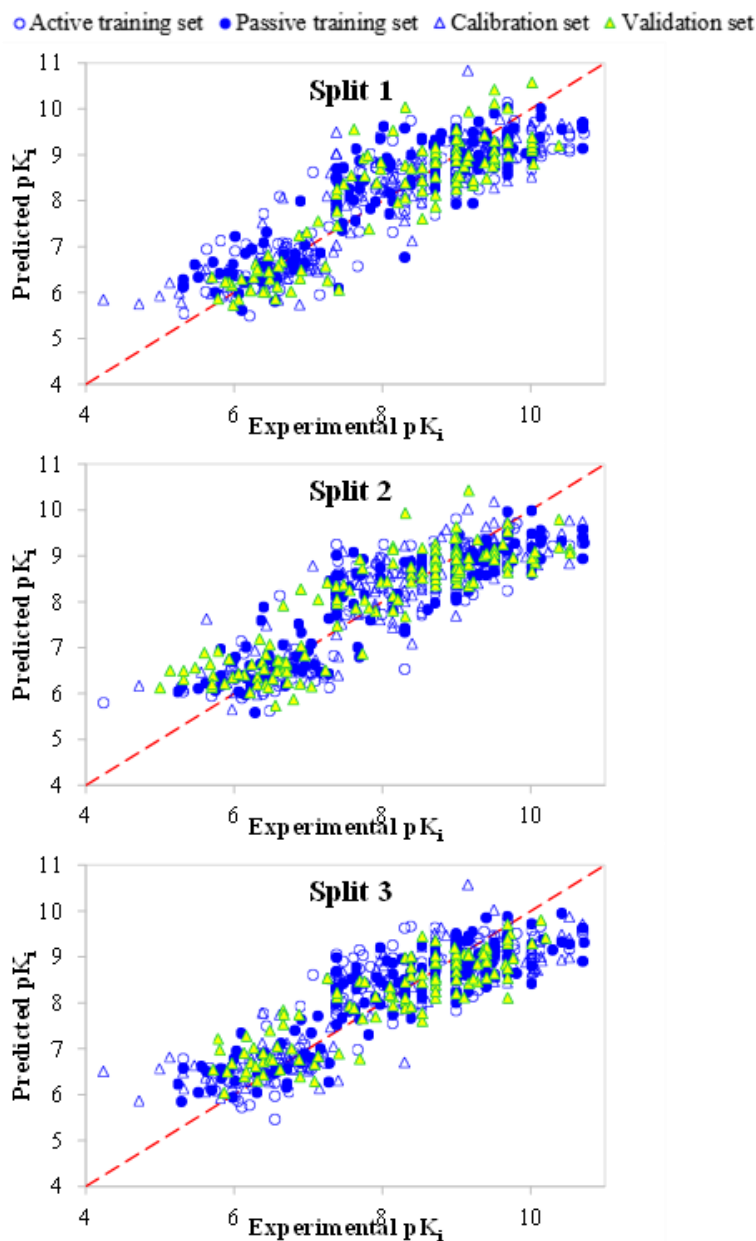
نتایج این مدل‌های QSAR ایجادشده در جدول شماره ۲ خلاصه شده است. همه مدل‌های طراحی شده QSAR از نظر آماری قابل اعتماد بودند و مقادیر عددی کمیت‌های آماری در محدوده‌های قابل قبولی که در منابع گزارش شده است، قرار داشتند (۳۰، ۳۱). اضافه کردن IIC به مدل توانایی آن را در پیش‌بینی pk_i افزایش می‌دهد. نتیجه IIC به مقدار عددی ضریب همبستگی (R) و میانگین خطای مطلق (MAE) بستگی دارد؛ بنابراین، IIC مؤلفه مهم‌تری نسبت به R و MAE است؛ از این رو، می‌توان بیان کرد که مدل‌های QSAR ساخته شده با IIC از نظر آماری معنادار و قوی‌تر هستند. نتایج نشان داد که هر سه مدل QSAR برای مجموعه‌های آموزش، کالیبراسیون، آزمون و اعتبارسنجی از نظر کیفیت آماری قابلیت پیش‌بینی قوی دارند و با الزامات مؤلفه‌های اعتبارسنجی

جدول شماره ۲. مشخصات آماری برای مدل‌های QSAR ایجاد شده برای شکافت‌های مختلف ۱ تا ۳

