



دانشگاه صنعتی شیراز

دانشکده علوم گروه شیمی

پایان نامه کارشناسی ارشد شیمی فیزیک

رشته شیمی گرایش شیمی فیزیک

پیش بینی خواص بیولوژیکی و ترمودینامیکی برخی داروها و
ترکیبات آلی با استفاده از روش QSAR و QSPR

بوسیله

مریم لطفی

استاد راهنما

دکتر محمد مهدی پاپری

دکتر محمد نوید سلطانی راد

استاد مشاور

دکتر عبدالرضا نکوئی

بهمن ماه ۱۳۸۹

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

بسمه تعالی

پیش بینی خواص بیولوژیکی و ترمودینامیکی برخی داروها و ترکیبات آلی با استفاده از روش
QSPR و QSAR

پایان نامه ارائه شده به عنوان بخشی از فعالیت‌های تحصیلی

توسط
مریم لطفی

برای اخذ درجه کارشناسی ارشد

گروه شیمی دانشکده علوم پایه
دانشگاه صنعتی شیراز

ارزیابی پایان نامه توسط هیات داوران با درجه:

دکتر محمد مهدی پاپری دانشیار رشته شیمی-شیمی فیزیک (استاد راهنما)
دکتر محمد نوید سلطانی راد استادیار رشته شیمی-شیمی آلی (استاد راهنما)
دکتر عبدالرضا نکوئی استادیار رشته شیمی-شیمی فیزیک (استاد مشاور)
دکتر جلیل مقدسی، استاد دانشگاه شیراز (داور)
دکتر دلارا محمد آقایی استادیار دانشگاه صنعتی شیراز (داور)

مدیر امور آموزشی و تحصیلات تکمیلی دانشگاه:

حق چاپ محفوظ و مخصوص به دانشگاه صنعتی شیراز است

تقدیم به

پدر و مادر دلسوز و مهربانم که

مرا بندگی حق،

آزاد زیستن،

و الفبای زندگی آموختند.

و تقدیم به همه آنان که مرا علم آموختند.

سپاسگزاری

سپاس و ستایش معلمی را که به من اندیشیدن را آموخت، نه اندیشه ها را.

از جناب آقای دکتر پاپری و جناب آقای دکتر سلطانی راد که در تک تک لحظات شکل گیری و پیشرفت این پروژه تحقیقاتی، راهنمایی صبور و همراهی دلسوز برای این جانب بودند، کمال تشکر را دارم.

چکیده

پیش بینی خواص بیولوژیکی و ترمودینامیکی برخی داروها و ترکیبات آلی با استفاده از روش

QSPR و QSAR

به وسیله:

مریم لطفی

در این رساله در صدد هستیم که با استفاده از روش QSAR ارتباط منطقی بین ساختار و فعالیت ۳۸ عدد از ترکیبات ضد قارچ را بر روی فعالیت بیولوژیکی حداقل غلظت بازدارنده (MIC₅₀) آن ها بدست آوریم. در این تحقیق توصیف کننده های مختلف برای هر مولکول محاسبه شده است. در نهایت معادله ای شامل ۶ پارامتر بین MIC₅₀ و توصیف کننده های ضریب توزیع (logP), RDF100V, Total valence connectivity, G_{1v}, Mor04v, Mwc10, مولکول ها برقرار شد. همچنین با استفاده از روش QSPR ارتباط منطقی بین ساختار با دمای بحرانی (T_c) و فشار بحرانی (P_c) برای دسته ای از داده های DECHEMA به دست آورده شد. که در نهایت معادله ای شامل شش پارامتر بین دمای بحرانی و توصیف کننده های Mor3u R5, MATS, BELp2 و Mor2v مولکولها و معادلهای شامل پنج پارامتر بین فشار بحرانی و توصیف کننده های ضریب توزیع (logp), Sum of Dgree, Mv, H, Mor21u برقرار شد. به علاوه با استفاده از روش رگرسیون فازی اقدام به مدل سازی برای به دست آوردن ارتباط بین دمای بحرانی و خواص ساختاری کردیم.

