

رقم الترتيب:

رقم التسلسل:

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة الشهيد حمه لخضر - الوادي
كلية العلوم الدقيقة
مذكرة تخرج لنيل شهادة

الماستر أكاديمي

قسم: علوم المادة

شعبة: الكيمياء

تخصص: كيمياء عضوية تحليلية

من إعداد الطالبة: طعيلي كوثر

بعنوان:

نمذجة العلاقة الكمية بين البنية الكيميائية (الجزئية) للمركبات
العضوية ومعاملات التوزيع (LogK_{ow}) المرتبطة بها

تحت إشراف الأستاذ: سويحي بلقاسم

أمام اللجنة المكونة من:

الاسم واللقب	الرتبة	الجامعة	الصفة
د. بوشقرا سماح	أ. مساعدة (أ)	جامعة الشهيد حمه لخضر - الوادي	رئيساً
د. مازري راضية	أ. محاضرة (ب)	جامعة الشهيد حمه لخضر - الوادي	ممتحناً
د. كراسع عائشة	أ. محاضرة (أ)	جامعة الشهيد حمه لخضر - الوادي	ممتحناً
أ. سويحي بلقاسم	أ. مساعد (أ)	جامعة الشهيد حمه لخضر - الوادي	مشرفاً

الموسم الجامعي: 2017/2018



رقم الترتيب :.....
رقم التسلسل:.....

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة الشهيد حمّة لخضر - الوادي
كلية العلوم الدقيقة
مذكرة تخرج لنيل شهادة

الماستر أكاديمي

قسم :علوم المادة

شعبة :الكيمياء

تخصص :كيمياء عضوية تحليلية

من إعداد الطالبة: طعبل كوتر

بعنوان:

نمذجة العلاقة الكمية بين البنية الكيميائية (الجزئية) للمركبات
العضوية ومعاملات التوزيع (LogK_{ow}) المرتبطة بها

تحت إشراف الأستاذ:سويحي بلقاسم

أمام اللجنة المكونة من:

الاسم واللقب	الرتبة	الجامعة	الصفة
د.بوشقرا سماح	أ.مساعدة (أ)	جامعة الشهيد حمّة لخضر - الوادي	رئيساً
د.مازري راضية	أ.محاضرة (ب)	جامعة الشهيد حمّة لخضر - الوادي	ممتحناً
د.كراسع عائشة	أ.محاضرة (أ)	جامعة الشهيد حمّة لخضر - الوادي	ممتحناً
أ.سويحي بلقاسم	أ.مساعد (أ)	جامعة الشهيد حمّة لخضر - الوادي	مشرفاً

الموسم الجامعي: 2017/2018

شكر و عرفان

الحمد لله الذي وفقني ومنحني القوة والصبر للإتمام هذا العمل

أقدس بالشكر الجزيل إلى:

الإستاد المتوفى «سويدي بلقاسم» الذي تفضل بالإشراف على هذا البحث

وقد لي الدعم والتوجيه والنصائح التي طالما أفادتني وعلى كرمه وعظيم جهده.

بكل مولاي الحبيب والاحترام أقدس بالشكر الجزيل والعرفان إلى أساتذتي الأفاضل في كلية

العلوم الرقبة.

وأيضا أقدس مكتبة «الرسالة» بالنظر على طباعة هذه المذكرة.

وأشكر والعرفان لجميع أصدقائي وزملائي في هذه المرحلة وكل من مد لي يد العون في

إتمام هذا العمل كما أقدس بالشكر والإستاء إلى (أبي وأمي وأخوتي).

المخلص

توصلنا في هذا العمل المنجز إلى بناء علاقة رياضية تربط بين البنية الكيميائية المعبر عنها بالواصفات الجزيئية (3D)، والخاصية الفيزيوكيميائية المتمثلة في معامل التوزيع LogK_{ow} ، وذلك للنتبؤ بقيم LogK_{ow} لمركبات جديدة. شملت هذه الدراسة 7 كتل من الواصفات الجزيئية (3D)، حيث تم فيها حساب الكتل كل على حد وبعده اتم حساب الكتل مجمعة. وتمت هذه الدراسة على 992 مركب كيميائي بتطبيق تقنية QSPR وبطريقة الانحدار الخطي المتعدد (MLR) وهذا باستخدام البرمجية الإحصائية XStat. (النسخة 2013) وكانت أفضل النتائج للكتل مجمعة وهي: ($R^2 = 0.894$), ($S^2 = 0.535$) و ($F = 73.298$).

الكلمات المفتاحية: النمذجة الجزيئية، LogK_{ow} ، الواصفات الجزيئية، تقنية QSPR، الانحدار الخطي المتعدد (MLR).

قائمة الاختصارات

الترجمة إلى العربية	المعنى	الاختصار
معامل التوزيع (أوكتانول - ماء)	Octanol -Water Partition Coefficient	Kow
لوغاريتم معامل التوزيع (أوكتانول - ماء)	Octanol-Water Partition Coefficient logarithme	LogKow
العلاقات الكمية (البنية الكيميائية - الفعالية)	Quantitative structure - activity relationships	QSAR
العلاقات الكمية (البنية الكيميائية - الخاصية)	Quantitative Structure - Property Relationships	QSPR
كروماتوغرافيا سائلة عالية الأداء	High-performance liquid chromatography	HPLC
الإنحدار الخطي المتعدد	Multiple Linear Regression	MLR
الإنحدار الخطي البسيط	Simple Linear Regression	SLR
الإنحدار المكونات الرئيسية	Principal component regression	PCR
الإنحدار المربعات الصغرى	Partial Least Square Regression	PLSR
ميكانيكا الجزيئية	Molecular mechanics	MM⁺
معيار نموذج رقم 3	Parameterized Model number 3	PM3
نموذج 1 ل Austin	Austin Model 1	AM1
متوسط الانحرافات المطلقة	Average Absolute Deviation	AAD
الانحراف المعياري	Standard Deviation	SD
مواصفات الإدخال النصي المبسط للجزيئات	Simplified molecular-input line-entry system	SMILES
إهمال معدلة التداخل ثنائي الذرة	Modified Neglect of Diatomic Overlap	MNDO
إهمال كامل لفرق التداخل	Complete Neglect of Differential Overlap	CNDO
تحليل العنصر الرئيسي	Principal Component Analysis	PCA
لغة تعريف البيانات	Data Definition Language	DDL
خطأ تربيعي متوسط	Mean Squared Error	MSE
خطأ تربيعي متوسط الجذر	Root-Mean-Squared Error	RMSE
النسبة المئوية لمتوسط الخطأ المطلق	Mean absolute percentage error	MAPE
الإحصائية Durbin - Watson	Durbin-Watson statistic	DW
Mallows ل Cp	Mallows's Cp	Cp
معيار Akaike للمعلومة	Akaike information criterion	AIC
معيار Bayesian Schwarz	Bayesian Schwarz criterion	SBC
معيار Amemiya للتنبؤ	Amemiya's Prediction Criterion	PC

قائمة الرموز

الرمز	المعنى
C_o	تركيز المادة في الطور العضوي (أوكتانول)
C_w	تركيز المادة في الطور المائي
2D	ثنائي البعد
3D	ثلاثي البعد
r	معامل الارتباط الخطي
R^2	معامل التحديد (مربع معامل الارتباط الخطي)
$R^2 \text{ ajusté}$	معامل التحديد المضبوط
F	الاختبار الإحصائي (لقياس معنوية النموذج)
π	إلكترونات الرابطة الثنائية
S^2	الانحراف المعياري
σ	متوسط الانحرافات المطلقة

قائمة الأشكال

الصفحة	العنوان	الرقم
الفصل الأول: عموميات حول معامل التوزيع (أوكتانول - ماء)		
8	علاقة معامل التوزيع Kow مع بعض الخصائص الفيزيوكيميائية الأخرى.	الشكل (1-I)
13	خطوات النمذجة الجزيئية.	الشكل (2-I)
14	علاقة QSPR/QSAR بتقييم المعايير.	الشكل (3-I)
18	مبدأ طريقة QSPR.	الشكل (4-I)
الفصل الثاني: طرق الحساب والبرامج المستعملة		
20	الخطوات المتبعة في التنبؤ.	الشكل (1-II)
25	درجة قوة علاقة معامل الارتباط.	الشكل (2-II)
26	شكل انتشار الارتباط في الانحدار الخطي المتعدد.	الشكل (3-II)
32	البرنامج ChemOffice Ultra.	الشكل (4-II)
33	البرنامج HyperChem.	الشكل (5-II)
34	البرنامج DRAGON.	الشكل (6-II)
36	حساب الواصفات الجزيئية.	الشكل (7-II)
37	البرنامج XLSTAT.	الشكل (8-II)
الفصل الثالث: قاعدة البيانات		
45	الخطوات التجريبية للعمل.	الشكل (1-III)
الفصل الرابع: النتائج والمناقشة		
52	المنحنيات التي تبين الواصفات الجزيئية والقيم العددية لمعاملاتها.	الشكل (1-IV)
58	بيان يوضح تغيرات LogKow بدلالة Préd(LogKow) الخاصة بالكتل مجمعة.	الشكل (2-IV)

قائمة الجداول

الصفحة	العنوان	الرقم
الفصل الأول: عموميات حول معامل التوزيع (أوكتانول - ماء)		
16	برامج النمذجة المستخدمة في علاقة (QSPR/QSAR) لإيجاد قيم LogK_{ow} .	الجدول (1-I)
الفصل الثاني: طرق الحساب والبرامج المستعملة		
35	أهم كتل الواصفات الجزيئية.	الجدول (1-II)
الفصل الثالث: قاعدة البيانات		
42	يوضح الجدول التالي جزءا من قاعدة البيانات.	الجدول (1-III)
44	نتائج حساب الواصفات الجزيئية المكونة لقاعدة البيانات باستعمال Dragon.	الجدول (2-III)
الفصل الرابع: النتائج والمناقشة		
49	نتائج الكتل المدروسة كل على حد ومجمعة.	الجدول (1-IV)
50	معاملات التعديل الخاصة بالكتل مجمعة.	الجدول (2-IV)
51	تحليل الفروق الخاص بالكتل مجمعة.	الجدول (3-IV)
51	ضبط معاملات النموذج الخاص بالكتل مجمعة.	الجدول (4-IV)
59	الواصفة ورمزها للواصفات الجزيئية الأقل ارتباطا (بالكتل مجمعة).	الجدول (5-IV)
59	الواصفة ورمزها للواصفات الجزيئية الأكثر ارتباطا (بالكتل مجمعة).	الجدول (6-IV)
61	أمثلة للمركبات التي وجدت خارج مجال الثقة.	الجدول (7-IV)

الفهرس

أ	الملخص
ب	قائمة الاختصارات
ج	قائمة الرموز
د	قائمة الأشكال
هـ	قائمة الجداول
الصفحة	العنوان
2	مقدمة عامة
الجزء النظري: حوصلة ببليوغرافية	
الفصل الأول: عموميات حول معامل التوزيع (أوكتانول - ماء)	
6	I. عموميات حول معامل التوزيع (أوكتانول - ماء)
6	1.1. لمحة تاريخية حول معامل التوزيع
7	2.1. تعريف معامل التوزيع (أوكتانول - ماء) K_{ow}
7	3.1. لوغاريتم معامل التوزيع (أوكتانول - ماء) $\text{Log}K_{ow}$
8	4.1. علاقة معامل التوزيع K_{ow} مع بعض الخصائص الفيزيوكيميائية الأخرى
9	5.1. الطرق التجريبية لحساب لوغاريتم معامل التوزيع (أوكتانول - ماء)
9	1.5.1. طريقة القارورة الهزازة
10	2.5.1. طريقة الكروماتوغرافيا السائلية العالية الأداء
10	3.5.1. طريقة التقلب البطيء
10	4.5.1. طريقة العمود المولد
11	6.1. النمذجة
12	1.6.1. ما هي النمذجة الجزيئية؟
12	2.6.1. النموذج الرياضي
12	3.6.1. مزايا استخدام النماذج الرياضية
13	7.1. العلاقات الكمية (البنية الكيميائية - الخاصية/الفعالية) (QSPR/QSAR)
14	1.7.1. نبذة تاريخية عن (QSPR/QSAR)
15	2.7.1. مشكلة نمذجة العلاقات الكمية (بنية كيميائية - الخاصية/الفعالية)
16	3.7.1. استخدام علاقات (QSPR/QSAR) لتقدير قيمة $\text{Log}K_{ow}$
17	4.7.1. تعريف QSPR

17	5.7.I. مبدأ QSPR
18	6.7.I. الطرق الرياضية المستخدمة في النموذج QSPR
الفصل الثاني: طرق الحساب والبرامج المستعملة	
20	II. طرق الحساب والبرامج المستعملة
20	1.II. الخطوات المتبعة في التنبؤ
21	2.II. طرق الحساب
21	1.2.II. الانحدار الخطي المتعدد
21	2.2.II. تعريف الانحدار الخطي المتعدد
22	3.2.II. نموذج الانحدار الخطي المتعدد
24	4.2.II. الارتباط الخطي
24	5.2.II. تعريف الارتباط
25	6.2.II. معامل الارتباط الخطي
25	1.6.2.II. نوع العلاقة
25	2.6.2.II. قوة العلاقة
26	3.6.2.II. شكل الانتشار
26	4.6.2.II. قياس الارتباط
27	7.2.II. معامل التحديد (R^2)
28	8.2.II. الاختبار الإحصائية (F)
28	9.2.II. متوسط الانحرافات المطلقة (AAD)
29	10.2.II. الانحراف المعياري (SD)
30	3.II. البرامج المستعملة
30	1.3.II. أهم الطرق التي تحدد البنية الكيميائية للجزيئات
31	2.3.II. برنامج (ChemOffice Ultra)
32	3.3.II. برنامج (HyperChem)
32	1.3.3.II. طريقة MM^+ (Molecular mechanics)
33	2.3.II. طريقة (Austin Model 1)AM1
33	3.3.II. طريقة (Parameterized Model number 3)PM3
34	4.3.II. برنامج (DRAGON)
34	1.4.3.II. تعريف الواصف الجزيئي
35	2.4.3.II. كتل الواصف الجزيئي

36	3.4.3.II حساب الواصفات الجزيئية
37	5.3.II برنامج (XLSTAT)
38	III.عموميات حول المركبات العضوية
38	1.III. تعريف المركبات العضوية
38	2.III. تصنيف المركبات العضوية
39	3.III. ما سبب اهتمامنا بالمركبات العضوية (مركبات الكربون)
39	4.III. ما هي أسباب كثرة المركبات العضوية (مركبات الكربون)
الجزء التطبيقي : الحساب والنتائج	
الفصل الثالث :قاعدة البيانات	
42	III. قاعدة البيانات
42	1.III. مكونات قاعدة البيانات
43	4.III. كيفية معالجة وحساب الواصفات الجزيئية المكونة لقاعدة البيانات
45	5.III. الخطوات التجريبية للعمل
46	1.5.III. شرح الخطوات التجريبية للعمل
الفصل الرابع :النتائج والمناقشة	
49	IV. النتائج
49	1.IV. نتائج الحسابات
50	1.IV. معادلة النموذج
58	2.IV. تحليل نتائج الكتل مجمعة
58	3.IV. تحليل باقي (Residues) الكتل مجمعة
58	4.IV. مناقشة نتائج الكتل مجمعة
63	خاتمة عامة
65	المراجع
69	الملاحق
	الملخص

مقدمة عامة

إن التطور التكنولوجي والصناعي السريع على مدى العقود الماضية يهدف إلى زيادة رفاة البشرية . ومع ذلك، فإنه لا يخلو من عواقب على صحة الإنسان والبيئة . وهذا بسبب إختلاف تأثير المواد الكيميائية على البيئة.

ترتكز الأبحاث العلمية حالياً في مجال الكيمياء حول دراسة الخاصية الفيزيوكيميائية معامل التوزيع (أوكتانول - ماء) الذي هو عبارة عن نسبة تركيز المادة المذابة في طور أوكتانول والطور المائي، و من أهم تطبيقاته استعماله في كراهية الماء للمركبات العضوية و له عدة تطبيقات أخرى في العديد من المجالات العلمية كالكيمياء التحليلية، الصيدلة، الكيمياء الحيوية ، ذوبانية الدهون....الخ.

يعبر لوغاريتم معامل التوزيع بين أوكتانول وماء، عن معيار أساسي لتقدير ألفة مادة ما للدهون، فهو مؤشر أساسي في تقدير المصير البيئي للمواد.

إن التطور المهم والسريع لوسائل الإعلام الآلي خلال السنوات الأخيرة سمح للكيميائيين باقتناء أداة حاسوبية جديدة تمكنهم من فهم المشاكل المتعلقة بالمجال الجزيئي بشكل أفضل، ألا وهي النمذجة الجزيئية.

إن النمذجة الجزيئية هي مجال يشمل جميع طرق الحساب النظري ، حيث تسمح بتحديد التمثيل البياني لهندسة أو هيئات ذرات الجزيء و تقدير الخصائص الفيزيوكيميائية الموافقة.

تعد برامج النمذجة الجزيئية من أهم البرامج المساهمة في تصوير المركب الكيميائي على هيئة ثلاثي الأبعاد وتعطينا خصائص الجزيئات، وفي هذه الدراسة استخدمنا مجموعة من البرمجيات وهي: البرنامج (ChemDraw Ultra) لرسم البنية الجزيئية لهذه المركبات العضوية، والبرامج (Dragon) و (HyperChem) لنمذجة الجزيئات وحساب الواصفات، والبرنامج الرياضي (Xlstat) للحساب الاحصائي.

إن الهدف من هذا العمل هو إنشاء نموذج رياضي (معادلة رياضية) على أساس قاعدة البيانات المتاحة، وهذا لدراسة وتحديد الخاصية الفيزيوكيميائية وهي معامل التوزيع (أوكتانول - ماء) LogK_{ow} ومن خلال هذه الدراسة نستطيع أن نتنبأ بمركبات كيميائية أخرى مجهولة، دون اللجوء إلى تحديدها مخبرياً.

خطة الدراسة:

- ✓ مقدمة عامة حيث تحدثنا فيها عن أهمية النمذجة الجزيئية والتعرف إلى البرامج المستعملة في الدراسة وإلى أهمية الخاصية الفيزيوكيميائية المختارة.
- ✓ الجزء الأول حوصلة ببليوغرافية وينقسم هذا الجزء إلى فصلين:
- الفصل الأول: تطرقنا فيه إلى عموميات حول معامل التوزيع (أوكتانول- ماء LogK_{ow} وتقنية العلاقات الكمية (بنية كيميائية - خاصية/فعالية) (QSPR/QSAR).
- الفصل الثاني : تحدثنا في هذا الفصل على طرق الحساب والبرامج المستعملة في هذه الدراسة.
- ✓ الجزء الثاني الحسابات والنتائج حيث تم في هذا الجزء تحليل ومناقشة النتائج التي توصلنا إليها، وهو ينقسم إلى فصلين كذلك:
- الفصل الثالث: تطرقنا في هذا الفصل إلى مكونات قاعدة البيانات ، وكيفية معالجة وحساب الواصفات الجزيئية المكونة لقاعدة البيانات ثم شرح خطوات العمل التجريبية بالتفصيل.
- الفصل الرابع: في هذا الفصل تم عرض النتائج المتحصل عليها ومناقشتها.
- ✓ وأخيرا خلاصة الدراسة متبوعة بالمقترحات والتوصيات.

