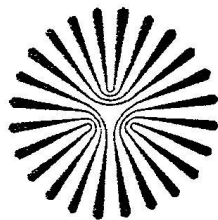


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه پیام نور مرکز اردبیل

پایان نامه برای دریافت درجه کارشناسی ارشد

در رشته شیمی تجزیه

دانشکده علوم پایه

گروه علمی شیمی

عنوان

پیش بینی چرخش ویژه ی برخی از ترکیبات آلی به وسیله ی شبکه های

عصبی مصنوعی

استاد راهنما

دکتر ایوب پارچه باف جدید

استاد مشاور

دکتر علیرضا بنایی

نگارش

سلماز کریمی

اردیبهشت ماه ۱۳۸۹

تقدیم به:

پدر بزرگوارم

که قاموس شرف است و ناموس فضیلت

تقدیم به:

مادر فداکارم

که چشمانش حکایت دریای عاطفه است و دستانش آیتی ز گل مهر

تقدیم به:

همسر عزیزم

او که ستاره درخشان آسمان قلبم است و امید زندگیم

تشکر و سپاس گزاری

اکنون که به فضل و یاری ایزد منان، توفیق پیمودن مسیری کوتاه از راه بی منتهای علم و دانش نصیب گردید لازم می دانم مراتب سپاس و قدردانی خود را خدمت عزیزانی ابراز نمایم که از همکاری و همراهی سودمندشان بهره مند بوده ام.

صمیمانه ترین سپاس خود را به محضر استاد گرامی آقای دکتر ایوب پارچه باف جدید به پاس رهنمود های ارزنده شان تقدیم می دارم.

همچنین از جناب آقای دکتر علیرضا بنایی به جهت کوشش های بی شائبه ایشان قدر دانی می نمایم.

از محضر استاد محترم داور خانم دکتر ماندانا امیری نهایت تشکر و سپاس گزاری را ابراز می دارم.

سلامت و بهروزی همه این بزرگواران را از خداوند متعال خواستارم.

از کارکنان محترم دانشگاه پیام نور مرکز اردبیل و همچنین از دوستان و همکلاسی های مهربانم به خاطر همدلی صمیمانه شان بی نهایت سپاسگزارم.

نام خانوادگی دانشجو : کریمی	نام : سلماز
عنوان پایان نامه : پیش بینی چرخش ویژه ی برخی از ترکیبات آلی به وسیله ی شبکه های عصبی مصنوعی	
استاد راهنما : آقای دکتر ایوب پارچه باف جدید	
استاد مشاور : آقای دکتر علیرضا بنایی	
مقطع تحصیلی : کارشناسی ارشد	رشته : شیمی
گرایش : تجزیه	دانشگاه : پیام نور اردبیل
دانشکده : علوم پایه	گروه علمی : شیمی
تاریخ فارغ التحصیلی :	تعداد صفحه :
کلید واژگان : ارتباط کمی ساختار- چرخش ویژه، الگوریتم ژنتیکی، الکل، شبکه عصبی مصنوعی، چرخش ویژه نوری	
چکیده :	
<p>QSPR برای مدل سازی و پیش بینی چرخش ویژه نوری بکار رفته است. در این پروژه، ابتدا ۵۹ مولکول الکل که دارای چرخش ویژه بودند انتخاب شدند و شکل مولکولی آنها با استفاده از نرم افزار Hyperchem 7.0 ترسیم شد. سپس فایل شکل های رسم شده به نرم افزار Draogn انتقال یافت که تعداد ۱۴۸۱ توصیف گر مولکولی را محاسبه می کند. با توجه به اینکه یک رابطه غیر خطی میان توصیف گرها وجود داشت از الگوریتم ژنتیک برای انتخاب بهترین توصیف گر استفاده شد. ابتدا مدل سازی با روش های الگوریتم ژنتیک- رگرسیون خطی چندگانه (GA-MLR) و الگوریتم ژنتیک- شبکه عصبی مصنوعی (GA-ANN) مدلی خطی با هشت توصیف گر ایجاد کرد که R^2 به ترتیب ۰,۷۷۳ و ۰,۹۲۶ حاصل شد. برای اعتبار سازی مدل از روش Leave-One-Out در سری آموزشی استفاده شد و چهار پارامتر بهینه شدند. در ادامه برای بدست آوردن ارتباط غیر خطی مولفه های اصلی از روش GA-ANN جهت محاسبه چرخش ویژه نوری الکل ها استفاده شد. در این روش ورودی شبکه های عصبی مصنوعی مولفه های اصلی انتخاب شده در الگوریتم ژنتیک بودند. بدین ترتیب یک شبکه از نوع ۱-۴-۸ بکار گرفته شد. در نهایت با توجه به مقادیر تجزیه و تحلیل واریانس و جداول و منحنی ها می توان نتیجه گرفت که مدل ریاضی بدست آمده توانایی پیش بینی R را داراست.</p>	

