

پیشگویی ضریب تقسیم اکتانول-آب حشره کش های آکالوئید کینولین با استفاده از توصیف کننده های مولکولی و روش رگرسیون خطی چند متغیره

شهاب الدین محبی^۱، فاطمه شفیعی^{۲*}، طاهره مومنی اصفهانی^۳، مهدی احمدی سابق^۴

۱- دانشجوی دکتری شیمی فیزیک، گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

۲- استادیار شیمی فیزیک، گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

۳- استادیار شیمی تجزیه، گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

۴- استادیار شیمی آلی، گروه شیمی، واحد اهر، دانشگاه آزاد اسلامی، اهر، ایران

چکیده

آکالوئیدهای کینولین و مشتقات آنها کاربردهای پزشکی و کشاورزی گسترده ای دارند. در این تحقیق از رابطه کمی ساختار-خاصیت (QSPR) برای پیش بینی ضریب تقسیم اکتانول-آب ۷۶ مشتق آکالوئید کینولین کمپوتوسین (CPT) به عنوان حشره کش با استفاده از روش الگوریتم ژنتیک و روش رگرسیون خطی چند متغیره برگشتی و توصیف کننده های مولکولی استفاده شده است. برای رسم ساختار شیمیایی ترکیبات مورد مطالعه از نرم افزار گوس و یواسفاده شد. بهینه سازی هندسی ترکیبات توسط نرم افزار گوسین ۰۹ با استفاده از نظریه تابعی چگالی B3YLP با مجموعه پایه G(d,p) ۳۱۱-۶ انجام شد. برای هر یک از ساختارهای بهینه شده توصیف کننده های مولکولی توسط نرم افزار دراگون محاسبه گردید. به منظور کاهش و انتخاب بهترین توصیف کننده ها از روش الگوریتم ژنتیک استفاده شد. همبستگی بین توصیف کننده ها در بهترین مدل با استفاده از ضریب پیرسون و ضریب نفوذ پذیری انجام پذیرفت. برای ارزیابی توانایی پیش بینی مدل از انواع مختلف اعتبار سنجی داخلی، خارجی و ضرایب آماری بهره گرفته شده است. بهترین مدل QSPR با مقدار مجذور ضریب همبستگی ($R^2 = 0/901$)، مجذور ضریب همبستگی اعتبار سنجی تقاطعی یکی بیرون ($Q^2_{LOO} = 0/919$)، و ریشه میانگین مربع خطا ($RMSE = 0/706$) به دست آمده است. نتایج نشان داد ضرایب آماری و اعتبارسنجی مدل خطی ساخته شده رضایت بخش است و لگاریتم ضریب تقسیم اکتانول-آب

مشتقات مورد مطالعه تحت تأثیر توصیف کننده خود همبستگی دو بعدی (ATS8e) است. این اطلاعات می تواند برای طراحی مشتقات جدید آکالوئید کینولین کمپوتوسین (CPT) به عنوان حشره کش مورد استفاده قرار گیرد.

واژه های کلیدی: "آکالوئید کینولین کمپوتوسین"؛ "ضریب تقسیم اکتانول-آب"؛ "توصیف کننده خود همبستگی دو بعدی"؛ "اعتبار سنجی"؛ "حشره کش".

* نویسنده رابط، پست الکترونیکی: Shafa38@yahoo.com

تاریخ دریافت مقاله: ۱۴۰۱/۱۰/۷ - تاریخ پذیرش مقاله: ۱۴۰۱/۱۲/۱۲



مقدمه

ترکیبات زیست کشی یک منبع جالب از آفت کش های زیست محیطی هستند که به طور طبیعی در گیاهان به عنوان مواد آلوشیمیایی (متابولیت های ثانویه) یافت می شوند و به گیاهان در مقاومت، تحمل یا جبران تنش های وارد شده ناشی از آفات و حشرات کمک می کنند. گروه بزرگی از متابولیت های ثانویه گیاهی شامل آلکالوئیدها، ساپونین ها، فنل ها و ترپن ها، امیدوارکننده ترین ترکیبات برای مدیریت آفات و حشرات هستند. متابولیت های ثانویه که از گونه های گیاهی خاص تولید می شوند می توانند روش های جدید و سازگار با محیط زیست را برای کنترل آفات و حشرات مضر فراهم کنند.

کینولین، از ترکیبات آلی آروماتیک هتروسیکلیک با ساختار دو حلقه ای متشکل از یک حلقه بنزن و یک حلقه پیریدین که در دو اتم کربن مجاور جوش خورده اند می باشد. حلقه بنزن دارای شش اتم کربن است، در حالی که حلقه پیریدین حاوی پنج اتم کربن و یک اتم نیتروژن است. چندین آلکالوئید (ترکیبات آلی قلیایی تولید شده در گیاهان) از خانواده کینولین دارای فعالیت های زیستی قابل توجهی هستند. کینین و کمپتوتسین دو مورد از معروف ترین و مهم ترین آلکالوئیدهای کینولین هستند که مشتقات و آنالوگ های اصلاح شده آن ها با طیف گسترده ای از فعالیت های زیستی، از جمله فعالیت های ضد تومور، ضد مالاریا، ضد باکتری و ضد قارچ، ضد انگلی و حشره کش، ضد ویروسی، ضد پلاکتی، ضد التهابی، علف کش، آنتی اکسیدانی و سایر فعالیت (Shang, et al., 2018; Chisholm, 1911; Wang, 2019; Roy, 2017) ها مورد بررسی قرار گرفته است

کینولین^۱ اولین بار در سال ۱۸۳۴ توسط شیمیدان آلمانی فریدلیب فردیناند رونگ^۲ از قطران زغال سنگ استخراج قطران زغال سنگ منبع اصلی کینولین تجاری است (Collin & Höke, 2005).

در سال ۱۸۴۲، چارلز گرهارت^۳، شیمیدان فرانسوی، ترکیبی را با تقطیر خشک کینین، استریکنین با پتاسیم هیدروکسید به دست آورد (Gerhardt, 1842)، او این ترکیب را کینولین نامید.

کینولین به عنوان یک محصول طبیعی ضد مالاریا از حشره چوبی پرویی^۴ به دست آمد. این حشرات دارای یک جفت غدد سینه ای هستند که در صورت اختلال، مایع بدبویی حاوی کینولین از آن ها خارج می شود (Eisner, et al., 1997; Chisholm, 1911).

کینین در سال ۱۸۲۰ از پوست درخت سینچونا^۵ جدا شد و به دلیل قیمت پایین آن، تجویز تزریقی، و اثربخشی بالا به طور گسترده ای برای درمان مالاریا در سراسر جهان استفاده شد (White, et al., 1983).

مالاریا کشنده ترین عفونت انگلی انسان است که توسط پنج گونه انگل تک یاخته ای ایجاد می شود و هر ساله باعث جان باختن تعداد زیادی از انسان ها در سراسر جهان می شود. تا به امروز بیش از ۷۰ آلکالوئید طبیعی برای درمان این بیماری استفاده شده است. بعضی از این آلکالوئید ها از نوع کینولین یا کوبینازولین هستند (Shang, et al., 2018).

مانند سایر ترکیبات هتروسیکلیک نیتروژن دار، مانند مشتقات پیریدین، کینولین اغلب به عنوان یک آلاینده محیطی مرتبط با تأسیسات پردازش نفت شیل یا زغال سنگ گزارش می شود و همچنین در سایت های تصفیه چوب های قدیمی یافت شده است. کینولین به دلیل حلالیت نسبتاً بالایی در آب دارای پتانسیل قابل توجهی برای تحرک در محیط است که

^۱- Quinoline

^۲- Friedlieb Ferdinand Runge

^۳- Charles Gerhardt

^۴- Oreophoetes peruana

^۵- Cinchona

ممکن است باعث آلودگی آب شود. کینولین به آسانی توسط میکروارگانیسم های خاصی مانند گونه ردوکوکوس^۱ که از خاک و لجن کارخانه کاغذ جدا می شود، تجزیه می گردد (O'Loughlin, et al., 1996)

کینولین به عنوان یک هسته اصلی برای برخی از مواد شیمیایی مهندسی شده مهم کشاورزی به عنوان ضد مالاریا استفاده می شود. این ترکیب بر مسیر تجزیه هموگلوبین انگل تأثیر می گذارد (Blake, 2016) همچنین مشخص گردید مشتقات ۴ آمینو کینولین نیز دارای خاصیت ضد مالاریا می باشند (Manohar, et al., 2014).

آنالوگ های ۴ و ۷ دی کلرو کینولین دارای خواص لاروکشی و سفیره کشی قابل توجهی هستند و به عنوان داروهای ضد مالاریا و ضد انگل های تب جنگل و تب دانگ مورد سنتز قرار گرفته اند (Murugan, et al., 2022).

سنتز شیمیایی، فعالیت بیولوژیکی و ارتباط کمی ساختار-فعالیت مجموعه ای از ایزوکسازولین های کینولین و ایزوکینولین به عنوان آفت کش برای حفاظت از محصولات کشاورزی گزارش شده اند (Xu, et al., 2014).

