

Introduction générale

L'informatique joue un rôle croissant dans la recherche en Chimie. Des secteurs très variés de la recherche fondamentale ou appliquée nécessitent des spécialités du traitement informatique, de l'information chimique, de la modélisation moléculaire ou de la chimie théorique.

La chimie se prête à un traitement informatique car elle est complexe et nécessite des capacités d'acquisition, de traitement et d'archivage considérables. D'importantes bases de données se constituent à travers le monde pour permettre aux chercheurs de suivre quasiment en temps réel l'avancement de la chimie. Le but est de montrer l'implication de l'informatique dans différentes applications de la chimie. Ce domaine est appelé "*Chemoinformatique*".

Le terme "*chemoinformatique* " est apparu il y a quelques années et a rapidement gagné l'utilisation répandue, Greg Paris a proposé une définition beaucoup plus large [1]

Chemoinformatique est un terme générique qui entoure la conception, création, organisation, gestion, récupération, analyse, diffusion, visualisation, et utilisation d'information chimique.

Clairement, la transformation des données en information et d'information en connaissance est un effort requis dans n'importe quelle branche de la chimie non seulement dans la conception de médicament. Nous partageons donc l'opinion que des méthodes de chemoinformatique sont nécessaires dans tous les secteurs de la chimie et adhérer à une définition beaucoup plus large :

Chemoinformatique est l'application des méthodes d'informatique pour résoudre des problèmes chimiques.

Pourquoi nous devons employer des méthodes d'informatique dans la chimie ? Beaucoup de problèmes en chimie sont trop complexes pour être résolus par des méthodes basées sur les premiers principes par des calculs théoriques. C'est vrai, premièrement, pour la relation entre la structure d'un composé et son activité biologique. Cependant, il s'applique également à beaucoup d'aspects de la réactivité chimique comme l'influence d'un catalyseur ou d'une température.

Dans cette situation, nous devons analyser des données expérimentales connues et établir un modèle pour les relations qui nous intéressent, comme entre la structure moléculaire et l'activité biologique ou la réactivité chimique.

Il y a une autre raison pour laquelle nous avons besoin des méthodes d'informatique dans la chimie : la chimie produit une quantité énorme de données, c'est particulièrement vrai pour des méthodes présentées dans la dernière décennie comme la chimie combinatoire et le criblage de haut débit. Juste le fait que plus de 45 millions de composés sont déjà connus et plusieurs millions de composés sont découverts tous les ans souligne le besoin de traitement d'information électronique pour gagner une vue d'ensemble de chimie connue. Ainsi, le stockage de l'information chimique dans les bases de données a une longue histoire. Nous avons une grande variété de bases de données dans la chimie, par exemple, bases de données sur la littérature, structures, réactions, spectres, et sur des données physiques, chimiques, ou biologiques.

Par exemple, toutes les structures de rayon X des composés organiques et organométalliques sont stockées dans la structure de fichier de données de Cambridge [2] qui contiennent actuellement plus de 300.000 structures. De même, des spectres sont stockés dans des bases de données de spectres avec le plus grand contenant, par exemple, 200.000 spectres (IRS) infrarouges. Grand en tant que ce nombre pourrait sembler, il est petit en effet par rapport au nombre de composés connus. En effet, nous connaissons les structures 3D moins de 1% de tous les composés, et nous avons moins de 1% des spectres IRS des composés connus en forme électronique. Ainsi, d'une part, nous avons beaucoup d'information mais, d'autre part, pour beaucoup de problèmes nous manquons de l'information nécessaire.

La question est donc, pouvons-nous apprendre assez des données que nous avons et pouvons-nous gagner la connaissance de ces données et d'information pour faire des prévisions pour le cas où l'information nécessaire n'est pas directement disponible ? C'est exactement l'un des secteurs où les méthodes de chimoinformatique peuvent entrer. Une grande partie des données a eu besoin, comme l'activité biologique d'un composé ou le taux ou le rendement d'une réaction chimique dans des conditions spécifiques, ne peut pas encore être directement calculé à partir des premiers principes en employant des méthodes théoriques dans un processus de l'étude déductive.