فهرست مطالب

صفحه

عنوان

فصل اول - مقدمه

- ۱-۱- مقدمه ۲
- ۲-۱- انتخاب سری مولکول ۳
- ۳-۱- محاسبه توصیف کننده ها ۴
- ۱-۳-۱- طبقه بندی های مهم توصیف کننده های مولکولی تئوری ۷
- ۲-۳-۱- ویژگی توصیف کننده مولکولی ۸
- ۳-۳-۱- توصیف گرهای جزئی ۹
- ۴-۳-۱- توصیف گرهای توپولوژی ۹
- ۵-۳-۱- شمارش گر قدم های مولکولی ۱۰
- ۶-۳-۱- توصیف گرهای تطبیقی ۱۰
- ۷-۳-۱- ضرایب بار توپولوژی گالوز ۱۰
- ۸-۳-۱- هم بستگی های دو بعدی ۱۱
- ۹-۳-۱- توصیف گرهای بار ۱۱
- ۱۰-۳-۱- ضرایب آروماتیسیت ۱۲
- ۱۱-۳-۱- برش مولکولی راندیک ۱۲

- ۱۲-۳-۱- توصیف گرهای هندسی..... ۱۲
- ۱۳-۳-۱- توزیع کننده های توزیع شعاعی..... ۱۳
- ۱۴-۳-۱- توصیف گرهای سه بعدی..... ۱۳
- ۱۵-۳-۱- توصیف گرهای تصویری..... ۱۴
- ۱۶-۳-۱- توصیف گرهای جانمایی توپولوژی- هندسی..... ۱۴
- ۱۷-۳-۱- گروه های عاملی..... ۱۴
- ۱۸-۳-۱- بخش های متصل به اتم مرکزی..... ۱۵
- ۱۹-۳-۱- توصیف گرهای تجربی..... ۱۵
- ۲۰-۳-۱- خواص..... ۱۵
- ۴-۱- انتخاب توصیف کننده های مناسب..... ۱۵
- ۵-۱- مراحل ایجاد رابطه ساختار با فعالیت / خاصیت..... ۱۶
- ۶-۱- رگرسیون خطی چندگانه..... ۱۸
- ۷-۱- الگوریتم ژنتیکی..... ۱۸
- ۱-۷-۱- محدوده کاربرد الگوریتم های ژنتیکی..... ۱۹
- ۲-۷-۱- ترکیب مجدد..... ۲۱
- ۳-۷-۱- ترکیب مجدد n نقطه ای..... ۲۲
- ۴-۷-۱- ترکیب مجدد همسان..... ۲۳
- ۵-۷-۱- جهش..... ۲۴

- ۶-۷-۱- جایگزین و ادامه..... ۲۵
- ۷-۷-۱- اندازه کارآیی یک نسل..... ۲۶
- ۸-۷-۱- کاربردهای الگوریتم ژنتیک..... ۲۶
- ۹-۷-۱- نتیجه گیری..... ۲۷
- ۸-۱- شبکه های عصبی مصنوعی (ANN)..... ۲۷
- ۱-۸-۱- تشابهات..... ۳۲
- ۲-۸-۱- انتظارات..... ۳۲
- ۳-۸-۱- ویژگی های شبکه عصبی مصنوعی..... ۳۴
- ۴-۸-۱- ویژگی های مهم شبکه عصبی مصنوعی..... ۳۴
- ۵-۸-۱- مزایا و معایب شبکه عصبی مصنوعی..... ۳۵
- ۶-۸-۱- توابع انتقال..... ۳۶
- ۹-۱- تکنیک Cross-Validation..... ۴۰
- ۱۰-۱- ساختار الکل ها..... ۴۰
- ۱-۱۰-۱- خواص فیزیکی الکل ها..... ۴۰
- ۱۱-۱- کایرالیت..... ۴۱
- ۱۲-۱- چرخش ویژه..... ۴۱