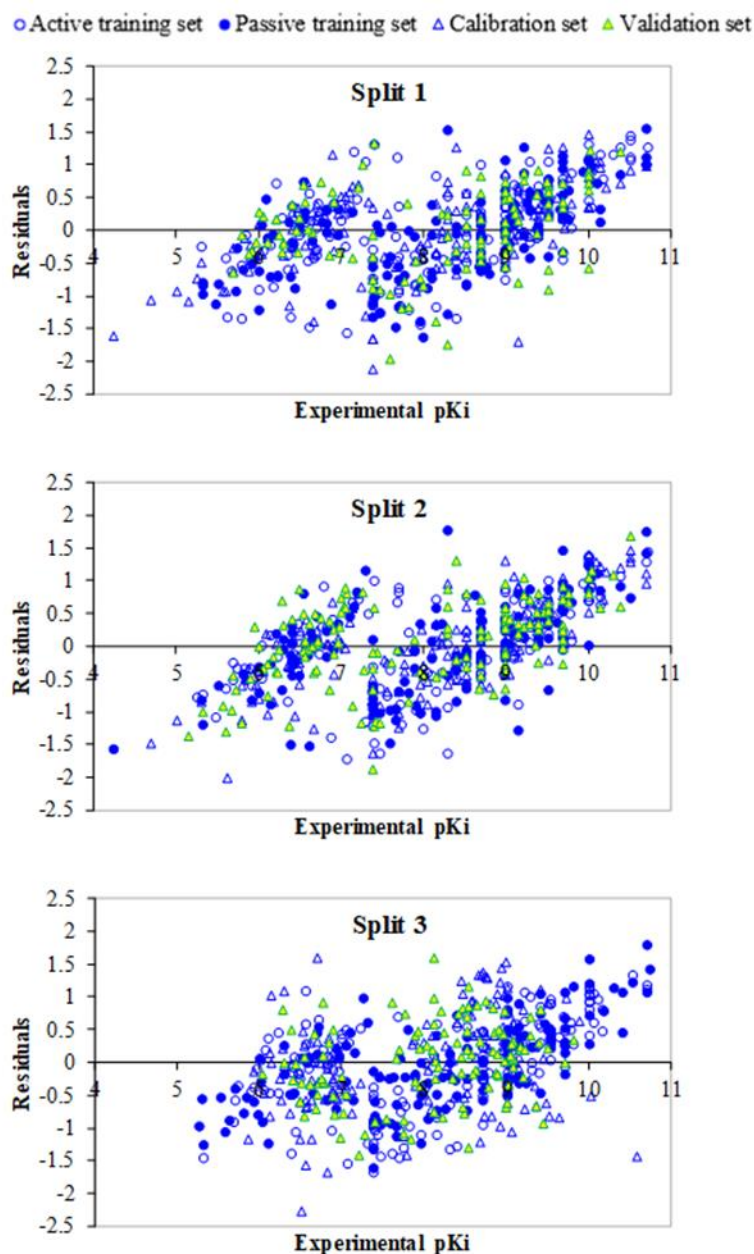
F	MAE	S	Δr_m^2	\bar{r}_m^2	CR_p^2	R_m^2	Q_{F3}^2	Q_{F2}^2	Q_{F1}^2	Q^2	HC	CCC	R^2	n	دسته	شکافت
۴۹۸	۰/۵۲۴۰	۰/۶۶۰			۰/۷۵۸۴					۰/۷۵۵۷	۰/۸۲۹۵	۰/۸۶۴۶	۰/۷۶۱۵	۱۵۸	آموزش	۱
۴۷۵	۰/۵۵۴۰	۰/۶۸۵			۰/۷۵۶۶				۰/۷۵۵۲	۰/۷۹۹۰	۰/۸۵۸۷	۰/۷۶۱۴	۱۵۱	منفعل آموزش		
۵۴۰	۰/۵۲۱۰	۰/۶۶۵	۰/۶۷۷۳	۰/۶۶۳۶	۰/۷۹۲۰	۰/۶۶۳۶	۰/۷۶۶۸	۰/۷۹۳۹	۰/۷۹۵۰	۰/۷۸۹۶	۰/۸۹۲۴	۰/۸۷۹۵	۰/۷۹۶۴	۱۴۰	کالیبراسیون	
۳۴۹	۰/۴۵۸۷	۰/۶۳۶	۰/۰۰۵۱	۰/۶۸۳۵		۰/۶۷۷۳				۰/۷۶۲۰	۰/۸۱۱۳	۰/۸۷۷۵	۰/۷۷۰۴	۱۰۶	آزمون	
۳۸۰	۰/۵۴۶۰	۰/۶۸۸			۰/۸۳۲۸					۰/۷۳۵۰	۰/۸۱۱۶	۰/۸۵۲۱	۰/۷۴۲۲	۱۳۴	آموزش	۲
۴۰۳	۰/۵۵۶۰	۰/۶۹۱			۰/۷۶۶۱					۰/۷۳۴۷	۰/۷۸۳۷	۰/۸۴۹۸	۰/۷۴۲۳	۱۴۲	آموزش منفعل	
۵۴۹	۰/۵۴۲۰	۰/۶۸۲	۰/۱۴۶۲	۰/۷۲۳۰	۰/۷۷۰۳	۰/۵۶۳۱	۰/۷۴۷۴	۰/۷۶۸۱	۰/۷۶۹۹	۰/۷۸۹۰	۰/۸۹۰۶	۰/۸۵۰۵	۰/۷۹۴۵	۱۴۴	کالیبراسیون	
۴۸۷	۰/۵۳۵۳	۰/۶۴۹	۰/۱۹۰۶	۰/۶۷۸۷		۰/۶۷۸۷				۰/۷۷۹۳	۰/۸۳۷۹	۰/۸۷۴۵	۰/۷۸۵۵	۱۳۵	آزمون	
۴۰۱	۰/۵۶۳۰	۰/۷۰۳			۰/۷۲۱۵					۰/۷۷۹۳	۰/۸۴۰۸	۰/۸۷۴۵	۰/۷۲۵۳	۱۵۴	آموزش	۳
۵۰۴	۰/۵۳۴۰	۰/۶۵۴			۰/۷۷۳۱					۰/۷۹۸۰	۰/۸۵۷۳	۰/۸۶۴۱	۰/۷۷۶۶	۱۴۷	آموزش منفعل	
۵۲۵	۰/۵۹۲۰	۰/۷۳۸	۰/۲۳۹۳	۰/۵۵۹۰	۰/۷۸۵۸	۰/۵۵۹۰	۰/۷۰۱۲	۰/۷۶۷۱	۰/۷۷۰۸	۰/۷۴۰۳	۰/۸۶۹۹	۰/۸۵۰۵	۰/۷۸۷۱	۱۴۴	کالیبراسیون	
۳۹۴	۰/۴۸۵۹	۰/۶۰۱	۰/۲۲۷۶	۰/۵۸۳۹		۰/۵۸۳۵				۰/۷۷۷۷	۰/۷۸۱۰	۰/۸۵۷۵	۰/۷۸۴۸	۱۱۰	آزمون	