Dans cette situation, nous devons voir si nous pouvons apprendre assez des données déjà disponibles pour obtenir la connaissance qui permet la prévision de nouvelles données.

Dans cet effort nous devons mettre des données dans le contexte, nous devons rapporter différents types de données l'un avec l'autre pour créer l'information. Comme exemple, la seule valeur de l'activité biologique d'un composé ne nous aide pas beaucoup ; seulement quand nous connaissons également la structure chimique du composé à ce moment nous avons l'information valable. Avec beaucoup de morceaux d'information, par exemple, avec un ensemble de paires de structures chimiques et de leurs activités biologiques associées, nous pouvons essayer de généraliser, développer un modèle des relations entre la structure chimique et l'activité biologique, et obtenir ainsi la connaissance (la connaissance des relations entre la structure d'un composé et son activité biologique).

Ce processus de la connaissance dérivé des données et des observations s'appelle étude inductive. Les méthodes informatiques sont maintenant devenues disponibles pour l'étude inductive, comme des méthodes d'identification de modèle, des réseaux de neurones artificiels, ou des méthodes d'extraction de données. L'étude inductive a une longue histoire dans la chimie.

Problématique

L'informatique appliquée à la chimie et à la biologie requiert d'énormes capacités de calcul, de moyens de traitement, de stockage et de logiciels. Le domaine de la chimio-informatique est un domaine interdisciplinaire en plein essor. Il vise à apporter des solutions informatiques à des problèmes liés au traitement de l'information chimique (stockage, recherche, acquisition et exploitation de connaissances).

Par exemple, c'est en mêlant biologie, chimie, technologie et informatique que l'on cherche à mettre au point les médicaments de demain. La chemoinformatique est un secteur prometteur qui ouvre des horizons très larges dans la conception de nouveaux traitements. Il trouve aujourd'hui de nombreuses applications, en particulier dans les stratégies de recherche de nouvelles molécules bio-actives, susceptibles de devenir les médicaments de demain.

Développer un médicament coûte très cher. De plus, l'opération dure longtemps et se révèle parfois inefficace. Pour réduire la facture, les industriels préfèrent avant de lancer la synthèse chimique très onéreuse, réaliser une synthèse virtuelle (c'est-à-dire de concevoir et évaluer sur ordinateur) des molécules qui deviendront les médicaments de demain.

Cette simulation *in silico* permet de réduire le temps et le coût de développement des médicaments en décelant très tôt les plus prometteurs tout en limitant plus précocement les molécules qui échoueraient en développement.

C'est pourquoi la chimio-informatique occupe désormais une place de choix. Par exemple, dans le processus de fabrication de médicaments, elle s'intègre dans les diverses étapes du processus de découverte des futurs médicaments : conception et synthèse des molécules, identification des cibles thérapeutiques, évaluation des effets secondaires, etc.

Objectifs de la thèse

L'objectif de ce travail est de montrer comment des problèmes complexes, posés par un monde réel comme la chimie, peuvent se traduire en problématiques de recherche en informatique, qui fourniront à leur tour des réponses pertinentes du point de vue chimique.

Principalement, il s'agit de montrer comment les problématiques de la chimoinformatique peuvent conduire au développement de thèmes de recherche fondamentale en informatique dont les résultats fournissent des réponses pertinentes du point de vue chimique. On soulignera le caractère interdisciplinaire de ce travail qui nécessite le dialogue entre informaticiens, chimistes et biologistes, tant au stade de la modélisation du domaine qu'à celui de la validation des résultats. L'accent sera mis sur les possibilités offertes par la théorie des graphes pour représenter, comparer et classer les objets chimiques que sont les molécules et leurs réactions.