فصل دوم- پیشینه پژوهش

۱-۲- مطالعات تئوری انجام شده..... ۴۴

فصل سوم- کارهای تجربی

۱-۳- مجموعه داده مورد استفاده..... ۴۸

۳-۱-۱- دسته بندی داده ها..... ۵۲

۳-۲- استفاده از نرم افزار Hyper Chem 7.0..... ۵۲

۳-۳- استفاده از نرم افزار Dragon..... ۵۴

۳-۴- استفاده از نرم افزار Excel ۲۰۰۷..... ۵۶

۳-۵- استفاده از نرم افزار Neural Power..... ۵۷

فصل چهارم- نتایج و بحث

۴-۱- انتخاب توصیف کننده های مناسب با استفاده از الگوریتم ژنتیک..... ۶۶

۴-۲- ایجاد مدل با استفاده از ANN و MLR..... ۶۷

۴-۳- توضیح توصیف گرها..... ۶۹

۴-۴- نتیجه گیری..... ۷۹

۴-۵- پیشنهادات..... ۷۹

۴-۶- منابع..... ۸۰

فصل اول

مقدمه

۱-۱) مقدمه

یکی از مهمترین زمینه های کاربرد روش های کمومتریکس مطالعه ارتباط بین خاصیت یا فعالیت مولکول ها با ویژگی های ساختاری آنهاست. این نوع مطالعات که با عنوان ارتباط کمی فعالیت و ساختار^۱ (QSAR) یا ارتباط کمی خاصیت و ساختار^۲ (QSPR) معروف شده اند، به بررسی نحوه ارتباط بین خاصیت یا فعالیت های مختلف ترکیبات شیمیایی با مشخصات ساختاری و ذاتی آنها و به نحوه ارتباط بین فعالیت بیولوژیکی و ساختار شیمیایی ترکیبات می پردازد. QSAR و QSPR زیر مجموعه ای از همبستگی های ساختار/ پاسخ^۳ (SRC) می باشند که هدف آن پیدا کردن رابطه ای میان ساختار مولکول و پاسخ مولکولی اندازه گیری شده است.

بررسی ارتباط بین ساختار شیمیایی و فعالیت ترکیبات، پیش بینی فعالیت ترکیبات جدید را بر اساس اطلاعات مربوط به ساختار شیمیایی آنها امکان پذیر می سازد. از آنجایی که مقادیر پارامترهای ترمودینامیکی، فیزیکی و شیمیایی بسیاری از ترکیبات در دسترس نیست و اندازه گیری آنها گران قیمت یا وقت گیر است، از طرفی مقادیر این پارامترها برای طراحی، کنترل و نیز فهم فرآیند های شیمیایی، مطالعه رفتار زیست محیطی آنها و نیز طراحی دارو ها نیاز است، استفاده از روش هایی برای تخمین این پارامترها ضروری بنظر می رسد. ایده برقرار کردن ارتباط کمی بین خاصیت یا فعالیت و ساختار، اولین بار توسط براون^۴ در سال ۱۹۶۵ مطرح شده است [۱]. وی از این طریق ساختار شیمیایی و فعالیت های فیزیولوژیکی را به هم مرتبط کرد. سپس اوردن^۵ و میر^۶ رابطه بین خاصیت آرام بخشی ترکیبات آلی و ضریب تقسیم آنها را مطالعه نمودند و در نهایت دریافتند که این خاصیت ترکیبات به میزان آب گریزی آنها بستگی دارد [۲].

