Pejouhesh dar Pezeshki (i.e., Research in Medicine) **Original Article**

The Quarterly Journal of School of Medicine, Shahid Beheshti University of Medical Sciences 2025; Vol. 49; No. 2

Quantitative Structure- Activity Relationship Study of a Series of Monoketoacid Derivatives as Anti-HIV Agents

Nasrin Zarei[1], Tannaz Zebardast[2*], Zahra Hajimahdi[3*], Afshin Zarghi[3]

1. Department of Pharmaceutical Chemistry, TeMS.C., Islamic Azad University, Tehran, Iran.
2. Department of Medicinal Chemistry, TeMS.C., Islamic Azad University, Tehran, Iran.
3. Department of Pharmaceutical Chemistry, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Received: March 30, 2025; Accepted: October 07, 2025

Abstract

Background and Aim: Due to the complexity of drug discovery processes and the high cost of biological testing, computational methods in medicinal chemistry have gained attention as effective tools for streamlining drug design. One such approach is Quantitative Structure-Activity Relationship (QSAR) analysis, which uses mathematical and statistical models to predict the biological activity of chemical compounds based on their structural features. By extracting molecular descriptors and establishing mathematical relationships between chemical structure and biological activity, QSAR enables the prediction of biological behavior without the need for synthesis and experimental evaluation. This strategy not only reduces time and cost but also plays a significant role in optimizing drug structures and enhancing their efficacy.

In the present study, QSAR analysis was performed on a series of monoketo acid derivatives previously reported to exhibit inhibitory activity against HIV. The resulting model, with strong predictive capability, may be used for the rational design of new compounds with potential anti-HIV activity and serves as a valuable tool in the development of HIV inhibitors.

Methods: This study was conducted using computational techniques. First, the chemical structures of the selected compounds were drawn and their energy levels optimized. Molecular descriptors were calculated using Dragon and PaDEL software. Constant, near-constant, and highly correlated descriptors were removed to improve model quality. The dataset was randomly divided into training and test sets. Important descriptors were selected using the stepwise method, and a multiple linear regression (MLR) model was constructed. For model validation, both internal and external validation techniques were applied. The coefficient of determination (R^2) for the training and test sets, as well as the Leave-One-Out cross- validation coefficient (Q^2_LOO), were calculated. Acceptable models were defined as those with R^2 and Q^2 values greater than 0.6.

Results: The coefficient of determination (R^2) for the training and test sets was 0.834 and 0.808, respectively. The cross-validation coefficient (Q^2_LOO) was 0.681, indicating good internal predictive power. The descriptors selected in the final model included R7e, RDF155e, AATS6s, and maxaaCH, belonging to the GETAWAY, RDF, AC, and Electrotopological classes, respectively.

Conclusion: The developed QSAR model appears statistically significant and suitable for predicting the inhibitory activity of novel monoketo acid derivatives against HIV-1. This model can be considered a valuable computational tool for the rational design of new anti-HIV agents.

Keywords: Monoketoacid; QSAR; Descriptors; Stepwise; Multiple Linear Regression; HIV

Please cite this article as: Zarei N, Zebardast T, Hajimahdi Z, Zarghi A. Quantitative Structure- Activity Relationship Study of a Series of Monoketoacid Derivatives as Anti-HIV Agents. Pejouhesh dar Pezeshki. 2025;49(2):57-69.

* **First Corresponding Author:** Zahra Hajimahdi; **Email:** z.hajimahdi@sbmu.ac.ir
Department of Pharmaceutical Chemistry, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

* **Second Corresponding Author:** Tannaz Zebardast; **Email:** t.zebardast@iautmu.ac.ir
Department of Medicinal Chemistry, TeMS.C., Islamic Azad University, Tehran, Iran.

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

فصلنامه علمی-پژوهشی دانشکده پزشکی، دانشگاه علوم پزشکی و خدمات بهداشتی-درمانی شهید بهشتی
دوره ۴۹، شماره ۲، سال ۱۴۰۴ شمسی، صفحات ۵۷ تا ۶۹

مطالعه رابطه کمی بین ساختار گروهی از ترکیب‌های مونوکتواسید با فعالیت ضد ویروس HIV-1 آن‌ها

نسرین زارعی[۱]، طناز زبردست[۲*]، زهرا حاجی مهدی[۳*]، افشین زرقی[۳]

۱- گروه شیمی دارویی، واحد علوم پزشکی تهران، دانشگاه آزاد اسلامی، تهران، ایران.
۲- گروه شیمی دارویی، واحد علوم پزشکی تهران، دانشگاه آزاد اسلامی، تهران، ایران.
۳- گروه شیمی دارویی، دانشکده داروسازی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

تاریخ دریافت: ۱۴۰۴/۰۱/۱۰ تاریخ پذیرش: ۱۴۰۴/۰۷/۱۵

چکیده

سابقه و هدف: با توجه به پیچیدگی‌های فرآیند کشف داروهای جدید و هزینه‌های بالای آزمایش‌های زیستی، استفاده از روش‌های محاسباتی در شیمی دارویی به‌عنوان ابزاری مؤثر در کوتاه‌سازی مسیر طراحی دارو مورد توجه قرار گرفته است. یکی از این روش‌ها، مطالعه رابطه کمی بین ساختار و فعالیت (QSAR) است که با بهره‌گیری از مدل‌های ریاضی و آماری، امکان پیش‌بینی فعالیت زیستی ترکیب‌های شیمیایی بر اساس ویژگی‌های ساختاری آنها را فراهم می‌سازد. در این روش، با استخراج توصیف‌گرهای مولکولی و ایجاد یک معادله ریاضی بین ساختار شیمیایی و فعالیت بیولوژیکی، می‌توان رفتار زیستی ترکیب‌های جدید را بدون نیاز به سنتز و آزمون تجربی پیش‌بینی کرد. این رویکرد نه‌تنها سبب صرفه‌جویی در زمان و هزینه می‌شود، بلکه در بهینه‌سازی ساختار داروها و افزایش اثربخشی آنها نیز نقش مهمی ایفا می‌کند.

در مطالعه حاضر، تحلیل QSAR بر روی مجموعه‌ای از ترکیب‌های مونوکتواسید که قبلاً فعالیت مهارکنندگی علیه ویروس HIV داشته‌اند، انجام شد. مدل حاصل، با قابلیت پیش‌بینی مناسب، می‌تواند در طراحی منطقی ترکیب‌های جدید با پتانسیل مهار ویروس HIV استفاده شود و به‌عنوان ابزاری مؤثر در توسعه داروهای ضد HIV به کار گرفته شود.

