

کاربرد الگوریتم ژنتیک به همراه رگرسیون خطی چند گانه برای پیش بینی

فعالیت دارویی مشتقات پیرازول ها

مهدی نکوئی*^۱، مجید محمدحسینی^۱، اسلام پوربشیر^۲، مهدی مهام^۳

۱- دانشگاه آزاد اسلامی، واحد شاهرود، گروه شیمی، شاهرود، ایران

۲- دانشکده شیمی، دانشگاه تهران، تهران، ایران

۳- دانشگاه آزاد اسلامی، واحد علی‌آباد کتول، گروه شیمی، علی‌آباد کتول، ایران

تاریخ دریافت: ۹۰/۷/۱۳

تاریخ پذیرش: ۹۰/۸/۲۸

چکیده:

مطالعه ارتباط کمی ساختار- فعالیت برای پیش بینی فعالیت دارویی مشتقات پیرازول ها، با استفاده از توصیف کننده های ساختاری و روش رگرسیون خطی چند گانه توسعه داده شده است. توصیف کننده های مولکولی با استفاده از الگوریتم ژنتیک انتخاب شدند. سپس یک مدل ساده، قوی، قابل تفسیر و با خطای کم و ضریب همبستگی بالا ساخته شد. نتایج نشان می دهد که تکنیک های خطی مانند رگرسیون خطی چند گانه که با یک روش انتخاب متغیر مناسب کوپل شده باشد، قادر است مدل‌های مناسبی برای پیش بینی فعالیت دارویی ترکیبات ارائه نماید. مدل ارائه شده پارامترهای آماری بالایی را نشان داد که می تواند جهت پیش بینی فعالیت دارویی ترکیبات مشابه مورد استفاده قرار گیرد.

کلمات کلیدی: ارتباط کمی ساختار، فعالیت - مشتقات پیرازول‌ها- رگرسیون خطی چندگانه- الگوریتم ژنتیک.

مقدمه:

یکی از مشکلاتی که جامعه بشری همیشه با آن روبرو بوده، مقابله با انواع بیماری هایی است که سلامت انسان ها را به مخاطره انداخته و همواره یکی از مهمترین دغدغه های محققان یافتن داروهای موثر، برای رفع این معضل و یا کاهش عوارض این بیماری ها بوده است. بروز انواع بیماری ها از قبیل سرطان، بیماری هولناک ایدز، بیماری های مشترک انسان و دام و مقاوم شدن ویروسها در برابر آنتی بیوتیک ها همه از جمله مواردی هستند که ذهن دانشمندان را در جهت یافتن

کاربرد الگوریتم ژنتیک به همراه رگرسیون خطی چند گانه برای پیش بینی ...

داروهای موثر و کارآمد برای مقابله با این بیماری ها معطوف خود نموده اند.^۱ روندی که در گذشته منجر به کشف و توسعه داروهای جدید می شد به روش آزمون و خطا صورت می گرفت که روشی وقت گیر و هزینه بر است. مشکل دیگری که در این راه دانشمندان را آزار می دهد، عدم اطلاع آنها از فعالیت داروئی ترکیبات، قبل از انجام سنتز و بررسی تجربی آنها بوده و به همین دلیل یکی از مهم ترین اهداف شیمیدان ها و محققان دارویی پیش بینی فعالیت ترکیبات دارویی، قبل از سنتز آن ها می باشد. چرا که انجام بسیاری از آزمایشات مستلزم صرف زمان و هزینه های زیادی است. از این رو نیاز به استفاده از روش های تئوری و محاسباتی که بدون انجام آزمایش بتوانند ویژگی و یا فعالیت ترکیبات دارویی را پیش بینی کنند ضروری به نظر می رسد. ظهور علم کمومتریکس توانسته راه حلی برای رفع این مشکلات باشد.^{۲،۳}

