**الجمهورية الجزائرية الديمقراطية الشعبية**

**وزارة التعليم العالي والبحث العالي**

**جامعة الشهيد حمه لخضر – الوادي**

**كلية التكنولوجيا**

**مذكرة نهاية الدراسة لنيل شهادة**

# ماستر أكاديمي

**ميدان :** التكنولوجيا

**شعبة:** هندسة الطرائق

**تخصص:** هندسة كيميائية

## إعداد الطلبة:

عبد الجواد الشيماء

عدايكة عفاف

مياطة صفاء

## الموضوع

# الدراسة النظرية لتاريخ تطور نماذج العلاقة الكمية بين البنية والنشاط من (QSAR) وصولا إلى (QSARINS).

**نوقشت يوم:2020/06/17 أمام اللجنة المكونة من:**

| | | | |
|---|---|---|---|
| بن ميه عمار | أستاذ(ة) محاضر(ة) | جامعة الشهيد حمه لخضر الوادي | رئيسا. |
| سروطي عبد الغني | أستاذ(ة) محاضر(ة) | جامعة الشهيد حمه لخضر الوادي | مناقشا. |
| منصر سهيلة | أستاذ(ة) محاضر(ة) | جامعة الشهيد حمه لخضر الوادي | مؤطرا. |

السنة الجامعية:2021/2020

# الإهداء

اهدي ثمرة جهدي هذه.....

إلي ملاكي في الحياة ..إلي معنى الحب إلى معنى الحنان والتفاني ..إلى بسمة الحياة وسر الوجود إلى من كان دعاؤها سر نجاحي وحنانها باسم جراحي

إلى اغلي الحبايب ....أمي الغالية

إلى من كلله الله بالهيبة والوقار ..إلى من احمل اسمه بافتخار ..أرجو من الله أن يمد في عمره لترى ثمارا قد حان قطافها بعد طول انتظار وستبقى كلماتك نجوم اهتدي بها اليوم وفي الغد والى الأبد ..أبي الغالي

إلى من شاركوني الدمعة والابتسامة ..إلى من كانوا متكئا استند عليه وخير معين الجأ إليه ..إلى من لم يبخلوا علي بالنصح والإرشاد .....أخواتي وأخوتي

عفاف وجهاد وئام وإيناس مريم وشيماء ريان مروة ويحي لؤي عبد الحق احمد....

إلى من تذوقت معهم أجمل اللحظات..إلى من جعلهم الله إخوتي بالله ..إلى من أحببتهم بالله .....طلاب شعبة الهندسة الكيميائية

إلى من لم اعرفهم ...ولن يعرفوني

إلى من أتمنى أن تبقى صورتهم ...في عيوني

شيماء & عفا ف & صفاء

# شكر وعرفان

الحمد والشكر لله الذي لا معبود سواه . الذي رزقنا العقل وحسن التوكل عليه سبحانه وتعالي وألهمنا الصبر ووفقنا لإتمام هذا العمل المتواضع .والصلاة على نبيه محمد واله وصحبه .....

وقبل أن نمضي نقدم اسمي آيات الشكر والامتنان والتقدير والمحبة إلى الذين مهدو لنا طريق العلم والمعرفة وسهروا من اجل ارتقائنا وفي مقدمتهم عميد كلية العلوم والتكنولوجيا ورئيس قسم هندسة طرائق وال بترو كيمياء ومسؤول شعبة هندسة الكيميائية وكل الأساتذة كلية العلوم والتكنولوجيا...

يسرنا أن نقدم بالشكر إلى السادة والمؤطرين والمشرفين على دفعة الهندسة الكيميائية الذين ساهموا في تكويننا ولم يبخلوا علينا بتوجيهاتهم.....

وعلى رأسهم الأستاذة الفضيلة "منصر سهيلة" التي أشرفت على انجاز هذا العمل وكانت نورا يضيء الظلمة التي كانت أحيانا في طريقنا فجزاها الله على الخير كله فلها منا الشكر والتقدير ...

كذلك اشكر كل من ساعدنا وقدم لنا العون وزرع التفاؤل في دربنا ومد لنا يد العون

شيماء

## الملخص

في هذا العمل قمنا بدراسة مراحل تطور **الطريقة الحسابية QSAR** منذ نشأتها إلى يومنا **هذا**,حيث شهدت توسعا كبيرا شمل العديد من المجالات منها الصيدلة الكيميائية,الكيمياء الفلاحية و علم الاحياء ...الخ حيث عرفت تطورا ملحوظا في المنهجيات والتقنيات الحسابية التي سمحت بتحديد العديد من المتغيرات و**الأساليب المستخدمة** و**البيانات** المتعلقة بطريقة النمذجة الجزيئية.في وقتنا الحالي تم الوصول لبرنامج **QSARINS** المتمثل في تطوير طرق نماذج **QSAR** **والتحقق من صحتها** بتحديد بياناتها الحديثة الآتية $Q^{2}_{F1}$ , $Q^{2}_{F2}$, $Q^{2}_{F3}$ وCCC و VIF التي تساهم في تحسين وتحقيق قوة التنبؤ لهذه النماذج و التأكد من صحتها ونجاعتها.

الكلمات المفتاحية : QSARINS, QSAR المركبات الكيمائية ، الواصفات الجزيئية ، النشاط البيولوجي

## Résumé :

Dans ce travail, nous nous sommes intéressés par l'étudedes différentes étapes de développement qu'a connu la méthode Quantitative Structure Activity Relationship **QSAR**, depuis sa création jusqu'à nos jours. Cette étude a fait révéler que cette méthode QSAR a connu une étendue significative dans certains domaines tels que la chimie pharmaceutique, chimie d'agriculture, biologie... Etc. Ce développement, en particulier ,dans les méthodologies et techniques de calcul a permis l'identification et la détermination de nombreuses variables, et données de la modélisation. Depuis plusieurs années, la méthodologie QSAR ne cesse de se développer pour qu'elle devienne ces jours-ci **QSARINS**. Cette dernièrea fait apparaitre des nouvelles données à savoir : $\boldsymbol{Q^2_{F3}}$; $\boldsymbol{Q^2_{F2}}$ ; $\boldsymbol{Q^2_{F1}}$**le CCC** et le **VIF** qui contribuent à l'amélioration du pouvoir de prédiction des modèles ainsi la confirmation de leur validité.


**Mot clé :** QSARINS, QSAR, composés chimiques, descripteur moléculaire, activité biologique.

**فهرس المحتويات**


| | |
|---|---|
| فهرس الاختصارات | |
| العنوان | الصفحة |
| مقدمة عامة | 4-1 |
| **الجزء النظري** | |
| **الفصل النظري: مراحل تطور QSAR** | |
| **I**-مراحل تطور QSAR | 6 |
| I-1 النمذجة الجزيئية | 6 |
| I-2 النموذج الرياضي | 7 |
| I-3 مزايا النموذج الرياضي. | 8 |
| I-4 تاريخ تطور QSAR. | 8 |
| I-4-1 نشأةQSAR. | 9 |
| I-4-1-1 مرحلة الأولى. | 9 |
| I-4-1-2 مرحلة الثانية. | 9 |
| I-4-1-3 مرحلة الثالثة. | 10 |
| I-4-1-4 مرحلة الرابعة. | 10 |
| I-4-1-5 تعريف QSAR. | 11 |
| I-4-1-6 مرحلة الخامسة. | 11 |
| I-4-1-7 مرحلة السادسة. | 12 |

| | |
|---|---|
| I-4-1-8 مرحلة السابعة. | 13 |
| I-4-1-9 مرحلة الثامنة . | 15 |
| I-4-1-10 مرحلة التاسعة. | 15 |
| I-5 أبعاد QSAR. | 17 |
| I-6 طرق QSAR الحديثة. | 19 |
| I-7 الإحصاء. | 21 |
| I-7-1 تعريف الإحصاء. | 21 |
| I-7-2 تفسير الواصفات. | 22 |
| I-8 البيانات الاحصائية لبداية بناء النموذج. | 24 |
| I-9 طرق نمذجة QSPR/QSAR. | 25 |
| I-10 طرق نمذجة QSARالأخرى. | 26 |
| I-11 برامج QSAR. | 29 |
| I-12 تطبيقات QSAR. | 30 |
| I-13 المبادئ الأساسية QSAR | 30 |
| I-14 الوصول إلىQSAR الحديثة (QSARINS) | 31 |
| I-14-1(QSAR-INSUBRIA) QSARINS | 33 |
| I-14-2 QSARINS | 33 |
| المراجع | 39-34 |
| **الجانب العملي** | |

| الفصل الثاني: طريقة حساب نماذج QSAR | |
|---|---|
| II- طريقة الحساب نماذج QSAR. | 41 |
| II-1 الهيكلي للمركبات وحساب الطاقة الفراغية Es. | 42 |
| II-1-1 الأساليب التجريبية. | 42 |
| II-1-2 طرق الكم. | 43 |
| II-2 توليد الواصفات. | 43 |
| II-3 حساب الواصفات الجزيئية. | 44 |
| II-4 تطوير نموذج : | 45 |
| II-4-1 اختيار الواصفات: | 45 |
| II-4-2 طرق التصفية | 46 |
| II-5 بناء نموذج QSAR: | 47 |
| II-5-1 مجموعات بيانات التدريب والاختبار | 47 |
| II-6 الأساليب المستخدمة في QSARالحديثة (QSARINS). | 48 |
| II-6-1 النماذج الإحصائية. | 48 |
| II-7 طرق تحليل البيانات | 49 |
| II-7-1الانحدار الخطي المتعدد | 49 |
| II-7-1-1 نموذج الانحدار الخطي. | 50 |
| II-8 المربعات الصغرى الجزئية (PLS) | 53 |
| II-8-1 المربعات الصغرى الجزئية الوراثية (G / PLS) | 54 |

| | |
|---|---|
| II-8-2 المربعات الصغرى الجزئية لتحليل العامل (FA-PLS) | 54 |
| II-8-3 المربعات الصغرى الجزئية لتصحيح الإشارة المتعامدة (OSC-PLS) | 54 |
| II-9 الشبكة العصبية الاصطناعيةANN | 55 |
| II-10 التحقق من صحة النموذج | 56 |
| II-10-1 البيانات والاختبارات الإحصائية القياسية. | 57 |
| II-10-1-1 معامل الارتباط r. | 57 |
| II-10-1-2 معامل التحديد $R^2$. | 57 |
| II-10-1-3 لمتوسط الخطأ المطلقMAE. | 58 |
| II-10-1-4 SD الانحراف المعياري الخطأ القياسي في التنبؤS. | 59 |
| II-10-1-5 عامل التضخم الخاص بـالتباينVIF | 59 |
| II-10-1-6 الاختبار الاحصائية F | 59 |
| II-10-1-7 le test-t de Studentti. | 60 |
| II-11 مقاييس التحقق من الصحة من نوعين بناءً على نوع نموذجQSAR. | 61 |
| II-11-1 مقاييس التحقق من الصحة للطرق القائمة على الانحدار. | 61 |
| II-11-1-1 مقاييس التحقق من الصحة للتحقق الداخلي. | 61 |
| II-11-1-1-1 Least Square Fitting | 61 |
| II-11-1-1-2 الخطأ التربيعي لمتوسط الجذر (RMSE) | 62 |
| II-13-1-1-3 عبر المصادقة(LGO) | 62 |
| II-13-1-1-3-1التحقق من صحة الإجازة مرة واحدة (LOO) والإجازة عن بعض (LSO) | 62 |

| | |
|---|---|
| II-13-1-1-3-2 التحقق من صحة الإجازة (LSO) | 63 |
| II-11-1-1-4مقاييس $Q^2$ و $r^2m$ | 64 |
| II-11-1-2مقاييس التحقق من الصحة الخارجية. | 64 |
| II-11-1-2-1معامل الارتباط التربيعي في مجموعة التدريب: $Q^2_{F1}$ | 64 |
| II-11-1-2-2 معامل الارتباط التربيعي التنبئي في مجموعة الاختبار $Q^2_{F2}$ | 65 |
| II-11-1-2-3 القدرة التنبؤية الخارجية: $Q^2_{F3}$ | 65 |
| II-11-1-2-4 معايير Golbraikh و Tropsha . | 65 |
| II-11-1-2-5 معامل الارتباط التوافق. | 66 |
| II-11-2 مقاييس التحقق من الصحة للطرق القائمة على التصنيف. | 67 |
| II-11-2-1 Wilks lambda (λ)الإحصاء. | 67 |
| II-11-2-2 مؤشر canonical index (Rc). | 67 |
| II-11-2-3 Chi-square(χ2) للتحقق من جودة النموذج القائم على التصنيف | 68 |
| II-12 تفسير وتحليل المجال التطبيق | 68 |
| المراجع | 69-74 |
| خاتمة عامة | 76 |

## فهرس الأشكال والصور


| | | الصفحة |
|---|---|---|
| **الشكل I -1** | خطوات النمذجة الجزيئية | 8 |
| **الشكلII -1** | المراحل الرئيسية في بناء نموذج QSAR | 42 |
| **الصورII- 2** | تبعثر للنشاط المرصود والمحسوب لنموذج MLR | 50 |
| **الصور II-3** | تمثيل تخطيطي لـ ANN | 56 |

## فهرس الرموز

| الرمز | المعنى |
|---|---|
| ia | هي مساهمة كل ميزة هيكلية ؛ |
| AB | هي النشاط البيولوجي ، |
| Φ | نظام بيولوجي |
| Π | الكراهية النسبية للماء |
| 50C | تركيز المادة السامة بنسبة 50% |
| CCC | معامل ارتباط التوافق |
| 1D1 | احادي البعد |
| D2 | ثنائي البعد |
| D3 | ثلاثي البعد |
| D4 | رباعي البعد |
| D5 | خماسي البعد |
| D6 | سداسي البعد |
| Es | الطاقة الفراغية |
| I | مصطلح إلكتروني |
| I2 | I2 = متغير المؤشر لعدم التشبع |
| F | الاختبار المعياري لحساب معنوية النموذجFisher parameter |
| | |

| | |
|---|---|
| PX | معامل التقسيم للمشتق |
| PH | معامل التقسيم للجزيء الأصلي |
| $Q^{2}_{F1}$ | معامل الارتباط التربيعي في مجموعة التدريب |
| $Q^{2}_{F2}$ | معامل الارتباط التربيعي في مجموعة الاختيار |
| $Q^{2}_{F3}$ | القدرة التنبؤية الخارجية |
| LMO-$Q^2$ | التحقق من خلال ترك الكثير للخارج تحقق من صحتها بحذف مجموعة من الملاحظات |
| $Q^2_{CV-LOO}$ | معامل الارتباط الناتج للتحقق المتقاطع. |
| $Q^2$ LOO | للتحقق المتقاطع من خلال عدم الخروج للتحقق المتقاطع بحذف ملاحظه من ملاحظه. معامل التنبؤ |
| $r^2m$ | بناءً على القيم المقاسة للنشاط المرصود والمتوقع |
| r | معامل الارتباط الخطي |
| $R^2$ | معامل التحديد مربع معامل الارتباط الخطي |
| $R^2$Ajusté | معامل التحديد المضبوط |
| Rc | مؤشر canonical index(Rc) |
| λ | Wilks lambda (λ)معامل الاحصاء |
| λi | القيمة الأصلية للمصفوفة |
| ($\chi$2) | Chi-square |
| s | مصطلح فاصل عادةً ما يكون ثابت بديل Taft، Es |

| | |
|---|---|
| u | هي متوسط مساهمة الجزيء الأصل ، |
| xi | وجود بنية معينة لجزء |
| X | الواصفات الجزيئية |
| Y | المتغير التابع |
| y | y: القيمة التجريبية .Valeur observéeالمرصودة |
| $\widehat{Y}_i$: | $\widehat{Y}_i$: القيمة المحسوبة للنشاط.Valeur estimée المقدره |
| $\overline{Y}$ | $\overline{Y}$ : متوسط قيمة القيم المحسوبة للنشاطValeur prédite المتوقعة |
| n-p | عدد درجات الحرية |
| t | de Student. عدد الملاحضات |
| ti | Résidu studentisé externe |
| n | العدد الاجمالي |
| S | خطأ قياسي في التنبؤ |
| p | عدد طبيعي للواصفات التي تعد هناك ثابت |
| MRn | الانكسارية المولارية للمادة في الموضع n |

**فهرس الاختصارات :**

| الاختصار | المعنى | الترجمة الى العربية |
|---|---|---|
| 1AM | Austin Model 1. | نموذج 1Austin |
| AD | discriminant analysis | التحليل التمييزي |
| ANN | Artificial Neural networks | الشبكات العصبية الاصطناعية |
| CV | validation croisée | التحقق المتقاطع |
| CoMFA | Comparative Molecular Field Analysis | المجال الجزيئي المقارن التحليل |
| CoMSIA | Comparative Molecular Similiarity Indices Analysis | تحليل مؤشرات التشابه الجزيئي المقارن |
| CoMBINE | Comparative Binding Energy Analysis | تحليل الطاقة الملزمة المقارن |
| CoRIA | Comparative Residue Interaction Analysis | تفاعل المخلفات المقارن التحليل |
| CoMMA | Comparative Molecular Moment Analysis | تحليل اللحظات الجزيئية المقارن |
| CoSA | Comparative Spectral Analysis | المقارن التحليل الطيفي |
| CART | Classification and Regression Tree | شجرة التصنيف والانحدار |
| CAS | chemical abstracts service | دائرة المستخلصات الكيميائية |
| ClogP | Coefficient de partage calculé par le logiciel CLOGP | معامل التقسيم المحسوب بواسطة برنامج CLOGP |
| CNDO | Complete Neglect of Differential Overlap | الإهمال الكامل للتداخل التفاضلي |
| DFT | La Théorie de la Fonctionnelle de la Densité. | نظرية الكثافة الوظيفية |

| | | |
|---|---|---|
| المربعات الصغرى الجزئية لتحليل العوامل | Partial least squares factorization | F / PLS |
| المربعات الصغرى الجزئية الجينية | genetic partial least squares | G/PLS |
| أعلى مداري جزيئي مشغول | Highest Occupied Molecular Orbital. | HOMO |
| تلميح تحليل مجال التفاعل | Hint Interaction Field Analysis | HIFA |
| | Holo-QSAR | HQSAR |
| إهمال متوسط للتداخل التفاضلي. | Intermediate Neglect of Differential Overlap | INDO MINDO |
| معامل التوزيع الاوكتانول –الماء | Octanol -Water Partition Coefficient | KOW |
| | (k-Nearest Neighbours) | k-NN |
| لوغاريتم معامل التوزيع الاوكتانول – الماء | Octanol -Water Partition Coefficient logarithme | LogKOW |
| معامل التقسيم اوكتانول– الماء | le coefficient de partage Octanol-Eau | LogP |
| أدنى مداري جزيئي غير مشغول. | Lowest Unoccupied Molecular Orbital. | LUMO |
| اترك لمرة واحدة | Leave-one-out (LOO) | Loo |
| اترك الكثير للخارج | Leave-many-out (LMO) | LMO |
| اترك بعضا | leave-some-out (LSO) | LSO |
| مغادرة المجموعة | Leave-Group-Out | LGO |
| الانحدار الخطي المتعدد | Multiple Linear Regression | MLR |

| | | |
|---|---|---|
| صنع القرار المتعدد اللغات | multicriteria decision making | MCDM |
| ميكانيكا الجزيئية | Molecular mechanics | MM+ |
| الخطأ المطلق | mean absolute error | MAE |
| المربعات الصغرى الجزئية وتصحيح الإشارة المتعامدة. | Partial least squares and orthogonal sign correction. | OSC/PLS |
| منظمة التعاون الاقتصادي والتنمية | Organisation de coopération et de développement économiques | OCDE |
| انحدار المربعات الصغيرة | Partial Least Square Regression | PLSR |
| انحدار المكونات الرئسية | Principal component regression | PCR |
| بيانات التنبؤ "مجموعة الاختبار | Prediction data 'test set | PSET |
| المتراكم البيولوجي والسمي | Bio accumulative and toxic | PBT |
| المجموع المتوقع المتبقي لـ المربعات | the predicted residual sum of squares (PRESS) | PRESS |
| معيار نموذج رقم3 | Parameterized Model number 3 | 3PM |
| العلاقة الكمية البنية الكيميائية –الفعالية | Quantitative structure - activity relationships | QSAR |
| العلاقةالكمية البنيةالكيميائية–الخاصية | Quantitative Structure - Property Relationships | QSPR |
| تنسيق تقارير نموذج QSAR | The QSAR model reporting format | QMRF |
| بيانات التعلم "مجموعة التدريب | Training Set | TSET |
| عامل التسامح TF | facteur de tolérance | TF |
| تسجيل تقييم وترخيص المواد الكيميائية | Registrement, Evaluation et Autorisation des produits CHimiques. | REACH |

| | | |
|---|---|---|
| الانحراف الجذر التربيعي | The root-mean-square deviation | RMSD |
| نموذج Recife1 | Recife Model 1 | Rc |
| الانحدار الخطي البسيط | Simple Linear Regression | SLR |
| مواصفات الإدخال الخطي المبسط للجزيئات | Simplified molecular-input line-entry system | SMILES |
| النمذجة المستقلة الناعمة لتشبيه الفئة | Soft independent modeling of class analogy | SIMCA |
| آلات ناقلات الدعم | Support vector machines | SVM |
| الانحراف المعياري | standard deviation | SD |
| الانحراف المعياري لخطأ التنبؤ | standard deviation of error of prediction | SDEP |
| عامل تضخم التباين | the variance inflation factor | VIF |

# مقدمة عامة

## مقدمة عامة

إن البحث عن وجود أي ارتباط بين الأشياء أو الملاحظات المختلفة هو ميل بشري طبيعي. ومن ثم تم تطوير مفهوم نمذجة الكمية بين البنية والنشاط (QSAR) من فكرة الوجود المحتمل لعلاقة رياضية بين طبيعة المواد الكيميائية ومظاهرها السلوكية. بعد الملاحظات الأولية ، تم تصور "الطبيعة الكيميائية" واستكشافها من حيث المعلومات المستخرجة من هيكل وخصائص المواد الكيميائية. تتأثر الخصائص الناتجة عن المواد الكيميائية بشكل كبير بطبيعة تركيبها الكيميائي، بالإضافة إلى الخصائص الفيزيائية والكيميائية المختلفة.