روش کار: این مطالعه به صورت محاسبه‌ای انجام می‌شود، در ابتدا ساختار ترکیب‌های مورد نظر رسم و سطح انرژی آنها بهینه شد. پس از آن توصیف‌گرها با استفاده از نرم‌افزار دراگون و پدل محاسبه شدند. سپس توصیف‌گرهای ثابت، تقریباً ثابت و هم راستا حذف شدند. در این مرحله داده‌ها به روش تصادفی به دو دستهی آموزش و آزمون تقسیم شد. با استفاده از روش گام به گام (stepwise)، توصیف‌گرهای مهم در مدل‌سازی انتخاب و با روش رگرسیون خطی چندگانه مدل‌سازی انجام شد. در اعتبارسنجی مدل، دو روش اعتبارسنجی داخلی و خارجی استفاده شد و میزان R^2 برای گروه آموزش و آزمون و Q^2_{LOO} محاسبه شد که باید مقدارآنها بیشتر از ۰/۶ باشند.

یافته‌ها: مقدار R^2 برای گروه آموزش برابر با ۰/۸۳۴ و برای گروه آزمون برابر با ۰/۸۰۸ به دست آمد. همچنین در روش اعتبارسنجی متقاطع، مقدار Q^2_{LOO} برابر ۰/۶۸۱ به دست آمد. توصیف‌گرهای به‌دست آمده در این مدل شامل R7e، RDF155e، AATS6s، maxaaCH، به ترتیب از دسته GETAWAY، RDF، AC و Electrotopological هستند.

نتیجه‌گیری: به نظر می‌رسد که مدل QSAR توسعه‌یافته از نظر آماری معنادار و برای پیش‌بینی فعالیت مهارکنندگی رشد HIV-1 در ترکیب‌های جدید مونوکتواسید قابل استفاده است. این مدل می‌تواند به‌عنوان ابزاری ارزشمند برای طراحی منطقی عوامل جدید ضد HIV استفاده شود.

واژگان کلیدی: مونوکتواسید؛ رابطه کمی بین ساختار- فعالیت؛ توصیف‌گر؛ گام به گام؛ رگرسیون خطی چندگانه؛ HIV

به این مقاله، به صورت زیر استناد کنید:

Zarei N, Zebardast T, Hajimahdi Z, Zarghi A. Quantitative Structure- Activity Relationship Study of a Series of Monoketoacid Derivatives as Anti-HIV Agents. Pejouhesh dar Pezeshki. 2025;49(2):57-69.

[*] **اولین نویسنده مسئول مکاتبات:** زهرا حاجی‌مهدی؛ **آدرس پست الکترونیکی:** z.hajimahdi@sbmu.ac.ir
گروه شیمی دارویی، دانشکده داروسازی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.
[*] **دومین نویسنده مسئول مکاتبات:** طناز زبردست؛ **آدرس پست الکترونیکی:** t.zebardast@iautmu.ac.ir
گروه شیمی دارویی، واحد علوم پزشکی تهران، دانشگاه آزاد اسلامی، تهران، ایران.

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

مقدمه

ایدز یا سندرم نقص ایمنی اکتسابی، یک بیماری جدی و کشنده است که در اثر ویروس نقص ایمنی انسانی (HIV) ایجاد می‌شود (۱). این ویروس که به گروه لنتی‌ویروس‌ها و خانواده رتروویروس‌ها تعلق دارد، از طریق مخاط وارد بدن می‌شود و پس از ورود به اندام‌های لنفاوی و سلول‌های لنفوسیت T، سیستم ایمنی بدن را به تدریج تضعیف می‌کند. در صورت عدم درمان، کاهش تعداد سلول‌های T فرد را مستعد ابتلا به عفونت‌های فرصت‌طلب و سرطان‌های مرتبط با عفونت می‌کند. HIV به مراحل مختلفی از چرخه زندگی خود وابسته است که شامل ورود به سلول، تبدیل RNA ویروسی به DNA، ادغام با DNA میزبان و تولید ویروس‌های جدید است. این ویروس در نهایت سبب نقص ایمنی و ابتلا به ایدز می‌شود که مرحله نهایی عفونت HIV است (۲، ۳).

این بیماری برای نخستین بار در سال ۱۹۸۱ در ایالات متحده گزارش شد و به سرعت به یک اپیدمی جهانی تبدیل شد (۴). گزارش‌های سازمان بهداشت جهانی (WHO)، در حال حاضر حدود ۴۰ میلیون نفر در سراسر جهان به HIV مبتلا هستند (۵). درمان HIV عمدتاً بر اساس استفاده از داروهای ضد رتروویروسی است که در مراحل مختلف چرخه زندگی ویروس مداخله می‌کنند. تاکنون بیش از ۳۰ دارو در چندین دسته اصلی برای درمان HIV توسعه یافته‌اند. این داروها شامل مهارکننده‌های هم‌جوشی، مهارکننده‌های ترانس‌کریپتاز معکوس، مهارکننده‌های پروتئاز و مهارکننده‌های اینتگراز هستند که هر کدام یک مرحله مشخص از تکثیر ویروس را هدف قرار می‌دهند (۶، ۷). در حال حاضر، درمان بسیار فعال ضد رتروویروسی (HAART) به عنوان یک رویکرد درمانی استاندارد در حال استفاده است. کاهش مرگ‌ومیرهای مرتبط با ایدز نتیجه پیشرفت‌های اخیر در داروها و رژیم‌های ضد HIV است، اما همچنان نیاز به تلاش بیشتری وجود دارد. ویروس HIV به دلیل ویژگی‌های ژنتیکی خود به سرعت جهش می‌کند و مقاومت دارویی ایجاد می‌کند، که این مسئله ضرورت کشف و توسعه داروهای جدید را افزایش می‌دهد (۸).

ترکیب‌های دی و مونوکتو اسید از ترکیب‌های شناخته شده به عنوان مهارکنندگان آنزیم اینتگراز ویروس HIV هستند که تأثیر قوی در مهار رشد ویروس از خود نشان داده‌اند. این ترکیب‌ها به دلیل داشتن گروه‌های کربونیل مجاور هم یون‌های Mg^{2+} داخل جایگاه فعال آنزیم اینتگراز را شلاته و آنزیم اینتگراز را مهار می‌کنند. از آنجا که آنزیم اینتگراز الحاق ژنوم ویروس به ژنوم سلول میزبان را کاتالیز می‌کند، مهار آن مانع رشد و تکثیر ویروس می‌شود. امروزه ترکیب‌های متنوع زیادی که داری قسمت مونو یا دی‌کتو اسید هستند به عنوان ترکیبت ضد ویروس HIV شناخته و معرفی شده‌اند (۹).