شکل شماره ۳ نمایش گرافیکی مقادیر داده‌های تجربی pK_i در برابر pK_i محاسبه شده سه مدل را نشان می‌دهند. مشاهده می‌شود که pK_i محاسبه شده برای مجموعه داده‌ها با مقادیر تجربی pK_i از مدل‌های QSAR محاسبه شده با همبستگی خوبی دارد. این نتایج تأیید می‌کنند که همه مدل‌های QSAR ساخته شده قابل اعتمادند و به خوبی برازش شده‌اند.

شکل شماره ۳ نمایش گرافیکی مقادیر داده‌های تجربی pK_i در برابر pK_i محاسبه شده سه مدل را نشان می‌دهند. مشاهده می‌شود که pK_i محاسبه شده برای مجموعه داده‌ها با مقادیر تجربی pK_i از مدل‌های QSAR محاسبه شده با همبستگی خوبی دارد. این نتایج تأیید می‌کنند که همه مدل‌های QSAR ساخته شده قابل اعتمادند و به خوبی برازش شده‌اند.



شکل شماره ۳. نمایش گرافیکی همبستگی میان pK_i تجربی و پیش‌بینی شده توسط مدل‌های ۱ تا ۳، مربوط به دسته‌های آموزش فعال، آموزش منفعل، کالیبراسیون و آزمون



شکل شماره ۴. نمایش گرافیکی باقیمانده‌ها برحسب pKi تجربی برای مدل‌های ۱ تا ۳، مربوط به دسته‌های آموزش فعال، آموزش منفعل، کالیبراسیون و آزمون

تفسیر مکانیکی مدل: اصل پنجم سازمان همکاری و توسعه اقتصادی (OECD) مربوط به «تفسیر مکانیکی» مدل ایجاد شده است، اگر تفسیر مدل امکان‌پذیر باشد که نشان می‌دهد آیا از چنین مدل‌هایی می‌توان برای به‌دست آوردن ویژگی‌های مولکولی مسئول افزایش و کاهش یک نقطه پایانی استفاده کرد. هدف این ارزیابی تعامل مکانیکی میان توصیفگرهای استفاده‌شده در یک مدل و نقطه پایانی پیشگویی شده است و این تعاملات باید در صورت امکان،

شکل شماره ۴ نمایش گرافیکی باقیمانده‌ها، اختلاف مقدار مشاهداتی و پیشگویی شده توسط مدل را برای مدل‌های ۱ تا ۳ نشان می‌دهد. همان‌طور که در هر سه شکل مشاهده می‌شود، باقیمانده‌ها در اطراف محور x در یک نوار افقی متمرکز در اطراف صفر و واریانس ثابت توزیع شده‌اند. این شکل نشان می‌دهد که در مدل‌سازی خطای سیستماتیک وجود ندارد و خطای مدل توزیع نسبتاً یکنواختی برای همه مقادیر pKi دارد.

افزاینده و کاهنده تمایل اتصال آمده است. باید توجه داشت که فراوانی این ویژگی‌های مولکولی در مجموعه آموزش و همچنین مجموعه کالیراسیون باید مورد توجه قرار گیرد. ظاهراً ویژگی ساختاری با فرکانس‌های کوچک در آموزش و مجموعه کالیراسیون را نمی‌توان به‌عنوان شاخص قابل اعتماد برای پیش‌بینی تمایل اتصال ترکیبات در برابر BCR-ABL تیروزین کیناز در نظر گرفت.

توصیفگرهای NNC-C...312، NNC-C...330، PT2-N...4...، VS2-C...9...، c...2...c...، VS2-C...8...، PT3-C...6...، c...c...2...، PT3-N...4... و C6...AH.3... از جمله توصیفگرهای افزاینده هستند؛ همچنین مهم‌ترین توصیفگر کاهنده مشترک در هر سه شکافت C6...AH.4...، 2...c...n... و O...C...C... است که تفسیر آن‌ها در جدول شماره ۳، در ستون آخر آمده است.

مستند کردند. هنگامی که تفسیر مکانیکی یک مدل QSAR ساخته‌شده با نظریه‌ها و سازوکارهای موجود سازگار باشد، به طراحی و توسعه داروها یا مولکول‌های جدید کمک می‌کند. در روش بهینه‌سازی مونت کارلو در نرم‌افزار کورال، با داشتن اطلاعات مربوط به چندین اجرا از بهینه‌سازی مونت کارلو می‌توان سه دسته از توصیفگرهای مولکولی استخراج کرد (۳۴، ۳۳): ۱. توصیفگرهای افزاینده فعالیت: توصیفگرهایی که ضریب همبستگی آن در همه اجراها مثبت است؛ ۲. توصیفگرهای کاهنده فعالیت: توصیفگرهایی که ضریب همبستگی آن در همه اجراها منفی است؛ ۳. توصیفگرهایی که نقش مشخصی ندارند: توصیفگرهایی که در بعضی از اجراهای مونت کارلو دارای ضریب مثبت و در برخی از اجراها دارای ضریب منفی اند (۳۵).

