رقم الترتيب:
رقم التسلسل:

الجمهورية الجزائرية الديمقراطية الشعبية

وزراء التعليم العالي والبحث العلمي

جامعة الشهيد حمّه لخضر - الوادي

كلية العلوم الدقيقة

مذكرة تخرج لنيل شهادة

الماستر أكاديمي

قسم: علوم المادة

شعبة: الكيمياء

تخصص: كيمياء عضوية تحليلية

من إعداد الطالبة: **بن عمارة سناء**

بعنوان:

نمذجة العلاقة (البنية الكيميائية - معامل التوزيع $LogK_{ow}$) باستعمال طريقة QSPR للمركبات العضوية

سلمت يوم: 2015/10/01

أعضاء لجنة المناقشة:

مازري راضية	أستاذة محاضرة (ب)	جامعة الشهيد حمّه لخضر بالوادي	رئيسة
سويعي بلقاسم	أستاذ مساعد (أ)	جامعة الشهيد حمّه لخضر بالوادي	مؤطرا
زواري أحمد رشيدة	أستاذة مساعدة (أ)	جامعة الشهيد حمّه لخضر بالوادي	مناقشة
بوثقرة سماح	أستاذة مساعدة (أ)	جامعة الشهيد حمّه لخضر بالوادي	مناقشة

الموسم الجامعي: 2015/2014

الإهداء

أهدي ثمرة جهودي هذه . . .

إلى ملاكي في الحياة . . إلى معنى الحب وإلى معنى الحنان والتفاني . . إلى بسمة الحياة وسر الوجود إلى من كان دعاؤها سر نجاحي وحنانها بلسم جراحي إلى أغلى الحبايب . . .**أمي الحبيبة**

إلى من كلله الله بالهيبة والوقار . . إلى من أحمل اسمه بكل افتخار . . أرجو من الله أن يمد في عمرك لترى ثمارا قد حان قطافها بعد طول انتظار وستبقى كلما تك نجوم أهتدي بها اليوم وفي الغد وإلى الأبد . . .**أبي العزيز**

إلى من شاركوني الدمعة والابتسامة . .إلى من كانوا متكئا أستند عليه وخير معيل ألجأ إليه . .إلى من لم يبخلوا علي بالنصح والإرشاد . . .**إخوتي وأخواتي**

رجاء . . هناء . . صفاء . . خنساء . . حسناء . .خليفة . . هيفاء . . محمد . . هيثم

إلى من تذوقت معهم أجمل اللحظات . . إلى من جعلهم الله أخوتي بالله . . وإلى من أحببتهم في بالله . . . **طلاب شعبة الكيمياء**

إلى من لم أعرفهم . . ولن يعرفوني

إلى من أتمنى أن تبقى صورهم . . في عيوني

سناء

شكر وعرفان

الحمد والشكر لله الذي لا معبود سواه، الذي رزقني العقل وحسن التوكل عليه سبحانه وتعالى وألهمني الصبر ووفقني إلى إتمام هذا العمل المتواضع، والصلاة السلام على نبيه محمد وآله وصحبه . . .

وقبل أن نمضي نقدم أسمى آيات الشكر والامتنان والتقدير والمحبة إلى الذين مهدوا لنا طريق العلم والمعرفة وسهروا من أجل ارتقائنا وفي مقدمتهم عميد كلية العلوم والتكنولوجيا ، رئيس قسم علوم المادة، مسؤول شعبة الكيمياء . . .

يسرني أن نتقدم بالشكر إلى السادة المؤطرين والمشرفين على دفعة الكيمياء لسنة 2015 الذين ساهموا في تكويننا ولم يبخلوا علينا بتوجهاتهم . . .

وعلى رأسهم الأستاذ الفاضل **سويعي بلقاسم** الذي أشرف على إنجاز هذا العمل، وكان نورا يضيء الظلمة التي تقف أحيانا في طريقي ، فجزاه الله عن الخير كله فله مني الشكر والتقدير كليهما على مجهوده ونصائحه القيمة وتوجيهاته السديدة التي كانت عونا لي في إتمام هذا العمل .

كما أشكر أعضاء اللجنة المناقشة لقبولها للمناقشة هذا العمل بداية برئيستها الدكتورة **مازري راضية** وكذلك المناقشين الأستاذة **زواري أحمد رشيدة** والأستاذة **بوشقرة سماح**.

كذلك أشكر كل من ساعدني وقدم لي العون وزرع التفاؤل في دربي ومد لي يد العون وزودني بالمعلومات اللازمة لإتمام هذا العمل وأخص بالذكر الأستاذة: **منصر سهيلة** والأستاذة **عائشة كراسة** وكذلك مفتش اللغة العربية **هيمه عبد الكريم**.

سناء

الملخص

في هذا العمل تحصلنا على علاقة رياضية تربط بين البنية الكيميائية المعبر عنها بالواصفات الجزيئية (2D)، والخاصية الفيزيوكيميائية المتمثلة في معامل التوزيع **$LogK_{ow}$**، وذلك للتنبؤ بقيم **$LogK_{ow}$ لمركبات أخرى**. شملت هذه الدراسة 13 كتلة من الواصفات الجزيئية (2D) وهذا وفق مرحلتين أولاهما هي حساب الكتل كل على حد, أما الثانية فهي حساب كل الكتل معا (مجمعة). وتمت هذه **الدراسة على** 2274 مركب كيميائي بتطبيق تقنية QSPR **وبطريقة الانحدار الخطي المتعدد** (MLR) وهذا باستخدام البرمجية الإحصائية Xlstat. **وكانت أحسن نتائج المرحلة الأولي هي (R^2 = 0,8556), (S^2 = 0.6286) و(F = 148.328) بالنسبة إلى كتلة (أجزاء مراكز الذرة).** أما بالنسبة للمرحلة الثانية (كل الكتل معا (النموذج العلم)) هي **(R^2 = 0.6887), (S^2 = 0.9678) و (F = 48.722).**

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

قائمة الاختصارات

الاختصار	المعنى	الترجمة إلى العربية
Kow	Octanol -Water Partition Coefficient	معامل التوزيع (أوكتانول ـ ماء)
LogKow	Octanol-Water Partition Coefficient logarithme	لوغاريتم معامل التوزيع (أوكتانول ـ ماء)
QSAR	Quantitative structure - activity relationships	العلاقات الكمية (البنية الكيميائية – الفعالية)
QSPR	Quantitative Structure - Property Relationships	العلاقات الكمية (البنية الكيميائية – الخاصية)
HPLC	High-performance liquid chromatography	كروماتوغرافيا سائلة عالية الأداء
MLR	Multiple Linear Regression	الإنحدار الخطي المتعدد
SLR	Simple Linear Regression	الإنحدار الخطي البسيط
PCR	Principal component regression	الإنحدار المكونات الرئيسية
PLSR	Partial Least Square Regression	الانحدار المربعات الصغرى
$^{+}$MM	Molecular mechanics	ميكانيكا الجزيئية
PM3	Parameterized Model number 3	معيار نموذج رقم 3
RM1	Recife Model 1	نموذج 1 لـ Recife
AM1	Austin Model 1	نموذج 1 لـ Austin
AAD	Average Absolute Deviation	متوسط الانحرافات المطلقة
SD	Standard Deviation	الإنحراف المعياري
SMILES	Simplified molecular-input line-entry system	مواصفات الإدخال النصي المبسط للجزيئات
MNDO	Modified Neglect of Diatomic Overlop	إهمال معدلة التداخل ثنائي الذرة
CNDO	Complete Neglect of Differential Overlap	إهمال كامل لفرق التداخل
PCA	Principal Component Analysis	تحليل العنصر الرئيسي
DDL	Data Definition Language	لغة تعريف البيانات
MSE	Mean Squared Error	خطأ تربيعي متوسط
RMSE	Root-Mean-Squared Error	خطأ تربيعي متوسط الجذر
MAPE	Mean absolute percentage error	النسبة المئوية لمتوسط الخطأ المطلق
DW	Durbin–Watson statistic	الإحصائية Durbin –Watson
Cp	Mallows's Cp	Mallows لـ Cp
AIC	Akaike information criterion	معيار Akaike للمعلومة
SBC	Bayesian Schwarz criterion	معيار Bayesian Schwarz
PC	Amemiya's Prediction Criterion	معيار Amemiya للتنبؤ

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 ajusté	معامل التحديد المضبوط
F	الاختبار الإحصائي (لقياس معنوية النموذج)
π	إلكترونات الرابطة الثنائية
S^2	الانحراف المعياري
σ	متوسط الانحرافات المطلقة

قائمة الأشكال

الرقم	العنوان	الصفحة
الفصل الأول: عموميات حول معامل التوزيع (أوكتانول - ماء)		
الشكل(I-1)	رسم تخطيطي يوضح معامل التوزيع (أوكتانول - ماء).	6
الشكل(I-2)	علاقة معامل التوزيع K_{OW} مع بعض الخصائص الفيزيوكيميائية الأخرى.	8
الشكل(I-3)	خطوات النمذجة الجزيئية.	12
الشكل(I-4)	علاقة QSPR/QSAR بتقييم المعايير.	13
الشكل(I-5)	مبدأ طريقة QSPR.	17
الفصل الثاني: طرق الحساب والبرامج المستعملة		
الشكل(II-1)	الخطوات المتبعة في التنبؤ.	19
الشكل(II-2)	خط الانحدار.	20
الشكل(II-3)	شكل انتشار الارتباط في الانحدار الخطي المتعدد.	25
الشكل(II-4)	البرنامج ChemOffice Ultra.	31
الشكل(II-5)	البرنامج HyperChem.	32
الشكل(II-6)	البرنامج DRAGON.	34
الشكل(II-7)	حساب الواصفات الجزيئية.	35
الشكل(II-8)	البرنامج XLSTAT.	37
الفصل الثالث: قاعدة البيانات		
الشكل(III-1)	الخطوات التجريبية للعمل.	44
الفصل الرابع: النتائج والمناقشة		
الشكل(IV-1)	بيان يوضح أسماء معاملات الواصفات الجزيئية الخاصة بنموذج الكتلة (أجزاء مراكز الذرة).	51
الشكل(IV-2)	بيان يوضح تغيرات $LogK_{OW}$ بدلالة $Préd(LogK_{OW})$ الخاصة بكتلة (أجزاء مراكز الذرة).	52
الشكل(IV-3)	بيان يوضح أسماء الواصفات الجزيئية الخاصة بنموذج جميع الكتل معا.	57
الشكل(IV-4)	بيان يوضح تغيرات $LogK_{OW}$ بدلالة $Préd(LogK_{OW})$ الخاصة بجميع الكتل معا.	60

قائمة الجداول

الرقم	العنوان	الصفحة
الفصل الأول: عموميات حول معامل التوزيع (أوكتانول - ماء)		
الجدول(I-1)	برامج النمذجة المستخدمة في علاقة (QSPR/QSAR) لإيجاد قيم $LogK_{ow}$.	15
الفصل الثاني: طرق الحساب والبرامج المستعملة		
الجدول(II-1)	درجة قوة علاقة معامل الارتباط.	24
الجدول(II-2)	أهم كتل الواصفات الجزيئية.	36
الفصل الثالث: قاعدة البيانات		
الجدول(III-1)	يوضح الجدول التالي جزءا من قاعدة البيانات.	42
الجدول(III-2)	نتائج حساب الواصفات الجزيئية المكونة لقاعدة البيانات باستعمال Dragon.	43
الفصل الرابع: النتائج والمناقشة		
الجدول(IV-1)	نتائج الكتل المدروسة كل على حد.	48
الجدول(IV-2)	معاملات التعديل الخاصة بالكتلة (أجزاء مراكز الذرة).	49
الجدول(IV-3)	ضبط معاملات نموذج الكتلة (أجزاء مراكز الذرة).	50
الجدول(IV-4)	تحليل الفروق الخاصة بالكتلة (أجزاء مراكز الذرة).	51
الجدول(IV-5)	الواصفة ورمزها للواصفات الجزيئية الأقل ارتباطا (أجزاء مراكز الذرة).	54
الجدول(IV-6)	الواصفة ورمزها للواصفات الجزيئية الأكثر ارتباطا (أجزاء مراكز الذرة).	54
الجدول(IV-7)	نتائج جميع الكتل المدروسة معا.	55
الجدول(IV-8)	معاملات التعديل الخاصة بجميع الكتل معا.	55
الجدول(IV-9)	تحليل الفروق الخاص بجميع الكتل معا.	55
الجدول(IV-10)	ضبط معاملات النموذج الخاص بجميع الكتل معا.	56
الجدول(IV-11)	الواصفة ورمزها للواصفات الجزيئية الأقل ارتباطا (جميع الكتل معا).	61
الجدول(IV-12)	الواصفة ورمزها للواصفات الجزيئية الأكثر ارتباطا (جميع الكتل معا).	62
الجدول(IV-13)	أمثلة للمركبات التي وجدت خارج مجال الثقة.	64
الجدول(IV-14)	نتائج عملية التحقق الخارجي (test).	65
الجدول(IV-15)	أسماء الـ 20 مركبا المحددة للعملية للتحقق الخارجي لـ (test).	66

قائمة المحتويات

الملخص	
قائمة الاختصارات	
قائمة الرموز	
قائمة الأشكال	
قائمة الجداول	
العنوان	**الصفحة**
مقدمة عامة	2
الجزء النظري: حوصلة بيبليوغرافية	
الفصل الأول: عموميات حول معامل التوزيع (أوكتانول - ماء)	
I. عموميات حول معامل التوزيع (أوكتانول ـ ماء)	6
I.1. تعريف معامل التوزيع (أوكتانول ـ ماء) K_{ow}	6
I.2. لوغاريتم معامل التوزيع (أوكتانول ـ ماء) $LogK_{ow}$	7
I.3. علاقة معامل التوزيع K_{ow} مع بعض الخصائص الفيزيوكيميائية الأخرى	7
I.4. الطرق التجريبية لحساب لوغاريتم معامل التوزيع (أوكتانول ـ ماء)	9
I.4.1. طريقة القارورة الهزازة	9
I.4.2. طريقة الكروماتوغرافيا السائلية العالية الأداء	9
I.4.3. طريقة التقليب البطيء	10
I.4.4. طريقة العمود المولد	10
I.5. النمذجة	11
I.5.1. ما هي النمذجة الجزيئية؟	11
I.5.2. النموذج الرياضي	12
I.5.3. مزايا استخدام النماذج الرياضية	12
I.6. العلاقات الكمية (البنية الكيميائية – الخاصية/الفعالية) (QSPR/QSAR)	13
I.6.1. نبذة تاريخية عن (QSPR/QSAR)	14
I.6.2. مشكلة نمذجة العلاقات الكمية (بنية كيميائية – الخاصية/الفعالية)	15
I.6.3. استخدام علاقات (QSPR/QSAR) لتقدير قيمة LogKow	15
I.6.4. تعريف QSPR	16
I.6.5. مبدأ QSPR	16
I.6.6. الطرق الرياضية المستخدمة في النموذج QSPR	17
الفصل الثاني: طرق الحساب والبرامج المستعملة	
II. طرق الحساب والبرامج المستعملة	19
II.1. الخطوات المتبعة في التنبؤ	19

2.II. طرق الحساب	20
1.2.II. الانحدار الخطي المتعدد	20
2.2.II. تعريف الانحدار الخطي المتعدد	21
3.2.II. نموذج الانحدار الخطي المتعدد	21
4.2.II. الارتباط الخطي	23
5.2.II. تعريف الارتباط	23
6.2.II. معامل الارتباط الخطي (r)	24
1.6.2.II. نوع العلاقة	24
2.6.2.II. قوة العلاقة	24
3.6.2.II. شكل الانتشار	25
4.6.2.II. قياس الارتباط	25
7.2.II. معامل التحديد (R^2)	26
8.2.II. الاختبار الإحصائية (F)	27
9.2.II. متوسط الانحرافات المطلقة (AAD)	27
10.2.II. الانحراف المعياري (SD)	28
3.II. البرامج المستعملة	29
1.3.II. أهم الطرق التي تحديد البنية الكيميائية للجزيئات	29
2.3.II. برنامج (ChemOffice Ultra)	30
3.3.II. برنامج (HyperChem)	31
1.3.3.II. طريقة MM^+ (Molecular mechanics)	31
2.3.3.II. طريقة RM1 (Recife Model 1)	32
3.3.3.II. طريقة PM3 (Parameterized Model number 3)	32
4.3.II. برنامج (DRAGON)	33
1.4.3.II. تعريف الواصف الجزيئي	34
2.4.3.II. كتل الواصف الجزيئي	35
3.4.3.II. حساب الواصفات الجزيئية	35
5.3.II. برنامج (XLSTAT)	37
الجزء التطبيقي: الحساب والنتائج	
الفصل الثالث: قاعدة البيانات	
III. قاعدة البيانات	40
1.III. مكونات قاعدة البيانات	40
2.III. ما سبب اهتمامنا بالمركبات العضوية (مركبات الكربون)؟	40
3.III. عموميات حول المركبات العضوية	41
1.3.III. تعريف المركبات العضوية	41

2.3.III. تصنيفها	41
4.III. كيفية معالجة وحساب الواصفات الجزيئية المكونة لقاعدة البيانات	43
5 .III. الخطوات التجريبية للعمل	44
6.III. شرح الخطوات التجريبية للعمل	45
الفصل الرابع: النتائج والمناقشة	
IV. النتائج	48
1.IV. نتائج المرحلة الأولى	48
1.1.IV. الكتلة أجزاء مراكز الذرة (Atom-centered fragments)	49
2.1.IV. تحليل نتائج المرحلة الأولى	53
3.1.IV. تحليل البقايا (Residues) نتائج المرحلة الأولى	53
4.1.IV. مناقشة النتائج للمرحلة الأولى	53
2.IV. نتائج المرحلة الثانية	55
1.2.IV. تحليل نتائج المرحلة الثانية	61
2.2.IV. تحليل البقايا (Residues) نتائج المرحلة الثانية	61
3.2.IV. مناقشة النتائج للمرحلة الثانية	61
خاتمة عامة	68
المراجع	70
الملاحق	74

مقدمة عامة

مقدمة عامة

إن التطور التكنولوجي والصناعي السريع على مدى العقود الماضية يهدف إلى زيادة رفاهية البشرية. ومع ذلك، فإنه لا يخلو من عواقب على صحة الإنسان والبيئة. وهذا بسبب المواد الكيميائية المختلفة للتطوير والذي يسبب تسرب المواد السامة في البيئة.

في يومنا هذا زاد اهتمام الباحثين في مجال الكيمياء بدراسة الخاصية الفيزيوكيميائية معامل التوزيع (أوكتانول - ماء) الذي هو عبارة عن نسبة تركيز مادة مذابة بين سائلين أحدهما مائي والآخر عضوي، ومن أهم تطبيقاته هو استعماله في قياس كراهية الماء للمركبات العضوية وله عدة تطبيقات أخرى في مختلف المجالات منها: طرق الفصل، الاستخلاص سائل – سائل، الكروماتوغرافيا والصيدلية مثلا [21].

يعبر لوغاريتم معامل التوزيع بين أوكتانول وماء، عن معيار أساسي لتقدير ألفة مادة ما للدهون [13].

إن النمذجة الجزيئية هي مجال يشمل جميع الأساليب النظرية والتقنيات الحسابية المستخدمة في تصميم نموذج أو تتبع سلوك الجزيئات. يتم استخدام التقنيات في مجالات الكيمياء الحسابية، تصميم الأدوية، علم الأحياء الحسابي وعلوم المواد لدراسة النظم الجزيئية التي تتراوح بين أنظمة كيميائية صغيرة إلى جميع أنحاء الجزيئات البيولوجية والتجمعات المادية.

تعد برامج النمذجة الجزيئية من أهم البرامج المساهمة في تصوير المركب الكيميائي على هيئة ثلاثي الابعاد وتعطينا خصائص الجزيئات، وفي هذه الدراسة استخدمنا مجموعة من البرمجيات وهي البرنامج (ChemDraw Ultra) لرسم البنية الجزيئية لهذه المركبات العضوية، والبرامج (HyperChem) و(Dragon) لنمذجة الجزيئات وحساب المواصفات، 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 حسابها مخبريا.

وقد أجريت هذه الدراسة وفق الخطة التالية:

- ✓ مقدمة عامة حيث تحدثنا فيها عن أهمية النمذجة الجزيئية والتعرف إلى البرامج المستعملة في الدراسة وإلى أهمية الخاصية الفيزيوكيميائية المختارة.
- ✓ الجزء الأول حوصلة بيبليوغرافية 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 إلى المركبات العضوية المكونة لقاعدة البيانات من حيث تعريفها، أهميتها، تصنيفها، 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.
- ✓ وفي الأخير خلاصة عامة متبوعة بتوصيات واقتراحات.

الجزء النظري

حوصلة بيبليوغرافية

الفصل الأول

عموميات حول معامل التوزيع (أوكتانول - ماء)

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

مقدمة:

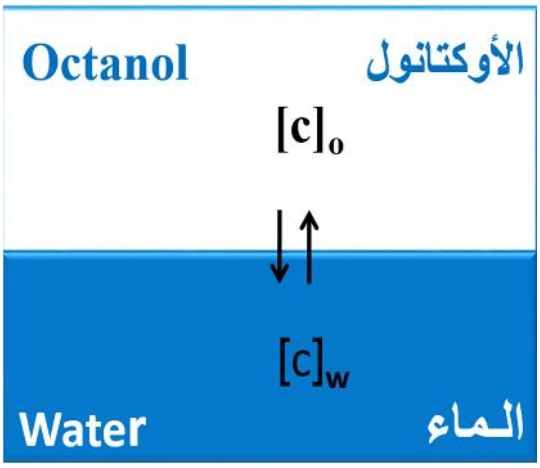
اهتم الباحثون في العقود الأخيرة بدراسة معامل التوزيع (أوكتانول - ماء) الذي هو عبارة عن نسبة تركيز مادة مذابة بين سائلين مختلفين لا يمتزجان، ومن أهم تطبيقاته هو استعماله في قياس كراهية الماء للمركبات العضوية وله عدة تطبيقات أخرى في مختلف مجالات العلوم نذكر منها: طرق الفصل، الاستخلاص سائل – سائل، الكروماتوغرافيا والعلوم الصيدلية. الكيمياء الحيوية والسموم وفي تقييم الخصائص البيولوجية ذات الصلة بفعالية الأدوية، ذوبانية الدهون، الامتصاص الخلوي، حيث يعتبر معامل التوزيع مؤشرا جيدا رابطا بين بنية المركب وخصائصه الفيزوكيميائية وله دور مهم في دراسة العلاقات (QSPR/QSAR). ونظرا لصغر مجال قيم K_{OW} يعوض بـ $LogK_{OW}$ [21].