یکی از روش‌های پیشرفته برای طراحی و توسعه دارو، استفاده از رابطه کمی ساختار- فعالیت (QSAR) است. مدل QSAR مناسب نشان می‌دهد که چگونه فعالیت زیستی یا ویژگی‌های یک مجموعه از مولکول‌ها به عنوان تابعی از توصیف‌کننده‌های مولکولی به‌دست‌آمده از ساختار شیمیایی تغییر می‌کند. روش‌های QSAR نسبت به آزمون‌های درون تن و برون تن کم‌هزینه‌تر و سریع‌تر هستند. در روش QSAR، مدل‌های معتبری با استفاده از تحلیل‌های آماری برای تعیین روابط خطی یا غیرخطی بین ساختار و فعالیت داروها ایجاد می‌شود. با استفاده از مدل‌های QSAR برای پیش‌بینی کمی فعالیت ساختارهای جدید، می‌توان از هزینه‌های اضافی مانند سنتز و ارزیابی زیستی در فرآیند طراحی و کشف دارو جلوگیری کرد (۱۰، ۱۱).

هدف از تحقیق حاضر، مدل‌سازی کمی بین اثر مهارکنندگی یک سری از ترکیب‌های مونوکتو اسید بر روی رشد ویروس HIV و توصیف‌گرهای آنها است که با استفاده از روش مدل‌سازی رگرسیون خطی چندگانه مرحله‌ای (Stepwise- MLR) انجام می‌شود. نتایج این تحقیق می‌تواند به شناسایی عوامل کلیدی در ساختار ترکیب‌های شیمیایی کمک کند که بر فعالیت مهارکنندگی رشد ویروس HIV تأثیر

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

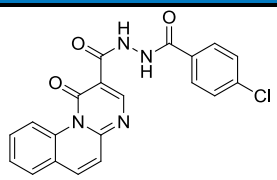
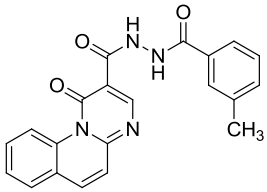
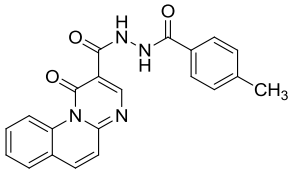
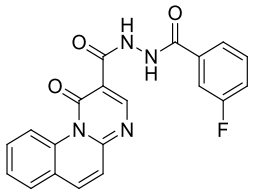
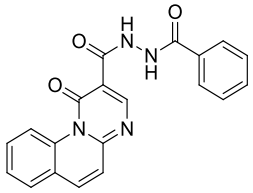
می‌گذارند. این دستاوردها می‌توانند راهنمایی برای طراحی داروهای جدید و مؤثرتر علیه HIV باشند.

روش کار

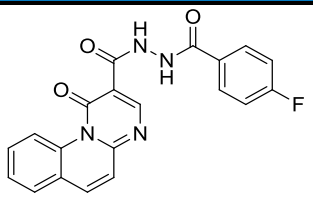
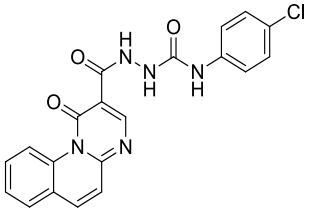
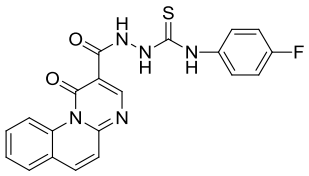
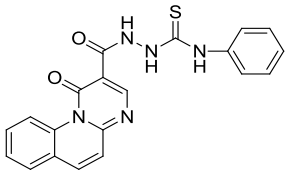
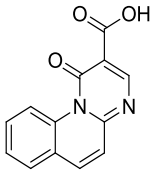
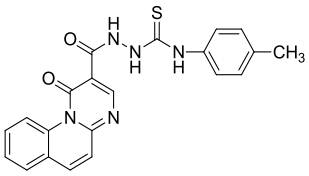
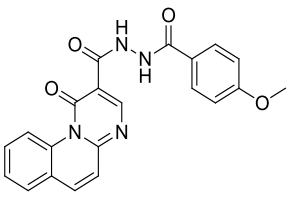
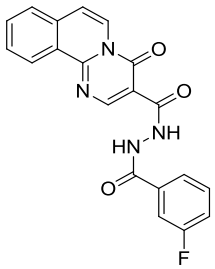
مجموعه داده‌ها

در این تحقیق با کد اخلاق IR.SBMU.PHARMACY.REC.1400.177 مطالعه رابطه کمی ساختار- فعالیت (QSAR) بر روی ترکیب‌های مونوکتو اسید انجام شد. این ترکیب‌ها در آزمایشگاه شیمی دارویی دانشکده داروسازی شهید بهشتی به عنوان مهارکنندگان رشد ویروس HIV سنتز و ارزیابی شدند (۱۲، ۱۳). جدول ۱ ساختارهای شیمیایی ترکیب‌های مورد مطالعه را نشان می‌دهد. فعالیت مهار رشد ویروس HIV به صورت درصد مهار p24 در غلظت ۱۰۰ میکرومولار و فرم لگاریتمی آن در جدول ۱ آورده شده است. آنتی‌ژن p24 یکی از آنتی ژن‌های اختصاصی ویروس است که مهار رشد آن نشان‌دهنده مهار رشد ویروس در محیط کشت سلولی است. فرم لگاریتمی درصد مهار (LogIR) به عنوان متغیر پیوسته در مطالعه به کار رفت.