فهرست مطالب

صفحه	عنوان
	فصل اول: روش QSAR/QSPR
1	۱-۱ مقدمه.....
6	۲-۱ بررسی روش QSAR/QSPR.....
6	۱-۲-۱ مدل QSAR/QSPR.....
7	۳-۱ مراحل کار در QSAR.....
8	۱-۳-۱ انتخاب سری داده‌ها.....
8	۲-۳-۱ محاسبه توصیف کننده‌ها.....
9	۴-۳-۱ تجزیه ، تحلیل و ارزیابی آماری توصیف کننده‌ها.....
۱۰	۴-۳-۱ آنالیز مدل‌های آماری و انتخاب مدل مناسب.....
10	۵-۳-۱ ارزیابی مدل به دست آمده.....
	فصل دوم: آغاز و پیدایش QSAR/QSPR
12	۱-۲ پیدایش QSAR.....
12	۲-۲ روابط خطی انرژی آزاد.....
۱۵	۳-۲ مدل هانش.....
16	۴-۲ طراحی دارو.....
16	۵-۲ طراحی دارو با کمک کامپیوتر.....

فصل سوم: توصیف کننده ها

- ۱-۳ انتخاب و محاسبه توصیف کننده ها..... ۱۷
- ۲-۳ انواع توصیف کننده ها..... ۱۷
- ۱-۲-۳ توصیف کننده های توپولوژیکی..... ۱۸
- ۱-۲-۳-۱ توصیف کننده های جزء..... ۱۸
- ۲-۳-۱-۲ اندیس های ارتباط مولکولی..... ۱۹
- ۳-۲-۱-۳ توصیف کننده های زیرساختاری..... ۱۹
- ۴-۲-۱-۳ توصیف کننده های محیطی..... ۱۹
- ۲-۲-۲ توصیف کننده های هندسی..... ۲۰
- ۱-۲-۲-۲ روش های ۲ بعدی..... ۲۰
- ۲-۲-۲-۲ روش های سه بعدی..... ۲۱
- ۳-۲-۲-۳ توصیف کننده های مساحت سایه مولکول..... ۲۱
- ۴-۲-۲-۳ توصیف کننده های اندیس کاپا..... ۲۱
- ۵-۲-۲-۳ توصیف کننده های فاکتور شکل..... ۲۲
- ۶-۲-۲-۳ توصیف کننده های مساحت و حجم واندروالس مولکول ها..... ۲۲
- ۷-۲-۲-۳ توصیف کننده های محورهای اصلی چرخش مولکول ها..... ۲۲
- ۸-۲-۲-۳ توصیف کننده های ممان اینرسی اصلی مولکول ها..... ۲۳
- ۳-۲-۳ توصیف کننده های الکترونی..... ۲۳

عنوان	صفحه
۱-۳-۲-۳ توصیف کننده‌های قطبی.....	۲۳
۲-۳-۲-۳ توصیف کننده‌های دارای انرژی.....	۲۶
۱-۲-۳-۲ شاخص‌های واکنش پذیری.....	۲۶
۴-۲-۳ توصیف کننده‌های فیزیکوشیمیایی.....	۲۷
۱-۴-۲-۳ لگاریتم ضریب تقسیم ($\log P$).....	۲۷
۲-۴-۲-۳ قطبش پذیری ملکولی.....	۲۷
۵-۲-۳ توصیف کننده‌های هیبریدی.....	۲۸
۶-۲-۳ شاخص‌های تشابه کوانتومی.....	۲۸
فصل چهارم: روش‌های آماری و تکنیک‌های آنالیز چند متغیره	
۱-۴ مقدمه.....	۲۹
۲-۴ بررسی‌ها برای انتخاب روش.....	۳۰
۱-۲-۴ روش‌های رگرسیون.....	۳۰
۱-۱-۲-۴ رگرسیون خطی ساده.....	۳۰
۲-۱-۲-۴ رگرسیون چندخطی.....	۳۱
۳-۱-۲-۴ رگرسیون چند خطی مرحله‌ای.....	۳۱
۲-۲-۴ روش هیوریستیک.....	۳۲
۳-۲-۴ الگوریتم ژنتیک بر اساس رگرسیون چند خطی (GA-MLR).....	۳۳
۴-۲-۴ تحلیل مؤلفه‌های اصلی.....	۳۳
۵-۲-۴ رگرسیون مؤلفه‌های اصلی.....	۳۴