الجزء النظري

حوصلة بييليوغرافية

الفصل الأول

عموميات حول معامل التوزيع

(أوكتانول - ماء)

I. عموميات حول معامل التوزيع (أوكتانول – ماء):

مقدمة:

يعتبر معامل التوزيع K_{ow} في نظام أوكتانول - ماء من أهم الخصائص الفيزيوكيميائية الرئيسية، ويعرف على أنه عبارة عن نسبة تركيز مادة مذابة بين سائلين مختلفين لا يمتزجان، و من أهم تطبيقاته استعماله في قياس كراهية الماء للمركبات العضوية و له عدة تطبيقات أخرى في العديد من المجالات العلمية كالكيمياء الزراعية، الكيمياء التحليلية، البيئة، الصيدلة، الكيمياء الحيوية و السموم وكذلك في تقييم الخصائص البيولوجية ذات الصلة بفعالية الأدوية ، ذوبانية الدهون و الامتصاص الخلوي. يعتبر معامل التوزيع مؤشرا جيدا للربط بين بنية المركب و فعاليته البيولوجية وله الدور الريادي في علاقات (QSPR/QSAR) ، ونظرا لصغر مجال قيم K_{ow} يعوض بالصيغة $\log K_{ow}$ [1].

I.1. لمحة تاريخية حول معامل التوزيع:

منذ القدم كان معامل التوزيع محط اهتمام العديد من العلماء، حيث منذ 100 سنة العالمان Crum و Fraser أول من افترضا أن النشاط البيولوجي مرتبط بالخاصية الفيزيائية، أطلقا عليها الذوبانية في الوسط المائي، وفي نفس الفترة سنة 1869 بينت تجارب Richardson أن الأثر التخديري للكحولات الأولية متغير بتغير كتلتها المولية [1].

في سنة 1893 أكد Richet فرضية كل من Crum و Fraser من خلال إظهار أنسمية مجموعة متنوعة من المركبات البسيطة مثل (الايثيرات، الكحولات و الكيتونات) تتناسب عكسيا مع خاصية ذوبانها في الماء وأول من عرف معامل التوزيع العالمان Jungfleisch و Berthelot بدراسة نسبة توزيع مادة مذابة بين طورين سائلين غير ممتزجين و كانا قد بدأ الدراسة حول تغير معامل التوزيع بدلالة درجة الحرارة.

في سنة 1909 نشر Hertz صيغة تربط بين معامل التوزيع و عدد الاستخلاصات المتتالية المطبقة على المادة.

الجزء النظري/ الفصل الأول: عموميات حول معامل التوزيع (أوكتانول – ماء)

وفي عام 1951 وسع Hansch و العديد من العلماء العمل أكثر ، و في سنة 1971 نشر Albert Leo أول مقال كليا لمعاملات التوزيع، من ثم أصبح معامل التوزيع متداول كثير خلال العقود الثلاثة الماضية و أصبح خاصية فيزيوكيميائية أساسية في دراسة علاقات (QSPR/QSAR)^[2] .

2.I. تعريف معامل التوزيع (أوكتانول – ماء) K_{ow} :

هو نسبة تركيز المادة المذابة في طور أوكتانول و الطور المائي، حيث يعبر عن تراكيز من حيث الوزن أومولات المواد الكيميائية لكل وحدة حجم من السائل، و يحسب كما هو موضح في العلاقة التالية^[3] :

$$K_{ow} = \frac{C_{octanol}}{C_{water}} \quad (1.I)$$

$C_{octanol}$: تركيز المادة المذابة في الطور العضوي (أوكتانول).

C_{water} : تركيز المادة المذابة في الطور المائي.

3.I. لوغاريتم معامل التوزيع (أوكتانول – ماء) $\text{Log}K_{ow}$:

مفاهيم عامة :

يعبر لوغاريتم معامل التوزيع بين الأوكتانول والماء عن مقدار ألفة مادة ما للدهون وعلى هذا النحو، يمثل $\text{Log}K_{ow}$ مؤشرا أساسيا في تقدير المصير البيئي للمواد. وتشتق عمليات توزيع عديدة باستخدام لوغاريتم معامل التوزيع، مثل الامتصاص في التربة والترسبات، والتركيز في الكائنات الحية. وتعيين معامل التوزيع بين الأوكتانول والماء هو شرط في مجموعة البيانات الأساسية التي تقدم للمواد الجديدة والمواد القائمة التي تتمتع بأولوية في التبليغ بالاتحاد الأوروبي. ونظرا لأن التعيين العملي لقيمة لوغاريتم معامل التوزيع (أوكتانول – ماء) غير ممكنة على الدوام، على سبيل المثال في حالة المواد الشديدة الذوبان في الماء والمواد الشديدة الألفة للدهون، فإنه يمكن استخدام قيمة لمعامل التوزيع K_{ow} مشتقة من قيمة العلاقات الكمية (البنية الكيميائية – الخاصية)^[3] .

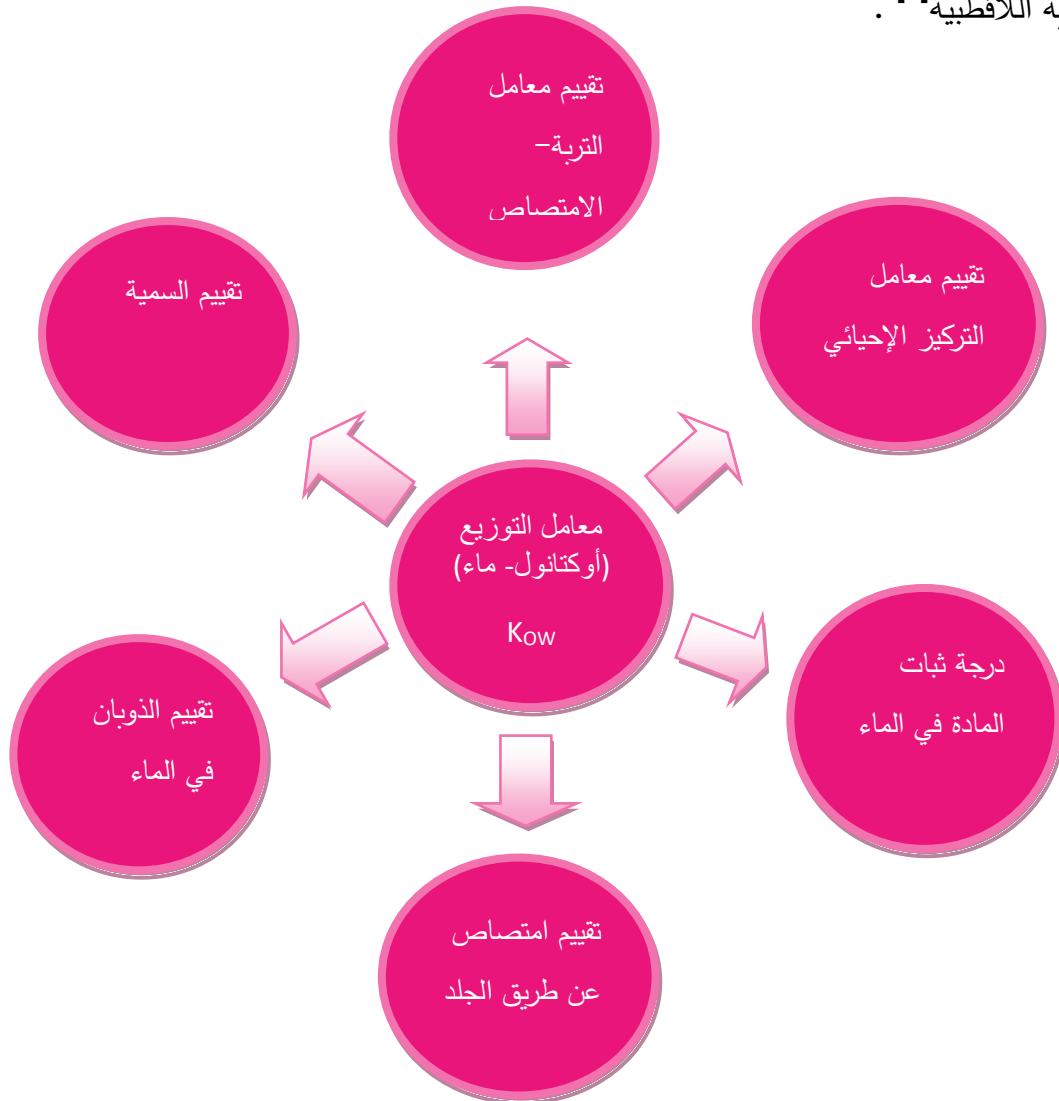
$$K_{ow} = \left(\frac{C_o}{C_w} \right) \quad (2.I)$$

$$\log K_{ow} = \log \left(\frac{C_o}{C_w} \right) \quad (3.I)$$

4.I. علاقة معامل التوزيع K_{ow} مع بعض الخصائص الفيزيوكيميائية الأخرى:

يعتبر معامل التوزيع (أوكتانول - ماء) $\log K_{ow}$ الخاصية المشاركة في حساب الخصائص الأخرى مثل (الذوبانية في الماء، التركيز الحيوي، الامتصاص في التربة والرواسب والنباتات...).

ويعتبر كذلك معيارا جيدا للتنبؤ بمعامل التركيز الحيوي لأنواع معينة من المركبات العضوية، مثل المواد العضوية اللاقطبية^[3].



الشكل (1-I): علاقة معامل التوزيع K_{ow} مع بعض الخصائص الفيزيوكيميائية الأخرى.

يتمثل أساس العلاقة بين التركيز الحيوي وقيمة $\text{Log}K_{ow}$ في تشابه عملية التوزع بين الطور الليبيدي في الأسماك والماء مع عملية التوزع بين معامل التوزيع (أوكتانول – ماء) . وينشأ سبب استخدام K_{ow} من قدرة الأوكتانول على أن يكون بديلا مرضيا لليبيدات الموجود في أنسجة الأسماك [3] .

هناك علاقات واضحة بدرجة عالية بين قيمة $\text{Log}K_{ow}$ وذوبانية المواد في زيت كبد الحوت والتريولين، وهذا الأخير هو أحد أوفر مركبات ثلاثي أسيل غليسول التي توجد في ليبيدات أسماك المياه العذبة [3] .

5.I. الطرق التجريبية لحساب لوغاريتم معامل التوزيع (أوكتانول – ماء) :

يرد في المبادئ التوجيهية القياسية طرائق مختلفة عديدة للتقدير العملي لقيم K_{ow} ، كطريقة القارورة الهزاة والكروماتوغرافيا السائلة العالية الأداء (HPLC)، مثل: OECD 117 (1983) ، OECD ، EPA- FIFRA ، ASTM (1993) ، EPA- OTS (1982) ، EEC A.8 (1992) ، 107(1995) (1982) . ولا تقتصر البيانات الموصى بها على البيانات الناتجة من استخدام طريقة القارورة الهزاة أو طريقة الكروماتوغرافية وفقا للتوجيهات . إذ أنه بالنسبة للمواد الشديدة الألفة للدهون التي تذوب ببطء في الماء تعتبر البيانات الناتجة من استخدام طريقة التقلب البطيء أكثر موثوقية بصفة عامة. ويجري في وقت الحاضر اختيار دائري لطريقة التقلب البطيء تمهيدا لوضع توجيه نهائي لمنظمة التعاون والتنمية [3] .

1.5.I. طريقة القارورة الهزاة :

المبدأ الأساسي للطريقة هو قياس ذوبان المادة في طورين مختلفين الماء والأوكتانول . و لتعيين معامل التوزيع يجب بلوغ حالة الاتزان بين جميع المكونات المتفاعلة في النظام، ومن ثم يعين تركيز المواد الذائبة في الطورين .تطبق طريقة القارورة الهزاة عندما تقع قيم لوغاريتم معامل التوزيع (أوكتانول – ماء) في النطاق من 2 إلى 4 .ولا تنطبق هذه طريقة إلا على المواد النقية الذوابة في الماء والأوكتانول، ويجب إجراؤها تحت درجة حرارة ثابتة في النطاق 20 إلى 25°C [3] .

2.5.I. طريقة الكروماتوغرافيا السائل عالية الأداء :

تجرى الكروماتوغرافيا السائل عالية الأداء (HPLC) في أعمدة تحليلية معبأة بطور صلب تجاري يحتوي سلاسل هيدروكربون طويلة مربوطة كيميائياً على السيليكا وتتحرك المواد الكيميائية المحقونة بطول العمود بمعدلات مختلفة بسبب اختلاف درجات التوزع بين الطور المائي المتحرك وطور الهيدروكربون الثابت. ولا تنطبق الطريقة الكروماتوغرافيا على الأحماض والقلويات أو المركبات الفلزية المعقدة أو المواد الخافضة للتوتر السطحي أو المواد التي تتفاعل مع المذيب. وتطبق الطريقة الكروماتوغرافيا عندما تقع قيمة لوغاريتم معامل التوزيع (أوكتانول – ماء) في النطاق من 0 إلى 2 وهذه الطريقة ذات حساسية أقل لوجود الشوائب في المركب المختبر بالمقارنة بطريقة القارورة الهزازة [3].

3.5.I. طريقة التقلب البطيء :

تتيح طريقة التقلب البطيء الحصول على تعيين دقيق وصحيح لقيمة معامل التوزيع (أوكتانول – ماء) للمركبات التي يصل في ها لوغاريتم معامل التوزيع (أوكتانول – ماء) إلى 7,2. وفي حالة المركبات الشديدة الألفة للدهون، تكون طريقة القارورة الهزازة معرضة لحدوث أخطاء تجريبية، وفي الطريقة الكروماتوغرافية، يلزم عمل استكمال خارجي خارج حدود نطاق التدرج من أجل الحصول على تقديرات لقيمة K_{ow} [3].

لتعيين معامل التوزيع يخلط الماء والأوكتانول والمركب المختبر حتى يتحقق الاتزان بينهما، ومن ثم يقاس تركيز المركب المختبر في كلا الطورين. ويمكن التغلب في طريقة التقلب البطيء على الصعوبات التجريبية المرتبطة بتكوين القطيرات الدقيقة، إلى حد ما، بتحقيق حالة الاتزان بين الماء والأوكتانول والمركب المختبر في مفاعل يقلب بهدوء. ويولد التقلب تدفقاً انسيابياً إلى حد ما بين الأوكتانول والماء، ويعجل التبادل بين الطورين بدون تكوين قطيرات دقيقة [3].

4.5.I. طريقة العمود المولد :

هناك طريقة أخرى متعددة المزايا لقياس لوغاريتم معامل التوزيع (أوكتانول – ماء) هي طريقة العمود المولد. وفي هذه الطريقة، يستخدم عمود مولد لتوزيع المادة المختبرة بين طوري الأوكتانول والماء حيث يعبأ العمود بمادة صلبة ويشبع بتركيز ثابت من المادة المختبرة في الأوكتانول.

الجزء النظري / الفصل الأول: عموميات حول معامل التوزيع (أوكتانول – ماء)

وتشطف المادة المختبرة من العمود المولد المشبع بالأوكتانول باستخدام الماء. ويمثل المحلول المائي الذي يخرج من العمود تركيز اتزان المادة المختارة التي هجرت طور الأكتانول إلى طور الماء.

والميزة الأساسية لطريقة العمود المولد بالمقارنة بطريقة القارورة الهزازة هي أن الأولى تتجنب تماما تكوين المستحلبات الدقيقة. لذلك، فإن هذه الطريقة مفيدة بوجه خاص لقياس قيم لوغاريتم معامل التوزيع (أوكتانول – ماء) الأقل من 4,5. ويتمثل عيب طريقة العمود المولد في المولد في أنها تتطلب أجهزة معقدة [3].

6.I. النمذجة:

مقدمة:

تعرف النمذجة بأنها عبارة عن خطوات متتالية لبناء نموذج يتناول الواقع وتترجم أجزاءه في متغيرات ومعادلات تربط بينها علاقات معينة.

مع قدوم الطرق النظرية للحساب، الأكثر فالأكثر تقدما، ووسائل الحساب الأكثر سهولة، تعد الكيمياء الحاسوبية (la chimie informatique) الآن الأداة الأكثر استعمالا في الصناعة وفي الوسط الأكاديمي. إن نمذجة جزيء بواسطة جهاز الحاسوب تعطي عموما عرض بياني لهندسة أو شكل الجزيء بعد تطبيق طريقة نظرية [4].

استخدام النمذجة يرجع للأسباب الآتية:

- قلة التكلفة .
- سهولة التنفيذ ودقة الأداء .
- وفرة الجهد والوقت .
- المرونة .
- إمكانية تطبيق النمذجة على مدى واسع من مجالات الحياة المختلفة [5].

1.6.I. ما هي النمذجة الجزئية ؟ :

كلمة "الجزئية" تعني بوضوح التعامل مع الجزئيات. ويعرف قاموس أوكسفورد النموذج بأنه "وصف مبسط أو مثالي لنظام أو عملية، في المصطلحات الرياضية كثيرا ما يستخدم لتسهيل العمليات الحسابية والتوقعات". تهتم النمذجة الجزئية بتقليد سلوك أنظمة الجزئي والجزئيات.

كما ترتبط هذه النمذجة بشكل ثابت بالنمذجة الحاسوبية. ومع ذلك، أحدثت التقنيات الحاسوبية ثورة في النمذجة الجزئية إلى درجة أن غالبية الحسابات لا يمكن أن تتجز بدون استعمال الحاسوب. هذا لا يعني أن نمودجا أكثر تطورا هو بالضرورة أفضل من أي واحد بسيط، ولكن أجهزة الكمبيوتر لديها بالتأكيد مجموعة أوسع من النماذج التي يمكن النظر فيها والنظم التي يمكن تطبيقها^[5].