تعد نمذجة العلاقة الكمية بين البنية والنشاط (QSAR) هي النهج الرئيسي لتنسيقات المسار لاستكشاف واستغلال الاعتماد على أنواع الأنشطة أو الخصائص الكيميائية أو البيولوجية أو السمية أو غيرها من الأنشطة;وواحدة من الأدوات الحسابية الرئيسية فقد نالت الثناء على مدار تاريخها بأكمله، فقد أثبتت نجاعتها بموثوقيتها وحدودها ونجاحاتها[1].

وتم استخدام نمذجة QSAR تقليديًا كنهج تحسين رئيسي في محاولة لتلبية حاجة الكيمياء الطبية ورغبتها في التنبؤ بالاستجابة البيولوجية [2] ووجدت طريقها إلى ممارسة الكيمياء الزراعية ، والكيمياء الصيدلانية ، وتصميم الأدوية، علم السموم ، الكيمياء الصناعية والبيئية ومعظم جوانب الكيمياء[1].وتمارس نمذجة QSARعلى نطاق واسع في الأكاديميات والصناعة والمؤسسات الحكومية حول العالم [3]و تجد نماذج QSARتطبيقًا واسعًا لتقييم التأثيرات المحتملة للمركبات الكيميائية والمركبات نانو على صحة الإنسان والأنظمة البيئية .

وفي السنوات الأخيرة، وجدت نمذجة QSAR تطبيقات أوسع في الاكتشافات الناجحة والرائدة وفي تقييم المخاطر الكيميائية. تم تمكين هذه التطورات من خلال البروتوكولات المحسنة لتطوير النموذج والأهم من

ذلك، التحقق من صحة النموذج الذي يتركز على تطوير النماذج مع قوة تنبؤ خارجية وتم استكشاف حديث لبرنامج QSARINS لتطوير طرق نماذج QSARالتحقق من صحتها بشكل مستقل.

الهدف من هذا العمل هو :

1) تقديم نظرة عامة لتاريخ تطور QSAR .

2) معرفة البيانات والاختبارات الإحصائية القياسية الحديثة .

3) معرفة طرق التحقق من صحة نماذج QSAR الحديثة .

4) التطرق للعديد من الأساليب و التطبيقات الجديدة والناشئة لنمذجةQSAR .

## المراجـع


1. Artem Cherkasov , Eugene N. Muratov, Denis Fourches , QSAR Modeling: Where have you been Where are you going to? Published in final edited form as: J Med Chem. 2014 June 26; 57(12): 4977–5010. doi:10.1021/jm4004285J Med Chem. Author manuscript; available in PMC 2015 June 26.
2. Hansch CLA. *Substituent constants for correlation analysis in chemistry and biology*. New York: John Wiley and Sons; 1979.
3. Paola Gramatica, QSAR Research Unit in Environmental Chemistry and Ecotoxicology, DBSF, A SHORT HISTORY OF QSAR EVOLUTION Insubria University, Varese, Italy Paola.gramatica@uninsubria.it; http://www.qsar.itt. January 2011.

# الجانب النظري

# الفصل الأول

## مراحل تطور QSAR

## I - مراحل تطور QSAR:

مع وصول أساليب الحساب النظرية المتطورة بشكل متزايد و المزيد من الموارد الحسابية التي يمكن الوصول إليها ، أصبحت النمذجة الجزيئية الآن أداة مفيدة بشكل متزايد لكل من الصناعة والأوساط الأكاديمية [1]. تتضمن النمذجة الحاسوبية للجزيء عمومًا عرض رسومي للهندسة أو تكوين ذرات الجزيء متبوعًا بتطبيق الطريقة النظرية [2]. تعرف النمذجة بأنها عبارة عن خطوات متتالية لبناء نموذج يتناول الواقع و تترجم أجزاءه في متغيرات ومعادلات تربط بينهما علاقة معينة.

- ترجع أسباب استخدام النمذجة إلى :

- ✓ قلة التكلفة .
- ✓ سهولة التنفيذ ودقة الأداء.
- ✓ وفرة الجهد والوقت.
- ✓ المرونة.
- ✓ إمكانية تطبيق النمذجة على مدى واسع من مجالات الحياة المختلفة[3].

### I-1 النمذجة الجزيئية:

كلمة "الجزئية" تعني وضوح التعامل مع الجزيئات، ويعرف قاموس اوكسفورد النموذج بأنه"وصف مبسط أو مثالي لنظام أو عملية .في المصطلحات الرياضية كثيرا ما يستخدم لتسهيل العمليات الحسابية والتوقعات". تهتم النمذجة الجزئية بتقليد سلوك أنظمة الجزيئي والجزيئات[3].

النمذجة الجزيئية مصطلح عام يشمل تقنيات مختلفة الرسومات الجزيئية والكيمياء الحسابية لعرض ومحاكاة، تحليل وحساب وتخزين خصائص جزيئات[4]. الغرض من النمذجة الجزيئية هو التنبؤ بهيكل وتفاعل الجزيئات أو أنظمة الجزيئات [5]. غالبًا ما يرتبط البحث عن مركبات كيميائية وكيميائية حيوية

جديدة وتوليفها بدراسة عن طريق النمذجة الجزيئية. تشمل طرق النمذجة الجزيئية ما يلي: ميكانيكا الكم والميكانيكا الجزيئية والديناميك الجزيئية.

النمذجة الجزيئية (الميكانيكا والديناميكية الجزيئية) هي طريقة يسمح بتحديد بنية وطاقة الكيانات الجزيئية. [5]. يمكن النظر إلى النمذجة الجزيئية على أنها مجموعة من التقنيات طرق حسابية تعتمد على طرق الكيمياء النظرية والبيانات التجريبية التي يمكن استخدامها إما لتحليل الجزيئات والأنظمة الجزيئية أو للتنبؤ بالخصائص الجزيئية والكيميائية والبيوكيميائية [6]. إنه بمثابة جسر بين النظرية والخبرة من أجل:

1. استخراج النتائج لنموذج معين.
2. مقارنة النتائج التجريبية للنظام.
3. مقارنة بين التنبؤات النظرية للنموذج.
4. المساعدة في فهم وتفسير الملاحظات التجريبية.
5. الارتباط بين التفاصيل المجهرية على المستوى الذري والجزيئي والخصائص المرئية.
6. تقديم معلومات غير متوفرة من تجارب حقيقية. [7]

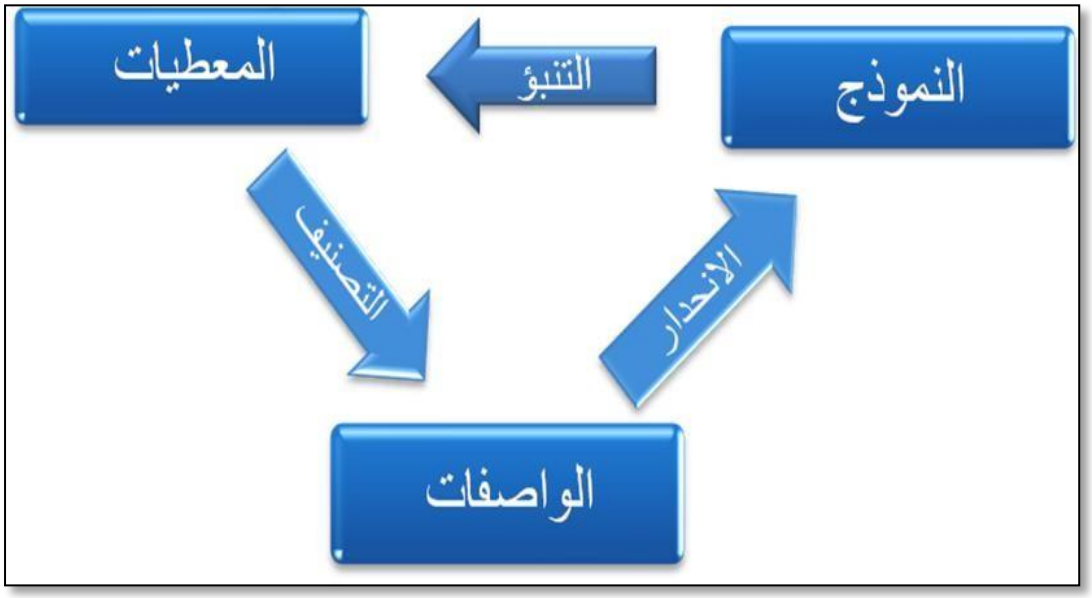
**I-2 النموذج الرياضي:**

النموذج الرياضي عبارة عن مجموعة من المعادلات التي تصف سلوك النظام وهذه المعادلات يكون عددها مساويا لعدد التغيرات فيها وعامة هذه المعادلات مبرمجة حيث تحل بواسطة الكمبيوتر وهذه الطريقة الأنسب للتعامل مع المعادلات الرياضية التي تتطلب حلا عدديا. [3].

**I-3 مزايا استخدام النماذج الرياضية:**

توجد مزايا كثيرة لاستخدام النماذج الرياضية ومنها:

- ✓ تساعد الباحث على ترتيب المعتقدات النظرية والملاحظات الميدانية 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.
- ✓ اقل تكلفة من النظام.
- ✓ تساعد النماذج الرياضية على سرعة التحليل.
- ✓ تساعد النماذج الرياضية على فهم النظام.
- ✓ الاهتمام بالوصف من خلال التفصيل والعرض. [3]




**الشكل (I-1) : خطوات النمذجة الجزيئية**

### I-4 تاريخ تطور :QSAR

تستند QSAR (الكمية الهيكلية – علاقات النشاط) إلى افتراض أن هيكل الجزيء (أي خصائصه الهندسية والفلكية والإلكترونية) يجب أن يحتوي على السمات المسؤولة عن خواصه الفيزيائية والكيميائية والبيولوجية ، وعلى القدرة على تمثل كيميائي بواحد أو أكثر من الواصفات العددية. من خلال نماذج

QSAR، يمكن الاستدلال على النشاط البيولوجي (أو الخاصية، والتفاعلية ،...الخ) لمادة كيميائية جديدة أو غير مختبرة من التركيب الجزيئي لمركبات مماثلة تم بالفعل تقييم أنشطتها. [8]

ويتم استخدام اختصار QSPR (البنية الكمية – علاقة الملكية) عند نمذجة خاصية ما. لقد مضى ما يقرب من 40 عامًا منذ أن تم استخدام نمذجة QSAR أولاً في ممارسة الكيمياء الزراعية وتصميم الأدوية وعلم السموم والكيمياء الصناعية والبيئية. قد تعزز قوتها المتزايدة في السنوات التالية أيضًا إلى التطور السريع والواسع في المنهجيات والتقنيات الحسابية التي سمحت بتحديد وصقل المتغيرات والأساليب العديدة المستخدمة فيمجال النمذجة. [8]

**I-4-1 نشأة QSAR :**

**I-4-1-1مرحلة الأولى:**

بدأت المحاولة الأولى لربط الخصائص النظرية للسلوك الكيميائي والفيزيائي مع الهياكل منذ حوالي 2500 عام مع أفلاطون. وصف أفلاطون تفاعلgasification التغويز كتحول للنار إلى هواء في محلول مائي، بافتراض معرفة بنية المواد المتفاعلة [9] كما هو موضح في معادلة (I-1)

$$2\text{Fire} \xrightarrow{\text{محلول}} 1air \begin{cases} \text{fire} = \text{pyramid} \\ water = \text{icosahedron} \\ \text{air} = \text{octahedron} \end{cases}$$

ويمكن أيضا اعتبار Mendeleev من أسلاف QSAR، نتيجة للتنبؤ بالعناصر الجديدة وخصائصها التي قادته إلى بناء الجدول الدوري للعناصر في عام 1869 [10].

**I-4-1-2 مرحلة الثانية:**

ولكن بعد تتبع تاريخ QSAR بشكل معقول، تم الإبلاغ عن التجارب الأولى للربط بين استجابة بيولوجية أو خاصية فيزيائية كيميائية لسلسلة من المركبات ذات السمات الهيكلية في القرن التاسع عشر.

والتحليلات الكمية حيث يرتبط النشاط الحيوي رياضيًا بمجموعة من العوامل باستخدام الاعتبارات الإحصائية جاءت لاحقًا [9].

**I-4-1-3 مرحلة الثالثة:**

نشأت نمذجة QSAR في مجال علم السموم. في الواقع، كانت بمحاولات تحديد العلاقات بين التركيب الكيميائي وفعالية السمية الحادة جزءًا من طبيعة السمية لأكثر من 100 عام. دفاعا عن أطروحته بعنوان "Action de l'alcohol amylique sur l'organisme " في كلية الطب، جامعة Strasbourgفي فرنسا في 9 يناير 1863، لاحظ كروس وجود علاقة بين سمية الكحوليات الأليفاتية الأولية وقابليتها للذوبان في الماء. أظهرت هذه العلاقة البديهية المركزية للهيكل ونمذجة السمية وسمية المواد تخضع لخصائصها، والتي بدورها تتحدد من خلال تركيبها الكيميائي. لذلك، هناك علاقات متبادلة بين البنية والخصائص والسمية [8]. أن سمية الكحول في الثدييات زادت مع انخفاض قابلية ذوبان الكحول في الماء [11].

**I-4-1-4 مرحلة الرابعة :**

منذ أكثر من قرن من الزمان،أعرب Crum-Brown وFraser [12] عن فكرة أن الفزيولوجيا لعمل مادة ما في نظام بيولوجي (Φ) هو وظيفة (f) في تركيبها الكيميائي .

$$.\boldsymbol{Q = \mathbf{f}(\mathbf{c})(\mathbf{2} - \mathbf{I})}$$

و بالتالي، فإن أي تغيير في التركيب الكيميائي C، سينعكس من خلال تغيير في المواد البيولوجية نشاط $\Delta \mathbf{Q}$ .

واقترحوا معادلة تربط التغييرات في النشاط البيولوجي بالتغيرات في التركيب الكيميائي لكنهم لم يظهروا طريقة لتوصيف التركيب الكيميائي من الناحية الكمية. من هذا الأساس، عبر Richardson

عن التركيب الكيميائي كدالة للذوبان [13]. فتح مجال QSPR، الذي يعتبر متضمنًا داخل QSAR ، Mills تم تطويره في عام 1884 على QSPR للتنبؤ بنقاط الانصهار والغليان في سلسلة متجانسة ودقيقة إلى أفضل من درجة [14]. في وقت لاحق، ربط Richet سمية مجموعة من مشتقات الكحول والإيثرو الكيتونات بقابلية الذوبان في الماء [15] مما أسس المبدأ التجريبي "كلما زادت قابليتها للذوبان، كانت أقل سمية " ، أي "أكثر قابلية للذوبان، وأقل سمية".

**I-4-1-5 تعريف QSAR :**

نموذج QSAR البنية الكمية –علاقات النشاط (هو ارتباط يعتمد على حساب إحصائي للعوامل الفيزيائية والكيميائية مع النشاط البيولوجي QSAR [16]. تعتمد طرق QSAR على افتراض أن نشاط أو خاصية يرتبط المركب الكيميائي بتركيبته، وبصورة أدق ينص هذا النهج على ارتباط نشاط وتركيب المركب الكيميائي بخوارزمية معينة رياضيا، ويعتمد على الافتراض الأساسي "المركبات الكيميائية مماثلة لها أنشطة مماثلة" بالإضافة إلى ذلك، عندما تكون العوامل الجزيئية معبراً عنها بالأرقام، يمكننا اقتراح علاقة رياضية أو علاقة نشاط الهيكل الكمي، فيما بينهما [17]. الفائدة الأساسية لهذه المعادلة هي أنه يجب أن تجعل من الممكن تحديد قيم المعلمات التي تتوافق مع أقصى نشاط وبالتالي التنبؤ بنشاط الجزيئات التي لم يتم تصنيعها بعد [18] .

بحكم التعريف، فان QSARهو نموذج رياضي يربط واحد أو أكثر العوامل الكمية المشتقة من التركيب الكيميائي، إلى مقياس كمي منننشاط. [17]

**I-4-1-6 مرحلة الخامسة:**

في الخمسينيات من القرن الماضي، كانت قوة سيطرت معادلة Hammett لحساب الفروق التفاعلية على تفسير تأثيرات بديلة، تضمنت مقالة كتبها Jaffe عام 1953 عدة مئات من هذه المعادلات [19].

في أواخر الخمسينيات من القرن الماضي، وسع taft مفهوم علاقات الطاقة الحرة الخطية لاقتراح وتناسب معادلة لا تشمل التأثيرات الإلكترونية للبدائل فحسب، بل تشمل أيضًا التأثيرات الاحصائية[20] في المقابل ، كان علماء الصيدلة البيوكيميائية في ذلك الوقت يركزون على تأثير التقسيم عاملات الجزيئات على امتصاص الدواء . [21]

**I-4-1-7 مرحلة السادسة :**

في مطلع القرن العشرين، اقترح Meyer وOverton، [22] [23] بشكل مستقل أن التأثير المخدر (المثبط) لمجموعة من المركبات العضوية يتوازى مع معاملات تقسيمolive oil حمض الاوليك/ الماء. أن الفاعلية المخدرة للمركبات العضوية تعتمد على قابليتها للدهون. واستناداً إلى التجارب البيولوجية، فقد ربطوا معاملات التقسيم بفاعلية التخدير.من جانب أخر حدد Overton أيضًا تأثير المجموعات الوظيفية في زيادة أو نقصان لمعاملات التقسيم [9].

بعد ذلك استمر Lazarev لازاريف في St Petersburg من حيث توقف Overton وMeyer، مع تطبيق معاملات التقسيم لتطوير معايير الصحة الصناعية. أبلغ Lazarev عن الارتباطات على مقياس لوغاريتمي، وطور نظامًا لتقدير معاملات التقسيم من التركيب الكيميائي. [9].

في السنوات التالية على الطريقة العضوية الفيزيائية، أدى العمل الأساسي ل Hammett إلى ظهور($\sigma - \rho$) [25] [24] في تحديد التأثيرات البديلة على التفاعلات العضوية، في حين ابتكرtaft طريقة لفصل التأثيرات القطبية والتأثيرية والرنينية وإدخال عامل الفراغي الأول ،Es [26].الطاقة الفراغية لكن أول صياغة رياضية تُنسب إلىFerguson، الذي أعلن عن مبدأ السمية. لاحظ زيادة فاعلية التخدير عند الصعود في متماثل سلسلة من الألكانات أو الألكانولات إلى نقطة حيث فقد الفاعلية، أو على الأقل لم تحدث زيادة أخرى، باستخدام الخصائص الفيزيائية مثل الذوبان في الماء، والتوزيع بين الأطوار، والشعيرات الدموية وضغط البخار [27] .

**I-4-1-8 مرحلة السابعة :**

في عام 1962 نشر Hansch et al دراستهم حول علاقات التركيب والنشاط لمنظمات نمو النبات واعتمادهم على ثوابت Hammett والكارهة للماء [28]. باستخدام الأوكتانول / الماء النظام ، تم قياس سلسلة كاملة من معاملات التقسيم ، وبالتالي تم إدخال مقياس جديد كاره للماء. تم تعريف العامل π ، وهي الكراهية النسبية للماء كبديل ، بطريقة مماثلة لتعريف sigma.

$$\pi = \log PX - \log PH \qquad (3-I)$$

*تمثل* $P$ X و $P$ H معاملات التقسيم للمشتق والجزيء الأصلي ،على التوالى.

لقد أرست مساهمات Hammett وtaft معًا الأساس لتطوير نموذج QSAR بواسطة Hansch وFujita ، والذي جمع الثوابت الكارهة للماء مع ثوابت Hammett الإلكترونية لإنتاج معادلة Hansch الخطية و أشكالها العديدة و الممتدة . هناك إجماع بين علماء السموم التنبؤية الحاليين على أن Corwin Hansch هو مؤسس QSAR الحديث. في المقالة الكلاسيكية [29] ، تم توضيح أنه ، بشكل عام ، النشاط البيولوجي لمجموعة يمكن من وصف المواد الكيميائية "المتجانسة" بنموذج شامل

$$\log I/C\,50 = a\,\pi + b\,\varepsilon + cS + d\,(4-I)$$

حيث C ، تركيز المادة السامة التي تظهر عندها نقطة النهاية (أو التأثير بنسبة 50٪) ،

π ، يرتبط بمصطلح كارهة للماء (هذا ثابت بديل يشير إلى الاختلاف في الكراهية للماء بين المركب الأصلي والنظير البديل ، تم استبداله بالمصطلح الجزيئي الأكثر عمومية ogl معامل تقسيم 1-أوكتانول / الماء ، **log Kow**) ،

I مصطلح إلكتروني، (في الأصل ثابت بديل هاميت ، s ) ومصطلح فاصل ، S (عادةً ما يكون ثابت بديل Taft، Es). بسبب العلاقة المنحنية، أو ثنائية الخطية، بين logI / C50 و الكراهية للماء الموجودة عادة في اختبارات الجرعة الواحدة ، تم إدخال المصطلح التربيعي $\pi 2$ لاحقًا إلى النموذج.

الأساس المنطقي للمعادلة (4) من McFarland [30].الذي افترض أن النشاط النسبي يعتمد على جزيئات النشاط البيولوجي ، مثل مادة سامة ، على:

(1) احتمال (**Pr1**) أن تصل المادة السامة إلى موقع عملها .

(2) احتمال (**Pr2**) أن المادة السامة ستتفاعل مع الهدف في هذا الموقع .

(3) التركيز الخارجي أو الجرعة.

تم تعديله بواسطة Kubinyi في عام (1977) ، الذي طور نموذجًا خطيًا للنشاط مقابل logP بدلاً من نموذج القطع المكافئ.[31]

أدى تحديد هذه النماذج إلى تطور هائل في تحليل QSAR والأساليب ذات الصلة [32]. إلى جانب طريقة Hansch ، تم تطوير منهجيات أخرى أيضًا لمعالجة أسئلة الهيكل والنشاط. تتناول طريقة Free–Wilson دراسات النشاط البنيوي في سلسلة متجانسة كما هو موضح في المعادلة (4).

$$\mathbf{AB = \Sigma a i x i + u} \qquad (\mathbf{5 - I})$$

عندما تكون AB هي النشاط البيولوجي .

u هي متوسط مساهمة الجزيء الأصل .

ai هي مساهمة كل ميزة هيكلية .