Contribution

Le travail effectué au cours de ce mémoire s'inscrit dans le cadre de *la prédiction de propriété physico-chimique*, pour cela nous proposons une méthodologie pour la construction d'un modèle prédictif à base de la méthode de modélisation par apprentissage statistique appelée *graph machines*. En partant du principe très général suivant : plutôt que la modélisation des molécules à partir des descripteurs qui conduit à une perte d'information et qui génère pour toutes les observations un seul modèle lors de l'apprentissage, il serait

envisageable de représenter les molécules directement à partir de leurs structures (par des graphes acycliques orientés) afin de s'affranchir des problèmes décrits précédemment.

Organisation du mémoire

Ce mémoire est structuré de la manière suivante :

- Chapitre 1 : l'objectif de ce chapitre est d'apporter une présentation du domaine de la chimoinformatique, ses spécificités ainsi que ses particularités. Pour cela, après avoir donné la définition de la chimoinformatique et les concepts de base pour comprendre le monde chimique, nous présentons brièvement quelques méthodes et outils ainsi que les logiciels utilisés dans ce domaine. Nous proposons ensuite un panorama des applications de la chimoinformatique existantes en mettant particulièrement l'accent sur les applications de prédiction de propriétés et d'activités des molécules. En montrant l'importance des méthodes informatiques pour la résolution des problèmes de la chimie.
- Chapitre 2 : Présente les méthodes traditionnelles de QSAR et de QSPR. Nous rappelons les principaux types de descripteurs, les problèmes liés à leur calcul et à leur sélection, ainsi que les principales techniques conventionnelles d'apprentissage et de sélection de modèle.
- Chapitre 3 : Introduit la notion de graphe, et décrit la genèse des *graph machines*, fonctions de même structure mathématique que les graphes auxquels elles sont associées.

Nous montrons alors comment ces fonctions permettent d'établir une relation entre des données structurées et des nombres.

- Le chapitre 4 : Présente l'intérêt de la méthode des *graph machines* par rapport aux méthodes traditionnelles, pour cela nous proposons une méthodologie de modélisation à l'aide des *graph machines* basée sur un codage qui tient compte directement de la structure des molécules que nous désignons par QSAR-GM (GM pour graph

machines). Dans ce codage, chaque molécule est représentée par un graphe acyclique orienté dont les nœuds sont associés aux atomes et les arêtes aux liaisons.

À la fin de ce document, une conclusion générale fait le bilan sur l'ensemble de ces travaux de recherche et indique des perspectives et des défis de la prédiction de propriétés chimiques dans le domaine de la chimoinformatique.

La chemoinformatique est une discipline scientifique qui a évolué dans les 10 dernières années à l'interface entre la chimie et l'informatique. Il a été constaté que, dans de nombreux domaines de la chimie, l'énorme quantité de données et d'informations produites par la recherche en chimie ne peut être traitée et analysée que par les méthodes assistées par ordinateur. Ainsi, les méthodes ont été développées pour la construction des bases de données sur les composés chimiques et leurs réactions, pour la prédiction des propriétés physiques, chimiques et biologiques des composés et des matériaux à l'échelle de l'atome et de la molécule, dans tous les secteurs de l'activité humaine, la conception des médicaments, pour élucider la structure, pour la prévision des réactions chimiques et de la conception de synthèse organique. La recherche et le développement sont essentiels en chemoinformatique d'une part pour accroître notre compréhension des phénomènes chimiques et d'autre part pour que l'industrie reste compétitive dans une économie mondiale.

C'est une branche de la chimie et/ou de la physico-chimie qui utilise les lois de la chimie théorique exploitées dans des codes informatiques spécifiques afin de calculer structures et propriétés d'objets chimiques (molécules, solides, clusters, surfaces ou autres), en appliquant autant que possible ces programmes à des problèmes chimiques réels.

Nous commençons par la définition de la chemoinformatique et les concepts de base pour comprendre le monde chimique, par la suite nous présentons brièvement quelques méthodes et outils ainsi que les logiciels utilisés dans ce domaine. En fin nous présentons un panorama des applications de la chemoinformatique existantes en mettant particulièrement l'accent sur les applications de prédiction de propriétés et d'activités des molécules.