عنوان	صفحه
۶-۲-۴ حداقل مربعات جزئی.....	۳۴
۷-۲-۴ کمترین مربعات جزئی ژنتیکی.....	۳۴
۸-۲-۴ آنالیز فاکتور حداقل مربعات جزئی (FA-PLS).....	۳۵
۹-۲-۴ شبکه های عصبی.....	۳۵
۱۰-۲-۴ آماده سازی داده.....	۳۵
۱۱-۲-۴ مقیاس سازی و متمرکز سازی.....	۳۵
۱۲-۲-۴ تصحیح و متراکم سازی داده.....	۳۶
۱۳-۲-۴ ساخت مدل پیشگویی کننده:تحلیل رگرسیون خطی.....	۳۶
۱۴-۲-۴ تجزیه و تحلیل رگرسیون چندخطی.....	۳۷
۱۵-۲-۴ سایر تبدیلات.....	۳۹
<i>فصل پنجم:نه کارگیری رایانه در طراحی دارو ها</i>	
۱-۵ مقدمه.....	۴۱
۲-۵ مکانیک کوانتومی و مولکولی.....	۴۱
۱-۲-۵ مکانیک مولکولی.....	۴۱
۲-۲-۵ مکانیک کوانتومی.....	۴۲
۳-۲-۵ انتخاب روش.....	۴۳
۱-۳-۲-۵ رسم ساختارهای شیمیایی.....	۴۳
۲-۳-۲-۵ ساختارهای 3D.....	۴۴
۳-۳-۲-۵ کمینه کردن انرژی.....	۴۴
۴-۳-۲-۵ مشاهده ساختارهای سه بعدی.....	۴۴
۵-۳-۲-۵ ابعاد مولکولی.....	۴۵

۴۵.....	۶-۳-۲-۵- خواص مولکولی.....
صفحه	عنوان
۴۵.....	۳-۵ بسته نرم‌افزاری ChemOffice.....
۴۵.....	۴-۵ بسته‌های نرم‌افزاری HyperChem.....
۴۶.....	۵-۵ بسته نرم‌افزاری Dragon.....
۴۶.....	۶-۵ نرم‌افزار Gaussia 03.....
۴۷.....	۷-۵ نرم افزار MATLAB.....
۴۷.....	۸-۵ بسته نرم‌افزاری SPSS/PC.....
۴۷.....	۹-۵ محاسبه و استخراج توصیف گرهای مولکولی.....
۴۸.....	۱-۹-۵ توصیف گرهای به دست آمده از HyperChem.....
۴۷.....	۲-۹-۵ توصیف گرهای به دست آمده از Gaussian 03.....
۴۸.....	۳-۹-۵ توصیف گرهای بدست آمده از Dragon.....
<i>فصل ششم: شیمی کوانتومی و کاربرد آن در QSPR/QSAR</i>	
۵۰.....	۱-۶ مقدمه.....
۵۰.....	۲-۶ روش های شیمی کوانتومی.....
۵۰.....	۱-۲-۶ روش های آغازین.....
۵۲.....	۲-۱-۲-۶ روش هارتری - فاک.....
۵۲.....	۲-۲-۶ روش های نیمه تجربی.....
۵۳.....	۱-۲-۲-۶ تئوری تابعی چگالی DFT.....
۵۴.....	۲-۲-۲-۶ ETH.....
۵۴.....	۳-۲-۲-۶ CNDO.....
۵۴.....	۴-۲-۲-۶ MINDO و INDO.....

PM3.AM1, MNDO	۵۵
یک برنامه اربیتال مولکولی نیمه تجربی	۵۶
روش‌های نیمه تجربی در میک	۵۶
بهینه سازی هندسه مولکولی	۵۸
خلاصه‌ای از قابلیت‌های میک	۵۹
کمیت‌های قابل محاسبه	۵۹
واحد‌ها	۶۰
توصیف گره‌های شیمی کوانتومی	۶۱
بارهای اتمی	۶۲
انرژی‌های اوربیتال مولکولی	۶۳
دانسیتیه‌های اوربیتال جبهه‌ای	۶۵
قطبش پذیری های اتم-اتم	۶۴
قطبش پذیری مولکولی	۶۶
ممان دو قطبی و شاخص‌های قطبیت	۶۶
انرژی	۶۷
توصیف کننده‌ها	۶۸
نتیجه گیری‌ها	۶۸