^۱ Quantitative Structure Activity Relationship

^۲ Quantitative Structure Property Relationship

^۳ Structure Response Correlation

^۴ Brown

^۵ Urden

^۶ Mier

استفاده از این روش روز به روز گسترش یافته است و امروز در پیش بینی خواص متفاوتی به کار برده می شود. از کاربردهای QSPR/QSAR می توان به محاسبه زمان بازداري ترکیبات در ستون کروماتوگرافی، حلالیت، گرانی، ثابت های اسیدی و بازی ترکیبات اشاره کرد [۱۷-۳].

استفاده از چنین مدلی برای پیش بینی خاصیت یا فعالیت ترکیبات مختلف مانع از اتلاف وقت و هزینه خواهد شد [۱۸]. ایجاد ارتباط بین خاصیت یا فعالیت یک ترکیب و ساختار آن با استفاده از چند توصیف کننده^۱ مناسب امکان پذیر است. توصیف کننده ها کمیت هایی هستند که به ساختار مولکول ارتباط دارند و برای هر مولکولی مقادیر خاصی را به خود اختصاص می دهند. انتخاب نوع توصیف گرها در بدست آوردن مدل مناسب بسیار تعیین کننده می باشد.

یافتن رابطه ریاضی بین یک خاصیت یا فعالیت ترکیب و توصیف کننده های مولکولی با استفاده از تکنیک های آماری مختلف امکان پذیر است. توسعه روابط QSPR/QSAR با کمک توصیف گرهای نظری نه تنها برای پیش بینی تجربی خواص یا فعالیت های فیزیکی، شیمیایی و بیولوژیکی ترکیبات ابزاری قوی می باشد بلکه برای فهم عمیق تر جزئیات مکانیسم های موجود در سیستم های پیچیده نیز لازم هستند. به طور کلی مطالعات QSPR/QSAR شامل چهار بخش اصلی است: انتخاب سری مولکولی، محاسبه توصیف کننده ها، انتخاب توصیف کننده های مناسب و ایجاد مدل جهت پیش بینی خاصیت یا فعالیت ترکیبات جدید. در ادامه این چهار بخش به ترتیب بررسی خواهند شد.

۱-۲) انتخاب سری مولکولی

اولین مرحله در مطالعات QSPR/QSAR جمع آوری و انتخاب تعدادی مولکول است که خاصیت یا فعالیت مورد نظر آنها بطور تجربی در دسترس باشد. سپس سری داده ها به دو قسمت سری آموزشی^۲ و سری ارزیابی^۳ تقسیم می شوند. عملیات مدل سازی بر روی سری آموزشی که اکثر

^۱ Descriptor

^۲ Training set

^۳ Test set

مولکول ها را در بر می گیرد، انجام می شود. سری ارزیابی به نحوی انتخاب می شوند که نماینده جمعیت مولکول های سری آموزشی باشند. لازم به ذکر است که مولکول های سری ارزیابی در هیچ یک از مراحل مدل سازی شرکت ندارند. در واقع سری ارزیابی برای ارزیابی اعتبار مدل های حاصله است. هر قدر که سری داده ها بزرگتر و متنوع تر باشد مدل حاصل از آن معتبرتر بوده و قدرت پیش بینی بالاتری خواهد داشت. همچنین خاصیت یا فعالیت مورد نظر ترکیبات انتخاب شده باید در شرایط یکسان اندازه گیری شده باشد. بدیهی است که هر چه اطلاعات تجربی قابل دسترس برای طراحی مدل بیشتر باشد، بدون تردید مدل کارایی بهتری خواهد داشت. QSPR نسبت ساختار یک ترکیب را با خصوصیت مورد نظر می رساند. ساختار ترکیبات، توصیف کننده های مولکولی را بیان می کند که خصوصیات مولکول داده شده است. توصیف کننده های مولکولی، مانند ضرایب هدایتی همانند سائیز، شاخه دار بودن، غیر اشباع بودن، حاوی هترو اتم بودن، حلقوی بودن مفید می باشند.