در اوایل دهه ۱۹۶۰، مهم ترین و معروف ترین آکالوئید کینولین، کمپتوتسین^۲ از درخت چینی کامپوتوتکا آکومیناتا^۳ جدا و شناسایی شد (Wall, et al., 1966). کمپتوتسین دارای طیف وسیعی از فعالیت ها می باشد و بعضی از مشتقات و آنالوگ های آن دارای فعالیت ضد سرطانی قوی هستند (Thomas, et al., 2004; Tanizawa, et al., 1994; Pan, 2003).

ضریب تقسیم اکتانول-آب^۴ (Po/w)

آب گریزی^۵ یا چربی دوستی یکی از ویژگی های مهم مولکولی در طراحی دارو، شیمی دارویی و اندازه گیری آلاینده ها می باشد. ضریب تقسیم اکتانول-آب معیاری برای میزان انحلال ماده مورد نظر در دو فاز آلی و آبی به حالت تعادل است و به عنوان یک ثابت فیزیکی در نظر گرفته می شود. این ضریب یکی از خواص شیمی فیزیکی قابل قبول، برای اندازه گیری آب گریزی ترکیبات شیمیایی است (Arnott & Planey, 2012; Rutkowska, et al., 2013; Abdolrahimi, et al., 2014) ترکیبات آب گریز با نسبت اکتانول به آب بالا معمولاً در نواحی آب گریز بدن مانند غشاء دو لایه لیپیدی سلول توزیع می شوند. در مقابل، ترکیبات آب دوست (نسبت اکتانول به آب پایین) در درجه نخست در نواحی آبی بدن مانند پلاسمای خون دیده می شوند (Scheytt, et al., 2005, Shargel, et al., 2012).

این شاخص نقش مهمی در تعیین خواص و فعالیت های بیولوژیکی ایفا می کند و همچنین یک جزء حیاتی در درک جذب، توزیع، متابولیسم و حذف یک ترکیب (دارو، سم و...) در سطح بدن می باشد (Kwon, 2001; Lipinski, et al., 2001) و

همچنین گزینه مفیدی برای تخمین توزیع یک ترکیب (دارو، سم و...) در سطح بدن می باشد. ترکیبات آب گریز با نسبت اکتانول به آب بالا معمولاً در نواحی آب گریز بدن مانند غشاء دو لایه لیپیدی سلول توزیع می شوند. در مقابل، ترکیبات آب دوست (نسبت اکتانول به آب پایین) در درجه نخست در نواحی آبی بدن مانند پلاسمای خون دیده می شوند (Scheytt, et al., 2005, Shargel, et al., 2012).

³- Rhodococcus

¹- camptothecin (CPT)

²- Camptotheca acuminata

³- Octanol/ water partition coefficient

⁴- Lipophilicity

ضریب تقسیم اکتانول-آب را می توان با استفاده از روش های تجربی شناخته شده اندازه گیری کرد، اما اخیراً روش های شیمی محاسباتی به طور گسترده ای برای پیش بینی $\log P$ استفاده شده است (Rageh, et al., 2017; Kujawski, et al., 2011; Ghasemi, et al., 2011).

مطالعه ارتباط کمی ساختار- فعالیت (QSAR) و ساختار- خاصیت (QSPR) یک روش قدرتمند برای پیش بینی خواص فیزیکی - شیمیایی (Hu, et al., 2010; Narender, et al., 2012; Lopez, et al., 2019; Han, et al., 2011)، فعالیت بیولوژیکی (Ruiz, et al., 2012; Murugan, et al., 2018; Kondaparia, et al., 2017) و متابولیسم اثر دارو & (Braga Andrade, 2012; Wenlock & Carlsson, 2015) است. با استفاده از این روش ها می توان در زمان و هزینه صرفه جویی نمود، به محیط زیست آسیب نرساند و موجودات آزمایشگاهی کمتری را از بین برد.

شاخص های فیزیکی - شیمیایی مانند لگاریتم ضریب تقسیم اکتانول-آب ($\log P$)، قطبش پذیری (Pol)، انکسار مولی (MR)، و انرژی هیدراتاسیون (HE) برای پیش بینی سمیت سلولی (pIC_{50}) مشتقات آلكالوئید كینولین کمپوتوسین استفاده شده است (Hussain, et al., 2016; Padron, et al., 2012; Hansch & Kurup, 2003).

روش کروماتوگرافی لایه نازک فاز معکوس در شرایط مختلف pH برای محاسبه مقادیر $\log P$ و $\log D$ آلكالوئید های ایزو کینولین استفاده شده است (Czeszak, 2020).

هدف از این پژوهش مطالعه ارتباط کمی ساختار-خاصیت برای پیش بینی $\log P$ مشتقات حشره کش آلكالوئید كینولین کمپوتوسین با استفاده از روش الگوریتم ژنتیک و رگرسیون خطی چندگانه MLR و توصیف کننده های مولکولی می باشد.

مواد و روش های محاسباتی

جهت انجام این تحقیق ابتدا ساختار اولیه تمامی سموم با استفاده از نرم افزار گوس و یو^۱۰۵ ترسیم شده و با کمک نرم افزار گوسین^۲۰۹ به روش تئوری تابع چگالی (B3LYP) و مجموعه پایه 6-311G (d,p) اقدام به بهینه سازی ترکیبات گردید. استخلاف ها در موقعیت های ۱ الی ۶ در ساختار کلی مشتقات کمپوتوسین مورد مطالعه در جدول (۱) ثبت شده است.

مقادیر تجربی $\log P$ بدست آمده از منبع معتبر علمی (Web search engine developed by ChemAxon) در جدول (۲) درج شده است.

رفتار شیمیایی یک ترکیب به ساختار مولکولی آن وابسته است لذا به منظور برقراری ارتباط کمی ساختار-خاصیت از توصیف کننده های مولکولی بهره گرفته شده است. به منظور محاسبه توصیف کننده های مولکولی (Todeschini & Consonni, 2000 & Consonni, 2009) از نرم افزار دراگون^۳ (نسخه ۵/۴، ۲۰۰۶) استفاده شده است. این نرم افزار برای هر ترکیب مورد مطالعه تعداد زیادی توصیف کننده محاسبه می نماید لذا به منظور کاهش و تعیین مناسب ترین توصیف کننده ها از روش الگوریتم ژنتیک^۴ (Hou, et al., 1999) در محیط برنامه متلب^۵ ۲۰۱۰ استفاده گردید.

¹- Gauss View 05

²- Gaussian 09

³- Dragon

²- Genetic Algorithm (GA)

³- MATLAB 2010a

جهت ساخت مدل های QSPR از روش رگرسیون خطی چندگانه برگشتی^۱ (Goodarzi, et al., 2012) در نرم افزار بسته آماری برای علوم اجتماعی^۲ نسخه ۲۲ استفاده شده است (Saxena & Prathipati, 2003; Pourbasheer, et al., 2014; Nekoei, et al., 2011).

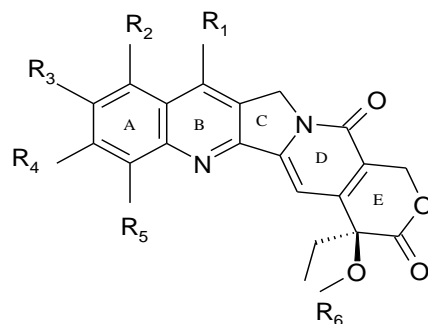
توصیف کننده های مولکولی

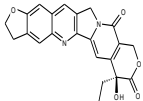
توصیف کننده ها که از آن ها به عنوان متغیر های مستقل نیز یاد می شود، اطلاعات فیزیکی و شیمیایی مولکول ها را با استفاده از روش های مختلف نظیر روابط ریاضی، نظریه گراف و ساختار هندسی به یک عدد تبدیل می نمایند و دارای انواع گوناگونی مانند توصیف کننده های توپولوژیکی^۳، ساختاری^۴، خودهمبستگی دوبعدی^۵، هندسی^۶ و توصیف کننده های مبتنی بر ماتریس دو بعدی^۷ می باشند (Todeschini & Consonni, 2008).

طیف وسیعی از توصیف کننده های مولکولی در شیمی نظری و در مطالعات دارویی مورد بررسی قرار گرفته است (Shao, et al., 2018; Zhao, et al., 2020; Devinyak, et al., 2014; Zapadka, et al., 2022; Mauri & Desc, 2020).

جدول ۱- استخلاف ها در موقعیت های ۱ الی ۶ در ساختار کلی مشتقات کمپتوسین مورد مطالعه

Table 1. Substitutions at the 1-6 positions in the template structure of the studied camptothecin derivatives.