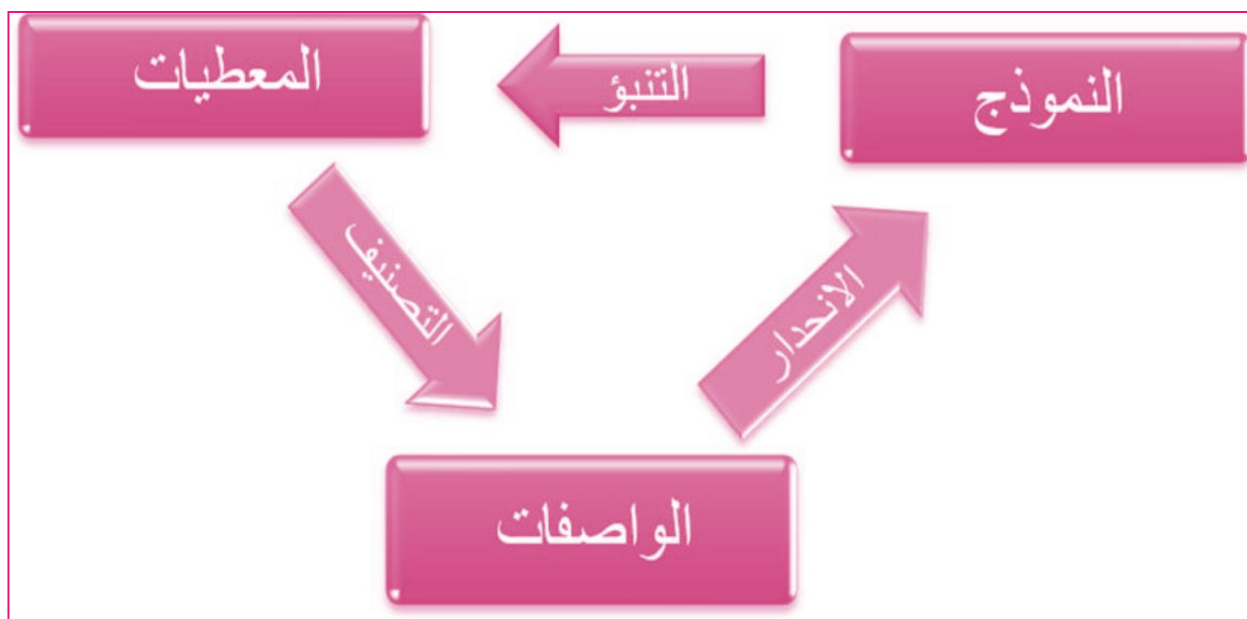
2.6.I. النمذج الرياضي :

النموذج الرياضي عبارة عن مجموعة من المعادلات التي تصف سلوك النظام تحت الدراسة وهذه المعادلات يكون عددها مساويا لعدد المتغيرات فيها، وعادة هذه المعادلات تكون مبرمجة حيث تحل بواسطة الكمبيوتر وهذه هي الطريقة الأنسب للتعامل مع المعادلات الرياضية التي تتطلب حلا عدديا^[5].

3.6.I. مزايا استخدام النماذج الرياضية :

هناك العديد من مزايا استخدام النماذج الرياضية نذكر منها:

- تساعد الباحث على ترتيب المعتقدات النظرية والملاحظات المبدئية عن النظام واستنباط النتائج المنطقية لهذا الترتيب.
- تسمح بالتحكم في مصادر التغير أكبر مما تسمح به الدراسة المباشرة.
- تصف المشكلة بطريقة مختصرة وهذا أفضل من الوصف الشفهي.
- تساعد على سرعة التحليل.
- تساعد على فهم النظام.
- أقل تكلفة من النظام^[5].



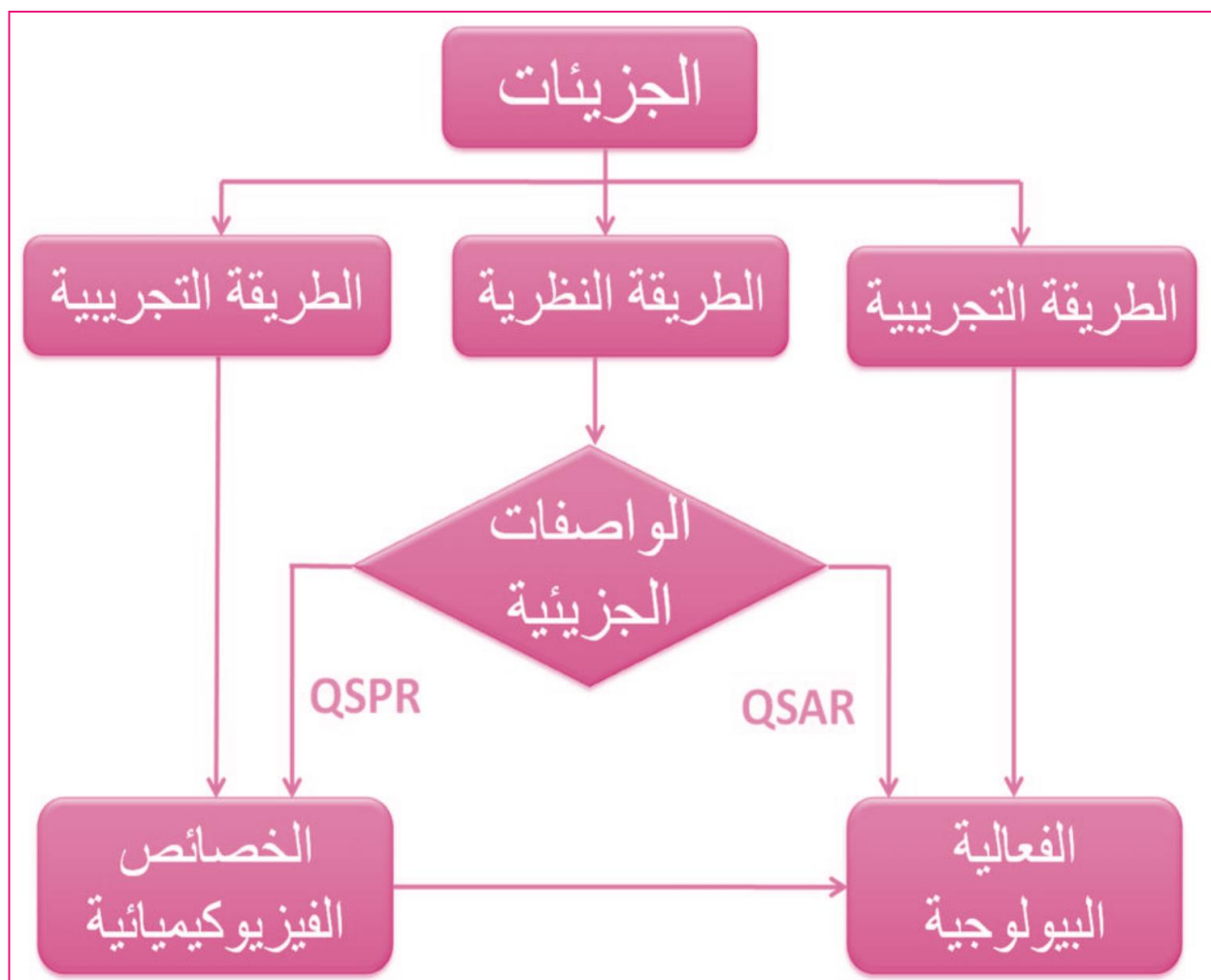
الشكل (I-2): خطوات النمذجة الجزيئية.

7. I العلاقات الكمية (البنية الكيميائية - الخاصية/الفعالية QSAR/QSPR) :

مقدمة:

العلاقة الكمية هي علاقة تربط بين البنية الكيميائية للجزيئات ونشاطها البيولوجي أو خصائصها الفيزيوكيميائية.

يتم استخدام العلاقة الكمية (البنية الكيميائية- الخاصية/الفعالية) بشكل متزايد، بسبب نمو وسائل الحساب. ففي الآونة الأخيرة يعطي إدخال لائحة REACH الأوروبية الجديدة ، التي توصي باستخدام هذه العلاقة من أجل تطوير النماذج التنبؤية حيث استمر استخدامها في النمو حتى أصبح لا غنى عنه في الكيمياء الصيدلانية وتصنيع الأدوية. ان تطورها في مجموعة واسعة من التطبيقات يفتح آفاقا كبيرة من أجل بناء نماذج رياضية بين البنية الكيميائية والنشاط البيولوجي أو الخصائص الفيزيوكيميائية، وهذا ما يسمى بالعلاقة الكمية بين الفعالية/الخاصية والبنية الكيميائية [3].



الشكل (3-I) : علاقة QSPR/QSAR بتقييم المعايير.

1.7.I. نبذة تاريخية عن (QSPR/QSAR) :

يرجع تاريخ استخدام العلاقات الكمية (QSPR/QSAR) في علم السموم المائية إلى الأعمال التي أنجزها في أواخر القرن التاسع عشر الباحثان كل من (أوفرتون في زيوريخ) و (ماير في ماربورغ) حيث أوضح هذان الباحثان أن قدرة المواد على إحداث تخدير في صغار الضفادع وصغار الأسماك ترتبط ارتباطاً مباشراً بمعامل توزعها المقاس بين زيت الزيتون والماء. وطرح "أوفرتون" في دراسته في سنة 1901 فرضية بأن هذا الارتباط يظهر السمية التي تحدث في تركيز أو حجم جزيئي (مولي) القياسي في نطاق موقع جزيئي معين في داخل الكائن العضوي.

كما خلص إلى أن هذا يناظر التركيز أو الحجم نفسه في مختلف الكائنات بغض النظر عما إذا كان الامتصاص من الماء أو عن طريق استنشاق غازي. ويعرف هذا الارتباط في علم التخدير باسم نظرية" ماير -أوفرتون".

اقترح كورفن هانس ومعاونوه في كلية بومونا استخدام نظام معامل التوزيع K_{ow} كنظام قياسي للتوزيع، ووجد أن معاملات التوزيع يمكن تقديرها مباشرة على أساس التركيب الكيميائي. و ذكر هؤلاء الباحثين كذلك أنه يمكن تصميم نماذج للعلاقات (QSPR/QSAR) على أساس تحليل إحصائي للانحدار في النتائج. في عام 1982 أعلن الباحثون أنهم بتطبيق هذا المنهج حققوا 138 نموذجا للعلاقات (QSPR/QSAR) على شكل:

$$\text{Log}(1/C) = A \text{Log}k_{ow} + B \quad (4-I)$$

K_{ow} : هو معامل التوزيع (الأكتانول – الماء) .

C : التركيز المولي للمادة الكيميائية الذي يسبب استجابة بيولوجية قياسية لتأثير المركبات العضوية غير الإلكتروليتية وغير المتفاعلة في حيوانات كاملة، أو أعضاء، أو خلايا، بل وأنزيمات نقية.

وكانت الانحدارات ومقاطع محاور الإحداثيات متطابقة تقريبا في خمس من هذه المعادلات التي سميت خمسة كحولات بسيطة أحادية الهيدروكسيل لخمسة أنواع من الأسماك، وكانت في الواقع مطابقة تقريبا لما توصل إليه كونمان في 1971، الذي يبدو أنه لم يكن على علم بأعمال هانس السابقة.

وأثبت كونمان وآخرون أن كل هذه المواد البسيطة غير الإلكتروليتية وغير المتفاعلة تمل من خلال آلية تخدير في اختبار السمية الحادة في الأسماك، مما يؤدي إلى سمية دنيا أو سمية خط الأساس [3].

2.7.I. مشكلة نمذجة العلاقات الكمية (البنية الكيميائية الخاصة/الفعالية) :

إن اختيار قيمة مناسبة للعلاقات QSPR/QSAR يعني أن يعطى النموذج تنبؤا موثوقا لسمية مادة كيميائية لم تختبر. وبصفة عامة، تتناقص الموثوقية مع تزايد تعقيد التركيب الكيميائي، ما لم تكن قيمة العلاقات QSPR/QSAR مشتقة لمجموعة ضيقة من المواد الكيميائية تشبه المادة المرشحة في التركيب. في المعتاد، تستخدم نماذج لعلاقة (QSPR/QSAR) مشتقة لرتب ضيقة محدودة من المواد الكيميائية في تطوير المستحضرات الصيدلانية متى تم تعيين مركب رائد جديد.

الجزء النظري/ الفصل الأول: عموميات حول معامل التوزيع (أوكتانول – ماء)

ومتى كانت هناك حاجة إلى إدخال تعديلات تركيبية طفيفة لزيادة نشاطه وتقليل سميته، والهدف إجمالاً هو عمل تقديرات عن طريق الاستيفاء الداخلي أكثر منه عن طريق الاستكمال الخارجي [3]

3.7.I. استخدام علاقات (QSPR/QSAR) لتقدير قيمة LogKow:

عند وجود قيمة تقديرية للمعامل LogKow يجب أخذ طريقة التقدير في عين الاعتبار، وقد وضعت ولا تزال توضع قيم عديدة للعلاقات (QSPR/QSAR) بغرض تقدير قيمة المعامل LogKow. وتستخدم في أحيان كثيرة أربعة برامج حاسوبية متاحة تجارياً (CLOGP, LOGKOW, AUTOLOGP, SPARC) لتقدير المخاطر إذا لم تتوفر بيانات مستقاة من التجربة.

وتقوم البرامج AUTOLOGP و LOGKOW و CLOGP على أساس إضافة مساهمات المجموعات التفاعلية، بينما يقوم برنامج SPARC على أساس خوارزمية نظرية بدرجة أكبر لمحاكاة التركيب الكيميائي.

ويوصى باستخدام برنامج LOGKOW و CLOGP لإغراض التصنيف بسبب موثوقيتهما وتوفرهما تجارياً، وسهولة استخدامهما [3].

الجدول (1-I): برامج النمذجة المستخدمة في علاقة (QSPR/QSAR) لإيجاد قيم LogKow [3].

المواد المعنية	مجال قيم LogKow	برنامج النمذجة
يحسب هذا البرنامج قيم LogKow للمركبات العضوية، التي تحتوي الذرات التالية: C, H, N, O, الهالوجين, Si, Li, Na, K P, Se, و/أو Hg.	$9 > \text{LogKow} > 0$	CLOGP
يحسب هذا البرنامج قيم LogKow للمواد العضوية التي تحتوي الذرات التالية: C, H, N, O, الهالوجين, Si, P, Na, K Li, Se, و/أو Hg.	$8 > \text{LogKow} > 4$	LOGKOW (KOWWIN)
يحسب البرنامج قيم LogKow للمواد الكيميائية العضوية التي تحتوي على ذرات H, N, O, C الهالوجين, P و S	$\text{LogKow} < 5$	AUTOLOGP

4.7.I. تعريف QSPR:

(QSPR): Quantitative Structure - Property Relationship

QSPR هي العملية التي يتم من خلالها إنشاء روابط كمية بين التركيب الجزيئي لمجموعة من المركبات مع خاصية فيزيوكيميائية. ويمكن وصف مراحل التطوير الرئيسية لنموذج QSPR على النحو التالي:

- اختيار واصفات متكيفة مع إشكالية البنية-الخاصية.
- استغلال قيم الواصفات الجزيئية كمتغيرات لتحديد العلاقة التي تربطها مع الخاصية المدروسة، وذلك باستخدام آليات خاصة.
- إنشاء معايير الأداء والتحقق من شأنها أن تساعد في اختيار أفضل النماذج لمشكلة معينة وتقدير أخطاء التنبؤ^[6].

5.7.I. مبدأ QSPR:

إن مبدأ العلاقة QSPR هو ، كما يوحي اسمها ، إقامة علاقة رياضية كميًا تربط بين الخصائص الجزيئية الإلكترونية والهندسية على حد سواء ، والتي يطلق عليها اسم الواصفات الجزيئية (النشاط البيولوجي السمية ، الخاصية الفيزيائية الكيميائية ، وما إلى ذلك) ، لسلسلة من المركبات الكيميائية المماثلة باستخدام أساليب تحليل البيانات. كما أن الشكل العام للنموذج هو:

(الواصفات) = f الخاصية

ولذلك فإن الهدف من هذه الطريقة هو تحليل البيانات البنوية من أجل الكشف عن العوامل المحددة للخاصية المقاسة. و للقيام بذلك ، يمكن استخدام أنواع مختلفة من الأدوات:

- الانحدار متعدد الخطوط (MLR) .
- انحدارات المربعات الصغرى الجزيئية (PLS) .
- أشجار القرار (DT).
- الشبكات العصبية (NN).
- الخوارزميات الجينية (GA).

حالما يتم تنفيذ هذه العلاقة والتحقق من صحتها ، يمكن استخدامها بعد ذلك للتنبؤ بخاصية جزيئات جديدة ، والتي لا تتوفر لها القيمة التجريبية^[3].



الشكل (I-4): مبدأ طريقة QSPR.

6.7.I الطرق الرياضية المستخدمة في النموذج QSPR:

- الانحدار الخطي البسيط (SRL).
- الانحدار الخطي المتعدد (MRL).
- الانحدار المكونات الرئيسية (PCR).
- الانحدار المربعات الصغرى (PLS) [6].

الفصل الثاني

طرق الحساب والبرامج المستعملة

II. طرق الحساب والبرامج المستعملة:

II.1. الخطوات المتبعة في التنبؤ:



الشكل (II-1): الخطوات المتبعة في التنبؤ.

2.II. طرق الحساب:

مقدمة:

سمح التطور المهم والسريع لوسائل الإعلام الآلي للكيميائيين باقتناء أداة حاسوبية جديدة تمكنهم من فهم المشاكل المعقدة بالمجال الجزيئي، ألا وهي النمذجة الجزيئية. حيث تعطي هذه الأخيرة معلومات حول الحركات الفعالة للجزيئات وطاقاتهم والتنبؤ الكيفي بخصائصهم (الشحنة، الإستقطابية، الحجم والسطح الجزيئي...)، كما تستعمل كذلك طريقة النمذجة الجزيئية في ميدان الكيمياء العضوية في تقدير درجة الارتباط المتعلقة ب (بنية - فعالية (QSAR) أو بنية - خاصية (QSPR))، وفي السنوات الأخيرة تطور استخدام النمذجة الجزيئية في حساب نماذج QSPR/QSAR باستخدام الانحدار الخطي المتعدد.

1.2.II. الانحدار الخطي المتعدد:

الانحدار هو أداة إحصائية تقوم ببناء نموذج إحصائي وذلك لتقدير العلاقة بين متغير تابع الذي يمثل الخاصية الفيزيوكيميائية ومجموعة من المتغيرات المستقلة، والتي تتمثل الواصفات الجزيئية، بحيث ينتج معادلة إحصائية توضح العلاقة بين المتغيرات [7].

ويفيدنا في:

- تحديد شكل العلاقة بين المتغيرات رياضياً وبيانياً (خط الانحدار).
- توضيح اتجاه العلاقة بين المتغيرات.
- التنبؤ بقيمة المتغير التابع بدلالة المتغيرات الأخرى.

2.2.II. تعريف الانحدار الخطي المتعدد:

يعرّف الانحدار الخطي المتعدد على أنه تقنيّة إحصائية تسمح بتقييم العلاقة بين المتغير التابع (Y) ومجموعة من المتغيرات المستقلة X_1, X_2, \dots, X_K ، لذا فهو يستخدم للتنبؤ بقيمة المتغير التابع من خلال مجموعة من المتغيرات المستقلة، أي تعتمد فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أو الانتشار [7].

إن الانحدار الخطي المتعدد ليس مجرد أسلوب واحد وإنما مجموعة من الأساليب التي يمكن استخدامها لمعرفة العلاقة بين متغير تابع مستمر وعدد من المتغيرات المستقلة التي عادة ما تكون مستمرة [7].

هذه المعادلة تتضمن (k+1) من المعلومات المطلوب تقديرها علما بأن الحد الأول منها a يمثل الحد الثابت، الأمر الذي يتطلب اللجوء إلى المصفوفات والمتجهات لتقدير تلك المعلومات. وعليه يمكن صياغة هذه المعادلات في صورة مصفوفات كالآتي^[7]:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix} \quad (3-II)$$

وباختصار يمكن كتابتها كالآتي:

$$Y = XB + U \quad (4-II)$$

حيث أن:

Y: متجه عمودي أبعاده (n+1) ويحتوي على المتغير التابع.

X: مصفوفة أبعادها (n × k+1) وتحتوي على المتغيرات المستقلة.