يشير xi إلى وجود (xi = 1) أو غياب xi = 0 بنية معينة لجزء.

### I-4-1-9 مرحلة الثامنة:

في السنوات التي أعقبت الستينيات أدت الحاجة إلى حل المشكلات الجديدة ، جنبًا إلى جنب مع مساهمات العديد من الباحثين الآخرين ، إلى توليد آلاف الاختلافات في طريقة Hansch لنمذجة QSAR ، بالإضافة إلى الأساليب الجديدة رسميًا. تم مشاركة Hans Konemann و Gilman Veith [33] [34]، اللذان طورا في أوائل الثمانينيات من القرن الماضي ، نماذج متعددة الطبقات تعتمد على المواد الكيميائية العضوية الصناعية ، في الفضل في إحياء QSAR.

### I-4-1-10 مرحلة التاسعة :

يعد علم QSAR، الذي تأسس على الاستخدام المنهج للنماذج الرياضية وعلى وجهة نظر متعددة المتغيرات ، أحد الأدوات الأساسية لتصميم الأدوية والمبيدات الحديثة وله دور متزايد في العلوم البيئية.

توجد نماذج QSAR عند تقاطع الكيمياء والإحصاء وعلم الأحياء ، في دراسات السموم. يتطلب تطوير نموذج QSAR هذه المكونات الثلاثة:

1) مجموعة بيانات يوفر مقاييس تجريبية لنشاط بيولوجي لمجموعة من المواد الكيميائية .

2) التركيب الجزيئي أو بيانات الخصائص (أي الواصفات أو المتغيرات أو المتنبئين) لهذه المجموعة من المواد الكيميائية .

3) الطرق الإحصائية لإيجاد العلاقة بين مجموعتي البيانات.

العامل المحدد في تطوير QSARs هو توافر بيانات تجريبية عالية الجودة. في تحليل QSAR ، من الضروري أن تكون بيانات الإدخال دقيقة لتطوير ملف نموذج هادف.

في الواقع ، يجب أن ندرك أن أي نموذج QSAR ناتج يتم تطويره يكون صالحًا إحصائيًا فقط مثل البيانات التي أدت إلى تطويره [8] .

ويتم الحصول على البيانات المستخدمة في تقييمات QSAR إما من الطبيعة أو تم إنشاؤها خصيصًا لتحليلات من نوع QSAR. يمكن أن تتكون هذه البيانات من سلسلة متجانسة من المواد الكيميائية أو ضمان التنوع الهيكلي حتى داخل فئة كيميائية. سمح هذا التنوع بتعميم أكثر قوة QSARs قابلة للتطبيق على نطاق واسع. يتم تعريف نموذج الهيكل والنشاط ويقيد بطبيعة وجودة البيانات المستخدمة في تطوير النموذج ويجب تطبيقه فقط ضمن مجال تطبيق النموذج.

يجب على QSAR المثالي أن:

(1) يأخذ بعين الاعتبار عددًا مناسبًا من الجزيئات للحصول على إحصائية كافية التمثيل.

(2) لديها مجموعة واسعة من قوة نقطة النهاية الكمية (أي عدة أوامر من حيث الحجم لنماذج الانحدار أو التوزيع المناسب للجزيئات في كل فئة (أي نشط وغير نشط) لنماذج التصنيف.

(3) قابلة للتطبيق للتنبؤات الموثوقة للمواد الكيميائية الجديدة (التحقق من الصحة ومجال التطبيق)

(4) يسمح بالحصول على معلومات آلية حول نقطة النهاية النموذجية.

تشمل الواصفات الكيميائية المعلمات التجريبية أو الكيميائية الكمية أو غير التجريبية،يمكن قياس أو تقدير الواصفات وتشمل الخصائص الفيزيائية والكيميائية مثل .(logP)

يمكن أن تعتمد الواصفات غير التجريبية على الذرات الفردية أو البدائل أو لكل جزيء، فهي عادة سمات هيكلية .يمكن أن تستند إلى نظرية الطوبولوجيا أو الرسم البياني، وكذلك على هذا النحو، تم تطويرها من معرفة البنية ثنائية الأبعاد ، أو يمكن حسابها من ثلاثي الأبعاد التشكيلات الهيكلية للجزيء .كما تم استخدام مجموعة متنوعة من الخصائص في نمذجة QSAR،

وتشمل هذه الخصائص الفيزيائية والكيميائية، الخصائص الكيميائية الكمية، وخصائص الارتباط .من أمثلة الخصائص الجزيئية توزيع الإلكترون،التصرف المكاني (التشكل والهندسة والشكل) والحجم الجزيئي فيزيائي كيميائي،تتضمن الخصائص واصفات للخصائص الكارهة للماء والإلكترونية

والجزيئية للجزيء بالإضافة إلى الخصائص الأخرى بما في ذلك الذوبان وثوابت التأين .الخصائص الكيميائية الكمية تشمل قيم الشحن والطاقة .تتضمن خصائص الارتباط الجزيئات البيولوجية الكبيرة وهي مهمة في الاستجابات بواسطة المستقبل.[8]

المشكلة الكبيرة تتعلق بالواصفات الجزيئية هي قابليتها للتكاثر: يمكن للقيم التجريبيةأنتختلف اختلافًا كبيرًا حتى عند الإشارة إلى نفس المركب [35] تم تطوير العديد من الأساليب الحساب النظري لـ logP [36][8] [37] ولكن أيضًا في هذه الحسابات ليس من غير المألوف أن يكون لديها اختلافات في عدة أوامر من حيث الحجم [38] .

**I-5 الأبعاد QSAR:**

تشمل التطورات الأخيرة في QSAR استخدام نمذجة QSAR عالية الأبعاد، لذلك من المناسب النظر في ما تعنيه الأبعاد في مصطلحات QSAR. أعطى [39] التعريفات التالية:

1D-QSAR: نشاط على غرار الخصائص الفيزيائية الكيميائية البسيطة مثل logP ؛

2D-QSAR: نشاط على غرار الأنماط الهيكلية مثل الروابط الجزيئية ؛

3D-QSAR: نشاط على شكل هيكل ثلاثي الأبعاد. في نمذجة المستقبلات، يعتمد QSAR على تفاعلات الترابط مع مستقبلات ثلاثي الأبعاد sutrogate.

4D-QSAR: كما هو الحال مع 3D ، ولكن مع تمثيلات متعددة لتشكيل / اتجاه الترابط.

5D-QSAR: كما هو الحال مع 4D ، ولكن مع تمثيلات متعددة لسيناريوهات الملاءمة المستحثة.

6D-QSAR: كما هو الحال مع 5D ، ولكن مع تمثيلات متعددة لنماذج الذوبان. [40]

بدأت QSAR بـ 0D وتطورت إلى 6D. نشأ كل بُعد كحاجة للتغلب على حدود الأبعاد السابقة وأن تكون أكثر إفادة من السابقة.

قام 1D-QSAR بحساب الخصائص الجزيئية مثل الإلكترونية ، وكاره للماء ، والفراغ.

D2-QSAR يأخذ بعين الاعتبار المعلمات الهندسية ، المؤشرات الطوبولوجية ، بصمات الجزيئية ، مساحة السطح القطبية لكنها تستثني الخصائص الفراغية [41].

تركز تقنية QSAR -D3 على المكاني خصائص المجمع. لذلك، يتم معالجة عيوب D2-QSARبطريقةQSAR-3D . تعتمد طرق التوصيف المستخدمة في 3D QSAR على المحاذاة. والطرق المعتمدة على المحاذاة هي:

- ✓ CoMFA (التحليل المجال الجزيئي المقارن) .
- ✓ CoMSIA (تحليل مؤشرات التشابه الجزيئي المقارن)
- ✓ CoMBINE (تحليل الطاقة الملزمة المقارن) .
- ✓ CoRIA (التحليل تفاعل المخلفات المقارن) .
- ✓ HIFA (تلميح تحليل مجال التفاعل) وما إلى ذلك. المحاذاة مستقلة تتضمن الطرق تحليل اللحظات الجزيئية المقارن (CoMMA) .

المقارن التحليل الطيفي (CoSA) ، HQSAR (Holo-QSAR).

يستخدم 3D QSAR أساليب مثل الشبكات العصبية الاصطناعية (ANN) ، طريقة المربعات الصغرى الجزئية (PLS) ، تحليل الكتلة ، و تحليل المكون الرئيسي ، وغيرها لاختيار الواصف يجعله أكثر قوة من 2D QSAR. حتى مع ذلك تواجه تقنية 3D QSAR صعوبات عندما تكون عدد المركبات في مجموعة البيانات كبيرة، ويجب أن تتنازل مع دقة التنبؤ. [42]

ولتغلب على هذه العوائق QSAR-D4 ، تم تطوير QSAR-D4 لمعالجة قضاياا3D QSAR ويمكن تطبيق 4D-QSAR لكل من المستقبلات المستقلة والمستقبلات التحليل المعتمد [42].

علاوة على ذلك، تطورت 5D-QSAR مع إضافة بعد جديد لـ 4D-QSAR، البعد الجديد هو طوبولوجيا الترابط المتعدد. تقوم مجموعات التمثيلات المتعددة المنتشرة في 5D-QSAR بهذا أقل تحيزًا عند مقارنتها بـ [43].

6D-QSAR يحسن استراتيجية 5D-QSAR السابقة من خلال تضمين بُعد آخر للحل وظيفة تساعد في تحليل نماذج الذوبان المختلفة. [44]

**I-6 طرقQSAR الحديثة:**

في طرق QSAR الحديثة ، أصبح من الشائع استخدام مجموعة واسعة من النظريات و الواصفات الجزيئية من أنواع مختلفة ، قادرة على التقاط جميع الجوانب الهيكلية لمادة كيميائية وترجمة التركيب الجزيئي إلى أرقام. تعد الواصفات المختلفة طرقًا أو وجهات نظر مختلفة لعرض الجزيء ، مع مراعاة السمات المختلفة لتركيبه الكيميائي ، ليس فقط أحادي الأبعاد مثل التعداد البسيط للذرات والمجموعات ، ولكن أيضًا ثنائي الأبعاد من الرسم البياني الطوبولوجي أو ثلاثي الأبعاد من الحد الأدنى من تشكيل الطاقة. تحسب الكثير من البرامج مجموعات واسعة من الواصفات النظرية المختلفة، من SMILES والرسوم البيانية ثنائية الأبعاد إلى إحداثيات x-3D و y و z. بعض البرامج أكثر استخداما مذكورة كالتالي: [45] ADAPT[46] و OASIS[47] و CODESSA[48] و MolConnZ[49] و DRAGON[50].

الميزة العظيمة للواصفات النظرية هي أنه يمكن حسابها بشكل متجانس بواسطة برنامج محدد لجميع المواد الكيميائية، حتى تلك التي لم يتم تصنيعها بعد، والحاجة الوحيدة هي بنية كيميائية مفترضة، وبالتالي فهي قابلة للتكرار.

طرق النمذجة المستخدمة في تطوير QSARs هي من نوعين فيما يتعلق بالاستجابة النموذجية:

فعالية نقطة النهاية (قيمة محددة لـ EC50) أو فئة / فئة (على سبيل المثال Mutagen / Not mutagen). لنمذجة الفاعلية.

فإن الأسلوب الرياضي الأكثر استخدامًا هو تحليل الانحدار المتعدد (MRA). تحليل الانحدار هو نهج بسيط يؤدي إلى نتيجة يسهل فهمها، ولهذا السبب، يتم اشتقاق معظم QSAR باستخدام تحليل الانحدار.يعد تحليل الانحدار وسيلة قوية لإنشاء ارتباط بين المتغيرات المستقلة (الواصفات الجزيئية X) والمتغير التابع Y ، مثل النشاط البيولوجي:

$$\mathbf{Y = b + aX1 + c\,X2 + \cdots (6 - I)}$$

لنمذجة الفئات، يمكن تطبيق نماذج كمية مختلفة من التصنيف. توجد مجموعة واسعة من طرق التصنيف، بما في ذلك:

التحليل التمييزي (DA ؛ الخطي ، التربيعي و DA المنتظم) ،

SIMCA (النمذجة المستقلة الناعمة لتشبيه الفئة) ،

k–NN (k–Nearest Neighbours) ،

CART (شجرة التصنيف والانحدار) ،

الشبكة العصبية الاصطناعية، آلة المتجهات الداعمة،الخ.

في هذه التقنيات ، المصطلح " الكمي "يشير إلى القيمة العددية للمتغيرات (الواصفات الجزيئية) اللازمة لتصنيف المواد الكيميائية في الأصناف النوعية. ويتضح من تحليل طبيعة أن عالم QSAR

قد خضع لتغييرات عميقة منذ العمل الرائد "Corwin Hansch"، الذي يعتبر مؤسس نمذجة QSAR الحديثة. [28] [29] [32]ينعكس التغيير الرئيسي في نمو نهج مفاهيم مواز ومختلف تمامًا لـ نمذجة العلاقات بين بنية مادة كيميائية ونشاطها / خصائصها. في طريقة Hansch، الذي لا يزال مطبقًا

على نطاق واسع ويتبعه العديد من مصممي QSAR، على سبيل المثال .Schultz et al[51]، [52] Veith and Mekenyan،Benigni [53] .

يتم تمثيل التركيب الجزيئي بعدد قليل فقط منالواصفات الجزيئية (عادةً logKow ، وثوابت هاميت ، HOMO / LUMO ، بعض المعلمات الفراغية) يتم اختيارها شخصيًا بواسطة العارض وإدراجها في معادلة QSAR لنمذجة نقطة النهاية المدروسة. ، وفي نهج مختلف يتم تمثيل التركيب الكيميائي ،

بخطوة التمهيدية أولى وتكون بواسطة عدد كبير من الواصفات الجزيئية النظرية، وفي الخطوة الثانية ، اختيار طرق القياس الكيميائي المختلفة كأفضل ارتباط مع الاستجابة والمضمنة في نموذج QSAR(الخوارزمية).

والهدف الأساسي هو تحسين أداء النموذج للتنبؤ وفقًا طريقة Hansch ، يتم توجيه اختيار الواصف من خلال اقتناع المصمم بالحصول على معرفة مسبقة بآلية النشاط / الخاصية المدروسة ، وافتراض تعيين معنى ميكانيكي لأي واصف جزيئي مستخدم يختاره المصمم من بين مجموعة محدودة من متغيرات النمذجة المحتملة ، المعروفة عادةً جيدًا والمستخدمة بشكل متكرر (على سبيل المثال: logKow هو معامل عالمي تحاكي تغلغل غشاء الخلية ، وبالتالي فهي تستخدم في الكثير من نماذج السمية ، ولكنها مرتبطة أيضًا بمعاملات التقسيم المختلفة مثل التركيز الأحيائي / التراكم الحيوي، معامل امتصاص التربة ؛ يتم اختيار HOMO / LUMO دائمًا لنمذجة التفاعل الكيميائي .....الخ).

**I-7الإحصاء :**

**I-7-1 تعريف النهج الإحصائي:**

إن الطريقة"الإحصائية" أو المقياس الكيميائي، وهي طريقة موازية لما يسمى بالطريقة "الميكانيكية" السابقة، و يستند إلى الاقتناع الأساسي بأن مصمم QSAR يجب ألا يؤثر، بشكل مسبق وشخصي، على اختيار الواصف من خلال افتراضات آلية، ولكن يجب تطبيق أدوات رياضية غير متحيزة لتحديد،

من بين مجموعة واسعة من واصفات المتداخلة، تلك الواصفات الأكثر ارتباطًا بالاستجابة المدروسة. يجب أن يكون عدد وتصنيفات واصفات المتداخلة المتاحة واسعًا ومختلفًا قدر الإمكان من أجل ضمان إمكانية تمثيل أي جانب من جوانب التركيب الجزيئي. تعد الواصفات المختلفة طرقًا أو وجهات نظر مختلفة لعرض الجزيء،ولكن يجب تطوير النماذج مع مراعاة مبدأ التباين، المسمى بـ Ockham's Razor "لا يجب مضاعفة الكيانات بما يتجاوز الضرورة" أو "تجنب التعقيد إذا لم يكن ضروريًا" . غالبًا ما تتم إعادة صياغة هذا المبدأ على أنه "الحل الأبسط أي الأفضل". وبالتالي يجب إجراء اختيار الواصف من خلال تطبيق الأساليب الرياضية (مثل التقنيات التطورية، والخوارزميات الجينية، ...الخ) مع الهدف النهائي والحاسم لتعظيم، كمعامل تحسين، القوة التنبؤية لنموذج QSAR، حيث تعتبر كأداة حقيقية. من أي نموذج يعتبر تنبؤاته [40].

**I-7-2 تفسير الواصفات:**

فيما يتعلق بقابلية تفسير الواصفات، من المهم أن نأخذ في الاعتبار أن الاستجابة النموذجية غالبًا ما تكون نتيجة لسلسلة من الآليات البيولوجية أو الفيزيائية والكيميائية المعقدة، وبالتالي من الصعب للغاية الاختزال لإعطاء أهمية كبيرة للمعنى الميكانيكي لـ الواصفات الجزيئية المستخدمة في نموذج QSAR. ويجب أيضًا تسليط الضوء على أنه في النماذج متعددة المتغيرات مثل نماذج MLR ، على الرغم من أن تفسير الواصف الجزيئي المفرد يمكن أن يكون مفيدًا بالتأكيد ، إلا أن مجموعة الواصفات المختارة هي فقط القادرة على نمذجة النهاية المدروسة. [40]

إذا كان الهدف الرئيسي لنمذجة QSAR هو سد الفجوات في البيانات المتاحة، فينبغي تركيز انتباه مصمم النماذج على جودة النموذج. فيما يتعلق بهذه النقطة، يقول Livingstone [53]: "تعتمد الحاجة إلى القابلية للتفسير على التطبيق، نظرًا لأن النموذج الرياضي الذي تم التحقق من صحته والذي يربط خاصية الهدف بالسمات الكيميائية قد يكون، في بعض الحالات، كل ما هو ضروري ، على الرغم من

أنه من المستحسن محاولة بعض التفسيرات "للآلية" من الناحية الكيميائية ، ولكنها غالبًا ما تكون غير ضرورية ، في حد ذاتها ".

و اتخذ Zefirov وPalyulin نفس الموقف حيث يميزان بين تقارير QSAR التنبؤية ، حيث يتعلق الاهتمام بشكل أساسي بأفضل جودة للتنبؤ ، من الوصفQSARs حيث يتم إيلاء اهتمام كبير لقابلية التفسير التوصفي[55].

في الواقع ، يجب أن يكون الهدف الأول لأي مصمم نموذج هو التحقق من صحة التطبيق التنبئي لنموذج QSAR لكل من المنهج الآلي والإحصائي. "Kubinyi Paradox " الشهيرة [56] [57] ، والتي أكد عليها أيضًا "Tropsha" في أوراقهم الشهيرة: يجب الحذر من $Q^2$ .[58] وأهمية أن تكون النموذج جيدًا [59] هي:

النماذج "الأفضل" ليست هي الأفضل للتنبؤ! في الواقع، يجب أن يكون نموذج QSAR، أولاً وقبل كل شيء، نموذجًا حقيقيًا وقويًا وتوقعيًا ، حتى يتم اعتباره نموذجًا موثوقًا به [60] ؛ فقط نموذج مستقر وتنبؤ يمكن تفسيره بشكل مفيد لمعناه الآلي ، ومع ذلك ، فإن هذا ليس دائمًا سهلاً أو ممكنًا.

أقرت مجموعات خبراء منظمة التعاون الاقتصادي والتنمية بالتحقق من صحة نموذج QSAR كنقطة حاسمة وعاجلة في السنوات الأخيرة ، وقد أدى ذلك إلى تطوير "مبادئ منظمة التعاون والتنمية في الميدان الاقتصادي للتحقق من نماذج (QSAR) [61] ، لأغراض تنظيمية. ترجع الحاجة إلى هذا الإجراء المهم بشكل أساسي إلى سياسة المواد الكيميائية الجديدة الأخيرة للمفوضية الأوروبية ( REACH تسجيل، تقييم وترخيص المواد الكيميائية) [62]، التي تنص صراحة على الحاجة إلى استخدام نماذج QSAR لتقليل الاختبارات التجريبية (بما في ذلك التجارب على الحيوانات). من الواضح، لتلبية متطلبات تشريعات REACH من الضروري استخدام نماذج QSAR التي تنتج تقديرات موثوقة، أي نماذج

QSAR التي تم التحقق من صحتها. وبالتالي، يجب أن يرتبط نموذج QSAR الموثوق به بالمعلومات التالية:

1. نقطة نهاية محددة.
2. خوارزمية لا لبس فيها.
3. مجال محدد للتطبيق.
4. مناسب مقاييس جودة الملاءمة والمتانة والتنبؤ.
5. تفسير آلي إن أمكن. تعتمد الحاجة إلى القابلية للتفسير على التطبيق، حيث قد يكون كل ما هو ضروري لنموذج رياضي تم التحقق منه يربط خاصية الهدف بالسمات الكيميائية، لا سيما عند التنبؤ.

هناك حاجة إلى البيانات لفحص مكتبات كبيرة من المواد الكيميائية، على الرغم من أنه من الواضح أنه من المرغوب فيه المحاولة بعض التفسيرات "للآلية" من الناحية الكيميائية.

**I-8 البيانات الاحصائية لبداية بناء نموذج:**

ولكي نستفيد من نموذج QSAR فلابد وأن يكون لديه الآن تدابير مناسبة لخير الملاءمة والقوة والتنبؤ [63] (منظمة التعاون الاقتصادي والتنمية، 2007). وأبلغ العمل المبكر ل QSAR عن عدد قليل من الإحصاءات، وفي بعض الأحيان لم يبلغ عن أي إحصاءات. وعادة ما تم الإبلاغ عنQSARمع معامل الارتباط (r) وخطأ قياسي في التنبؤ (s) أو مقياس آخر للخطأ مثل الجذر متوسط الخطأ المربع .(RMSE) وهذا يعطي دلالة على الخير من الملاءمة، ولكن ليس من القوة أو التنبؤ. وإحدى طرق تقييم الصلابة هي استخدام إجراء المصادقة الداخلية مثل الإجازة، حيث يتم إزالة مركب واحد، وإعادة تطوير QSARعلى المركبات المتبقية وإعادة حساب معامل الترابط. ثم يعاد المركب المزال ويزال مركب ثان، ويتكرر الإجراء، وهكذا حتى كل المركبات وقد أزيلت واحدة تلو الأخرى، وتم حساب

معامل ترابط عام. وتجدر الإشارة إلى أنه لا يزال هناك بعض الجدل بشأن قيمة Q، على الأقل فيما يتعلق بقدرته التنبؤية. [64-66]

وتكتسي القدرة على التنبؤ أهمية حيوية، لأن أحد المتطلبات الرئيسية لنموذج QSPR /QSAR هو أنه ينبغي أن يكون قادرا على التنبؤ بدقة معقولة بالنشاط/الممتلكات اللازمة للمركبات غير المستخدمة في تطوير النموذج.