1.I. تعريف معامل التوزيع (أوكتانول - ماء) K_{ow}:

معامل التوزيع (أوكتانول - ماء) K_{ow} هو نسبة تركيز المادة المذابة في طور أوكتانول وفي الطور المائي. يعبر عن تراكيز من حيث الوزن أومولات المواد الكيميائية لكل وحدة حجم من السائل [7].

$$K_{ow} = \frac{C_{OCTANOL}}{C_{WATER}} \qquad (1.I)$$

$C_{OCTANOL}$: تركيز المذاب في الطور الأوكتانول.
C_{WATER}: تركيز المذاب في الطور المائي.



الشكل(I-1): رسم تخطيطي يوضح معامل التوزيع (أوكتانول - ماء).

معامل التوزيع K_{ow} هو الخاصية الفيزيوكيميائية التي يمكن قياسها في المختبر وتستخدم للتمييز الكمي للمركبات العضوية الكارهة للماء. وهو المؤشر الأولي للفصل بين الماء والمواد العضوية.

يفضل في حالة المواد العضوية استخدام قيم عالية الجودة مستمدة من التجربة لمعامل التوزيع (أوكتانول - ماء)، أو قيم سبق تقييمها في الدراسات باعتبارها "القيم الموصى بها".

في حالة عدم توفر بيانات تجريبية عالية الجودة، يمكن لأغراض التصنيف استخدام بيانات معتمدة من قبيل العلاقات الكمية (بنية كيميائية – خاصية/فعالية) (QSPR/QSAR) لتقييم $LogK_{ow}$ [13].

2.I. لوغاريتم معامل التوزيع (أوكتانول - ماء) $LogK_{ow}$:

تعاريف عامة:

يعبر لوغاريتم معامل التوزيع بين الأوكتانول والماء عن مقدار ألفة مادة ما للدهون وعلى هذا النحو، يمثل $LogK_{OW}$ معيارا أساسيا في تقدير المصير البيئي للمواد. وتشتق عمليات توزيع عديدة باستخدام لوغاريتم معامل التوزيع، مثل الامتصاص في التربة والترسبات، والتركيز في الكائنات العضوية [13].

أما تعيين معامل التوزيع بين الأوكتانول والماء هو شرط في مجموعة البيانات الأساسية التي تقدم للمواد الجديدة والمواد القائمة التي تتمتع بأولية 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) [13].

$$K_{ow} = \left(\frac{Co}{Cw}\right) \qquad (2.I)$$

إذا:

$$\log K_{ow} = \log\left(\frac{Co}{Cw}\right) \qquad (3.I)$$

3.I. علاقة معامل التوزيع K_{OW} مع بعض الخصائص الفيزيوكيميائية الأخرى:

يعتبر معامل التوزيع (أوكتانول - ماء) $LogK_{OW}$ الخاصية المشاركة في حساب الخصائص الأخرى مثل (الذوبانية في الماء، التركز الحيوي، الامتصاص في التربة والرواسب والنباتات...). ويعتبر كذلك معيارا جيدا للتنبؤ بمعامل التركز الحيوي لأنواع معينة من المركبات العضوية، مثل المواد العضوية اللاقطبية [13].

الشكل(I-2): علاقة معامل التوزيع K_{OW} مع بعض الخصائص الفيزيوكيميائية الأخرى.

يتمثل أساس العلاقة بين التركز الحيوي وقيمة $LogK_{OW}$ 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 [13].

هناك علاقات واضحة بدرجة عالية بين قيمة $LogK_{OW}$ وذوبانية المواد في زيت كبد الحوت والتريولين، وهذا الأخير هو أحد أوفر مركبات ثلاثي أسيل غليسرول التي توجد في ليبيدات أسماك المياه العذبة [13].

I.4. الطرق التجريبية لحساب لوغاريتم معامل التوزيع (أوكتانول - ماء):

وردت طريقتان مختلفتان للتعيين العملي لقيم معامل التوزيع (أوكتانول - ماء)، هما طريقة القارورة الهزازة وطريقة الكروماتوغرافيا السائل عالية الأداء (HPLC). في توجيهات قياسية مثل:

EPA-OTS (1982)، EEC A.8 (1992)، OECD 117 (1983) ، OECD 107(1995)، EPA-FIFRA (1982)، ASTM (1993). ولا تقتصر البيانات الموصى بها على البيانات الناتجة من إستخدام طريقة القارورة الهزازة أو طريقة الكروماتوغرافية وفقا للتوجيهات. إذ أنه بالنسبة للمواد الشديدة الألفة للدهون التي تذوب ببطء في الماء تعتبر البيانات الناتجة من استخدام طريقة التقليب البطيء أكثر موثوقية بصفة عامة. ويجري في وقت الحاضر اختيار دائري لطريقة التقليب البطيء تمهيدا لوضع توجيه نهائي لمنظمة التعاون والتنمية [13].

I.4.1. طريقة القارورة الهزازة:

المبدأ الأساسي للطريقة هو قياس ذوبان المادة في طورين مختلفين، الماء والأوكتانول. ولتعيين معامل التوزع يجب بلوغ حالة الأتزان بين جميع المكونات المتفاعلة في النظام، ومن ثم يعين تركيز المواد الذائبة في الطورين. تطبق طريقة القارورة الهزازة 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° [13].

I.4.2. طريقة الكروماتوغرافيا السائل عالية الأداء:

تجرى الكروماتوغرافيا السائل عالية الأداء (HPLC) في أعمدة تحليلية معبأة بطور صلب تجاري يحتوي سلاسل هيدروكربون طويلة مربوطة كيميائيا 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، أوالمركبات الفلزية المعقدة أوالمواد الخافضة للتوتر السطحي أوالمواد التي تتفاعل مع المذيب. وتنطبق الطريقة الكروماتوغرافيا عندما تقع قيمة لوغاريتم معامل التوزيع (أوكتانول ـ ماء) في النطاق من 0 إلى 2 وهذه الطريقة ذات حساسية أقل لوجود الشوائب في المركب المختبر بالمقارنة بطريقة القارورة الهزازة [13].

I.3.4. طريقة التقليب البطيء:

تتيح طريقة التقليب البطيء الحصول على تعيين دقيق وصحيح لقيمة معامل التوزيع (أوكتانول - ماء) للمركبات التي يصل فيها 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 التدريج من أجل الحصول على تقديرات لقيمة K_{OW}.

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 التقليب البطيء على الصعوبات التجريبية المرتبطة بتكوين القطيرات الدقيقة، إلى حد ما، بتحقيق حالة الاتزان بين الماء والأوكتانول والمركب المختبر في مفاعل يعرض لتقليب بطيء. ويولد التقليب تدفقا انسيابيا إلى حد ما بين الأوكتانول والماء، ويعجل التبادل بين الطورين بدون تكوين قطيرات دقيقة [13].

I.4.4. 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. ويعبأ العمود بمادة مازة صلبة ويشبع بتركيز ثابت من المادة المختبرة في الأوكتانول.

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 [13].

I.5. النمذجة:

مقدمة:

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

ويرى الباحث الأمريكي راسل راي (Rassyl Ray) أن النمذجة تتراوح بين النمذجة التحليلية والنمذجة التركيبية. ويقصد بالنمذجة التحليلية حيث يتم التوصل إلى نماذج جزئية من النموذج الكلي، أما النمذجة التركيبية الوصول للنموذج الكلي باستخدام العديد من النماذج الجزئية أوالمصغرة.

كما يصف العالم الأمريكي كلير (Klir) الصعوبة التي يواجهها صانع النموذج بأنها تتمثل في تقبله للواقع المراد نمذجته وإعادة بناؤها بناء رياضيا ومنطقيا. لذلك الهدف الحقيقي من النمذجة هو تجريد تبسيطي للواقع، هذا التجريد يساعد على حل مشكلة معينة [10].

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 [10].

I.5.1. ما هي النمذجة الجزيئية؟

كلمة "الجزيئية" تعني بوضوح التعامل مع الجزيئات. ويعرف قاموس أوكسفورد النموذج بأنه "وصف مبسط أو مثالي لنظام أوعملية، في المصطلحات الرياضية كثيرا ما يستخدم لتسهيل العمليات الحسابية والتوقعات". تهتم النمذجة الجزيئية بتقليد سلوك أنظمة الجزيئي والجزيئات.

كما ترتبط هذه النمذجة بشكل ثابت بالنمذجة الحاسوبية. ومع ذلك، أحدثت التقنيات الحاسوبية ثورة في النمذجة الجزيئية إلى درجة أن غالبية الحسابات لا يمكن أن تنجز بدون استعمال الحاسوب. هذا لا يعني أن نموذجا أكثر تطورا هو بالضرورة أفضل أي واحد بسيط، ولكن أجهزة الكمبيوتر لديها بالتأكيد مجموعة أوسع من النماذج التي يمكن النظر فيها والنظم التي يمكن تطبيقها [10].

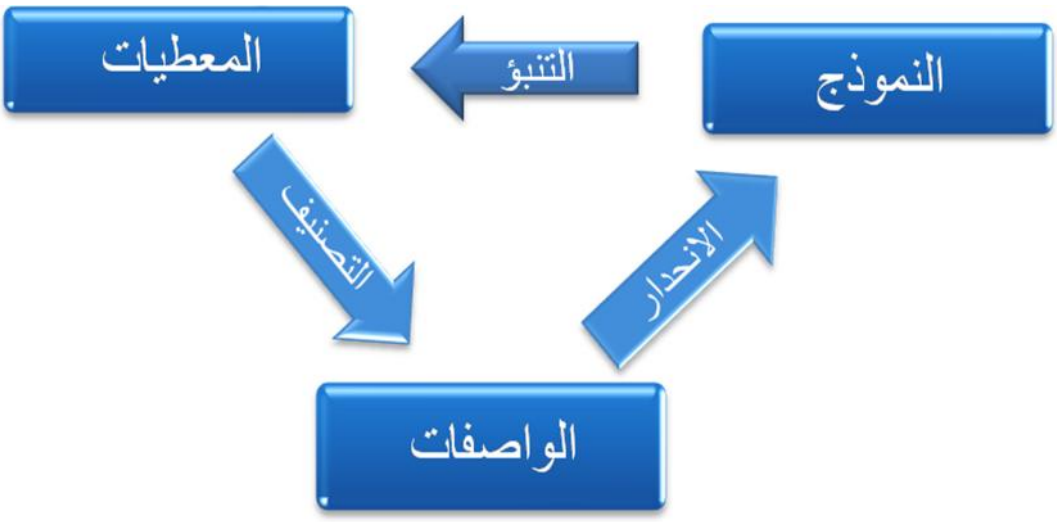
I.2.5. النموذج الرياضي:

النموذج الرياضي عبارة عن مجموعة من المعادلات التي تصف سلوك النظام تحت الدراسة وهذه المعادلات يكون عددها مساويا لعدد المتغيرات فيها وعامة هذه المعادلات مبرمجة حيث تحل بواسطة الكمبيوتر وهذه هي الطريقة الأنسب للتعامل مع المعادلات الرياضية التي تتطلب حلا عدديا [10].

I.3.5. مزايا استخدام النماذج الرياضية:

توجد مزايا كثيرة لاستخدام النماذج الرياضية ومنها:

- ✓ تساعد الباحث على ترتيب المعتقدات النظرية والملاحظات المبدئية عن النظام واستنباط النتائج المنطقية لهذا الترتيب.
- ✓ تصف النماذج الرياضية المشكلة بطريقة مختصرة وهذا أفضل من الوصف الشفهي.
- ✓ تسمح النماذج الرياضية بالتحكم في مصادر التغير أكبر مما تسمح به **الدراسة المباشرة**.
- ✓ **أقل تكلفة من النظام**.
- ✓ تساعد النماذج الرياضية على سرعة التحليل.
- ✓ تساعد النماذج الرياضية على فهم النظام.
- ✓ الاهتمام بالوصف من خلال التفصيل والعرض [10].



الشكل(I-3): خطوات النمذجة الجزيئية.

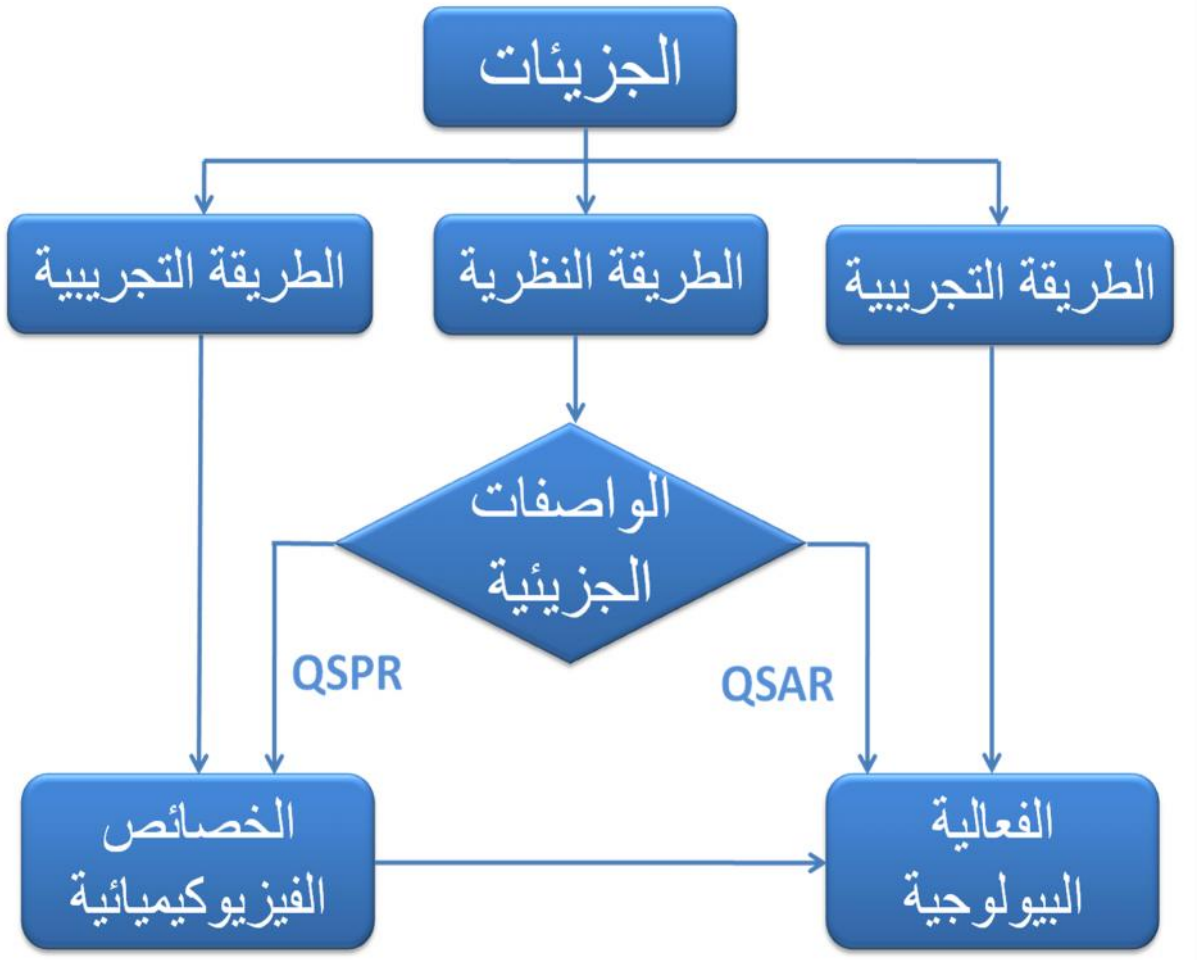
I.6. العلاقات الكمية (البنية الكيميائية – الخاصية/الفعالية) (QSPR/QSAR):

مقدمة:

العلاقة الكمية هي علاقة بين البنية الكيميائية وبين نشاطه البيولوجي أوخصائصه الفيزيوكيميائية. يسمح تحليل علاقة الكمية بالبنية بتحديد الزمر الكيميائية المسؤولة عن استثارة التأثير البيولوجي المستهدف في الكائن الحي، وهذا يسمح بتعديل تأثير مركب فعال بيولوجيا (عادة دواء) عن طريق تغيير بنيته الكيميائية.

يستخدم المختصون في الكيمياء الدوائية طرائق الاصطناع الكيميائي لإدخال زمر كيميائية جديدة في المركب واختبار التغيرات في تأثيراتها البيولوجية.

جرى تطوير هذه الطريقة لبناء نماذج رياضية بين البنية الكيميائية والنشاط البيولوجي والخصائص الفيزيوكيميائية، وهذا ما يسمى بالعلاقة الكمية بين الفعالية/الخاصية والبنية الكيميائية [13].



الشكل(I-4): علاقة QSPR/QSAR بتقييم المعايير.

1.6.I. نبذة تاريخية عن (QSPR/QSAR):

يرجع تاريخ استخدام العلاقات (QSPR/QSAR) في علم السموم المائية إلى الأعمال التي أنجزت في أواخر القرن التاسع عشر من قبل الباحثين (أوفرتون في زيورخ) و(ماير في ماربورغ) والتي توضح أن قدرة المواد على إحداث تخدير في صغار الضفادع وصغار الأسماك ترتبط ارتباطا مباشرا بمعامل توزعها المقاس بين زيت الزيتون والماء. طرح أوفرتون في درسته عام 1901 فرضية بإن هذا الإرتباط يعكس السمية التي تحدث في تركيز أوفي حجم الجزيئي القياسي، في نطاق موقع جزيئي معين في داخل الكائن العضوي.

كما خلص إلى أن هذا يناظر التركيز أوالحجم نفسه في مختلف الكائنات، بغض النظر عما إذا كان الامتصاص من الماء أوعن طريق استنشاق غازي. ويعرف هذا الارتباط في علم التخدير بإسم نظرية ماير- أوفرتون.

اقترح كورفن هانس ومعاونوه في كلية بومونا استخدام نظام معامل التوزيع K_{ow} كنظام قياسي للتوزيع، ووجد أن معاملات التوزيع هذه قيم جمعية تركيبية يمكن تقديرها مباشرة على أساس التركيب الكيميائي. وذكر هؤلاء الباحثين كذلك أنه يمكن تصميم نماذج للعلاقات (QSPR/QSAR) على أساس تحليل إحصائي للانحدار في النتائج. أعلن الباحثون عن طريق تطبيق هذا المنهج في عام 1982 ما مجموعة 138 نموذجا للعلاقات (QSPR/QSAR) على شكل:

$$\mathbf{Log(1/C) = A \log K_{ow} + B} \qquad \mathbf{(4.I)}$$

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

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

وكانت الإنحدارات ومقاطع محاور الإحداثيات متطابقة تقريبا في خمس من هذه المعادلات التي بسمية 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 تعمل من خلال آلية تخدير في اختبار السمية الحادة في الأسماك، مما يؤدي إلى سمية دنيا أوسمية خط الأساس [13].

I.2.6. مشكلة نمذجة العلاقات الكمية (البنية الكيميائية - الخاصية/الفعالية):

إن اختيار قيمة مناسبة للعلاقات QSPR/QSAR يعني أن يعطى النموذج تنبؤا موثوقا لسمية مادة كيميائية لم تختبر أولنشاطها الحيوي. 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.

ومتى كانت هناك حاجة إلى إدخال تعديلات تركيبية طفيفة لزيادة نشاطه وتقليل سميته والهدف الإجمالي هو عمل تقديرات عن طريق الاستيفاء الداخلي أكثر منه عن طريق الاستكمال الخارجي [13].

I.3.6. استخدام علاقات (QSPR/QSAR) لتقدير قيمة $LogK_{ow}$:

عند وجود قيمة مقدرة للمعامل $LogK_{ow}$ يجب أخذ طريقة التقدير في عين الاعتبار، وقد وضعت ولا تزال توضع قيم عديدة للعلاقات (QSPR/QSAR) بغرض تقدير قيمة المعامل $LogK_{ow}$. وتستخدم أغلب الأحيان برامج حاسوبية منها:(CLOGP, LOGKOW, AUTOLOGP) لتقدير المخاطر إذا لم تتوفر بيانات مشتقة من التجربة.

وتقوم البرامج CLOGP, LOGKOW, AUTOLOGP على إضافة مساهمات المجموعات التفاعلية.

ويوصى باستخدام برنامج CLOGP وLOGKOW لإغراض التصنيف بسبب موثوقيتهما وتوفرهما تجاريا، وسهولة استخدامهما [13].

الجدول(I-1): برامج النمذجة المستخدمة في علاقة (QSPR/QSAR) لإيجاد قيم $LogK_{ow}$ [13].

برنامج النمذجة	مجال قيم $Log\ K_{ow}$	المواد المعنية
CLOGP	$9 > LogK_{ow} > 0$	يحسب هذا البرنامج قيم $LogK_{ow}$ للمركبات العضوية، التي تحتوي الذرات التالية: C, H, N, O, الهالوجين Si, P, Se, و/أو Li, Na, K و/أو Hg.
LOGKOW (KOWWIN)	$8> LogK_{ow} > 4$	يحسب هذا البرنامج قيم $LogK_{ow}$ للمواد العضوية التي تحتوي الذرات التالية: C, H, N, O, الهالوجين Si, P, Se, و/أو Li, Na, K و/أو Hg. يمكن أيضا حساب قيم $LogK_{ow}$ للأملاح وللمركبات التي تحمل شحنات شكلية (بإستثناء مركبات النترو وأكسيد النتروجين).
AUTOLOGP	$LogK_{ow}< 5$	يحسب البرنامج $LogK_{ow}$ للمواد الكيميائية العضوية التي ذرات H, N, O, C الهالوجين، P وS.

I.4.6. تعريف QSPR:

(QSPR): Quantitative Structure - Property Relationship

QSPR هي العملية التي من خلالها يتم تأسيس العلاقات الكمية بين البنية الجزيئية لمجموعة مركبات مع الخاصية الفيزيوكيميائية. وذلك بالمراحل الرئيسية الثلاثة لبناء نموذج (حساب) QSPR وهي:

- ✓ اختيار الواصفات الجزيئية المناسبة مع إشكالية بنية الكيميائية – الخاصية.
- ✓ استغلال قيم الواصفات الجزيئية كمتغيرات لتحديد العلاقة التي تربطها مع الخاصية المدروسة، وذلك باستخدام برامج خاصة.
- ✓ وضع معايير الأداء والتحقق من شأنها أن تساعد في اختيار أفضل النماذج للإشكالية المطروحة وتقدير أخطاء التنبؤ [9].