1.1 Définition

La **chemoinformatique**, également dénommée chemoinformatique (anglicisme) ou chimio-informatique, est le domaine de la science qui consiste en l'application de l'informatique aux problèmes relatifs à la chimie. Elle a pour objectif de fournir des outils et des méthodes pour l'analyse et le traitement des données issues des différents domaines de la chimie.

Elle est notamment utilisée en **pharmacologie** pour la découverte de nouvelles molécules actives et la prédiction de propriétés à partir de structures moléculaires.

Certaines applications de la chemoinformatique reposent sur les équations de la **physique quantique**. Elles permettent ainsi de modéliser les conformations des **molécules** [3].

Cette très belle image qui représente des atomes liés par des liaisons chimiques n'est qu'une des innombrables possibilités de la chemoinformatique.

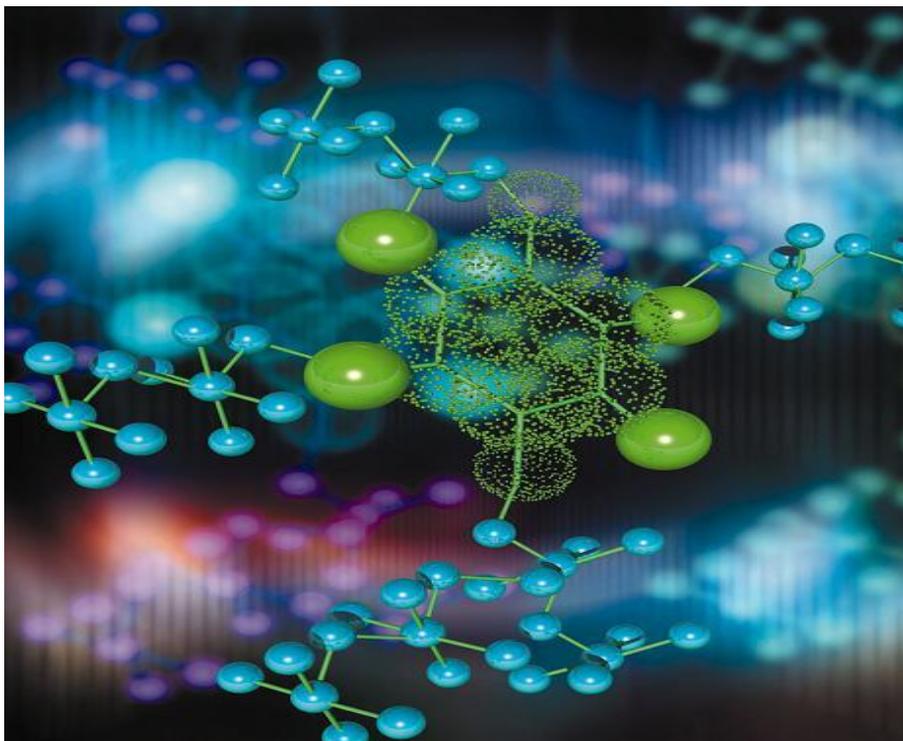


Fig.1.1 : Atomes liés par des liaisons chimiques.

La recherche et le développement en chemoinformatique sont essentiels pour :

- L'amélioration de la compréhension des phénomènes chimiques.
- Permettre à l'industrie chimique de rester compétitive dans le marché mondial [4].

1.2 Historique

La chemoinformatique est une nouvelle discipline apparue il y a environ 40 ans. Au début des années soixante, dans le but d'élucider la structure de composés chimiques inconnus, les données provenant des méthodes existantes (spectroscopie) ont été mises en commun sur informatique, C'était la naissance de la chemoinformatique. Le projet DENDRAL initié en 1964 à l'université de Stanford a été le premier à développer des générateurs de structures chimiques à partir de spectres de masses.

Ce n'est qu'à la fin des années 60 que Sasaki à l'université de technologie de Toyohashi et Munk à l'université d'Arizona ont utilisé plusieurs méthodes de spectroscopie afin d'élucider la structure chimique de leurs composés.