۱-۳) محاسبه توصیف کننده ها

توصیف کننده ها پارامترهای عددی هستند که ویژگی های متفاوتی از مولکول را بیان می کنند. به عبارت دیگر، هر یک از این توصیف کننده ها، اطلاعات خاصی از مولکول را در اختیار می گذارند. تعیین و انتخاب توصیف کننده ها از اهمیت به سزایی برخوردار است. در صورتی که توصیف کننده های انتخاب شده مناسب نباشند، عمل مدل سازی با مشکل مواجه شده و مدل مناسب به دست نخواهد آمد. بنابراین در انتخاب توصیف کننده ها باید دقت کرد و توصیف کننده هایی می بایست انتخاب شوند که در ارتباط با ویژگی خاص مولکول باشند. تئوری های کوانتومی روشی را برای محاسبه دقیق توصیف گرهای مولکولی فراهم می کنند. در حال حاضر روش های نظری که برای مطالعه خواص مولکول ها بکار می روند، عبارتند از:

۱. محاسبات آغازین^۱

۲. محاسبات نیمه تجربی^۲

۳. محاسبات مکانیک مولکولی^۳

روش های آغازین براساس پیش گویی های اولیه کوانتومی استوار بوده و با استفاده از آن می توان خصوصیات حالت های گذار و حالت های برانگیخته ترکیبات را محاسبه نمود. تنها عیب این روش زمان بسیار زیاد محاسبات آن است. در حال حاضر محدودیت اساسی از نقطه نظر بزرگی یک مولکول در توانایی برای یک محاسبه دقیق آغازین به محدودیت از لحاظ ظرفیت حافظه و سرعت عملیات محاسباتی کامپیوترها مربوط می شود.

با به بازار آمدن کامپیوتر های با توانایی بالا به همان اندازه امیدواری برای محاسبات خواص مولکول های بزرگتر بیشتر می شود. در هر حال پیشرفت قابل توجه شیمی کوانتومی در سال های اخیر محاسبات مکانیک کوانتومی را به صورت ابزار با ارزشی برای پاسخگویی به تعداد زیادی از مسائل با اهمیت شیمیایی در آورده است. روش های مکانیک مولکولی کاملاً^۱ سریع بوده و برای سیستم های خیلی بزرگ مانند آنزیم ها به آسانی قابل استفاده می باشند و دقت برخی از این روش ها برای هیدرو کربن ها با دقت مهمترین روشهای آغازین برابری می کنند ولی این روشها عموماً^۲ برای حالت پایه مولکول ها فرمول بندی شده اند و نمی توانند برای حالت هایی که شامل شکسته شدن پیوند ها می باشند، استفاده گردند.

روش های نیمه تجربی بین دو روش فوق قرار دارند. در این روش ها مانند روش های مکانیک مولکولی از پارامترهای تجربی استفاده می کنند و نظیر روش های آغازین براساس مکانیک کوانتومی هستند و تنها تفاوت اساسی بین روش های نیمه تجربی و روش های آغازین، استفاده گسترده از تقریب ها در محاسبات نیمه تجربی است. در حالی که روش های آغازین روش های بسیار دقیقی در

^۱ Ab initio

^۲ Semi-empirical

^۳ Molecular Mechanics

این مورد از محاسبات هستند ولی در این روش ها زمان محاسبات کامپیوتری متناسب با توان های بالایی از تعداد الکترونهاى مولکول ها هستند. در نتیجه این محاسبات بیشتر محدود به مولکول های کوچکتر می باشند. با توجه به تقریب های بکار رفته در روش های نیمه تجربی، زمان محاسبه کامپیوتری به شدت پایین می آید. توصیف گرهای تئوری استخراج شده از روش های شیمی کوانتومی اساساً^۱ با مقادیر بدست آمده از تجربه متفاوت هستند، اگرچه به صورت طبیعی همپوشانی بین آنها وجود دارد. برعکس داده های تجربی، در این مقادیر خطای آماری وجود ندارد. اما در هر حال یک سری خطای ذاتی به علت فرضیات بکار رفته جهت ساده کردن محاسبات در آنها وجود دارد. همچنین یکی از ضعف های اساسی توصیف گرهای تئوری بدست آمده از شیمی کوانتوی در نظر نگرفتن اثرات حلال است، چرا که این محاسبات برای مولکول های منزوی می باشند. توصیف کننده های مولکولی یا تئوری نقش اساسی در شیمی، علوم دارو سازی، علوم زیست محیطی و مطالعات روی سلامتی و کنترل کیفیت دارند. این توصیف کننده ها به ما امکان یافتن ارتباط پاسخ ساختار و یافتن مشابهت ها^۱ و زیر ساختار های ترکیبات را می دهند.