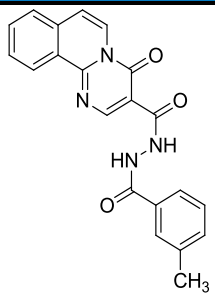
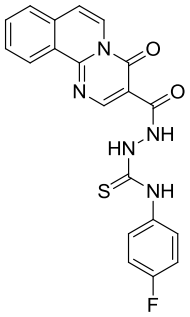
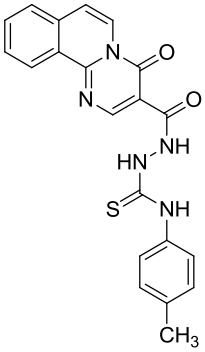
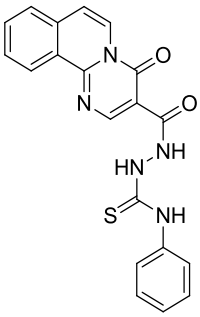
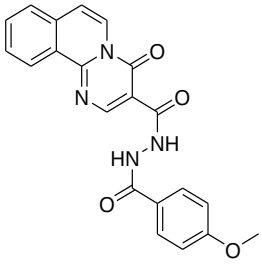
جدول ۱- ساختار ترکیب‌های مجموعه داده‌ها

شماره	ترکیب	درصد مهار در غلظت ۱۰۰ میکرومولار	لگاریتم درصد مهار
۱		۵۶	۱/۷۵
۲		۷۶	۱/۸۸
۳		۵۴	۱/۷۳
۴		۱۸	۱/۲۶
۵		۵۱	۱/۷۱

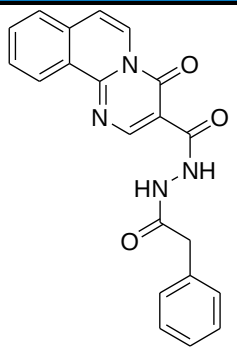
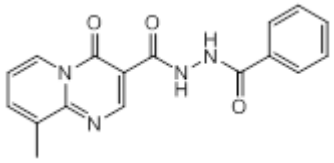
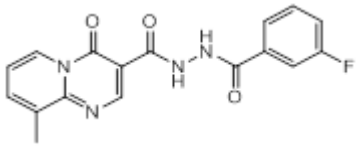
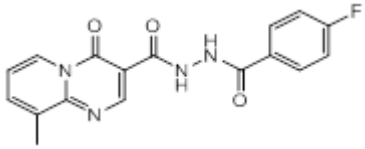
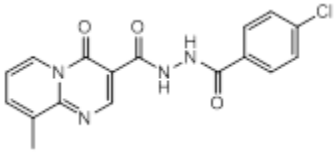
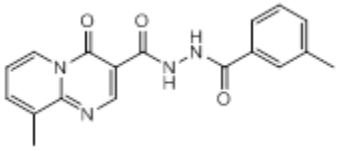
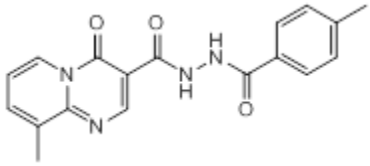
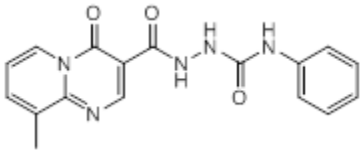
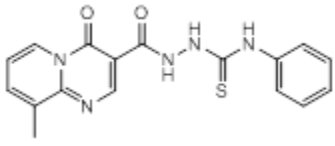
[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

شماره	ترکیب	درصد مهار در غلظت ۱۰۰ میکرومولار	لگاریتم درصد مهار
۶		۱۵	۱/۱۸
۷		۵۱	۱/۷۱
۸		۷۵	۱/۸۸
۹		۶۶	۱/۸۲
۱۰[a]		۱۴	۱/۱۵
۱۱[a]		۷۸	۱/۸۹
۱۲		۲۷	۱/۴۳
۱۳[a]		۳۷	۱/۵۷

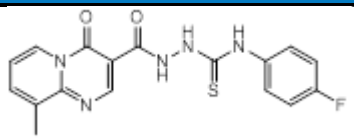
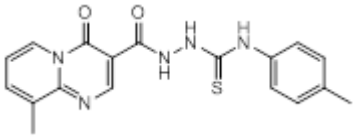
[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

شماره	ترکیب	درصد مهار در غلظت ۱۰۰ میکرومولار	لگاریتم درصد مهار
۱۴[a]		۳۲	۱/۵۱
۱۵		۲۴	۱/۳۸
۱۶		۹۹	۲/۰۰
۱۷		۹۸	۲/۰۰
۱۸[a]		۲۷	۱/۴۳

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

شماره	ترکیب	درصد مهار در غلظت ۱۰۰ میکرومولار	لگاریتم درصد مهار
۱۹		۱۶	۱/۲۰
۲۰		۲۱	۱/۳۲
۲۱[a]		۵۰	۱/۷۰
۲۲		۱۵	۱/۱۸
۲۳		۲۳	۱/۳۶
۲۴		۹	۰/۹۵
۲۵		۲۱	1.32
۲۶		۶	۰/۷۸
۲۷		۵	۰/۷۰

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

شماره	ترکیب	درصد مهار در غلظت ۱۰۰ میکرومولار	لگاریتم درصد مهار
۲۸		۱۲	۱/۰۸
۲۹		۲۳	۱/۳۶

[a] **test molecules**

رسم و بهینه‌سازی ساختارها و محاسبه توصیف‌گرها

در ابتدا، ساختار مولکولی ۲۹ ترکیب با نرم‌افزار (۱۴) HyperChem 8.0 رسم شد. پیش‌بهینه‌سازی با روش میدان نیروی مکانیک مولکولی (MM+) انجام شد. سپس، کانفورمرهای کم‌انرژی با استفاده از روش نیمه‌تجربی AM1 و الگوریتم پولاک- ریبر (Polak- Ribiere) به دست آمدند. بهینه‌سازی تا زمانی ادامه یافت که جذر میانگین مربعات گرادیان انرژی (RMS gradient) به‌یک صدم کیلوکالری بر مول برسد.