I.5.6. مبدأ QSPR:

مبدأ طريقة QSPR هي إنشاء العلاقة الرياضية التي تربط كميا خصائص الجزيئات الإلكترونية والهندسية، تعرف بالواصفات الجزيئية (الفعالية البيولوجي، سمية، وخصائص الفيزيوكيميائية، وغيرها) لسلسلة من المركبات الكيميائية متشابهة باستخدام طرق تحليل المعطيات. والشكل العام للنموذج كالتالي:

$$\text{الخاصية} = f(\text{الواصفات}) \qquad (5.I)$$

الهدف من هذه الطريقة هو تحليل البيانات البنيوية للكشف عن عوامل خاصية قياسها. وللقيام بذلك، يمكن استخدام الأدوات التالية:

- ✓ الانحدار الخطي.
- ✓ أشجار القرار.
- ✓ الشبكات العصبية.
- ✓ الخوارزميات الوراثية.

بمجرد إنشاء هذه العلاقة والتأكد من صلاحيتها، يمكننا أن نتنبأ بالخصائص جديدة للمركبات مجهولة، والتي هي غير معروفة تجريبيا [9].

الشكل(I-5): مبدأ طريقة QSPR.

I.6.6. الطرق الرياضية المستخدمة في النموذج QSPR:

- ✓ الانحدار الخطي البسيط (SRL).
- ✓ الانحدار الخطي المتعدد (MRL).
- ✓ الانحدار المكونات الرئيسية (PCR).
- ✓ الانحدار المربعات الصغرى (PLS) [9].

الفصل الثاني

طرق الحساب والبرامج المستعملة

II. طرق الحساب والبرامج المستعملة:

1.II. الخطوات المتبعة في التنبؤ:

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

2.II. طرق الحساب:

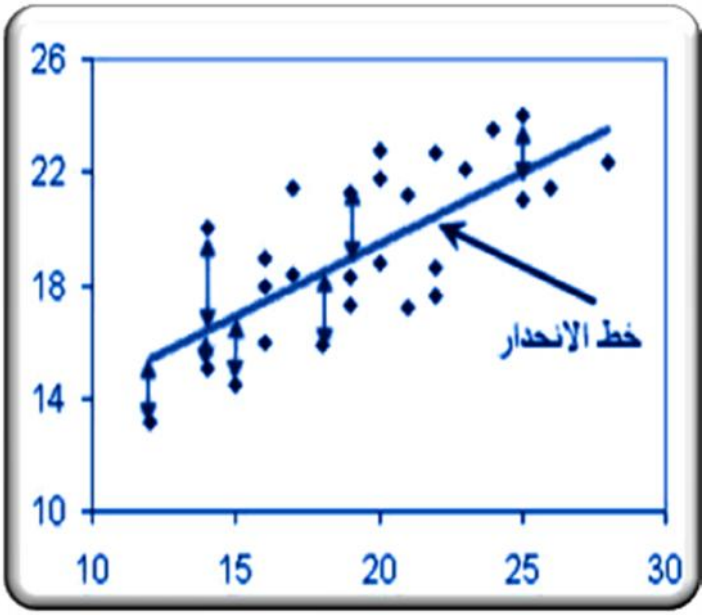
مقدمة:

إن نمذجة جزيئية بواسطة جهاز الحاسوب تعطي معلومات حول هندسة وشكل هذا الجزئي وبعض الخصائص (الشحنة، الإستقطابية، الحجم والسطح الجزيئي...)، كما تسمح أيضا بتحديد بنية الجزيئ وطاقة استقراره، وتستعمل كذلك طريقة النمذجة الجزيئية في ميدان الكيمياء العضوية وهذا لتقدير درجة الارتباط المتعلقة بـ (بنية - فعالية (QSAR) أوبنية - 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 يربط علاقة بين المتغير التابع الذي يمثل الخاصية الفيزيوكيميائية والعديد من المتغيرات المستقلة، والتي تتمثل في الواصفات الجزيئية.

ويستخدم الانحدار الخطي للتنبؤ بقيمة المتغير التابع، من خلال تمثيل العلاقة بين المتغير التابع والمتغيرات المستقلة [17].



الشكل(II-2): خط الانحدار [11].

يعد الانحدار الخطي المتعدد من الأساليب الإحصائية المتقدمة والتي تضمن دقة الاستدلال من أجل تحسين نتائج البحث عن طريق الاستخدام الأمثل للبيانات في إيجاد علاقات سببية بين الظواهر موضوع البحث [16].

2.2.II. تعريف الانحدار الخطي المتعدد:

الانحدار الخطي المتعدد هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة وللتنبؤ بقيم مستقبلية، وهو أيضا عبارة عن انحدار للمتغير التابع (**Y**) على العديد من المتغيرات المستقلة $X_1 , X_2 , ...X_K$ لذا فهو يستخدم في التنبؤ بتغيرات المتغير التابع الذي تؤثر فيه عدة متغيرات مستقلة أي تعتمد فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أوالانتشار [16].

إن الانحدار الخطي المتعدد ليس مجرد أسلوب واحد وإنما مجموعة من الأساليب التي يمكن أستخدامها لمعرفة العلاقة بين متغير تابع مستمر وعدد من المتغيرات المستقلة التي عادة ما تكون مستمرة [16].

3.2.II. نموذج الانحدار الخطي المتعدد:

يتمثل هذا النموذج في المعادلة الرياضية التي تمكننا من الحصول على أفضل معادلة انحدار تربط العلاقة بين المتغير التابع والمتغيرات المستقلة وذلك إذا توافرت لدينا قاعدة بيانات من عينة حجمها n فيكون النموذج الرياضي الخطي بالشكل التالي:

$$Y_j = a + b_1X_{1j} + b_2X_{2j} + \dots + b_{k-1}X_{kj} + U_j \quad ،j=1, 2,\dots, n \qquad (1\text{-II})$$

حيث أن:

Y_j: المتغير التابع.

a: ثابت الانحدار الخطي.

b_1: ميل الانحدار y على المتغير المستقل الأول.

b_2: ميل الانحدار y على المتغير المستقل الثاني.

X_{1j}: المتغير المستقل الأول.

X_{2j}: المتغير المستقل الثاني [16].

U_j: المتغير العشوائي.

في الواقع الأمر فإن هذه المعادلة واحدة من جملة معادلات يبلغ عددها (n) ويكون نظام المعادلات كالأتي [16]:

$$\begin{cases} Y_1 = a + b_1X_{11} + b_2X_{21} + \dots + b_kX_{k1} + U_1 \\ Y_2 = a + b_1X_{12} + b_2X_{22} + \dots + b_kX_{k2} + U_2 \\ . \quad . \quad . \quad . \quad . \quad . \quad \dots \quad . \quad . \quad . \\ Y_n = a + b_1X_{1n} + b_2X_{2n} + \dots + b_kX_{kn} + U_n \end{cases} \qquad \textbf{(2-II)}$$

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 كالأتي [16]:

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

وباختصار يمكن كتابتها كالأتي:

$$Y = XB + U \qquad \textbf{(4-II)}$$

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 [16].

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 توزيعا طبيعيا للمتغيرات المستقلة والمتغير التابع.
3. يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل [16].

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

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

حيث أن:

$\hat{Y}_j$: 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 [16].

بعد الحصول على نتائج معادلة الانحدار يجب علينا أن نبين هل أن هذه المعاملات مقبولة من الناحية الإحصائية أي معنوية إحصائيا مع التنويه بأن المعنوية تكون لكل معامل على حده.

يعتبر معامل الارتباط الخطي مقياسا جيدا لدراسة العلاقة بين متغيرين والتي تأخذ الشكل الخطي وسوف تقتصر دراستنا على الارتباط بين متغيرين (ظاهرتين) فقط [20].

كما سيتم الحصول على إحصائيات تستخدم لمعرفة المعنوية الإجمالية للنموذج ومنها معامل الارتباط الخطي (r)، معامل التحديد (R^2)، الاختبار الإحصائي (F).

4.2.II. الارتباط الخطي:

يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها من متغير وأحد، وننتقل من التعامل مع متغير وأحد إلى التعامل مع متغيرين أوأكثر، وسنتناول هذه الوحدة دراسة وتحليل العلاقة بين متغيرين، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، وذلك لدراسة العلاقة بين متغيرين (X,Y) [26].

5.2.II. تعريف الارتباط:

الارتباط يعنى وجود علاقة بين ظاهرتين أو متغيرين بمعنى أن التغير في أحد المتغيرين يؤدى إلى التغير في المتغير الأخر سواء بالزيادة أوالنقصان، وإذا كان المتغيران يزيدان معا وينقصان معا فإن العلاقة بينهما (طردية)، أما إذا كان أحدهما ينقص بزيادة المتغير الأخر [26] فإن العلاقة بينهما (عكسية).

II.6.2. معامل الارتباط الخطي (r):

معامل الارتباط هو مقياس كمي نسبي يستخدم لتحديد نوع وقوة العلاقة بين متغيرين، ويرمز له بالرمز **(r)** [26].

II.6.2.1. نوع العلاقة:

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

- ✓ إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الآخر، والعكس.
- ✓ إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الآخر، والعكس.
- ✓ إذا كان معامل الارتباط قيمته تساوى صفرا ($r = 0$) دل ذلك على عدم وجود علاقة (أي ارتباط) بين المتغيرين [26].

II.6.2.2. قوة العلاقة:

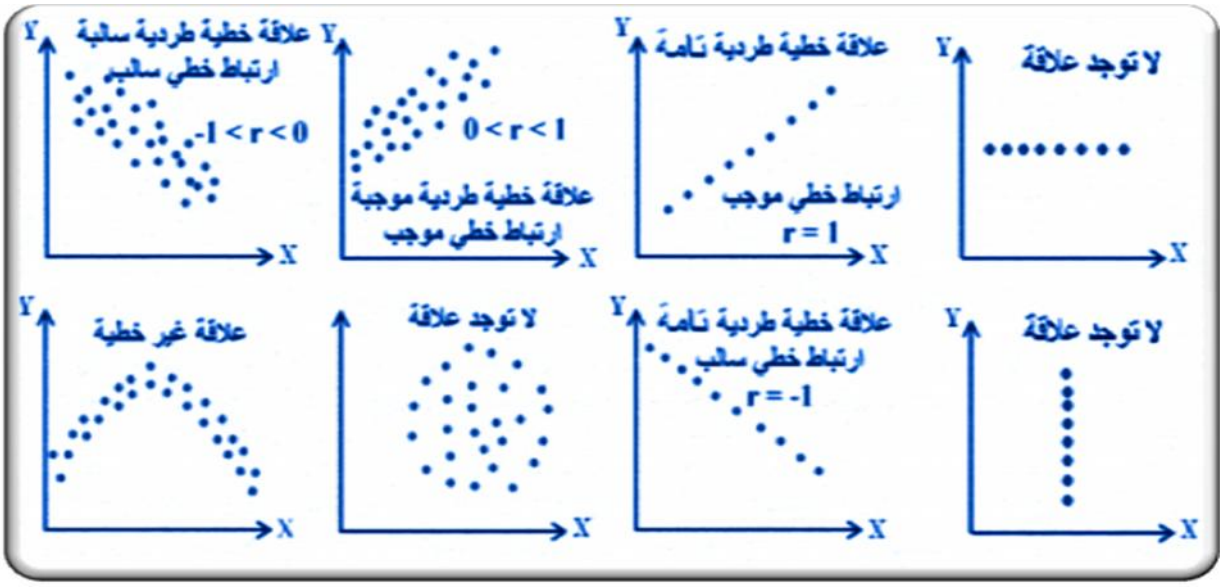
يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن الصفر، حيث أن قيمة معامل الارتباط تقع في المدى ($-1 \leq r \leq 1$)، وقد صنف بعض الإحصائيين درجات لقوة العلاقة التي يمكن تمثيلها على الجدول التالي [26]:

الجدول(II-1): درجة قوة علاقة معامل الارتباط [26].

قيمة معامل الارتباط	المعنى
+1	ارتباط طردي تام
(من 0.90 إلى 0.99)	ارتباط طردي قوي جدا
(من 0.70 إلى 0.89)	ارتباط طردي قوي
(من 0.50 إلى 0.69)	ارتباط طردي متوسط
(من 0.30 إلى 0.49)	ارتباط طردي ضعيف
(من 0.01 إلى 0.29)	ارتباط طردي ضعيف جدا
0	لا يوجد ارتباط
(من 0.01- إلى 0.29-)	ارتباط عكسي ضعيف جدا
(من 0.30- إلى 0.49-)	ارتباط عكسي ضعيف
(من 0.50 إلى 0.69-)	ارتباط عكسي متوسط
(من 0.70- إلى 0.89-)	ارتباط عكسي قوي
(من 0.90- إلى 0.99-)	ارتباط عكسي قوي جدا
-1	ارتباط عكسي تام

3.6.2.II. شكل الانتشار:

شكل الانتشار هو عبارة عن تمثيل أزواج قيم الظاهرتين المتغيرين (X, Y) بمجموعة من النقاط على الرسم. فيتم تمثيل المتغير X على المحور الأفقي، والمتغير Yعلى المحور الرأسي حيث نرصد لكل زوج من قيم المتغيرين بنقطة على الرسم فنحصل على شكل الانتشار [26].



الشكل(II-3):شكل انتشار الارتباط في الانحدار الخطي المتعدد [5].

4.6.2.II. قياس الارتباط:

يمكن قياس الارتباط بين المتغيرين[23] بمعامل الارتباط الخطي بدلالة القراءات لبيانات المتغيرين باستخدام الصيغة التالية [26]:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} \qquad (6-II)$$

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 [26].

7.2.II. معامل التحديد (R^2):

أما R^2 فهو يسمى بمعامل التحديد والذي يستخدم لمعرفة القوة التفسيرية للنموذج المقدر (المعادلة المقدرة) في حالة الانحدار الخطي المتعدد.

تعريف:

يعرف معامل التحديد بأنه مربع معامل الارتباط ويرمز له بالرمز R^2 حيث: $R^2 = r^2$ [18].

وهو معامل يستخدم لمعرفة مدى جودة التوفيق لخط الانحدار المقدر للمتغير التابع على المتغير المستقل، وبما أنه عبارة عن مربع معامل الارتباط لذلك فقيمته تكون موجبة محصورة بين الصفر والواحد، أي: $0 \leq R^2 \leq 1$ [18].

هو يقيس نسبة التغير في المتغير التابع نتيجة تغير المتغير المستقل، أي أنه يوضح نسبة مساهمة المتغير المستقل في التغير الحاصل في المتغير التابع [18].

- ✓ 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 [18].

يمكن إيجاد معامل التحديد إما من خلال معامل الارتباط أو من خلال ما يلي [18]:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y)^2}{\sum_{i=1}^{n}(y_i - y)^2} \qquad (7 - II)$$

حيث:

$\hat{y}_i$: القيمة التقديرية لمتغير التابع.

y_i: المتغير التابع.

y: متجه البعد n [18].

8.2.II. الاختبار الإحصائية (F):

نستخدم أيضا إحصائية **(F)** للحكم على معنوية النموذج المقدر ككل عند مستوى معنوية معين [16].

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 المفسرة $X_1, X_2, \ldots X_K$ حيث تستخرج قيمة **(F)** كالآتي[23]:

$$F = \frac{\sum_{i=1}^{n}(\hat{y}_i - y)^2 / k}{\sum_{i=1}^{n}(y_i - y)^2 / (n-k-1)} \qquad (8-II)$$

حيث:

$\hat{y}_i$: القيمة التقديرية لمتغير التابع.

y_i: المتغير التابع.

y: متجه البعد n.

n: العدد الإجمالي.

k: عدد طبيعي متغير [23].

ويتم حساب التشتت حول المتوسط الحسابي من خلال ما يلي:

9.2.II. متوسط الانحرافات المطلقة (AAD):

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

$$\sigma = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n} \qquad (9-II)$$

حيث أن:

$\bar{x}$: الوسط الحسابي.

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

n: عدد الفئات [12].

10.2.II. الانحراف المعياري(SD):

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:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (10 - II)$$

حيث أن:

$\bar{x}$: الوسط الحسابي.

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

n: عدد الفئات [12].

3.II. البرامج المستعملة:

مقدمة:

إن برمجيات الكيمياء النظرية تهتم بتطبيق الأسس النظرية والرياضية في حل المسائل الكيميائية، وتعد برمجيات النمذجة الجزيئية إحدى أهم هذه التطبيقات فيمكن من خلال البعض منها معرفة: شكل الجزيئة، تعيين طاقة التنشيط، التوزيع الشحني في الجزيئة وتحديد الجانب الفعال، الألفة الإلكترونية، الخصائص الفيزيوكيميائية.

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

توجد العديد من التقنيات الطيفية التي تستعمل للحصول على معلومات حول البنيات المجهولة للجزئيات الكيميائية ومنها:

- ✓ إن تفسير طيف الأشعة تحت الحمراء يمكنه إعلامنا بوجود أوغياب المجموعات الوظيفية النشطة.
- ✓ طيف الرنين المغناطيسي النووي يمكنه تحديد عدد وطبيعة ومحيط ذرات الهيدروجين لجزيء ما.
- ✓ طيف الكتلة يزودنا بمعلومات حول الكتلة المولية والصيغة الجزيئية ووضعية المجموعات الوظيفية داخل الجزيء.
- ✓ الأطياف الإلكترونية تزودنا بمعلومات حول البنية الإلكترونية للجزيئات.

ومن بين الطرق التجريبية غير الطيفية هي طريقة الانعراج بواسطة الأشعة السينية وهي الأكثر استعمالا لأنها توفر جميع المعطيات الهندسية للجزيء.

من أهم الطرق التي وضعها الكيميائيون النظريون هي الطرق الكمية، وهناك طرق أخرى تسمي بالميكانيكا الجزيئية أوحقل القوة، وتزودنا بمعلومات حول البنية الثلاثية الأبعاد والخصائص الفيزيائية والكيميائية.

هناك بعض الطرق الكمية الأكثر سهولة وهي طريقة هيكل (Huckel) التي تأخذ بعين الاعتبار الإلكترونات (π).

الطرق النصف تجريبية مثل (MNDO, CNDO, PM3, RM1) التي تأخذ بعين الإعتبار الكترونات التكافؤ وتهمل بعض التكاملات وتعوضها ببرامترات موافقة لها.

الطرق الكمية الأكثر تطورا هي (ab-initio) التي تهتم بكل الإلكترونات وتعاني هذه الطريقة من بعض العقبات، والتي يتطلب فيها الحساب وقتا طويلا حتى بالنسبة للجزيئات المتوسطة الحجم. وعلى الرغم من ذلك فهي تستغرق أقل وقت مقارنة ببقية الطرق، وهذا يفسر كثرة استعمالها [19].

II.2.3. برنامج (ChemOffice Ultra):

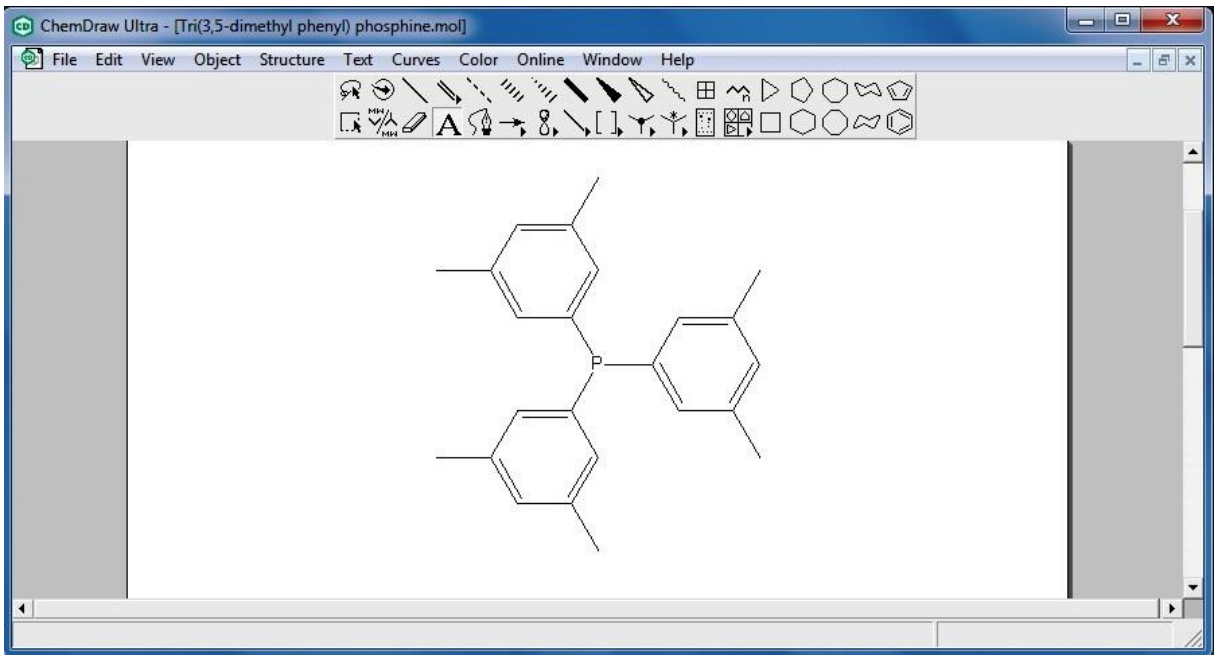
تعريف:

هو عبارة عن أوفيس (Office) خاص بالكيميائيين والعلماء والباحثين والدارسين [2]. ويتميز بالمرونة وسهولة الاستخدام [8]. 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 [2].

ChemOffice Ultra يساعد الكيميائيين والبيولوجيين على الحفاظ على كفاءة مسار عملهم، تصور واكتساب فهم أعمق لنتائجها وربط النشاط البيولوجي مع البني الكيميائية [2].

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.
- ✓ ChemDraw Ultra لبرنامج Excel يضيف الاستخبارات الكيميائية لجداول البيانات إكسل بحيث يستخدم الكيميائيون الـ 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 [2].
- ✓ ChemFinder يستخدمه العلماء لتنظيم مركبات قاعدة بيانات الشخصية مع الخصائص وتحويل البيانات إلى رسوم بيانية سهلة الفهم لتبين بسهولة الهيكل والنشاط المركبات.



الشكل(II-4): البرنامج ChemOffice Ultra.

II.3.3. برنامج (HyperChem):

تعريف:

هو أحد البرامج المتقدمة في مجال النمذجة الجزيئية ذو قدرات حسابية عالية تتعلق بالأنظمة الجزيئية البسيطة والمعقدة، وهو قادر على رسم الجزيئة ببعدين وثلاثة أبعاد وبإحداثيات خطية وصولا إلى أفضل شكل مستقر للجزيئة.

برنامج (HyperChem) عبارة عن تقنية لرسم الجزيئات، وفي هذا الجزء من العمل استعملنا النمذجة الجزيئية التي تتضمن طرق الحساب النظرية (الميكانيكا الجزيئية MM^+ والطرق نصف التجريبية PM3 وRM1) التي تمكننا من تحديد هندسة الجزيئات وتقدير الخصائص الفيزيوكيميائية [22].

II.3.3.1. 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 [10].