توصیف کننده های مولکولی بیان ریاضی مولکول بوده و در واقع نتیجه نهایی منطق و روش ریاضی هستند. این توصیف کننده ها اطلاعات شیمیایی مولکولی را که با یک بیان نمادین نمایش داده شده اند به اعداد قابل استفاده تبدیل می کنند. اولین توصیف های مولکولی که در سال ۱۹۴۷ شناخته شده اند عدد واینر^۲ و عدد پلات^۳ می باشند. طبق تعریف بالا توصیف کننده های مولکولی به دو دسته تقسیم می شوند:

الف) توصیف کننده های حاصل از اندازه گیری های شیمیایی (نظیر ضریب شکست مولی، ممان دو قطبی، قطبش پذیری، لگاریتم ضرب تقسیم آب - اکتانول و خصوصیات شیمی فیزیکی و...)

ب) توصیف کننده های مولکولی تئوری از بیان نمادین مولکول ناشی می شوند و تعدا اطلاعات شیمیایی حاصل از آنها به نوع نمایش بستگی دارد. ساده ترین نمایش مولکولی، فرمول شیمیایی است که فهرستی از انواع اتم ها می باشد.

^۱ Similarity

^۲ Wiener number

^۳ Platt number

اختلاف بنیادی میان این دو دسته این است که توصیف کننده های تئوری دارای خطای آماری با خطای آزمایشی نمی باشد، محاسبه آسان شده است و خطا شامل یک ثابت تقریبی می باشد. مزیت توصیف کننده های مولکولی تئوری در هزینه، وقت و توانایی کاربر می باشد.

۱-۳-۱) طبقه بندی های مهم توصیف کننده های مولکولی تئوری عبارتند از:

۱. توصیف کننده های بدست آمده از فرمول شیمیایی، که توصیف کننده های 0D نامیده می شوند. مانند تعداد اتم، وزن مولکولی، شمارش نوع اتم (توصیف کننده های جزئی و شمارشی).

۲. توصیف کننده های 1D. مانند شکست های ساختاری و اثر انگشتی.

۳. چگونگی اتصال اتم ها به مولکول، نمایش دو بعدی می باشد که بصورت 2D نشان داده می شود. مانند ثابت های نمودار.

۴. اجازه نمایش ترکیب فضایی مولکول را نمایش سه بعدی می نامند که نمایش هندسی هم نامیده می شود و بصورت 3D نشان داده می شود. مانند سه بعدی، تصویری، جا نمایی توپولوژی-هندسی، شیمی کوانتومی.

توصیف کننده های مولکولی مطابق تغییر ناپذیری شکلی، به چهار طبقه تقسیم می شوند که توسط چارتن پیشنهاد شد:

الف) وابستگی شکلی نداشته باشند: که این نوع، همه توصیف کننده هایی را که به هندسه سه بعدی مولکول وابسته نیستند را شامل می شود. مانند وزن مولکولی، توصیف کننده های شمارشی.

ب) وابستگی شکلی کمی داشته باشند: دسته ای از توصیف کننده های مولکولی را شامل می شود، مقادیری را که تغییرات کوچکی را که فقط در حضور بارهای شکلی مربوطه است را نشان می دهد. مانند پیکربندی های cis و trans که معمولاً توصیف کننده های بار می باشند.