B: متجه عمودي أبعاده (k+1 × 1) يحتوي على المعالم المطلوب تقديرها.

U: متجه عمودي أبعاده (n × 1) يحتوي على الأخطاء العشوائية^[7].

باستخدام الفروض السابقة تصبح أفضل معادلة لتقدير قيمة المتغير التابع بدلالة المتغيرات المستقلة هي:

$$\hat{Y}_j = \hat{a} + \hat{b}_1 X_{1j} + \hat{b}_2 X_{2j} + \dots + \hat{b}_{k-1} X_{kj} \quad , j=1, 2, \dots, n \quad (5-II)$$

حيث أن:

\hat{Y} : القيمة التقديرية لمتغير التابع.

\hat{a} : القيمة التقديرية لثابت الانحدار الخطي.

b_1 : ميل الانحدار y على القيمة التقديرية لمتغير المستقل الأول.

b_2 : ميل الانحدار y على القيمة التقديرية لمتغير المستقل الثاني.

X_{1j} : المتغير المستقل الأول.

X_{2j} : المتغير المستقل الثاني [7].

يمكن استخدام الانحدار الخطي المتعدد في حالة توافر الشروط التالية:

1. أن تكون العلاقة خطية بين المتغيرات المستقلة والمتغير التابع.

2. أن تكون البيانات موزعة توزيعاً طبيعياً للمتغيرات المستقلة والمتغير التابع.

3. يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل [7].

بعد الحصول على نتائج معادلة الانحدار يجب علينا أن نبين هل أن هذه المعاملات مقبولة من الناحية

الإحصائية أي معنوية إحصائياً مع التنويه بأن المعنوية تكون لكل معامل على حده.

يعتبر معامل الارتباط الخطي مقياساً جيداً لدراسة العلاقة بين متغيرين والتي تأخذ الشكل الخطي [8]، توجد العديد من الطرق أو الأساليب التي يتم من خلالها اختيار أفضل نموذج للتعبير عن العلاقة بين المتغيرات المستقلة والمتغير التابع نذكر منها: معامل الارتباط الخطي (r)، معامل التحديد (R^2)، الاختبار الإحصائي (F).

II.4.2. الارتباط الخطي:

يمكن من خلاله وصف شكل توزيع البيانات التي تم جمعها من متغير واحد، وننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، وسنتناول في هذه الوحدة دراسة وتحليل العلاقة بين متغيرين، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، وذلك لدراسة العلاقة بين متغيرين (X)، (Y) [9].

II.5.2. تعريف الارتباط:

يبحث تحليل الارتباط في دراسة العلاقة بين متغيرين أو أكثر لمعرفة ما إذا كان تغير أحدهما أو مجموعة منها مرتبط بتغير الآخر، و بعبارة أخرى فإن الارتباط يعني قياس قوة واتجاه العلاقة بين متغيرين أو أكثر [10].

يعني الارتباط تماشي قيم متغير مع متغير آخر في الاتجاه الطردي أو المعاكس وبذلك يصحب التغير في أحدهما بتغير الآخر [11].

6.2.II. معامل الارتباط الخطي (r):

يعرف معامل الارتباط والذي يرمز له بالرمز r بأنه عبارة عن مقياس رقمي يقيس قوة الارتباط بين متغيرين [5].

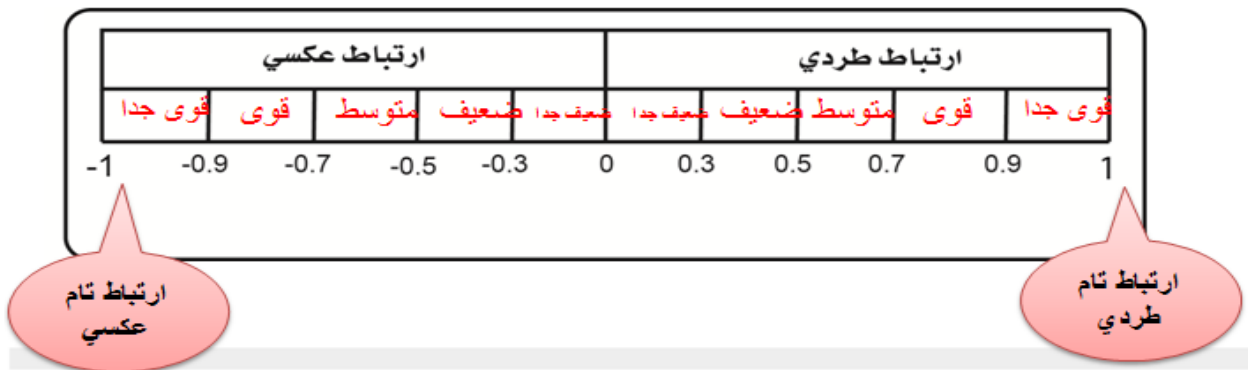
1.6.2.II. نوع العلاقة:

تأخذ العلاقة بين المتغيرين ثلاث صور مختلفة حسب إشارة معامل الارتباط كما يلي:

- $r = -1$ ارتباط عكسي تام (كل المشاهدات تقع على خط مستقيم ذي ميل سالب)، أي المتغيرين لا يتغيران في نفس الاتجاه.
- $r = +1$ ارتباط طردي موجب تام (كل المشاهدات تقع على خط مستقيم ذي ميل موجب)، أي المتغيرين يتغيران في نفس الاتجاه.
- $r = 0$ عدم وجود أي ارتباط بين المتغيرين [12].

2.6.2.II. قوة العلاقة:

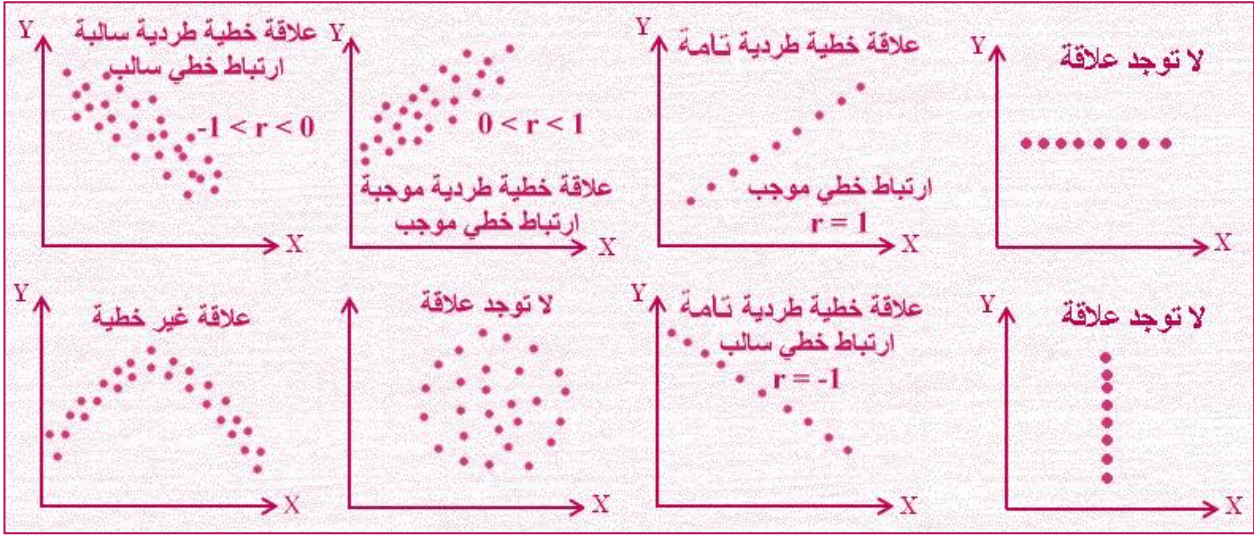
عندما يكون $(-1 \leq r \leq 1)$ تعطى علاقة أساسية للارتباط وتدل القيمة على وجود ارتباط إحصائي بين المتغيرين، وتكون قوته أو ضعفه حسب قرب أو بعد r من ± 1 ومنه يكون التمييز بين الارتباط الطردي أو العكسي، ويمكن تمثيل درجات قوة العلاقة على الشكل التالي [13]:



الشكل (II-2): درجة قوة علاقة معامل الارتباط.

3.6.2.II. شكل الانتشار:

شكل الانتشار عبارة عن تمثيل قيم متناظرة للمتغيرين (X)، (y) بمجموعة من النقاط ، فيتم تمثيل المتغير x على المحور الأفقي ، و المتغير y على المحور العمودي، حيث ينتج عن الأزواج المرتبة (x) ، (y) نقطة على الرسم فنحصل على شكل الانتشار [9].



الشكل (3-II): شكل انتشار الارتباط في الانحدار الخطي المتعدد.

4.6.2.II. قياس الارتباط:

يمكن قياس معامل الارتباط الخطي بدلالة القراءات لبيانات المتغيرين x و y، باستخدام الصيغة التالية:

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad (6-II)$$

حيث:

n: عدد الأزواج قيم المتغيرين.

$\sum XY$: مجموع حاصل ضرب Y في x .

$\sum X$: مجموع قيم X.

$\sum Y$: مجموع قيم Y.

$\sum X^2$: مجموع مربعات قيم X.

$\sum Y^2$: مجموع مربعات قيم Y [14].

7.2.II. معامل التحديد (R^2):

R^2 هو معامل التحديد والذي يستخدم لمعرفة القوة التفسيرية للنموذج المقدر في حالة الانحدار الخطي المتعدد لأنه يأخذ بعين الاعتبار المتغيرات المستقلة.

تعريف:

يعرف معامل التحديد بأنه مربع معامل الارتباط ويرمز له بالرمز R^2 حيث: $R^2 = r$ [15].

وهو معامل يستخدم لمعرفة مدى جودة التوفيق لخط الانحدار المقدر للمتغير التابع على المتغير المستقل، وبما أنه عبارة عن مربع معامل الارتباط لذلك فقيمه تكون موجبة محصورة بين الصفر والواحد، أي: $0 \leq R^2 \leq 1$ [15].

هو يقيس نسبة التغير في المتغير التابع نتيجة تغير المتغير المستقل، أي أنه يوضح نسبة مساهمة المتغير المستقل في التغير الحاصل في المتغير التابع [15].

- يكون التقدير جيدا كلما اقتربت قيمة R^2 من الواحد.
- إذا كان: $R^2 = 1$ يعني أنه يوجد علاقة خطية تامة بين المتغيرين التابع والمستقل أي أن المتغير المستقل يساهم بنسبة 100% في التغير الحاصل في المتغير التابع.
- كلما اقترب R^2 من الصفر كلما دل ذلك على ضعف مساهمة المتغير المستقل في تفسير التغير الحاصل في المتغير التابع [15].

يمكن إيجاد معامل التحديد إما من خلال معامل الارتباط أو من خلال ما يلي:

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - Y)^2}{\sum_{i=1}^n (Y_i - Y)^2} \quad (7-II)$$

حيث:

\hat{Y}_1 : القيمة التقديرية لمتغير التابع.

Y_i : المتغير التابع.

Y : متجه البعد n [15].

8.2.II. الاختبار الإحصائية (F):

نستخدم إحصائية f للحكم على معنوية النموذج المقدر ككل عند مستوى معنوية^[17].

تعريف:

لقياس معنوية النموذج يتم استخدام اختبار (F) لاختبار معنوية النموذج وذلك بتحديد ما إذا كانت هناك علاقة معنوية بين المتغير المعتمد (Y_j) والمتغيرات المستقلة X₁، X₂، ... X_K حيث تستخرج قيمة (F) كالآتي^[16]:

$$F = \frac{\sum_{i=1}^n (\hat{Y}_1 - Y)^2 / K}{\sum_{i=1}^n (Y_i - Y)^2 / (n - K - 1)} \quad (8-II)$$

حيث :

\hat{Y}_1 : القيمة التقديرية لمتغير التابع.

y_i : المتغير التابع.

y : متجه البعد n.

n : العدد الإجمالي.

k : عدد طبيعي متغير^[17].

يتم حساب التشتت حول المتوسط الحسابي من خلال ما يلي:

9.2.II. متوسط الانحرافات المطلقة (Average Absolute Deviation) AAD:

يقيس انحراف القيم عن وسطها الحسابي بغض النظر عن إشارة ذلك الانحراف حيث يتم حسابه، ويرمز له بالرمز σ ، ويمكن حساب متوسط الانحرافات المطلقة من خلال المعادلة التالية:

$$\sigma = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n} \quad (9-II)$$

حيث:

\bar{X} : الوسط الحسابي.

X_i : مركز الفئة i وهي تساوي (الحد الأعلى للفئة + الحد الأدنى للفئة) / 2.

n : عدد الفئات [18].

10.2.II. الانحراف المعياري SD (Standard Deviation):

نعرف الانحراف المعياري لعينة حجمها n مسحوبة من مجتمع ما بأنه الجذر التربيعي لتباين هذه البيانات

وبالتالي فإن الانحراف المعياري للبيانات X_1, X_2, \dots, X_n والتي وسطها الحسابي \bar{X} ، ويرمز له بالرمز S^2

ويمكن حساب الانحراف المعياري من خلال المعادلة التالية:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (10-II)$$

حيث:

\bar{X} : الوسط الحسابي.

X_i : مركز الفئة i وهي تساوي (الحد الأعلى للفئة + الحد الأدنى للفئة) / 2.

n : عدد الفئات [18].

3.II. البرامج المستعملة:

مقدمة:

مع ظهور المزيد من أساليب الحساب النظرية وموارد الحوسبة التي يمكن الوصول إليها أكثر أصبحت كيمياء الكمبيوتر الآن أداة مفيدة لفهم أفضل للمشاكل المتعلقة بالمجال الجزيئي، وذلك باستخدام برمجيات النمذجة الجزيئية حيث تساعدنا على معرفة: شكل وهندسة الجزيء، الخصائص الفيزيوكيميائية، تحديد بنية و طاقة الجزيء، الألفة الالكترونية.

1.3.II. أهم الطرق التي تحدد البنية الكيميائية للجزيئات:

توجد العديد من التقنيات الطيفية التي تستعمل إما مجتمعة مع بعضها أو منفصلة، للحصول على معلومات بنيوية حول البنية المجهولة منها:

✓ إن تفسير طيف الأشعة تحت الحمراء يمكنه إعلامنا بوجود أو غياب مجاميع وظيفية نشطة.

✓ انطلاقا من طيف الرنين المغناطيسي النووي (RMN) يمكننا تحديد عدد وطبيعة ومحيط ذرات الهيدروجين لجزيء ما.

✓ طيف الكتلة يزودنا بمعلومات حول الكتلة المولية والصيغة الجزيئية ووضعية المجاميع داخل الجزيء.

✓ الأطياف الالكترونية تزودنا بمعلومات حول البنية الالكترونية للجزيئات.

ومن بين الطرق التجريبية غير الطيفية، الانعراج (la diffraction) بواسطة الأشعة السينية (RX) وهي

الأكثر استعمالا لأنها توفر جميع المعطيات الهندسية للجزيء.

الطرق التي وضعها الكيميائيون النظريون هي الطرق الكمية بينما الطرق المسماة بطرق حقل القوة أو

"الميكانيكا الجزيئية" فتقودنا إلى معرفة البنية ثلاثية الأبعاد 3D والخصائص الفيزيائية والكيميائية الموافقة لها.

من بين الطرق الكمية الأكثر بساطة طريقة هيكل (Huckel) التي لا تأخذ بعين الاعتبار إلا الالكترونات

. (π)

الطرق النصف تجريبية مثل (MNDO)،(CNDO)،PM3 لا تأخذ بعين الاعتبار إلا الكثرونات التكافؤ و تهمل بعض التكاملات و تعوضها ببرامترات موافقة لها.

الطرق الكمية الأكثر تطورا هي المسماة (**ab-initio**) التي تهتم بكل الالكترونات و لكنها تعاني من بعض الحدود المعينة، و يتطلب الحساب وقتا طويلا حتى بالنسبة للجزيئات المتوسطة الحجم. ومهما كانت الطريقة الكمية المستعملة فالاختلاف يكمن في الزمن المستغرق للحساب بالميكانيكا الجزيئية فهو وقت أقل بالنسبة لبقية الطرق، وهذا الاختلاف يزداد بدلالة أبعاد الجزيء هذا ما يفسر اهتمام الصناع بهذه الطريقة^[19].

II.3.2. برنامج (ChemOffice Ultra) :

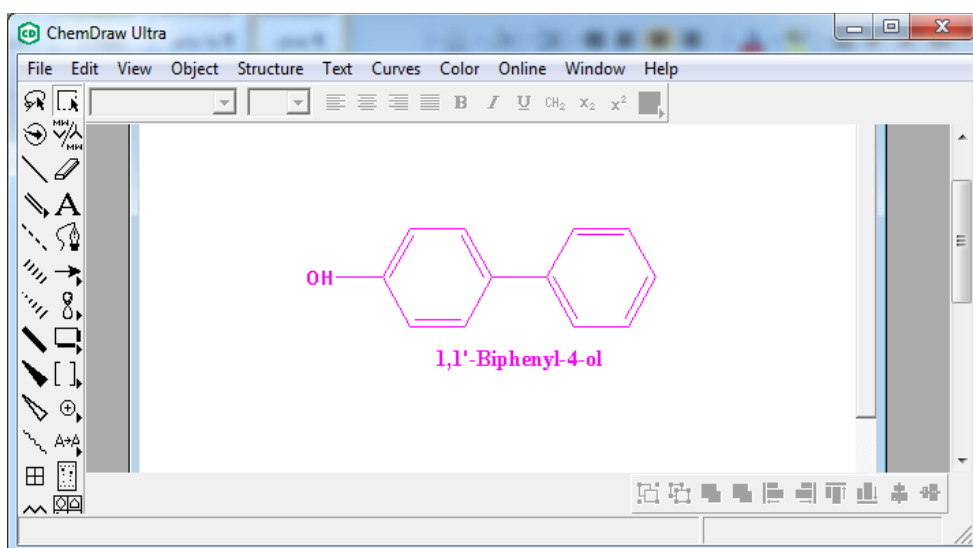
تعريف:

هو عبارة عن أوفيس (Office) خاص بالكيميائيين والعلماء والباحثين والدارسين .ويتميز بالمرونة و سهولة الاستخدام^[15]. هو أيضا عبارة على مجموعة متكاملة من أدوات الإنتاجية الشخصية التي تمكن العلماء والباحثين من التقاط وتخزين واسترجاع وتبادل البيانات والمعلومات عن المركبات، وردود الفعل والمواد وخصائصها^[4].

يحتوي ChemOffice Ultra على التطبيقات التالية:

- ✓ يستخدم ChemOffice Ultra من قبل مئات الآلاف من العلماء في جميع أنحاء العالم بسرعة وفعالية لرسم الجزيئات والتفاعلات البيولوجية لاستخدامها في أجهزة الكمبيوتر ووثائق مختبر الإلكترونية، إلى البحث في قواعد البيانات، وأسماء لتوليد دقيقة من الهياكل وخصائص للتنبؤ والأطياف.
- ✓ Chem Draw Ultra يضيف لبرنامج Excel الاستخبارات الكيميائية لجداول البيانات، بحيث يستخدم الكيميائيون ال Excel في التحليل والفرز وتنظيم أدوات للتلاعب وإثراء مجموعات أخرى من المركبات والبيانات واستكشاف العلاقات بين الهيكل والنشاط .
- ✓ Chem3D Ultra يولد نماذج 3D بحيث يمكن الكيميائيين من عرض مركباتها في ثلاثة أبعاد لتقييم الشكل والخصائص لزيادة نشاط أو خصوصية.