اخیراً "مطالعه روی مشتقات پیرازول ها به عنوان بازدارنده های آنزیم p38 α برای اهداف درمانی در معالجه روماتیسم مفصلی و دیگر بیماری های استخوانی و مفصلی به طور صعودی افزایش یافته است. این بازدارنده ها فعالیت آنزیم های p38 α را متوقف می کنند.^۴ این آنزیم ها توسط عوامل مختلف از جمله شوک و ضربه های عصبی، پرتوهای مضر، سایش های مکانیکی و تحریک سیتوکین ها ایجاد می شوند. تولید این آنزیم ها باعث آزاد شدن TNF α و دیگر سیتوکین ها و مهاجرت آنها به سلولهای سفید خونی و ایجاد ورم و التهاب در مفاصل می گردد. بنابراین مطالعه روی بازدارنده های آنزیم p38 α به عنوان اهداف درمانی حوزه مطالعاتی جدیدی می باشد.^{۵،۸} بازدارنده های آنزیم p38 α با بلوکه کردن فعالیت TNF α باعث درمان التهاب مفاصل یا همان روماتیسم می شود. هدف از انجام این تحقیق پیش بینی فعالیت دارویی مشتقات پیرازول ها با استفاده از ژنتیک الگوریتم و رگرسیون خطی چندگانه می باشد.

انتخاب سری داده ها

در این کار تعداد ۴۴ ترکیب از مشتقات پیرازول ها به عنوان بازدارنده های آنزیم p38 α برای درمان التهاب مفاصل یا همان روماتیسم توسط روشهای کمومتریکس مورد بررسی قرار گرفت.^{۹،۱۰} در این مقاله قدرت بازدارندگی این ترکیبات به صورت IC₅₀ گزارش شده است. IC₅₀ عبارتست از مینیمم غلظتی از ترکیب که باعث ۵۰٪ بازدارندگی بیماری می شود. این مقادیر به مقیاس لگاریتمی تبدیل شده $-\log(\text{IC}_{50}) = \text{pIC}_{50}$ و مورد استفاده قرار گرفته است.

در این کار این ترکیبات به صورت تصادفی به دو گروه سری آموزش و سری پیش بینی تقسیم شده است، سری آموزش شامل ۳۵ مولکول و سری پیش بینی شامل ۹ مولکول می باشد. مقادیر pIC₅₀ به عنوان متغیر وابسته و توصیف کننده ها به عنوان متغیر مستقل انتخاب شد. سری آموزش جهت ایجاد یک مدل مناسب و سری پیش بینی جهت ارزیابی مدل مورد استفاده قرار گرفت.

محاسبه توصیف کننده‌ها

در ابتدا برای محاسبه توصیف کننده‌ها، ساختار ترکیبات به کمک نرم افزار Hyperchem رسم شدند. سپس ساختارهای مولکولی رسم شده، به وسیله الگوریتم AM1 بهینه شدند. با استفاده از این نرم افزار می توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچش، بار اتم ها، انرژی تشکیل مولکول و... را بدست آورد. ساختارهای بهینه شده به نرم افزار دراگون منتقل و توصیف کننده‌ها به تعداد ۱۴۹۷ مورد به کمک این نرم افزار محاسبه شدند. این نرم افزار برای محاسبه هیجده دسته از توصیف کننده‌های مولکولی مورد استفاده قرار می گیرد.

کاهش تعداد توصیف کننده‌های نظری

یکی از مشکلاتی که در هنگام ایجاد مدل‌های QSAR با آن مواجه می شویم ، تعداد زیاد متغیرهای مستقل می باشد. در اغلب موارد تعداد توصیف کننده‌ها از تعداد مولکول‌ها بسیار بیشتر است. در این صورت استفاده از روش‌های حداقل مربعات باعث ایجاد مشکلاتی نظیر انتخاب شانسی و همبستگی تصادفی می گردد. با توجه به این که بعضی از متغیرهای مستقل، ثابت بوده و همچنین برخی دیگر با یکدیگر همبستگی نشان می دهند، لذا به روش زیر بعضی از متغیرها حذف شدند.

- ۱- توصیف کننده‌هایی که مقادیر ثابت و یا تقریباً ثابت دارند (بیش از ۹۰٪ داده‌های ثابت دارند)، حذف شدند. در این مرحله تعداد ۳۱۰ توصیف کننده حذف و بدین ترتیب ۱۱۷۱ توصیف کننده باقی ماند.
- ۲- توصیف کننده‌هایی که همبستگی بالای ۰/۹ با یکدیگر دارند مورد بررسی قرار گرفتند و بین آنها، توصیف کننده‌ای که همبستگی کمتری با متغیر مستقل داشت حذف گردید. بدین ترتیب تعداد ۵۹۱ توصیف کننده، کنار گذاشته شد و تعداد ۵۸۰ توصیف کننده باقی ماند.