فصل هفتم: بحث و نتیجه گیری

مطالعات QSAR بر روی ضد قارچ‌های آزولی	۷۰
مطالعات ارتباط کمی فعالیت - ساختار (QSAR) بعضی از مشتقات ایمیدازول به عنوان عوامل ضد قارچ	۷۰
انتخاب مجموعه داده‌ها	۷۲
رسم ساختار شیمیایی مولکول‌ها و بهینه‌سازی هندسه مولکولی	۷۹

صفحه

عنوان

۷۹.....	۱-۲-۱-۱-۷	ارسم ساختار شیمیایی مولکول‌ها و بهینه‌سازی هندسه مولکولی توسط روش نیمه‌تجربی.....
۸۱.....	۲-۲-۱-۱-۷	ارسم ساختار شیمیایی مولکول‌ها و بهینه‌سازی هندسه مولکولی توسط روش آغازین.....
۸۱.....	۳-۱-۱-۷	محاسبه مقادیر توصیف‌کننده‌ها با استفاده از نرم افزارهای مختلف و روش‌های مختلف.....
۸۱.....	۱-۳-۱-۱-۷	محاسبه مقدار توصیف‌گرها توسط روش‌های نیمه‌تجربی.....
۸۲.....	۴-۱-۱-۷	مقایسه مقادیرهای توصیف‌کننده‌های به دست آمده از نرم افزارهای مختلف.....
۸۲.....	۵-۱-۱-۷	حذف توصیف‌گرهای دارای همبستگی بالا.....
۸۶.....	۶-۱-۱-۷	تقسیم داده‌ها به دو دسته Training و Prediction.....
۸۶.....	۷-۱-۱-۷	مدل‌سازی توسط روش‌های آماری.....
۹۰.....	۸-۱-۱-۷	توصیف‌کننده‌های به دست آمده از مدل‌سازی.....
۹۰.....	۱-۸-۱-۱-۷	الگاریتم ضریب توزیع (logP).....
۹۰.....	۸-۱-۱-۷	RDF100V.....
۹۱.....	۳-۸-۱-۱-۷	Total valence connectivity.....
۹۱.....	۸-۱-۱-۷	G _{1v} ۴.....
۹۱.....	۸-۱-۱-۷	Mor04V۵.....
۹۱.....	۸-۱-۱-۷	Mwc10.....
۹۲.....	۲-۷	نتیجه حاصل از انجام مطالعات QSPR.....
۹۲.....	۱-۲-۷	مطالعه کمی خاصیت-ساختار (QSPR) برخی از ترکیبات آلی.....
۹۳.....	۱-۱-۲-۷	انتخاب مجموعه داده‌ها.....
۹۴.....	۲-۱-۲-۷	مدل‌سازی مولکولی و جمع‌آوری توصیف‌کننده‌ها.....
۹۴.....	۱-۲-۱-۲-۷	ارسم ساختار شیمیایی مولکول‌ها و بهینه‌سازی هندسه مولکولی.....
۹۷.....	۲-۲-۱-۲-۷	محاسبه مقدار توصیف‌گرها با استفاده از نرم افزارهای مختلف.....
صفحه		عنوان

۱۰۲.....۳-۲-۱-۲-۷ مقایسه مقدار توصیف‌گرهای به دست آمده از نرم افزارهای مختلف.....

۱۰۲.....۱-۳-۲-۱-۲-۷ مقایسه مقدار توصیف‌کننده‌های به دست آمده از نرم افزارهای هایپرکیم و کیم‌دی‌آلترا.....

۱۱۵.....۳-۱-۲-۷ مدل‌سازی توسط روش‌های آماری.....

۱۱۵.....۱-۳-۱-۲-۷ مدل‌های QSPR به دست آمده برای تخمین فشار بحرانی.....

۱۱۶.....۱-۱-۳-۱-۲-۷ توصیف‌کننده‌های به دست آمده از مدل‌سازی برای فشار بحرانی.....

۱۱۷.....۲-۳-۱-۲-۷ مدل‌های به دست آمده برای تخمین دمای بحرانی.....

۱۱۷.....۱-۲-۳-۱-۲-۷ توصیف‌کننده‌های به دست آمده از مدل‌سازی.....

فصل هشتم: به کارگیری رگرسیون خطی فازی در QSPR

۱۱۹.....۱-۸ مقدمه.....

۱۲۰.....۲-۸ مدل QSPR.....