در جدول شماره ۳ نمونه‌هایی از توصیفگرهای

جدول شماره ۳. توصیفگرهای مولکولی افزاینده pKi و کاهنده pKi مشترک در هر سه مدل ارائه‌شده

شماره	توصیفگر	توضیحات
توصیفگرهای افزاینده		
۱	NNC-C...330.	کد نزدیک‌ترین همسایگی برای کربن برابر با ۳۳۰
۲	NNC-C...312.	کد نزدیک‌ترین همسایگی برای کربن برابر با ۳۱۲
۳	VS2-C...7...	پوسته ظرفیت درجه دوم برای اتم کربن برابر با ۷
۴	PT2-N...4...	تعداد مسیرهای به طول ۲ که برابر با ۴ است و از رأس اتم نیتروژن شروع شده است.
۵	VS2-C...9...	پوسته ظرفیت درجه دوم برای اتم کربن برابر با ۹
۶	c...2...c...	حضور دو کربن آروماتیک جدا در حلقه شماره ۲
۷	VS2-C...8...	پوسته ظرفیت درجه دوم برای اتم کربن برابر با ۸
۸	PT3-C...6...	تعداد مسیرهای به طول ۳ که برابر با ۶ است و از رأس اتم کربن شروع شده است.
۹	c...c...2...	حضور دو کربن آروماتیک متوالی در حلقه شماره ۲
۱۰	PT3-N...4...	تعداد مسیرهای به طول ۳ که برابر با ۴ است و از رأس اتم نیتروژن شروع شده است.
۱۱	C6...AH.3...	وجود سه حلقه شش عضوی آروماتیک دارای هترواتم
توصیفگرهای کاهنده		
۱	C6...AH.4...	وجود ۴ حلقه شش عضوی آروماتیک دارای هترواتم
۲	c...n...2...	حضور کربن و نیتروژن آروماتیک در حلقه دوم
۳	O...C...C...	حضور اکسیژن آلیفاتیک و به دنبال آن، دو کربن آلیفاتیک متوالی

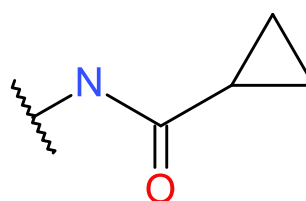
pKi بین ۴/۲۴ و ۱۰/۷۳ داشتند، مشخص شد مشتقاتی که ۳ شش ضلعی آروماتیک و هترواتم دارند، معمولاً دارای pki بالایی (۸/۸ تا ۱۰) هستند. تعدادی از این ساختارها در جدول شماره ۴ گزارش شده است.

به منظور بررسی تأثیر توصیفگرهای افزایشده و کاهشده بر روی فعالیت ترکیبات، در ادامه تعدادی از این توصیفگرها با مثال شرح داده شده است. یکی از توصیفگرهای افزایشده، حضور C6...AH.3 یعنی حضور ۳ عدد شش ضلعی آروماتیک است که هترواتم دارد و با بررسی ۵۵۵ داده که

جدول شماره ۴: نمونه‌هایی از ترکیبات دارای سه حلقه شش‌تایی آروماتیک شامل هترواتم (C6...AH.3) از دسته توصیفگرهای افزایشده

شماره	ساختار	pKi	شماره	ساختار	pKi
۲۸۵		۸/۷۷	۳۴۰		۸/۷۹
۵۲۴		۸/۹۸	۲۸۸		۹
۲۸		۹/۴	۱۸۲		۹/۵۲
۱۹۵		۹/۶۶	۳۰۶		۹/۷
۱۱۴		۹/۷۷	۱۱۶		۱۰

از مؤلفه‌های مهم دیگر توصیفگر NNC-C312 است. این توصیفگر به این معنی است که ۳ اتم در همسایگی اتم کربن وجود دارد که تنها ۱ عدد آن اتم کربن باشد. این توصیفگر افزاینده در بیشتر ترکیبات با pKi بالا، یعنی از pKi 7 به بالا وجود دارد. قطعه مولکولی نمایش داده شده در شکل شماره ۵، نمونه واضحی از توصیفگر NNC-C312 است. جدول شماره ۴ تعدادی از ترکیبات با pKi بالا را نشان می‌دهد که این موضوع را تأیید می‌کند.



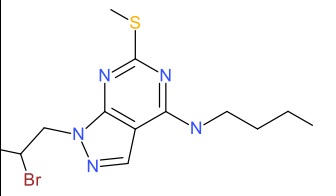
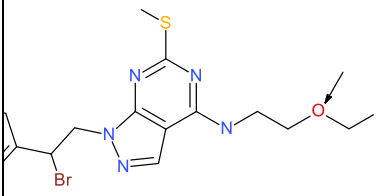
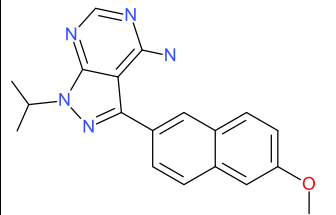
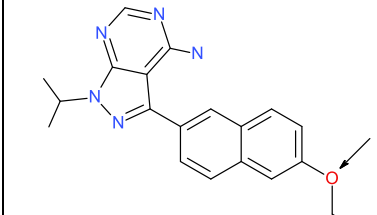
شکل شماره ۳. قطعه مولکولی با اثر افزایشی بر روی pKi ترکیبات