✓ ChemFinder يستخدمه العلماء لتنظيم مركبات قاعدة بيانات الشخصية مع الخصائص وتحويل البيانات إلى رسوم بيانية سهلة الفهم لتبين بسهولة الهيكل ونشاط المركبات [20].



الشكل (4-II): برنامج ChemOffice Ultra.

3.3.II. برنامج (HyperChem):

تعريف:

HyperChem يجمع في واجهة واحدة مجموعة من الأدوات المخصصة للنمذجة الجزيئية المعروفة بجودتها ومرونتها وسهولة استخدامها.

HyperChem هو برنامج يسمح برسم وتحسين الجزيئات، فهو يحتوي على طرق الحساب النظرية (الميكانيكا الجزيئية MM^+ والطرق نصف تجريبية AM1 و PM3)، التي تمكننا من تحديد هندسة الجزيئات وتقدير الخصائص الفيزيوكيميائية المرافقة لها.

وقد أجريت الحسابات بواسطة برنامج (HyperChem) الذي يعتمد على حقل القوة MM^+ (Molecular mechanics)، حيث يتم حفظ كل جزي في ملف على صيغة HIN^[21].

1.3.3.II. طريقة MM^+ (Molecular mechanics):

الميكانيك الجزيئي ويسمى أيضا حقل القوة وهو الطريقة الأساسية لحسابات التمثيل الجزيئي الخاص بالجزيئات، وهو مطور لكي يسمح بتحديد شكل الجزيئات وخصائصها الترموديناميكية، الفيزيوكيميائية وأطياف اهتزازها، يمكن اعتباره كنظام خبير مؤلف من نموذج رياضي^[22].

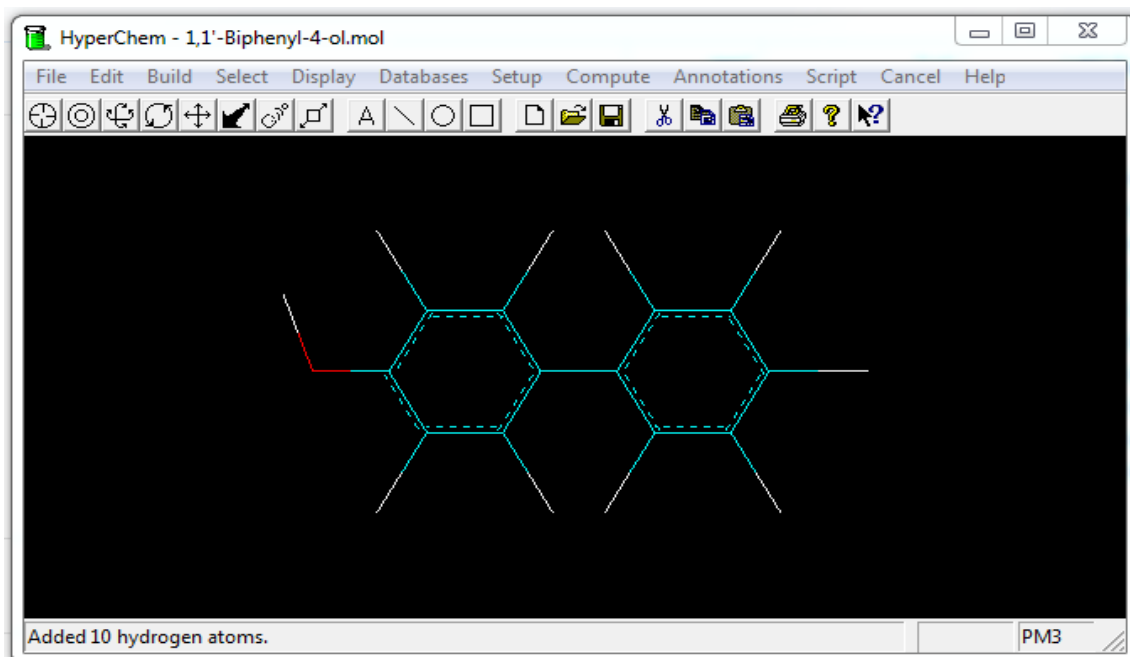
تستخدم دائما الميكانيكا الجزيئية لتنفيذ العمليات الحسابية على الأنظمة التي تحتوي على أعداد كبيرة من الذرات [5].

2.3.II. طريقة AM1 (Austin Model 1):

هي طريقة نصف تجريبية وهي تحسين لطريقة MNDO، حيث تستخدم AM1 على نطاق واسع في نمذجة المركبات العضوية بسبب الأداء وقلة الوقت في الحساب، كما لها القدرة على بناء الروابط الهيدروجينية وحساب طاقة التنشيط للتفاعلات [23].

3.3.II. طريقة PM3 (Parameterized Model number 3):

هي طريقة حساب كيميائية نصف تجريبية و هي امتداد لطريقة AM1 (Austin Model 1) وهي تحسين للطريقة النصف تجريبية MNDO (Modified Neglect of Diatomic Overlap)، (إهمال تعديل التداخل ثنائي الذرة) و يمكننا بواسطة PM3 حساب الخصائص البنيوية، الطاقة الكلية، حرارة التشكل. و هي عموما من أدق الطرق النصف تجريبية التي يتضمنها برنامج HyperChem [24]. الهدف من هذه الدراسة الكمية والكيفية لقاعدة البيانات هو البحث على احتمالات تواجد هذه المركبات في مستويات امتثالي، على أساس هندسي و طاقي. إن عرض الخصائص الامتثالية لهذه المركبات يسمى بالتحليل الامتثالي.



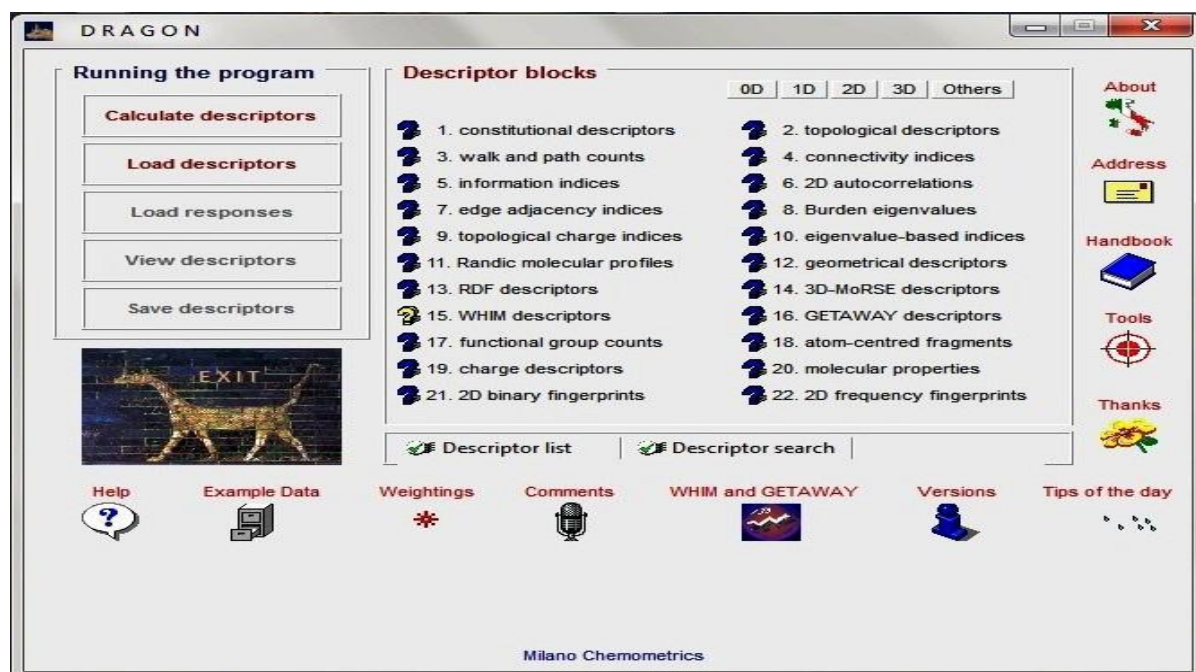
الشكل (II-5): برنامج HyperChem.

4.3.II. برنامج (DRAGON):

تعريف:

برنامج (DRAGON) هو عبارة عن تطبيق لحساب الواصفات الجزيئية، يمكن استخدام هذه الواصفات لتقييم تأثير البنية الجزيئية أو العلاقات الكمية ، وكذلك لتحليل التماثل وفحص قواعد بيانات الجزيئات.

يقوم برنامج (DRAGON) بحساب أنواع أخرى من الواصفات الجزيئية كنوع الذرة، المجموعات الوظيفية وكذلك الواصفات الطوبولوجية والهندسية وبعض الخصائص الفيزيوكيميائية مثل الشحن الجزيئية ، المرنة، الانكسارية ، H-المانحة ، H-الجاذبة^[6].



الشكل (II-6): برنامج DRAGON

1.4.3.II. تعريف الواصف الجزيئي (Molecular descriptor):

الواصف الجزيئي هو النتيجة النهائية لإجراء رياضي لتحويل المعلومات الكيميائية إلى رموز ومعطيات عددية، حيث تلعب الواصفات الجزيئية دور أساسي في الكيمياء ، العلوم الصيدلانية، حماية البيئة، الصحة ومراقبة الجودة.

اقترح العديد من الواصفات بالاعتماد على النظريات المختلفة بهدف توقع نشاط والخواص الفيزيوكيميائية للجزيئات^[25].

يعتمد محتوى معلومات الواصفات الجزيئية على نوع التمثيل الجزئي وعلى الخوارزمية المعرفة لحسابها.

1. فهناك واصفات جزيئية بسيطة مشتقة لحساب بعض أنواع الذرات أو الأجزاء الهيكلية في الجزيئية.

2. هناك واصفات طوبولوجية أو تسمى الواصفات ثنائية البعد (2D) مشتقة من خوارزميات طوبولوجية (رسم بياني جزئي) .

3. واصفات هندسية وتسمى الواصفات ثلاثية البعد (3D) مشتقة من التمثيل الهندسي.

إن أغلب الواصفات لا تستطيع تمثيل التعقيد الجزئي لجزء ما، أو تمثيل كل الاستجابات الفيزيوكيميائية كنتيجة للتفاعلات الحيوية، تعتبر معظمها تقريبات لكن يمكن اتخاذها كنماذج لدراساتها.

II.2.4.3. كتل الواصف الجزيئي:

يحتوي برنامج (DRAGON) على 22 كتلة منطقية تشمل العديد من الواصفات الجزيئية منها :

الطوبولوجية ، الهندسية ، الالكترونية ، و الواصفات المتعلقة بالفاعلية البيولوجية والصيدلانية^[25].

إن كتل الواصفات الجزيئية يمكن أن تختار سويًا مع الواصفات المهمة. والجدول التالي يبين أهم الكتل.

الجدول (II-1): أهم كتل الواصفات الجزيئية^[25].

العدد	الترجمة إلى العربية	اسم الكتلة بالإنجليزية	رقم الكتلة	البعد
48	الواصفات الأساسية	Constitutional descriptors	1	2D
119	الواصفات الطوبولوجية	Topological descriptors	2	
47	مسار وطرق العد	Walk and path counts	3	
33	مؤشرات الربط	Connectivity indices	4	
47	مؤشرات المعلومة	Information indices	5	
96	2D الارتباط الذاتي	2D autocorrelations	6	
107	مؤشرات جوار الحافة	Edge adjacency indices	7	
64	واصفات القيم الحرة ل Burden	Burden Eigen value descriptors	8	
21	مؤشرات الشحن الطوبولوجية	Topological charge indices	9	
44	مؤشرات أساسها قيم حرة	Eigen value-based indices	10	
154	مجاميع المجموعات الوظيفية	Functional group counts	17	

120	أجزاء مراكز الذرة	Atom-centered fragments	18	3D
29	الخصائص الجزيئية	Molecular properties	20	
41	المقاطع الجانبية للجزيئات Randic	Randic molecular profiles	11	
74	الواصفات الهندسية	Geometrical descriptors	12	
150	واصفات RDF	RDF descriptors	13	
160	واصفات 3D-MORSE	D-MORSE descriptors ³	14	
99	واصفات WHIM	WHIM descriptors	15	
197	واصفات GETAWAY	GETAWAY descriptors	16	
14	الواصفات الشحنية	Charge descriptors	19	

3.4.3.II. حساب الواصفات الجزيئية :

لحساب الواصفات الجزيئية للمركبات العضوية المشكلة لقاعدة البيانات استعملنا برنامج (DRAGON) الذي يحوي 1664 واصفة مقسمة على 22 كتلة منطقية، كل كتلة منطقية تتضمن مجموعة من الواصفات الخاصة (طوبولوجية، هندسية، إلكترونية، خصائص فيزيوكيميائية... الخ) [25].

ID	Symbol	Value	Meaning
1	MW	346.280	molecular weight
2	AMW	9.890	average molecular weight
3	Sv	22.330	sum of atomic van der Waals volumes (scaled on Carbon atom)
4	Se	35.490	sum of atomic Sanderson electronegativities (scaled on Carbon atom)
5	Sp	25.320	sum of atomic polarizabilities (scaled on Carbon atom)
6	Ss	39.780	sum of Kier-Hall electrotopological states
7	Mv	0.640	mean atomic van der Waals volume (scaled on Carbon atom)
8	Me	1.010	mean atomic Sanderson electronegativity (scaled on Carbon atom)
9	Mp	0.720	mean atomic polarizability (scaled on Carbon atom)
10	Ms	2.090	mean electrotopological state
11	nAT	35	number of atoms
12	nSK	19	number of non-H atoms
13	nBT	37	number of bonds
14	nBO	21	number of non-H bonds
15	nBM	6	number of multiple bonds
16	SCBO	24	sum of conventional bond orders (H-depleted)

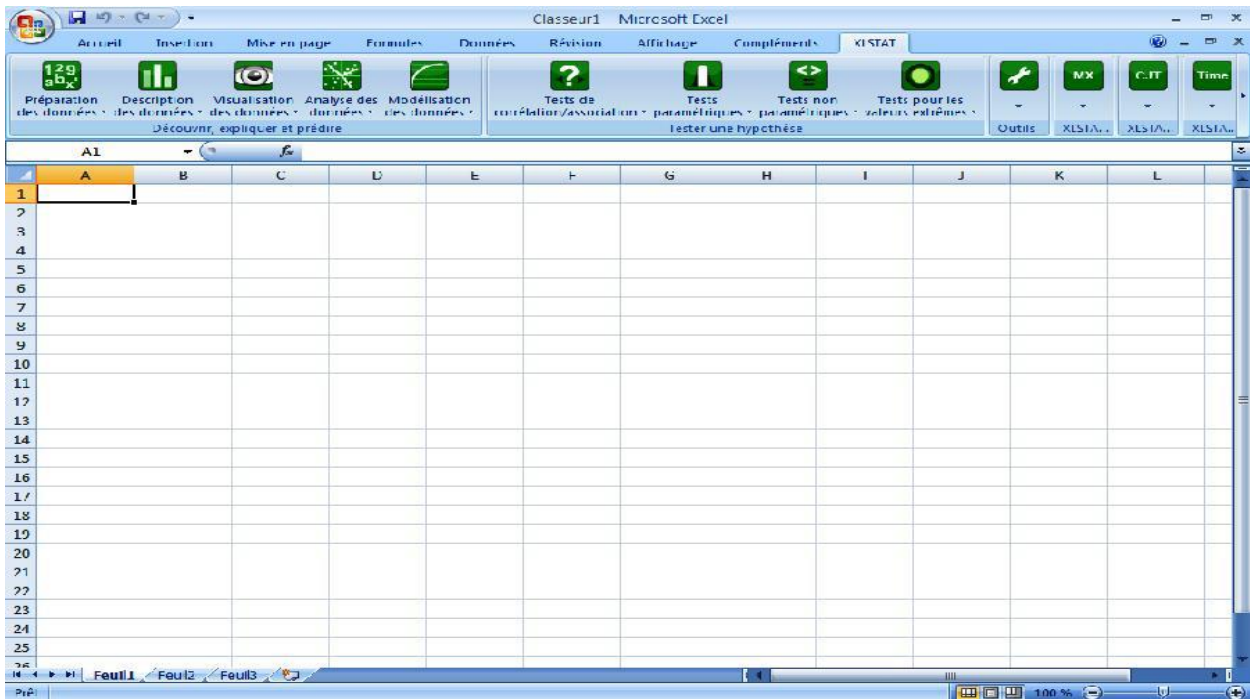
الشكل (7-II): حساب الواصفات الجزيئية.

5.3.II. برنامج (XLSTAT):

XLSTAT هو برنامج يقوم بتحليل البيانات إحصائياً وهو تابع لبرنامج Microsoft Excel. تطويره بفضل Thierry Fahmy مؤسس Addinsoft. البرنامج متاح على شبكة الإنترنت منذ عام 1996 [26].

نختار هذا البرنامج لأنه يحتوي على العديد من الميزات:

- XLSTAT هو برنامج موثوق للغاية ومستقر جداً.
- يسمح لنا بإعادة تنظيم هيكل قاعدة البيانات وتسهيل تحليلها.
- مع XLSTAT، يصبح تحليل المكون الأساسي (PCA)، وذلك بطريقة الانحدار الخطي (بسيطة أو متعدد).
- يتيح لنا الوصول إلى عدد كبير من الاختبارات، تتضمن نتائج الاختبار تفسيراً واضحاً.
- XLSTAT يقدم مجموعة واسعة من الاختبارات الإحصائية التي تسمح لنا باختبار الفرضيات.
- XLSTAT يسمح لنا بتقييم جودة البيانات، ويحدد أخطاء الحساب، مما يساعدنا على تحسين عملنا.
- XLSTAT يشمل بعض الأدوات التي تسهل تصور البيانات [26].



الشكل (II-8): البرنامج XLSTAT.

III. عموميات حول المركبات العضوية:

III.1. تعريف المركبات العضوية:

تتكون المركبات العضوية من عنصر الكربون كعنصر أساسي بالإضافة إلى بعض العناصر الأخرى التي تدخل في تركيبها مثل : الهيدروجين والأكسجين والنتروجين والكبريت والفسفور وغيرها من العناصر [27].

نتتبع مسار المركب العضوي في الوسط الطبيعي على عدة عوامل منها تركيبه الكيميائي، قابليته للذوبان في الماء، قدرته على الالتصاق بالسطوح الأخرى، ثابت تفككه وقابليته للتحويل إلى بخار. فضلا عن عوامل أخرى لا تقل أهمية كالتفكك في الماء، التفكك بالضوء الشمسي وتفاعلات الأكسدة والاختزال ثم التحلل البيولوجي أي التحلل العضوي. كما أنه لا يغرب عن البال تأثيرات خصائص الوسط البيئي نفسه كدرجة الحرارة، ومعدلات سقوط الأمطار، وشدة الرياح والعوامل الجغرافية والمناخية [7].