ساختارهای حاصل به نرم‌افزارهای PaDEL و Dragon۳ (۱۶، ۱۵) منتقل شدند تا توصیف‌گرها محاسبه شوند. برای هریک از مولکول‌ها در نرم‌افزار دراگون ۱۴۹۷ توصیف‌گر در ۲۰ طبقه مختلف و در نرم‌افزار پدل ۱۴۴۴ توصیف‌گر محاسبه شد. پس از ادغام داده‌های حاصل از دو بسته نرم‌افزاری، در مجموع ۲۹۴۱ توصیف‌گر برای تمام مولکول‌ها محاسبه شد. توصیف‌گرهایی که مقادیر یکسانی برای کل مولکول‌ها داشتند (توصیف‌گرهای ثابت) از فهرست توصیف‌گرها حذف شدند. همچنین توصیف‌گرهایی که برای ۹۰ درصد مولکول‌ها یکسان بودند (تقریباً ثابت)، حذف شدند. سپس با بررسی همبستگی بین متغیرها از میان هر دو توصیف‌گری که ضریب همبستگی بیش از ۰/۹ داشتند (توصیف‌گرهای هم راستا) ، موردی که همبستگی کمتری با متغیر وابسته داشت حذف شد. توصیف‌گرهای باقی‌مانده در یک ماتریس داده (D) n×m جمع‌آوری شدند که در آن n=۲۹ تعداد مولکول ها و m=۵۰۲ تعداد توصیف‌گرهای باقیمانده است. داده‌ها به صورت کاملاً تصادفی به دو گروه آموزش (۲۳ ترکیب) و آزمون (۶ ترکیب) تقسیم شدند. گروه آموزش برای ساخت مدل و گروه آزمون برای ارزیابی قدرت پیش‌بینی مدل استفاده شد.

انتخاب توصیف‌گرهای مؤثر و مدل‌سازی

یکی از مشکلات اصلی در QSAR، ابعاد بالای فضای توصیف‌گرهاست. بنابراین، انتخاب توصیف‌گرها مهم‌ترین مرحله در این مطالعه محسوب می‌شود. هدف از انتخاب توصیف‌گرها، برگزیدن زیرمجموعه‌ای از متغیرهای ورودی از طریق حذف توصیف‌گرهای غیرمرتبط یا فاقد اطلاعات پیش‌بینی‌کننده است. این فرآیند با هدف استخراج حداکثر اطلاعات ممکن از تعداد محدودی توصیف‌گر انجام می‌شود تا در نهایت، بهترین زیرمجموعه از متغیرها برای مدل نهایی QSAR تعیین شود.

چندین الگوریتم برای انتخاب توصیف‌گرها وجود دارد که هرکدام دارای مزایا و معایب خاص خود هستند. روش گام‌به‌گام (Stepwise) ترکیبی از انتخاب رو به جلو (Forward Selection) و حذف رو به عقب (Backward Elimination) است که به‌صورت خودکار، توصیف‌گرهای آماری معنادار را انتخاب می‌کند. رگرسیون گام‌به‌گام بر اساس دو استراتژی مختلف عمل می‌کند در انتخاب رو به جلو ، مدل با هیچ متغیری شروع می‌شود و سپس با افزودن هر متغیر، بهبود برازش مدل ارزیابی می‌شود. در حذف رو به عقب، مدل با همه متغیرها شروع شده و سپس با حذف تدریجی متغیرهای غیرمؤثر، بهبود مدل بررسی می‌شود (۱۷). در این تحقیق با استفاده از روش گام به گام (SW) بهترین مجموعه

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

توصیف‌گرهای محاسبه شده انتخاب و به روش رگرسیون خطی چندگانه (MLR) مدل‌سازی انجام شد. برای انجام مدل‌سازی از نرم‌افزار SPSS17 استفاده شد (۱۸). رگرسیون خطی چندگانه رایج‌ترین شکل تحلیل رگرسیون خطی است. در این روش، رابطه بین یک متغیر وابسته پیوسته و دو یا چند متغیر مستقل توضیح داده می‌شود. به عبارت دیگر، این روش فرض می‌کند که بین متغیر وابسته و متغیرهای مستقل، رابطه خطی وجود دارد.

اعتبارسنجی مدل‌ها

برای ارزیابی مدل‌ها از روش‌های اعتبارسنجی داخلی و خارجی استفاده شد. اعتبارسنجی داخلی شامل تحلیل پارامترهای آماری مانند ضریب رگزسیون (R^2)، خطای استاندارد تخمین، اعتبارسنجی متقاطع با روش کنار گذاشتن یک داده (Leave-One-Out) و تصادفی سازی متغیر وابسته (Y-randomization) هستند. در اعتبارسنجی خارجی از گروه تست برای بررسی دقت مدل در پیش‌بینی داده‌های جدید استفاده شد (۱۹).

یافته‌ها

پس از تقسیم مجموعه داده‌ها به مجموعه آموزشی و مجموعه آزمون، روش گام به‌گام برای انتخاب توصیف‌گرهای مرتبط با فعالیت، بر اساس نمونه‌های موجود در مجموعه آموزشی، انجام شد. تحلیل رگرسیون خطی چندگانه با انتخاب گام به گام برای مدل‌سازی رابطه ساختار- فعالیت به‌کار گرفته شد. فرآیند SW-MLR اجرا شد و از میان مدل‌های به‌دست‌آمده، مدلی با بالاترین کیفیت آماری و شامل چهار توصیف‌گر مرتبط برای ساخت مدل انتخاب شد (معادله ۱).

معادله خطی حاصل از توصیف‌گرهای انتخاب‌شده بر اساس SW-MLR به‌صورت زیر است:

معادله ۱:

LogIR =1.0807(±0.758) +5.360(±0.627) ***R7e**-0.48(±0.114) ***AATS6s**+0.068(±0.017) ***RDF155e**-0.939(±0.32) * **MaxaaCH**

پارامترهای آماری این مدل در جدول ۲ نشان داده شده‌اند. مقدار R^2 مدل SW-MLR برابر با ۰/۸۴ برای مجموعه آموزشی و ۰/۸۱ برای مجموعه آزمون به‌دست آمد. مدل ایجادشده با استفاده از روش اعتبارسنجی متقاطع Leave- One- Out (LOO) ارزیابی و اعتبارسنجی شد. در روش LOO، یک نقطه داده از مجموعه حذف شده و مدل دوباره محاسبه می‌شود. سپس، فعالیت پیش‌بینی‌شده برای آن نقطه با مقدار واقعی مقایسه می‌شود. این فرآیند به‌طور مکرر تکرار می‌شود تا هر نقطه داده یک‌بار از مجموعه حذف شود. پارامترهای اعتبارسنجی متقاطع در جدول ۲ نشان داده شده‌اند. توصیف‌گرهای انتخاب‌شده توسط مدل MLR و تعریف آنها در جدول ۳ ارائه شده‌اند.