وقد أصبح من المقبول الآن على نطاق واسع أن يتم التنبؤ بحركة من هذا النوع عن طريق المصادقة الخارجية. أي أن QSAR تستخدم للتنبؤ بأنشطة مجموعة اختبارو تتألف من مركبات مماثلة لتلك المستخدمة في مجموعة التدريب، ولكنها لم تستخدم في مجموعة التدريب. ويتم هذا عادة من خلال تقسيم مجموعة بيانات إلى مجموعتين، مجموعة تدريب (والتي سوف يتم تطوير QSAR معها) ومجموعة اختبار ، عادة بنسبة 80 %/20%. كما يوصىGramatica[60] ألا يتم التحقق الخارجي من صحة الآليات على QSAR التي تم تطويرها باستخدام مجموعات البيانات الصغيرة "25 مركباً".

ومن بين الإحصائيات المفيدة الإضافية: $r^2$ لأخطاء القياسية لمعامل كل وصفي، وتعطي إشارة إلى الإدراج الصحيح للوصف في منطقة QSAR ؛ وإحصائية فيشر ، التي تعطي مؤشراً على احتمال أن يكون هناك ارتباط بين نماذج QSAR [66] .

**I-9طرق نمذجة QSPR/QSAR:**

مع توفر العديد من الواصفات، وأن أي أحدا بدائيًا لا يعرف في كثير من الأحيان أي واصف (R) سيعطي نموذجًا جيدًا، فإن العديد من ممارسي QSAR يبدؤون بمجموعة كبيرة من الواصفات ويستخدمون تقنيات المتغيرات ، مثل التخلص من الأزواج العالية – العلاقة الخطية المتداخلة، لتقليل الرقم. ومع ذلك، لا يزال من الضروري اختيار تلك الأوصاف القليلة التي ستعطي أفضل نموذج واستخدام

تقنيات مثل الانحدار التدريجي، واختيار أفضل اختيار مجموعات فرعية أو وراثية الخوارزميات للقيام بذلك.

أظهر Topliss و Costello [67] أنه إذا كانت نسبة عدد المركبات في مجموعة التدريب المستخدمة لتطوير QSAR إلى عدد الواصفات في QSAR أقل من خمسة ، فهناك خطر كبير من أن الارتباط كان بالصدفة ولم يكن صحيح . أظهر Topliss and Edwards في(1979) لاحقًا أنه إذا تم اختيار الواصفين من بين مجموعة كبيرة جدًا ، فإن ذلك يزيد أيضًا من مخاطر. ويمكن أن تنشأ ارتباطات اخرى.[62]

من المثير للقلق أنه لا يزال يتم تجاهل كلا التوجيهين في كثير من الأحيان. تم نشر إرشادات تطوير نماذج QSPR/QSAR [68–70].

يتم تجاهل العديد من الإرشادات الخاصة بتطوير QSAR والتحقق من صحتها أو استخدامها بشكل غير صحيح، الأمر الذي دفع Dearden وآخرون في عام (2009) [64] . للإبلاغ عن وجود 21 نوعًا مختلفًا من الأخطاء في نشر أوراق QSAR، بما في ذلك، على سبيل المثال لا الحصر ، التركيبات الكيميائية غير الصحيحة، وتكرار المركبات، والأخطاء في قيم الواصف والنقطة النهائية ، والبيانات غير الكافية ، والإحصاءات غير الكافية أو غير الصحيحة ، وغير المعترف بها إغفال نقاط البيانات والإفراط في تجهيز البيانات والتحقق غير الصحيح.

**I–10 طرق QSAR الأخرى :**

إن الطرق البسيط نسبياً الذي ترتبط فيه مجموعة من الأنشطة البيولوجية بواحد أو أكثر من أجهزة الوصف يطلق عليه الانحدار الخطي المتعدد (MLR) ولا يزال هذا النظام منتشراً على نطاق واسع ، ولكنه يعاني من عدد من العيوب :

1 – لا يمكنه التعامل مع العلاقات غير المستقيمة بين النشاط البيولوجي ؛

2 – حساسة جدا للقيم المتطرفة.

3 – وينبغي أن تكون الوصفات مستقلة عن بعضها البعض.

ومن ثم، فقد تم على مدار سنوات عديدة استخدام العديد من تقنيات النمذجة في تحليل QSARI QSPR. تم تطوير العديد من هذه في الأصل لأغراض أخرى غير تحليلQSAR .الشبكات العصبية الاصطناعية (ANN) ، لأنها مشابهة من الناحية المفاهيمية للعصبية وتستخدم الشبكات في الكائنات الحية على نطاق واسع إلى حد ما في نمذجة QSPR/QSAR [71] . وهي تتمتع بميزة أنها يمكن أن تتضمن وظائف غير متطورة ، مثل الشروط المتبادلة و المربعية. أما مساوئها فتتلخص في أنها يمكن بسهولة أن تكون مدربة تدريبا مفرطا، الأمر الذي من شأنه أن يقلل من قدرتها على التنبؤ، وأنها لا تفضي إلى معادلة QSPR / QSAR و الشبكات الوظيفية ذات الأساس الإشعاعي هي بديل ANNS فقد استخدمهم Modarresi, Modarress, و Dearden (2007) [72].

لتطوير نموذج QSPR لثبات قانون هنري (معامل تقسيم الهواء والماء)الشبكات العصبية والمنطق (ART) عبارة عن توليف لنظرية الرنين التكيفي Fuzzy ARTMIP الضبابي وكان أحد الأمثلة على استخدامه في نماذج QSAR هو التنبؤ بقابلية المركبات العضوية للذوبان المائي [73] .وتجدر الإشارة إلى أن الخطأ القياسي للتنبؤ في ذلك العمل كان في حدود 0.08 وحدة LOg، في حين أن الخطأ القياسي لقياس قابلية الذوبان المائي هو حوالي 0.6 وحدة سجل. ولقد أشار Livingstoneفي عام (1995) إلى أنه إذا كان الخطأ القياسي في التنبؤ أقل من الخطأ في القياس ، فقد حدث فرط في ملاءمة النموذج.[74]تحليل المكونات الرئيسية (PCA) عبارة عن تقنية تستخدم تحولاً تقويمياً لتحويل مجموعة من القيم من المتغيرات التي قد ترتبط ببعضها البعض (الوصفية) إلى مجموعة من قيم المتغيرات غير المرتبطة ببعضها البعض تسمى المكونات الرئيسية. ويتم تعريف التحول بحيث يكون للحاسوب الشخصي الأول أكبر تباين ممكن (أي أنه يشكل أكبر قدر ممكن من التباين في البيانات) ،

كل مكون لاحق بدوره لديه اعلى تباين ممكن في إطار القيد بأنه تقويم للمكونات السابقة. وهو بالتالي مفيد في تقليل مجموعة وصفية كبيرة إلى أبعاد يمكن التحكم فيها .التحليل التمييزي هو أحد التقنيات التي يمكن استخدامها في البيانات القاطعة أو التصنيفية، على سبيل المثال النشطة (0/1) ، أو القوية/المعتدلة/الضعيفة. وفي الواقع، فإنه يضع فرط الحركة متعدد الأبعاد بين الفئات ، على الرغم من أنه غالبًا ما وجد أن التمييز الطبقي أقل من 100٪ وقد يُعتقد أنه شبه كمي ، ولكن بما أن قيم الوصف المستخدمة كمية ، فهي كذلك مصنفة على أنها تقنية QSAR/QSPR

إن Logistic regression الانحدار اللوجستي مماثل لتحليل الصمامات؛ وهي تقدر الاحتمالات باستخدام وظيفة لوجستية. لقد ناقشت كل من القيمة استخدام كل من التكتل في تطوير نماذج التصنيف للتأثيرات الصحية البشرية[75].

واستخدم Barratt (1995) [76] كل من PCA والتحليل التمييزي في دراسة عن تآكل الجلد للأحماض العضوية. ومن بين أجهزة الوصف التي تسجل P ، والحجم الجزيئي ، ونقطة الانصهار ، وال .pKa ، حصل على اثنين من الحواسيب الشخصية التي تميز بين الأحماض العضوية المؤكسدة وغير المؤكسدة.

يشبه تحليل المراسلات من الناحية المفاهيمية تحليل المكونات الرئيسية، ولكنه ينطبق على البيانات الفئوية بدلاً من البيانات المستمرة. وقد استخدمه Cronin و Dearden (1997) لتحديد إمكانية التحسس الجلدي للمركبات العضوية [77].إن آلات ناقلات الدعم (SVM) تخضع للإشراف على نماذج التعلم مع خوارزميات التعلم المرتبطة بها والتي تتعرف على أنماط التصنيف وتحليل الانحدار.

وقد استخدمت هذه النماذج في تطوير نماذج QSAR للتنبؤ بنشاط مضاد لفيروس نقص المناعة البشرية في عام (2010) [78]. فهي تسمح بحدود مفرطة أكثر تعقيدا مقارنة بالتحليل التمييزي البسيط . وتستخدم المربعات التكيفية الأقل وظيفة تمييزية واحدة لاتخاذ قرارات بشأن التمييز الجماعي المطلوب.

ولديها ميزة أنها يمكن أن تنظر في وقت واحد في أي عدد من الفئات، وبالتالي يمكن استخدامها في كل من التصنيف والتراجع؛ كما أنه مفيد في التعامل مع البيانات التي تنطوي على أنشطة بتركيزات واحدة أو تركيزين فقط ، عندما لا يمكن حساب وحدة مراجعة الحسابات أو LD50 . لقد قدم Schaper [79] وSaxena في عام (1991) عدة أمثلة على أدائها الواعد في تحليل الانحدار . ويستخدم التحليل العنقودي لتصنيف البيانات إلى مجموعات أو فئات، استنادا إلى الخواص الكيميائية و/أو البيولوجية، قبل إجراء التحليل QSAR . ولذلك يمكن استخدامه كأداة لخفض البيانات. ومن المفيد عندما تتضمن مجموعة البيانات مواد كيميائية ذات آليات عمل مختلفة. وقد استخدمه Yuan وParrill في عام (2005) في التحليل 3 D-QSARلمثبطات فيروس نقص المناعة البشرية(HIV-I) [80].

**I-11 برامج QSAR** :

بما كان أول برنامج QSAR هو برنامج Tile ClogP لحساب قيم Plog، الذي تم تطويره في كلية بومونا في عام(1979)[81]. منذ ذلك الحين، كان هناك العديد من برامج التنبؤ بتسجيل Pتم تطويره ، بما في ذلك بعض برامج شركات الأدوية الداخلية. اختبر Mannhold و Dross في عام (1996)[82] القدرات التنبؤية لـ 14 برنامجًا من هذا القبيل، باستخدام مجموعة من 138 مركبًا متنوعًا من الحيوانات. قام Dearden وآخرون في عام(2003) بتحديث تلك الدراسة[83] ، مرة أخرى باختبار 14 برنامجًا. اختبر Dearden أيضًا عددًا من البرامج للتنبؤ بنقطة الانصهار ونقطة الغليان وضغط البخار في عام (2003) والذوبان المائي في عام (2006) [84] Dearden, وآخرون في (2007) [85],واختبر Liao وNicklaus في 2009 عددًا من برامج البرامج للتنبؤ بـ pKa [86].

في السنوات الأخيرة، تم تطوير المزيد من برامج QSPR/QSAR للتنبؤ بمجموعة واسعة من التأثيرات السامة بالإضافة إلى الخصائص الفيزيائية والكيميائية. ناقش Dearden في(2010 ، 2012)

البرامج المتاحة للتنبؤ بالسمية والخصائص الفيزيائية والكيميائية. مازال مفقود ، خاصة بالنسبة لبرامج السمية ، هو التقييم المستقل للأداء التنبئي المقارن لحزم البرامج المختلفة .[87] [88]

### I-12تطبيقات QSAR:

هناك عدد كبير من تطبيقات نماذج QSAR في الإعدادات الصناعية، في مجال البحث الجامعي والاقتصاد وغيرها من المجالات [89].

نورد أدناه بعض التطبيقات الممكنة لنماذج [92-90] .

- تعظيم النشاط الدوائي أو المبيد الحيوي أو مبيدات الآفات.
- التصميم العقلاني للعديد من المنتجات مثل المواد الخافضة للتوتر السطحي، العطور والأصباغ والمواد الكيميائية الدقيقة.
- التعرف على المركبات الخطرة في المراحل الأولى من التطور.
- التنبؤ بالسمية والآثار الجانبية للمركبات الجديدة.

### I-13 المبادئ الأساسية :

تهدف طريقة QSAR القائم على التمثيل العددي للتركيب الكيميائي إلى فهم كيف يؤثر التباين الهيكلي على النشاط البيولوجي لمجموعة صنف من المركبات. الافتراض الرئيسي هو أن العوامل التي تحكم الأحداث في النظام البيولوجي يتم تمثيلها من خلال الواصفات التي تميز المركبات، والتي يتم التعبير عن نشاطها البيولوجي من خلال نفس الآلية. لذلك يحاول QSAR معرفة خصائص الجزيء التي تؤثر على نشاطه وما يمكن تعديله لتحسين خصائصه. ومن ثم بالنسبة لسلسلة من الجزيئات النشطة بيولوجيًا، من المتوقع أن ينعكس أي اختلاف منهجي في التركيب الكيميائي من واحد إلى آخر في تباين مماثل نسبي في استجابة. يعتمد نموذج QSAR على افتراض وجود علاقة أساسية بين التركيب الجزيئي والنشاط البيولوجي والتي تنشأ من هذا الاختلاف المنهجي أيضًا.

ومن المفترض أن الوصف الفيزيائي الكيميائي متعدد المتغيرات لمجموعة المركبات يكشف عن هذه التشابهات. جميع الخصائص الفيزيائية والكيميائية والبيولوجية لمادة كيميائية يمكن حساب المادة من تركيبها الجزيئي، وتشفيرها في شكل رقمي بمساعدة العديد من الواصفات. ومع ذلك من المهم ملاحظة الفرق بين الارتباط والسببية. لا يعني ارتباط QSAR المرضي أن واصفًا معينًا يتسبب في التأثير الفعال للمركب. قد يتم استكمال نقص الأدلة على السببية بمعلومات إضافية حول الآليات المختلفة المؤدية إلى النشاط البيولوجي.[9]

في الواقع، يرتكز QSAR على مبدأ التشابه المعروف، والذي ينص على تشابه مركبات الأنشطة المماثلة. بالإضافة إلى ذلك، في العلاقات الخطية ...الخ. بالإضافة إلى مبدأ التشابه المذكور أعلاه، ينطبق مبدأ الجوار أيضًا. ينص هذا المبدأ على أن الجزيئات الموجودة في نفس المنطقة من مساحة الواصف تقدم نشاطًا مشابهًا [93].

### I-14 الوصول إلى QSAR الحديثة (QSARINS)

في الآونة الأخيرة، تم تنفيذ العديد من الأفكار حول نمذجة QSAR استنادًا إلى حوالي 20 عامًا من الخبرة في Insubria ، في برنامج QSARINS لتطوير نماذج QSAR والتحقق من صحتها ، بناءً على الخوارزمية الجينية للاختيار المتغير. توجد العديد من أدوات القياس الكيميائي (PCAMCDM...الخ) لتحليل واختيار أفضل النماذج. تتوفر العديد من العوامل وأدوات التحقق من الصحة للتحقق الدقيق، بشكل أساسي على المجموعات الخارجية. تتوفر رسوم بيانية مختلفة لمخطط الانحدار، والتحليل المتبقي، ومجال التطبيق واكتشاف القيم المتطرفة، وما إلى ذلك. يتم أيضًا تنفيذ نمذجة الإجماع بواسطة العديد من النماذج. تجمع وحدة تسمى QSARINS-Chem الهياكل الكيميائية (D3)

والبيانات التي صممتها مجموعة Insubria Group في هذه السنوات. العديد من نماذج Insubria ، المستندة إلى واصفات PaDEL ، والتي يمكن حسابها مباشرة من QSARINS أيضًا للمركبات الجديدة المتاحة وتقابلة للتطبيق على المواد الكيميائية الجديدة التنبؤ و بالخصائص الفيزيائية والكيميائية أو سلوك PBT أو السميات المختلفة. يتم أيضًا الإبلاغ عن QMRF لكل نموذج. وظهور أهم البيانات التي زادت في تحسين قوة التنبؤ .

ويود مطورو نموذج QSAR أن يحصلوا في خطوة واحدة، على مجموعات بيانات كاملة من المواد الكيميائية ، بما في ذلك الملفات البنيوية ولابد وأن تكون قاعدة البيانات هذه في متناول الجميع من خلال مدخلات مختلفة ، على سبيل المثال جهاز تعريف المستخلصات الكيميائية ، ونظام الدخول الخطي المبسط (SMELES) ، والأسماء الكيميائية . ومن الأفضل أن تكون المادة الكيميائية المستجوبة مرئية في ثلاثة أبعاد 3D مع نقاط النهاية المتاحة.وعلى هذا، ولكي نكون مفيدين لنماذج QSAR استخدمنا البرمجيات المفتوحة المصدر المجانية على الإنترنت لحساب الواصفات.

يمكن تطبيق هذه النماذج الجديدة، المصادق عليها من الخارج، من قبل أي شخص يستخدم QSARINS–Chem، بعد أن قام أولا بالحساب الإلكتروني لأجهزة وصف النماذج. ومن الممكن التحقق من مجال انطباق المواد الكيميائية الجديدة من خلال الرسم البياني لإنسوبريا. ويمكن أيضاً الحصول على نموذج الإبلاغ المناظر (QMRF) QSAR، الأمر الذي يسهل إمكانية تسجيل المادة الكيميائية المدروسة في "ريتش" . ومن الممكن أيضاً الحصول على هذا النموذج. وقد تم تضمين بعض أدوات القياس الكيميائي، مثل التناظر بين المكونات الرئيسية (PCA) وصنع القرار المتعدد اللغات (MCDM) في النسخة السابقة للتطبيقات المختلفة في جزء النمذجة (PCA) للتحليل الاستكشافي لمجموعات البيانات واختيار النماذج الأكثر اختلافاً لنماذج الإجماع اللاحقة ، MCDM لاختيار أفضل النماذج). [93]

### QSARINS (QSAR-INSUBRIA) 1-14-I

هو برنامج جديد لتطوير والتحقق من صحة نماذج الانحدار الخطي المتعددة العلاقة بين البنية والنشاط (QSAR) بطريقة المربعات الصغرى العادية والخوارزمية الجينية لاختيار المتغيرات .يركز هذا البرنامج بشكل أساسي على التحقق الخارجي من نماذج QSAR و أدوات متنوعة للتحليل الاستكشافي لمجموعات البيانات عن طريق تحليل المكونات الرئيسي، والتقليل المسبق من الواصفات الجزيئية المدخلة ، وتقسيم مجموعات البيانات إلى مجموعات تدريب وتنبؤ ، واكتشاف القيم المتطرفة والتنبؤات المستكملة أو المستنبطة ، والتحقق الداخلي والخارجي من خلال العوامل مختلفة ، ونمذجة الإجماع ومختلف لتصورات QSARINS .

### QSARINS 2-14-I:

هي منصة سهلة الاستخدام لنمذجة QSAR وفقًا لمبادئ منظمة التعاون الاقتصادي والتنمية ولتحليل موثوقية البيانات المتوقعة التي تم الحصول عليها .يتم تنفيذ نموذج مؤشر Insubria المتراكم البيولوجي والسمي (PBT) للتنبؤ بالسلوك التراكمي للمواد الكيميائية الجديدة أثناء تنفيذ PBTs.بالإضافة إلى ذلك ، يسمح QSARINS للمستخدم بالتحقق من صحة النماذج الفريدة المطورة مسبقًا باستخدام برامج مختلفة أيضًا.

## المراجع الأجنبية :


1. S. Belaidi, Thèse de doctorat, Université de Batna, 2002.
2. T. Clark, ‘’Handbook of Computational Chemistry’’, Edition,Wiley,
3. Samar Bakoben&AhlamHouda, 16/07/2011, MEGBI Training Course Molecular Modelling, Institute for Genetic Engineering, Ecology and Health (IGEEH) Karlsruhe, Germany.120p.
4. P. Kollman, Acc. Chem. Res., 1996, 29, 462.
5. LEMCHOUNCHI, Merzaka (2009) *Etude comparative des pénicillines et des thiazoles antibiotiques par la méthode QSAR.* Master thesis, Université Mohamed Khider Biskra.
6. H.-D. Höltje, G. Folkeis, Molecular Modeling: Basic Principles and Applications. VCH, New-York, 1997.
7. MELKEMI Nadjib,ThèseDOCTORAT EN SCIENCES, *Etude des corrélations entre les paramètres structurauxet les propriétés bioactives de quelques composésorganiques*.
8. Paola Gramatica, QSAR Research Unit in Environmental Chemistry and Ecotoxicology, DBSF, A SHORT HISTORY OF QSAR EVOLUTIONInsubriaUniversity, Varese, Italy Paola.gramatica@uninsubria.it; http://www.qsar.itt.January 2011.
9. Mendeleev, D.I. Principles of Chemistry; Vol. 2, 1868–71, tr. 1905.
10. Borman, S. New QSAR Techniques Eyed for Environmental Assessments. Chem. Eng. News,68, 1990, 20-23.
11. A. Crum-Brown, T.R. Fraser, *Trans. R. Soc. Edinburgh* 1868–1869, *25*, 151
12. Richardson, B.J. Medical Times and Gazette, 2, 1868, 703.
13. Mills, E.J. Philosophical Magazine, 17, 1884, 173-187.
14. Richet, C. Compt. Rend. Soc. Biol. (Paris), 45, 1893, 775-776.
15. V. N. Viswanadhan , A. K. Ghose et J. N. Weinstein, *Biochimica et Biophysica Acta* , 1990, 1039, 356.
16. C. Hansch and E.J. Lien, ―Structure-activity relationships in antifungial agents. A survey‖, Journal of Medicinal Chemistry, 14(8), 1971, 653–670.
17. MEHELLOU Mohammed Nadjib,MASTER ACADEMIQUE, Etude des relations structure/activité quantitatives (QSAR/2D) d'une série de dérivés de Triazolothiadiazoles.2017/2018.
18. Jaffe HH. A Reexamination of the Hammett Equation. Chem. Rev. 1953; 53:191–261.