توجد المركبات العضوية الطبيعية والمصنعة في كافة أنواع التراب وفي المياه وفي الغلاف الجوي لكن بدرجات متفاوتة. وتتحول هذه المركبات العضوية بيئيا عادة إما بالعوامل الكيميائية أو بالعوامل البيولوجية أو بكليهما معا [7].

III.2. تصنيف المركبات العضوية:

بعد اضمحلال نظرية القوة الحيوية أصبح بالإمكان تصنيع المركبات العضوية بكثرة مما استوجب ضرورة تصنيفها لتسهيل عملية دراستها حيث أعتمد هذا التصنيف على عدة أسس منها:

1. تصنيف يعتمد على بناء الهيكل الكربوني للمركبات (حيث صنفنا إلى مركبات حلقة ومركبات غير حلقة).

2. تصنيف يعتمد على الخواص الفيزيائية (صنفنا إلى مركبات عطرية ومركبات دهنية).

3. تصنيف يعتمد على الخواص الكيميائية (صنفنا إلى مركبات مشبعة وغير مشبعة).

4. تصنيف يعتمد على نوع المجموعة الوظيفية في جزئ المركب العضوي) حيث أعتبر هذا التصنيف الأهم والأكثر فائدة وذلك لتشابه المركبات التي لها نفس المجموعة الوظيفية في الخواص الفيزيائية والكيميائية بالإضافة إلى سهولة تسميتها [28].

III.3. ما سبب اهتمامنا بالمركبات العضوية (مركبات الكربون)؟:

يرجع اهتمامنا بمركبات الكربون لسببين:

1. كثرة مركبات الكربون، حيث يبلغ عدد مركبات الكربون المعروفة اليوم ما يقارب 3 مليون مركب، وهذا العدد أكبر من عدد مركبات العناصر الأخرى.
2. أهمية مركبات الكربون (العضوية) بالنسبة للإنسان، فهي تحتل مكانة مهمة في الحياة، فالغذاء والسكن والملابس والأدوية والمنظفات والمبيدات الحشرية وأدوات التجميل معظمها من المركبات العضوية [29].

III.4. ما هي أسباب كثرة المركبات العضوية (مركبات الكربون)؟:

1. قدرة ذرات الكربون على الارتباط ببعض بروابط قوية مكونة سلاسل جزيئية مختلفة الأشكال والأحجام.
2. قدرة ذرات الكربون على الارتباط بذرات عناصر أخرى كالأكسجين والنيتروجين و الهالوجينات وغيرها.
3. اختلاف طريقة ارتباط العناصر في الجزيئات العضوية ، وهذا ما يعرف بظاهرة التشكل [29].

الجزء التطبيقي

الحسابات والنتائج

الفصل الثالث

قاعدة البيانات

III. قاعدة البيانات:

مقدمة:

تتكون قاعدة البيانات من مجموعة معتبرة من المركبات الكيميائية العضوية وكذلك القيم التجريبية لمعامل التوزيع LogK_{ow} الخاصة بكل مركب. حيث تم تحضير هذه القيم في المخبر وذلك باستعمال طرق خاصة منها: طريقة القارورة الهزازة، طريقة الكروماتوغرافيا السائلة العالية الأداء (HPLC) High-performance (liquid chromatography), طريقة التقلب البطيء وطريقة العمود المولد.

1.III. مكونات قاعدة البيانات:

تحتوي قاعدة البيانات على 992 مركب عضوي يمثل مختلف العائلات الكيميائية (هيدروكربونات، الأمينات، الأميدات، كحولات، مركبات صيدلانية...) ، ولكل مركب قيمة معامل التوزيع (الأكتانول - الماء) LogK_{ow} خاصة به بالإضافة إلى 7 كتل ثلاثية البعد 3D للوصفات الجزيئية المحسوبة بواسطة برمجية Dragon.

الشكل (1-III): يوضح الجدول التالي جزءا من قاعدة البيانات.

الرقم	LogKow	الاسم	SMILES
1	3.94	Acenaphthylene	c1ccc2cccc3c2c1C=C3
2	-0.53	Acetamide,2-chloro	C=C(N)CCl
3	-1.11	Acetic acid, hydroxy	O=C(O)CO
4	3.2	Azulene	C1=CC=C2C=CC=C2C=C1
5	1.73	Barbituric acid,5-butyl-5-ethyl	CCCCC1(CC)C(=O)NC(=O)NC1=O
6	1.76	Benzaldehyde,4-methoxy	O=Cc(ccc(OC)c1)c1
7	0.59	1,4-Benzenediol	Oc(ccc(O)c1)c1
8	0.9	Benzenamine	Nc(cccc1)c1
9	0.51	Chlorofluoromethane	FCCl
10	0.84	Cyclohexanone,oxime	N(O)=C(CCCC1)C1
11	3.82	Diazene,diphenyl	N(=Nc(cccc1)c1)c(cccc2)c2
12	-2.32	D-Ribose	OCC1OC(O)C(O)C1O
13	0.58	Ethanamine,N-ethyl	N(CC)CC

14	0.37	Ethanol, 2,2-dichloro	OCC(Cl)Cl
15	-2.95	Hexanoic acid,6-amino	O=C(O)CCCCCN
16	-1.52	L-Leucine	O=C(O)C(N)CC(C)C
17	-2.54	L-Proline	O=C(O)C(NCC1)C1
18	1.7	m-Cyanophenol	Oc1cccc(C#N)c1
19	1.94	3-Methoxysalicylic acid	Coc1cccc(C(O)=O)c1O
20	0.06	N-Phenylsuccinimide	O=C1CCC(=O)N1c2ccccc2
21	1.16	Oxindole	c1ccc2CC(=O)Nc2c1
22	0.49	Pentanoic acid, 4-oxo	O=C(O)CCC(=O)C
23	0.27	1,4-Dioxane	O(CCOC1)C1
24	-2.17	Uric acid	O=C1NC(=O)NC2=C1NC(=O)N2
25	2.46	Zoxazolamine	c1c(Cl)cc2nc(N)cc2c1

2.III. كيفية معالجة وحساب الواصفات الجزيئية المكونة لقاعدة البيانات:

يعطينا برنامج (DRAGON) 1664 واصفة جزيئية مقسمة على 22 كتلة، ولإنقاص هذا الكم الهائل

من الواصفات نقوم بما يلي :

1. نحذف كل الواصفات الجزيئية الثابتة بحيث أنه يمكن أن نجد هناك جزيئات في قاعدة البيانات لها نفس القيم في نفس الواصفة.

2. نحذف كل الواصفات الجزيئية التي تشكل دالة خطية مع الواصفات الجزيئية الأخرى تسمى هذه الطريقة بإزالة الارتباطات الخطية.

3. ضبط قاعدة البيانات : نجري على قاعدة البيانات عملية الضبط (Normalisation) وذلك بحصر البارامترات (الواصفات) بين $(-1 \leq 0 \leq 1)$, هذا لتحقيق التجانس بين مختلف القيم، والطريقة هذه تؤدي إلى إنقاص الخطأ المحتمل.

بعد هذه الخطوات تحصلنا على عدد الواصفات (برامتر Paramètres)^[25] وهي موضح كما يلي:

الجدول (III-2): نتائج حساب الوصفات الجزيئية المكونة لقاعدة البيانات باستعمال Dragon.

عدد الوصفات بعد الحساب	عدد الوصفات في الكتلة	الكتل
0	41	المقاطع الجانبية للجزيئات Randic
0	74	الوصفات الهندسية
0	150	وصفات RDF
0	160	وصفات 3D-MoRSE
0	99	وصفات WHIM
0	197	وصفات GETAWAY
0	14	الوصفات الشحنية
0	735	العدد الإجمالي

3.III. الخطوات التجريبية للعمل:



الشكل (1-III): الخطوات التجريبية للعمل.

III.3.1. شرح الخطوات التجريبية للعمل:

في هذا الجزء من العمل قمنا بتطبيق برمجيات مختلفة، وقد تضمن هذا العمل مجموعة من الخطوات وهي مدونة كالتالي:

1. تم اختيار مجموعة من المركبات العضوية مع القيم التجريبية للخاصية الفيزيوكيميائية والمتمثلة في معامل التوزيع (الأكتانول - الماء) LogK_{ow} من المراجع العلمية وهي التي تمثل قاعدة البيانات وتتكون من 993 مركب ،

2. رسم البنية الجزيئية لهذه المركبات على برنامج ChemDraw Ultra ويحفظ الملف على صيغة .mol (* .MDL Molfile).

- وقد أجري ببرنامج (ChemDraw Ultra 2008).

3. فتح الملف في برنامج HyperChem وهذا قصد نمذجة المركبات باستعمال طرق الحساب النظرية (الميكانيكا الجزيئية MM^+ والطرق نصف التجريبية AM1 و PM3)، ويحفظ الملف على صيغة (* .HIN).

- وقد أجري ببرنامج (HyperChem 8.0).

4. حساب الواصفات الجزيئية باستعمال برنامج Dragon وذلك باختيار 7 كتل ثلاثية البعد (3D).

- وقد أجريت الحسابات ببرنامج (Dragon 5.4).

5. إنشاء جداول تجمع بين الخاصية المختارة LogK_{ow} والواصفات الجزيئية التي تم حسابها وذلك على صفحات برنامج Excel وهذا لكل كتلة وبعد ذلك تم إنشاء جدول يجمع بين كل الواصفات الجزيئية للكتل والخاصية المختارة LogK_{ow} .

6. تطبيق تقنية QSPR وهذا للحساب الإحصائي بطريقة الانحدار الخطي المتعدد (MLR) باستخدام البرامج الإحصائي Xistat وهذا لدراسة درجة الارتباط وحساب قيم معاملات التحديد (R^2) التي تدلنا على درجة الألفة بين القيم التجريبية ($\text{LogK}_{\text{ow}} (\text{exp})$) والقيم المحسوبة (النظرية) $\text{LogK}_{\text{ow}} (\text{préd})$ بالإضافة إلى نوعية وشكل معادلة النموذج. إن معامل التحديد يأخذ القيم بين 0 و 1، حيث كلما اقترب معامل التحديد من 1، كلما كان النموذج أفضل .

7. تم تصميم معادلتني للنموذج الرياضي الأولى خاصة بكل كتلة على حد, أما الثانية خاصة بجميع الكتل معا ,حيث كل معادلة تربط بين الخاصية المختارة ومجموعة الوصفات الجزيئية المرتبطة بها.
8. يتم اختيار أفضل نموذج رياضي للانحدار الخطى على أساس أعلى قيمة لمعامل التحديد (R^2), وكذلك على أقل عدد من الوصفات الجزيئية المتشكلة لمعادلة النموذج.

الفصل الرابع

النتائج والمناقشة

IV.النتائج:

مقدمة:

في هذا العمل قمنا بإجراء مرحلتين لدراسة 7 كتل من الواصفات الجزيئية ثلاثية البعد (3D) وذلك باستعمال 992 مركب كيميائي، بتطبيق تقنية QSPR وهذا بالحساب الإحصائي بطريقة الانحدار الخطي المتعدد (MLR) خطوة خطوة (step Wise) في البرنامج الإحصائي (XIstat (V.2015).

لقد أستعمل 992 مركب كيميائي حيث 795 التي تمثل (70%) لبناء النموذج. و197 مركب للتحقق الخارجي وتمثل (30% Validation)، ولقياس صلاحية والتنبؤ باستعمال هذا النموذج يمكن حساب الخاصية **Logkow** لمركبات مجهولة لم تدخل في مجموعة المعطيات.

1.IV. نتائج الحسابات: النتائج مدونة على الجدول التالي:

الجدول (1-IV): نتائج الكتل المدروسة كل على حد ومجموعة.

F	S ²	R ²	الترجمة إلى العربية	اسم الكتلة بالانجليزية	البعد
60.136	1.437	0.207	المقاطع الجانبية للجزيئات	Randic molecular profiles	3D
40.025	1.033	0.578	الواصفات الهندسية	Geometrical descriptors	
75.635	0.834	0.722	واصفات RDF	RDF descriptors	
67.970	0.742	0.783	واصفات 3D-MoRSE	3D-MoRSE descriptors	
54.760	0.937	0.652	واصفات GETAWAY	GETAWAY descriptors	
41.516	1.276	0.327	الواصفات الشحنة	Charge descriptors	
43.354	1.10	0.587	واصفات WHIM	WHIM descriptors	
73.298	0.535	0.894	الكتل مجمعة	The blocks are grouped	

نلاحظ من الجدول السابق أن الكتل مجمعة لها أكثر قيمة لمعامل التحديد $R^2=0.89$ وأقل قيمة لمعامل الانحراف المعياري $S^2=0.535$ وقيمة الإختبار الإحصائي $(F)=73.298$.

بناء على هذه النتيجة يمكن القول بأن النموذج له القدرة على التنبؤ بخاصية **Logkow** لمركبات جديدة .

1.IV. معادلة النموذج:

يمثل نموذج الانحدار الخطي المتعدد العلاقة بين متغير كمي وهو المتغير التابع وعدة متغيرات كمية أخرى وهي المتغيرات المستقلة. بافتراض وجود متغير تابع ومجموعة متغيرات مستقلة فإنه يمكن صياغة الشكل العام لمعادلة النموذج على النحو التالي:

$$\text{Logk}_{ow} = a + \sum (b_j \times x_j)$$

حيث:

a: قيمة الثابت.

b_j: قيمة المعامل.

x_j: الوصفة.

$$\begin{aligned} \text{LogKow} = & -a + b_1 \times \text{PCWTe} - b_2 \times \text{Mor02u} - b_3 \times \text{Mor23u} - b_4 \times \text{Mor26u} + b_5 \times \text{Mor27u} - \\ & b_6 \times \text{Mor28u} - b_7 \times \text{Mor02m} + b_8 \times \text{Mor08m} - b_9 \times \text{Mor09m} + b_{10} \times \text{Mor17m} - b_{11} \times \text{Mor24m} \\ & - b_{12} \times \text{Mor27m} + b_{13} \times \text{Mor06v} + b_{14} \times \text{Mor16v} - b_{15} \times \text{Mor18v} - b_{16} \times \text{Mor22v} + b_{17} \times \text{Mor32v} + \\ & b_{18} \times \text{Mor02e} - b_{19} \times \text{Mor07e} + b_{20} \times \text{Mor12e} + b_{21} \times \text{Mor20e} + b_{22} \times \text{Mor26e} - b_{23} \times \text{Mor29e} + \\ & b_{24} \times \text{Mor02p} - b_{25} \times \text{Mor04p} - b_{26} \times \text{Mor12p} + b_{27} \times \text{Mor14p} + b_{28} \times \text{Mor25p} + b_{29} \times \text{Mor29p} - \\ & b_{30} \times \text{Mor32p} + b_{31} \times \text{SPAN} - b_{32} \times \text{AROM} + b_{33} \times \text{QXXm} - b_{34} \times \text{G(N..O)} + b_{35} \times \text{G(N..S)} - \\ & b_{36} \times \text{G(O..S)} + b_{37} \times \text{G(O..F)} + b_{38} \times \text{G(Cl..Cl)} - b_{39} \times \text{H6m} + b_{40} \times \text{HATS6m} - b_{41} \times \text{H2v} + b_{42} \times \text{H3v} + \\ & b_{43} \times \text{H0e} + b_{44} \times \text{H2p} - b_{45} \times \text{HATS5p} + b_{46} \times \text{R3u} + b_{47} \times \text{R2m} - b_{48} \times \text{R4m} + b_{49} \times \text{R4m}^+ + \\ & b_{50} \times \text{R5v}^+ + b_{51} \times \text{R8v}^+ + b_{52} \times \text{R2e} - b_{53} \times \text{R3e} + b_{54} \times \text{R5e}^+ - b_{55} \times \text{RTe}^+ - b_{56} \times \text{R4p}^+ + b_{57} \times \text{RTp}^+ - \\ & b_{58} \times \text{SP05} - b_{59} \times \text{RDF020u} - b_{60} \times \text{RDF105u} + b_{61} \times \text{RDF125u} - b_{62} \times \text{RDF150u} - b_{63} \times \text{RDF010m} - \\ & b_{64} \times \text{RDF030m} + b_{65} \times \text{RDF020v} + b_{66} \times \text{RDF020e} - b_{67} \times \text{RDF030e} + b_{68} \times \text{RDF045e} - \\ & b_{69} \times \text{RDF020p} - b_{70} \times \text{RDF060p} - b_{71} \times \text{RDF065p}. \end{aligned}$$

الجدول (2-IV): معاملات التعديل الخاصة بالكتل مجمعة.

Observations	Somme des poids		DDL	R ²	R ² ajusté	MCE
692.000	692.000		620.000	0.894	0.881	0.286
RMCE	MAPE	DW	Cp	AIC	SBC	PC
0.535	59.043	1.786	549.490	-797.100	470.250	0.131

الجدول(3-IV): تحليل الفروق الخاص بالكتل مجمعة.

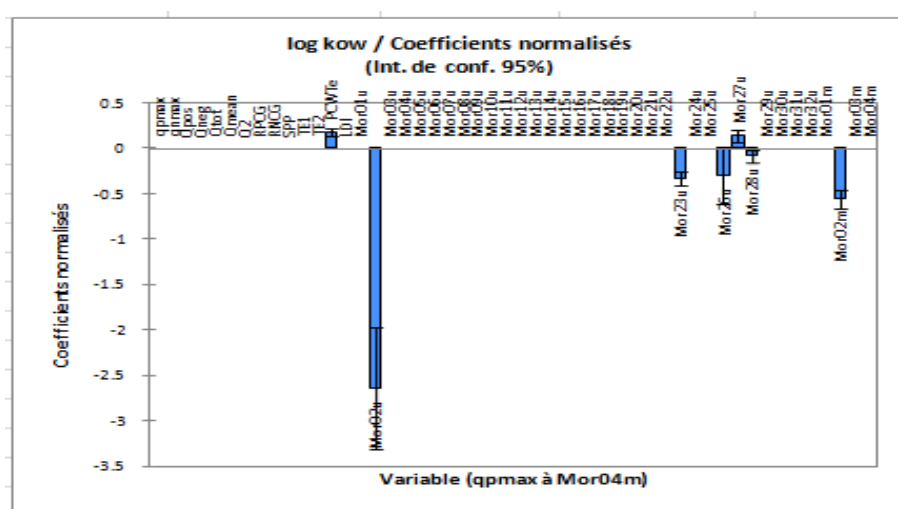
Pr > F	F	متوسط المربعات	مجموع المربعات	DDL	مصدر النموذج
< 0.0001	73.298	20.998	1490.861	71	النموذج
		0.286	177.614	620	الخطأ
			1668.476	691	مجموع التصحيح

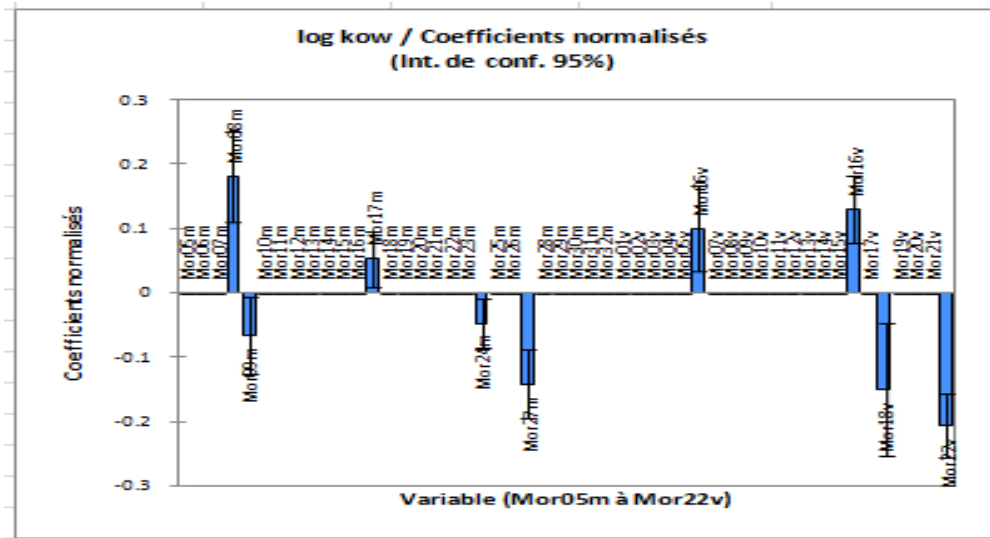
الجدول(4-IV): ضبط معاملات النموذج الخاص بالكتل مجمعة.