جدول ۲- نتایج ارزیابی آماری مدل

	R^2 گروه آزمون	خطای معیار تخمین SEE	نسبت واریانس F	Q^2_{LOO}	تخمین ریشه مربعات خطا RMSE	R^2 گروه تست
مدل	۰/۸۳۴	۰/۱۶۹	۲۲/۵۹	۰/۶۸۱	۰/۲۱۳	۰/۸۰۸

جدول *Error! No text of specified style in document.*- توصیف توصیف‌گرهای به‌دست آمده

توصیف‌گر	تعریف	کلاس
R7e	از دسته توصیف‌گرهای R-GETAWAY در فاصله توپولوژی ۷ قرار دارد و با الکترونگاتیویته وزن دار شده	GETAWAY
RDF155e	تابع توزیع شعایی در فاصله ۱۵/۵ که با الکترونگاتیویته وزن دار شده است.	RDF
AATS6s	توصیف‌گر AST (Autocorrelation Moreau-Broto) در فاصله ۶ در حالت ۱ وزن دار شده است.	AC
maxaaCH	توصیف‌گرهای الکتروتوپولوژیکال از نوع اتمی	Electrotopological

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

مدل همچنین با استفاده از آزمون Y-randomization اعتبارسنجی شد. در این آزمون، چندین بازآرایی تصادفی روی متغیر وابسته انجام شد و مدل‌های QSAR جدید ساخته شدند. مقادیر پایین در مدل‌های به‌دست‌آمده نشان داد که نتایج مطلوب مدل اصلی، ناشی از همبستگی تصادفی نبوده است (جدول ۴).

جدول۴- نتایج تصادفی‌سازی متغیر وابسته

(*Y-randomization*)

مدل	Q^{2}**	R^{2}*
اصلی	0.681276	0.833885
۱	-0.67258	0.044697
۲	-0.61906	0.098455
۳	-0.75751	0.10219
۴	-0.75302	0.108635
۵	-0.56846	0.03237
۶	-0.45883	0.276491
۷	-0.1133	0.253346
۸	0.115753	0.424402
۹	-0.22036	0.343342
۱۰	-0.42914	0.27091

* ضریب رگرسیون گروه آموزش

** ضریب رگرسیون اعتبارسنجی متقاطع

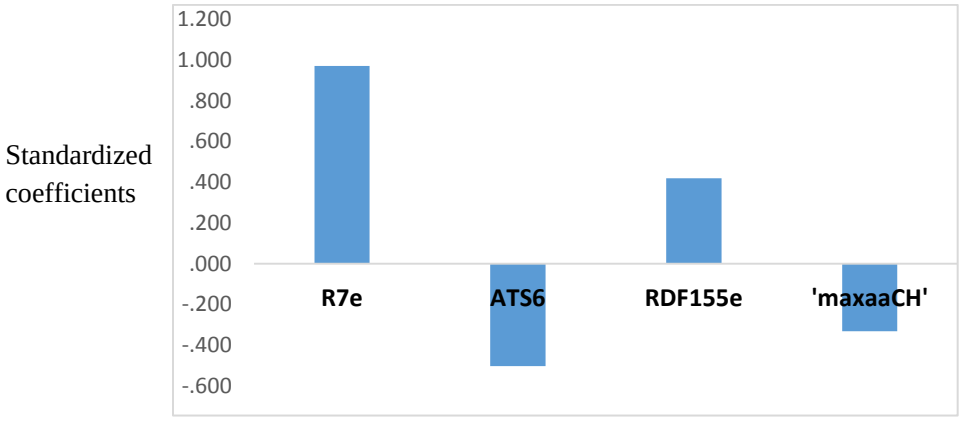
بحث

تحقیق نشان داد که ارتباط معنادار بین ویژگی‌های ساختاری ترکیب‌های مونوکتواسید و فعالیت مهارکنندگی آنها علیه ویروس HIV موجود داشته است. مقدار بالای R^2 برای گروه آموزش (۰/۸۳۴) و گروه آزمون (۰/۸۰۸) بیانگر دقت مناسب مدل در توصیف داده‌های موجود و توانایی آن در پیش‌بینی فعالیت ترکیب‌های جدید است. همچنین مقدار Q^2 LOO برابر با ۰/۶۸۱ نشان‌دهنده اعتبار قابل قبول مدل در اعتبارسنجی متقاطع و عدم بیش‌برازش آن استع. این موضوع اهمیت استفاده از روش‌های اعتبارسنجی داخلی و خارجی را در ارزیابی کیفیت مدل‌های QSAR برجسته می‌سازد.

یکی از اهداف کلیدی در مطالعه‌های QSAR، شناسایی عوامل ساختاری مؤثر بر فعالیت زیستی ترکیب‌هاست. در این مطالعه، تفسیر توصیف‌گرهای منتخب در مدل نهایی اطلاعات ارزشمندی درباره نقش ویژگی‌های مولکولی در مهار ویروس HIV ارائه می‌دهد. توصیف‌گر R7e از دسته GETAWAY، که بیشترین اثر مثبت را بر فعالیت مهاری نشان داده، بیانگر توزیع فضایی الکترون‌ها در مولکول است و می‌تواند در تعاملات فضایی با جایگاه فعال آنزیم نقش کلیدی ایفا کند. توصیف‌گر RDF155e نیز از دسته توزیع شعاعی (Radial Distribution Function) بوده و نشان‌دهنده نحوه پراکندگی اتم‌ها در ساختار سه‌بعدی مولکول است؛ این ویژگی در برهم‌کنش‌های غیرکووالانسی با آنزیم اهمیت دارد. در مقابل، توصیف‌گرهای AATS6s و maxaaCH اثر منفی بر فعالیت مهاری دارند. AATS6s از دسته توصیف‌گرهای خودهمبستگی وزن‌دار بوده و بیانگر توزیع خواص سطحی در طول زنجیره مولکولی است، در حالی که maxaaCH نشان‌دهنده بیشینه تعداد پیوندهای فعال C–H در مولکول است. افزایش این توصیف‌گرها ممکن است موجب کاهش توانایی مولکول در ایجاد پیوندهای هیدروژنی یا تعاملات مؤثر با جایگاه فعال آنزیم شود. اهمیت نسبی هر توصیف‌گر بر اساس ضرایب رگرسیون استانداردشده تعیین شده است، چرا که واحدهای متفاوت توصیف‌گرها مانع از مقایسه مستقیم ضرایب خام می‌شود. نمودار ارائه‌شده در شکل ۱ به‌وضوح نشان می‌دهد که کدام ویژگی‌های ساختاری بیشترین تأثیر را بر فعالیت زیستی دارند.