فهرست جدول ها

عنوان	شماره صفحه
جدول شماره ۱: نام و نشانه برخی از توصیف کننده‌های جزء.....	۱۹
جدول شماره ۲: نام و نشانه برخی از توصیف کننده‌های الکترونی.....	۲۴
جدول شماره ۳: توصیف کننده های موجود در نرم افزار دراگن.....	۴۹
جدول شماره ۴: تغییر استخلاف ها در ۳۸ ترکیب مشابه آزول ها.....	۷۳
جدول شماره ۵: مقایسه مقادیر Lumo به دست آمده از دو نرم افزار مختلف.....	۸۳
جدول شماره ۶: مقایسه مقادیر HOMO به دست آمده از دو نرم افزار مختلف.....	۸۴
جدول شماره ۷: مقایسه مقادیر جرم به دست آمده از دو نرم افزار مختلف.....	۸۵
جدول شماره ۸: مقایسه مقادیر به دست آمده از مدل سازی با مقادیر تجربی برای دسته Training.....	۸۹
جدول شماره ۹: نام ترکیبات گرفته شده از داده های DEHEMA جهت انجام مطالعات QSPR.....	۹۸
جدول شماره ۱۰: مقایسه مقدار جرم به دست آمده از دو نرم افزار مختلف.....	۱۰۳
جدول شماره ۱۱: مقدار HOMO به دست آمده از دو نرم افزار مختلف.....	۱۰۷
جدول شماره ۱۲: مقایسه مقدار LUMO به دست آمده از دو نرم افزار مختلف.....	۱۱۱
جدول شماره ۱۳: مقادیر عددی توصیفگرها و دمای بحرانی.....	۱۲۲
جدول شماره ۱۴: مقادیر R^2 برای مدل حاصل از هر h.....	۱۲۶

فهرست شکل ها

عنوان	صفحه
شکل شماره ۱. ساختار عمومی ضد قارچ‌های به کار رفته در این مطالعه.....	۸۰
شکل شماره ۲. ساختار فضا پرکن مولکول ۴۰ با بیشترین فعالیت.....	۸۰
شکل شماره ۳. ساختار فضا پرکن مولکول ۲۷ با کمترین فعالیت.....	۸۰
شکل شماره ۴. نمودار مقدار مشاهده شده (واقعی) بر حسب مقدار مورد انتظار برای فعالیت بیولوژیکی برای داده های Training.....	۸۸
شکل شماره ۵. نمودار مقدار مشاهده شده (واقعی) بر حسب مقدار مورد انتظار برای فعالیت بیولوژیکی برای داده های prediction.....	۸۸
شکل شماره ۶. ساختار فضا پرکن مولکول اتان با کمترین مقدار دمای بحرانی.....	۹۵
شکل شماره ۷. ساختار فضا پرکن مولکول آنتراسن با بیشترین مقدار دمای بحرانی.....	۹۵
شکل شماره ۸. ساختار فضا پرکن مولکول فنول با بیشترین مقدار فشار بحرانی.....	۹۶
شکل شماره ۹. ساختار فضا پرکن مولکول ۱ و ۲ و ۳-تری نیترو بنزن با کمترین مقدار فشار بحرانی.....	۹۶

فصل اول: روش QSAR

مقدمه

برای درک مکانیسم فرآیندهای مختلف شیمیایی، کشف و توسعه مواد جدید، حفظ محیط زیست و زمینه‌های دیگر شیمی، هنوز توانایی حل مسائل به طور کامل وجود ندارد و برای عملی کردن بعضی از مسائل نیاز به سیستم‌های بسیار پیچیده‌ای می‌باشد که انجام آن‌ها در گرو صرف هزینه‌های بسیار و مطالعات گسترده است.

تدابیر مختلفی می‌تواند در طراحی داروها استفاده شود بعضی از این تدابیر مستلزم تغییر شکل دارو هستند به طوری که دارو بهتر در موقعیت گیرنده قرار بگیرد. بقیه تدابیر شامل یک تغییر در خواص فیزیکی داروست؛ به نحوی که توزیع، متابولیسم و یا برهم‌کنش‌های اتصال گیرنده تحت تأثیر قرار بگیرند. این تدابیر اغلب مستلزم سنتز مولکول‌های مشابه حاوی تعدادی استخلاف بر روی حلقه‌های آروماتیک یا غیر آروماتیک و یا گروه‌های عاملی در دسترس است. چنان که بتوانیم مولکول‌های مشابهی را با همه استخلاف‌ها و ترکیب استخلاف‌های ممکن سنتز کنیم، آن‌گاه تعداد نامحدودی از مولکول‌های مشابه ممکن را می‌توان ساخت. بنابراین اگر روشی اتخاذ شود که به واسطه آن بتوان تصمیم گرفت که کدام استخلاف استفاده شود، بسیار مفید خواهد بود [۱].