یکی از توصیفگرهای کاهنده حضور اکسیژن

آلیفاتیک و به دنبال آن دو کربن آلیفاتیک متوالی (...C...C...O...) است. مثال‌هایی از اثر این توصیفگر در مولکول‌هایی که تنها اختلاف آن‌ها در این توصیفگر است، در جدول شماره ۵ آمده است. این توصیفگر کاهنده در جدول شماره ۵، با علامت فلش روی مولکول‌ها مشخص شده است. ساختارهایی که این توصیفگر را دارند، با ساختارهای مشابه خودشان در جدول شماره ۵ مقایسه شده‌اند؛ به عنوان مثال، مشتقات شماره ۳۸۳ و شماره ۸ از نظر ساختاری مشابه هستند و تفاوت آن‌ها در حضور اکسیژن در حلقه سیکلوپنتان است. حضور اکسیژن که باعث به وجود آمدن توصیفگر ...C...C...O... در ساختار ۳۸۳ شده است، موجب کاهش pKi از مقدار ۹/۳۰ به مقدار ۸/۷۸ گردیده است. درباره جفت ساختارهای مشابه دیگر مانند جفت ساختار ۲۸۶ و ۴۰۰، ۲۳۸، ۲۷۷ و ۲۵۶ و ۲۶، به همین ترتیب حضور این توصیفگر باعث کاهش چشمگیری در مقدار pKi شده است.

جدول شماره ۵. مثال‌هایی از تأثیر کاهندگی حضور اکسیژن آلیفاتیک و به دنبال آن، دو کربن آلیفاتیک متوالی (...C...C...O...) بر روی مقدار

شماره	ساختار	pKi	شماره	ساختار	pKi
۳۸۳		۸/۷۸	۸		۹/۳۰
۲۸۶		۷/۳۸	۴۰۰		۸/۱۱

۶/۴۹		۲۷۷	۶/۲۶		۲۳۸
۶/۱۲		۲۶	۵/۸۰		۲۵۶

شماره	توصیفگر	توضیحات
توصیفگرهای افزاینده		
۱	NNC-C...330.	کد نزدیک‌ترین همسایگی برای کربن برابر با ۳۳۰
۲	NNC-C...312.	کد نزدیک‌ترین همسایگی برای کربن برابر با ۳۱۲
۳	VS2-C...7...	پوسته ظرفیت درجه دوم برای اتم کربن برابر با ۷
۴	PT2-N...4...	تعداد مسیرهای به طول ۲ که برابر با ۴ است و از رأس اتم نیتروژن شروع شده است.
۵	VS2-C...9...	پوسته ظرفیت درجه دوم برای اتم کربن برابر با ۹
۶	c...2...c...	حضور دو کربن آروماتیک جدا در حلقه شماره ۲
۷	VS2-C...8...	پوسته ظرفیت درجه دوم برای اتم کربن برابر با ۸
۸	PT3-C...6...	تعداد مسیرهای به طول ۳ که برابر با ۶ است و از رأس اتم کربن شروع شده است.
۹	c...c...2...	حضور دو کربن آروماتیک متوالی در حلقه شماره ۲
۱۰	PT3-N...4...	تعداد مسیرهای به طول ۳ که برابر با ۴ است و از رأس اتم نیتروژن شروع شده است.
۱۱	C6...AH.3...	وجود سه حلقه شش عضوی آروماتیک دارای هترواتم
توصیفگرهای کاهنده		
۱	C6...AH.4...	وجود ۴ حلقه شش عضوی آروماتیک دارای هترواتم
۲	c...n...2...	حضور کربن و نیتروژن آروماتیک در حلقه دوم
۳	O...C...C...	حضور اکسیژن آلیفاتیک و به دنبال آن، دو کربن آلیفاتیک متوالی

طراحی مولکول‌های جدید:

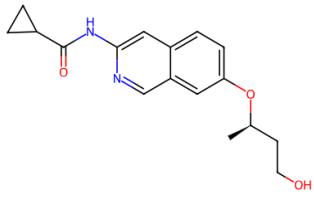
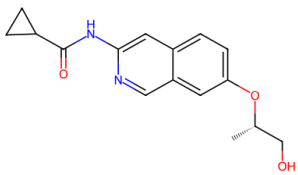
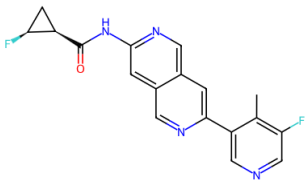
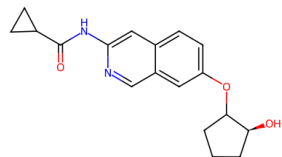
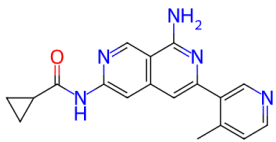
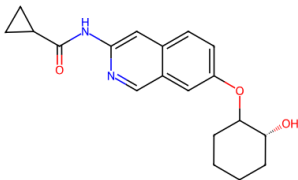
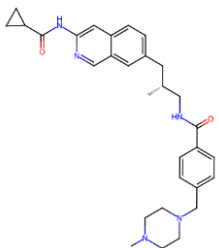
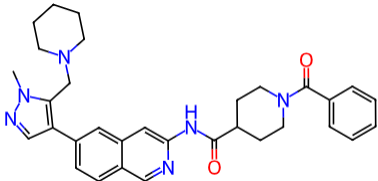
بر اساس نتایج تفسیر مدل و به کمک ساختار کلی مشتقاتی از ایماتینیب در مجموعه داده‌ها که pKi بالایی داشتند، تعدادی از مشتقات ایماتینیب که در داده‌هایی

گزارش شده، وجود نداشتند و دارای pKi میانگین پیش‌بینی شده بر اساس سه مدل بالاتر از ۸ بودند، از بانک داده‌های ChEMBL (<https://www.ebi.ac.uk/chembl>) استخراج و به عنوان نتایج مدل‌سازی گزارش شدند (جدول

شماره ۶). (تجربی برای آنها گزارش نشده است. می‌توان با همه ترکیبات پیش‌تر توسط سایر محققان سنتز گردیده‌اند؛ اما مقادیر تمایل اتصال تیروزین کیناز (Ki) کاملی از روش سنتز این ترکیبات به‌دست آورد.