پ) وابستگی شکلی متوسطی داشته باشند: توصیف کننده های مولکولی را که شامل اختلافات کوچکی در تغییرات شکلی دارند را شامل می شود. مانند توصیف کننده هایی که مبنای آنها توزیع جرمی می باشد، توصیف کننده های EVA

ت) وابستگی شکلی بالایی داشته باشند: این طبقه از توصیف کننده ها با مقادیر خیلی حساس به تغییر شکلی در مولکول مربوط می شود. مانند مقادیر انرژی برخورد که از میدان های برخورد مولکولی بدست می آیند، توصیف کننده های Shape, G-WHIM, WHIM, 3D-MORSE. در مجموع، بعضی از خصوصیات تغییر ناپذیر مانند تغییر ناپذیری در تعداد اتم که برای توصیف کننده های مولکولی بکار رفته در مدل های QSAR/QSPR الزامی می باشد.

۱-۳-۲) تعدادی از ویژگی های لازم برای یک توصیف کننده مولکولی عبارتند از:

۱) ساده بودن

۲) مستقل بودن

۳) تفسیر ساختار مولکول (غنی بودن از نظر اطلاعات)

۴) عدم وابستگی با سایر توصیف کننده ها

۵) تغییر منظم با تغییر تدریجی در ساختارها

۶) وابستگی صحیح به اندازه مولکول

۷) تمایز بین ایزومرهای مختلف مولکول.

محاسبات توصیف کننده ها معمولاً با استفاده از محاسبات تئوری یا آزمایشی های عملی انجام می گیرد. امروزه نرم افزارهای مختلفی نیز طراحی شده اند که براساس محاسبات دقیق تئوری می توانند توصیف کننده های بیشتری را با دقت خوب و سرعت زیاد محاسبه کنند [۱۹]. در این پروژه از نرم افزار دراگون^۱ با ویرایش ۲,۱ استفاده شده است، که توصیف کننده های محاسبه شده توسط این

^۱ Dragon

نرم افزار به ۱۸ دسته کلی تقسیم می شوند که در زیر به ترتیب معرفی می شوند.

۱-۳-۳) توصیف کننده های جزئی^۱

این توصیف کننده ها اطلاعات اساسی در مورد ساختار مولکول ارائه می کنند و به توصیف کننده های بدون بعد یا صفر معروف هستند. در دسته بندی های کلی تر این گروه از توصیف کننده ها، زیر مجموعه توصیف کننده توپولوژی هستند. این سری به اتصالات مولکولی و به طور کلی به ساختار صورت بندی مولکول مربوط نمی شوند و تنها به خواص جمعی مولکول مورد نظر بستگی دارند. از جمله این توصیف کننده ها می توان به جرم مولکولی، تعداد اتم های مختلف، تعداد پیوند های متفاوت و تعداد اجزای ساختاری مثل حلقه های مختلف اشاره کرد. در نرم افزار دراگون تعداد چهل و هفت توصیف کننده جزئی قابل محاسبه است که بسته به نوع ترکیب انتخابی و ارتباط آن با خاصیت مورد نظر می توان تعدادی از این دسته توصیف کننده را انتخاب نمود [۱۹].

۱-۳-۴) توصیف کننده های توپولوژی^۲

این توصیف کننده ها برخلاف دسته قبل کاملاً وابسته به ساختار مولکول می باشند و از گراف های مولکولی بدست می آیند. جزء ساده ترین نوع توصیف کننده ها می باشند و تنها به نوع اتم، نوع پیوندها و نحوه ارتباط اتم ها به یکدیگر وابسته است. این پارامترها را می توان بدون بهینه کردن ساختار مولکول محاسبه کرد. این توصیف کننده ها اصطلاحاً "دو بعدی هستند و به صورت بندی فضایی مولکول بستگی ندارند. برای محاسبه این توصیف کننده ها ساختار مولکول بدون در نظر گرفتن هیدروژن ها مورد بررسی و محاسبه قرار می گیرد. در نرم افزار دراگون با استفاده از اطلاعات تئوری و ساختار ورودی و بهینه سازی شده مولکول تعداد ۲۶۲ توصیف کننده از این را محاسبه می شود. از جمله این توصیف کننده ها می توان ضرایب زاگراب^۳، ضرایب مربوط به اطلاعات پیوندی،

^۱ Constitutional Descriptors

^۲ Topological Descriptors

^۳ Zagreb indices

اطلاعات مربوط به اندازه مولکول، حجم مولکول و اطلاعاتی کلی در رابطه با اجزای اتمی را نام برد [۱۹].