بنابراین نیاز مبرمی به مدل‌های پیشگویی کننده خواص ترکیبات مفید و طراحی داروهای جدید با خواص بهبود یافته و کاهش عوارض جانبی و همچنین تشخیص سلامت مواد شیمیایی هستیم. علاوه بر این تشخیص خطر مواد شیمیایی که به محیط زیست وارد می‌شوند و رشد روش‌های سنتزی جایگزین و بی‌خطر، به شدت احساس می‌شود.

گذشته از این به روش‌هایی احتیاج داریم که جایگزین استفاده از حیوانات آزمایشگاهی شود و یا دست کم استفاده از آن‌ها را به حداقل برساند. مدت‌های مدیدی متخصصین شیمی دارویی با روش‌های معین، ترکیبات رهبر^۱ را با این انگیزه که سنتز، آزمایش و درک فرآیندهای آن‌ها امکان پذیر باشد اصلاح می‌کردند. با استفاده از روش‌های سنتی ممکن بود ماه‌ها سنتز یک ترکیب جدید برای آزمایش‌های بیولوژیک طول بکشد. بنابراین در طول قرون گذشته تلاش‌های بسیار و مستمری برای

¹ Lead compounds

طراحی ترکیبات جدیدی که از لحاظ بیولوژیکی فعال هستند با هدف کاهش هزینه تحقیقات آزمایشگاهی صورت گرفته است [۲].

از جمله روش‌های محاسباتی کمومتریکس^۱ است. تجزیه و تحلیل آماری و ریاضی داده‌های شیمیایی معمولاً تحت عنوان کمومتریکس یاد می‌شود. به عبارت دیگر کمومتریکس یک روش کارآمد برای خلاصه کردن اطلاعات مفید از یک سری داده‌های مشخص و پیش‌بینی سری دیگر داده‌هاست.

در حقیقت هدف کمومتریکس بهبود بخشیدن فرآیندهای اندازه‌گیری و استخراج اطلاعات مفیدتر از داده‌های اندازه‌گیری شده فیزیکی و شیمیایی می‌باشد. یکی از زمینه‌های مهم کاربرد کمومتریکس در مطالعاتی است که خواص مولکول‌ها را به ویژگی‌های ساختاری آنها نسبت می‌دهد. از نظر شیمی‌دانان فعالیت و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. این نوع از مطالعات به بررسی کمی ارتباط ساختمان با فعالیت (QSAR) و همچنین بررسی کمی ارتباط ساختمان با ویژگی (QSPR) معروف می‌باشند. هدف از انجام این مطالعات، پیدا کردن رابطه‌ای است که بین رفتار فیزیکوشیمیایی یک مولکول و پارامترهای ساختاری آن وجود دارد. نتایج این مطالعات علاوه بر شفاف‌سازی نحوه ارتباط بین خواص مولکول‌ها و ویژگی‌های ساختمانی آنها به پژوهشگران در پیش‌بینی رفتار مولکول‌های جدید بر اساس رفتار مولکول‌های مشابه کمک می‌کند. در واقع از جمله روش‌های مدل‌سازی است که شامل ارتباط میان داده‌های آزمایشگاهی با همان خواص فیزیکی یا غیر آن، در قالب سیستم مدل‌سازی است؛ که این روش‌ها دارای تاریخچه-ای طولانی در مدل‌سازی سیستم‌های بیولوژیک می‌باشند و اکنون نیز نقش مؤثری در تحلیل نتایج و استفاده از نتایج مدل‌سازی مولکولی در صنایع شیمی دارند.

به عنوان مثال نمونه‌ای را با یک سری خواص فیزیکی در نظر بگیرید که می‌خواهید آن را بهینه کنید (سودمندی دارو، ارتباط آنزیمی، مقاومت کششی یک پلیمر)؛ از جمله تازه‌ترین روش‌هایی که به این منظور مورد استفاده قرار می‌گیرد روش هانش^۲ می‌باشد (۱۹۷۱). که مبتنی است بر ارتباط فعالیت‌های بیولوژیک مولکول در قالب پارامتر «آبگریزی» آن که به عنوان ثابت اکتانول به آب معرفی می‌شود.