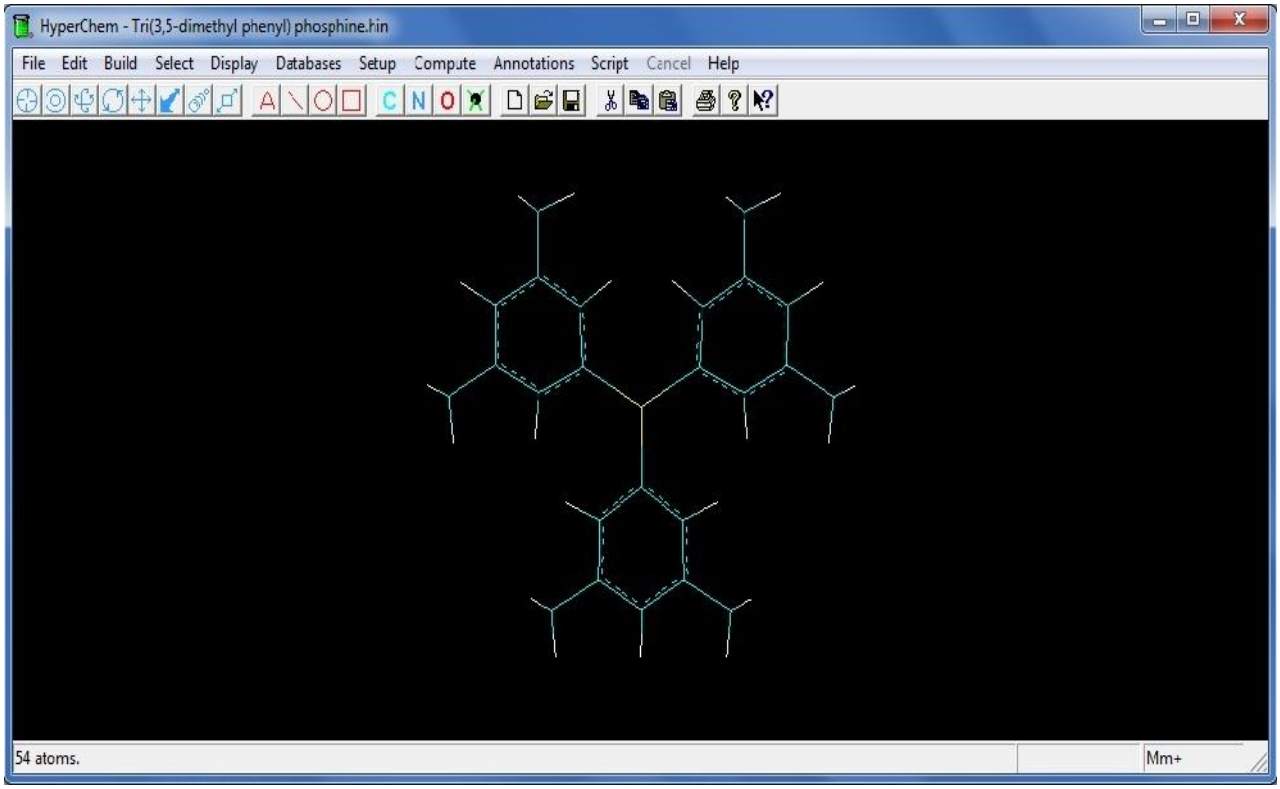
II.3.3.2. طريقة RM1 (Recife Model 1):

هي طريقة نصف تجريبية جديدة وهي تطور لطريقة AM1 ذات نطاق واسع، فهي تقدم المعايير والثوابت الحديثة للخصائص المستخدمة في إجراء البارامترات: مع ارتفاع درجات الحرارة من تشكيل لحظات ثنائي القطب، إمكانات التأين والمتغيرات الهندسية طول الروابط والزوايا [4].

II.3.3.3. طريقة PM3 (Parameterized Model number 3):

هي طريقة حساب كيميائية نصف تجريبية وهي امتداد لطريقة AM1 طريقة نصف تجريبية وهي تحسين لطريقة النصف تجريبية MNDO (Modified Neglect of Diatomic Overlop) إهمال تعديل التداخل ثنائي الذرة ويمكننا بواسطة PM3 حساب الخصائص البنيوية، الطاقة الكلية حرارة التشكل وهي من أدق الطرق النصف تجريبية التي يتضمنها برنامج (HyperChem) [4].

الهدف من هذه الدراسة الكمية والكيفية لقاعدة البيانات هو البحث على احتمالات تواجد هذه المركبات في مستويات امتثالية، على أساس هندسي وطاقوي. إن عرض الخصائص الامتثالية لهذه المركبات يسمى بالتحليل الامتثالي.



الشكل(II-5): البرنامج HyperChem.

II.4.3. برنامج (DRAGON):

تعريف:

برنامج (DRAGON) عبارة عن تقنية مطبقة لحساب الواصفات الجزيئية، أنجز من طرف العالم (Milano)، ومن مجموعة البحث (QSAR/QSPR). هذه الواصفات تستعمل لتقييم الجزئي (النشاط البنيوي أوالخصائص البنيوية) بالإضافة إلى تحليل التشابه وفحص قواعد البيانات الجزيئية [19].

- ✓ وقد تم تصميم (DRAGON) باعتباره برنامج سهل الاستعمال الذي يتولى إجراء العمليات الحسابية للواصف وفقا لتسلسل منطقي بسيط.
- ✓ الواصفات الجزيئية الناتجة عن برنامج (DRAGON) تستعمل لحساب وتقدير الخصائص الفيزيوكيميائية كمعامل التوزيع K_{OW} **مثلا**.
- ✓ إن قدرة برنامج (DRAGON) تمكنه من حساب حتى 250251 جزيئة.
- ✓ الإصدار الأول لبرنامج (DRAGON) يعود إلى سنة 1997، وفي الوقت الراهن أضيفت له عدة واصفات جزيئية جديدة [1].
- ✓ صمم برنامج (DRAGON) للاستعمال في كلا النظامين Windows و Linux حيث أن هناك إصدارات في Windows يسميان (DRAGON Professional) و(DRAGON PLUS) وإصدار واحد في Linux ويدعى (DRAGON X).
- ✓ يستطيع برنامج (DRAGON) حساب عدة أنواع من الواصفات الجزيئية كنوع الذرة، المجموعات الوظيفية وكذلك الواصفات الطوبولوجية والهندسية وبعض خصائص الجزيئات الفيزيوكيميائية مثل $LogK_{ow}$، الشحن الجزيئية، المرونة، الانكسارية، عدد الروابط القابلة للتدوير، السطوح والحجوم الجزيئية، وكذا الأبعاد بين مختلف ذرات الجزيئة.

الشكل(II-6): البرنامج DRAGON [3].

II.2.4.3. تعريف الواصف الجزيئي:

الواصف الجزيئي هو النتيجة النهائية لإجراء رياضي لتحويل المعلومات الكيميائية إلى رموز ومعطيات عديدة، حيث تلعب الواصفات الجزيئية دورا أساسيا في الكيمياء، العلوم الصيدلية، حماية البيئة، الصحة ومراقبة الجودة. اقترح العديد من الواصفات بالاعتماد على النظريات المختلفة يهدف توقع الفعالية والخواص الفيزيوكيميائية للجزيئات [1].

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

1. هناك واصفات جزيئية بسيطة مشتقة لحساب بعض أنواع الذرات أوالأجزاء الهيكلية في الجزيئة.
2. هناك واصفات طوبولوجية أوتسمى الواصفات ثنائية البعد(2D) مشتقة من خوارزميات طوبولوجية.
3. واصفات هندسية وتسمى الواصفات ثلاثية الأبعاد(3D) مشتقة من التمثيل الهندسي.

إن أغلب الواصفات لا تستطيع تمثيل التعقيد الجزئي لـجزيء ما، أو تمثيل كل الاستجابات الفيزيوكيميائية كـنتيجة للـتفاعلات الحيوية، تعتبر معظمها تقريبات لكن يمكن اتخاذها كنماذج لدراستها.

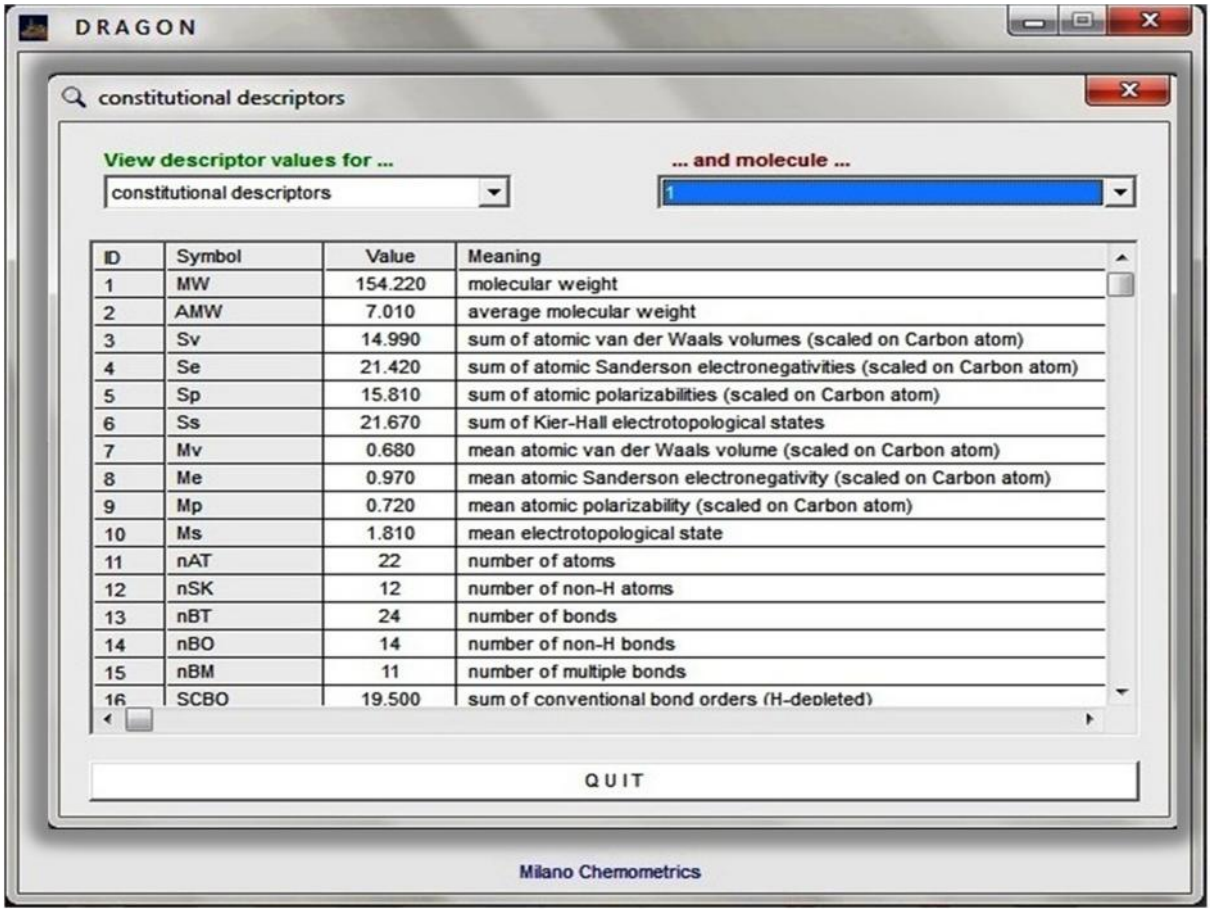
II.3.4.3. كتل الواصف الجزيئي:

يحتوي برنامج (DRAGON) على 22 كتلة منطقية تشمل العديد من الواصفات الجزيئية منها الطوبولوجية، الهندسية، الإلكترونية، والواصفات المتعلقة بالفاعلية البيولوجية والصيدلانية.

إن كتل الواصفات الجزيئية يمكن أن تختار سويا مع الواصفات المهمة.

II.1.4.3. حساب الواصفات الجزيئية:

لحساب الواصفات الجزيئية لمركبات العضوية المشكلة لقاعدة البيانات استعملنا برنامج (DRAGON) الذي يحوي 1664 واصفة مقسمة على 22 كتلة منطقية **(جدول 2)**، كل كتلة منطقية تتضمن مجموعة من الواصفات الخاصة (طوبولوجية، هندسة، إلكترونية، خصائص فيزيوكيميائية...الخ) [1].



ID	Symbol	Value	Meaning
1	MW	154.220	molecular weight
2	AMW	7.010	average molecular weight
3	Sv	14.990	sum of atomic van der Waals volumes (scaled on Carbon atom)
4	Se	21.420	sum of atomic Sanderson electronegativities (scaled on Carbon atom)
5	Sp	15.810	sum of atomic polarizabilities (scaled on Carbon atom)
6	Ss	21.670	sum of Kier-Hall electrotopological states
7	Mv	0.680	mean atomic van der Waals volume (scaled on Carbon atom)
8	Me	0.970	mean atomic Sanderson electronegativity (scaled on Carbon atom)
9	Mp	0.720	mean atomic polarizability (scaled on Carbon atom)
10	Ms	1.810	mean electrotopological state
11	nAT	22	number of atoms
12	nSK	12	number of non-H atoms
13	nBT	24	number of bonds
14	nBO	14	number of non-H bonds
15	nBM	11	number of multiple bonds
16	SCBO	19.500	sum of conventional bond orders (H-depleted)

الشكل(II-7): حساب الواصفات الجزيئية.

الجدول(II-2): أهم كتل الواصفات الجزيئية [19] [1].

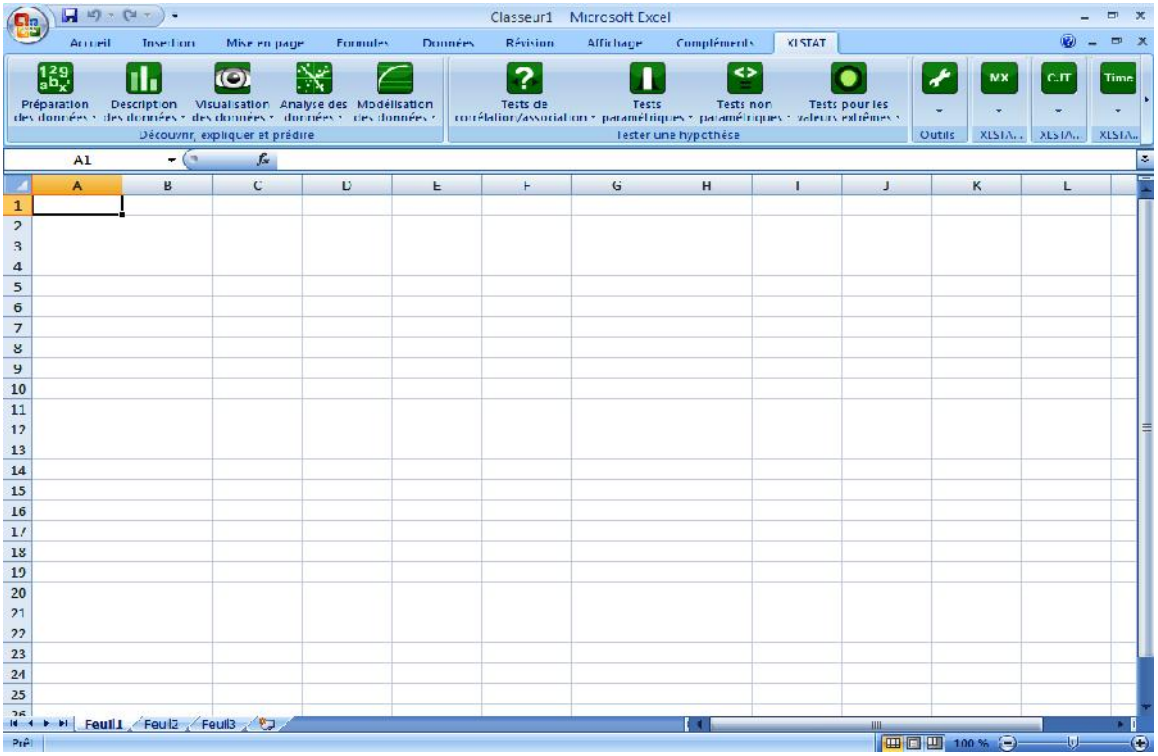
البعد	رقم الكتلة	إسم الكتلة بالإنجليزية	الترجمة إلى العربية	عددها
2D	1	Constitutional descriptors	الواصفات الأساسية	48
	2	Topological descriptors	الواصفات الطوبولوجية	119
	3	Walk and path counts	مسار وطرق العد	47
	4	Connectivity indices	مؤشرات الربط	33
	5	Information indices	مؤشرات المعلومة	47
	6	2D autocorrelations	الارتباط الذاتي 2D	96
	7	Edge adjacency indices	مؤشرات جوار الحافة	107
	8	Burden Eigen value descriptors	واصفات القيم الحرة لـ Burden	64
	9	Topological charge indices	مؤشرات الشحن الطوبولوجية	21
	10	Eigen value-based indices	مؤشرات أساسها قيم حرة	44
	17	Functional group counts	مجاميع المجموعات الوظيفية	154
	18	Atom-centered fragments	أجزاء مراكز الذرة	120
	20	Molecular properties	الخصائص الجزيئية	29
3D	11	Randic molecular profiles	المقاطع الجانبية للجزيئات Randic	41
	12	Geometrical descriptors	الواصفات الهندسية	74
	13	RDF descriptors	واصفات RDF	150
	14	D-MoRSE descriptors3	واصفات 3D-MoRSE	160
	15	WHIM descriptors	واصفات WHIM	99
	16	GETAWAY descriptors	واصفات GETAWAY	197
	19	Charge descriptors	الواصفات الشحنية	14

II.5.3. برنامج (XLSTAT):

XLSTAT هو برنامج يقوم بتحليل البيانات إحصائيا وهو تابع لبرنامج Microsoft Excel. تطوير بفضل Thierry Fahmy، مؤسس Addinsoft. البرنامج متاح على شبكة الإنترنت منذ عام 1996 [6].

نختار هذا البرنامج لأنه يحتوي على العديد من الميزات:

- ✓ XLSTAT هو برنامج موثوق به جدا ومستقر جدا.
- ✓ يمكننا من إعادة تنظيم هيكل قاعدة البيانات وتسهيل تحليلها.
- ✓ وفي XLSTAT نستطيع تحليل المكونات الرئيسية (PCA) وذلك بطريقة الانحدار الخطي (بسيطة أو متعدد).
- ✓ XLSTAT يمكننا من تقييم جودة البيانات، ويحدد أخطاء حسابية تساعدنا على تحسين عملنا.
- ✓ يتضمن XLSTAT بعض الأدوات التي تسهل التصور من البيانات.
- ✓ XLSTAT يقدم تشكيلة واسعة من الاختبارات الإحصائية التي تسمح لنا باختبار الفرضيات [6].



الشكل(II-8): البرنامج XLSTAT.

الجزء التطبيقي

الحساب والنتائج

الفصل الثالث

قاعدة البيانات

III. قاعدة البيانات:

مقدمة:

تتمثل قاعدة البيانات في مجموعة معتبرة من المركبات الكيميائية العضوية وكذلك القيم التجريبية لمعامل التوزيع $LogK_{OW}$ لكل مركب. وهذه القيم تم تحضيرها في المخبر باستعمال طرق خاصة منها طريقة القارورة الهزازة، طريقة الكروماتوغرافيا السائلية العالية الأداء (HPLC)، طريقة التقليب البطيء وطريقة العمود المولد.

1.III. مكونات قاعدة البيانات:

قاعدة البيانات تحتوي على 2471 مركب عضوي يمثل مختلف العائلات الكيميائية (الهيدروكربونات، الأمينات، الأميدات، كحولات، مركبات صيدلانية...)، ولكل مركب قيمة معامل التوزيع (الأكتانول - الماء) $LogK_{OW}$ خاصة به بالإضافة إلى 13 كتلة (**مجموعة**) ثنائية البعد (2D) للواصفات الجزيئية المحسوبة بواسطة برمجية Dragon.

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

يرجع اهتمامنا بالمركبات العضوية للأسباب التالية:

1. كثرة هذه المركبات، حيث يبلغ عدد المركبات العضوية المعروفة لحد اليوم ما يقارب 3 مليون مركب [25]، ويعود ذلك لقدرة ذرات الكربون على الارتباط ببعضها البعض بروابط قوية مكونة لسلاسل جزئية مختلفة الأشكال والأحجام وقدرتها أيضا على الارتباط بذرات أخرى كالأكسجين والنيتروجين والهالوجينات وغيرها [24].
2. أهمية مركبات الكربون (العضوية) بالنسبة للإنسان، فهي تحتل مكانة مهمة في الحياة، فالغذاء والسكن والملابس والأدوية والمنظفات والمبيدات الحشرية وأدوات التجميل مثلا معظمها من المركبات العضوية [25].
3. تميز المركبات العضوية بخصائص فيزيائية وكيميائية خاصة تميزها عن غيرها من المركبات الأخرى.
4. تميز تفاعلات المركبات العضوية حيث لوحظ أن التفاعل في الجزيء العضوي يحدث في جزء معين داخل الجزيء بينما يبقى معظم الجزيء بعيدا عن التفاعل [24].

3.III. عموميات حول المركبات العضوية:

1.3.III. تعريف المركبات العضوية:

تتألف المركبات العضوية من عنصر الكربون كعنصر أساسي ولكن هناك العديد من العناصر الأخرى التي تدخل في تركيبها مثل الهيدروجين والأكسجين والنيتروجين والكبريت والفسفور وغيرها من العناصر [24].

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 للتحول 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 [14].

تحول المركبات العضوية بفعل تأثير الكائنات الحية عموما وعمليات تحلل هذه المركبات بواسطة الأحياء المجهرية الدقيقة كالبكتيريا والفطريات المختلفة هي على رأس طرق الاضمحلال البيولوجي لهذه المركبات في الطبيعة، من حيث كمية المادة المتحولة ودرجة التحول أوالتحلل الذي بلغته. 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 المذكورة آنفا أو بكليهما معا [14].

2.3.III. تصنيفها:

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

1. تصنيف يعتمد على بناء الهيكل الكربوني للمركبات (حيث صنفت إلى مركبات حلقية ومركبات غير حلقية).
2. تصنيف يعتمد على الخواص الفيزيائية (صنفت إلى مركبات عطرية ومركبات دهنية).
3. تصنيف يعتمد على الخواص الكيميائية (صنفت 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) [15].