Comp. No	R1	R2	R3	R4	R5	R6
1(CPT)	H	H	H	H	H	H
2	H	H	OH	H	H	H
3	H	H	NH ₂	H	H	H
4	H	H	OH	F	H	H
5	H	H	CH ₂ OH	H	H	H
6	C ₂ H ₅	H	H	H	H	H
7	C ₂ H ₅	H	OH	H	H	H
8	H	NO ₂	H	H	H	H
9		--	--	-	-	-
10	H	H	OCH ₃	OH	H	H
11	H	H	H	OCH ₃	OH	H

⁴- Backward Multiple Linear Regression(BW-MLR)

⁵- Statistical Package for Social Science (SPSS)Version 22

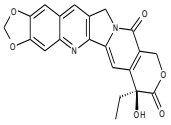
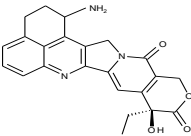
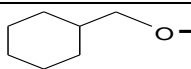
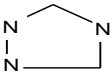
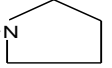
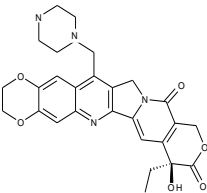
³- Topological

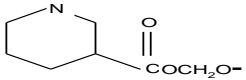
⁴- Constitutional

⁵- 2D autocorrelations

⁶- Geometrical

⁷- 2D matrix-based descriptors

12		--	--	--	--	--
13	NO ₂	H	H	H	H	H
14	NH ₂	H	OCH ₃	F	H	H
15	CH ₃	H	OCH ₃	F	H	H
16	CH ₂ N(CH ₃) ₃	H	OH	H	H	H
17	Si(CH ₃) ₃	H	OH	H	H	H
18	CHO ₂ (CH ₃) ₂	H	H	H	H	H
19	Si(CH ₃) ₃	H	H	F	H	H
20	Si(CH ₃) ₃	H	NH ₂	H	H	H
21	Si(CH ₃) ₃	H	H	NH ₂	H	H
22	C ₂ H ₄ N(CH ₃) ₂	H	H	H	H	H
23		--	--	--	--	--
24	C ₂ H ₄ Si(CH ₃) ₃	H	H	H	H	H
25	H	N(CH ₃) ₂	OH	H	H	H
26	Si(CH ₃) ₃	H	NH ₂	F	H	H
27	Si(CH ₃) ₃	F	NH ₂	H	H	H
28	Si(CH ₃) ₃	H	H	F	F	H
29	Si(CH ₃) ₃	F	F	H	H	H
30	C ₂ H ₄ Si((CH ₃) ₂ OH)	H	H	H	H	H
31	C ₂ H ₄ Si((CH ₃) ₂ C ₂ H ₅)	H	H	H	H	H
32	C ₂ H ₄ Si((CH ₃) ₂ CH ₃ I)	H	H	H	H	H
33	C ₂ H ₄ Si((CH ₃) ₂ CH ₃ OH)	H	H	H	H	H
34	H	H	O(CH ₂ CF ₃)	H	H	H
35	C ₂ H ₄ Si(CH ₃) ₃	H	H	H	H	H
36	Si(CH ₃) ₃	H	OCOCH ₃	H	H	H
37	Si(CH ₃) ₃	H	H	F	Si(CH ₃) ₃	H
38	C ₂ H ₄ Si((CH ₃) ₂ C ₃ H ₆ OH)	H	H	H	H	H
39	C ₂ H ₄ Si((CH ₃) ₂ C ₃ H ₆ Br)	H	H	H	H	H
40	C ₂ H ₄ Si((CH ₃) ₂ C ₃ H ₆ I)	H	H	H	H	H
41	C ₂ H ₄ Si((CH ₃) ₂ CH ₂ N ₃)	H	H	H	H	H
42	C ₂ D ₄ Si(CH ₃) ₃	H	H	H	H	H
43	C ₂ H ₄ Si(CH ₃) ₃	H	H	H	H	OCH ₂ COOCH ₃
44	C ₂ H ₄ Si((CH ₃) ₂ CH ₂ OCOCH ₃)	H	H	H	H	OCH ₂ COOCH ₃
45	C ₂ H ₄ Si((CH ₃) ₂ NH ₂ CH ₃)	H	H	H	H	OCH ₂ COOCH ₃
46	C ₂ H ₅	H		H	H	H
47	C ₂ H ₄ Si((CH ₃) ₂ C ₃ H ₆ N(CH ₃) ₂)	H	H	H	H	H
48	C ₂ H ₄ Si(CH ₃) ₂ CH ₂ -N 	H	H	H	H	H
49	C ₂ H ₄ Si(CH ₃) ₂ CH ₂ -N 	H	H	H	H	H
50		--	--	--	--	--
51	H	H	H	H	H	COO(CH(CH ₃)(NH ₂))
52	Si(CH ₃) ₃	H	NHCOO(O(CH ₃) ₃)	H	H	H
53	C ₂ H ₄ Si((CH ₃) ₂ C ₃ H ₆ Br)	H	H	H	H	OCOCH ₃
54	C ₂ H ₄ Si((CH ₃) ₂ CH ₂ N ₃)	H	H	H	H	OCOCH ₃
55	C ₂ H ₄ Si((CH ₃) ₂ CH ₂ NHCOON(CH ₃) ₂)	H	H	H	H	H

56	C_2H_5	H		H	H	H
57	$Si((CH_3)_3)$	H	$NHCO_2C(CH_3)_3$	F	H	H
58	$Si((CH_3)_3)$	F	$NHCO_2C(CH_3)_3$	H	H	H
59	$C_2H_4Si(CH_3)_2C_3H_6^-$	H	H	H	H	H
60	$C_2H_4Si(CH_3)_2C_3H_6^-$	H	H	H	H	H
61	$C_2H_4Si(CH_3)_2CH_2-N$	H	H	H	H	H
62	$C_2H_4Si(CH_3)_2CH_2S$	H	H	H	H	H
63	$CH_2 NHCH_2Si(CH_3)_2C_2H_4 -$	H	H	H	H	H
64	$C_2H_4Si((CH_3)_2C_6H_6PH(OCH_3)_2)$	H	H	H	H	H
65	$C_2H_4Si((CH_3)_2C_6H_6N_3)$	H	H	H	H	$OCOOCH_3$
66	$C_2H_4Si((CH_3)_2C_6H_6NHCOON(CH_3)_2)$	H	H	H	H	H
67	$C_2H_4Si((CH_3)_2CH_2NHCOON(CH_3)_2)$	H	H	H	H	H
68	$C_3H_5 Si(CH_3)_2C_2H_4 -$	H	H	H	H	H
69	$C_2H_4Si((CH_3)_2CH_2NHCOOC(F)_3)$	H	H	H	H	H
70	$C_2H_4Si((CH_3)_2CH_2NHCOON(CH_3)_2)$	H	H	H	H	$COOCH_3$
71	$C_2H_4Si((CH_3)_2CH_2N(C_2H_5)COONH(CH_3))$	H	H	H	H	H
72	$C_2H_4Si((CH_3)_2C_3H_6NHCOON(CH_3)_2)$	H	H	H	H	$COOCH_3$
73	$C_2H_4Si((CH_3)_2NHN(CH_3))$	H	H	H	H	$COOCH_3$
74	$C_2H_4Si((CH_3)_2CH_2NHCOON(C_2H_5)_2)$	H	H	H	H	$COOCH_3$
75	$C_2H_4Si((CH_3)_2C_3H_6NHCOOCF_3)$	H	H	H	H	$COOCH_3$
76	$C_2H_4Si((CH_3)_2C_3H_6NHCOON(CH_3)_2)$	H	H	H	H	$COOCH_3$

نتایج و بحث

مدل‌های اولیه بدست آمده برای لگاریتم ضریب تقسیم اکتانول-آب ($\log P$)

برای بررسی ارتباط کمی بین ساختار شیمیایی ترکیبات مورد مطالعه و $\log P$ آن‌ها، از روش الگوریتم ژنتیک - رگرسیون چند متغیره خطی (GA-MLR) استفاده شد. ابتدا ۷۶ مشتق CPT به طور تصادفی به یک مجموعه آموزشی شامل ۶۰ مولکول (۸۰٪ داده‌ها) و یک مجموعه آزمون شامل ۱۶ مولکول (۲۰٪ داده‌ها) تقسیم شد. مجموعه آموزشی برای ساخت مدل و مجموعه آزمون برای آزمودن مدل ساخته شده استفاده گردید.

برای ارزیابی عملکرد مدل خطی از ضرایب آماری معمول نظیر: ریشه میانگین مربعات خطا (RMSE)، مجذور ضریب همبستگی (R^2)، آماره دوربین واتسون (DW)، ضریب تعیین تعدیل شده (R^2_{adj})، سطح معناداری (Sig) و ضریب فیشر (F) استفاده شد (Chatterjee & Simonoff, 2013; Hyndman & Koehler, 2006).