این روش در طول انرژی آزاد خطی در شیمی فیزیک زیستی قرار می‌گیرد. (روابط هامت^۳ (۱۹۵۳)) که انرژی آزاد (لگاریتم ثابت تعادل)، به نوع دیگری از انرژی یا خواص ماده که بر اساس انرژی آزاد سنجیده می‌شود، وابسته است. همان طور که پیشنهاد شد، ساده‌ترین چنین روش‌هایی

¹ Chemometrics

² Hansch

³ Hammett

به عنوان مدل‌های خطی مطرحند و رگرسیون خطی برای آن‌ها پیشنهاد می‌شود. به همین ترتیب که مدل پیچیده‌تر می‌شود، مدل به سمت رگرسیون غیرخطی میل می‌کند (کوال اسکی^۱) و به منظور توالی چنین مدل‌هایی تا حصول روابطی میان خواص فیزیکی ماده و میزان فعالیت‌های آن می‌بایستی مدل‌سازی به روش «شبکه‌های عصبی» به کار گرفته شود. علاوه بر آب‌گریزی، سایر پارامترهای مؤثر و مرتبط به عنوان متغیرهای مستقل در چنین مدل‌هایی، عبارتند از: نسبت آرایش اتم‌ها به نیروهای واندروالس و خواص الکترواستاتیک یک روش که در برگیرنده اطلاعات سه بعدی در قالب یک مدل باشد، مانند «آنالیز میدان‌های مولکولی تطبیقی» که به اختصار COMFA خوانده می‌شود. که توسط پترسون^۲، کرامر^۳، و بانس^۴ در سال ۱۹۸۸ پایه گذاری شد. این پارامترها و متغیرهای وابسته به آن، عمومی می‌باشند و در بهینه‌سازی مدل مؤثرند. علاوه بر آن ارتباطات میان مولکولی را در جهت توجیه ساختار ماکرومولکول‌ها در ابعاد سه بعدی توجیه می‌کند. (به این معنی که نیروهای الکترواستاتیک و پتانسیل‌های واندروالس را لحاظ می‌کند).

تکنیک‌های فوق به نام (QSAR) خوانده می‌شود. و از مهمترین این تکنیک‌ها می‌توان به (QSPR) اشاره کرد که این تکنیک زمانی که بهینه سازی تعدادی از خواص مکانیکی بر اساس فعالیتشان مطلوب باشد، مورد توجه قرار می‌گیرد.

یک بار دیگر می‌توان این عمل را در زمینه مدل‌های خطی و غیرخطی انجام داد و می‌توان خواص متعددی از مجموعه ای از مولکول‌ها را در چنین مدلی لحاظ کرد. بنابراین برای نتیجه‌گیری خواص الکترواستاتیک یا خواص الکترونی ماده، به منظور آزمایش این که آیا آن‌ها واقعاً دارای چنین روابطی هستند، مکانیک کوانتوم محاسباتی استفاده می‌شود. از آنجائیکه هدف نهایی، دستیابی به ساختار دقیق الکترونی ماده نمی‌باشد و دقت کمتر و کارایی بیشتر مطلوب است، (روش‌های نیمه تجربی) روش‌های مکانیک کوانتومی برای چنین اهدافی مطلوب است.

تکنیک‌های دوگانه مدل‌سازی مولکولی و روش‌های کمی آماری می‌توانند در توضیح اطلاعات ساختاری ترکیبات فعال مفید باشند. به دلیل این که اثر بیولوژیکی به ندرت به یک یا دو خاصیت شیمیایی وابسته است، بنابراین اغلب در این روش‌ها یک مسأله چند بعدی برای تعداد زیادی عامل مطرح می‌شود که قادر به توجیه یک فضای وسیعی از پارامترها است.

برای این که بتوانیم با مجموعه‌ای از داده‌های پیچیده با فعالیت‌های بیولوژیکی^۵ متفاوت و تعداد زیاد توصیف‌کننده‌ها^۱ کار کنیم نیاز به ابزارهای پیشرفته آماری و محاسباتی در روش داریم. امروزه

1 Kowal Ski

2 Patterson

3 Cramer

4 Bance

5 Biological activities