انتخاب توصیف کننده‌های موثر

مهم ترین بخش در ایجاد یک مدل کار آمد، انتخاب توصیف کننده‌های مناسب است. پس از محاسبه توصیف کننده‌های مختلف، تعدادی از آنها به عنوان توصیف کننده‌های مناسب برای ساخت مدل انتخاب می شوند. این مرحله شامل یافتن توصیف کننده‌های حاوی اطلاعات مفید است به طوری که قدرت پیش بینی مدل در سطح قابل قبولی باشد.^{۱۱،۱۴} در این کار از الگوریتم ژنتیک برای انتخاب مناسب ترین توصیف کننده‌ها استفاده شد. کاربرد الگوریتم ژنتیک در مدل‌سازی QSAR جستجو در میان توصیف کننده‌ها به منظور یافتن مؤثرترین توصیف کننده‌ها جهت انجام مدل‌سازی

کاربرد الگوریتم ژنتیک به همراه رگرسیون خطی چند گانه برای پیش بینی ...

است. الگوریتم ژنتیک، یک روش جستجوی هوشمند و تصادفی است که با بکارگیری عملگرهای ژنتیک از یک فرایند تکامل تدریجی تبعیت می کند.

از الگوریتم ژنتیک در شیمی بطور گسترده ای در زمینه مدل سازی QSAR و کمومتری کس استفاده شده.^{۱۵،۱۷} با این روش از بین ۵۸۰ توصیف کننده باقی مانده تعداد ۶ توصیف کننده به عنوان مناسب ترین آنها انتخاب شد. لیست توصیف کننده های انتخاب شده توسط الگوریتم ژنتیک به همراه توصیف مختصری از آنها در جدول (۱) آورده شده است.

جدول ۱: توصیف کننده های انتخاب شده توسط الگوریتم ژنتیک و توصیف آنها

Descriptor	Chemical meaning	MF ^a
Constant	Intercept	-
X2sol	Solvation connectivity index chi-2	-۲/۹۲۹
BEHv8	Highest eigenvalue n. 8 of Burden matrix / weighted by atomic van der Waals volumes	۸/۲۹۴
JGI9	Mean topological charge index of order9	۰/۹۹۷
GATS4p	Geary autocorrelation - lag 4 / weighted by atomic polarizabilities	۴/۴۳۸
HATS8u	Leverage-weighted autocorrelation of lag 8 / unweighted	۰/۹۳۹
R4m+	R maximal autocorrelation of lag 4 / weighted by atomic masses	۰/۱۳۱

ارزیابی توصیف کننده های انتخاب شده

به منظور ارزیابی توصیف کننده های انتخاب شده مبنی بر مستقل بودن از همدیگر در جدول (۲) ضرایب همبستگی توصیف کننده های انتخاب شده نسبت به یکدیگر آورده شده است. همانطور که از جدول مشاهده می شود بیشترین ضریب همبستگی بین توصیف کننده X2sol و توصیف کننده BEHv8 با مقدار ضریب همبستگی ۰/۶۹۲ می باشد این نتایج نشان می دهد که بین توصیف کننده های انتخاب شده همبستگی وجود نداشته و توصیف کننده ها مستقل از هم هستند و نتایج بدست آمده از مدل سازی دال بر وابستگی توصیف کننده ها نمی باشد.

جدول ۲: ماتریس ضرایب همبستگی توصیف کننده های انتخاب شده

	X2sol	BEHv8	JGI9	GATS4p	HATS8u	R4m+
X2sol	۱					
BEHv8	۰/۶۹۲	۱				
JGI9	۰/۱۹۴	۰/۱۴۴	۱			
GATS4p	۰/۰۵۲	۰/۳۱۹	۰/۰۹۹	۱		
HATS8u	۰/۰۹۳	۰/۳۴۵	۰/۱۵۷	۰/۰۷۳	۱	
R4m+	-۰/۱۷۲	۰/۲۹۸	-۰/۰۲۶	۰/۲۰۶	-۰/۵۰۳	۱

ایجاد مدل با استفاده از MLR

پس از انتخاب مناسب ترین توصیف کننده ها توسط الگوریتم ژنتیک، مرحله بعدی ایجاد مدل، میان توصیف کننده های انتخاب شده و فعالیتهای ترکیبات دارویی می باشد. از نرم افزار SPSS برای این منظور استفاده گردید. بین توصیف کننده ها و فعالیتهای بازدارندگی مشتقات پیرازولها و سری آموزش با استفاده از روش MLR رابطه زیر بدست آمد:

$$pIC_{50} = 10.452(\pm 3.759) + 0.458 (\pm 0.042) \mathbf{X2sol} - 6.257(\pm 1.658) \mathbf{BEHv8} + 241.370$$

$$(\pm 24.793) \mathbf{JGI9} + 9.659 (\pm 0.969) \mathbf{GATS4p} - 5.189 (\pm 0.586) \mathbf{HATS8u} - 6.182 (\pm 2.227)$$

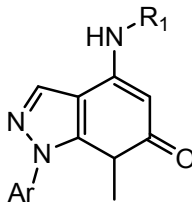
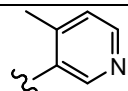
$$\mathbf{R4m}^+$$

سپس از معادله بدست آمده برای پیش بینی فعالیت سری تست استفاده گردید.