جهت پیش بینی $\log P$ با استفاده از روش GA-MLR، ۱۴ مدل با ۱۲ تا ۲۵ توصیف کننده بدست آمد (جدول ۳ را ببینید). با توجه به اینکه ضرایب آماری این مدل‌ها تقریباً مشابه می‌باشند لذا مدل ۱۴ که کمترین تعداد توصیف کننده را

دارد به عنوان مدل اولیه انتخاب شده است. مدل QSPR (معادله (۱)) و ضرایب آماری برای مدل ۱۴ که دارای ۱۲ دوازده توصیف کننده می باشد به صورت زیر است:

$$\text{LogP} = -142.782 + 0.253(\text{Sv}) - 17.009(\text{Jhetv}) + 11.275(\text{Jhetp}) - 0.007(\text{D}/\text{Dr05}) + 8.912(\text{MWC08}) + 46.055(\text{PCR}) - 0.173(\text{PCD}) - 3.078(\text{ATS8e}) + 2.647(\text{ATS8p}) - 3.545(\text{MATS4m}) - 1.159(\text{GATS1m}) - 4.101(\text{GATS1e}) \quad (1)$$

$$N=60 \quad R = 0.974 \quad R^2=0.949 \quad R^2_{\text{adj}}=0.967 \quad DW=1.910 \quad RMSE=0.448 \quad F = 265.639 \quad \text{Sig}=0.000$$

جدول ۳- نام توصیف کننده ها و ضرایب آماری مدل های بدست آمده برای logP مشتقات CPT با روش GA-MLR

Table 3. The name of descriptors and statistical coefficients of the QSAR models derived by GA-MLR analysis for logP_{OW} of CPTs.

Model	Independent variables	R	R ²	R ² _{adj}	RMSE	F	Sig
1	GATS4p, MAXDP, Jhetv, MATS4m, AMW, PCR, DDr05, GATS1e, MATS8p, GATS1m, MATS4p, GATS5m, DDr10, GATS3m, SMTIV, GATS3p, X1v, MWC08, ATS8e, piID, ATS8p, Jhetp, Sv, MW, PCD	0.999	0.998	0.967	0.474	51.748	0.000
2	GATS4p, MAXDP, Jhetv, MATS4m, AMW, PCR, DDr05, GATS1e, MATS8p, GATS1m, GATS5m, DDr10, GATS3m, SMTIV, GATS3p, X1v, MWC08, ATS8e, piID, ATS8p, Jhetp, Sv, MW, PCD	0.999	0.998	0.968	0.468	67.164	0.000
3	GATS4p, MAXDP, Jhetv, MATS4m, AMW, PCR, DDr05, GATS1e, MATS8p, GATS1m, GATS5m, GATS3m, SMTIV, GATS3p, X1v, MWC08, ATS8e, piID, ATS8p, Jhetp, Sv, MW, PCD	0.999	0.998	0.969	0.461	69.611	0.000
4	GATS4p, MAXDP, Jhetv, MATS4m, AMW, PCR, DDr05, GATS1e, MATS8p, GATS1m, GATS5m, GATS3m, SMTIV, GATS3p, X1v, MWC08, ATS8e, ATS8p, Jhetp, Sv, MW, PCD	0.999	0.998	0.969	0.455	75.095	0.000
5	GATS4p, MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, MATS8p, GATS1m, GATS5m, GATS3m, SMTIV, GATS3p, X1v, MWC08, ATS8e, ATS8p, Jhetp, Sv, MW, PCD	0.999	0.998	0.970	0.449	77.625	0.000
6	GATS4p, MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, MATS8p, GATS1m, GATS3m, SMTIV, GATS3p, X1v, MWC08, ATS8e, ATS8p, Jhetp, Sv, MW, PCD	0.999	0.998	0.971	0.445	86.170	0.000
7	GATS4p, MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, MATS8p, GATS1m, SMTIV, GATS3p, X1v, MWC08, ATS8e, ATS8p, Jhetp, Sv, MW, PCD	0.998	0.996	0.971	0.441	98.720	0.000
8	GATS4p, MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, GATS1m, SMTIV, GATS3p, X1v, MWC08, ATS8e, ATS8p, Jhetp, Sv, MW, PCD	0.996	0.992	0.972	0.438	100.012	0.000
9	GATS4p, MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, GATS1m, SMTIV, GATS3p, MWC08, ATS8e, ATS8p, Jhetp, Sv, MW, PCD	0.99	0.990	0.972	0.435	109.308	0.000
10	MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, GATS1m, SMTIV, GATS3p, MWC08, ATS8e, ATS8p, Jhetp, Sv, MW, PCD	0.993	0.986	0.972	0.434	149.951	0.000
11	MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, GATS1m, SMTIV, GATS3p, MWC08, ATS8e, ATS8p, Jhetp, Sv, PCD	0.988	0.976	0.972	0.438	171.530	0.000
12	MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, GATS1m, SMTIV, MWC08, ATS8e, ATS8p, Jhetp, Sv, PCD	0.985	0.970	0.972	0.439	201.983	0.000
13	MAXDP, Jhetv, MATS4m, PCR, DDr05, GATS1e, GATS1m, MWC08, ATS8e, ATS8p, Jhetp, Sv, PCD	0.979	0.958	0.971	0.445	225.600	0.000
14	Jhetv, MATS4m, PCR, DDr05, GATS1e, GATS1m, MWC08, ATS8e, ATS8p, Jhetp, Sv, PCD	0.974	0.949	0.970	0.449	265.639	0.000

بررسی چند خطی^۱ یا همبستگی بین متغیرهای مستقل

هنگامی که دو یا چند متغیر مستقل (توصیف کننده های مولکولی) در یک مدل رگرسیون چندگانه با یکدیگر همبستگی دارند، چند خطی بودن یا هم خطی بودن باید بررسی شود (Perie, et al., 2008 & Pourbasheer, et al., 2017). همبستگی نزدیک به صفر بین توصیف کننده ها گویای مستقل بودن آنهاست در حالیکه همبستگی نزدیک به یک مشخص می نماید بین توصیف کننده ها همبستگی کاملی وجود دارد و یکی از دو توصیف کننده ها باید حذف گردد.

¹- Multicollinearity

همبستگی بین توصیف کننده های مولکولی در مدل (۱) با محاسبه ضریب نفوذ پذیری یا تاثیر پذیری^۱ (VIF) و ضریب پیرسون^۲ (PCC) با استفاده از برنامه SPSS بررسی شده است (O'Brien, 2007; Reisfeld, *et al.*, 2013; Craney & Surles, 2002).

در صورتی که VIF عددی بین یک و ده را داشته باشد مشخص کننده این است که همبستگی کمی بین توصیف کننده ها وجود دارد اما ضریب پیرسون هر چه به صفر نزدیکتر باشد گویای عدم همبستگی بین دو متغیر مستقل می باشد. اگر VIF بالاتر از ۱۰ یا کوچکتر از ۱ باشد، نشان می دهد که مشکل چند خطی وجود دارد و برخی از متغیرها باید حذف شوند (Prévost, 2004).

در جدول (۴) ضرایب فوق به همراه تولرانس^۳ (عکس ضریب نفوذ پذیری) برای دوازده توصیف کننده مولکولی logP که عبارتند از: Sv, Jhetp, Jhetv, D/Dr05, MWC08, PCR, PCD, ATS8e, ATS8p, MATS4m, GATS1m و GATS1e درج گردیده است. همانطور که در جدول ۴ مشاهده می گردد، مقدار ضریب پیرسون بین دو توصیف کننده (Jhetp, Jhetv)، (ATS8e, ATS8p) و (PCR, PCD) نزدیک به یک است، و VIF برای این توصیف کننده ها بزرگتر از ۱۰ می باشد، بنابراین بین این توصیف کننده ها چند خطی وجود دارد. پس از حذف توصیف کننده Jhetv با بیشترین مقدار VIF (۲۲۲/۱۵۷) و ضریب پیرسون (۰/۹۶۸) با توصیف کننده Jhetp از معادله (۱) و در مراحل بعدی حذف توصیف کننده های Sv و ATS8p از این مدل، معادله جدید به صورت زیر حاصل گردید:

$$-3.673 + 1.468 (\text{ATS8e}) \quad (2) \log P =$$

$$F= 555.376 \quad \text{Sig}=0.000 \quad N=60 \quad R= 0.949 \quad R^2=0.901 \quad R^2_{\text{adj}}=0.927 \quad DW=2.009 \quad \text{RMSE}=0.706$$

جدول ۴- نتایج همبستگی بین توصیف کننده ها در معادله (۱).

Table 4. The results of correlation between the molecular descriptors in Equation (1).