جدول شماره ۶. ساختار مشتقات ایماتینیب جدید، ID مربوطه و نتایج پیشگویی pKi (میانگین حاصل از سه مدل ارائه‌شده) برای آنها که از بانک داده‌های ChEMBL جستجو شده‌اند.

ساختار و ID مربوط به ChEMBL	pKi (میانگین)	ساختار و ID مربوط به ChEMBL	pKi (میانگین)
CHEMBL3689721 	۹/۱۲	CHEMBL3689568 	۹/۲۵
CHEMBL3689564 	۸/۶۳	CHEMBL3689603 	۹/۴۶
CHEMBL3689722 	۹/۰۱	CHEMBL3689605 	۹/۱۴
CHEMBL3685170 	۸/۷۹	CHEMBL3689727 	۸/۸۴
CHEMBL3689577 	۸/۸۳	CHEMBL3689578 	۸/۸۳

<p>CHEMBL3685189</p> 	۸/۳۷	<p>CHEMBL3685158</p> 	۸/۳۵
<p>CHEMBL3689718</p> 	۸/۹۸	<p>CHEMBL3689569</p> 	۸/۱۱
<p>CHEMBL5195701</p> 	۹/۰۶	<p>CHEMBL3689570</p> 	۸/۱۶
<p>CHEMBL3685193</p> 	۸/۵۰	<p>CHEMBL4751301</p> 	۸/۹۷

بحث و نتیجه گیری

روش‌های محاسباتی یک روش قدرتمند برای طراحی/پیش‌بینی فعالیت داروها در مقابل بیماری‌های متفاوت، توجه بیشتری را به خود جلب کرده‌اند. در این مطالعه، مدل‌های QSAR برای پیش‌بینی تمایل اتصال (pKi) 555 ترکیب از مشتقات ایماتینیب به‌عنوان مهارکنندهٔ BCR-ABL تیروزین کیناز در درمان بیماری سرطان لوسمی مزمن، با استفاده از روش مونت کارلو و بر اساس معیار شاخص ایدئال همبستگی (IIC) ساخته شده است. ساخت مدل‌های QSAR توسط نرم‌افزار CORAL و با استفاده از توصیفگر هیبریدی بهینه که ترکیبی از SMILES و HFGs است، انجام گردیده است. در پایان، برخی از ویژگی‌های آماری مدل‌های CORAL به‌دست آمده برای توصیف pKi با برخی از مؤلفه‌های آماری مدل‌های دیگر که از قبل گزارش شده است،

مقایسه گردیده است (۱۰). مقایسه نشان می‌دهد که ویژگی‌های آماری مدل‌های CORAL طراحی شده از نظر کیفی، قابل مقایسه یا حتی بهتر از مدل‌های پیشنهاد شده در گزارش پیشین است. در مدل‌های QSAR پیشین از شبکه‌های عصبی (DNN) و مجموعه‌های ترکیبی توصیفگرهای D/3D/MD۲ به‌منظور پیش‌بینی تمایل اتصال ترکیبات استفاده شده است. آن‌ها نشان دادند که مدل‌های رگرسیون DNN برای مجموعه داده‌ها $R2 \geq 0.71$ و $MAE \leq 0.85$ را دارد. در این تحقیق، $R2 \geq 0.74$ و $MAE \leq 0.59$ برای مدل‌ها به‌دست آمد. علاوه بر این، مدل‌های QSAR ایجاد شده در این کار، با استفاده از معیارهای قوی مانند IIC، Q2F2، Q2F1، Q2F3، $cR2p$ و $R2m$ ارزیابی شده‌اند. در اینجا، کل مجموعه داده به چهار مجموعه (آموزش، آموزش منفعل، مجموعه کالیبراسیون و آزمون) و سه شکاف تقسیم گردید،

مشارکت نویسندگان

این مقاله برگرفته از رساله دکتری خانم شراره وردست باغ می‌باشد که آقایان دکتر علی الماسی‌راد و دکتر شهرام لطفی اساتید راهنما و سرکار خانم دکتر شهین احمدی استاد مشاور در تهیه و تدوین آن مشارکت داشتند.

درحالی‌که در گزارش پیشین، مجموعه داده به دو مجموعه (مجموعه آموزشی و آزمایشی) و یک شکاف تقسیم شد؛ در نتیجه، ما نشان داده‌ایم که پتانسیل پیش‌بینی مدل جهانی QSAR توسعه یافته می‌تواند با مدل‌های QSAR گزارش شده محلی مرتبط باشد.