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

الرقم	LogKow	الإسم	SMILES
1	0.35	Formaldehyde	O=C
2	-1.88	L-Ascorbic acid	O1C(=O)C(O)=C(O)C1C(O)CO
3	0.61	Thiotepa	S=P(N1CC1)(N2CC2)N3CC3
4	4.23	Indomethacin	COc1ccc2n(C(=O)c3ccc(Cl)cc3)c(C)c(CC(=O)O)c2c1
5	1.94	Carbon disulfide	C(=S)=S
6	0.29	Tabun	CCOP(=O)(C≡N)N(C)C
7	3.43	Peroxide, dibenzoyl	O=C(OOC(=O)c(cccc1)c1)c(cccc2)c2
8	3.67	Thioperoxydicarbonic diamide, tetraethyl-	N(C(=S)SSC(N(CC)CC)=S)(CC)CC
9	0.91	Cyclohexanone, oxime	N(O)=C(CCCC1)C1
10	1.22	Carbonic acid, diethyl ester	O=C(OCC)OCC
11	0.08	1,3-Dioxolan-2-one, 4-methyl-	O=C(OCC1C)O1
12	-0.6	Propanedinitrile	N≡CCC≡N
13	1.57	Benzeneethanol,	OCCc(cccc1)c1
14	-1.76	L-Tyrosine,	O=C(O)C(N)Cc(ccc(O)c1)c1
15	1.05	Ethane, 1,1'-oxybis-	O(CC)CC
16	-2.62	Hexanoic acid, 6-amino-	O=C(O)CCCCCN
17	-1	Hydrazine, methyl-	N(N)C
18	-1.16	Acetamide,	O=C(N)C
19	0.28	Dimethoate,	O=C(NC)CSP(OC)(OC)=S
20	4.49	Promethazine,	c1ccc2Sc3ccccc3N(CC(C)N(C)C)c2c1
21	1.17	Niridazole,	O=C1NCCN1c2ncc(s2)N(=O)=O
22	2.06	Zoxazolamine,	c1c(Cl)cc2nc(N)oc2c1
23	-1.59	L-Leucine,	O=C(O)C(N)CC(C)C
24	1.69	Benzoic acid, 4-nitro-	O=C(O)c(ccc(N(=O)=O)c1)c1

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

يحتوي برنامج (DRAGON) على 1664 واصفة جزيئية مقسمة على 22 كتلة، ولحساب هذه الواصفات الجزيئية بواسطة برنامج (DRAGON) نتتبع الخطوات التالية:

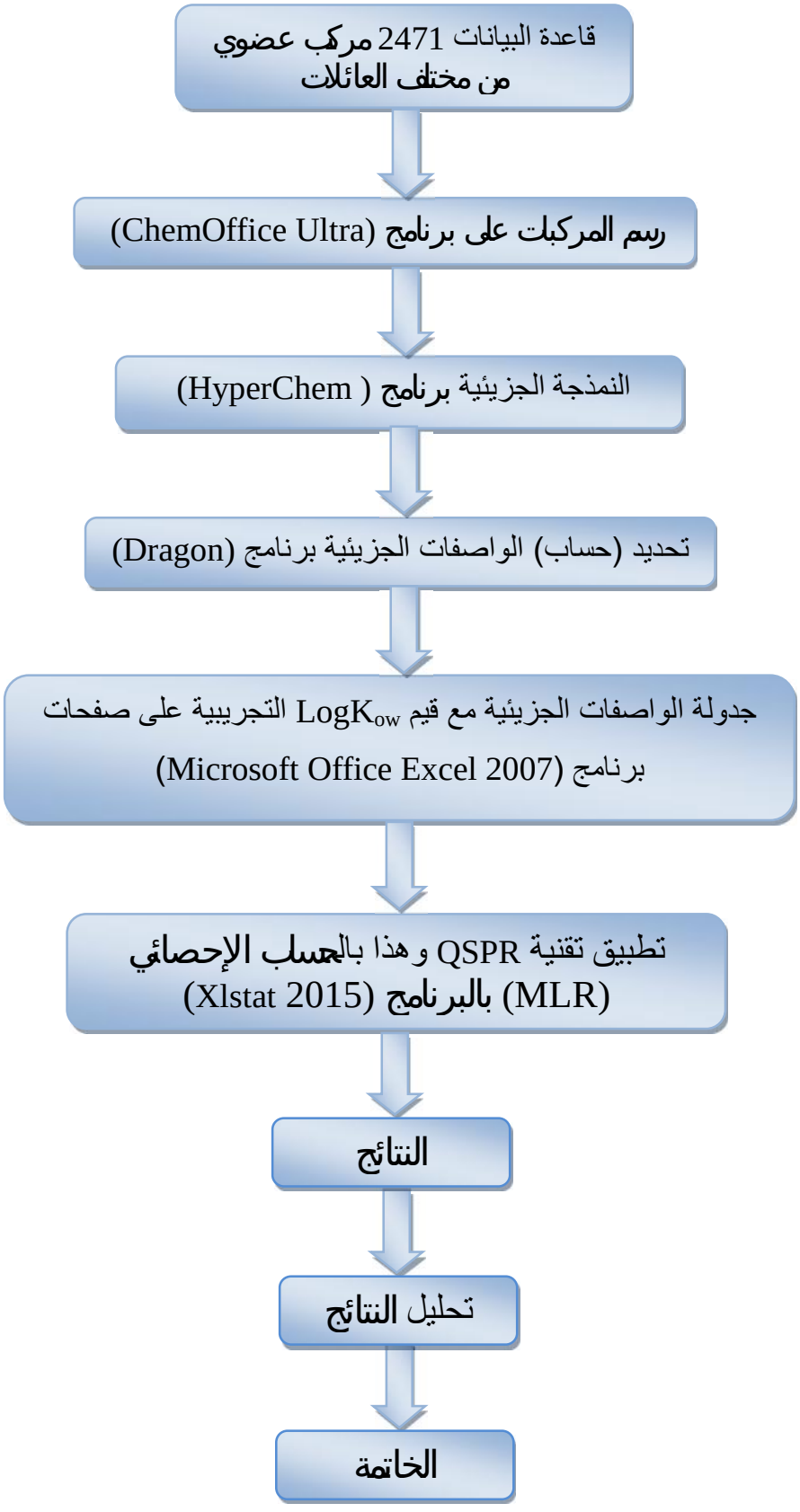
1. نحذف كل الواصفات الجزيئية الثابتة بحيث أنه يمكن أن نجد هناك جزيئات في قاعدة البيانات لها نفس القيم في نفس الواصفة.
2. نحذف كل الواصفات الجزيئية التي تشكل دالة خطية مع الواصفات الجزيئية الأخرى تسمى هذه الطريقة بإزالة الارتباطات الخطية.
3. ضبط قاعدة البيانات: نجري على قاعدة البيانات عملية الضبط (Normalisation) وذلك بحصر البارامترات (الواصفات) بين ($1 \leq 0 \leq -1$)، هذا لتحقيق التجانس بين مختلف القيم، والطريقة هذه تؤدي إلى الخطأ المحتمل.

بعد هذه الخطوات تحصلنا على 592 من 929 واصفة (برا متر Paramètres) [19] وهي موضح كما يلي:

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

الكلي	عدد الواصفات في الكتلة	عدد الواصفات بعد الحساب
الواصفات الأساسية	48	40
الواصفات الطوبولوجية	119	80
مسار وطرق العد	47	15
مؤشرات الربط	33	19
مؤشرات المعلومة	47	30
الإرتباط الذاتي 2D	96	74
مؤشرات جوار الحافة	107	40
واصفات القيم الحرة لـ Burden	64	29
مؤشرات الشحن الطوبولوجية	21	21
مؤشرات أساسها قيم حرة	44	8
مجاميع المجموعات الوظيفية	154	111
أجزاء مراكز الذرة	120	98
الخصائص الجزيئية	29	27
العدد الإجمالي	**929**	**592**

III.5. الخطوات التجريبية للعمل:



الشكل(III-1): الخطوات التجريبية للعمل.

III.1.5. شرح الخطوات التجريبية للعمل:

لإنجاز هذا العمل قمنا بعدة حسابات تجريبية علمية وهذا باستعمال برمجيات مختلفة، وقد تضمن هذا العمل مجموعة من الخطوات والمتمثلة في التسلسل التالي:

1. تم جمع قائمة من المركبات العضوية مع القيم التجريبية للخاصية الفيزيوكيميائية المختارة وهي معامل التوزيع (الأكتانول - الماء) $LogK_{OW}$ من المراجع العلمية وهي التي تمثل قاعدة البيانات وتتكون من 2471 مركب.

2. رسم البنية الجزيئية لهذه المركبات على برنامج ChemDraw Ultra ويحفظ الملف بشكل MDL Molfile (* .mol).
   - وقد أجري الرسم ببرنامج (ChemDraw Ultra 2008).

3. فتح الملف في برنامج النمذجة الجزيئية HyperChem **لحساب** طاقات الاستقرار وذلك بطرق (الميكانيكا الجزيئية MM^+ والطرق نصف التجريبية PM3 وRM1)، ويحفظ الملف بشكل (*.HIN).
   - وقد أجريت النمذجة ببرنامج (HyperChem 8.0).

4. حساب الواصفات الجزيئية باستعمال برنامج Dragon وذلك باختيار 13 كتلة (مجموعة) ثنائية الأبعاد (2D).
   - 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 $LogK_{ow}$.
   - وقد أجريت الجدولة على برنامج (Microsoft Office Excel 2007).

6. تطبيق تقنية QSPR وهذا للحساب الإحصائي بطريقة الانحدار الخطي المتعدد (MLR) باستخدام البرامج الإحصائي Xlstat وهذا لدراسة درجة الإرتباط وحساب قيم معاملات التحديد (R^2) التي تدلنا على درجة الألفة بين القيم التجريبية $LogK_{ow}$ والقيم المحسوبة (النظرية) لـ $LogK_{ow}$ بالإضافة إلى نوعية وشكل معادلة النموذج. إن معامل التحديد يأخذ القيم بين (0، 1) حيث كلما اقترب معامل التحديد إلى 1 يعني هذا وجود علاقة جيدة بين الخاصية المختارة والواصف، وذلك يؤدي إلى الوصول لنموذج أفضل.

- وقد أجريت الحسابات ببرنامج (Xlstat 2015).

7. تم تصميم معادلة النموذج الرياضي الخاص لكل كتلة على حد والذي يربط بين الخاصية المختارة ومجموعة الواصفات الجزيئية المرتبطة بها، وكذلك معادلة خاصة بجميع الكتل معا.

8. يتم اختيار أفضل نموذج رياضي للانحدار الخطى على أساس أعلى قيمة لمعامل التحديد (R^2)، وكذلك على أقل عدد من الواصفات الجزيئية المتشكلة لمعادلة النموذج.

الفصل الرابع

النتائج والمناقشة

IV. النتائج:

مقدمة:

في هذا العمل أجرينا مرحلتين لدراسة 13 كتلة (مجموعة) من الواصفات الجزئية ثنائية البعد (2D) **باستعمال** 2471 مركب كيميائي عضوي، بتطبيق تقنية QSPR وهذا للحساب الإحصائي بطريقة الانحدار الخطي المتعدد (MLR) خطوة خطوة (step Wise) في البرنامج الإحصائي Xlstat. تم بناء (حساب) النموذج باستعمال 1851 مركب و600 مركب للتحقق الداخلي (Validation 30%)، أما بالنسبة للتحقق الخارجي لـ (test) استعملنا 20 مركبا باختيار عشوائي.

- ✓ **المرحلة الأولى:** قمنا بحساب كل كتلة على حد.
- ✓ **المرحلة الثانية:** قمنا بحساب كل الكتل معا (مجمعة).

1.IV. نتائج المرحلة الأولى: سجلت النتائج على الجدول التالي:

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

البعد	اسم الكتلة بالإنجليزية	الترجمة إلى العربية	R^2	S^2	σ
2D	Constitutional descriptors	الواصفات الأساسية	0.6311	0.98	0.00021
	Topological descriptors	الواصفات الطوبولوجية	0.5919	1.07	0.00026
	Walk and path counts	مسار وطرق العد	0.2177	1.44	0.00031
	Connectivity indices	مؤشرات الربط	0.2661	1.39	0.00012
	Information indices	مؤشرات المعلومات	0.4219	1.26	0.0002
	2D autocorrelations	الارتباط الذاتي 2D	0.4316	1.22	0.00073
	Topological charge indices	مؤشرات الشحن الطوبولوجية	0.1975	1.49	0.00018
	Edge adjacency indices	مؤشرات أساسها قيم حرة	0.3394	1.46	0.00026
	Functional group counts	مجاميع المجموعات الوظيفية	0.7541	0.79	0.00015
	Atom-centered fragments	أجزاء مراكز الذرة	0.8556	0.63	0.00039
	Burden Eigen value descriptors	واصفات القيم الحرة لـ Burden	0.4247	1.25	0.00044
	Eigen value-based indices	مؤشرات جوار الحافة	0.1872	1.32	0.00049
	Molecular properties	الخصائص الجزيئية	0.7854	0.76	0.00053

من الجدول السابق نلاحظ أن الكتلة أجزاء مراكز الذرة (Atom-centered fragments) لها أكثر قيمة لمعامل التحديد R^2 =0,8556 وأقل قيمة لـ S^2 = 0.6286.

IV.1.1. الكتلة أجزاء مراكز الذرة (Atom-centered fragments):

معادلة النموذج:

يهتم نموذج الانحدار الخطي المتعدد بتقدير 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:

$$Logk_{ow} = a + \Sigma (b_j \times x_j)$$

حيث:

a: قيمة الثابت.

b_j: قيمة المعامل.

x_j: الواصفة.

$$LogK_{ow} = a - b_1\times\textbf{C-001} - b_2\times\textbf{C-003} - b_3\times\textbf{C-006} + b_4\times\textbf{C-011} + b_5\times\textbf{C-013} + b_6\times\textbf{C-014} + b_7\times\textbf{C-015} - b_8\times\textbf{C-017} - b_9\times\textbf{C-018} + b_{10}\times\textbf{C-019} + b_{11}\times\textbf{C-020} + b_{12}\times\textbf{C-024} + b_{13}\times\textbf{C-025} + b_{14}\times\textbf{C-026} + b_{15}\times\textbf{C-027} + b_{16}\times\textbf{C-028} + b_{17}\times\textbf{C-029} + b_{18}\times\textbf{C-032} + b_{19}\times\textbf{C-034} + b_{20}\times\textbf{C-035} - b_{21}\times\textbf{C-036} + b_{22}\times\textbf{C-038} + b_{23}\times\textbf{C-041} + b_{24}\times\textbf{C-044} + b_{25}\times\textbf{H-046} - b_{26}\times\textbf{H-049} - b_{27}\times\textbf{H-050} - b_{28}\times\textbf{H-051} + b_{29}\times\textbf{H-052} + b_{30}\times\textbf{H-053} + b_{31}\times\textbf{H-054} - b_{32}\times\textbf{O-056} - b_{33}\times\textbf{O-058} - b_{34}\times\textbf{O-059} + b_{35}\times\textbf{O-060} + b_{36}\times\textbf{O-061} + b_{37}\times\textbf{Se-064} - b_{38}\times\textbf{N-066} - b_{39}\times\textbf{N-067} - b_{40}\times\textbf{N-068} - b_{41}\times\textbf{N-069} - b_{42}\times\textbf{N-072} + b_{43}\times\textbf{N-073} - b_{44}\times\textbf{N-075} - b_{45}\times\textbf{N-076} - b_{46}\times\textbf{N-077} + b_{47}\times\textbf{N-078} - b_{48}\times\textbf{N-079} + b_{49}\times\textbf{F-085} + b_{50}\times\textbf{Cl-086} + b_{51}\times\textbf{Cl-087} + b_{52}\times\textbf{Cl-088} + b_{53}\times\textbf{Cl-089} + b_{54}\times\textbf{Br-091} + b_{55}\times\textbf{Br-093} + b_{56}\times\textbf{Br-094} - b_{57}\times\textbf{Br-095} + b_{58}\times\textbf{I-096} + b_{59}\times\textbf{I-099} - b_{60}\times\textbf{I-100} + b_{61}\times\textbf{S-106} + b_{62}\times\textbf{S-107} + b_{63}\times\textbf{S-108} - b_{64}\times\textbf{S-109}$$

الجدول (IV-2): معاملات التعديل الخاصة بالكتلة (أجزاء مراكز الذرة).

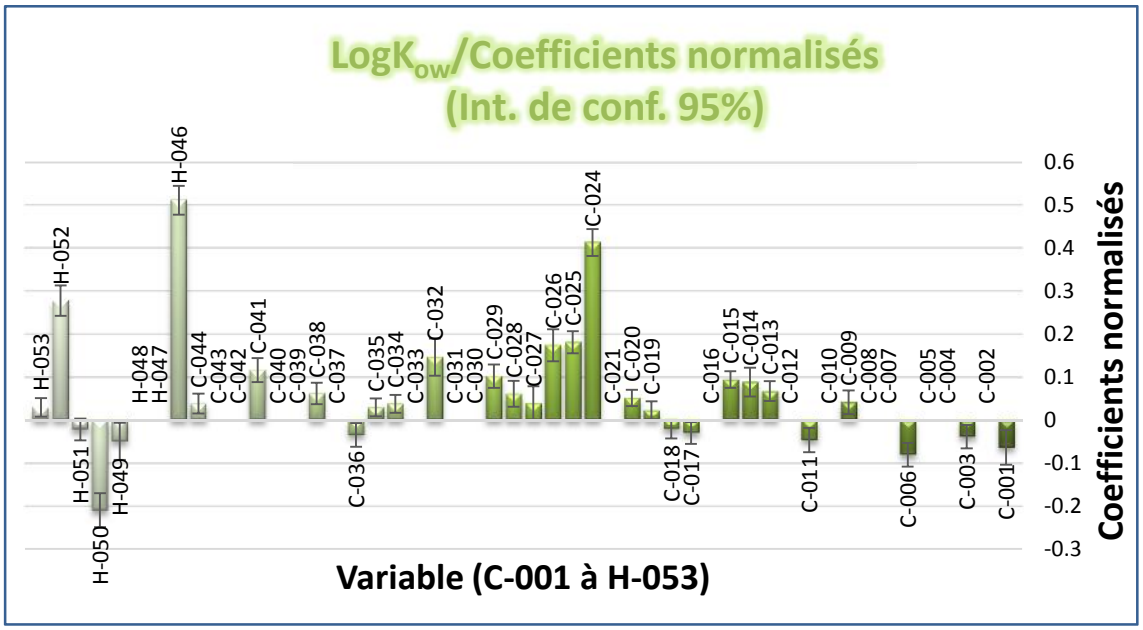
Observations	Somme des poids		DDL	R^2	R^2 ajusté	MCE
1693.000	1693.000		1627.000	0.856	0.850	0.396
RMCE	**MAPE**	**DW**	**Cp**	**AIC**	**SBC**	**PC**
0.630	74.552	1.681	99.129	-1502.749	-1143.088	0.156

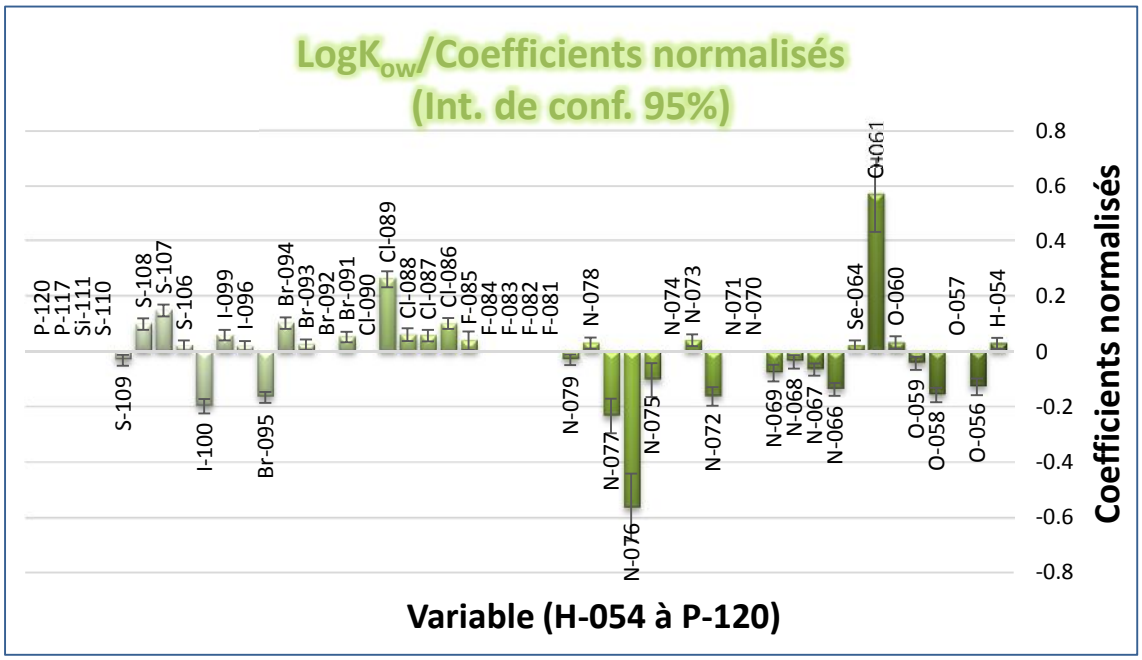
الجدول (IV-3): ضبط معاملات نموذج الكتلة (أجزاء مراكز الذرة).