19. Taft, R. Separation of Polar, Steric, and Resonance Effects in Reactivity. In: Newman, M., editor.Steric Effects in Organic Chemistry. Wiley; New York: 1956. p. 556.
20. Hogben CAM, Tocco DJ, Brodie BB, Schanker LS. On the Mechanism of Intestinal Absorption ofDrugs. J. Pharmacol. Exp. Ther. 1959; 125:275–282. [PubMed: 13642268]
21. H. Meyer, *Arch. Exp. Pathol. Pharmakol*. 1899, *42*, 109.
22. C.E. Overton, *Studien Uber die Narkose*, Fischer, Jena,Germany, 1901.
23. L. P. Hammett, *Chem. Rev*., 1935, *17*, 125
24. L. P. Hammett, *Physical Organic Chemistry*, 2nd ed., McGraw-Hill, New York, 1970
25. R. W. Taft, *J. Am. Chem. Soc*., 1952, *74*, 3120.
26. Ferguson, J. The use of chemical potentials as indices of toxicity. Proc. Roy. Soc. Lond. B. Biol.Sci., 127, 1939, 387–404
27. C. Hansch, P. P. Maloney, T. Fujita, and R. M. Muir, *Nature*, 1962, *194*, 178
28. C. Hansch, T. Fujita, *J. Am. Chem. Soc.* 1964, *86*, 1616.
29. J.W. McFarland, *J. Med. Chem.* 1970, *13*, 1092.
30. Kubinyi, H. (1977). Quantitative structure-activity relationships. 7. The bilinear model, a new model lornonlineardependene of biological atcivity on hydrophobic cbaracter. Journal of Medicinal Chemistry, 20(5}, 625-629 .doi: 10.1 02 IIjm00215a002 PMID:857018
31. C. Hansch, A. Leo, *Exploring QSAR, Fundamentals and Applications in Chemistry andBiology,* ACS Professional Reference Book, American Chemical Society, Washington, DC,1995.
32. H. Ko¨nemann, *Toxicolog*, 1981, *19*, 209.
33. G.D. Veith, D.J. Call, L.T. Brooke, *Can. J. Fish Aquat. Sci*. 1983, *40*, 743.
34. R. Renner, *Environ. Sci. Technol.* 2002, *36*, 410A.
35. R. Mannhold, A. Petrauskas, *QSAR Comb. Sci.* 2003, *22*, 466. ;
36. G. Klopman, J. K. Li, S.Wang, M. Dimayuga, *J. Chem. Inf. Comput. Sci*. 1994, *34*, 752.
37. E. Benfenati, G. Gini, N. Piclin, A. Roncaglioni, M. R. Varı`, *Chemosphere*, 2003, *53*, 1155.
38. Greener, M. (2007,lune). QSAR: prediction beyond the fourth dimension. Drug Disco very Gild Development,
39. Jolm C. Dearden, School ofPharnlacy and Biomolecular Sciences, Liverpool JolmMoores University, Liverpool, UK,The History and Development ofQuantitative Structure-ActivityRelationships (QSARs**)**,International Journal of Quantitative Structure-Property RelationshipsVolume 1 - Issue 1 • January-June 2016

40. SwathikClarancia Peter, Jaspreet Kaur Dhanjal, Vidhi Malik, Navaneethan Radhakrishnan2, MannuJayakanthan and DuraiSundar,Quantitative structure-activity relationship (QSAR): modeling approachesto biological applications,January 2018,DOI: 10.1016/B978-0-12-809633-8.20197-0.
41. Hopfinger, A., Wang, S., Tokarski, J. S., Jin, B., Albuquerque, M., Madhav, P. J. &Duraiswami, C. (1997). Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. Journal of the American Chemical Society, **119,** 10509-10524.
42. Vedani, A. &Dobler, M. (2002). 5D-QSAR: the key for simulating induced fit? Journal of medicinal chemistry, **45,** 2139-2149.
43. Polanski, J. (2009). Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. Current medicinal chemistry, **16,** 3243-3257
44. A.J. Stuper P.C., Jurs, *J Chem Inf Comput Sci*, 1976, *16*, 99.
45. . http://research.chem.psu.edu/pcjgroup/ADAPT.html.
46. O. Mekenyan, D. Bonchev, , *Acta Pharm Jugosl.,* 1986, *36*, 225.
47. A.R., Katritzky, V.S., Lobanov, CODESSA, Version 5.3, University of Florida, Gainesville .1994.
48. MolConnZ, Ver. 4.05, 2003, Hall Ass. Consult., Quincy, MA.
49. R. Todeschini, V. Consonni, A. Mauri, M. Pavan, DRAGON—Software for the calculation of molecular descriptors. Ver. 5.4 for Windows, 2006, Taletesrl, Milan, Italy.
50. T.W Schultz, M.T.D.Cronin, T.I. Netzeva , A.O. Aptula, , *Chem. Res. Toxicol*., 2002, *15*, 1602.
51. G.D Veith, O.G. Mekenyan, , *Quant. Struct-Act. Rel.,* 1993, *12*,349.
52. R. Benigni, *Chem. Rev*., 2005, *105*, 1767.
53. D.J. Livingstone, *J. Chem. Inf. Comput. Sci.*, 2000, *40*, 195.
54. N.S. Zefirov and V.A. Palyulin, *J. Chem. Inf. Comput. Sci*., 2001, *41*, 1022.
55. J. H. van Drie, *Curr. Pharm.Des.* 2003, *9*, 1649.
56. J. H. van Drie, in: Computational Medicinal Chemistry for Drug Discovery, P. Bultinck et al., Eds., Marcel Dekker, 2004 .
57. A. Golbraikh, A. Tropsha, *J. Mol.Graph Mod. 2*002, *20*, 269.
58. A. Tropsha, P. Gramatica, V.J. Gombar, *QSAR Comb. Sci.*, 2003, *22*, 69.
59. P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR Comb.Sci.*2007, 26(5), 694-701.
60. http://www.oecd.org/document/23/0,2340,en_2649_201185_33957015_1_1_1_1,00.html.
61. http://europa.eu.int/comm/environment/chemicals/reach.htm.

62. OECD. (2007). Giudance document on the validulion of(quantiative)structure-activityrelationships: (Q)SARmodels. Retrieved fromwww.oecd.org/officialdocuments/publiedisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2.
63. Dearden, J. c., Cronin, M. T. D., & Kaiser, K.. L. E. (2009). How not to develop a quantitative structure-activityor structure-property relationship (QSAR/QSPR). SARand QSARin Environmental Research, 20(3-4), 241-266.doi: 10.10801106293609029495 67 PMID: 19544191.
64. Schon, u., Antel, 1., Bruckner, R., Messinger, J., Franke, R., &Gruska, A. (1998). Synthesis, pharmacologicalcharacterization, and quantitative structure-activity relationship analyses of 3,7,9,9tetraalkylbispidines:DerivaLives with specilic bradycardic activity. Journal ofMedicinal Chemistry, 41(3), 318-331. doi: 10.10211 jm970120q PMID:9464363.
65. Topliss, J. G., & Costello, R. 1. (1972) . Chance correlations in structure-activity studies using multiple regressionanalysis. JOUT/wi o/Medirillal Chemistry, 15(10),1066--1068. doi:10. 10211jm00280a017 PMlD:5069775
66. Topliss, J. G., & Edwards, R. P. (1979). Chance factors in studies ofquantitative structure-activity relationships .Journal ofMedicinal Chemistry, 22(10), 1238-1244. doi: I0.1 021 jm00196aOl7 PMID:513071.
67. Walker,1. D., Dearden, 1. c., Schultz, T. w., Jaworska, 1., & Comber, M. H. 1. (2003). QSARs for new practitioners.In J. D. Walker (Ed.), QSARsforpollutiom prevention, toxicity screening, risk assessment, and web applications(pp. 3-18). Pensacola, FL: SETAC Prcss.
68. Walker,1. D., Jaworska, 1., Comber, M. H. 1., Schultz, T. w., & Dearden, 1. C. (2003). Guidelines for developingand using quantitative structure-activity relationships. Environmental Txicologyamd Chemistry, 22(8), 1653-IM5. doi:10.1897/01-627 PMID:12924568
69. Livingstone, D. 1. (2004). Building QSAR models: a practical guide. In M. T. D. Cronin & D. J. Livingstone (Eds.) ,Predicting chemicaltoxicityamdfale (pp. 151-170). Boca Raton, FL: CRC Press. doi:10.120119780203642627.ch7
70. Dearden,1. c., & Rowe, P. H. (2015). Use ofartificial neural networks in the QSAR prediction ofphysicochemicalproperties and toxicities for REACH legislation. In H. Cartwright (Ed.), Artificial neural networks (pp. (i5-R8).New Yurk, NY: Springer. dui: 10.1007/978"14939-2239-0_5

71. Modarresi, H. , Mudarress, H., & Dearden, J. C. (2007). QSPR model of Henry's law constant for a diverse setofforganic chemicals based on genetic algorithm-radial basis function network approach. Chel1losphere, 66(11),2067-2076. doi 1016/j.chcmosphere.2006.09.049 PMID: 17113627.
72. Yaffe, D., Cohen, Y., Espinosa, G., Arenas, A., & Giralt, F. (2001). A fuzzy ARTMAP based on quantitativestructure-property relationships (QSPRs) for predicting aqueous solubility of organic compounds. Journal ofChemical Information and Computer Sciences, 41(5), 1177-1207. doi: 10.1021/ci010323u PMID: 11604019.
73. Livingstone, D. (1995) . Data analysisor chemicals. Oxford, UK: Oxford University Press.
74. Worth, A. P., & Cronin. M. T. D. (2003). The use ofdiscriminant analysis, logistic regression and classificationtree analysis in the development ofclassification models for human health effects. Journal ofMolecular StructureTHEOCHEM, 622(1-2), 97-111. doi: 1O.1016/S0166-1280(02)00622-X.
75. Barratt, M. D. (1995). The role ofstructure-activity relationships and expertsystems in alternative strategies forthe determination ofskin sensitisation, skin corrosivity and eye irritation. Altemativesto Laboratory Allimals,23,111-122.
76. Cronin, M. T. D., & Dearden, 1. C. (1997). Correspondence analysis ofthe skin sensitization potential oforganicchemicals. Quantitative Structure-Activity Relationships, /6(1), 33-37 . doi:10. 1002/qsar.19970 160 I06
77. Darnag, R., MostapbaMazouz, E. L., Schmitzer, A., Villemin, D., Jarid, A., & Cher4aoui, D. (1010). Suppurtvector machines: Development of QSAR models for predicting anti-lilV-I activities of TIEO derivatives.European Journal ofMedicinal Chemistry, 45(4), 1590-1597. doi: 10.1016/j.ejmech.2010.01.002 PMID:20110 136.
78. Schaper, K.-1., & Saxena, A. K. (1991). QSAR analysis of dose and property dependent activity classes ofhomologous drugs. In C. Silipo& A. Vittoria (Eds.), QSAR: rational approaches to the design ofbioactivecompounds (pp. 45-48). Amsterdam, The Netherlands: Elsevier.
79. Yuan, H., &Parrill, A. (2005). Cluster analysis and three-dimensional QSAR studies ofHIV-1 integrase inhibitors.Journal ofMolecular Gmphics&Modelling, 23(4), 317-328. doi: 10.101 6/j.jmgm.2004.10.003 PMID: 15670952
80. Hansch. c., & Leo, A. (1979). Substituent constantsfor correlation analysis in chemistry and biology. NcwYork, NY: John Wiley & Sons.

81. Mannbold, R., & Dross, K. (1996) . Calculation prucedures for molecular lipophilicity: A cumparative study.QuantitativeStructure--Activity Relationships, 15(5),403-409. doi: 10.1 002/qsar.19960150506.
82. Dearden, 1. C. (2003). Quantitative structure-property relarionships for prediction of boiling point, vaporpressure, and melting poinl. Environmental Toxicologyand Chemistry, 22(8), 1696-1709. dui: 10.1897/01-363PMID: 12924571
83. Dearden,1. C. (2006). III silico prediction ofaqueous solubility. £\"pert Opinioll 011 Drug Discovery, I( 1),31-52.dui: 10.1517/17460441.1.1.31 PMLD:23506031.
84. Dearden,1. c., Cronin, M. T. D., & Lappin, D. C. (2007). Acomparison of commercial available software fort.he prediction ofpKa. The Journal ofPharmacy and Pharmacology, 59(Suppl. I), A-7.
85. Liao, c., &Niddaus, M. C. (2009). Comparison of nine programs predicting pKa values of pharmaceuticalsubstances. Journal of Chemical Information and Modeling, 49( 12), 2801-2812 . doi:10. 10211ci900289xPMID: 19961204.
86. Dearden, J. C. (2010). Expert systems for toxicity prediction. In M. T. D. Cronin & 1. C. Madden (Eds.),III silico toxicology: principles al1d applications (pp. 478-507). London, UK: Royal Society of Chemistry.doi: I0.1 039/9781849732093-00478.
87. Dearden, 1. C. (2012). Prediction of physicochemical properties. In B. Reisfeld, & A.N. Mayeno A. (Eds.),Computational toxicology (pp. 93-138). New York, NY: Springer (Humana Press). doi: 10.1007/978-1-62703-050-2_6.
88. T. Puzyn, J. Leszczynski, M. T. D. Cronin, « Recent advances in QSAR studiesMethods and applications » *Springer Dordrecht Heidelberg London New York*, 2010
89. M. Tichý, *Science.Tot. Env*. 109, 1991, 407-410. Chapitre 2 : QSAR : Principe et Méthode 79.
90. B. Peric, J. Sierra, E. Martí, R. Cruañas, M. A. Garau, *Ecotoxico. Env. Saf*. 115, 2015, 257-262.
91. O. M. Bautista-Aguilera, G. Esteban, I. Bolea, K. Nikolic, D. Agbaba, I. Moraleda, I. Iriepa, A. Samadi, E. Soriano, M. Unzeta, J. Marco-Contelles, *Europ. J. Med. Chem*. 75, 2014, 82-95.
92. Paola Gramatica,* Stefano Cassani, and Nicola Chirico. QSARINS-Chem: Insubria Datasets and New QSAR/QSPRModels for Environmental Pollutant s in QSARINS. Journal of Computational Chemistry 2014, 35, 1036–1044.

# الجانب العملي

# الفصل الثاني

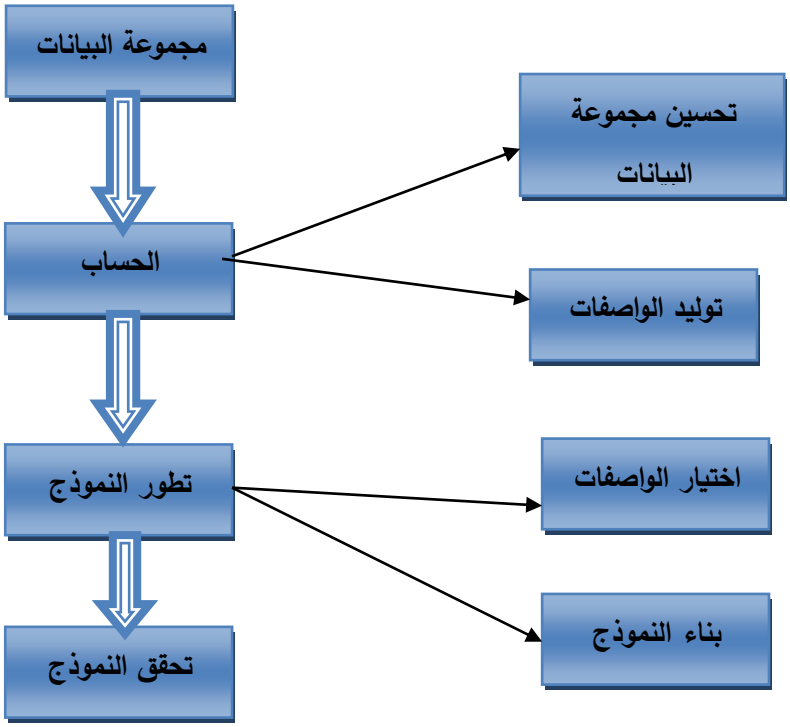
## طريقة حساب نماذج QSAR

## II- طريقة حساب نماذج QSAR :

للوصول لنموذج QSAR دقيق و موثوق هناك عدد من الشروط المتعلقة بـ الجوانب المختلفة التي تتوافق مع النموذج الذي يجب تحقيقه [2-1].

إن النشاط البيولوجي من سلسلة المركبات يجب أن تكون مرتبطة بالخصائص الفيزيائية والكيميائية و يجب أيضًا أن يتم استنباط أنشطة المواد الكيميائية الشاملة للآلية الواضحة و بنية كيميائية ذات صلة وهي ليست مطلوب بدقة ومع ذلك، غالبًا ما يكون من الصعب ، إن لم يكن من المستحيل ، تحديد آلية العمل ، في حين أنه عادة ما يكون إنشاء تشابه كيميائي أقل صعوبة. ومن ثم، فإن QSARs تم تطويره بشكل عام للسلسلة الجزيئية المتجانسة ، أي لمجموعات متشابهة كيميائيًا المركبات على أمل أن يكون لها أيضًا نفس آلية العمل. مع المركبات المحتملة أن تتناسب مع الارتباط بشكل سيئ فقط وأن تظهر على شكل القيم المتطرفة. لذلك، فإن الشروط الرئيسية المطلوبة لبناء نموذج QSAR صالح هي المجالات الرائدة المشاركة في تطوير مثل هذه النماذج [3] . كما موضح في الشكل 1.

والآن أصبح من السهل للغاية الوصول إلى مختلف قواعد البيانات من خلال موارد الإنترنت، ولكن بالنسبة لنمذجة البنية الكمية وعلاقات النشاط ، فمن المفيد للغاية أيضاً الحصول على مجموعات البيانات التي تتألف من ملفات البنية الكيميائية ، والنقاط النهائية ، والمعلومات الإضافية. فقدتم فحص بعض أكثر مجموعات البيانات المطلوبة من الملوثات البيئية في النسخة الجديدة من برامجنا الخاصة بتطوير نماذج منطقة QSAR الخطية المتعددة QSARINS. ويتم التعامل مع هذه الإضافة من خلال وحدة تسمى QSARINS-Chem ، والتي تشتمل أيضاً على نماذج QSAR/QSPR التي أعيد تطويرها حديثاً لبعض نقاط الاهتمام البيئية وغيرها من المجالات .




**الشكل ( II-1) : مخطط المراحل الرئيسية في بناء نموذج QSAR الحديثة .**

### II-1 التمثيل الهيكلي للمركبات وحساب الطاقة الفراغية Es:

تمثيل الهيكلي لجميع المركبات، الموجودة في المجموعة المراد دراستها، في شكل رسم ثنائي الأبعاد. هناك العديد من البرامج التي تمكننا من رسم الهياكل الجزيئية للمركبات العضوية، على سبيل المثال: ChemBioDraw ultra 16.0 أو Hyperchem وغيرها من البرامج [4] [5]. وحساب الطاقة الفراغية Es.

### II-1-1 الأساليب التجريبية:

مجموعة متنوعة من الطرق لحساب الطاقة الكامنة للأنظمة الجزيئات وفقًا لإحداثيات نواتها الذرية. هذه الطرق معالجة النوى على شكل جسيمات نيوتن تحت تأثير الوظيفة الطاقة الكامنة أو مجال القوة. جميع الطرق تستخدم البيانات التجريبية لتحديد ثوابت القوة الفردية (على سبيل المثال ، الثوابت من

الارتباط القابل للتوسيع) والتوازن لكل كمية هندسية (على سبيل المثال ،أطوال الارتباط). يوفر برنامج HyperChem هذه الطرق: الميكانيكا الجزيئية (MM+ و AMBER و BIO + و OPLS ..الخ) والديناميكي الجزيئية [6].

**II-1-2 طرق الكم:**

في ميكانيكا الكم نحن مهتمون بتوزيع الإلكترونات (المدارات) في الفضاء. تتضمن أفضل البرامج عمليات للتحسين الهندسة. الهدف الأساسي من ميكانيكا الكم هو تحديد الطاقة والتوزيع الإلكتروني. لذلك يتم حساب الطاقات الجزيئية باستخدام معادلة شرود نجر مع الشكلية المدارية الجزيئية (O.M). معادلة يمكن حل نظام شرود نجر الجزيئي عن طريق إدخال التقريبات. هذه الطرق هي تقنيات لحل معادلة شرود نغر تستخدم الأنظمة متعددة الإلكترونات البيانات المجهزة بالنتائج تجارب من أجل تبسيط العمليات الحسابية. تحتوي الطرق شبه التجريبية على القاعدة الدنيا بشكل افتراضي (STO-3G) من ناحية أخرى تستخدم طرق الكم الخالصة الأسس والارتباطات المختلفة تعتمد على نوع الحساب. برنامج HyperChem يوفر هذه الأساليب شبه التجريبية (CNDO، MINDO3 /INDO ،PM3. ، RM1.) والطرق الكمومية (Ab-initio وDFT)[6].

**II-2 توليد الواصفات:**

الواصفات هي قيم عددية تأكد الخصائص الفيزيائية والكيميائية للجزيئات من تمثيلاتها البنيوية. يعد إنشاء الواصفات خطوة حاسمة في تحليل QSAR وستعتمد جودة النموذج بشكل أساسي على طبيعة التوصيفات المتولدة. يمكن حساب معظم الواصفات ، التي تزيد حاليًا عن 4000 ، نظريًا من برامج مختلفة مثل أن: [7] flep ؛ [8] sepha ؛AutoQAR [9] ;DRAGON [10].