در کار حاضر، روش بهینه‌سازی مونت کارلو برای ساخت سه مدل QSAR برای پیش‌بینی تمایل اتصال (pKi) 555 ترکیب از مشتقات ایماتینیب به‌عنوان مهارکننده BCR-ABL تیروزین کیناز در درمان بیماری سرطان لوسمی مزمن، با استفاده از TF2 با IIC استفاده گردید. مدل‌های ساخته شده با استفاده از شاخص‌های اعتبارسنجی داخلی و خارجی و آزمون تصادفی سازی Y ارزیابی شدند که همه مدل‌های ساخته شده قابل اطمینان، پیش‌بینی پذیر و قابل تکرار بودند. مدل‌های QSAR پیشنهادی ضریب همبستگی (R²) رضایت‌بخش و در محدوده ۰/۸۷۷۵-۰/۸۵۷۵ برای دسته آزمون هستند. با استفاده از مدل‌های ساخته شده، توصیفگرهای افزاینده یا کاهنده تمایل اتصال (Ki) ترکیبات مطالعه شده، با بررسی وزن همبستگی ویژگی‌های مختلف مولکولی در چندین مرحله اجرای روش مونت کارلو به دست آمده است. با استفاده از نتایج مدل‌سازی، ترکیباتی جدید با فعالیت بیشتر علیه بیماری لوسمی میلوئید مزمن پیشنهاد شده است.

سپاس‌گزاری

از همه‌ی اشخاصی که در این پژوهش همکاری داشتند، تشکر و قدردانی می‌شود.

تعارض منافع

نویسندگان اعلام می‌نمایند که هیچگونه تعارض منافی وجود ندارد.

کد اخلاق

این مطالعه دارای تاییدیه اخلاق به شماره IR.IAU.PS.REC.1400.260 از کمیته اخلاق در پژوهش دانشگاه علوم پزشکی آزاد اسلامی تهران می‌باشد.

حمایت مالی

پژوهش حاضر از هیچ سازمان و ارگانی حمایت مالی دریافت نکرده است.

References

- Goudzal A, El Aissouq A, El Hamdani H, Ouammou A. QSAR modeling, molecular docking studies and ADMET prediction on a series of phenylaminopyrimidine-(thio) urea derivatives as CK2 inhibitors. *Mater Today* 2022;51:1851-62. doi: 10.1016/j.matpr.2020.08.044.
- Yang M, Xi Q, Jia W, Wang X. Structure-based analysis and biological characterization of imatinib derivatives reveal insights towards the inhibition of wild-type BCR-ABL and its mutants. *Bioorg Med Chem Lett* 2019;29:126758. doi:10.1016/j.bmcl.2019.126758.
- Oliveira A, Moura S, Pimentel L, Neto J, Dantas R, Silva-Jr F, et al. New imatinib derivatives with antiproliferative activity against A549 and K562 cancer cells. *Molecules* 2022;27:750. doi: 10.3390/molecules27030750.
- An X, Tiwari AK, Sun Y, Ding P-R, Ashby Jr CR, Chen Z-S. BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. *Leuk Res* 2010;34:1255-68. doi:10.1016/j.leukres.2010.04.016.
- Luo H, Quan H, Xie C, Xu Y, Fu L, Lou L. HH-GV-678, a novel selective inhibitor of Bcr-Abl, outperforms imatinib and effectively overrides imatinib resistance. *Leukemia* 2010;24:1807-9. doi:10.1038/leu.2010.169.
- Hamzehali H, Lotfi S, Ahmadi S, Kumar P. Quantitative structure-activity relationship modeling for predication of inhibition potencies of imatinib derivatives using SMILES attributes. *Sci Rep* 2022;12:21708. doi:10.1038/s41598-022-26279-8.
- Kumar P, Kumar A, Singh D. CORAL: Development of a hybrid descriptor based QSTR model to predict the toxicity of dioxins and dioxin-like compounds with correlation intensity index and consensus modelling. *Environ Toxicol Pharmacol* 2022; 93:103893. doi: 10.1016/j.etap.2022.103893.
- Toropova A, Toropov A, Viganò E, Colombo E, Roncaglioni A, Benfenati E. Carcinogenicity prediction using the index of ideality of correlation. *SAR QSAR Environ Res* 2022;33:419-28. doi:10.1080/1062936X.2022.2076736.
- Toropova AP, Toropov AA, Roncaglioni A, Benfenati E. The index of ideality of correlation improves the predictive potential of models of the antioxidant activity of tripeptides from frog skin (*Litoria rubella*). *Comput Biol Med* 2021;133:104370. doi:10.1016/j.combiomed.2021.104370.
- Kyaw Zin PP, Borrel A, Fourches D. Benchmarking 2D/3D/MD-QSAR Models for Imatinib Derivatives: How Far Can We Predict? *J Chem Inf Modl* 2020; 60:3342-60. doi:10.1021/acs.jcim.0c00200.
- Lotfi S, Ahmadi S, Kumar P. A hybrid descriptor based QSPR model to predict the thermal decomposition temperature of imidazolium ionic liquids using Monte Carlo approach. *J Mol Liq* 2021;338:116465. doi:10.1016/j.molliq.2021.116465.
- Duhan M, Sindhu J, Kumar P, Devi M, Singh R, Kumar R, et al. Quantitative structure activity relationship studies of novel hydrazone derivatives as α -amylase inhibitors with index of ideality of correlation. *J Biomol Struct Dyn* 2022;40: 4933-53. doi:10.1080/07391102.2020.1863861.
- Achary P, Toropova A, Toropov A. Combinations of graph invariants and attributes of simplified molecular input-line entry system (SMILES) to build up models for sweetness. *Food Res Int* 2019;122:40-6. doi:10.1016/j.foodres.2019.03.067.
- Ahmadi S, Azimi N. Quasi-SMILES-Based QSPR/QSAR Modeling. *QSPR/QSAR Analysis Using SMILES and Quasi-SMILES*: Springer; 2023. p. 191-210. doi:10.1007/978-3-031-28401-4_8.
- Ahmadi S. Mathematical modeling of cytotoxicity of metal oxide nanoparticles using the index of ideality correlation criteria. *Chemosphere* 2020;242:125192. doi:10.1016/j.chemosphere.2019.125192.
- Das NR, Sharma T, Mallick A, Toropova AP, Toropov AA, Achary P. Computational Approach in Designing and Development of Novel Inhibitors of AKR1C1. *Ambient Intelligence in Health Care*: Springer; 2023. p. 325-37. doi:10.1007/978-981-19-6068-0_32.
- Ghaedi A. Predicting the cytotoxicity of ionic liquids using QSAR model based on SMILES optimal descriptors. *J Mol Liq* 2015;208:269-79. doi: 10.1016/j.molliq.2015.04.049.
- Singh R, Kumar P, Devi M, Lal S, Kumar A, Sindhu J, et al. Monte Carlo based QSGFEAR: prediction of Gibb's free energy of activation at different temperatures using SMILES based descriptors. *New J Chem* 2022;46:19062-72. doi:10.1039/D2NJ03515D.
- Ahmadi S, Lotfi S, Kumar P. Quantitative structure-toxicity relationship models for predication of toxicity of ionic liquids toward leukemia rat cell line IPC-81 based on index of ideality of correlation. *Toxicol Mech Methods* 2022; 32:302-12. doi:10.1080/15376516.2021.2000686.
- Lotfi S, Ahmadi S, Kumar P. The Monte Carlo approach to model and predict the melting point of imidazolium ionic liquids using hybrid optimal descriptors. *RSC Adv*