الخطأ النوعي	القيمة	المعامل	الواصف	الخطأ النوعي	القيمة	المعامل	الواصف
0.019	-0.108	b ₃₆	G(O..S)	0.270	-1.699	a	ثابت
0.009	0.023	b ₃₇	G(O..F)	0.010	0.076	b ₁	PCWTe
0.008	0.057	b ₃₈	G(Cl..Cl)	0.075	-0.583	b ₂	Mor02u
0.325	-1.126	b ₃₉	H6m	0.178	-1.562	b ₃	Mor23u
0.194	0.826	b ₄₀	HATS6m	1.193	-2.179	b ₄	Mor26u
1.600	-6.150	b ₄₁	H2v	0.286	1.220	b ₅	Mor27u
0.613	1.895	b ₄₂	H3v	0.239	-0.587	b ₆	Mor28u
0.173	0.799	b ₄₃	H0e	0.022	-0.237	b ₇	Mor02m
1.184	5.164	b ₄₄	H2p	0.079	0.399	b ₈	Mor08m
0.576	-3.943	b ₄₅	HATS5p	0.096	-0.210	b ₉	Mor09m
0.371	2.076	b ₄₆	R3u	0.152	0.365	b ₁₀	Mor17m
0.049	0.156	b ₄₇	R2m	0.200	-0.474	b ₁₁	Mor24m
0.164	-1.170	b ₄₈	R4m	0.205	-1.075	b ₁₂	Mor27m
0.486	2.615	b ₄₉	R4m+	0.128	0.381	b ₁₃	Mor06v
4.504	12.657	b ₅₀	R5v+	0.321	1.562	b ₁₄	Mor16v
7.675	42.353	b ₅₁	R8v	0.353	-1.008	b ₁₅	Mor18v
0.199	-0.687	b ₅₂	R2e	0.393	-3.166	b ₁₆	Mor22v
0.376	-2.288	b ₅₃	R3e	2.113	7.776	b ₁₇	Mor32v
1.199	4.635	b ₅₄	R5e+	0.072	0.496	b ₁₈	Mor02e
0.411	-0.891	b ₅₅	Rte+	0.071	-0.428	b ₁₉	Mor07e

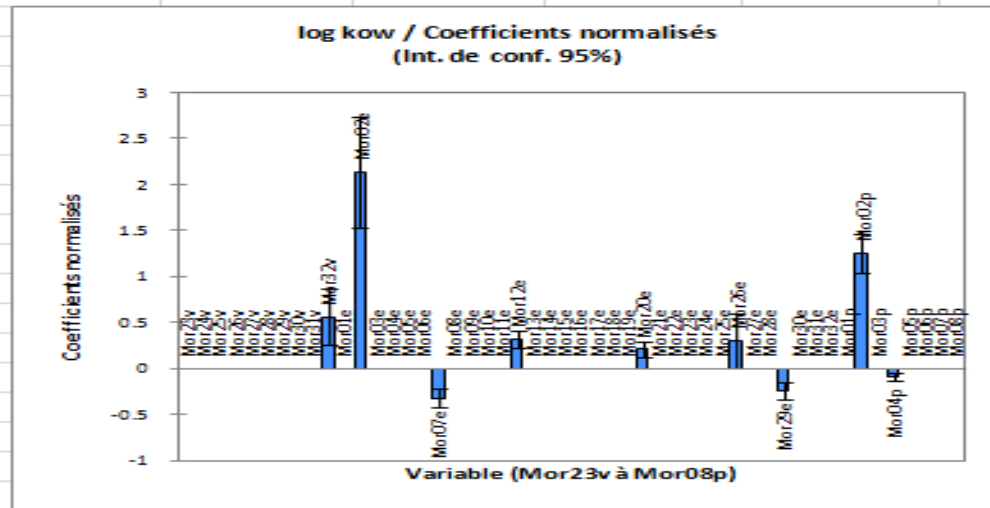
1.809	-7.141	b ₅₆	R4p+	0.156	1.094	b ₂₀	Mor12e
0.816	2.374	b ₅₇	RTp+	0.151	0.778	b ₂₁	Mor20e
0.061	-0.233	b ₅₈	SP05	1.121	2.272	b ₂₂	Mor26e
0.172	-1.056	b ₅₉	RDF020u	0.343	-1.647	b ₂₃	Mor29e
0.033	-0.119	b ₆₀	RDF105u	0.046	0.516	b ₂₄	Mor02p
0.080	0.341	b ₆₁	RDF125u	0.096	-0.350	b ₂₅	Mor04p
0.247	-1.230	b ₆₂	RDF150u	0.289	-1.451	b ₂₆	Mor12p
0.578	-1.463	b ₆₃	RDF010m	0.212	0.670	b ₂₇	Mor14p
0.014	-0.051	b ₆₄	RDF030m	0.369	1.561	b ₂₈	Mor25p
0.514	1.835	b ₆₅	RDF020v	0.666	1.656	b ₂₉	Mor29p
0.136	0.546	b ₆₆	RDF020e	1.870	-5.732	b ₃₀	Mor32p
0.016	-0.039	b ₆₇	RDF030e	0.077	0.349	b ₃₁	SPAN
0.013	0.095	b ₆₈	RDF045e	0.100	-0.600	b ₃₂	AROM
0.307	-0.910	b ₆₉	RDF020p	0.003	0.010	b ₃₃	QXXm
0.034	-0.091	b ₇₀	RDF060p	0.005	-0.013	b ₃₄	G(N..O)
0.027	-0.161	b ₇₁	RDF065p	0.032	0.110	b ₃₅	G(N..S)

الشكل (1-IV): المنحنيات التي تبين الواصفات الجزيئية والقيم العددية لمعاملاتها. يشار إلى أن الأشكال من (1 إلى 12) هي متكاملة.

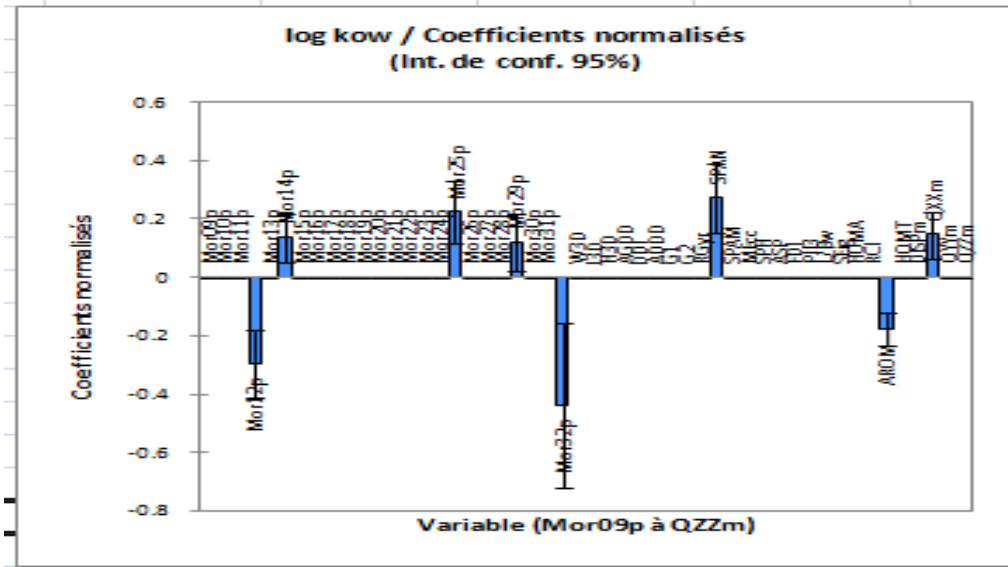




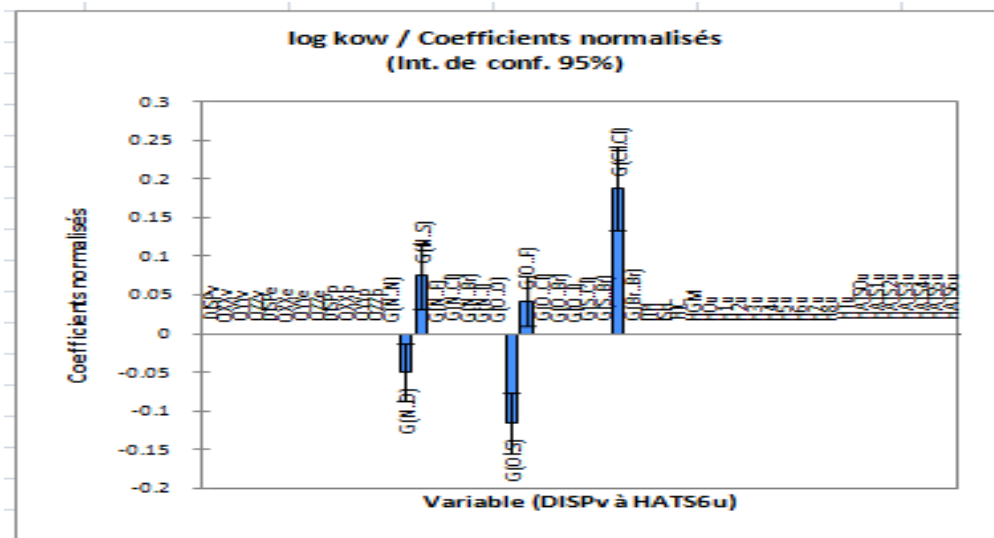
2



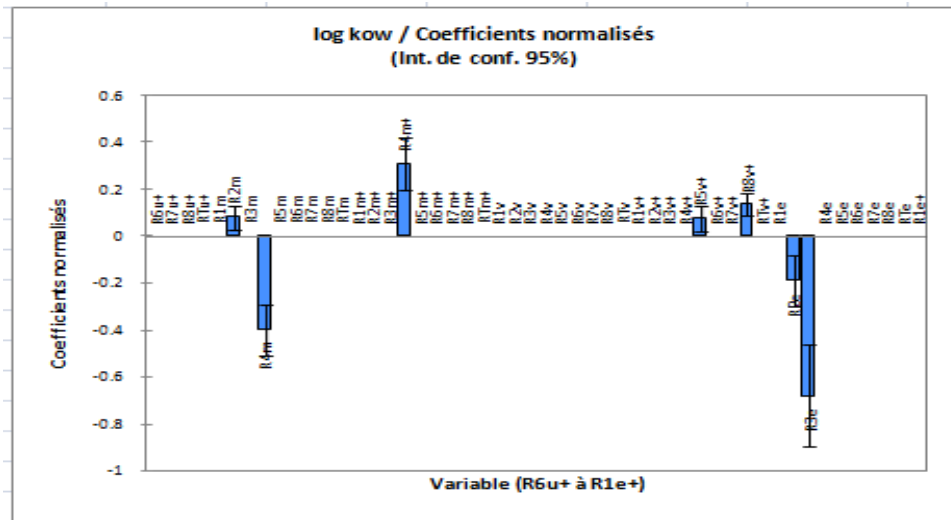
3



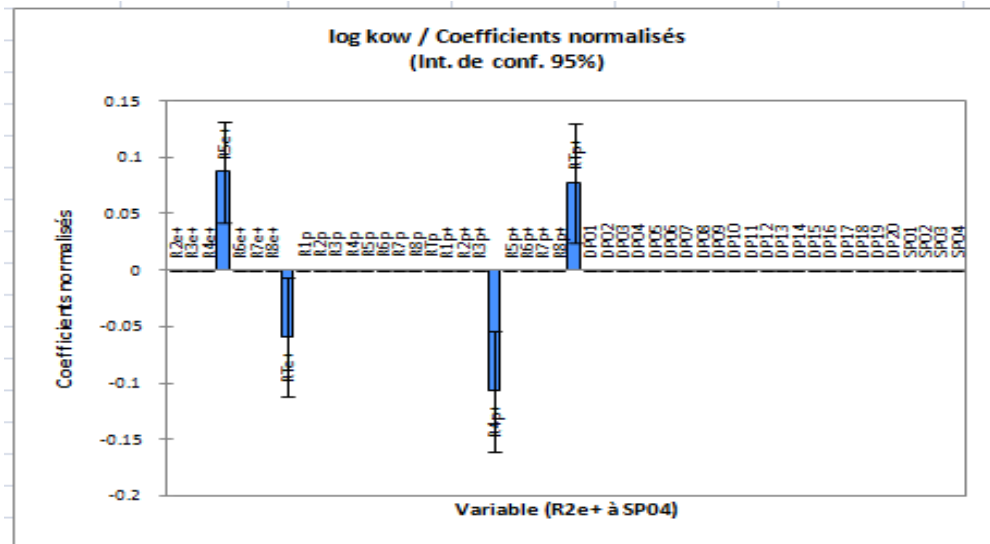
4



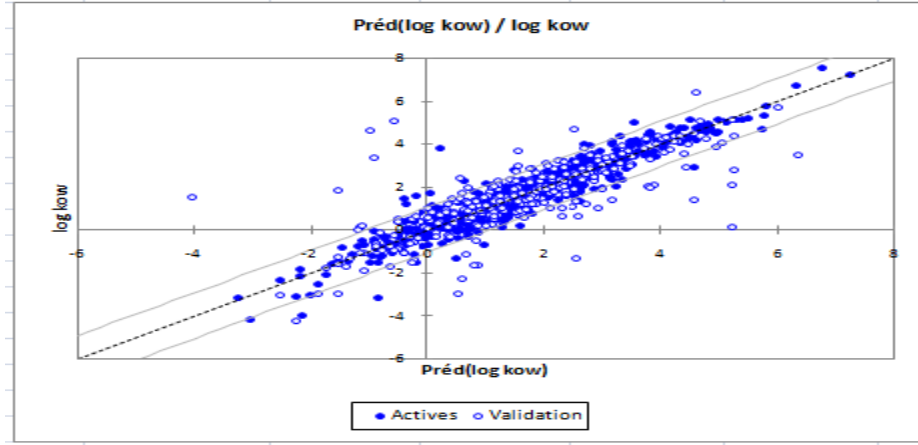
5



8



9



الشكل (2-IV): بيان يوضح تغيرات LogK_{ow} بدلالة $\text{Préd}(\text{LogK}_{\text{ow}})$

الخاصة بالكتل مجمعة.

2.IV. تحليل نتائج الكتل مجمعة:

من خلال نتائج الجدول (1-IV) نلاحظ أن معامل التحديد (R^2) يعطينا فكرة عن النسبة المئوية لتغير المتغير المنمذج موضح بالمتغيرات التفسيرية. حيث كلما اقترب معامل التحديد من 1 كلما كان النموذج أفضل.

في هذه الحالة لدينا $R^2 = 0.894$ و $R^2 \text{ ajusté} = 0.881$ ، وبالتالي فإن النسبة المئوية لتغير المتغير المنمذج = 89.4% .

3.IV. تحليل باقي الكتل (Residues) مجمعة:

$$\text{Residues} = 1 - R^2 = 1 - 0.894 = 0.106 = 100 \times 0.106 = 10.6\%$$

هذا يعني أن الباقي يوجد خارج مجال الثقة، وكل قيمة خارج هذا المجال غير موثوق فيها. المتغيرات التي تسمح بتفسير 89.4% من تغير LogK_{ow} تفسر النسبة المئوية الباقية وهي 10.6%

4.IV. مناقشة نتائج الكتل مجمعة:

دراسة ارتباط الوصفات الجزيئية مع بعضها البعض على شكل أزواج وذلك بالمقارنة بين أقصى وأدنى قيمة لها.

- الواصفات الجزيئية الأقل ارتباطا:

$$Q_{neg} = -0.850 \cdot (qp_{max})$$

$$HATS3u = -0.239 \cdot (HATS6u)$$

$$H4v = -0.034 \cdot (HTm)$$

الجدول (5-IV): الواصفة ورمزها للواصفات الجزيئية الأقل ارتباطا (بالكتل مجمعة).

رمز الواصفة	الواصفة
qpmax	Maximum positive charge
Qneg	total negative charge
HATS6u	leverage weighted outocorrelation of lag 6/unweighted
HATS3u	leverage weighted outocorrelation of lag 3/unweighted
HTm	H total index/weighted by atomic masses
H4v	H outocorrelation of lag 4/unweighted by atomic van der waals volumes

هذا يبين أنها لا تعطي نفس المعلومة المتعلقة بالخاصية المدروسة (قيد المحاكاة) $\log k_{ow}$ (كل واصفة لها معلومة تختلف عن معلومة الواصفة الأخرى).

- الواصفات الجزيئية الأكثر ارتباطا:

$$Mor16v = 0.111 \cdot (Mor6v)$$

$$Q_{xxm} = 0.186 \cdot (AROM)$$

$$R_{Te^+} = 0.236 \cdot (RTP^+)$$

الجدول (6-IV): الواصفة ورمزها للواصفات الجزيئية الأكثر ارتباطا (بالكتل مجمعة).

رمز الواصفة	الواصفة
AROM	Aromaticity index
Mor6v	3D-MORSE-signal 6/unweighted by atomic van der waals volumes
Mor16v	3D-MORSE-signal 16/unweighted by atomic van der waals volumes
Qxxm	Qxx COMMA2 value/weighted by atomic masses
R _{Te⁺}	R maximal index/weighted by atomic sanderson electronegativities
RTP ⁺	R maximal index/weighted by atomic polarizabilities

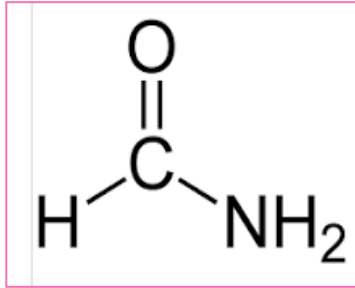
الجزء التطبيقي/ الفصل الرابع: النتائج والمناقشة

الرسم البياني $(\text{p}é\text{rd}(\log k_{\text{ow}})/\log k_{\text{ow}})$ يبين لنا أن هناك نقاط موجودة داخل المجال (مجال الثقة) الذي يمثل 95% وهذه النقاط تمثل مركبات مدروسة من قاعدة البيانات. ووفقا للتحليل الإحصائي للقيم والمنحى البياني $(\log k_{\text{ow}} \text{ p}é\text{rd} , \log k_{\text{ow}})$ ونتائج المقارنة بين الكتل يمكن ملاحظة أنه: يوجد نقاط خارج مجال الثقة ، هذه النقاط تمثل القيم الشاذة لإحداثيات مركبات غير مرتبطة).