از منظر کاربردی، مدل SW-MLR توسعه‌یافته در این مطالعه می‌تواند به‌عنوان ابزاری مؤثر در طراحی منطقی ترکیب‌های جدید با فعالیت ضد HIV استفاده شود. این مدل نه‌تنها در غربالگری اولیه ترکیب‌ها، بلکه در بهینه‌سازی ساختار داروها پیش از سنتز نیز کاربرد دارد. پیشنهاد می‌شود در مطالعه‌های آینده، از روش‌های مدل‌سازی پیشرفته‌تر مانند ماشین‌بردار پشتیبان (SVM)، درخت تصمیم یا شبکه‌های عصبی مصنوعی استفاده شود تا دقت پیش‌بینی افزایش یابد. همچنین، اعتبارسنجی زیستی ترکیب‌های پیشنهادی در شرایط آزمایشگاهی می‌تواند گام بعدی در تأیید اثربخشی مدل باشد و مسیر توسعه داروهای جدید را تسریع کند (۲۰، ۲۱).

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]



شکل ۱- ضرایب استاندارد شده بر حسب توصیف‌گرها

R7e

R7e از توصیف‌گرهای GETAWAY است و جزو توصیف‌گرهای سه بعدی طبقه‌بندی می‌شود. این توصیف‌گرها بیان‌کننده ویژگی‌های توپولوژیکی، هندسی و وزن مولکولی، مولکول هستند و از ماتریس تأثیر مولکولی[1] محاسبه می‌شوند. این ماتریس به آسانی از مختصات فضایی اتم‌های مولکول دریک صورت‌بندی مشخص به‌دست می‌آید که با معادله ۲ بیان می‌شود.

معادله ۲:

$$\mathbf{H} = \mathbf{M} \cdot \left(\mathbf{M^T} \cdot \mathbf{M}\right)^{-1} \cdot \mathbf{M^T}$$

در معادله ۲، M به ماتریس مختصات اتمی، T به ماتریس ترآن‌هاده[2] و H ماتریس تاثیر مولکولی اشاره دارد.

عناصر قطری (h_{ii}) ماتریس H لوریج[3] نام دارند که هریک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. به عنوان مثال اتم‌های دورتر نسبت به اتم‌های نزدیک‌تر به مرکز اتم، دارای h_{ii} بزرگ‌تر هستند. نخستین سری این توصیف‌گرهای مولکولی، H-GETAWAY نامیده می‌شود که تنها توسط اطلاعات به دست آمده توسط ماتریس نفوذ مولکولی محاسبه می‌شوند. در حالی که سری R-GETAWAY ترکیبی از اطلاعات ماتریس تأثیر- فاصله است که با معادله ۳ تعریف می‌شود.

معادله ۳:

$$[R]_{\bar{y}} = \left[\frac{\sqrt{h_{ii}h_{jj}}}{r_{ij}}\right]_{\bar{y}}$$

که h_{ii} و h_{jj} لوریج (قدرت نفوذ) دو اتم i و j و r_{ij} فاصله آن دو اتم است.

R7e از دسته توصیف‌گرهای R-GETAWAY از لوریج های دو اتم که در فاصله توپولوژی ۷ قرار دارند و با الکترونگاتیویته وزن‌دار شده‌اند، محاسبه شده‌اند. این توصیف‌گرها از آن جهت مهم هستند که به ساختار سه بعدی حساس هستند و معمولا برای مقایسه مولکول‌ها و حتی صورت بندی ها از نظر شکل، تقارن، اندازه اتم به کار می‌روند. این توصیف‌گر دارای علامت مثبت در مدل است و با افزایش این توصیف‌گر LogIR افزایش می‌یابد.

RDF155e

این توصیف‌گر از دسته توصیف‌گرهای RDF است. می‌توان تابع توزیع شعاعی‌یک ترکیب شامل A اتم را به احتمال‌یافتن‌یک اتم دریک حجم کروی با شعاع r تعریف کرد. فرم کلی این تابع به‌صورت معادله ۴ است.

معادله ۴:

$$g(R) = f \cdot \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} w_i \cdot w_j \cdot e^{-\beta \cdot (R - r_{ij})^2}$$

در معادله ۴، w به خاصیت اتمی (الکترونگاتیویته، جرم اتمی، قطبش پذیری اتمی و ...)، r_{ij} به شعاع بین اتم i و j، f به فاکتور مقیاس، β به فاکتور دما که سبب حرکت و جابه‌جایی اتم‌ها و A به تعداد اتم‌ها اشاره دارد.

RDF155e که با الکترونگاتیویته وزن‌دار شده در مدل انتخابی دیده می‌شود و جزو توصیف‌گرهای سه بعدی طبقه‌بندی

[1] Molecular Influence Matrix
[2] Transpose Matrix
[3] Leverage

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

می‌شود. با توجه به رابطه فوق، هر چه الکترونگاتیویته بیشتر شود، مقدار توصیف‌گر افزایش می‌یابد. علامت مثبت مربوط به ضریب این توصیف‌گر در مدل انتخابی نشانگر این مطلب است که استخلاف‌های الکترونگاتیو سبب فعالیت مهاری مناسب می‌شود.

AATS6S

AATS6S از توصیف‌گرهای دو بعدی همبستگی است. نخستین توصیف‌گرهای همبستگی مربوط به بوروتو و مورئو[1] هستند که تابع همبستگی را روی نمودار مولکولی اعمال می‌کنند. به این ترتیب توزیع خواص مولکولی بر حسب توپولوژی مولکول اندازه‌گیری می‌شود. پس از آن با تبدیل فاصله توپولوژیکی با فاصله بین اتمی r_{ij} این رویکرد به ساختار هندسی سه بعدی توسعه داده شد.

توصیف‌گرهای مولکولی بر پایه تابع همبستگی AC_i به طور کلی به صورت معادله ۵ تعریف می‌شوند.

معادله ۵:

$$AC_l = \int_a^b f(x) \cdot f(x+l) \cdot dx$$

در این معادله f(x) هر تابعی از x و l نشان دهنده فاصله x است. همان‌طور که گفته شد همبستگی بوروتو و مورئو (همبستگی ساختار توپولوژیکی،AST[2]) همبستگی ویژه‌ای است که روی نمودارمولکولی تعریف می‌شود. AST به صورت معادله ۶ تعریف می‌شود.