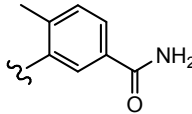
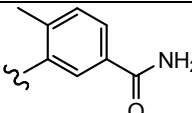
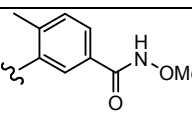
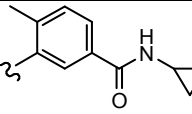
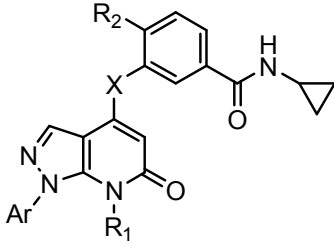
مقادیر واقعی و پیش بینی شده فعالیتها برای کلیه ترکیبات مجموعه آموزش و تست در جدول ۳ آورده شده است.

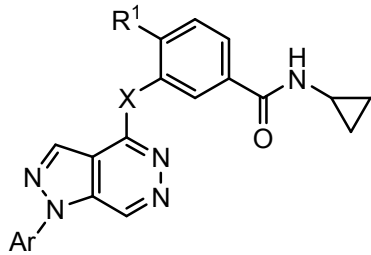
جدول ۳: مقادیر تجربی و محاسبه شده مقادیر pIC_{50} برای ترکیبات مختلف برای مجموعه های آموزشی و پیش بینی در مدل

GA-MLR

No.	Ar	R ₁	R ₂	X	Exp.	GA-MLR
						
۱	2,5-Di-F-Ph		-	-	۶/۳۵	۶/۳۱

کاربرد الگوریتم ژنتیک به همراه رگرسیون خطی چند گانه برای پیش بینی ...

No.	Ar	R ₁	R ₂	X	Exp.	GA-MLR
۲	2,4-Di-F-Ph		-	-	۸/۵۹	۸/۴۹
۳	2,5-Di-F-Ph		-	-	۸/۴۳	۸/۶۲
۴	2,4-Di-F-Ph		-	-	۸/۹۶	۸/۸۲
۵	2,4-Di-F-Ph		-	-	۸/۴۹	۸/۵۰
						
۶	3-F-Ph	Me	Me	NH	۸/۴۸	۸/۵۶
۷	3-F-Ph	Me	Me	O	۸/۰۷	۸/۰۶
۸ ^a	4-F-Ph	Me	Cl	NH	۸/۸۲	۸/۷۹
۹	4-F-Ph	Et	Me	NH	۸/۴	۸/۴۳
۱۰	4-F-Ph	Et	Cl	NH	۸/۹۲	۸/۸۶
۱۱	4-F-Ph	Et	Me	O	۸/۴	۸/۳۱
۱۲	2,4-Di-F-Ph	Me	Me	NH	۸/۴۹	۸/۴۸
۱۳	2,4-Di-F-Ph	Me	Cl	NH	۹	۸/۸۶
۱۴	2,4-Di-F-Ph	Et	Cl	NH	۸/۹۶	۸/۹۶
۱۵	2,4-Di-F-Ph	Me	F	NH	۷/۸۹	۸/۰۹
۱۶	2,4-Di-F-Ph	Me	Me	O	۸/۱۱	۸/۰۹
۱۷	2,5-Di-F-Ph	Me	Me	NH	۸/۵۴	۸/۵۱
۱۸	2,5-Di-F-Ph	Me	Cl	NH	۸/۹۶	۸/۹۱