Pearson Correlation for logP												
Descriptor	Jhetv	MATS4m	PCR	DDr05	GATS1e	GATS1m	MWC08	ATS8e	ATS8p	Jhetp	Sv	PCD
Jhetv	1.00											
MATS4m	0.380	1.00										
PCR	-0.617	-0.103	1.00									
DDr05	0.397	0.349	-0.180	1.00								
GATS1e	0.119	-0.088	-0.195	-0.033	1.00							
GATS1m	0.237	0.409	-0.152	0.058	-0.036	1.00						
MWC08	-0.552	-0.165	0.817	-0.319	-0.022	-0.234	1.00					
ATS8e	-0.385	0.035	0.089	0.226	-0.018	0.012	-0.074	1.00				
ATS8p	0.322	0.123	-0.065	-0.048	-0.041	0.058	-0.117	-0.950	1.00			
Jhetp	-0.968	-0.433	0.455	-0.369	-0.038	-0.288	0.430	0.476	-0.446	1.00		
Sv	-0.135	-0.245	0.459	-0.049	-0.268	-0.251	0.349	-0.217	-0.342	0.069	1.00	
PCD	0.691	0.118	-0.960	0.106	0.246	0.149	-0.717	-0.179	0.078	-0.528	-0.451	1.00

¹- Variance inflation factor

²- Pearson coefficient correlation

³- Tolerance

ادامه جدول ۴

Collinearity Statistical					Corrected model
Descriptor	Tolerance = $\frac{1}{VIF}$	VIF1	VIF2	VIF3	VIF4
Jhetv	0.005	222.157	-	-	-
MATS4m	0.438	2.282	1.675	-	-
PCR	0.029	34.600	9.923	1.059	-
DDr05	0.339	3.011	-	-	-
GATS1e	0.479	2.087	1.836	1.574	-
GATS1m	0.361	2.767	2.460	1.965	-
MWC08	0.081	12.320	6.874	11.559	-
ATS8e	0.028	36.282	19.438	-	1.00
ATS8p	0.023	43.618	25.597	13.471	-
Jhetp	0.007	149.596	-	-	-
Sv	0.026	38.221	35.263	-	-
PCD	0.008	129.222	20.502	-	-

اعتبار سنجی^۱

برای ارزیابی دقت نتایج و ارزیابی معتبر بودن مدل بدست آمده باید از روش های اعتبار سنجی (Cramer III, 1988 & (Votano, 2004; Gramatica, 2007, Roy & Mitra, 2011, Saghaie, *et al.*, 2013) در این تحقیق از روش اعتبار سنجی متقابل یا ضربدری^۲ و روش اعتبار سنجی خارجی^۳ (Consonni, *et al.*, 2010; Chirico & Gramatica, 2011) استفاده شده است. برای این منظور از همان ابتدا مشتقات مورد مطالعه به صورت تصادفی به دو دسته آموزشی (۶۰ ترکیب) و آزمون (۱۶ ترکیب) تقسیم شدند. در این مطالعه از اعتبار سنجی متقابل یکی بیرون Q^2_{LOO} (Vehtari, *et al.*, 2017; Consonni, *et al.*, 2009) استفاده شده است. برای این منظور هر بار یکی از ترکیبات مجموعه آموزشی به صورت تصادفی حذف و مقدار مجذور همبستگی با استفاده از معادله (۳) محاسبه گردید این عمل برای ۵۰٪ مجموعه آموزشی تکرار گردید و سپس میانگین مجذور همبستگی به صورت Q^2_{LOO} محاسبه شد. در معادله (۳)، Y_{Exp} بیانگر مقدار مشاهده شده یا مقدار تجربی، Y_{pred} بیانگر مقدار پیش بینی شده یا محاسبه شده و \bar{Y} بیانگر مقدار میانگین می باشد. مقدار Q^2_{LOO} برای لگاریتم ضریب تقسیم اکتانول-آب ۰/۹۱۹ محاسبه گردید.

$$(۳) \quad Q^2 = 1 - \frac{\sum(Y_{pred} - \bar{Y}_{Exp})^2}{\sum(Y_{pred} - \bar{Y})^2} \quad Q^2 \leq 1$$

اعتبار سنجی خارجی

برای ارزیابی خارجی و بررسی پایداری مدل QSPR بدست آمده (معادله (۲))، ضرایب رگرسیون مانند R^2 ، R^2_{adj} ، $RMSE$ ، DW ، F و Sig با استفاده از معادله (۲) برای ۱۶ ترکیب مجموعه آزمون (مجموعه پیش بینی) بدست آورده شد و با ضرایب آماری مجموعه آموزشی مقایسه انجام گردید. نتایج این بررسی در جدول ۵ ثبت شده است. همانطور که در این جدول مشاهده می شود، مقادیر ضرایب آماری مانند R^2 و $RMSE$ برای مجموعه های آموزشی و آزمون نزدیک به

^۱- Validation

^۲- Cross-validation (CV)

^۳- External validation

^۱- Leave - one - out (Q^2_{LOO})

یکدیگر هستند و مقادیر بالای R^2 ، R^2_{adj} و F به همراه مقادیر پایین $RMSE$ نشان می‌دهد که مدل ساخته شده به خوبی توانایی پیش‌بینی $\log P$ را دارد.

جدول ۵- ضرایب آماری به دست آمده برای مجموعه آزمون و آموزشی $\log P$ بر اساس معادله (۲).

Table 5. Statistical parameters obtained for the training and test sets of $\log P$ based on Equation (3).

Data set	Property	N	R	R^2	R^2_{adj}	RMSE	DW	F	sig
training	$\log P$	60	0.949	0.901	0.927	0.706	2.009	555.376	0.000
test	$\log P$	16	0.952	0.906	0.934	0.812	1.320	188.189	0.000

باقیمانده^۱

روش دیگر برای اعتبار سنجی مدل ها، مقایسه بین مقادیر مشاهده شده و پیش‌بینی شده است. در آمار، باقیمانده تفاوت بین مقادیر مشاهده شده و پیش‌بینی شده می‌باشد. یک نمودار باقیمانده نمایش مقادیر باقیمانده در روی محور عمودی و مقادیر مشاهده شده بر روی محور افقی می‌باشد. مقایسه بین مقادیر باقیمانده و مشاهده شده $\log P$ برای مجموعه های آموزشی و آزمون در شکل (۲) نشان داده شده است. همانطور که در شکل (۲) نشان داده شده است توزیع خطاها در دو طرف محور افقی دارای یک الگوی تصادفی می‌باشد که نشان می‌دهد هیچ خطای سیستماتیک یا سوگیری شده در مدل توسعه یافته QSPR وجود ندارد.

مقایسه بین مقادیر پیش‌بینی شده و مشاهده شده $\log P$ برای مجموعه های آموزشی و آزمون در شکل (۳) نشان داده شده است. این مجموعه داده ها روی یک خط می‌باشند هستند که نشان می‌دهد مقادیر پیش‌بینی شده بر اساس روش GA-MLR مطابقت خوبی با مقادیر مشاهده شده دارند.

در جدول (۲) مقادیر مشاهده شده، پیش‌بینی شده و باقیمانده برای لگاریتم ضریب تقسیم ۷۶ مشتق آکالوئید مورد مطالعه آورده شده است.

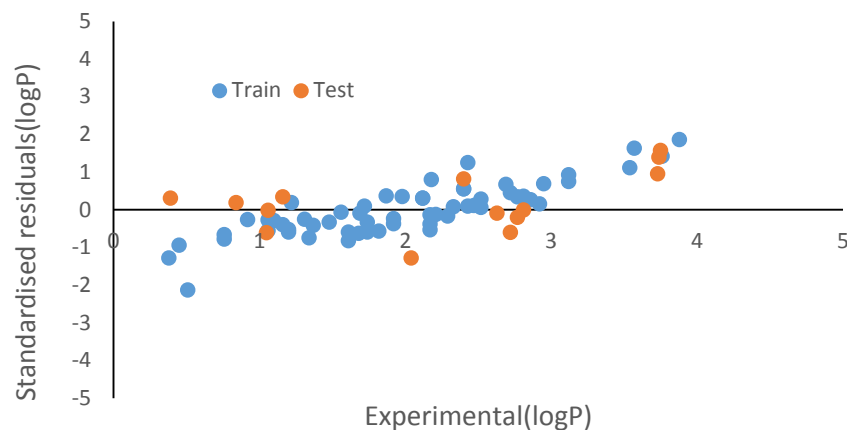
²- Residual

جدول ۲- مقادیر مشاهده شده، پیشگویی شده و باقیمانده برای لگاریتم ضریب تقسیم مشتقات کمپتوتسین مورد مطالعه با استفاده از روش

GA- MLR

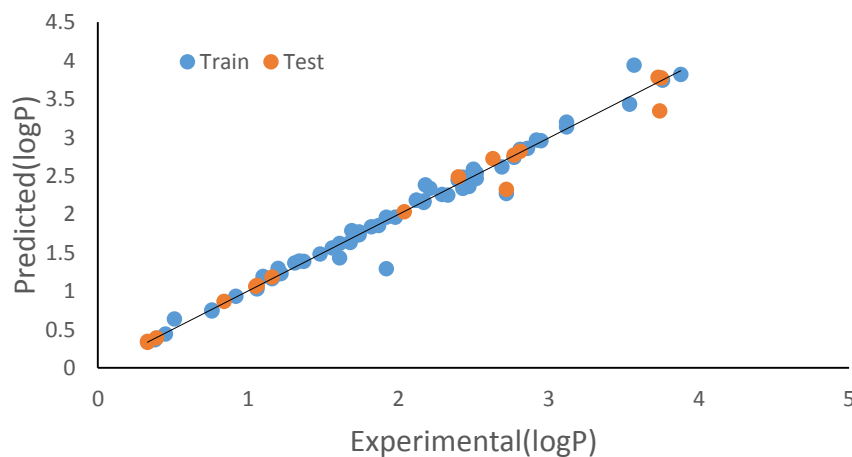
Table 2. The observed, predicted and residual values of logP of the studied camptothecin derivatives.