معادله ۶:

$$ATS_d = \sum_{i=1}^{A} \sum_{j=1}^{A} \delta_{ij} \cdot \left(w_i \cdot w_j\right)_d$$

در معادله ۶، W به خواص اتمی، δ_{ij} به Kronecker delta، A به عدد اتمی و d به فاصله توپولوژیکی اشاره دارد.

برای هر خاصیت اتمی (w)، مجموعه‌ای از ترم‌های همبستگی که برای تمام فاصله‌های موجود در نمودار تعریف می‌شود را توصیف‌گر AST می‌نامند، که به صورت معادله ۷ تعریف می‌شود.

معادله ۷:

$$\langle ATS_0, ATS_1, ATS_2, \dots, ATS_D \rangle_w$$

AATS6S ظاهر شده در معادله نشان‌دهنده میانگین توصیف‌گر AST (همبستگی Moreau-Broto) در حالت پایه[3] ۶ است.

maxaaCH

توصیف‌گرهای الکتروتوپولوژیکال از محیط الکترونی هر اتم به وجود می‌آیند و با توجه به خواص الکترونی ذاتی آن و اثر اتم در مولکول هستند.

maxaaCH از جمله این توصیف‌گرهاست که یک توصیف‌گر دو بعدی است. در این رابطه max نشان‌دهنده ماکزیمم، a نشان‌دهنده آروماتیسیته و CH نشان‌دهنده پیوند کربن و هیدروژن است. همان‌طور که در معادله دیده می‌شود ضریب این توصیف‌گر در معادله منفی است که این به معنای رابطه عکس این توصیف‌گر با فعالیت است.

نتیجه‌گیری

در مجموع نتایج انجام مدل‌سازی QSAR دو بعدی بر روی ترکیب‌های مونوکتو اسید نشان می‌دهد که مدل از نظری آماری معنادار و معتبر است و برای پیش‌بینی فعالیت ترکیب‌ها می‌تواند استفاده شود.

ملاحظات اخلاقی

این مطالعه، در دانشگاه علوم پزشکی شهید بهشتی بررسی و با کد اخلاق IR.SBMU.PHARMACY.REC.1400.177 ثبت شده است.

تشکر و قدردانی

این مطالعه حاصل بخشی از پایان‌نامه خانم نسرین زارعی برای دریافت درجه کارشناسی ارشد در رشته شیمی دارویی از واحد علوم پزشکی دانشگاه آزاد اسلامی بود.

تعارض منافع

نویسندگان، تعارض منافعی را گزارش نکرده‌اند.

[1] Moreau- Broto
[2] Autocorrelation of a Topological Structure
[3] Intrinsic Estate

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]

References

1. Sepkowit KA, AIDS-the first 20 years. N. Engl. J. Med. 2001;23(344):1761-72.

2. Parboosing R, Maguire G, Govender P. Nanotechnology and the treatment of HIV infection. Viruses. 2012;4(4):488-520.

3. Blattner w, Gallo RC, Temin HM. HIV causes AIDS. Science. 1998;241(4865):515-6.

4. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science. 1983;220(4599):865-7.

5. unaids. Global HIV & AIDS statistics [Available from: https://www. unaids.org/ en/resources /fact-sheet.

6. Liao C, Wang Q. Authentic HIV-1 integrase inhibitors for the treatment of HIV-1/AIDS. Privileged scaffolds in drug discovery. Elsevier, London, 2023, pp 377–390.

7. Zhan P, Pannecouque C, De Clercq E, Liu X. Anti-HIV drug discovery and development: current innovations and future trends: miniperspective. J. Med. Chem. 2016;59(7):2849–2878.

8. Piantadosi A, Chohan B, Panteleeff D, Baeten JM, Mandaliya K, Ndinya-Achola JO, et al. HIV-1 evolution in *gag* and *env* is highly correlated but exhibits different relationships with viral load and the immune response. AIDS. 2009;23(5):579-87.

9. Barik P, Gupta S, Singh G, Bharti SK, Asati V. Structural aspects of HIV-1 integrase inhibitors: SAR studies and synthetic strategies. Mol. Divers. 2024. https://doi.org/10.1007/s11030-024-11068-4

10. Gupta K, Kumar A, Patel R, et al. Predictive QSAR Models Followed by Toxicity, Molecular Docking, and Molecular Dynamics Simulation in Search of Azole Derivatives as AChE Inhibitors for the Treatment of Alzheimer's Disease.*J. Chemometr.* 2025*;* 39 (8): e70049.

11. Agarwal M, Rajkhowa S, Zaki MEA, Sinha S. Targeting PARP14: An in silico framework for identifying novel Competitive inhibitors via 3D-QSAR pharmacophore modeling and molecular dynamics, Comput. Biol. Med. 2025; 196 :110769.

12. Hajimahdi Z, Zarghi A, Zabihollahi R, Aghasadeghi MR. Synthesis, biological evaluation, and molecular modeling studies of new 1,3,4-oxadiazole- and 1,3,4-thiadiazole-substituted 4-oxo-4H-pyrido[1,2-a] pyrimidines as anti-HIV-1 agents. Med. Chem. Res. 2013;22: 2467–75

13. Hajimahdi Z, Zabihollahi R, Aghasadeghi MR, Zarghi A. Design, synthesis and docking studies of new 4-hydroxyquinoline-3-carbohydrazide derivatives as anti-HIV-1 agents. Drug Res. 2013;63: 192-7.

14. HyperChem(TM) Professional 8.0, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.

15. Mauri A, Consonni V, Pavan M, Todeschini R. DRAGON software: an easy approach to molecular descriptor calculations. MATCH Commun. Math. Comput. Chem. 2006; 56: 237–48.

16. Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 2011;32(7):1466-74.

17. Nathans LL, Oswald FL, Nimon K. Interpreting multiple linear regression: A guidebook of variable importance. Practical Assessment, Research & Evaluation. 2012;17: 1-19.

18. Cormick KM, Salcedo J, Poh A. SPSS Statistics for Dummies. 3rd edition. John Wily&Sons Inc. 2015. p. 10.

19. Gramatica P. Principles of QSAR models validation: Internal and external. QSAR Comb. Sci. 2007;26: 694-701.

20. Todeschini R and Consonni V, Handbook of Molecular Descriptors. Vol. 11, John Wiley & Sons. 2008.

21. Mauri A, Consonni V, Todeschini R. Molecular Descriptors, in Handbook of Computational Chemistry (ed. T. Puzyn), Springer. 2016

[Downloaded from pejouhesh.sbmu.ac.ir on 2026-11-06]