۱-۳-۵) شمارش گر قدم های مولکولی^۱

این توصیف کننده ها نیز زیر مجموعه توصیف کننده های دو بعدی هستند و بنابراین به ساختار مولکول وابسته اند. حرکت های جهت دار مولکولی در فواصل مختلف، به صورت پارامتری تحت عنوان فاصله در طول های مختلفی که اصطلاحاً "قدم نامیده می شود، اندازه گیری می شوند. این شمارش به صورت رفت و برگشت انجام می شود و نرم افزار دراگون ۲۱ توصیف کننده را محاسبه می کند. ده توصیف کننده مربوط به حرکت رفت، ده توصیف کننده مربوط به حرکت برگشت، و توصیف کننده نهایی مربوط به مجموعی کل قدم های شمارش شده در رفت و برگشت است [۱۹].

۱-۳-۶) توصیف کننده های تطبیقی^۲

این دسته از توصیف کننده ها، ویژه مقدار^۳ یک ماتریس اصلاح شده می باشد و نیز از محاسبات تئوری ریاضی به دست می آیند. با تشکیل ماتریس تطبیق براساس فواصل بدست آمده از ساختار مولکولی و با وزن دار کردن قطر این ماتریس با استفاده از پارامترهای مختلف نظیر جرم اتمی یا حجم واندوالسی اتم، ضرایب الکترو نگاتیویته یا پلاریزاسیون، می توان یک سری عملیات ریاضی را وارد محاسبات تئوری کرد، به این ترتیب مقادیر ویژه این ماتریس دسته ای شامل ۶۴ توصیف کننده را ارائه می دهد که بین این چهار پارامتر تقسیم می شود [۱۹].

۱-۳-۷) ضرایب بار توپولوژی گالوز^۴

این توصیف کننده ها اولین بار توسط گالوز و گارسیا تعریف شده است. این توصیف کننده نشان

^۱ Molecular walk counts

^۲ BCUT Descriptors

^۳ Eigenvalues

^۴ Galvez topological charge

دهنده توزیع بار مولکولی است و به نحوی انتقال بار بین جفت اتمها مربوط می شود. محاسبه قدر مطلق مقدار ویژه ماتریس تطبیق اصلاح شده^۱ و وزن دار شده توسط مقادیر بار مولکول ها، این دسته از توصیف کننده ها را ارائه می دهد. این دسته شامل ده ضریب اصلی، ده مقدار به عنوان متوسط مقادیر بدست آمده و یک مقدار عددی معروف به ضریب بار عملی است که مجموع تمام مقادیر اصلی است [۱۹].

۱-۳-۸ هم بستگی های دو بعدی^۲

این دسته شامل ۹۶ توصیف کننده می باشد که از حاصل جمع مربوط به جرم اتم های انتهایی در شرایط محاسباتی خاص بدست می آید. این ضرایب در چهار دسته به نام های ارائه دهندگانشان: مورکاوو^۳، بروتو^۴، موران^۵ و گیری^۶ طبقه بندی می شود، که هر دسته می تواند توسط مقادیر مختلف عددی مربوط به وزن مولکولی یا حجم وزن دار شده یا بدون وزن محاسبه شود [۱۹].

۱-۳-۹ توصیف کننده های بار^۷

این توصیف کننده ها بر مبنای ارائه بار تخمین زده شده در روش های مولکولی کوانتومی محاسبه می شوند و جنبه های الکترونی کل مولکول و نواحی خاص را شرح می دهد. همانند اتم ها، پیوند ها، شکست های مولکولی. این توصیف کننده بوسیله شیمی محاسبه ای محاسبه می شود. نرم افزار دراگون ۱۴ توصیف کننده را در این سری گزارش می کند که از جمله آنها می توان، به مجموع بار مثبت، مجموع بار منفی، مجموع بار کامل، پلاریزاسیون و دی پل ممان های لحظه ای اشاره کرد [۱۹].

^۱ Scaled adjacency matrix

^۲ 2D autocorrelation descriptors

^۳ Morcao

^۴ Broto

^۵ Moran

^۶ Geary

^۷ Charge descriptors