No	logP (Exp)	logP (pred)	logP (Res)	No	logP (Exp)	logP (pred)	logP (Res)
1	1.2200	1.0303	0.1897	39	3.1200	2.1991	0.9209
2	0.9200	1.1815	- 0.2615	40	3.7400	2.3471	1.3929
3	0.3900	0.3020	0.0888	41	2.5200	2.2372	0.2828
4	1.0600	1.3298	-0.2698	42	2.4000	1.8452	0.5548
5	0.7462	1.3900	-0.6438	43	3.1200	2.3738	0.7462
6	2.1800	2.7186	- 0.5386	44	2.1700	2.7086	- 0.5386
7	1.8700	1.5016	- 0.2409	45	1.9200	2.1609	- 0.2409
8	1.1600	0.8190	0.3410	46	3.7500	2.1759	1.5741
9	1.1000	1.3900	-0.2900	47	2.3300	2.2504	0.0796
10	0.7600	1.4194	- 0.5670	48	1.8200	2.3870	- 0.5670
11	0.7600	1.5412	-0.7812	49	2.4700	2.3635	0.1065
12	0.8400	0.6510	0.1890	50	0.5100	2.6381	- 2.1281
13	1.1600	1.4162	- 0.2562	51	1.3100	1.5662	- 0.2562
14	1.5636	1.6587	- 0.0951	52	2.6300	2.7251	- 0.0951
15	1.7200	1.6293	0.0907	53	3.5400	2.4311	1.1089
16	1.1876	1.1324	0.0552	54	2.5200	2.4648	0.0552
17	1.5600	1.6249	- 0.0649	55	1.7400	2.3385	- 0.5985
18	1.3700	1.7879	-0.4179	56	2.7200	2.2695	0.4505
19	1.9800	1.6352	0.3448	57	2.7700	2.4296	0.3404
20	1.0600	1.0742	- 0.0142	58	2.7700	2.9706	- 0.2006
21	1.0600	1.6117	-0.5517	59	2.1700	2.3062	- 0.1362
22	1.6900	1.7850	- 0.095	60	1.6800	2.3107	- 0.6307
23	1.4800	1.8099	- 0.3299	61	2.8100	2.4501	0.3599
24	2.4000	1.8452	0.5548	62	3.7600	2.3444	1.4156
25	0.9450	2.2200	- 1.2750	63	3.7300	2.7833	0.9467
26	1.2000	1.7248	- 0.5248	64	2.2100	2.3385	- 0.1285
27	1.2000	1.7938	- 0.5938	65	2.9500	2.2593	0.6907
28	2.1200	1.8143	0.3057	66	1.6100	2.4340	- 0.8240
29	2.1200	1.8143	0.3057	67	2.4300	2.3385	0.0915
30	1.0500	1.6622	- 0.6122	68	2.8100	2.8220	- 0.0120
31	2.6900	2.0170	0.6730	69	2.5000	2.3885	0.1115
32	3.8800	2.0184	1.8616	70	2.1700	2.5515	- 0.3815
33	1.7400	2.0684	- 0.3284	71	2.2900	2.4575	- 0.1675
34	2.4300	1.1815	1.2485	72	2.0400	3.3228	- 1.2828
35	2.4000	1.5879	0.8121	73	1.9200	2.2901	- 0.3701
36	1.3400	2.0919	- 0.7519	74	2.8600	2.5940	0.2660
37	3.5700	1.9406	1.6294	75	2.9200	2.7658	0.1542
38	1.6100	2.2064	- 0.5964	76	2.7200	3.3228	-0.6028



شکل ۱- مقادیر مشاهده شده بر حسب باقیمانده برای لگاریتم ضریب تقسیم اکتانول-آب برای مجموعه آموزشی و آزمون با روش GA-MLR.

Fig. (1). Plot of the residuals against the observed values of $\log P_{O/W}$ for the training and test sets by the GA-MLR method.



شکل ۲- مقادیر مشاهده شده بر حسب پیشگویی شده برای لگاریتم سمیت برای مجموعه آموزشی و آزمون با روش GA-MLR.

Fig. (2). Plot of the predicted values against observed values of $\log P_{O/W}$ for the training and test sets by the GA-MLR method.

توصیف کننده منتخب

همانطور که در معادله (۲) مشاهده می شود، تنها توصیف کننده ATS8e برای پیش بینی $\log P$ مشتقات CPT نسبت به سایر توصیف کننده ها مناسب می باشد. این توصیف کننده نحوه توزیع یک ویژگی را در طول ساختار توپولوژیکی توصیف می کند و در گروه توصیف کننده های خود همبستگی دو بعدی طبقه بندی می شود و با الکترونگاتیویته ساندرسون ارتباط دارد (Sharma, 2012; Velázquez-Libera, *et al.*, 2019). ساندرسون به همبستگی بین اندازه اتمی و الکترونگاتیویته اشاره کرده است (Sanderson, 1983). با آگاهی از الکترونگاتیویته ساندرسون و طول پیوند، انرژی پیوند

در طیف وسیعی از ترکیبات را می توان تخمین زد (Sanderson, 1983). الکترونگاتیویته ساندرسون برای تحقیقات مختلف در شیمی آلی / معدنی استفاده شد (Trofimov, et al., 2005 & Hajimahdi, et al., 2016).

ضریب توصیف کننده ATS8e دارای علامت مثبت در معادله (۲) می باشد که نشان می دهد $\log P$ مشتقات کمپوتوسین مستقیماً با این توصیف کننده مرتبط است و این ویژگی با افزایش الکترونگاتیویته ساندرسون در مولکول افزایش می یابد. بنابراین، الکترونگاتیویته ساندرسون نقش مهمی در پیش بینی ضریب تقسیم اکتانول-آب ($\log P$) این مشتقات دارد.

نتیجه گیری

مطالعات ارتباط کمی ساختار- خاصیت و ساختار - فعالیت ترکیبات شیمیایی به منظور پیش بینی خواص و فعالیت های ترکیبات جدید به کار می روند. در این مطالعه با استفاده از روش رگرسیون خطی چند متغیره- الگوریتم ژنتیک (GA-MLR) ارتباط کمی لگاریتم ضریب تقسیم اکتانول-آب ($\log P$) و ۷۶ مشتق کینولین آکالوئید کمپوتوسین (CPT) به عنوان حشره کش را با توصیف کننده های محاسبه شده توسط نرم افزار دراگون مورد ارزیابی قرار گرفته است. برای هر ترکیب مورد مطالعه تعداد زیادی توصیف کننده توسط این نرم افزار محاسبه گردید، که با کمک روش الگوریتم ژنتیک تعداد توصیف کننده ها تقلیل داده شد و با استفاده از نرم افزار SPSS و روش GA-MLR و ضرایب آماری مدل ها، مناسب ترین مدل برای فعالیت مورد مطالعه انتخاب گردید و با کمک دو ضریب پیرسون و نفوذ پذیری همبستگی بین توصیف کننده ها در مدل منتخب مورد بررسی قرار گرفت. این تحلیل نشان می دهد که توصیف کننده (ATS8e) که از گروه توصیف کننده های خود همبستگی دو بعدی می باشد همبستگی بالایی با $\log P$ ۷۶ مشتق CPT ایجاد نموده است. مقادیر ضرایب آماری مدل منتخب مانند: مجذور ضریب همبستگی ($R^2 = 0/901$)، ضریب تعیین تعدیل شده ($R^2_{adj} = 0/927$) و ریشه میانگین مجذور خطا ($RMSE = 0/706$) نشان می دهد ارتباط خطی مطلوبی بین متغیر وابسته و مستقل وجود دارد.

در نهایت جهت ارزیابی قدرت پیش بینی مدل منتخب از اعتبار سنجی های داخلی، خارجی و رسم نمودار استفاده گردید. نتایج اعتبار سنجی نشان می دهد که مدل توسعه یافته از استحکام و توانایی پیش بینی خوبی برخوردار است و می توان از آن برای پیش بینی $\log P$ مشتقات جدید CPT استفاده کرد و بدینوسیله در وقت، هزینه صرفه جویی نموده و به محیط زیست نیز آسیب نرساند.