- 2021;11:33849-57. doi: 10.1039/D1RA06861J.
21. Ahmadi S, Ketabi S, Qomi M. CO₂ uptake prediction of metal-organic frameworks using quasi-SMILES and Monte Carlo optimization. *New J Chem* 2022;46:8827-37. doi:10.1039/D2NJ00596D.
22. Kumar P, Kumar A. CORAL: QSAR models of CB1 cannabinoid receptor inhibitors based on local and global SMILES attributes with the index of ideality of correlation and the correlation contradiction index. *Chemo Intell Lab Syst* 2020;200:103982. doi: 10.1016/j.chemolab.2020.103982.
23. Shayanfar A, Shayanfar S. Is regression through origin useful in external validation of QSAR models? *Eur J Phar Sci* 2014;59:31-5. doi:10.1016/j.ejps.2014.03.007.
24. Consonni V, Ballabio D, Todeschini R. Comments on the definition of the Q₂ parameter for QSAR validation. *J Chem Inf Model* 2009;49:1669-78. doi: 10.1021/ci900115y.
25. Roy K, Kar S. The rm² metrics and regression through origin approach: Reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?'). *Eur J Pharm Sci* 2014;62:111-4. doi:10.1016/j.ejps.2014.05.019.
26. Lawrence I, Lin K. Assay validation using the concordance correlation coefficient. *Biometrics* 1992;599-604. doi:10.2307/2532314.
27. Rücker C, Rücker G, Meringer M. Y-randomization and its variants in QSPR/QSAR. *J Chem Inf Model* 2007;47:2345-57. doi: 10.1021/ci700157b.
28. Toropova AP, Toropov AA, Veselinović AM, Veselinović JB, Benfenati E, Leszczynska D, et al. Nano-QSAR: Model of mutagenicity of fullerene as a mathematical function of different conditions. *Ecotoxicol Environ Saf* 2016;124:32-6. doi:10.1016/j.ecoenv.2015.09.038.
29. Chirico N, Gramatica P. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J Chem Inf Model* 2012;52:2044-58. doi:10.1021/ci300084j.
30. Yordanova D, Schultz TW, Kuseva C, Tankova K, Ivanova H, Dermen I, et al. Automated and standardized workflows in the OECD QSAR Toolbox. *Comput Toxicol* 2019;10:89-104. doi:10.1016/j.comtox.2019.01.006.
31. Gatidou G, Vazaiou N, Thomaidis NS, Stasinakis AS. Biodegradability assessment of food additives using OECD 301F respirometric test. *Chemosphere* 2020;241:125071. doi: 10.1016/j.chemosphere.2019.125071.
32. Marzo M, Lavado G, Como F, Toropova A, Toropov A, Baderna D, et al. QSAR models for biocides: The example of the prediction of *Daphnia magna* acute toxicity. *SAR QSAR Environ Res* 2020;31:227-43. doi:10.1080/1062936X.2019.1709221.
33. Soleymani N, Ahmadi S, Shiri F, Almasirad A. QSAR and molecular docking studies of isatin and indole derivatives as SARS 3CLpro inhibitors. *BMC Chem* 2023;17:32. doi: 10.1186/s13065-023-00947-w.
34. Azimi A, Ahmadi S, Kumar A, Qomi M, Almasirad A. SMILES-based QSAR and molecular docking study of oseltamivir derivatives as influenza inhibitors. *PACs* 2023;43:3257-77. doi:10.1080/10406638.2022.2067194.
35. Toropova AP, Schultz TW, Toropov AA. Building up a QSAR model for toxicity toward *Tetrahymena pyriformis* by the Monte Carlo method: A case of benzene derivatives. *Environ Toxicol Pharmacol* 2016;42:135-45. doi:10.1016/j.etap.2016.01.010.