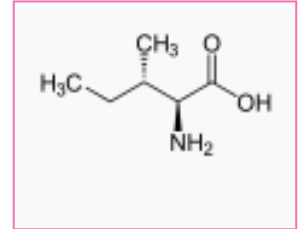
بالمقابل وجدنا 47 جزيء غير مرتبط أي خارج مجال الثقة لأن لديها هياكل كيميائية غير منتظمة. أما سبب عدم الانتظام يرجع لوجود الهالوجينات مثل (S)، (Cl)، (Br) والحلقات البنزينية أو المجموعات القادرة على إعطاء زوج الإلكترونات مثل (NH₂)، (Na).

أما جودة النموذج المتحصل عليها فتعتمد على عدد المتغيرات (الواصفات الجزيئية) وقيمة معامل التحديد R^2 . تتغير درجة تقدير (قيمة R^2) ل $\log k_{\text{ow}}$ وفقا لنوع الواصفات لكل كتلة.

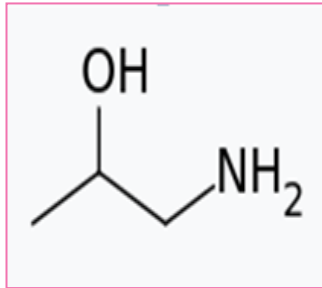
من أجل تحسين النتائج يجب: استبعاد الجزيئات غير المرتبطة مثلا المركبات التالية:



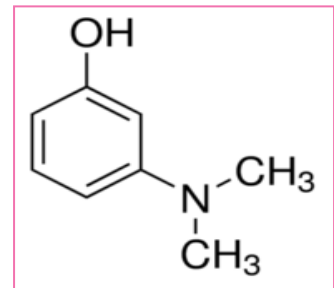
Formamide



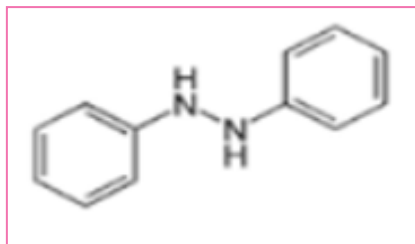
L-Isoleucine



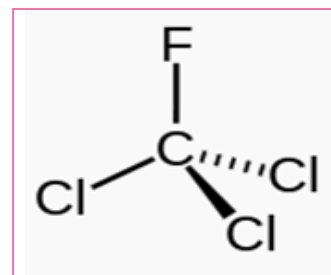
1-amino,2-Propanol



3-(dimethylamino),Phenol



2-diphenyl-1-Hydrazine



trichlorofluoro-Methane

الجدول (7-IV): أمثلة للمركبات التي وجدت خارج مجال الثقة.

رقم	LogKow	الإسم	SMILES
409	3.16	(isothiocyanatomethyl)-Benzene	<chem>N(=C=S)Cc1ccccc1</chem>
448	2.99	bromo-Benzene	<chem>c1ccccc1Br</chem>
277	4.61	4-chlorobiphenyl	<chem>Clc1ccc(cc1)-c2ccccc2</chem>
598	3.82	diphenyl-Diazene	<chem>N(=Nc1ccccc1)c2ccccc2</chem>
669	-1.51	Formamide	<chem>O=CN</chem>
687	2.94	2-diphenyl-1-Hydrazine	<chem>N(Nc1ccccc1)c2ccccc2</chem>
704	-1.7	L-Isoleucine	<chem>O=C(O)C(N)C(CC)C</chem>
750	2.53	trichlorofluoro-Methane	<chem>FC(Cl)(Cl)Cl</chem>
730	-0.57	N-methyl-N-nitroso-Methanamine	<chem>O=NN(C)C</chem>
859	1.56	3-(dimethylamino)-Phenol	<chem>Oc1ccc(N(C)C)cc1</chem>
214	-0.96	1-amino-2-Propanol	<chem>OC(CN)C</chem>
286	-0.84	4-i-Butylsemicarbazide	<chem>NNC(=O)NCC(C)C</chem>

خاتمة عامة

توصلنا في هذا العمل إلى تحقيق العلاقة الرياضية التي تربط بين الخاصية الفيزيوكيميائية معامل التوزيع LogK_{ow} والواصفات الجزيئية، وهي (الواصفات الجزيئية) $f = \text{الخاصية}$ بالشكل التالي:

$$\text{Logkow} = a + \sum (b_j \times x_j)$$

ومن خلال الحسابات و النتائج تبين لنا أن أفضل نموذج

هو خاص بالكتل مجمعة والذي تميز ب: $R^2=0.894, S^2=0.535, F=73.298$

ووفقا لهذه الدراسة وهذه النتائج يمكننا استنتاج ما يلي:

✓ تعتبر الدراسة بتقنية ال QSAR/QSPR باستعمال طريقة (MLR) حلا موثوقا لأنها تسمح

للكيميائيين بإيجاد قيم ل LogK_{ow} لمركبات مجهولة خاصة المركبات الكيميائية الخطيرة ذات السمية العالية التي يصعب التعامل معها مخبريا.

✓ درجة تقدير الخصائص تتغير وفقا لجودة واصفات كل كتلة.

✓ أن القدرة التنبؤية تعتمد على نوع وعدد المركبات الكيميائية المكونة لقاعدة البيانات.

ولتحسين النتائج يوصى ما يلي:

✓ استبعاد المركبات الغير مرتبطة.

✓ جمع أكثر عدد من الجزيئات لتوسيع قاعدة البيانات لإعطاء نتائج أفضل.

✓ يجب إجراء تحسين (Optimisation) كامل للمركبات للوصول إلى أقصى درجة استقرار.

المراجع

المراجع باللغات الأجنبية:

- [1] A.Crum Brown and T. R. Fraser, Trans. Roy. Soc., Edinburgh, 25 (1868-1869) 151-203.
- [2] A. Leo, C. Hanch, D. Elkins, Chem. Rev., 71 (1971)525- 616.
- [3] United Nations Publications, 2010 - Globally Harmonized System of Classification and Labelling of Chemicals (Ghs), 566p.
- [4] T. Clark, ‘‘Handbook of Computational Chemistry’’, Edition, Wiley, London,1985.
- [5] Samar Bakoben & Ahlam Houda, 16/07/2011, MEGBI Training Course MolecularModelling, Institute for Genetic Engineering, Ecology and Health (IGEEH) Karlsruhe, Germany.120p.
- [6] Gerd B. Rocha, Ricardo O. Freire, Alfredo M. Simas, and James J. P. Stewart,2003, RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I, Departamento de Química Fundamental - CCEN - UFPE, 50590-470 - Recife PE, Brazil. Stewart Computational Chemistry, 15210 Paddington, Colorado Springs CO 80921-2512, USA.172p.
- [12] Martin G., Laffort P., Odeurs et désodorisations dans l'environnement, Lavoisier, Tec&Doc, Paris. 1991.
- [13] M. Berthelot and E. Jungfleisch, Ann. Chim.Phys., (4) 26 (1872) 396-407.
- [14] Mr. Lanez Touhami, Cours de technique de purification en synthèse organique, Université de Kasdi Merbah Ouargla.
- [15] Prof. Dr. Stefan Bienz, Unknown 2015, Short Manual to the Chemical Drawing Program ChemDraw®, university of zurich IfC Department of Chemistry,22p.
- [16] REGADI Dahmane, MASTER, UNIVERSITE KASDI MARBAH OUARGLA /2013.
- [18] Statistics for Managers, Levine et al. Prentice Hall,1999.
- [19] Belaidi, Thèse de doctorat, Université de Batna,2002.
- [20] CambridgeSoft Corporation 100 CambridgePark Drive Cambridge, MA 02140 USA,2002, ChemOffice, ChemDraw, Chem3D, ChemFinder et ChemInfo sont des marques de commerce de CambridgeSoft ©.28

[21] [http: \\internet\www.chemcad.com.htm](http://internet/www.chemcad.com.htm).

[23] Hakan Kayi, PARAMETERIZATION OF THE AM1SEMIEMPIRICAL MOLECULAR ORBITAL METHOD FOR THE FIRST-ROW TRANSITION METALS AND OTHER ELEMENTS, Universität Erlangen-Nürnberg , Türkei , 2009 ,pag 4.

[24] HyperChem (Molecular Modelling System) Hypercube, Inc. , 1115 NW 4th Street, Gainesville, FL 32601 ; USA, 2000.

[26] Logiciel XLSTAT version 7.0, 2006, Addinsoft 40 rue Damrémont 75018 PARIS, 74p.

المراجع باللغة العربية:

[7] الانحدار الخطي المتعدد، أ.د. ثائر داود سلمان، فرع العلوم النظرية كلية التربية الرياضية/جامعة بغداد.

[8] تطبيقات إحصائية باستخدام SPSS, د. وليد محمد عفيفي محمد، جامعة أم القرى /معهد البحوث العلمية وإحياء التراث الإسلامي.

[9] مبادئ الإحصاء 102 كمى، الوحدة الخامسة: الارتباط الخطى البسيط، جامعة الملك خالد.

[10] علي مكيد، الاقتصاد القياسي دروس ومسائل محلولة، د.م.ج، الجزائر، 2007، ص43-44.

[11] علي لزعر، الإحصاء وتوفيق المنحيات، د.م.ج، الجزائر، 2002 ، ص97.

[15] تحليل الانحدار الخطي البسيط، أ.د/محمد راتول.

[16] أساسيات الكيمياء العضوية، الإدارة العامة لتصميم وتطوير المناهج/المملكة العربية السعودية.

[17] صياغة نموذج في إحصاء تقدير الإنتاج، م.م منى خضر اسطيفان، مجلة كلية بغداد للعلوم الاقتصادية جامعة العبد/العدد/2010 .

[22] زواري أحمد رشيدة، مذكرة ماجستير، دراسة البرمترات المحبة للماء والكارهة على السطح المشترك (محب للماء/ كاره للماء للمضادات الحيوية الماكروليدية ذات 16 ذرة)، جامعة ورقلة، 2005.

- [24] أبو سدرة، فتحي صالح، 1999 كتاب الإحصاء والاقتصاد القياسي، منشورات المركز القومي للدراسات والبحوث العلمية، دار الكتب الوطنية، بنغازي- ليبيا.
- [25] تخة مبركة، مساهمة في دراسة العلاقة) بنية كيميائية - رائحة (باستعمال طريقة آلات الشعاع الساند(SVM) (تطبيق على عائلة البيرازين)،مذكرة ماجستير، جامعة ورقلة/2008،ص28-32.
- [27] الكيمياء الحديثة، د. هاني كريدين، 1994 ص 391.
- [28] أسس الكيمياء العضوية، أ.أحمد الشريف على الدرسي، 2008 /دار الكتب الوطنية بنغازي-ليبيا.
- [29] الكيمياء العضوية، بيت معلمي الكيمياء بالعاصمة المقدسة 1421هـ.

الملاحق

مفهوم بعض الواصفات الجزيئية

1. مؤشرات الشحن الطوبولوجية:

هي واصفات مشتقة من مصفوفة غير متناظرة CT تعرف بالشكل التالي:

$$CT_{ij} = \begin{cases} \delta_i & \text{if } i = j \\ m_{ij} - m_{ji} & \text{if } i \neq j \end{cases}$$

δ_i : هي درجة القمة لكل ذرة، وبمعنى آخر عدد الذرات المرتبطة بالهيدروجين المتفرع.

m_{ij} : هي عناصر المصفوفة، نحصل بضرب مصفوفة الجوار بمصفوفة المسافة المربعة المتبادلة المقاطع القطرية.

لأجل كل مسار ذي الامتداد K يوجد ثابت أو واصف طوبولوجي GGLK يكون متعلقا بنصف مجموع الشحن الطوبولوجية، ويمثل عدد النهايات للحدود. وله رتبة من 0 إلى 10.

2. الواصفات الهندسية:

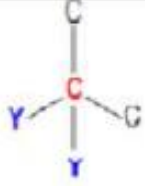
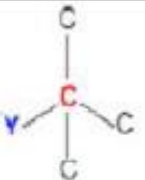
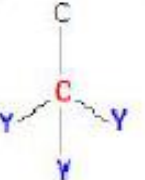



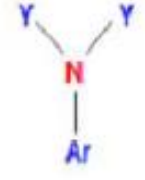
تعرف البنية ثلاثية الأبعاد 3D للجزيئة على وجه العموم، تتركز على الهندسة الجزيئية، كما تمثل طرق الكيمياء المعلوماتية. تبين هذه الواصفات التموضع النسبي للذرات في الفضاء، استنادا على الإحداثيات الكارتيزية، ووصف بنيات التشكل للجزيئات المتماثلة.

L'asphericity: ASP ثابت يصف كروية الجزيئية يكون متباين القياس، حيث يقيس انحراف الشكل الكروي للجزيئة.

- يتغير ASP من 0 في الجزيئات عالية الكروية إلى 1 في الجزيئات الخطية.
- ASP = 0.25 من أجل جزيئة على شكل سيجارة.
- ASP = 1 من أجل جزيئة على شكل قرص.

3. مجاميع المجموعات الوظيفية:

هي واصفات جزيئية سهلة وبسيطة تعرف عدد المجموعات الوظيفية الخاصة لكل جزيئة، تحسب عدد الارتباطات الجزيئية للمركبات والذرات نعرف منها ما يلي:

$Y = (H, X, OH, NH_2, \dots)$ أو ذرة متغايرة و $C(sp^3)$		nC_s
$Y = (H, X, OH, NH_2, \dots)$ أو ذرة متغايرة و $C(sp^3)$		nC_t
$Y = (H, X, OH, NH_2, \dots)$ أو ذرة متغايرة و $C(sp^3)$		nC_p
عدد حلقات البترين غير المستبدل $C(sp^2)$		NC_{bH}
عدد الكبريتيدات		$nRSR$
عدد الاثيرات العطرية $Y = Ar \neq C=O, C \neq N$		$nArOR$
عدد الامينات الثلاثية العطرية $Y = Ar, Al \neq C=O$		$nAr NR_2$

4. مؤشرات المعلومة:

هي واصفات تبرز محتوى المعلومات حول الجزئية، لها عدة معايير مختلفة تستعمل لتعريف الفروع المتكافئة في الجزئية كتعريف كيميائي لمجالات الالتصاق في الفضاء، وتبين طوبولوجية الجزيء وتناظر كل العناصر ذات العدد n .

$$I = n \log_2 n - \sum_{g=1}^G n_g \log_2 n_g$$

G: عدد الأصناف المكافئة المختلفة.

n_g: عدد العناصر للصنف g Th.

تكافؤ الصنف يعرف ببعض العلاقات في النظام، اللوغاريتم \log_2 مأخوذ في القاعدة 2 لقياس محتوى المعلومات الثنائية، ومن مؤشرات المعلومة ما يلي:

$$ISIZ = nAT \cdot \log_2 nAT$$

$$CICK = \log_2 nAT - ICK$$

IVDE محتوى المعلومات المتوسط على مساواة درجة القمة.

Xinde الثابت x ل Balaban.

5. واصفات RDE (Radial Distribution Function):

تعتمد أساسا على دالة التوزيع الشعاعي التي تبين احتمال التوزيع لتوجد ذرة في حجم كروي قطره R الشكل العام لدالة التوزيع الشعاعي يكون:

$$RDF \quad R_w = f \cdot \sum_{j=1}^{nAT-1} \sum_{j=i+1}^{nAT} w_i \cdot w_j \cdot e^{-\beta(R-f_{ij})^2}$$

f: معامل القياس.

W: الخصائص المميزة للذرات.

r_{ij} : المسافة ما بين الذرات تعطى بالأنغستروم.

nAT: عدد الذرات في الجزيئية.

b: يمثل احتمال توزيع المسافات الفردية، ويمكن أن يفسر عامل درجة الحرارة الذي يبين حركة الجزيئات.

هناك خمسة أنواع من الواصفات RDF:

1. (u) الحالة غير المرجحة.

2. (m) الكتلة الذرية.

3. (v) حجم فاندرفالس.

4. (e) كهروسالبية ساندرسن.

5. (p) قابلية الاستقطاب الذري.

المخلص

توصلنا في هذا العمل المنجز إلى بناء علاقة رياضية تربط بين البنية الكيميائية المعبر عنها بالواصفات الجزيئية (3D)، والخاصية الفيزيوكيميائية المتمثلة في معامل التوزيع LogK_{ow} ، وذلك للتعرف بقيم LogK_{ow} لمركبات جديدة. شملت هذه الدراسة 7 كتل من الواصفات الجزيئية (3D)، حيث تم فيها حساب الكتل كل على حد وبعدة اتم حساب الكتل مجمعة. وتمت هذه الدراسة على 992 مركب كيميائي بتطبيق تقنية QSPR وبطريقة الانحدار الخطي المتعدد (MLR) وهذا باستخدام البرمجية الإحصائية Xistat. (النسخة 2013) وكانت أفضل النتائج للكتل مجمعة وهي: ($R^2 = 0.894$)، ($S^2 = 0.535$) و ($F = 73.298$).

الكلمات المفتاحية: النمذجة الجزيئية، LogK_{ow} ، الواصفات الجزيئية، تقنية QSPR، الانحدار الخطي المتعدد (MLR).

Résumé :

Dans ce travail, nous avons acquis un modèle mathématique reliant la structure chimique qui est représenté par descripteurs moléculaires (3D), et de la propriété physico-chimique du coefficient de partage LogK_{ow} , ce modèle nous permet de prédire les valeurs de LogK_{ow} à d'autres produits chimiques. Cette étude, a été appliqué sur 07 blocs de descripteurs moléculaires (3D), où il a été calculé chaque bloc individuel, suivi par le calcul de tous les blocs entre eux (regroupées). Cette étude a été appliquée à 992 composés chimiques en utilisant la technique QSPR avec régression linéaire multiple (MLR) par le logiciel XLSTAT (version 2013). Le meilleur résultat a été pour les blocs regroupées est ($R^2 = 0.894$), ($S^2 = 0.535$) et ($F = 73.298$).

Mots-clés: La modélisation moléculaire, LogK_{ow} , descripteurs moléculaires, technique QSPR, régression (MLR).

Abstract :

In this work we have acquired a mathematical model linking the chemical structure which is represented by molecular descriptors (3D), and the physicochemical property the partition coefficient LogK_{ow} , this model allows us to predict the LogK_{ow} values for other chemicals. This study, was applied to 07 molecular descriptors (3D) blocks, where each individual block was calculated, followed by the calculation of all the blocks between them (grouped together).. This study was applied to 992 chemical compounds using QSPR technique with multiple linear regression (MLR) by xlstat (2013 version). The best result was for grouped blocks ($R^2 = 0.894$), ($S^2 = 0.535$) and ($F = 73.298$).

Key words: Molecular modeling, LogK_{ow} , molecular descriptors, QSPR technique, regression (MLR)