الواصف	المعامل	القيمة	الخطأ النوعي	الواصف	المعامل	القيمة	الخطأ النوعي
الثابت	a	0.460	0.050	O-058	b_{33}	-0.257	0.022
C-001	b_1	-0.092	0.030	O-059	b_{34}	-0.132	0.039
C-003	b_2	-0.142	0.051	O-060	b_{35}	0.101	0.034
C-009	b_3	0.316	0.107	O-061	b_{36}	1.309	0.156
C-011	b_4	-0.278	0.087	Se-064	b_{37}	0.846	0.368
C-013	b_5	0.652	0.114	N-066	b_{38}	-1.130	0.099
C-014	b_6	1.303	0.251	N-067	b_{39}	-0.375	0.066
C-015	b_7	0.831	0.087	N-068	b_{40}	-0.185	0.063
C-017	b_8	-0.198	0.087	N-069	b_{41}	-0.307	0.061
C-018	b_9	-0.186	0.098	N-072	b_{42}	-0.377	0.040
C-019	b_{10}	0.125	0.059	N-073	b_{43}	0.435	0.112
C-020	b_{11}	1.044	0.195	N-075	b_{44}	-0.280	0.086
C-024	b_{12}	0.242	0.009	N-076	b_{45}	-2.646	0.291
C-025	b_{13}	0.329	0.024	N-077	b_{46}	-2.385	0.327
C-026	b_{14}	0.237	0.026	N-078	b_{47}	0.159	0.052
C-027	b_{15}	0.156	0.080	N-079	b_{48}	-1.108	0.374
C-028	b_{16}	0.281	0.070	F-085	b_{49}	0.182	0.077
C-029	b_{17}	0.592	0.080	Cl-086	b_{50}	1.007	0.099
C-032	b_{18}	0.788	0.119	Cl-087	b_{51}	0.467	0.086
C-034	b_{19}	0.321	0.090	Cl-088	b_{52}	0.350	0.067
C-035	b_{20}	0.388	0.140	Cl-089	b_{53}	0.640	0.036
C-036	b_{21}	-0.358	0.147	Br-091	b_{54}	0.784	0.141
C-038	b_{22}	0.327	0.067	Br-093	b_{55}	0.564	0.215
C-041	b_{23}	0.460	0.056	Br-094	b_{56}	0.659	0.067
C-044	b_{24}	0.510	0.156	Br-095	b_{57}	-2.865	0.171
H-046	b_{25}	0.192	0.006	I-096	b_{58}	0.930	0.448
H-049	b_{26}	-0.160	0.071	I-099	b_{59}	0.984	0.156
H-050	b_{27}	-0.257	0.025	I-100	b_{60}	-2.613	0.176
H-051	b_{28}	-0.027	0.016	S-106	b_{61}	0.659	0.289
H-052	b_{29}	0.183	0.012	S-107	b_{62}	0.673	0.049
H-053	b_{30}	0.113	0.042	S-108	b_{63}	0.777	0.082
H-054	b_{31}	0.275	0.095	S-109	b_{64}	-0.464	0.149
O-056	b_{32}	-0.336	0.041				

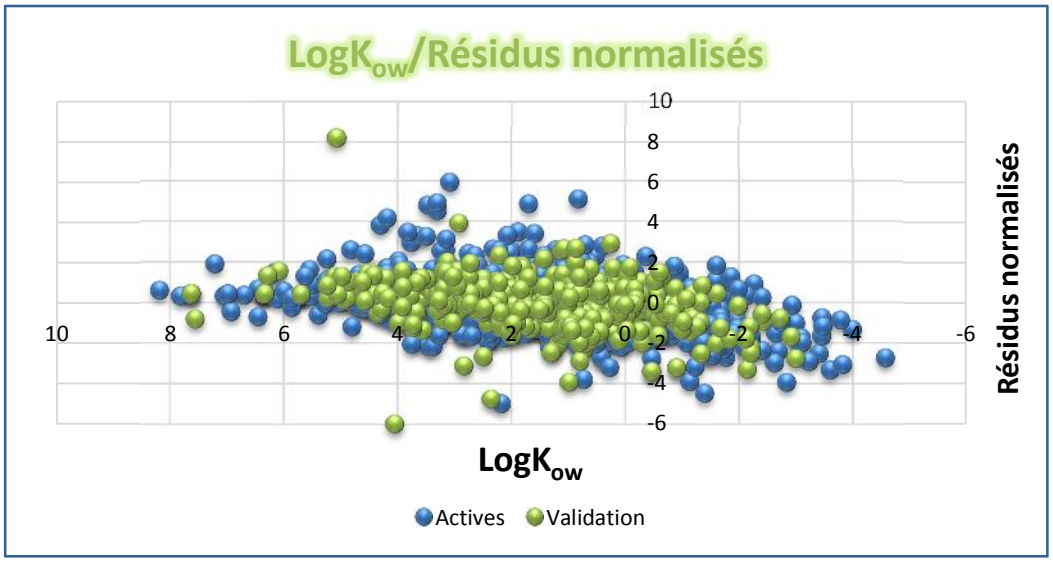
الجدول (IV- 4): تحليل الفروق الخاصة بالكتلة (أجزاء مراكز الذرة).

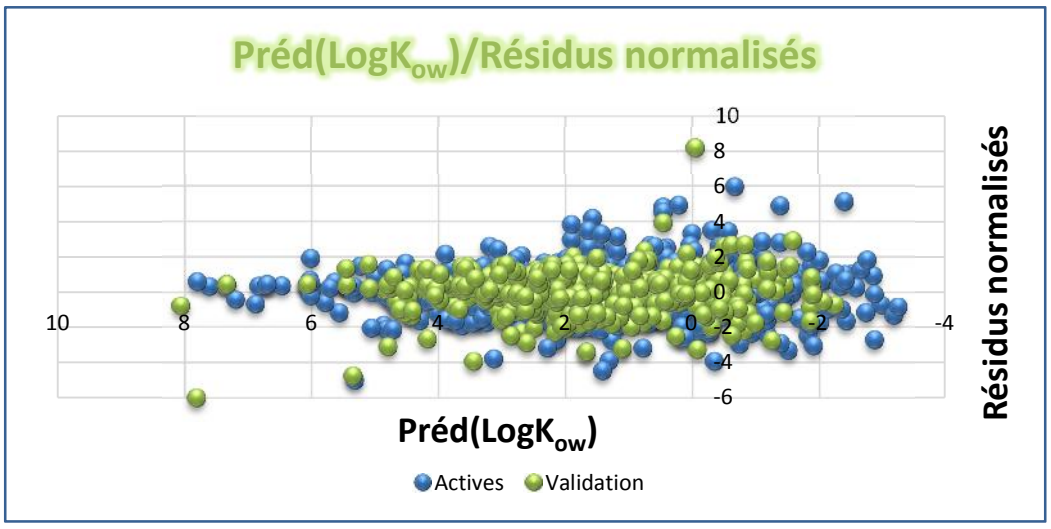
مصدر	DDL	مجموع المربعات	متوسط المربعات	F	Pr > F
النموذج	65	3822.179	58.803	148.328	< 0.0001
الخطأ	1627	645.003	0.396		
مجموع التصحيح	1692	4467.182			

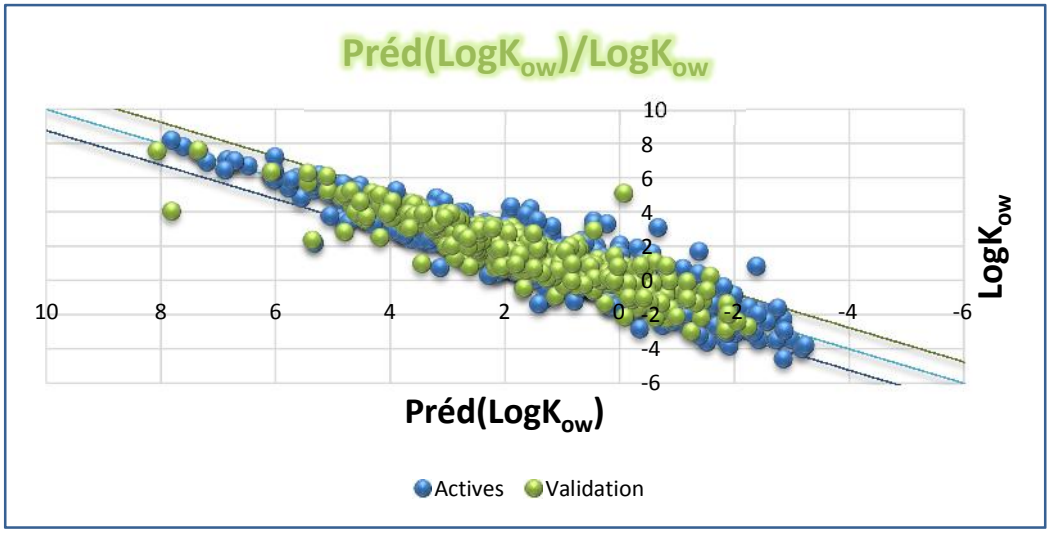




الشكل (IV-1): بيان يوضح أسماء معاملات الواصفات الجزيئية الخاصة بنموذج الكتلة (أجزاء مراكز الذرة).







الشكل(IV-2): بيان يوضح تغيرات LogKow بدلالة Préd(LogKow) الخاصة بكتلة (أجزاء مراكز الذرة).

2.1.IV. تحليل نتائج المرحلة الأولى:

معامل التحديد **(R^2)** يعطينا فكرة عن النسبة المئوية لتغير المتغير المنمذج موضح بالمتغيرات التفسيرية .حيث كلما اقترب معامل التحديد إلى 1، كلما كان النموذج أفضل.

في هذه الحالة لدينا **R^2 = 0.855612982801114** و **R^2 ajusté = 0.849844601659179** وبالتالي **فإن** النسبة المئوية لتغير المتغير المنمذج **= 85.56%**.

هذا يبين لنا أن النسبة المئوية **85.56%** من تغير $LogK_{OW}$ يفسر بواسطة الواصفات الجزيئية. والباقي من هذه النسبة يرجع إلى تأثير المتغيرات التفسيرية الأخرى التي لم تأخذ بعين الاعتبار أوالواصفات الجزيئية المختارة ليست وصفية بشكل كامل.

3.1.IV. تحليل البقايا (Residues) نتائج المرحلة الأولى:

Residue = 1- R^2 = 1 - 0.8556 = 0.1444 = 100 × 0.1444 =14.44%

هذا يعني أن **95%** من الباقي توجد في المجال (+2، -2) كل قيمة خارج هذا المجال غير موثوق فيها. المتغيرات التي تسمح بتفسير **85.56%** من تغير $LogK_{OW}$ تفسير النسبة المئوية للبقايا وهي **14.44% من** عدة عوامل تفسيرية أخرى.

4.1.IV. مناقشة النتائج للمرحلة الأولى:

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

- الواصفات الجزيئية الأقل ارتباطا:

(C-001, C-024) = - 0.31

(H-046, C-024) = - 0.28

(C-002, C-024) = - 0.27

الجدول (IV-5): الواصفة ورمزها للواصفات الجزيئية الأقل ارتباطا (أجزاء مراكز الذرة).

رمز الواصفة	الواصفة
C-001	CH3R / CH4
C-024	R--CH--R
C-002	CH2R2
H-046	H^a attached to C^0(sp3) no X attached to next C

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

- الواصفات الجزيئية الأكثر ارتباطا:

(O-061, N-076) = 0.87

(C-014, F-085) = 0.83

(C-027, H-049) = 0.80

الجدول (IV-6): الواصفة ورمزها للواصفات الجزيئية الأكثر ارتباطا (أجزاء مراكز الذرة).

رمز الواصفة	الواصفة
O-061	c--O
N-076	RO-NO / Ar-NO2 / R--N(--R)--Of
C-014	CX4
F-085	Fa attached to C2(sp2)-C4(sp2) / C1(sp) / C4(sp3) / X
C-027	R--CH--X
H-049	Ha attached to C3(sp3) / C2(sp2) / C3(sp2) / C3(sp)

الرسم البياني **(pérd ($logk_{ow}$)/$logk_{ow}$)** يبين لنا أن هناك **95%** من جزيئات قاعدة البيانات داخل مجال الثقة. ووفقا للتحليل الإحصائي للقيم (**$logk_{ow\ pred}$, $logk_{ow\ exp}$**) ونتائج المقارنة بين الكتل **يمكن ملاحظة** أن: يوجد قيم **موجودة خارج المجال** (-2 ، +2) هذه القيم تمثل النقاط الشاذة (جزيئات غير مرتبطة).

متغيرات الكتلة (**أجزاء مراكز الذرة**) تعطي تقريبا التمثيل المناسب للبنية الجزيئية قاعدة البيانات المستخدمة.

2.IV. نتائج المرحلة الثانية: وسجلت في الجدول التالي:

الجدول (IV-7): نتائج جميع الكتل المدروسة معا.

البعد	اسم الكتلة بالإنجليزية	الترجمة إلى العربية	R^2	S^2	σ
2D	All descriptors blokes together	كل الكتل معا	0.689	0.97	0.00041

معادلة النموذج:

$$\mathbf{LogK_{ow}} = a + b_1\times\mathbf{ATS1v} + b_2\times\mathbf{ATS2e} + b_3\times\mathbf{MATS1m} + b_4\times\mathbf{MATS5m} + b_5\times\mathbf{MATS1v} + b_6\times\mathbf{MATS5v} + b_7\times\mathbf{MATS1e} + b_8\times\mathbf{MATS2e} + b_9\times\mathbf{MATS3e} + b_{10}\times\mathbf{MATS8p} + b_{11}\times\mathbf{GATS8m} + b_{12}\times\mathbf{GATS1v} + b_{13}\times\mathbf{GATS8v} + b_{14}\times\mathbf{GATS1e} + b_{15}\times\mathbf{GATS2e} + b_{16}\times\mathbf{GATS5e} + b_{17}\times\mathbf{C\text{-}001} - b_{18}\times\mathbf{C\text{-}006} + b_{19}\times\mathbf{C\text{-}014} + b_{20}\times\mathbf{C\text{-}020} + b_{21}\times\mathbf{C\text{-}025} + b_{22}\times\mathbf{C\text{-}027} + b_{23}\times\mathbf{C\text{-}044} + b_{24}\times\mathbf{O\text{-}058} + b_{25}\times\mathbf{N\text{-}067} - b_{26}\times\mathbf{Cl\text{-}087} - b_{27}\times\mathbf{Cl\text{-}090} - b_{28}\times\mathbf{Br\text{-}095} - b_{29}\times\mathbf{I\text{-}100} + b_{30}\times\mathbf{S\text{-}110} + b_{31}\times\mathbf{Si\text{-}111} + b_{32}\times\mathbf{BEHm1} - b_{33}\times\mathbf{BEHm3} - b_{34}\times\mathbf{BEHp1} - b_{35}\times\mathbf{X2A} - b_{36}\times\mathbf{X3A} + b_{37}\times\mathbf{X4Av} + b_{38}\times\mathbf{Me} + b_{39}\times\mathbf{RBN} - b_{40}\times\mathbf{nN} + b_{41}\times\mathbf{nX} - b_{42}\times\mathbf{nR08} + b_{43}\times\mathbf{nR09} + b_{44}\times\mathbf{Eig1Z} + b_{45}\times\mathbf{nCconj} + b_{46}\times\mathbf{nArNCS} + b_{47}\times\mathbf{nArCOOH} - b_{48}\times\mathbf{nRCO} + b_{49}\times\mathbf{nCONN} + b_{50}\times\mathbf{nC(=N)N2} - b_{51}\times\mathbf{nRNH2} + b_{52}\times\mathbf{nN=N} + b_{53}\times\mathbf{nRNNOx} - b_{54}\times\mathbf{nROH} - b_{55}\times\mathbf{nOHs} - b_{56}\times\mathbf{nSO2OH} - b_{57}\times\mathbf{nCXr=} - b_{58}\times\mathbf{nOxiranes} - b_{59}\times\mathbf{LAI} - b_{60}\times\mathbf{Inflammat\text{-}50} - b_{61}\times\mathbf{GGI1} - b_{62}\times\mathbf{JGI3} + b_{63}\times\mathbf{PW2} - b_{64}\times\mathbf{PJI2} - b_{65}\times\mathbf{D/Dr03} + b_{66}\times\mathbf{D/Dr10} + b_{67}\times\mathbf{T(S,,Cl)} - b_{68}\times\mathbf{T(P,,Br)} + b_{69}\times\mathbf{T(Cl,,Cl)}$$

الجدول (IV-8): معاملات التعديل الخاصة بجميع الكتل معا.

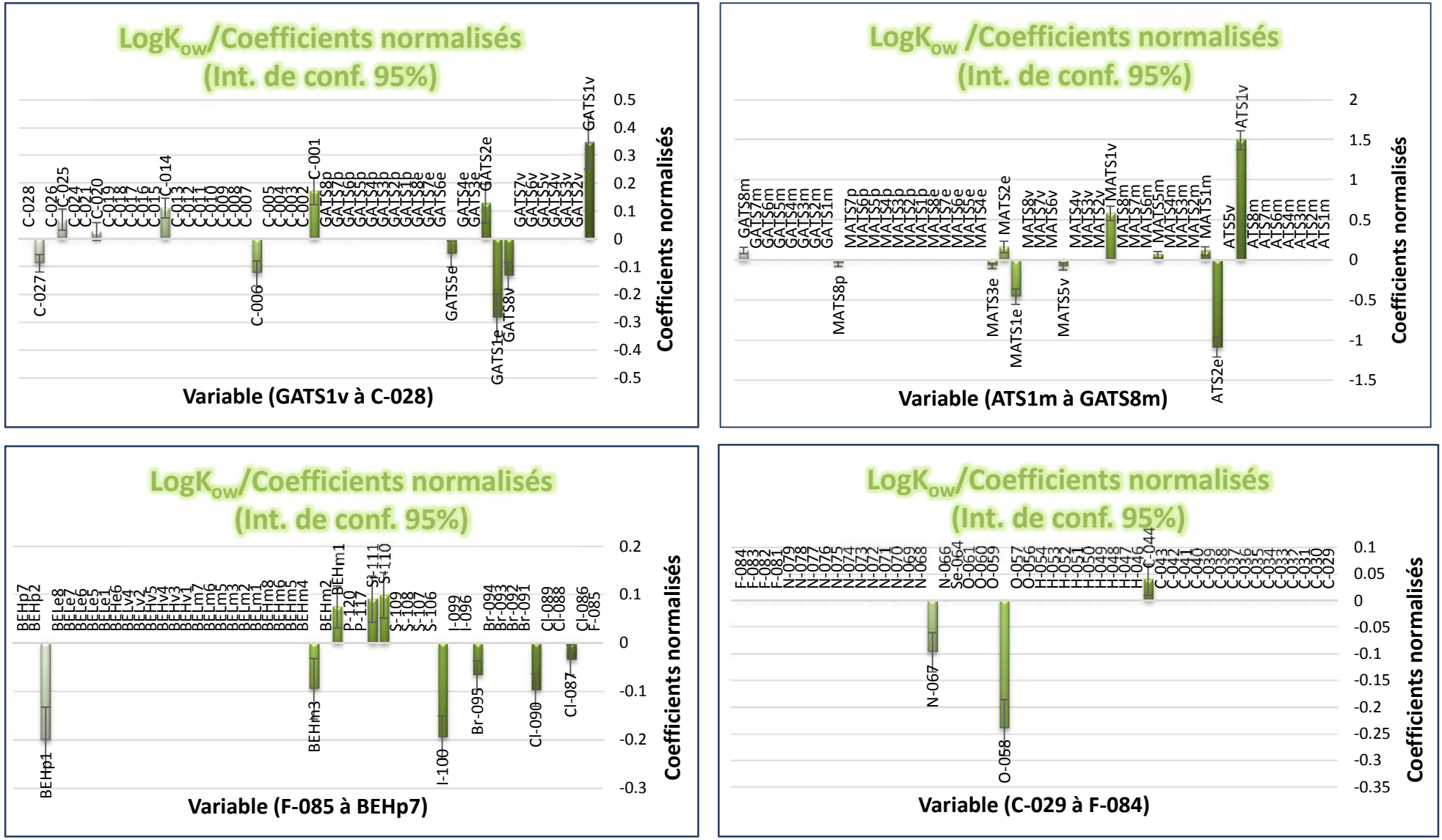
Observations	Somme des poids		DDL	R^2	R^2 ajusté	MCE
1589.000	1589.000		1519.000	0.689	0.675	0.867
RMCE	**MAPE**	**DW**	**Cp**	**AIC**	**SBC**	**PC**
0.931	115.423	2.037	202.572	-158.465	217.495	0.340

الجدول (IV-9): تحليل الفروق الخاص بجميع الكتل معا.

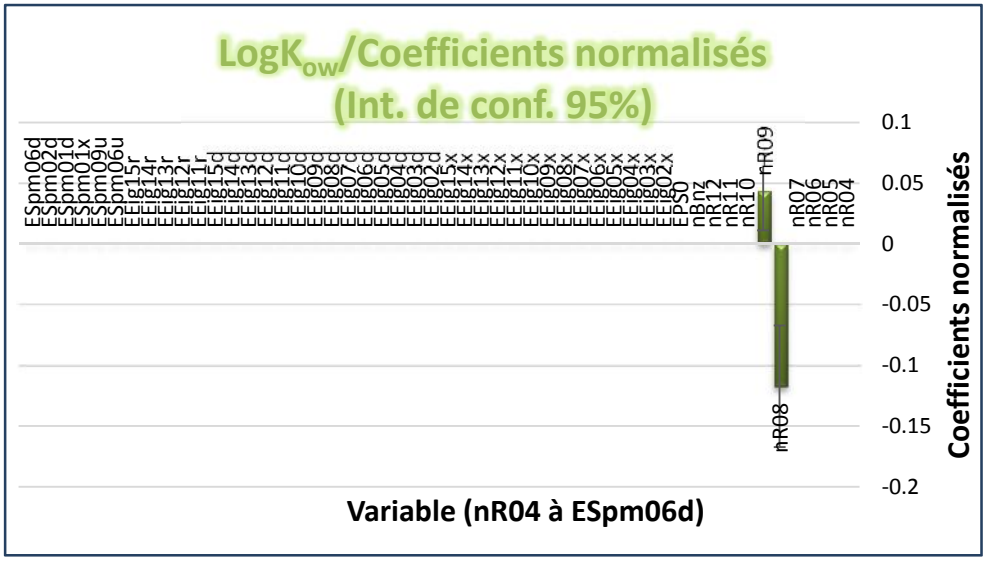
مصدر	DDL	مجموع المربعات	متوسط المربعات	F	Pr > F
النموذج	69	2914.504	42.239	48.722	< 0.0001
الخطأ	1519	1316.890	0.867		
مجموع التصحيح	1588	4231.394			

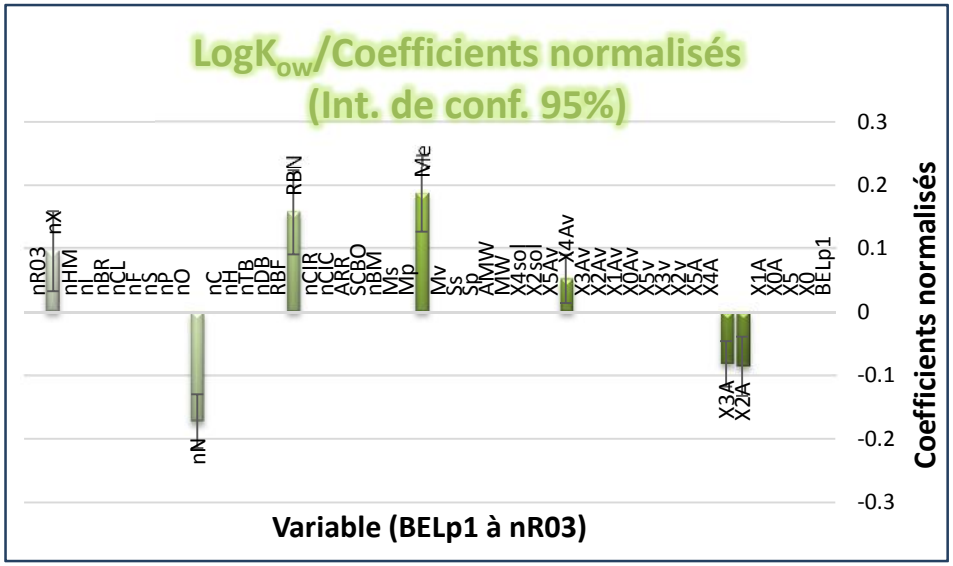
الجدول (IV-10): ضبط معاملات النموذج الخاص بجميع الكتل معا.