En 1969, Corey et Wipke ont présenté un travail similaire concernant les systèmes de représentation des molécules. Peu après d'autres groupes comme Ugi, Hendrickson et Gelernter ont développés des systèmes pour représenter des molécules organiques. La chemoinformatique (qui n'avait pas encore de nom) a été reconnue à la fin des années 60 comme une méthode utile d'analyse des données. Dès lors, beaucoup d'articles concernant cette discipline ont commencé à apparaître dans les journaux scientifiques. C'est seulement en 1998 que F.K Brown a défini pour la première fois cette discipline comme étant la chemoinformatique [5].

1.3 Concepts de base

1.3.1 La chimie

La **chimie** est une **science** de la nature divisée en plusieurs spécialités, à l'instar de la **physique** et de la **biologie** avec lesquelles elle partage des espaces d'investigations communs ou proches.

Pour reprendre un canevas de présentation proposée par la plus grande association de chimistes au monde, l'*American Chemical Society*, la chimie étudie [6] :

1. les éléments chimiques à l'état libre, **atomes** ou **ions** atomiques, et les innombrables et diverses associations par liaisons chimiques qui engendrent notamment des composés moléculaires stables ou des intermédiaires plus ou moins instables. Ces entités de **matière** peuvent être caractérisées par une identité reliée à des caractéristiques quantiques et des propriétés précises.
2. les processus qui changent ou modifient l'identité de ces particules ou molécules de matière, dénommés **réaction, transformation, interaction...**
3. les mécanismes intervenant dans les processus chimiques ou les équilibres physique entre deux formes. Leurs définitions précises permettent de comprendre ou d'interpréter avec des hypothèses l'évolution matérielle avec en vue une exploitation des résultats de façon directe ou induite.
4. les phénomènes fondamentaux observables en rapport avec les forces de la nature qui jouent un rôle chimique, favorisant les réactions ou synthèse, addition, combinaison ou décomposition, séparation de phases ou extraction. L'analyse permet de découvrir les compositions, le marquage sélectif ouvre la voie à un schéma réactionnel cohérent dans des mélanges complexes.

1.3.1.1 Une science segmentée en de multiples disciplines

La recherche et l'enseignement en chimie sont organisés en disciplines qui, souvent en absence de services, de coopération ou d'aides réciproques, s'ignorent et se développent en toute autonomie [7] :

- la **biochimie** qui étudie les réactions chimiques dans des milieux biologiques (cellules...) et/ou avec des objets biologiques (protéines...).

- la **chimie analytique** est l'étude des méthodes d'analyses qualitatives et/ou quantitatives qui permettent de connaître la composition d'un échantillon donné ; ses principaux domaines sont : la chromatographie et la spectroscopie;
- la **chimie des matériaux** est la préparation et l'étude de substances avec une application en tant que matériau. Ce domaine intègre des éléments des autres domaines classiques de la chimie avec un intérêt particulier pour les problèmes fondamentaux concernant les matériaux.
- la **chimie inorganique** ou chimie minérale, concerne la description et l'étude des éléments chimiques et des composés sans squelette carboné.
- la **chimie organique** est la description et l'étude des composés comportant un squelette d'atomes de carbone (composés organiques) ;
- la **chimie physique** dont l'objet est l'étude des lois physiques des systèmes et procédés chimiques ; ses principaux domaines d'étude comprennent : la thermochimie, la cinétique chimique, l'électrochimie, la radiochimie, la sonochimie et les spectroscopies.
- la **chimie théorique** est l'étude de la chimie à travers un raisonnement théorique fondamental (habituellement à l'aide des mathématiques et de la physique). En particulier, l'application de la mécanique quantique à chimie a donné naissance à la chimie quantique. Depuis la fin de la seconde guerre mondiale, le progrès des ordinateurs a permis le développement de la chimie numérique (ou computationnelle).

1.3.2 L'atome : est formé d'un noyau atomique contenant des nucléons qui maintient autour de lui un nombre d'électrons équilibrant la charge positive du noyau.