**II-3 حساب الواصفات الجزيئية:**

تم الحصول على المعلومات حول بنية الجزيئات، المحددة بواسطة الواصفات الجزيئية من تمثيلات مختلفة مثل 3D،2D ، وما إلى ذلك ، مضمنة في نموذج QSAR. يجب أن تكون المطابقات الجزيئية المستخدمة صحيحة من أجل نموذج تنبؤ يأفضل يوجد أنواع مختلفة من الواصفات.مثل واصفة العد (D0) ، بصمات الجزئية (D1) ، الطوبولوجي الواصفة (D2) والهندسية (D3) والشبكة القائمة (D4) ...الخ.

يزداد تعقيد المعلومات والقدرة على التمييز بين الهياكل المتشابهة على النحو الذي توفره الواصفات (dcscriptors) المختلفة مع زيادة الأبعاد.

- توفر الواصفتان D0 و D1 معلومات الأساسية مثل الوزن الجزيئي وعدد المكونات المشتقة مباشرة من الصيغة الجزيئية. تكون شحنة الجزيء هو واصفة D1.
- المؤشرات الطوبولوجية يتم حسابها من الصيغة البنائية هي واصفات ثنائية الأبعاد.

تستند هذه إلى الرسم البياني النظري وتعكس الروابط في الهيكل. الأكثر استخداما الطوبولوجي الواصف هو مؤشر الاتصال الذي اقترحه Randic،[11] [12] المؤشرات الطوبولوجية الأخرى هي مؤشر Wiener W [13] [14] ، مؤشرات الاتصال [15] وKier Shape [16]، مؤشر Balaban J ،[17] ومؤشرات Zagreb [18].

- تعتمد الواصفات ثلاثية الأبعاد على إحداثيات ثلاثية الأبعاد للذرات التي تتكون منها المركبات. بعضها شائع الاستخدام طرق حساب الواصفات ثلاثية الأبعاد هي CoMBINE, CoMSIA, CoMFA CoMMA,و HoloQSAR و CoSA.

- تعتبر الواصفات D4 و D5 وD6 متعددة الأبعاد، والتي تشمل العوامل المتضمنة في هيكل ومرونة موقع ربط المستقبلات بالتزامن مع طوبولوجيا. تعتمد واصفة D4 على الشبكات المرجعية والمحاكاة الديناميكية الجزيئية. تم حساب الواصفات باستخدام متعدد تشكل المطابقات والتوجهات وحالات البروتون والمتشابهات المتزامنة للرابط الواصفة D5. تشكل شروط الذوبان واصفة D6. يمكن أن تكون الواصفات الجزيئية محسوبة باستخدام برامج مختلفة. [19]

**II-4 تطوير نموذج:**

يبدأ تطوير النموذج بجمع البيانات تجارب موثوق وبأعداد كبيرة، فمن الضروري إنشاء ملف يكون فيه عدد من الواصفات، التي تميز الهياكل الجزيئية، المستخدمة في تطوير نموذج QSAR [20]. بمجرد بناء النموذج، يجب تقييمه بواسطة طرق التحقق من الصحة (عمليات التحقق الداخلية والخارجية) من أجل تقدير متانتها والقوة التنبؤية [23] [21]. أخيرًا، بالنسبة لأي نموذج، من المهم معرفة نوع جزيئات يمكن استخدام هذا النموذج (لتعرف على مجال التطبيق) [24]. هناك عدة أسباب عملية لاستخدام دراسات QSAR. [25]

- للتنبؤ بخصائص النشاط البيولوجي والفيزيائي الكيميائي عن طريق وسائل عقلانية.
- التوفير في تكلفة تطوير المنتج (مثال على الصناعة المستحضرات الصيدلانية والمبيدات الحشرية والمنظفات وما إلى ذلك).
- يمكن أن تقلل التوقعات من الحاجة إلى اختبارات طويلة ومكلفة.

**II-4-1 اختيار الواصفات:**

البحث عن جميع الواصفات التي تشكل النموذج الصحيح ومعبرة عن النشاط البيولوجي بتكلفة معقولة للحساب ، هو الخطوة الحاسمة ، لأن الحساب من بين جميع النماذج الممكنة ليس عمليًا نظرًا للعدد الكبير من الواصفات (أكثر من 4500 واصف).[26]

### II-4-2 طرق التصفية:

من بين عدد كبير من الواصفات المحسوبة، يتم اختيار عدد قليل فقط لتعريف النموذج. يؤدي تحديد الميزة بعد حساب الواصف إلى إزالة العلاقة الخطية المتداخلة بين أزواج الواصف. الهدف من التحليل الإحصائي على وجه التحديد هو "فصل" هذه الواصفات وتحديد تلك المرتبطة بالمتغير المستهدف، والتي تنتج الإشارة.

التحليل الإحصائي يجعل من الممكن تحديد الواصفات التي ترتبط ببعضها البعض للاحتفاظ بالواصفات الرئيسية وتقليل الواصفات الزائدة [27]. في المصطلحات الإحصائية، تسمى الخاصية المدروسة المتغير المستهدف أو المتغير التابع أو حتى المتغير Y.

والخاصية التي نبحث عن شرحها وتحسينها تسمى الواصفات، في هذه الحالة الكيميائية، المتغيرات التوضيحية والمتغيرات المستقلة أو المتغيرات X.

اختيار أكثر الميزات المناسبة تتم باستخدام طرق التصفية [28]. تتضمن الطرق تصفية الواصفات التالي تقلل حجم مجموعة الواصفات المعتمدة على الارتباطات بين المتغيرات. لذا، فإن الواصفات الجزيئية التي تظهر الارتباط المتبادل مع الاحتفاظ بواصف واحد فقط من الزوج [29]. تتم أيضًا إزالة الواصفات ذات التباين الأدنى. تستخدم طرق التصفية تقنيات مثل تحليل entropy Shannon,chi-square, odds ratio نسبة الأرجحية ،[30]. اختيار الميزات القائمة على أساس الارتباط [31] ،Fisher Score،احصائيات Kolmogorov–Smirnov [32] ، تحليل المكونات الأساسية. يتم أيضا تجميع الأساليب القائمة على المسافة مثل مقاييس المسافة الإقليدية ضمن طرق التصفية. تستخدم طرق الالتفاف طرقا قائمة على الانحدار لتحديد الواصفات. بشكل عام ، تتضمن طرق الالتفاف مزيدًا من القوة الحسابية و تؤدي بشكل افضل من طرق التصفية. [33] ، الاختيار متغير و النمذجة على أساس التنبؤ [34]، إزالة رجعية ، اختيار أمامي ، خوارزمية جينية ، الشبكة العصبية والمنتظمة

Bayesian وتحليل العوامل والبروتوكول التوافقي هو بعض من طرق الالتفاف شائعة الاستخدام . يتم أيضًا استخدام الطرق الهجينة التي تجمع بين طرق المرشح والالتفاف لاختيار الميزات [28]

**II-5 بناء نموذج QSAR:**

تعتمد جودة نموذج QSAR إلى حد كبير على البيانات المستخدمة فيه أعمال بناء. ومن ثم ، قبل تطوير النموذج ، من الضروري اكتساب رؤيا حول البيانات. يساعد الفهم الشامل للمشكلة والعوامل المؤثرة في التمييز علاقات هادفة. معلومات أساسية ذات صلة بالنظام قيد الدراسة يتم جمعها إما بيولوجية أو كيميائية عن طريق الأبحاث.

يجب اختيار مجموعات البيانات لبناء للنموذج بعناية لأن البيانات السيئة وغير المتسقة قد تؤدي إلى الفساد نموذج. هناك العديد من العوامل الأخرى مثل تقسيم البيانات إلى مجموعات بيانات التدريب والاختبار، الواصفات الجزيئية، الطرق الإحصائية لتطوير النماذج التي تؤثر على جودة نموذج QSAR [35].

**II-5-1 مجموعات بيانات التدريب والاختبار:**

يتم أولاً تقسيم البيانات التي تم جمعها بشكل عشوائي إلى جزأين:

- بيانات التعلم "مجموعة التدريب (TSET)": تستخدم لبناء النماذج النهائية
- بيانات التنبؤ "مجموعة الاختبار (PSET)": تستخدم للتحقق من صحة النماذج المدربة.
- من ناحية أخرى، لا تخضع هذه البيانات لأي معالجة أثناء إنشاء النماذج ويتم حجزها فقط لاختبار موثوقية النماذج [36]. يتم استخدام مجموعة التدريب لصياغة نموذج QSAR، بينما تستخدم مجموعة الاختبار لتقييم إمكانية التنبؤ بها ودقتها.

مجموعة البيانات هي بشكل عام مقسمة بطريقة تجعل كلتا المجموعتين تشغل مساحة الواصف بالكامل. تعمل تقنيات تقسيم البيانات المناسبة على تحسين التنبؤ بالنموذج. والمناهج المختلفة المتاحة لتقسيم مجموعة البيانات تشمل مجموعات k-means، بناءً على استجابة X، بناءً على الاستجابة Y، والاختيار العشوائي ، والتصميم الجزيئي الإحصائي، واستبعاد المجال ، واختيار Kennard وStone ، اختيار بطاقة Kohonen المنظمة ذاتيًا ، ومجموعة الاختبار الموجهة نحو الاستقراء الاختيار [37] .[38]

- التحديد العشوائي: يمكن تقسيم مجموعة البيانات من خلال عملية بسيطة لاختيار عشوائي لمجموعة المعايرة والاختبار (للتحقق الخارجي).
- بناءً على الاستجابة Y: يعتمد هذه الطريقة على اخذ عينات من النشاط (Yاستجابة). يتم تقسيم النطاق الكامل للاستجابة إلى صناديق والمركبات المنتمية لكل خزان يتم تعيينه للمعايرة أو مجموعات الاختبار بشكل عشوائي أو شخصي.
- بناءً على الإجابة X: تؤخذ الخصائص والتشابه البنيوي للجزيئات في الاعتبار لتجميع المركبات المتشابهة. ثم يتم جزء محدد مسبقًا من المركبات للمعايرة أو مجموعة الاختبار يدويًا أو على أساس منتظم.

## II-6 الأساليب المستخدمة في QSAR الحديثة (QSARINS):

### II-6-1النماذج الإحصائية:

يحدد التحليل الإحصائي الارتباطات بين الواصفات والمتغير الهدف. كما يشير أيضًا إلى المساهمة النسبية لكل واصف في الشرح العام للنشاط. النموذج الإحصائي هو معادلة تعطي القيمة

المتغير المستهدف كدالة لمجموع القيم المرجحة للواصفات.الأدوات الإحصائية الرئيسية للحصول على نموذج هي:

- الانحدار الخطي متعدد المتغيرات (MLR) ؛
- انحدار المكون الرئيسي (PCR) ؛
- انحدار المربعات الصغرى الجزئية (PLS) ؛
- الشبكة العصبية الاصطناعية (ANN)[39].
- آلات ناقلات الدعم(SVM)

**II-7 طرق تحليل البيانات :**

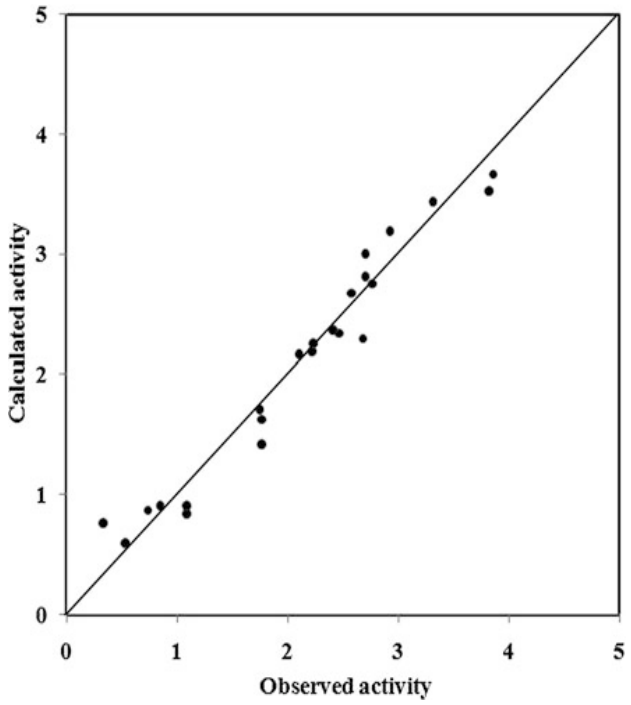
لتطوير نموذج QSARIN نحتاج إلى طريقة تحليل البيانات، تجعل هذه الطريقة من الممكن تحديد العلاقة الموجودة بين الكمية والنشاط والهيكل (الواصفات). هناك عدة طرق لبناء نموذج وتحليل البيانات الإحصائية لهذا الأخير، بعضها خطي مثل الانحدار الخطي متعدد (MLR) ، انحدار المربعات الصغرى الجزئية (PLS) ، والبعض الآخر غير خطي مثل أشجار القرار والشبكات العصبية وما إلى ذلك ، تتوفر هذه الأساليب في برامج مثل : Excel stat, Minitab, Statistica, Molegro Data Modeller, SPSS, R,....الخ [40].

**II-7-1الانحدار الخطي المتعدد:**

الانحدار الخطي المتعدد أو MLR هو أسلوب شائع الاستخدام في QSAR بسبب بساطته وشفافيته وقابليته للتكرار وسهولة تفسيره. [41]. هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة وللتنبؤ بقيم مستقبلية، وهو أيضا عن انحدار متغير التابع y على العديد من المتغيرات المستقلة x1. x2. xk 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 أو الانتشار [42].

نموذج MLR الذي يناسب البيانات المقدمة بشكل جيد سيؤدي إلى مخطط مبعثر (ملحوظة مقابل محسوبة) تظهر حدًا أدنى لانحراف النقاط عن الخط تناسب (الشكل 2.1). يتم تحديد جودة نموذج MRL من عدد من المقاييس كما هو موضح ادناه




**الشكل ( II−2): تبعثر للنشاط المرصود والمحسوب لنموذج MLR**

### II−7−1−1 نموذج الانحدار الخطي:

يتمثل هذا النموذج في معادلة رياضية التي يمكننا 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 :

$$\mathbf{Yj = b + b1 * X1j + b2 * X2j. + bk - 1 * Xkj + Uj(II.1)}$$

**j=1.2.....n**

حيث أن

Yj:المتغير التابع.

a:ثابت الانحدار الخطي .

b1:ميل الانحدار y على المتغير المستقل الأول.

b2: ميل الانحدار yعلى المتغير المستقل الثاني.

j1x: المتغير المستقل الأول.

2xj: المتغير المستقل الثاني. [42]

Uj: المتغير العشوائي.

في واقع الأمر فان هذه المعادلة واحدة من جملة معادلات يبلغ n ويكون نظام المعادلات كالأتي:

$$\begin{cases} \mathbf{y = a + b1X11 + b2X12 \ldots . . + bkXk1 + U1} \\ \mathbf{y = a + b1X21 + b2X22 \ldots . . + bkXk1 + U2} \\ \mathbf{y} = . + . + . \ldots . . + . + . \\ \mathbf{y = a + b1X1n + b2X2n \ldots . . + bkXkn + Un} \end{cases} \quad \mathbf{(II.2)}$$

هذه المعادلة تتضمن $k-1$ من المعلومات المطلوب تقديرها علما أن الحد الأول منها a يمثل الحد الثابت.

الأمر الذي يتطلب اللجوء إلى المصفوفات والمتجهات لتقدير تلك المعلومات.وعليه يمكن كتابة صياغة هذه المعادلات في صورة مصفوفات كالأتي:

$$\begin{bmatrix} \mathbf{y1} \\ \mathbf{y2} \\ . \\ . \\ \mathbf{yn} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{X11} & \mathbf{X21} \ldots\ldots\ldots\ldots & \mathbf{Xk} \\ \mathbf{1} & \mathbf{X21} & \mathbf{X22} \ldots\ldots . . & \mathbf{Xk} \\ & & . & \\ & & . & \\ \mathbf{1} & \mathbf{X1n} & \mathbf{X2n} \ldots\ldots\ldots & \mathbf{Xkn} \end{bmatrix} * \begin{bmatrix} \mathbf{b1} \\ \mathbf{b2} \\ . \\ . \\ \mathbf{bn} \end{bmatrix} + \begin{bmatrix} \mathbf{U1} \\ \mathbf{U2} \\ . \\ . \\ \mathbf{Un} \end{bmatrix} \mathbf{(II.3)}$$

وباختصار يمكن كتابتها كالأتي :

$$\mathbf{Y = XB + U(4.II)}$$

حيث أن :

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

X: مصفوفة ابعادها k*n+1 وتحتوي على المتغيرات المستقلة.

B: متجه عمودي أبعاده يحتوي على المعالم المطلوب تقديرها .

U: متجه عمودي أبعاده يحتوي على الأخطاء العشوائية. [42]

يمكن استخدام في حالة توافر الشروط التالية :

أن تكون العلاقة خطية بين المتغيرات المستقلة والمتغير التابع .

أن تكون البيانات موزعة توزيعا طبيعيا

يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل. [42]

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

$$\widehat{Yj} = \hat{a} + \hat{b}1 * X1j + \hat{b}2 * X2j + \hat{b}k - 1 * Xkj + Uj \qquad (II.5)$$
$$j=1\ 2\ .....n$$

حيث أن:

$\widehat{Yj}$ القيمة التقديرية لمتغير التابع .

$\hat{a}$ لثابت الانحدار الخطي .

b1: ميل الانحدار y على القيمة التقديرية المتغير المستقل الأول.

b2: ميل الانحدار yعلى القيمة التقديرية المتغير المستقل الثاني.

j1x: المتغير المستقل الأول.

2xj: المتغير المستقل الثاني.

Uj: المتغير العشوائي.

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

**II-8 المربعات الصغرى الجزئية (PLS):**

أثناء التعامل مع عدد كبير من الواصفات المتداخلة والصاخبة لفترة محدودة عدد نقاط البيانات PLS هو الخيار الأفضل على MLR، باعتباره تعميمًا لـ MLR [43]. يحاول استخراج المتغيرات الكامنة (LV) ، وهي وظائف المتغيرات الأصلية ، التي تمثل أكبر قدر من تباين العامل الأساسي ممكن أثناء النمذجة. قبل التحليل المتغيراتX و Y غالبًا ما يتم تحويل توزيعاتهم لجعل توزيعاتهم متماثلة إلى حد ما.

الإجابة عادة ما يتم تحويل المتغيرات لوغاريتميًا ويجب أن تكون المتغيرات X كذلك تحجيم مناسب. يجد PLS الخطي بعض المتغيرات الجديدة (المتغيرات الكامنة) ، وهي مجموعات خطية من المتغيرات الأصلية. عندما يكون عدد LVs هو يساوي عدد المتغيرات ، يصبح نموذج PLS هو نفسه نموذج MLR.

ومن الضروري إجراء اختبار دقيق للأهمية التنبؤية لكل مكون PLS ، و ثم إيقاف إضافة المكونات الجديدة عندما تبدأ المكونات في أن تكون غير مهمة. يعد التحقق المتقاطع (CV) طريقة عملية وموثوق لاختبار هذه الأهمية التنبؤية. يمكن التعبير عن معادلة PLS بنفس الشكل كما في MLR ؛ وبالتالي يمكن العثور بسهولة على مساهمات الواصفات الفردية للاستجابة. [44]

تم تطوير المفهوم الأساسي لانحدار PLS في الأصل بواسطة Wold. كشعبية و منهجية عملية PLS، يستخدم على نطاق واسع في مختلف المجالات [46] [45]. في مجال QSAR / QSPR PLS، هي تشتهر بتطبيقها على CoMFA و CoMSIA. في الآونة الأخيرة ، تطورت PLS بالاقتران مع طرق رياضية أخرى لإعطاء أداء أفضل في تحليلات QSPR / QSAR. هذه تطورت PLS إلى

- المربعات الصغرى الجزئية الجينية (G / PLS) ،
- المربعات الصغرى الجزئية لتحليل العوامل (FA–PLS)
- المربعات الصغرى الجزئية وتصحيح الإشارة المتعامدة **(OSC–PLS)**، تم تقديمها بإيجاز في الأقسام التالية.

**II–8–1 المربعات الصغرى الجزئية الوراثية (G / PLS):**

مشتق G / PLS من طريقتين لحساب QSAR تقريب الوظيفة الجينية (GFA) وPLS[47.48].

تستخدم خوارزمية G / PLS GFA لتحديد وظائف الأساس المناسبة لاستخدامها في ملف نموذج البيانات وانحدار PLS كأسلوب مناسب لوزن وظائف الأساس المساهمات النسبية في النموذج النهائي. وبالتالي فإن تطبيق G / PLS يسمح ببناء أكبر معادلات QSAR مع تجنب الإفراط في التركيب والقضاء على معظم المتغيرات. كما الانحدار الطريقة المستخدمة في تحليل المجال الجزيئي (MFA) ، أداة تحليل 3D–QSAR المعروفة ، G / PLS هي الأكثر استعمالا. [49].

**II–8–2 المربعات الصغرى الجزئية لتحليل العامل (FA–PLS):**

هذا هو مزيج من التحليل العملي (FA) و PLS ، حيث يتم استخدام FA للاختيار الأولي لـ واصفات ، وبعد ذلك يتم تنفيذ PLS. FA هي أداة لمعرفة العلاقات بين المتغيرات. يقلل المتغيرات إلى عدد قليل من العوامل الكامنة التي يتم من خلالها اختيار المتغيرات المهمة لـ PLS تراجع. في معظم الأحيان ، يتم استخدام طريقة عدم الخروج كأداة لاختيار الرقم الأمثل من مكونات PLS. يمكننا أن نجد أمثلة على FA–PLS المستخدمة في تحليل QSAR في [50–53].