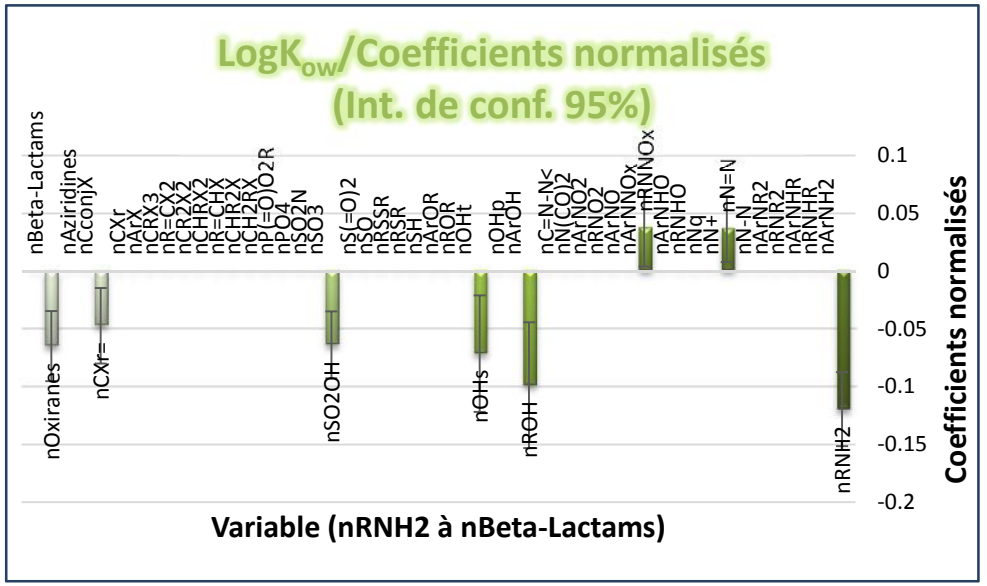
الواصف	المعامل	القيمة	الخطأ النوعي	الواصف	المعامل	القيمة	الخطأ النوعي
ثابت	a	-6.258	1.720	X2A	b_{35}	-2.306	0.643
ATS1v	b_1	5.187	0.208	X3A	b_{36}	-2.172	0.485
ATS2e	b_2	-2.991	0.173	X4Av	b_{37}	1.019	0.379
MATS1m	b_3	0.665	0.164	Me	b_{38}	8.120	1.330
MATS5m	b_4	0.241	0.084	RBN	b_{39}	0.114	0.025
MATS1v	b_5	2.699	0.211	nN	b_{40}	-0.215	0.027
MATS5v	b_6	-0.263	0.077	nX	b_{41}	0.149	0.050
MATS1e	b_7	-2.511	0.266	nR08	b_{42}	-0.941	0.205
MATS2e	b_8	0.853	0.190	nR09	b_{43}	0.280	0.106
MATS3e	b_9	-0.358	0.085	Eig1Z	b_{44}	0.007	0.002
MATS8p	b_{10}	-0.191	0.079	nCconj	b_{45}	0.068	0.029
GATS8m	b_{11}	0.187	0.031	nArNCS	b_{46}	1.315	0.475
GATS1v	b_{12}	1.522	0.210	nArCOOH	b_{47}	0.290	0.145
GATS8v	b_{13}	-0.287	0.053	nRCO	b_{48}	-0.328	0.099
GATS1e	b_{14}	-1.251	0.195	nCONN	b_{49}	0.217	0.105
GATS2e	b_{15}	0.705	0.198	nC(=N)N2	b_{50}	0.821	0.223
GATS5e	b_{16}	-0.111	0.049	nRNH2	b_{51}	-0.938	0.128
C-001	b_{17}	0.261	0.037	nN=N	b_{52}	1.369	0.546
C-006	b_{18}	-0.206	0.038	nRNNOx	b_{53}	0.327	0.148
C-014	b_{19}	1.672	0.270	nROH	b_{54}	-0.242	0.068
C-020	b_{20}	0.519	0.304	nOHs	b_{55}	-0.327	0.118
C-025	b_{21}	0.124	0.034	nSO2OH	b_{56}	-2.398	0.555
C-027	b_{22}	-0.309	0.056	nCXr =	b_{57}	-0.507	0.179
C-044	b_{23}	0.708	0.269	nOxiranes	b_{58}	-2.986	0.713
O-058	b_{24}	-0.421	0.047	LAI	b_{59}	-1.172	0.543
N-067	b_{25}	-0.613	0.119	Inflammat-50	b_{60}	-0.361	0.139
Cl-087	b_{26}	-0.286	0.133	GGI1	b_{61}	-0.135	0.041
Cl-090	b_{27}	-0.693	0.122	JGI3	b_{62}	-5.010	1.406
Br-095	b_{28}	-1.624	0.368	PW2	b_{63}	4.784	1.061
I-100	b_{29}	-2.603	0.303	PJI2	b_{64}	-0.340	0.139
S-110	b_{30}	0.548	0.133	D/Dr03	b_{65}	-0.012	0.004
Si-111	b_{31}	1.788	0.480	D/Dr10	b_{66}	0.005	0.001
BEHm1	b_{32}	0.082	0.024	T(S,,Cl)	b_{67}	0.062	0.014
BEHm3	b_{33}	-0.304	0.103	T(P,,Br)	b_{68}	-0.304	0.115
BEHp1	b_{34}	-1.327	0.229	T(Cl,,Cl)	b_{69}	0.016	0.005

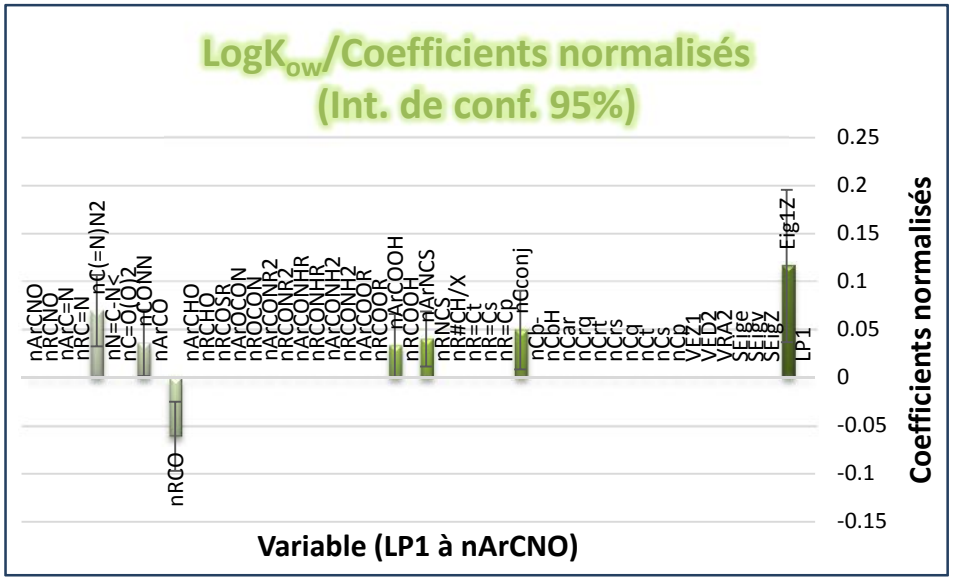


الشكل (3-IV): بيان يوضح أسماء الواصفات الجزيئية الخاصة بنموذج جميع الكتل **معا**.

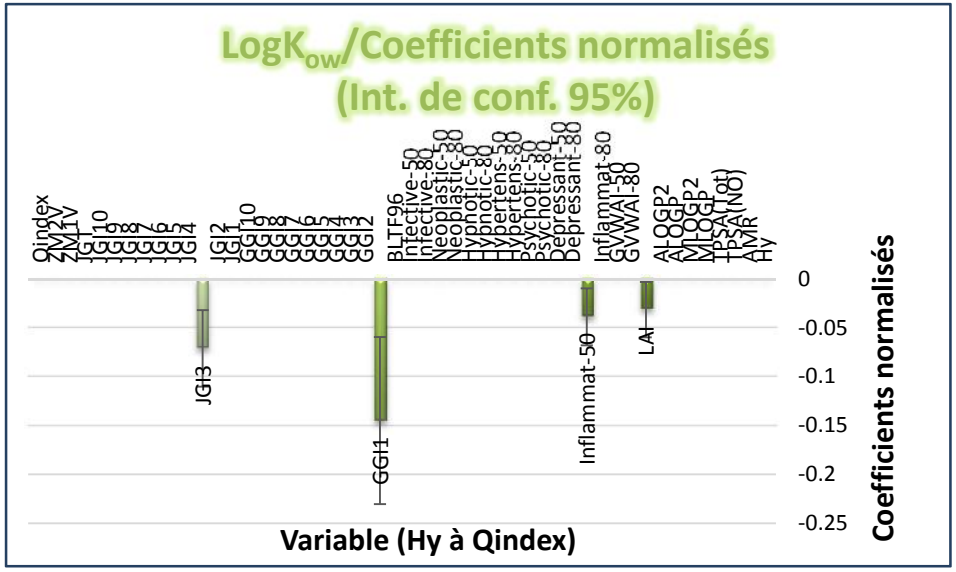


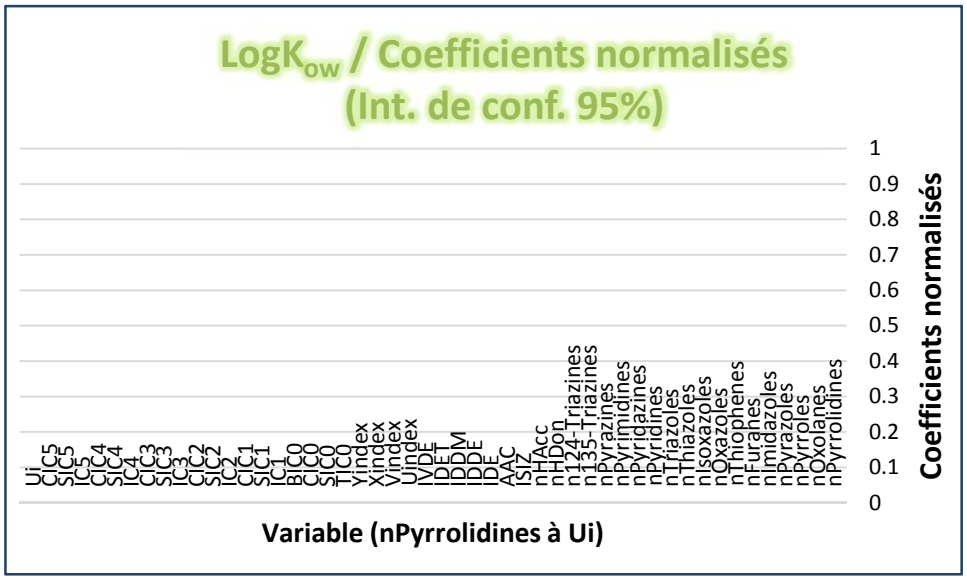


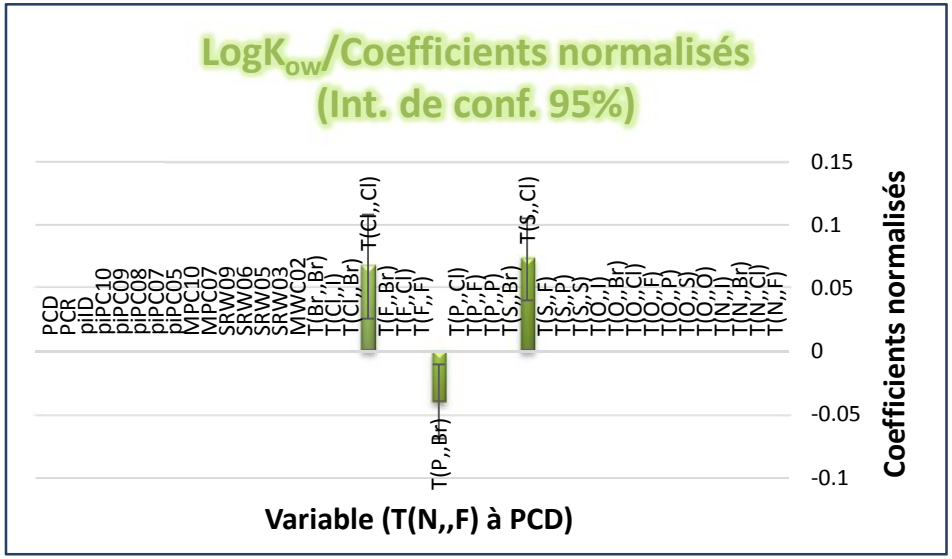


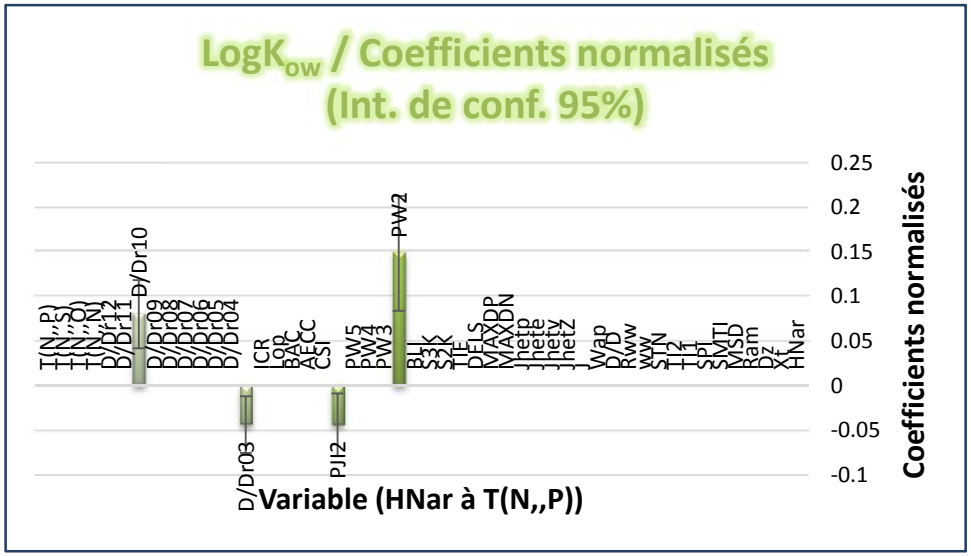


تابع الشكل (IV-1-3): بيان يوضح أسماء الواصفات الجزيئية الخاصة بنموذج جميع الكتل **معا**.

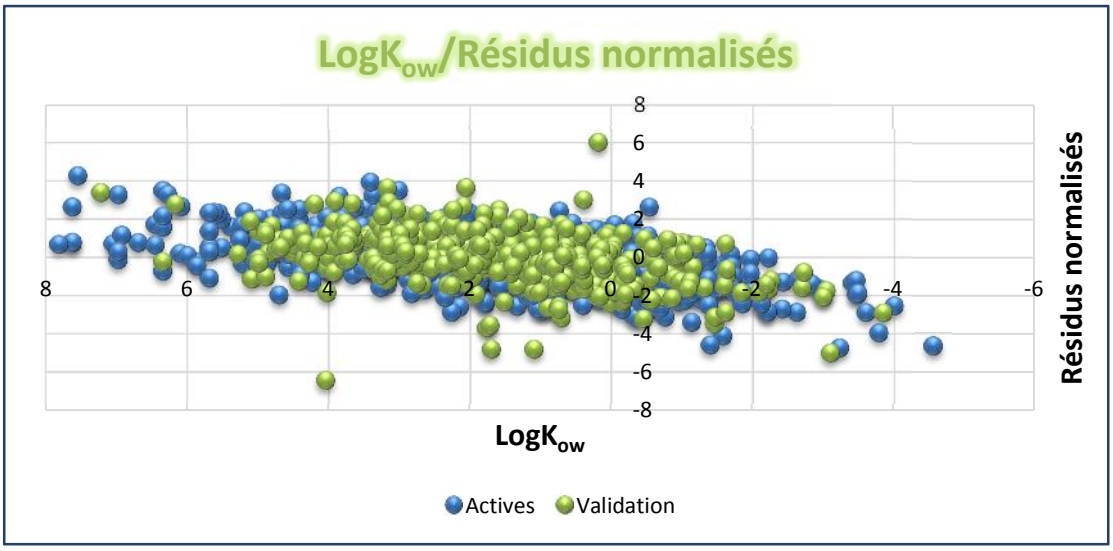


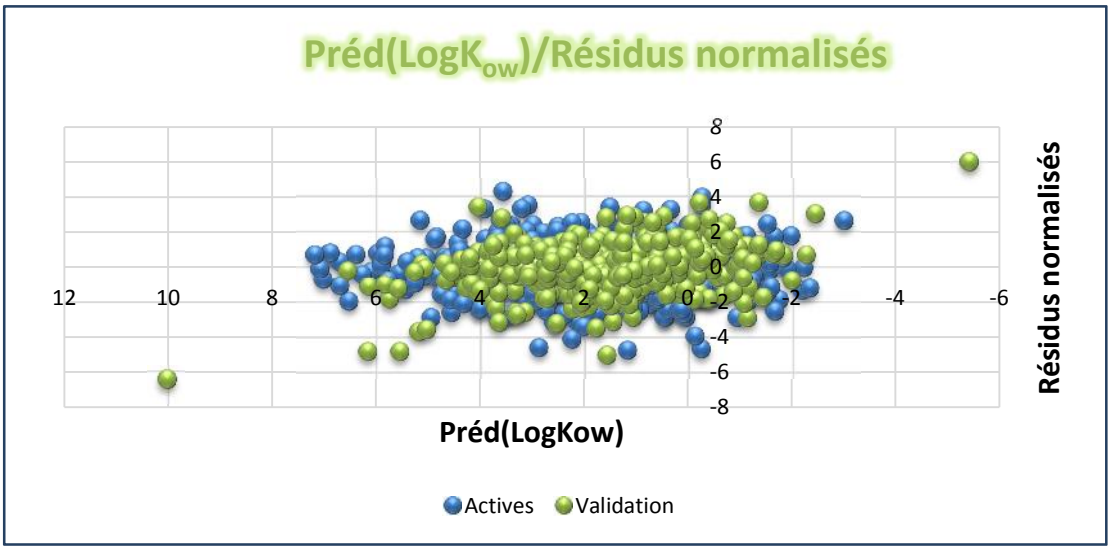


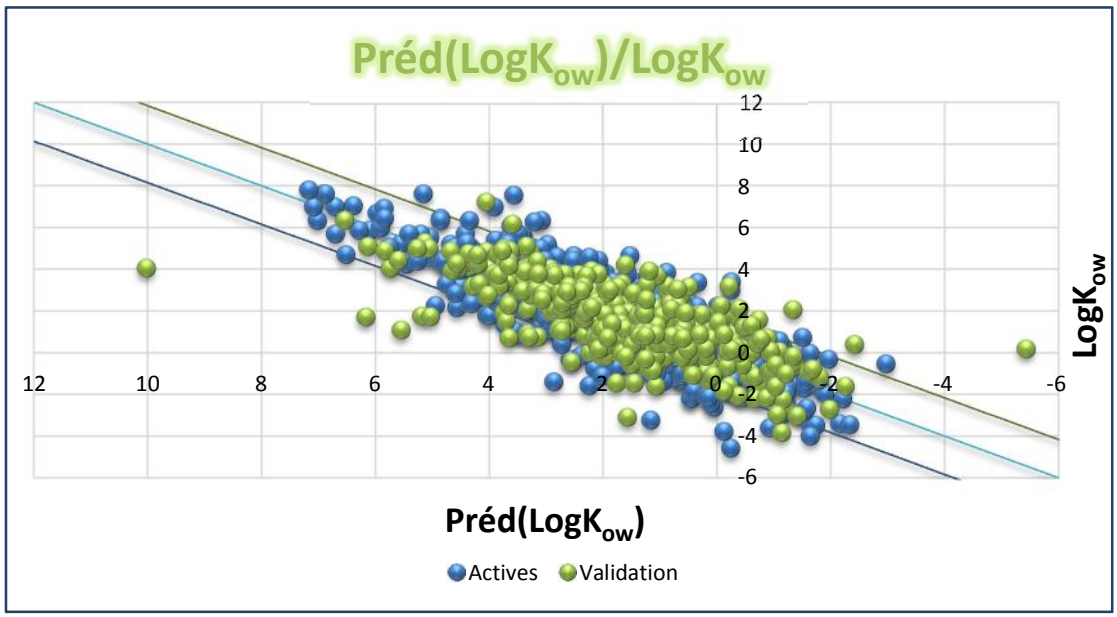




تابع الشكل (IV-3-2): بيان يوضح أسماء الواصفات الجزيئية الخاصة بنموذج جميع الكتل **معا**.







الشكل (4-IV): بيان يوضح تغيرات LogKow **بدلالة** Préd(LogKow) **الخاصة** بجميع الكتل معا.

1.2.IV. تحليل نتائج المرحلة الثانية:

معامل التحديد **(R^2)** يعطينا فكرة عن النسبة المئوية لتغير المتغير المنمذج موضح بالمتغيرات التفسيرية .حيث كلما اقترب معامل التحديد إلى 1، كلما كان النموذج أفضل.

في هذه الحالة لدينا **R^2 = 0.688780987926** و **R^2 ajusté = 0.67464398211092** **وبالتالي فإن** النسبة المئوية لتغير المتغير المنمذج **= 68.87%**.

2.2.IV. تحليل البقايا (Residues) نتائج المرحلة الثانية:

Residues = 1- R^2 = 1 - 0.6887 = 0.3113 = 100 × 0.3113 = 31.13 %

هذا يعني أن **95%** من الباقي توجد في المجال (+2، -2) كل قيمة خارج هذا المجال غير موثوق فيها. المتغيرات التي تسمح بتفسير **68.87%** من تغير $LogK_{OW}$ تفسر النسبة المئوية للبقايا وهي **13.31% من** عدة عوامل تفسيرية أخرى.

3.2.IV. مناقشة نتائج المرحلة الثانية:

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

- الواصفات الجزيئية الأقل ارتباطا:

(BEHm5, VED2) = - 0.93

(ATS1m, VED2) = - 0.72

(GATS5e, MATS6m) = - 0.69

الجدول(IV-11): الواصفة ورمزها للواصفات الجزيئية الأقل **ارتباطا** (جميع الكتل معا).