1.3.3 L'élément chimique : est l'ensemble des atomes qui ont un nombre donné de protons dans leur noyau. Ce nombre est son numéro atomique. Par exemple, tous les atomes avec 6 protons dans leurs noyaux sont des atomes de l'élément carbone. Ces éléments sont

représentés dans le tableau périodique (Figure 1.2), qui rassemble les éléments de propriétés similaires.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	H																	He	
2	Li	Be											B	C	N	O	F	Ne	
3	Na	Mg											Al	Si	P	S	Cl	Ar	
4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr	
5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe	
6	Cs	Ba	*	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
7	Fr	Ra	*	Lr	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Uut	Uuq	Uup	Uuh	Uus	Uuo
			↓																
			*	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb		
			*	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No		

Fig.1.2 : Tableau périodique des éléments chimiques.

1.3.4 La liaison chimique: est le phénomène qui lie les atomes entre eux en échangeant ou partageant un ou plusieurs **électrons** ou par des **forces électrostatiques** .

1.3.5 La molécule : est un ensemble électriquement neutre d'atomes associés par des **liaisons covalentes**.

Exemple : L'eau : trois atomes, deux éléments, deux **liaisons**, une molécule. Un atome d'oxygène (ici en rouge), se lie à deux atomes d'hydrogène (ici en blanc).

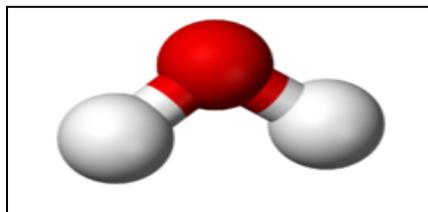


Fig.1.3 : Schéma de la molécule d'eau.

1.3.6 L'ion : est une **espèce chimique** (un atome ou une molécule) qui a perdu ou qui a gagné un ou plusieurs électrons. Il est appelé **cation** lorsqu'il est chargé positivement et **anion** lorsqu'il est chargé négativement.

1.3.7 Le composé chimique : est une substance issue de l'assemblage de plusieurs types d'atomes issus d'éléments chimiques différents dans des proportions définies. Il est caractérisé par sa **formule chimique**.

1.3.8 La réaction chimique : Une **réaction chimique** est une transformation de la matière au cours de laquelle les espèces chimiques (atomiques, ioniques ou moléculaires) qui constituent la matière sont modifiées : les espèces qui sont consommées sont appelées réactifs. Les espèces formées au cours de la réaction sont appelées produits (de réaction). Depuis les travaux de Lavoisier (1777), les scientifiques savent que la réaction chimique se fait sans variation mesurable de la masse : « Rien ne se perd, rien ne se crée, tout se transforme » qui traduit la conservation de la masse.

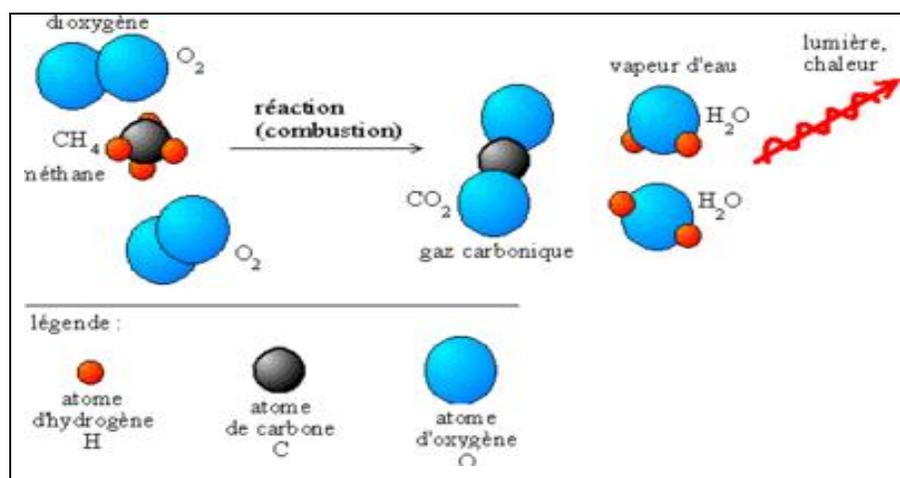


Fig.1.4 : Réaction chimique (échange d'atomes entre les composés, exemple de la combustion du méthane dans le dioxygène).