Referance

- Abdolrahimi, S., Nasernejad, B. and Pazuki, G. 2014.** Prediction of partition coefficients of alkaloids in ionic liquids based aqueous biphasic systems using hybrid group method of data handling (GMDH) neural network. *Journal of Molecular Liquids*, 191: 79–84.
- Arnott, J. A. and Planey, S.L. 2012.** The influence of lipophilicity in drug discovery and design. *Expert Opinion on Drug Discovery*, 7: 863–875.
- Blake, L. D. 2016.** Antimalarial Exoerythrocytic Stage Drug Discovery and Resistance Studies. *ProQuest Dissertations and Theses*. 172.
- Bray, P.G., Hawley, S.R., Mungthin, M. and Ward, S.A. 1996.** Physicochemical properties correlated with drug resistance and the reversal of drug resistance in *Plasmodium falciparum*. *Molecular Pharmacology*, 50(6): 1559-1566.
- Chatterjee, S. and Simonoff, J. S. 2013.** *Handbook of Regression Analysis*, John Wiley & Sons, Hoboken, NJ.
- Chirico, N.N. and Gramatica, P. 2011.** Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling*, 51; 2320- 2335.
- Chisholm, H. 1911.** Quinoline, *Encyclopædia Britannica*. Vol. 22 (11th ed.). Cambridge University Press.
- Collin, G., Höke, H. 2005.** Quinoline and Isoquinoline. *Ullmann's Encyclopedia of Industrial Chemistry*. Weinheim: Wiley-VCH.
- Consonni, V., Ballabio, D. and Todeschini, R. 2010.** Evaluation of model predictive ability by external validation techniques. *Journal of Chemometrics*, 24(3-4): 194- 201.
- Consonni, V., Ballabio, D. and Todeschini, R. 2009.** Comments on the definition of the Q2 parameter for QSAR validation. *Journal of Chemical Information and Modeling*, 49(7):1669-1678.
- Cramer III, R. D., Bunce, J. D., Patterson, D. E. and Frank, I. E. 1988.** Cross validation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Molecular Informatics*, 7: 18–25.
- Craney, T. A. and Surlles, J. G. 2002.** Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, 14: 391-403.
- Czeszak, A., Resztak, M., Czyrski, A. and Nowak, I. 2020.** Determination of partition coefficient of isoquinoline alkaloids from *Chelidonium Majus* by reversed phase thin layer chromatography. *New Journal of Chemistry*, 44: 7484-7489.
- Eisner, T., Morgan, R.C., Attygalle A.B., Smedley, S.R., Herath, K.B., Meinwald, J. 1997.** Defensive Production of quinoline by a phasmid insect (*Oreophoetes peruana*). *Journal of Experimental Biology*. 200: 2493–2500.
- Gerhardt, C. 1842.** Untersuchungen über die organischen Basen (Investigations of organic bases), *Annalen der Chemie und Pharmacie*, 42 : 310-313.
- Ghasemi, J. and Saaidpour, S. 2007.** Quantitative structure–property relationship study of n-octanol–water partition coefficients of some of diverse drugs using multiple linear regression. *Analytica Chimica Acta*, 604: 99–106.
- Goodarzi, M., Dejaegher, B. and Heyden, Y. V. 2012.** Feature selection methods in QSAR studies. *Journal of AOAC International*, 95(3): 636-651.
- Gramatica, P. 2007.** Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, 2007, 26(5), 694-701.
- Hajimahdi, Z., Safizadeh, F. and Zarghi, A. 2016.** QSAR Analysis for Some 1, 2-Benzisothiazol-3-one Derivatives as Caspase-3 Inhibitors by Stepwise MLR Method. *Iranian Journal of Pharmaceutical Sciences*, 15(2): 439-448.
- Hansch, C. and Kurup, A. 2003.** QSAR of chemical polarizability and nerve toxicity. *Journal for Chemical Information and Computer scientists*, 43: 1647-1651.
- Han, S.Y., Qiao, J.Q., Zhang, Y.Y., Yang, L.L., Lian, H.Z., Ge, X. and Chen, H.Y. 2011.** Determination of n-octanol/water partition coefficient for DDT-related compounds by RP-HPLC with a novel dual-point retention time correction. *Chemosphere*, 83(2): 131-136.

- Hou, T. J., Wang, J. M., Liao, N. and Xu, X. J. 1999.** Applications of genetic algorithms on the structure-activity relationship analysis of some cinnamamides. *Journal for Chemical Information and Computer scientists*, 39: 775- 781.
- Hu, Y. Unwalla, R. Denny, R.A. Bikker, J. Di, L. and Humblet, C. 2010.** Development of QSAR models for microsomal stability: Identification of good and bad structural features for rat, human and mouse microsomal stability. *Journal of Computer-Aided Molecular Design*, 24, 23-35.
- Hussain, I., Bania, K.K., Gour, N.K. and Deka, R.C. 2016.** Application of Physicochemical and DFT Based Descriptors for QSAR Study of Camptothecin Derivatives. *Chemistry Select*, 1(15): 4973-4978.
- Hyndman, R.J. and Koehler, A.B. 2006.** Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22 (4): 679–688.
- Kondaparla, S., Soni, A., Manhas, A., Srivastava, K., Puri, S.K and Katti, S.B. 2017.** Antimalarial activity of novel 4-aminoquinolines active against drug resistant strains. *Bioorganic Chemistry*, 70: 74–85.
- Kujawski, J., Bernard, M. K., Janusz, A. and Kuźma, W. 2011.** Prediction of log P: ALOGPS Application in Medicinal Chemistry Education. *Journal of Chemical Education*, 89(1): 64–67.
- Kwon, Y. 2001.** Partition and Distribution Coefficients, *Handbook of Essential Pharmacokinetics, Pharmacodynamics and Drug Metabolism for Industrial Scientists*. (secondary). New York: Kluwer Academic/Plenum Publishers.
- Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. 2001.** Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* , 46 (1–3): 3–26.
- Lopez, K., Pinheiro, S., and Zamora, W. J. 2021.** Multiple linear regression models for predicting the n-octanol/water partition coefficients in the SAMPL7 blind challenge. *Journal of computer-aided molecular design*, 35: 923-931.
- Manohar, S., Tripathi, M. and Rawat, D. S. 2014.** 4-Aminoquinoline based molecular hybrids as antimalarials: An overview. *Current Topics in Medicinal Chemistry*, 14: 1706–1733.
- Mauri, A. and Desc, A. 2020.** A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints. In: Roy, K. (eds) *Ecotoxicological QSARs. Methods in Pharmacology and Toxicology*. Humana, New York.
- Murugan, K., Panneerselvam, C., Subramaniam, J., Paulpandi, M., Rajaganesh, R., Vasanthakumaran, M., Madhavan, J., Shafi, S.S., Roni, M., Portilla-Pulido, J.S., Mendez, S.C., Duque, J.E., Wang, L., Aziz, A.T., Chandramohan, B., Dinesh, D., Piramanayagam, S. and Hwang, J.S. 2022.** Synthesis of new series of quinoline derivatives with insecticidal effects on larval vectors of malaria and dengue diseases. *Scientific Reports*, 12(1):4765-4772.
- Murugan, K. Madhavan, J. and Samidoss, C.M. 2018.** Bismuth oxyiodide nanoflakes showed toxicity against the malaria vector *Anopheles stephensi* and in vivo antiplasmodial activity. *Journal of Cluster Science*, 29(2), 337–344.
- Narender, S., Sidhartha, C., Ruifeng, L., Mohamed Diwan, M.A., Gregory, T. and Anders, W. 2012.** QSAR classification model for antibacterial compounds and its use in virtual screening. *Journal of Chemical Information and Modeling*, 52: 2559-2569.
- Nekoei, M., Salimi, M., Dolatabadi, M. and Mohammadhosseini, M. 2011.** Prediction of antileukemia activity of berbamine derivatives by genetic algorithm–multiple linear regression. *Monatshefte für Chemie*, 142; 943- 948.
- O'Brien, R.M. 2007.** A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41: 673-690.
- O'Loughlin, E. J., Kehrmeier, S. R. and Sims, G. K. 1996.** Isolation, characterization, and substrate utilization of a quinoline-degrading bacterium. *International Biodeterioration & Biodegradation*, 38 (2): 107-114.
- Padron, J.A., Carrasco, R. and Pellon, R.F. 2002.** Molecular descriptor based on a molar refractivity partition using Randic-type graph-theoretical invariant. *Journal of Pharmaceutical Sciences*, 5(3): 258-266.