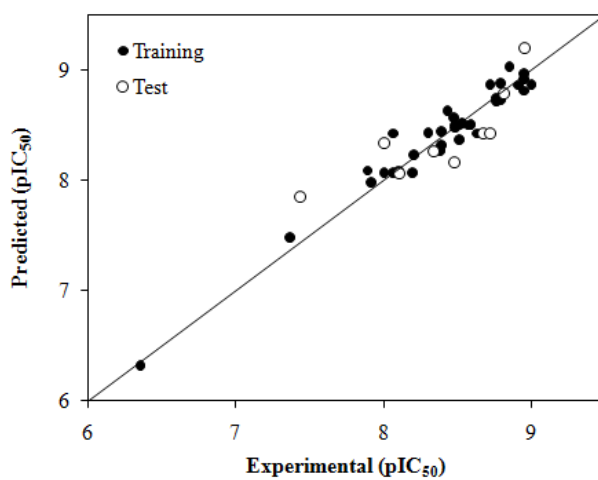
No.	Ar	R ₁	R ₂	X	Exp.	GA-MLR
۱۹	2,5-Di-F-Ph	Me	Me	O	۸/۲۱	۸/۲۲
۲۰ ^a	2,5-Di-F-Ph	Me	Cl	O	۸/۴۸	۸/۱۶
۲۱	2,6-Di-F-Ph	Me	Me	NH	۸/۷۲	۸/۸۶
۲۲ ^a	2,6-Di-F-Ph	Me	Cl	NH	۸/۹۶	۹/۲۰
۲۳ ^a	2,6-Di-F-Ph	Me	F	NH	۸/۷۲	۸/۴۲
						
۲۴	2-Cl-Ph	Cl	-	NH	۸/۸	۸/۸۷
۲۵ ^a	3-F-Ph	Me	-	NH	۸/۳۴	۸/۲۶
۲۶	3-F-Ph	Cl	-	NH	۸/۳	۸/۴۲
۲۷	4-F-Ph	Cl	-	NH	۸/۰۷	۸/۴۳
۲۸	2,4-Di-F-Ph	Me	-	NH	۷/۹۲	۷/۹۷
۲۹	2,4-Di-F-Ph	Me	-	NH	۸/۵۲	۸/۵۰
۳۰	2,4-Di-F-Ph	F	-	NH	۷/۳۷	۷/۴۷
۳۱	2,4-Di-F-Ph	Me	-	O	۸	۸/۰۶
۳۲ ^a	2,4-Di-F-Ph	Cl	-	O	۷/۴۴	۷/۸۵
۳۳ ^a	2,5-Di-F-Ph	Cl	-	NH	۸/۶۸	۸/۴۲
۳۴	2,6-Di-F-Ph	Me	-	NH	۸/۷۷	۸/۷۵
۳۵	2,6-Di-F-Ph	Cl	-	NH	۸/۸۵	۹/۰۲
۳۶	2,6-Di-F-Ph	Me	-	O	۸/۵۲	۸/۳۶

کاربرد الگوریتم ژنتیک به همراه رگرسیون خطی چند گانه برای پیش بینی ...

No.	Ar	R ₁	R ₂	X	Exp.	GA-MLR
۳۷	4-F-Ph	Me	H	NH	۸/۳۹	۸/۲۶
۳۸	4-F-Ph	Cl	H	NH	۸/۶۴	۸/۴۳
۳۹ ^a	4-F-Ph	Me	H	O	۸/۱۱	۸/۰۶
۴۰	2,4-Di-F-Ph	Me	H	NH	۸/۵۷	۸/۵۰
۴۱ ^a	2,4-Di-F-Ph	Me	F	NH	۸	۸/۳۳
۴۲	2,4-Di-F-Ph	Cl	H	NH	۸/۸	۸/۷۲
۴۳	2,4-Di-F-Ph	Me	H	O	۸/۲	۸/۰۶
۴۴	4-F-Ph	Me	H	NH	۸/۷۷	۸/۷۱

همچنین مقادیر فعالیت‌های محاسبه شده و تجربی برای ترکیبات براساس مدل MLR در دو مجموعه آموزشی و

تست در نمودار ۱ آورده شده است.



نمودار ۱: مقادیر pIC_{50} محاسبه شده برای مشتقات پیرازولها براساس مدل MLR در دو مجموعه آموزشی و تست برحسب

مقادیر تجربی

همانطور که در شکل دیده می شود نزدیکی نتایج به یکدیگر، اعتبار مدل را نشان می دهد. جهت ارزیابی اهمیت و میزان تاثیر توصیف کننده هایی که در مدل وارد شدند، اثر متوسط هر توصیف کننده به صورت زیر محاسبه شد.

$$ME_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \sum_i^n \beta_j d_{ij}}$$

در این معادله ME_j اثر متوسط توصیف کننده j ، ضریب توصیف کننده در معادله MLR، d_{ij} مقدار توصیف کننده j برای مولکول i ، m تعداد توصیف کننده های وارد شده در مدل و n تعداد مولکولهاست. مقادیر ME_j برای هر توصیف کننده در جدول ۱ نشان داده شده است.