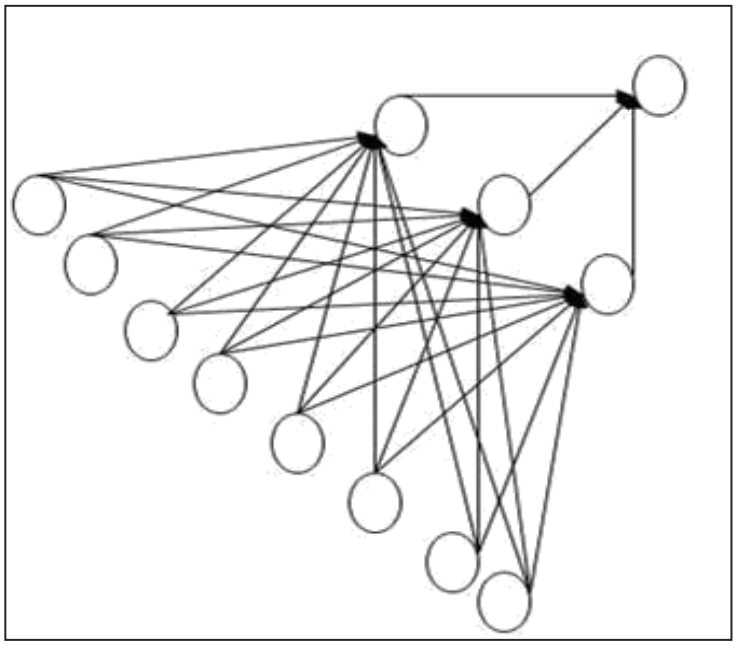
**II–8–3 المربعات الصغرى الجزئية لتصحيح الإشارة المتعامدة (OSC–PLS):**

تم تقديم تصحيح الإشارة المتعامد (OSC) بواسطة [54] WOld لإزالة منهجية الاختلاف من مصفوفة الاستجابة X غير ذات الصلة أو المتعامدة إلى مصفوفة الخاصية Y. لذلك ، يمكن للمرء أن

يتأكد من الاحتفاظ بالمعلومات المهمة المتعلقة بالتحليل. منذ ذلك الحين، تم نشر خوارزميات OSC المختلفة في محاولة لتقليل تعقيد النموذج عن طريق إزالة المكونات المتعامدة من الإشارة. في المجردة ، فإن المعالجة المسبقة مع OSC ستساعد التقليدية PLS للحصول على نموذج أكثر دقة ، كما ثبت في العديد من دراسات التحليل الطيفي [49]. حتى تاريخه، للأسف ، لا يوجد سوى عدد قليل من التقارير التي يتم فيها تطبيق OSC–PLS على دراسات QSPR / QSAR [57–55]، ولكن من المتوقع إجراء المزيد من أبحاث QSAR أو QSPR التي تتضمن تطبيق طريقة OSC–PLS المستقبل.

**II–9 الشبكة العصبية الاصطناعية ANN:**

في السنوات الأخيرة ، تم استخدام الشبكات العصبية الاصطناعية على نطاق واسع. يمكن استخدام الشبكات العصبية (ANN) لتوليد نماذج تنبؤية لعلاقات الكمية بين البنية و النشاط (QSAR) بين مجموعة من الواصفات الجزيئية من MLR والنشاط المرصود. و إنه تعميم النماذج الرياضية للنظام البيولوجي. العامل المهم في ANN هو القدرة على البناء نموذج المشكلة باستخدام البيانات من القياسات التجريبية لمجال المشكلة. تم تطبيق التقدم في مجال تصميم الأدوية ANN لحل مشكلة مختلف العملية الصيدلانية وتطوير المنتجات. يوضح (الشكل 5) الوارد أدناه التخطيطي تمثيل ANN. الشبكات العصبية هي أجهزة خرائط خالية من النماذج قادرة على ذلك التقاط العلاقات المعقدة غير الخطية في البيانات الأساسية التي غالبًا ما يفوتها التقليد يقترب QSAR. ومع ذلك ، من المعروف أن الشبكات العصبية غير مستقرة ، بمعنى أنها ثانوية التغييرات في بيانات التدريب و عوامل التدريب التي يمكن أن يكون لها عواقب وخيمة في قدرة التعميم للنماذج الناتجة. [58]



**الشكل ( II-3 ) : تمثيل تخطيطي لـ ANN**

## II-10 التحقق من صحة النموذج:

تهدف النمذجة إلى توفير نموذج ليس فقط ملائمًا للبيانات التجريبية، ولكن يمكن أيضًا تعميمها على أمثلة جديدة [59]. لهذا، عدة طرق للتحقق، مثل التحقق الداخلي، التحقق من الصحة اختبار خارجي واختبار العشوائية لتقدير موثوقية النموذج QSAR وتحديد مدى ملاءمتها لتطبيق معين. مؤشرات مختلفة الإحصائيات تستخدم أيضًا لتحديد جودة نموذج QSAR، يمكن التحقق من النماذج من خلال تقنيات مثل التحقق الداخلي والتحقق الخارجي والتحقق المتبادل [60]. في التحقق الداخلي، يتم توقع النشاط وتقدير العوامل لتحليل دقة التنبؤ بناءً على المركبات المستخدمة في بناء النموذج. هذا غير مناسب عند استخدام مجموعة اختبار جديدة من المركبات. لكن أسلوب التحقق الخارجي يعمل جيدًا حتى مع مجموعات البيانات الجديدة. في هذه الحالة، يتم تقسيم مجموعة البيانات إلى مجموعات بيانات الاختبار والتدريب. يتم التحقق من صحة النموذج من خلال مركبات مجموعة الاختبار المستقلة عن مجموعة التدريب [62-60]. ومع ذلك، فإن التحقق الخارجي لا يستحق لأنه يترك جزءًا كبيرًا من

مجموعة البيانات للاختبار [63]. اقترح الباحث Golbraikh في عام 2002 قيمة عالية لـ $R^2$ التحقق من صحتها لتكون واحدة من ذات قدرة تنبؤية عالية لنموذج QSAR [22].

**II-10-1 البيانات والاختبارات الإحصائية القياسية :**

غالبًا ما يتم تصور جودة النموذج على مخطط مبعثر، حيث يتم إدخال القيم المحسوبة للممتلكات (النشاط البيولوجي) ، اعتمادًا على تلك القيم تجريبي. جودة النمذجة تكون أفضل عند النقاط من هذا الرسم البياني قريبِة من خط. [60] تعديل النقاط لهذا الخط يمكن تقييمها بواسطة: فحص جودة الملاءمة بواسطة: معامل R التربيع $R^2$ للتحديد المتعدد ، والذي يمكن حسابه بسهولة بواسطة (المعادلة 1) و الخطأ المطلق (MAE) ؛ الانحراف المعياري (SD) ؛ الانحراف الجذر التربيعي (RMSD) عامل تضخم التباين (VIF) (عامل العلاقة الخطية المتداخلة) وعامل فيشر (F). تستخدم المؤشرات الإحصائية عادة في تحليل الانحدار المتعدد. [49]

**II-10-1-1 معامل الارتباط r:**

معامل الارتباط هو مقياس كمي نسبي يستخدم لتحديد نوع وقوة العلاقة بين متغيرين ويرمز له ب(r) [64] .

**II-10-1-2 معامل التحديد $R^2$:**

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

- **تعريف:**

هو معامل يستخدم لمعرفة مدى جودة التوفيق لخط الانحدار المقدر للمتغير التابع على المتغير المستقل، وبما انه عبارة عن مربع معامل الارتباط لذلك فقيمته تكون موجبة محصورة بين الصفر والواحد أي: $0 \leq R^2 \leq 1$ [65] .

هو يقيس نسبة التغير في المتغير التابع نتيجة تغير المتغير المستقل، أي انه يوضح نسبة مساهمة المتغير المستقل في التغير الحاصل في المتغير التابع. [65]

- يكون التغير جيدا كلما اقتربت قيمة $R^2$ من الواحد.
- إذا كان $R^2=1$ يعني انه يوجد علاقة خطية تامة بين المتغيرين التابع والمستقل أي أن المتغير المستقل يساهم بنسبة 100 في متغير 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. [66]

$$R^2 = 1 - \frac{\sum(Yi-\hat{Y}i)^2}{\sum(Yi-\bar{Y})^2} \quad (II.6)$$

أو :

y: القيمة التجريبية للنشاطyi هو المتغير التابع الملاحظ (الاستجابة التجريبية) ،

n: هو رقم المتغير التابع.

$\hat{Y}i$: القيمة المحسوبة للنشاط.

$\bar{Y}$ : متوسط قيمة القيم المحسوبة للنشاط. هي القيمة المتوسطة للمتغير التابع.

كلما كانت قيمة $R^2$ أقرب إلى 1 ، زادت ترابط القيم المحسوبة والملاحظة.

### II-1-1-3 متوسطالخطأ المطلق (MAE لمتوسطالخطأ المطلق):

لمتوسط الخطأ المطلق هو مؤشر مستخدم آخر محدد بالعلاقة :

$$MAE = \frac{\sum|\widehat{Yi} - yi|}{p} \quad (7.II)$$

**II-10-1-4** الخطأ القياسي في التنبؤS:

مؤشر إحصائي آخر يستخدم هو SD الانحراف المعياري كما هو موضح في هذه المعادلة **(II.8)**

$$S = \sqrt{\frac{(\hat{Y}_1 - Y)^2}{n-p-1}} \qquad \textbf{(II.8)}$$

حيث n هو عدد بيانات التعلم و p هو عدد الواصفات المستخدمة فيها بناء النموذج.[40]

**II-10-1-5 عامل التضخم الخاص بالتباينVIF:**

**يتم** فحص العلاقة الخطية المتعددة من خلال طريقة تسمى عامل التضخم الخاص بالتباين (VIF: عامل تضخم التباين) [67] المعبر عنه بالمعادلة

$$VIF_{xk} = \frac{1}{1-r_k^2} \qquad \textbf{(II.9)}$$

مع VIF (xk) هي قيمة عامل تضخم التباين للواصف xk (K = 1 ؛ 2 ؛ ؛ 5) ،

و$r_k^2$ هو معامل الارتباط التربيعي الناتج عن انحدار الواصف xk على باقي الواصفات.

تشير قيمة VIF $\geq$ 10 إلى وجود علاقة خطية متداخلة قوية بين الواصفات ، ويشار إلى عدم وجود علاقة خطية متعددة بقيمة VIF <10.

ويمكن أيضًا تأكيد وجود أو عدم وجود علاقة خطية متعددة قوية بواسطة كمية أخرى ، على غرار VIF ، يسمى عامل التسامح TF المعطى في المعادلة 3.3:

$$TF = \frac{1}{VIF_{xk}} \qquad (10.II)$$

**II-10-1-6الاختبار الاحصائية F:**

يستخدم مؤشر فيشر F أيضًا لقياس مستوى الأهمية الإحصاء النموذجي ، أي جودة اختيار الواصفات التي تشكل نموذج.

- **تعريف**

لقياس معنوية النموذج يتم استخدام اختبار F لاختبار معنوية النموذج وذالك بتحديد ما اذا كانت هناك علاقة معنوية المتغير المعتمد Y والمتغيرات المفسرة XK.... X2 X1 حيث تستخرج قيمة F كالاتي: [68]

$$\mathbf{F = \frac{\sum(\hat{Y}_I - \overline{Y})^2}{\sum(Y - \hat{Y}_I)^2} \frac{p - n - 1}{p}} \qquad (11.II)$$

y: القيمة التجريبية للنشاط

$\hat{Y}_I$: القيمة المحسوبة للنشاط

$\overline{Y}$ : متوسط قيمة القيم المحسوبة للنشاط.

n:العدد الاجمالي .

p:عدد طبيعي متغير.

**II-10-1-7 اختبار** le test-t de Student :

يتم تقييم أهمية الواصفات في النموذج من خلال اختبار ti للطالب عن طريق الاختبار الفرضية الصفرية التي تعتبر أن الواصف غير مهم. من أجل الانحدار المضاعف الخطي ، يُفترض أن المعامل ai المرتبط به هو صفر.

يتم رفض هذه الفرضية (بفاصل ثقة$\alpha$) إذا كانت النسبة ti بين ai و ti يصل الخطأ القياسي S (ai) إلى قيمة ترتيب الكسر $(1 - \alpha/2)$ لتوزيع الطالب عند $(p-n-2)$ درجات الحرية. [36]

$$\mathbf{|ti| = \left|\frac{a}{s(ai)}\right| > t_{1-\alpha/2}^{p-n-2}} \qquad (12.II)$$

## II-11 مقاييس التحقق من الصحة من نوعين بناءً على نوع نموذجQSAR:

1) نماذج QSAR القائمة على الانحدار.

2) نماذج QSAR القائمة على التصنيف.

كل من الأساليب القائمة على الانحدار والتصنيف لها مقاييسها الفريدة للتحقق من صحتها.

### II-11-1 مقاييس التحقق من الصحة للطرق القائمة على الانحدار:

يتم حساب مقاييس التحقق من الصحة للنماذج القائمة على الانحدار لكل من الداخلية والخارجية استراتيجيات التحقق من الصحة.

#### II-11-1-1 مقاييس التحقق من الصحة للتحقق الداخلي:

يستخدم التحقق الداخلي لنماذج QSAR استخدام الجزيئات من مجموعة التدريب للاختبار إمكانية التنبؤ بالنموذج.

بعض الطرق الأكثر شيوعًا المستخدمة للداخلية تم وصف التحقق من صحة نماذج QSAR هنا.

##### II-11-1-1-1 Least Square Fitting:

يتشابه التركيب المربّع الأصغر مع الانحدار الخطي وهو أكثر عمليات التحقق شيوعًا طريقة. إنه مقياس معامل الارتباط المربع ($R^2$) بين المتوقع و القيمة التجريبية للنشاط. يمكن إزالة القيم المتطرفة من مجموعة بيانات التدريب ، من أجل تحسين نموذج QSAR ، إذا كان الفرق بين $R^2$ و $R^2adj$ أقل من 0.3. [60]

$$\mathbf{R^2ext = 1 - \frac{\sum_{i=1}^{next}(yi-\hat{y})^2}{\sum_{i=1}^{next}(yi-\bar{y})^2}} \quad \textbf{(13.II)}$$

$$\boldsymbol{R_a^2 = \frac{(N-1)*R^2-p}{N-1-P}} \quad \textbf{(14.II)}$$

n, p عدد المتغيرات المتوقع في تطوير النموذج .

### II-11-1-1-2 الخطأ التربيعي لمتوسط الجذر (RMSE) أو (χ2):

تُستخدم قيم χ2 و RMSE لتقييم الجودة التنبؤية لنموذج ما. χ2 عروض القيمة الفرق بين قيم النشاط الحيوي المحددة تجريبياً والقيم المتوقعة حسب النموذج ، في حين أن قيمة RMSE هي تصوير الخطأ بين متوسط قيم النشاط التجريبية والمتوقعة. حتى بالنسبة للنماذج ذات قيمة $R^2$ الكبيرة (أي> = 7) ، يجب أن تكون قيم χ2 و RMSE أقل من 0.5 و 0.3 على التوالي ، من أجل التنبؤ الجيد قدرة النموذج [60].

$$\mathbf{RMSEp} = \sqrt{\frac{\sum(\mathbf{yobs(test)} - \mathbf{ypred(test)})}{\mathbf{n\ ext}}} \quad \textbf{(II.15)}$$

### II-11-1-1-3 عبر المصادقة(LGO): Cross validation :

تتضمن مناهج التحقق المتقاطعة للتحقق الداخلي من مغادرة المجموعةLeave-Group-Out(LGO) ، والتي يتضمن ترك جزيء أو مجموعة من الجزيئات أثناء إنشاء النموذج والتقييم إمكانية التنبؤ بالنموذج باستخدام الجزيئات المتبقية. بعض التدابير الهامة المستخدمة في التحقق المتقاطع الداخلي لنماذج QSAR مذكورة أدناه.

### II-11-1-1-3-1 التحقق من صحة الإجازة مرة واحدة (LOO) والإجازة عن بعض (LSO) :

**:Leave-One-Out (LOO) and Leave-Some-Out (LSO) cross validation**

في التحقق من صحة LOO المتقاطع ، يتم ترك مركب واحد ويتم إنشاء نموذج QSAR باستخدام المركبات المتبقية. يتم استخدام المركب المستبعد كاختبار للنموذج المتوقع. تتكرر هذه العملية لإزالة كل من المركبات في مجموعة البيانات واحدًا تلو الآخر. و يتم استخدام النتائج التي تم الحصول عليها من هذا لتقدير المعلمات المشاركة في التحقق من الصحة المقاييس. يتم تقييم إمكانية التنبؤ بالنموذج بواسطة PRESS (المجموع المتوقع المتبقي لـ المربعات) والتحقق من صحتها ($Q^2$) عندما SDEP (الانحراف المعياري لخطأ التنبؤ) تم الحصول عليها منPRESS[62] [61]

$$\mathbf{PRESS} = \sum(\mathbf{Yobs} - \mathbf{Ypred})^{\mathbf{2}} \qquad (\mathbf{II.16})$$

$$\mathbf{SDEP} = \sqrt{\frac{\mathbf{PRESS}}{\mathbf{n}}} \qquad (\mathbf{II.17})$$

$$\mathbf{Q^2} = \mathbf{1} - \frac{\sum(\mathbf{Yobs(train)} - \mathbf{Ypred(train)})^{\mathbf{2}}}{\sum(\mathbf{Yobs(train)} - \overline{\mathbf{Ytraıning}})^{\mathbf{2}}} = \mathbf{1} - \frac{\mathbf{PRESS}}{\sum(\mathbf{Yobs(train)} - \mathbf{Ytraining})^{\mathbf{2}}} \qquad (\mathbf{II.18})$$

في المعادلتين (16) – (17) تتوافق Yobs/Ypred مع قيم الخاصية الرصودة التي تتنبا بها Loo وn عدد الملاحظات Yobs الخاصية الرصوده Ypred raintالخاصية المتوقعة للجزيئات تدريبية تعتمد على تقنية loo القيمة الحدية ل $Q^2$هي 0.5.

**II-11-1-1-3-2 التحقق من صحة الإجازة (LSO):**

في حالة LSO (اترك بعضًا خارجًا) أو LMO (اترك كثير خارج) ، فإن مجموعة من مركبات البيانات هي يتم التخلص منها ويتم إنشاء النماذج مع بقية المركبات. المركبات المتروكة هي تستخدم للتحقق من إمكانية التنبؤ بالنموذج. على غرار طريقةLOO ، وطريقة LMO يتضمن أيضًا دورات متكررة من الإزالة وإنشاء النموذج كل مرة تم التعامل مع المركب كمجموعة اختبار[60] [61] [66]. عند الانتهاء من جميع دورات تدريب النموذج والاختبار .

القيمة الإجمالية Q² LMO يتم حسابها بناءً على قيم النشاط المتوقعة للمركبات.تعد طريقة LMO أكثر اتساقًا مقارنةً بطريقة LOO [60].

عادة ما تكون قيمة Q² أصغر من قيمة R². من أجل تجنب الإفراط في تركيب في النموذج ، يجب ألا يتجاوز الفرق بين R² وQ² (0.3). قد يكون النموذج المجهز أكثر من اللازم متوقع جيدًا لنشاط المركبات ولمجموعة التدريب ولكن التنبؤ بالمركبات الجديدة هو أفضل . [60]

### II-11-1-1-4 مقاييس $Q^2$ و $r^2m$:

True $Q^2$ الذي اقترحه Hawkins وآخرون ، يستخدم لمجموعات البيانات الصغيرة ويتم حساب مقياس $r^2m$ بناءً على القيم المقاسة للنشاط المرصود والمتوقع. لا ينبغي التعامل مع $Q^2$ كدليل نهائي على القدرة وعلى التنبؤ الجيد بالنماذج. لا ينبغي تفسير قيمة $Q^2$ الأعلى من 0.5 على أنها قدرة تنبؤية عالية لنماذج QSAR ؛ حتى يتم اختبار قدرة النموذج على التنبؤ بنشاط عدد كبير من المركبات التي لا تستخدم لتدريب النموذج. [22]

$$\overline{r_m^2} = \frac{\left(r_m^2 - r_m'^2\right)}{2} \qquad \textbf{(II.19)}$$

$$(\Delta r_m^2 = \left|r_m^2 - r_m'^2\right| \qquad \textbf{(II.20)}$$

$$r_m'^2 = r^2 * \left(1 - \sqrt{(r^2 - r'^20)}\right) \text{و} r_m^2 = r^2 * \left(1 - \sqrt{(r^2 - r^20)}\right) \qquad \textbf{(II.21)}$$

بعض المقاييس الأخرى المستخدمة للتحقق الداخلي هي $r^2m$ الحقيقي (LOO) و Y- Randomisation ، وهو مقياس لارتباط الصدفة. يكشف مقياس ($r^2m$ LOO) الحقيقي عن خصائص التحقق الخارجية حيث يتم اشتقاق قيمته من النموذج الذي تم تطويره بعد الدورات المتكررة لـ LOO. يتضمن اختبار التوزيع العشوائي Y عملية عشوائية ونموذج عشوائي للتحقق من صحة النموذج عن طريق تبديل قيم الاستجابة فيما يتعلق بالمصفوفة غير المعدلة. [61]

### II-11-1-2 مقاييس التحقق من الصحة الخارجية:

### II-11-2-1 $Q_{F1}^2$ معامل الارتباط التربيعي في مجموعة التدريب:

معامل الارتباط التربيعي في مجموعة التدريب $Q_{F1}^2$ الصيغة المحسوبة في المعادلة (II.22) [69].

R² التنبؤية التي يمكن أن تُعطى أيضًا $Q^2_{F1}$ حول الارتباط المرصود والبيانات المتوقعة. يُقال أن النموذج يتمتع بقدرة تنبؤية جيدة إذا كانت قيمة $Q^2_{F1}$ أكبر من 0.5 [66].

$$\mathbf{Q^2_{F1} = 1 - \frac{\sum_{i=1}^{next}(yi-\hat{y}i)^2}{\sum_{i=1}^{next}(yi-\overline{ytr})^2}(} \qquad \mathbf{(II.22)}$$

مع yi القيمة التجريبية للخاصية، i القيمة المتوقعة / المحسوبة للخاصية وTR متوسط قيم yi لمجموعة التدريب.

**II-11-2-2-1 معامل الارتباط التربيعي التنبئي في مجموعة الاختبار $Q^2_{F2}$ :**

معامل الارتباط ألتربيعي التنبؤي في مجموعة الاختبار $Q^2_{F2}$الصيغة المحسوبة في المعادلة (II.23)$Q^2_{F2}$[70] [69]. و$Q^2_{F3}$ باستخدام متوسط مجموعة بيانات الاختبار ومجموعة بيانات التدريب على التوالي [71].