- Pan, X.D. and Wang, C.Y. 2003.** Current status of camptothecin derivatives as natural antitumor agents. *Yaoxue Xuebao*, 38: 715-720.
- Perie, C. and Ouimet, R. 2008.** Organic carbon, organic matter and bulk density relationships in boreal forest soils. *Canadian Journal of Soil Science*, 88: 315-325.
- Pourbasheer, E., Aalizadeh, R., Ganjali, M.R., Norouzi, P., Shadmanesh, J. 2014.** QSAR study of ACK1 inhibitors by genetic algorithm-multiple linear regression (GA-MLR). *Journal of Saudi Chemical Society*, 18: 681- 688.
- Pourbasheer, E., Ahmadpour, S., Zare-Dorabei, R. and Nekoei, M. M. 2017.** Quantitative structure activity relationship study of p38a MAP kinase inhibitors. *Arabian Journal of Chemistry*, 10 (1): 33-40.
- Prévost, M. 2004.** Predicting soil properties from organic matter content following mechanical site preparation of forest soils. *Soil Science Society of America Journal*, 68: 943-949.
- Rageh, A. H., Atia, N. N. and Abdel-Rahman, H. M. 2017.** Lipophilicity estimation of statins as a decisive physicochemical parameter for their hepato-selectivity using reversed-phase thin layer chromatography. *Journal of Pharmaceutical and Biomedical Analysis*, 142: 7-14.
- Reisfeld, B. and Mayeno, A. N. 2013.** *Computational Toxicology: Volume 21, On the development and validation of QSAR models*, Springer: Science+Business Media, LLC.
- Roy, K. and Mitra. I. 2011.** On Various Metrics Used for Validation of Predictive QSAR Models with Applications in Virtual Screening and Focused Library Design. *Combinatorial Chemistry & High Throughput Screen*, 14: 450-474.
- Ruiz, P., Begluitti, G., Tincher, T., Wheeler, J. and Mumtaz, M. 2012.** Prediction of acute mammalian toxicity using QSAR methods: A case study of sulfur mustard and its breakdown products. *Molecules*, 17: 8982-9001.
- Rutkowska, E., Pająk, K. and Józwiak, K. 2013.** Lipophilicity: methods of determination and its role in medicinal chemistry. *Acta Poloniae Pharmaceutica* , 70: 3-18.
- Saghaie, L., Sakhi, H., Sabzyan, H., Shahlaei, M. and Shamshirian, D. 2013.** Stepwise MLR and PCR QSAR study of the pharmaceutical activities of antimalarial 3-hydroxypyridinone agents using B3LYP/6- 311++ G** descriptors. *Medicinal Chemistry Research*, 22: 1679-1688.
- Sanderson, R. T. 1983.** Electronegativity and bond energy. *Journal of the American Chemical Society*, 10: 2259- 2261.
- Saxena., A. K. and Prathipati. P. 2003.** Comparison of MLR, PLS and GA-MLR in QSAR analysis. *SAR and QSAR in Environmental Research.*, 14: 433-445.
- Shang, X.F., Morris-Natschke, S.L., Liu, Y.Q., Guo, X., Xu, X.S., Goto, M., Li, J.C., Yang, G.Z. and Lee, K.H. 2018.** Biologically active quinoline and quinazoline alkaloids part I. *Medicinal Research Reviews*, 38(3):775-828.
- Scheytt, T., Mersmann, P., Lindstädt, R. and Heberer, T. 2005.** 1-Octanol/water partition coefficients of 5 pharmaceuticals from human medical care: carbamazepine, clofibrac acid, diclofenac, ibuprofen, and propyphenazone. *Water, air, and soil pollution*, 165(1-4): 3-11.
- Shao, Z., Wu, P., Zhang, X., Dimitrov, D., Liu, J. B. 2018.** On the maximum ABC index of graphs with prescribed size and without pendent vertices, *IEEE Access*, 6: 27604-27616.
- Sharma N., Ethiraj, K.R., Yadav, M., Nayariseri, S. A., Chaurasiya, M., Vankudavath, R.N. and Rao, K.R. 2012.** Identification of LOGP values and Electronegativities as structural insights to model inhibitory activity of HIV-1 capsid inhibitors - a SVM and MLR aided QSAR studies. *Current Topics in Medicinal Chemistry*, 12(16): 1763-1774.
- Shargel, L., Susanna, W. and Yu, A. B. (2012).** "Chapter 10: Physiological Drug Distribution and Protein Binding". *Applied Biopharmaceutics & Pharmacokinetics. (secondary) (6th ed.)*. New York: McGraw-Hill Medical. p. 211.
- Tanizawa, A., Fujimori, A., Fujimori, Y. and Pommier, Y. 1994.** Comparison of topoisomerase I inhibition, DNA damage, and cytotoxicity of camptothecin derivatives presently in clinical trials. *Journal of the National Cancer Institute* , 86: 836-842.
- Thomas, C.J., Rahier, N.J. and Hecht, S.M. 2004.** Camptothecin: current perspectives. *Bioorganic & Medicinal Chemistry*, 12 (7): 1585-1604.

- Todeschini, R. and Consonni, V. 2000.** Handbook of Molecular Descriptors. Weinheim: Wiley-VCH.
- Todeschini, R. and Consonni, V. 2009.** Molecular descriptors for chemoinformatics. Alphabetical listing (2nd ed., Vol. 1). Weinheim: Wiley-VCH.
- Todeschini, R. and Consonni, V. 2008.** Handbook of molecular descriptors, vol 11. Wiley, New Jersey.
- Trofimov, M.I. and Smolenskii, E.A. 2005.** Application of the electronegativity indices of organic molecules to tasks of chemical informatics. Russian Chemical Bulletin, 54(9): 2235-2246.
- Velázquez-Libera, J. L., Caballero, J., Toropova, A. P. and Toropov, A. A. 2019.** Estimation of 2D autocorrelation descriptors and 2D Monte Carlo descriptors as a tool to build up predictive models for acetylcholinesterase (AChE) inhibitory activity. Chemometrics and Intelligent Laboratory Systems, 184: 14-21.
- Vehtari, A., Gelman, A. and Jonah Gabry, J. 2017.** Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing , 27(5):1413-1432.
- Votano, J. R., Parham, M., Hall, L. H., Kier, L. B., Oloff, S., Torpsha, A., Xie, Q. and Tonga, W. 2004.** Three new consensus QSAR models for the prediction of Ames genotoxicity. Mutagenesis, 19: 365- 377.
- Wang, X .H. 2019.** Design, synthesis, and biological activity evaluation of camptothecin-HAA-Norcantharidin conjugates as antitumor agents in vitro. Chemical Biology & Drug Design, 93 (6): 986-992.
- Wall, M.E., Wani, M. C., Cook, C. E., Palmer, K. H., McPhail, A. T. and Sim, G. A. 1966.** Plant Antitumor Agents. I. The isolation and structure of camptothecin, a Novel alkaloidal leukemia and tumor inhibitor from camptotheca acuminata. Journal of the American Chemical Society, 88: 3888-3890.
- Wenlock, M.C. and Carlsson, L.A. 2015.** How experimental errors influence drug metabolism and pharmacokinetic QSAR/QSPR models. Journal of Chemical Information and Modeling, 55:125-134.
- Web search engine developed by ChemAxon. 2019.** software available at "http:// WWW. Chemicalize. Org".
- White, N.J., Looareesuwan, S. and Warrel, D.A. 1983.** Quinine loading dose in cerebral malaria. American Journal of Tropical Medicine and Hygiene, 32:1-5.
- Xu, M., Wagerle, T., Long, J.K., Lahm, G.P., Barry, J.D. and Smith, R.M. 2014.** Insecticidal quinoline and isoquinoline isoxazolines. Bioorganic & Medicinal Chemistry Letters 24: 4026-4030.
- Zhao, D., Siddiqui, M. K., Cheema, I. Z., Muhammad, M. H., Rauf, A. and Ishtiaq, M. 2020.** On molecular descriptors of polycyclic aromatic hydrocarbon. Polycyclic Aromatic Compounds, 1-12.
- Zapadka, M., Dekowski, P. and Kupcewicz, B. 2022.** HATS5m as an Example of GETAWAY molecular descriptor in assessing the similarity/diversity of the structural features of 4-Thiazolidinone. International Journal of Molecular Sciences, 23: 6576-6595.

Prediction of octanol – water partition coefficient of quinoline alkaloid insecticides using Molecular Descriptors and MLR Method

*Sh. Mohebbi¹, F. Shafiei^{*2}, T. Momeni Isfahani³, M. Ahmadi Sabegh⁴*

1. PHD student, Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran
2. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran
3. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran
4. Assistant Professor, Department of Chemistry, Ahar Branch, Islamic Azad University, , Ahar, Iran

Abstract

Quinoline alkaloids and their derivatives have wide medical and agricultural applications. In this research, a quantitative structure- property relationship (QSPR) has been employed to predict the octanol-water partition coefficient (logP) of 76 quinoline alkaloid camptothecin (CPT) derivatives as antitumor potencies using GA-MLR method and molecular descriptors.

The Gauss View 05 software was used for drawing chemical structure of the studied compounds. The geometry optimizations of the studied compounds were done by the Gaussian 09W software at B3YLP density functional theory (DFT) with 6-311G (d,p) basis set.

Molecular descriptors for each of optimized structures were calculated by Dragon software in different category.

In order to reduce and select the best descriptors, the Genetic Algorithm technique and stepwise multiple linear regression method was used.

The pearson coefficient correlation (PCC) and the variance inflation factor (VIF) statistics were used to test the multicollinearity of descriptors in the best model.

The different types of internal and external validations were used to evaluate predictive model performance. The best QSPR model is obtained with R^2 value of 0.901, $Q^2_{LOO} = 0.919$, and $RMSE = 0.706$.

The results of statistical parameters and validations of the GA-MLR model generated were found to be satisfactory. The model revealed that octanol-water partition coefficient of CPT derivatives is influence by ATS8e (2D-autocorrelation) descriptor. This information could be used to design novel quinoline alkaloid camptothecin (CPT) derivatives as insecticide agents.

Keyword: "quinoline alkaloid camptothecin (CPT)" ; "the octanol-water partition coefficient (logP)" ; "2D-autocorrelation descriptor" ; " validation" ; " insecticide"

* Corresponding Author, E-mail: Shafa38@yahoo.com
Received: 28 Dec. 2022 – Accepted: 3 Mar. 2023