نتیجه گیری کلی:

در کار ارائه شده مطالعه ارتباط کمی ساختار-فعالیت روی مشتقات پیرازول که به عنوان داروهای ضد روماتیسم استفاده می شوند انجام شد. در ابتدا توصیف کننده های ساختاری ایجاد و مناسب ترین آنها که بیشترین ارتباط را با فعالیت دارویی مشتقات پیرازول داشتند به کمک الگوریتم ژنتیک انتخاب گردید و به روش رگرسیون خطی چندگانه مدل مناسب ایجاد شد. از این مدل می توان برای پیش بینی فعالیت ترکیبات مشابه استفاده کرد. همچنین این مدل نشان می دهد که حجم و اندروالس، بار کل مولکول، قطبیت و جرم اتمی فاکتورهایی هستند که روی فعالیت این ترکیبات تاثیر می گذارند بنابراین با بهینه کردن این پارامترها می توان خصلت دارویی این ترکیبات را افزایش داد.

قدردانی

نویسندگان از حوزه معاونت محترم پژوهشی دانشگاه آزاد اسلامی واحد شاهرود جهت حمایت از این کار پژوهشی تشکر و قدردانی می کند.

کاربرد الگوریتم ژنتیک به همراه رگرسیون خطی چند گانه برای پیش بینی ...

مراجع:

- 1- G. Melagraki, A. Afantitis, K. Makridima, H. Sarimveis and O. Igglessi-Markopoulou, *J. Mol. Model.*, 2006, **12**, 297.
- 2- A. Habibi-Yangjeh, E. Pourbasheer and M. Danandeh-Jenagharad, *B. Korean. Chem. Soc.*, 2008, **29**, 833.
- 3- A. Habibi-Yangjeh, E. Pourbasheer and M. Danandeh-Jenagharad, *Monatsh. Chem.*, 2009, **140**, 15.
- 4- S. Margutti and S. A. Laufer, *Chem. Med. Chem.*, 2007, **2**, 1116.
- 5- S. J. OKeefe, J. S. Mudgett, S. Cupo, J. N. Parsons, N. A. Chartrain, C. Fitzgerald, S. L. Chen, K. Lowitz, C. Rasa, D. Visco, S. Luell, E. Carballo-Jane, Owens and D. Zaller, *M. J. Biol. Chem.*, 2007, **282**, 34663.
- 6- L. H. Pettus and R. P. Wurz, *Curr. Top. Med. Chem.*, 2008, **8**, 1452.
- 7- G. Wagner and S. Laufer, *Med. Res. Rev.*, 2006, **26**, 1.
- 8- J. Westra and P. C. Limburg, *Mini-Rev. Med. Chem.*, 2006, **6**, 867.
- 9- R. P. Wurz, L. H. Pettus, S. Xu, B. Henkle, L. Sherman, M. Plant, K. Miner, H. McBride, L. M. Wong, C. J. M. Saris, M. R. Lee, S. Chmait, C. Mohr, F. Hsieh and A. S. Tasker, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 4724.
- 10- R. P. Wurz, L. H. Pettus, B. Henkle, L. Sherman, M. Plant, K. Miner, H. J. McBride, L. M. Wong, C. J. M. Saris, M. R. Lee, S. Chmait, C. Mohr and F. Hsieh, A. S. Tasker, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 1680.
- 11- J. Hunger and G. Huttner, *J. Comput. Chem.* 1999, **20**, 455.
- 12- S. Ahmad and M. M. Gromiha, *J. Comput. Chem.*, 2003, **24**, 1313.
- 13- C. L. Waller and M. P. Bradley, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 345.
- 14- J. Aires-de-Sousa and M. C. Hemmer, J. Casteiger,; *Anal. Chem.*, 2002, **74**, 80.
- 15- M. Bos and H. T. Weber, *Analytica Chimica Acta.* 1991, **247**, 97.
- 16- Z. Michalewicz, *Genetic Algorithms + Data Structures = Evaluation Programs*, Springer, Berlin, 1996.
- 17- R. M. Marques, P. J. Schoenmakers, C. B. Lucasius and G. Kateman, *Chromatographia.* 1993, **36**, 83.