رمز الواصفة	الواصفة
BEHm5	highest eigenvalue n. 5 of Burden matrix / weighted by atomic masses
VED2	average eigenvector coefficient sum from distance matrix
ATS1m	Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic masses
GATS5e	Geary autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities
MATS6m	Moran autocorrelation - lag 6 / weighted by atomic masses

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

- الواصفات الجزيئية الأكثر ارتباطا:

(ESpm02d, ESpm06d) = 1.00

(EEig06x, EPS0) = 0.99

(ATS2m, ATS1m) = 0.95

الجدول (IV-12): الواصفة ورمزها للواصفات الجزيئية الأكثر ارتباطا (جميع الكتل معا).

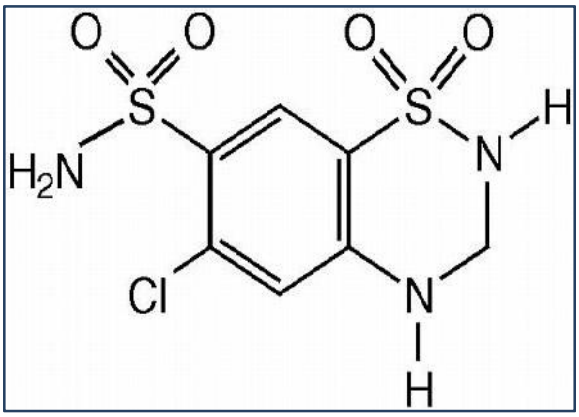
رمز الواصفة	الواصفة
ATS1m	Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic masses
ATS2m	Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic masses
EPS0	edge connectivity index of order 0
EEig06x	Eigenvalue 06from edge adj. matrix weighted by edge degrees
ESpm02d	Spectral moment 02 from edge adj. matrix weighted by dipole moments
ESpm06d	Spectral moment 06 from edge adj. matrix weighted by dipole moments

الرسم البياني ($logk_{ow}$/(pérd ($logk_{ow}$) يبين لنا أن هناك 95% من جزيئات قاعدة البيانات داخل مجال الثقة. ووفقا للتحليل الإحصائي للقيم والمنحى البياني ($logk_{ow}$, $logk_{ow\ pérd}$) ونتائج المقارنة بين الكتل يمكن ملاحظة أنه: يوجد قيم خارج مجال الثقة (-2 ، +2) هذه القيم تمثل النقاط الشاذة أي (جزيئات غير مرتبطة).

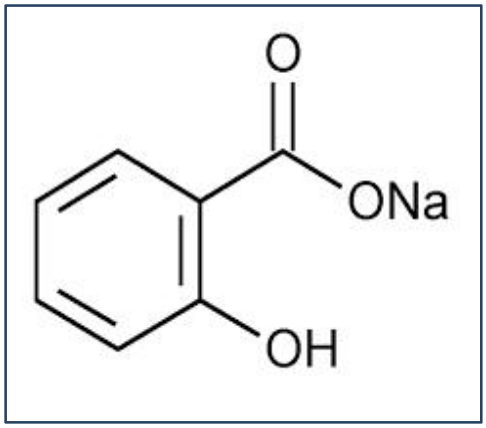
بالمقابل وجدنا 42 جزيء غير مرتبط أي خارج مجال الثقة لان لديها هياكل كيميائية غير منتظمة. أما سبب عدم الانتظام يرجع لوجود الهالوجينات مثل (N, Cl, S) والحلقات البنزينية والروابط الثلاثية أوالمجموعات القادرة على إعطاء زوج الإلكتروني مثل (NH_2, Na).

أما جودة النموذج المتحصل عليه فتعتمد على عدد المتغيرات (الواصفات الجزيئية) وقيمة معامل التحديد R^2. تتغير درجة تقدير (قيمة R^2) لـ $logk_{ow}$ وفقا لنوع الواصفات لكل كتلة.

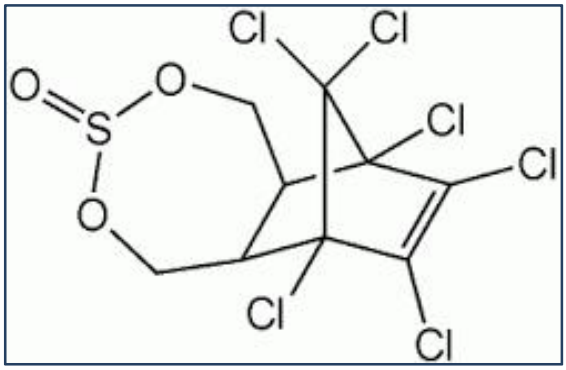
من أجل تحسين النتائج يجب: استبعاد الجزيئات غير المرتبطة مثلا المركبات التالية:



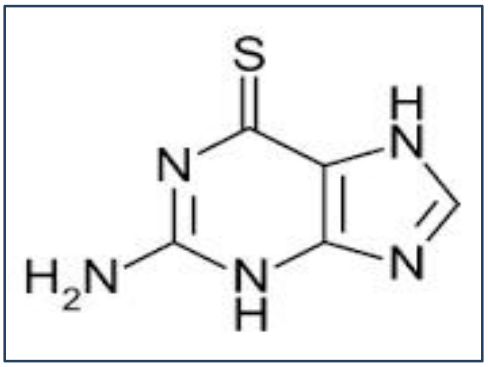
Hydrochlorothiazide



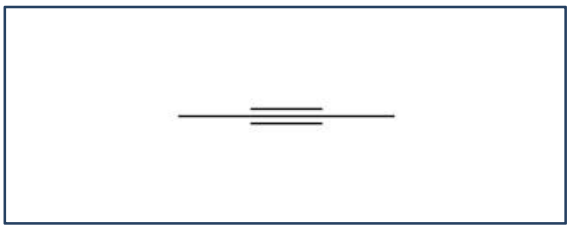
Sodium salicylate



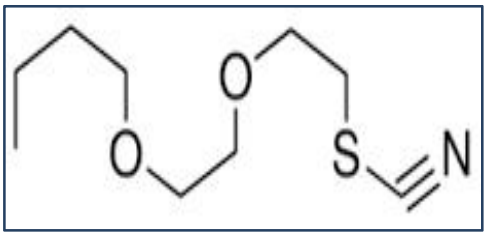
a-Endosulfan



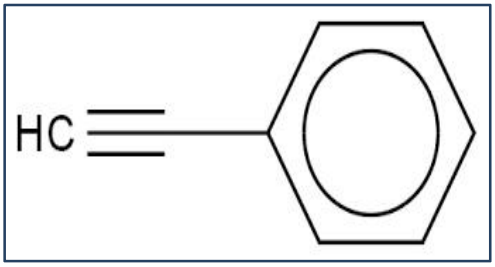
Thioguanine



2-Butyne



2(2-Butoxyethoxy)ethyl Thiocyanate



Benzene, ethynyl

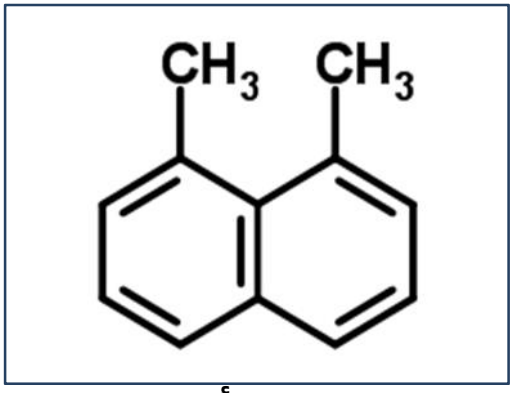
الجدول (IV-13): أمثلة للمركبات التي وجدت خارج مجال الثقة.

رقم	$LogK_{ow}$	الإسم	SMILES
82	4.05	Bis(tribuyl tin) ether	CCCC[Sn](CCCC)(CCCC)O[Sn](CCCC)(CCCC)CCCC
263	1.7	Cyclopropane	C(C1)C1
249	1.09	Methane, chloro	ClC
141	-3.83	L-Glutamic acid	O=C(O)C(N)CCC(=O)O
214	1.52	Methane, trichloro	C(Cl)(Cl)Cl
717	0.39	Methylmercuric chloride	Cl[Hg]C
1188	-3.09	Butylpyridinium bromide	c1ccccn1(CCCC)Br
1151	-1.46	Benzene, iodyl	O=I(=O)c1ccccc1
1128	7.22	Tetradecane	C(CCCCCCCCCCCC)C
1643	1.72	3-Isopropyl-4-hydroxyacetanilide	CC(=O)Nc1cc(C(C)C)c(O)cc1
1644	2.06	Enflurane	FC(F)OC(F)(F)C(F)Cl
1442	0.75	Sulfamethomidine	c1cc(N)ccc1S(=O)(=O)Nc2nc(C)nc(OC)c2
1750	-1.61	Streptozocin	CN(N=O)C(=O)NC1C(O)OC(CO)C(O)C1O
1827	0.18	6-Nitrohydrothiazide	NS(=O)(=O)c1c(N(=O)(=O))cc2NCNS(=O)(=O)c2c1
1984	4.2	Penbutolol	C1CCCC1c2ccccc2OCC(O)CNC(C)(C)C

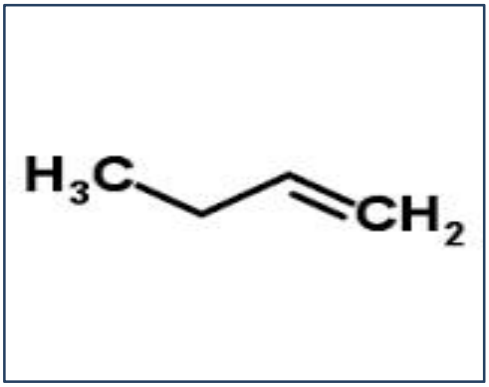
أما بالنسبة للتحقق الخارجي لـ (test):

أخذنا عشوائيا 20 مركبا **(الجدول 15)** من قاعدة البيانات وقمنا باستعمالها للتحقق الخارجي لـ (test) وذلك بتعويض قيم ومعاملات الواصفات الجزيئية في معادلة النموذج جميع الكتل معا (مجمعة). وطبقنا هذا العمل على مركبين فقط من الـ 20 مركب وهما:

- ✓ **المركب الأول:** 1,8-Diméthyl naphtalène.
- ✓ **المركب الثاني:** 1-Butène.



المركب الأول



المركب الثاني

بعد التعويض والحسابات كانت النتائج كمايلي:

الجدول(IV-14): نتائج عملية التحقق الخارجي (test).

	قيمة LogKow بعد الحساب	قيمة LogKow التجريبية
المركب الأول	4.18	4.26
المركب الثاني	1.93	2.17

هذه النتائج خاضعة لقوة النموذج المحسوب وهي: **68.87%**.

الجدول (IV-15): أسماء الـ 20 مركبا المحددة للعملية للتحقق الخارجي لـ (test).

الرقم	LogKow	الإسم	SMILES
4	4.23	Indomethacin	COc1ccc2n(C(=O)c3ccc(Cl)cc3)c(C)c(CC(=O)O)c2c1
88	0.97	Cyclophosphamide	C1CCNP(=O)(N(CCCl)CCCl)O1
151	2.42	Librium	CNC2=Nc1ccc(Cl)cc1C(=N(=O)C2)c3ccccc3
203	-0.32	Thymine	C1=C(C)C(=O)NC(=O)N1
323	0.58	Propanoic acid	O=C(O)CC
440	2.98	Benzoic acid, 4-hydroxy-, propyl ester	O=C(OCCC)c(ccc(O)c1)c1
514	2.14	Benzoic acid, 4-nitro-, ethyl ester	O=C(OCC)c(ccc(N(=O)=O)c1)c1
630	2.09	2-Butene,	C(=CC)C
914	4.73	111F3-22-di(p-MeOphenyl)ethane,	C(F)(F)(F)C(c1ccc(OC)cc1)c2ccc(OC)cc2
971	2.26	Benzene, ethynyl	C(c(cccc1)c1)≡C
1000	1.73	2,6-Dinitrophenol	Oc1c(cccc1N(=O)=O)N(=O)=O
1080	1.43	2-Bromoacetanilide	CC(=O)Nc1ccccc1Br
1143	0.02	1-(3-Carboxypropyl)benzotriazole	c1ccc2nnn(CCCC(=O)O)c2c1
1313	1.79	9-Pentyladenine	n1cnc2n(CCCCC)cnc2c1N
1516	1.7	1,4-Benzodiazp2-on,5(2NO2Ph)7NO2	c1c(N(=O)(=O))cc2C(c3ccccc3N(=O)(=O))=NCC(=O)Nc2c1
1660	-1.56	2-Methylpyridinium Iodide,n-hexyl	Cc1ccccn1(CCCCCC)I
1703	1.53	p-Nitroformanilide	c1cc(N(=O)(=O))ccc1NC=O
1913	0.35	2'-Nitrobenzoguanamine	n1c(N)nc(N)nc1c2ccccc2N(=O)(=O)
2000	-0.16	p-Amido-amphenicol	c1cc(C(=O)N)ccc1C(O)C(CO)NC(=O)C(Cl)Cl
2270	2.86	N(N""-Me-(4Me-Ph)Carbamylthio)methomyl	CC(SC)=NOC(=O)N(C)SN(C)C(=O)c1ccc(C)cc1

خاتمة عامة

خاتمة عامة

بعد الدراسة وإجراء الحسابات للمرحلتين (1) و(2) توصلنا إلى تحقيق العلاقة التالية:

(الواصفات الجزئية) f = الخاصية بالشكل التالي: $Logk_{ow} = a + \Sigma (b_j \times x_j)$ حيث كان أفضل نموذج للمرحلة (1) هو الخاص بالكتلة أجزاء مراكز الذرة **(Atom-centered fragments)** والذي تميز بـ **(R^2= 0.8558, F = 148.328, S^2 = 0.6286).**

أما بالنسبة لنتائج المرحلة (2) فتميزت بـ (F= 48.72, R^2= 0.6887, S^2 = 0.9678).

وبناء عن هذه النتائج يمكن **استنتاج ما** يلي:

- ✓ تعتبر الدراسة بتقنية الـ QSAR/QSPR باستعمال طريقة (MLR) حلا موثوقا لأنها تسمح للكيميائيين بإيجاد قيم لـ LogKow لمركبات مجهولة خاصة المركبات الكيميائية الخطيرة ذات السمية العالية التي يصعب التعامل معها مخبريا.
- ✓ أنه يمكن دراسة العلاقة بنية الكيميائية ـ معامل التوزيع (أوكتانول - ماء) LogKow **وذلك** باستخدام طريقة الانحدار الخطي المتعدد (MLR).
- ✓ أن القدرة التنبؤية تعتمد على نوع وعدد المركبات الكيميائية لمكونة لقاعدة البيانات.

ولتحسين النتائج يمكن **اقتراح ما** يلي:

- ✓ **إلغاء أو حذف** المركبات غير مرتبطة.
- ✓ جمع أكثر عدد من الجزيئات لتوسيع قاعدة البيانات لإعطاء نتائج أفضل.
- ✓ يجب إجراء تحسين (Optimisation) **كامل للمركبات للوصول إلى أقصى درجة لاستقرار.**

المراجع

المراجع

المراجع باللغات الأجنبية:

[1] Andrea Mauri*, V. Consonni, M. Pavan, and R. Todeschini, 2006, DRAGON SOFTWARE: AN EASY APPROACH TO MOLECULAR DESCRIPTOR CALCULATIONS, Milano Chemometrics and QSAR Research Group, Dept. of Environmental Sciences, University of Milano-Bicocca, P.za della Scienza 1 – 20126 Milano (Italy).

[2] CambridgeSoft Corporation 100 CambridgePark Drive Cambridge, MA 02140 USA, 2002, ChemOffice, ChemDraw, Chem3D, ChemFinder et ChemInfo sont des marques de commerce de CambridgeSoft ©.28p.

[3] DRAGON, for Windows and Linux 2006, Copyright © 2003-2006 by Talete srl, Milano, Italy. Help

[4] 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.

[5] Jmasi.com/ehsa/correlation/int.htm.

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

[7] Mr. Lanez Touhami, Cours de technique de purification en synthèse organique, Université de Kasdi Merbah Ouargla.

[8] Prof. Dr. Stefan Bienz, Unknown 2015, Short Manual to the Chemical Drawing Program ChemDraw®, university of zurich IfC Department of Chemistry,22p.

[9] REGADI Dahmane, MASTER, UNIVERSITE KASDI MARBAH OUARGLA /2013.

[10] Samar Bakoben & Ahlam Houda, 16/07/2011, MEGBI Training Course Molecular Modelling, Institute for Genetic Engineering, Ecology and Health (IGEEH) Karlsruhe, Germany.120p.

[11] Samehar.files.wordpress.com/2009/10/residuals.

[12] Statistics for Managers, Levine et al. Prentice Hall, 1999,

[13] United Nations Publications, 2010 - Globally Harmonized System of Classification and Labelling of Chemicals (Ghs), 566p.

المراجع باللغة العربية:

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

[15] **أسس** الكيمياء العضوية، أ.أحمد الشريف على الدرسي، 2008/دار الكتب الوطنية بنغازي-ليبيا.

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

[17] الانحدار والثبات، د. حنان ربيع.

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

[19] **تخة مبركة**، مساهمة في دراسة العلاقة (بنية كيميائية – رائحة) باستعمال طريقة آلات الشعاع **الساند** (SVM) (تطبيق على عائلة البيرازين)، **مذكرة** ماجستير، **جامعة ورقلة**/ 2008. **ص** 28-32.

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

[21] حمادي رضا، مذكرة ماجستر، حساب LogP **انطلاقا من مجموعة من الوسائط الخاصة** بالمذيب والمذاب، **جامعة ورقلة/** 2008. **ص** 1.

[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.

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

[24] الكيمياء الحديثة، د. هاني كريدين، 1994. **ص** 391.

[25] الكيمياء العضوية، **د. منال فوزي أبوطالب.**

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

الملاحق

الملاحق

مفهوم بعض الواصفات الجزيئية

1. الواصفات الهندسية:

تعرف البنية ثلاثية الأبعاد للجزيئة على وجه العموم، ترتكز 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.

ASP: L'asphericity ثابت يصف كروية الجزئية 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.

2. مؤشرات الشحن الطوبولوجية:

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

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

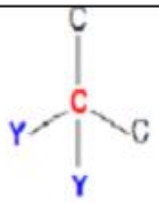
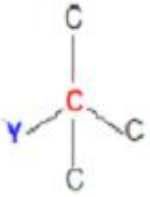
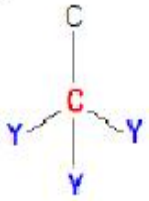
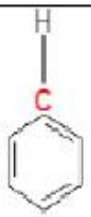
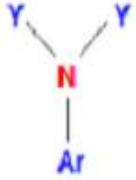
δ_i: هي درجة القمة لكل ذرة، وبمعنى آخر عدد الذرات المرتبطة بالهيدروجين المتفرع.

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.

3. مجاميع المجموعات الوظيفية:

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

nC_s		Y = (H, X, OH, NH2, ...) أو ذرة متغايرة .و$C(sp^3)$
nC_t		Y = (H, X, OH, NH2, ...) $C(sp^3)$ أو ذرة متغايرة .و
nC_p		Y = (H, X, OH, NH2, ...) $C(sp^3)$ أو ذرة متغايرة .و
NCbH		عدد حلقات البترين غير المستبدل $C(sp^2)$
nRSR		عدد الكبريتيدات
nArOR		عدد الاثيرات العطرية Y= Ar ≠ C=O,C #N
nAr NR2		عدد الأمينات الثلاثية العطرية Y = Ar ,Al ≠ C=O

4. مؤشرات المعلومة:

هي واصفات تبرز محتوى المعلومات حول الجزئية، لها عدة معايير مختلفة تستعمل لتعريف الفروع المتكافئة في الجزيئة 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 gTh.

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

ISIZ = nAT. $\log_2$ nAT

CICK = $\log_2$ nAT - ICK

IVDE محتوى المعلومات المتوسط على مساواة درجة القمة.

Xindex الثابت x لـ 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:

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

f: معامل القياس.

W: الخصائص المميزة للذرات i و j.

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

nAT: عدد الذرات في الجزئية.

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.

الملخص:

في هذا العمل تحصلنا على علاقة رياضية تربط بين البنية الكيميائية المعبر عنها بالواصفات الجزيئية (2D)، والخاصية الفيزيوكيميائية المتمثلة في معامل التوزيع **$LogK_{ow}$**، وذلك للتنبؤ بقيم **$LogK_{ow}$ لمركبات أخرى**. شملت هذه **الدراسة** 13 كتلة من الواصفات الجزيئية (2D) وهذا وفق مرحلتين أولاهما هي حساب الكتل كل على حد، أما الثانية فهي **حساب كل الكتل معا** (مجمعة). وتمت هذه الدراسة على 2274 مركب كيميائي بتطبيق تقنية QSPR وبطريقة الانحدار **الخطي المتعدد** (MLR) وهذا باستخدام البرمجية الإحصائية Xlstat. **وكانت أحسن نتائج المرحلة الاولي هي (R^2 = 0,8556), (S^2 = 0.6286) و(F = 148.328) بالنسبة إلى كتلة أجزاء مراكز الذرة. أما بالنسبة** للمرحلة الثانية **(كل الكتل معا (النموذج العلم)) هي (R^2 = 0.6887), (S^2 = 0.9678) و (F = 48.722).**

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

Résumé:

Dans ce travail, nous avons acquis un modèle mathématique reliant la structure chimique qui est représenté par descripteurs moléculaires (2D), et de la propriété physico-chimique du coefficient de partage **$LogK_{ow}$**, ce modèle nous permet de prédire les valeurs **$LogK_{ow}$** à d'autres produits chimiques. Cette étude, qui contient deux étapes a été appliqué sur 13 blocs de descripteurs moléculaires (2D). La première étape est le calcul de chaque bloc individuel, la seconde est la charge de tous les blocs entre eux (regroupées). Cette étude a été appliquée à 2274 composés chimiques en utilisant la technique QSPR avec régression linéaire multiple (MLR) par le logiciel XLSTAT. Le meilleur résultat de la première étape est **(R^2 = 0,8556), (S^2 = 0,6286) et (F = 148,328)** pour (fragments d'atomes centré). Pour la deuxième étape (tous les blocs ensemble (modèle général) est **(R^2 = 0,6887), (S^2 = 0,9678) et (F = 48,722)**.

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 (2D), and the physicochemical property the partition coefficient **$LogK_{ow}$**, this model allows us to predict the **$LogK_{ow}$** values for other chemicals. This study, which contains two stages, was applied to 13 molecular descriptors (2D) blocks. The first stage is the calculation of each individual block, second is the expense of all the blocks together (grouped). This study was applied to 2274 chemical compounds using QSPR technique with multiple linear regression (MLR) by xlstat software. The best result of the first stage is **(R^2= 0.8556), (S^2 = 0.6286)** and **(F = 148.328)** for (atom-centred fragments). For the second stage (all the blocks together (general model) is **(R^2=0.6887), (S^2 = 0.9678)** and **(F = 48.722)**.

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