للتحقق من نموذج QSAR ، القيمة الحدية 0.5 هي محددة لكلا المقياسين [66].

$$Q^2_{F2} = 1 - \left[\frac{\sum_{i=1}^{next}(yi-\hat{y}i)^2}{\sum_{i=1}^{next}(yi-\overline{y_{ext}})^2}\right]( \text{II.23})$$

**II-11-2-3 $Q^2_{F3}$ القدرة التنبؤية الخارجية:**

القدرة التنبؤية الخارجية في الصيغة الحسابية في معادلة (II.24) [73، 72] ، يجب أن تكون أعلى أو تساوي 0.6 لنموذج QSAR مقبول [74].

$$Q^2_{F3} = 1 - \left[\frac{\sum_{i=1}^{next}(yi-\hat{y}i)^2/n_{ext}}{\sum_{i=1}^{next}(yi-\overline{y_{ext}})^2/n_{tr}}\right].( \text{II.24})$$

**II-11-2-1-4 معايير Golbraikh و Tropsha:**

تضع معايير Golbraikhو Tropsha شرطا لاختيار التدريب ومجموعات بيانات الاختبار. للحصول على قوة تنبؤية جيدة، يجب أن يرضي نموذج QSAR الشروط التالية .[60][22] ;لذلك يجب التحقق

من العوامل الأخرى لهذا الغرض. هذه العوامل المعروفة باسم "معايير التحقق الخارجية" أو يشار إليها غالبًا باسم "معايير Trophsa".

- **معايير التحقق الخارجية (سلسلة الاختبار):**

$r^2_{test} > 0.6$ (المعيار 1) **(II.24)**

$Q^2_{training} > 0.5$ (المعيار 2) **(II.25)**

$\frac{r^2 - r^2_0}{r^2} < 0.1$ و $0.85 \leq K \leq 1.15$ (II.27) (II.28)

**(المعيار 4)** $\frac{r^2 - r^2_0}{r^2} < 0.1$ و $0.85 \leq k' \leq 1.15$ (II.29)(II.30)

**(المعيار 5)** $r^2 - r^2_0 \leq 0.3$ (II.31)

مع:

$r^2$: معامل الارتباط لجزيئات سلسلة الاختبار ؛

$r^2_0$: معامل الارتباط بين القيم المتوقعة والتجريبية لسلسلة الاختبار ؛

$r'^2_0$:معامل الارتباط بين القيم التجريبية والقيم المتوقعة لسلسلة الاختبار ؛

K: هو ثابت خط الارتباط (في الأصل) (القيم المتوقعة كدالة لـالقيم التجريبية) ؛

k': هو ثابت خط الارتباط (في الأصل) (القيم التجريبية كدالة القيم المتوقعة).

### II-11-1-2-5 معامل الارتباط التوافق ccc:

تشمل المقاييس الأخرى RMSEP (جذر متوسط الخطأ التربيعي للتنبؤ) إلى حساب خطأ التنبؤ لنموذج QSAR [66] . معامل الارتباط ، وهو الإجراء الأكثر تقييدًا واحترازًا ، مع الوضع المثالي قيمة 1[73] $r^2m$ (الترتيب) الذي يجعل توقعات الترتيب [66] ؛ و $r^2m$ (اختبار) لفهم العلاقة بين الملاحظة و القيم المتوقعة [66] [74].

معامل ارتباط التوافق (CCC) الذي يشبه معامل الارتباط ، ولكنه ، بالإضافة إلى ذلك ، يأخذ في الاعتبار القرب من القطر ؛ يجب أن يكون أعلى عند 0.85 [74–75] ؛ يتم حسابه وفقًا لـ المعادلة (32).

$$ccc = 2\frac{\sum_{n=1}^{next}(yi-\bar{y})(\hat{y}i-\bar{\hat{y}})}{\sum_{n=1}^{next}(yi-\overline{y_{ext}})^2+\sum_{n=1}^{next}(\hat{y}i-\bar{\hat{y}})^2+n_{ext}(\overline{y_{ext}}-\bar{\hat{y}})^2} \text{(32.II)}$$

**II–11–2 مقاييس التحقق من الصحة للطرق القائمة على التصنيف:**

مصفوفة التحقق المستخدمة في الأساليب القائمة على التصنيف هي

**II–11–2–1 Wilks lambda (λ) الإحصاء:**

يتم استخدامه لاختبار أهمية وظيفة النموذج المميز ويتم حسابه على أنه نسبة مجموع المربعات داخل الفئة إلى التشتت الكلي. تتراوح القيمة بين $0 < \lambda < 1$ ، بقيمة أقل تقابل مستوى أعلى من التمييز. [44]

$$\text{Wilks } \lambda = \frac{\text{مجموع المربعات ضمن مجموعة}}{\text{المجموع الكلي للمربعات}} \text{(II. 33)}$$

تمتد قيمة Wilks lambda من 0 إلى 1 ، حيث يتوافق 0 مع المستوى الجيد لـ التمييز و 1 يشير إلى عدم التمييز

**II–11–2–2 مؤشر canonical index (Rc) :**

علاوة على ذلك ، مؤشر canonical index (Rc) يستخدم لتقدير قوة العلاقة بين مختلف المعاملين والمتغيرات المستقلة. [44]

$$Rc = \sqrt{\frac{\lambda i}{1+\lambda i}} \text{(II. 33)}$$

هنا ، يشار إلى λi على أنها القيمة الأصلية للمصفوفة.

### II-11-2-3 (χ2) Chi-square للتحقق من جودة النموذج القائم على التصنيف :

ومسافة Mahalanobis ماهالانوبيس التربيعية مقياس محسوب باستخدام نقاط بيانات عشوائية . [66] [75]

$$\chi^2 = \sum_{i=1}^{t} \frac{(fi - Fi)^2}{Fi} \quad (II.34)$$

حيث يتم ملاحظة استجابة fi ، يتم توقع استجابة Fi ، و t هو عدد الملاحظات

### II-12 تفسير وتحليل المجال التطبيق:

يجب أن تكون العوامل أو الواصفات المستخدمة في النموذج قابلة للتفسير. يساعد تفسير نموذج QSAR المبني على فهم تأثير الواصفات في النشاط المتوقع. يساعد تحليل مجال التطبيق على فهم ما إذا كان يمكن استخدام نموذج QSAR المدمج لأي مجموعة من المركبات. تم بناء نموذج مجال التطبيق على المنطقة النظرية الموجودة في الفضاء الكيميائي للواصفات و نمذجة النشاط. تمكن من فهم جدوى النشاط أو التنبؤ بالاستجابة من خلال نموذج QSAR المركب لمجموعة معينة من المركبات. لذلك, فإن التنبؤ بنموذج QSAR لمجموعة من المركبات يمكن الاعتماد عليه فقط إذا كان الفضاء الكيميائي أو مجال قابلية تطبيق هذا المركب يقع ضمن مجال قابلية تطبيق المركبات المستخدمة لتدريب النموذج [66]. يتم تحديد المنطقة النظرية في الفضاء الكيميائي أو تقديرها بواسطة طرق مختلفة. يتم إجراء تقييم مجال التطبيق من خلال توزيع كثافة الاحتمال ، والطرق الهندسية ، والطرق القائمة على المسافة ، والنطاقات في مساحة الواصف التي يتم استخدامها ، ونطاق متغير الاستجابة عندما تكون مساحة الاستجابة النموذجية مستخدمة. [76]

**المراجع الأجنبية**


1. Wold, S.; Dunn, W.J. Multivariate quantitative structure-activity relationships (QSAR):conditions for their applicability. J. Chem. Inf. Comput. Sci., 23, 1983, 6-13.
2. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold S. Multi- and Megavariate Data Analysis. Principles and Applications. Umetrics: Umea, 2001
3. Girona, Doctoral Thesis for the obtaining of the degree of, Doctor in Theoretical and Computational Chemistry, MOLECULAR QUANTUM SIMILARITY IN QSAR: APPLICATIONS INCOMPUTER-AIDED MOLECULAR DESIGN. June 2004
4. M Froimowitz. Hyperchem : a software package for computational chemistry and molecular modeling. *Biotechniques*, 14(6) :1010–1013, 1993.
5. Robert E Buntrock. Chemofficeultra 7.0. *Journal of chemical information and computer sciences*, 42(6) :1505–1506, 2002.
6. T. Salah ; thèse master en chimie pharmaceutique intitulé :développement des modèles QSAR pour la prédiction des activités inhibitrices antitrypanosomiennes des dérivés de cryptolepine; université Med Khider Biskra; 2013.
7. Massimo Baroni, Gabriele Cruciani, Simone Sciabola, Francesca Perruccio, and Jonathan S Mason. A common reference framework for analyzing/comparing proteins andligands. fingerprints for ligands and proteins (flap) : theory and application. *Journal ofchemical information and modeling*, 47(2) :279–294, 2007.
8. Steven L Dixon, Alexander M Smondyrev, Eric H Knoll, Shashidhar N Rao, David EShaw, and Richard A Friesner. Phase : a new engine for pharmacophore perception, 3dqsar model development, and 3d database screening : 1. methodology and preliminaryresults. *Journal of computer-aided molecular design*, 20(10-11) :647–671, 2006.
9. Steven L Dixon, Jianxin Duan, Ethan Smith, Christopher D Von Bargen, Woody Sherman, and Matthew P Repasky. Autoqsar : an automated machine learning tool for bestpractice quantitative structure–activity relationship modeling. *Future medicinal chemistry*, 8(15) :1825–1839, 2016.
10. Andrea Mauri, Viviana Consonni, Manuela Pavan, and Roberto Todeschini. Dragon software : An easy approach to molecular descriptor calculations. 56 :237–248, 01 2006.
11. Randić, M. (2001). The connectivity index 25 years after. Journal of Molecular Graphics and Modelling, 20, 19-35.

12. Li, X. & Shi, Y. (2008). A survey on the Randic index. MATCH Commun. Math. Comput. Chem, 59, 127-156.
13. Wiener, H. (1947). Structural determination of paraffin boiling points. Journal of the American Chemical Society, 69, 17-20.
14. Ivanciuc, O. (2000). QSAR comparative study of Wiener descriptors for weighted molecular graphs. Journal of chemical information and computer sciences, 40, 1412-1422.
15. Kier, L. B., Hall, L. H., Murray, W. J. & Randi, M. (1975). Molecular connectivity I: Relationship to nonspecific local anesthesia. Journal of pharmaceutical sciences, 64, 1971-1974.
16. Kier, L. B. (1985). A shape index from molecular graphs. Molecular Informatics, 4, 109-116.
17. Balaban, A. T. (1982). Highly discriminating distance-based topological index. Chemical Physics Letters, 89, 399-404.
18. Gutman, I. &Trinajstić, N. (1972). Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons. Chemical Physics Letters, 17, 535-538.
19. G Damale, M., N Harke, S., A Kalam Khan, F., B Shinde, D. & N Sangshetti, J. (2014). Recent advances in multidimensional QSAR (4D-6D): a critical review. Mini reviews in medicinal chemistry, 14, 35-55.
20. M. Zhao, Z. Li, L. Peng, Y. R. Tang, C. Wang, Z. Zhang, S. Peng, *Bioorg. Med. Chem.*15, 2007, 2815–2826
21. A. Toropov, A. P. Toropova, E. Benfenati, G. Gini, D. Leszczynska, J.Leszczynski,G. De Nucci, *Biochem. Biophys. Res. Comm*. 432(2), 2013, 214–225.
22. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 20, 2002, 269-276.
23. A. Tropsha, P. Gramatica, K.V. Gombar, *QSAR Comb. Sci.* 22, 2003, 69-77.
24. P. Gramatica, E. Giani, E. Papa, *J. Mol. Graph. Mod*. 25, 2007, 755-766.
25. T. I. Oprea, "Chemoinformatics in Drug Discovery" *Ed. WILEY-VCH Verlag. Allemagne*, 2005.
26. *Mr. Mohamed AISSAOUI*,Étude théorique pour l'identification des molécules biologiquement actives (approche QSAR, DOCKING, et VIRTUEL SCREENING), 10 juillet 2018
27. A. Fortuné, Thèse de doctorat, Techniques de Modélisation Moléculaire Appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance, Université Joseph Fourier – Grenoble I, 2006

28. *Good*arzi, M., Dejaegher, B. &Heyden, Y. V. (2012). Feature selection methods in QSAR studies. Journal of AOAC International, 95, 636-651.
29. Roy, K., Kar, S. & Das, R. N. (2015a). A primer on QSAR/QSPR modeling: Fundamental concepts, Springer.
30. Liu, Y. (2004). A comparative study on feature selection methods for drug discovery. Journal of chemical information and computer sciences, 44, 1823-1828
31. Demel, M. A., Janecek, A. G., Gansterer, W. N. & Ecker, G. F. (2009). Comparison of Contemporary Feature Selection Algorithms: Application to the Classification of ABC-Transporter Substrates. Molecular Informatics, 28, 1087-1091.
32. Guyon, I., Weston, J., Barnhill, S. &Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine learning, 46, 389-422.
33. Xue, Y., Li, Z.-R., Yap, C. W., Sun, L. Z., Chen, X. & Chen, Y. Z. (2004). Effect of molecular descriptor feature selection in support vector machine classification of pharmacokinetic and toxicological properties of chemical agents. Journal of chemical information and computer sciences, 44, 1630-1638.
34. Liu, S.-S., Liu, H.-L., Yin, C.-S. & Wang, L.-S. (2003). VSMP: a novel variable selection and modeling method based on the prediction. Journal of chemical information and computer sciences, 43, 964-969.
35. SwathikClaranciaPeter, Jaspreet Kaur Dhanjal, Vidhi Malik, Navaneethan Radhakrishnan2, MannuJayakanthan and Durai Sundar ,Quantitative structure-activity relationship (QSAR): modeling approachesto biological applications, January 2018,DOI: 10.1016/B978-0-12-809633-8.20197-0.
36. M. SAIHI Youcef,ThèseDOCTORAT *,Etude de la relation quantitative structure-activitéinhibitrice des enzymes hydrolytiques : cas desalpha-glucosidases*.
37. Roy, K. & Mandal, A. S. (2008). Development of linear and nonlinear predictive QSAR models and their external validation using molecular similarity principle for anti-HIV indolyl aryl sulfones. Journal of enzyme inhibition and medicinal chemistry, 23, 980-995.
38. Roy, K., Roy, P. & Leonard, J. (2008). On some aspects of validation of predictive QSAR models. Chemistry Central Journal, 2, P9.
39. MELKEMI Nadjib, ThèseDOCTORAT EN SCIENCES ,Etude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelques composés organiques

40. BOUARRA Nabil,THÈSE Présentée *en vue de l'obtention du diplôme de Doctorat* en sciences Etudes QSPR des propriétés contrôlant l'évolution de quelques HAP dans l'environnement,2018.
41. Snedecor GW, Cochran WG (1967) Statistical methods. Oxford and IBH, New Delhi
43. Wold S, Sjöströ m M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. ChemomIntell Lab Syst 58:109– 130
44. K. Roy et al., A Primer on QSAR/QSPR Modeling,SpringerBriefs in Molecular Science, Author(s) 2015 DOI 10.1007/978-3-319-17281-1_2
45. Word, H. *Research Papers in Statistics*; Wiley: New York, NY, USA, 1966.
46. Jores-Kong, H.; Word, H. *Systems under Indirect Observation: Causality, structure, prediction*; North-Holland: Amsterdam, The Netherlands, 1982.
47. Rogers, D.; Hopfinger, A.J. Application of genetic function approximation to quantitative structure-activity-relationships and quantitative structure-property relationships. *J. Chem. Inf.Comput. Sci.* 1994, *34*, 854-866
48. Fan, Y.; Shi, L.M.; Kohn, K.W.; Pommier, Y.; Weinstein, J.N. Quantitative structure-antitumor activity relationships of camptothecin analogues: Cluster analysis and genetic algorithm-based studies. *J. Med. Chem.* 2001, *44*, 3254-3263.
49. Technologies and Materials for Renewable Energy, Environment and Sustainability, TMREES18, 19–21 September 2018, Athens, Greece.
50. Thomas Leonard, J.; Roy, K. Comparative QSAR modeling of CCR5 receptor binding affinity of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas. *Bioorg. Med. Chem. Lett.* 2006,*16*, 4467-4474.
51. Leonard, J.T.; Roy, K. Exploring molecular shape analysis of styrylquinoline derivatives as HIV- 1 integrase inhibitors. *Eur. J. Med. Chem.* 2008, *43*, 81-92.
52. Roy, K.; Ghosh, G. QSTR with extended topochemical atom (ETA) indices 8.(a) QSAR for the inhibition of substituted phenols on germination rate of Cucumis sativus using chemometric tools. *QSAR Comb. Sci.* 2006, *25*, 846-859.
53. . Leonard, J.T.; Roy, K. Comparative QSAR modeling of CCR5 receptor binding affinity of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas. *Bioorg. Med. Chem. Lett.* 2006, *16*, 4467-4474.
54. Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. Orthogonal signal correction of near-infrared spectra. *Chemometr. Intel. Lab. Syst.* 1998, *44*, 175-185.

55. Priolo, N.; Arribere, C.M.; Caffini, N.; Barberis, S.; Vazquez, R.N.; Luco, J.M. Isolation and purification of cysteine peptidases from the latex of Araujiahortorum fruits - Study of their esterase activities using partial least-squares (PLS) modeling. *J. Mol. Catal. B-Enzym.* 2001, *15*, 177-189.
56. Yang, S.B.; Xia, Z.N.; Shu, M.; Mei, H.; Lue, F.L.; Zhang, M.; Wu, Y.Q.; Li, Z.L. VHSEH Descriptors for the Development of QSAMs of Peptides. *Chem. J. Chinese Univ.* 2008, *29*, 2213- 2217.
57. Liang, G.Z.; Mei, H.; Zhou, Y.; Yang, S.B.; Wu, S.R.; Li, Z.L. Using SZOTT descriptors for the development of QSAMs of peptides. *Chem. J. Chinese Univ.* 2006, *27*, 1900-1902.
58. Modern drug design with advancement inQSAR: A review1224.
59. E. X. Esposito, A. J. Hopfinger, J. D. Madura, *Methods in Molecular Biology*. 275,2004, 131-213
60. Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P. & Agrawal, R. K. (2011). Validation of QSAR models-strategies and importance. International Journal of Drug Design & Discovery, 3, 511-519.
61. Roy, K., Kar, S.& Das, R. N. (2015b). Statistical methods in QSAR/QSPR. A primer on QSAR/QSPR modeling. Springer.
62. Roy, K. & Kar, S. (2015). How to judge predictive quality of classification and regression based QSAR models. Frontiers of Computational Chemistry; Haq, ZU, Madura, J., Eds, 71-120.
63. Hawkins, D. M., Basak, S. C. & Mills, D. (2003). Assessing model fit by cross-validation. Journal of chemical information and computer sciences, 43, 579-586.
66. Kar, S. & Roy, K. (2011). Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs
67. Edward R Mansfield and Billy P Helms. Detecting multicollinearity. *The American Statistician*, 36(3a) :158–160, 1982
69. Shi LM, Fang H, Tong W, Wu J, Perkins R, Blair RM, Branham WS, Dial SL, Moland CL, Sheehan DM . QSAR Models Using aLarge Diverse Set of Estrogens. *Journal of Chemical Information and Computer Sciences* 2001; 41 (1): 186-195.
70. Schümann, G., Ebert, R.-U., Chen, J., Wang, B. &Küne, R. (2008). External validation and prediction employing the predictive squared correlation coefficient mean vs training set activity mean. Journal of chemical information and modeling, 48, 2140-2145.

71. Consonni V, Ballabio D, TodeschiniR .Comments on definition of Q2 parameter for QSAR validation. *Journal of ChemicalInformation and Modeling* 2009; 49 (7): 1669-1678.
72. Consonni V, Ballabio D, TodeschiniR .Evaluation of model predictive ability by external validation techniques. *Journal OfChemometrics* 2010; 24(3): 194–201.
73. Chirico, N. &Gramatica, P. (2011). Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. Journal of chemical information and modeling, 51, 2320-2335.
74. Chirico N, GramaticaP . Real external predictivity of QSAR models. Part 2. New inter-comparable thresholds for different validationcriteria and the need for scatter plot inspection. *Journal of chemical information and modeling* 2011; 52(8): 2044-2058.
75. Roy, K. & Mitra, I. (2011). On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. Combinatorial chemistry & high throughput screening, 14, 450-474.
76. Jaworska, J., Nikolova-Jeliazkova, N. &Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set descriptor space: a review. ATLANOTTINGHAM-, 33, 445.

**المراجع العربية**

42. الانحدار الخطي المتعدد، أ.د. ثائر داود سلمان، فرع العلوم النظرية كلية التربيةالرياضية \جامعة بغداد.

64. مبادئ الإحصاء 102 كمى، الوحدة الخامسة: الارتباط الخطى البسيط، جامعة الملك خالد.

65. تحليل الانحدار الخطي البسيط ا د محمد راتول.

68. صياغة نموذج في إحصاء تقدير الإنتاج، م.م منى خضر اسطيفان، مجلة كلية بغدادللعلوم الاقتصادية جامعة العدد/ 2010

# خاتمة عامة

## خاتمة

بعد هذه الدراسة وجدنا أن برنامج QSARINS الذي يعكس ما توصلت إليه نماذج QSAR أضاف عدة مقاييس وبيانات الجديدة نسبيًا للحكم على جودة وقوة تنبؤ وتحقق لهذه النماذج إلى جانب مقاييس التحقق الأخرى ونأمل أن يساعد هذا المنظور التحليلي في التواصل بين الكيمياء الجزيئية والتجريبية نحو التطوير التعاوني واستخدام الأمثل لنماذج QSAR.

نعتقد أيضًا أن الإرشادات المقدمة هنا ستشجع على استخدام QSAR عالية الجودة والمصادق عليها لاتخاذ القرارات التنظيمية ومعايير علمية أكثر صرامة على المخطوطات التي تُبلغ عن دراسات QSARالجديدة. التي أثبتت أنها عملية تنبؤية وتشخيصية مقبولة على نطاق واسع لإيجاد ارتباطات بين الهياكل الكيميائية والنشاط البيولوجي.