1.3.9 Propriété physico-chimique de molécule (des protéines)

1.3.9.1 Dénaturation

Une protéine est dénaturée lorsque sa conformation tridimensionnelle spécifique est changée par rupture de certaines liaisons sans atteinte de sa structure primaire. Il peut s'agir, par exemple, de la désorganisation de zones en hélice α . La dénaturation peut être réversible ou irréversible. Elle entraîne une perte totale ou partielle de l'activité biologique. Elle produit très souvent un changement de solubilité de la protéine [8].

Les agents de dénaturation sont nombreux :

- agents physiques : chaleur, radiations, pH ;
- agents chimiques: solution d'urée qui forme de nouvelles liaisons hydrogène dans la protéine, solvants organiques, détergents...

1.3.10 Prédiction de propriété ou d'activité (Relation quantitative structure à activité)

Une relation quantitative structure à activité (en anglais : *Quantitative structure-activity relationship* ou QSAR, parfois désignée sous le nom de relation quantitative structure à propriété - en anglais : *quantitative structure-property relationship* ou QSPR) est le procédé par lequel une structure chimique est corrélée avec un effet bien déterminé comme l'activité biologique ou la réactivité chimique.

Ainsi par exemple l'activité biologique peut être exprimée de manière quantitative, comme pour la concentration de substance nécessaire pour obtenir une certaine réponse biologique. De plus lorsque les propriétés ou structures physico-chimiques sont exprimées par des chiffres, on peut proposer une relation mathématique, ou *relation quantitative structure à activité*, entre les deux.

L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de la réponse biologique pour des structures similaires.

La QSAR la plus commune est de la forme : activité = f (propriétés physico-chimiques et/ou structurales) [9].

1.4 Vue d'ensemble

La section suivante donne une vue d'ensemble de chemoinformatique, soulignant les problèmes et les solutions communs aux divers sous-domaines plus spécialisés. Ces matières constituent également les divers chapitres du manuel de Chemoinformatique [10] et du manuel de Chemoinformatique [11].

1.4.1 Représentation des composés chimiques

Une gamme entière des méthodes pour la représentation sur l'ordinateur des composés et des structures de produit chimique a été développée comprenant des codes linéaires, des tables de raccordement, et des matrices.

Des méthodes spéciales ont dû être conçues pour représenter uniquement une structure chimique, pour percevoir des dispositifs tels que l'aromaticité et pour traiter la stéréochimie, les structures 3D, ou les surfaces moléculaires.

1.4.2 Représentation des réactions chimiques

En manipulant des réactions chimiques il ne suffit pas d'indiquer seulement les produits de départ et les produits d'une réaction mais on doit également indiquer l'emplacement de réaction et les liens cassés et faits dans une réaction. En outre la stéréochimie des réactions doit être manipulée.

1.4.3 Données en chimie

Beaucoup de **connaissances chimiques ont été dérivées des données**. La chimie doit offrir une gamme riche des données sur les propriétés physiques, chimiques, et biologiques, par exemple, données binaires pour la classification, vraies données pour la modélisation, et données spectrales ayant une densité élevée de l'information. Ces données doivent être introduites dans une forme favorable à l'échange d'information facile et analyse de données.

1.4.4 Les sources de données et les bases de données

L'énorme quantité de données en chimie a mené au développement des bases de données pour stocker et disséminer les données en forme électronique. Par exemple, des bases de données ont été développées pour la littérature chimique, composés chimiques, structures 3D, réactions, et spectres. L'Internet est de plus en plus employé pour distribuer des données et l'information en chimie.

1.4.5 Méthodes pour calculer des données physiques et chimiques

Une variété de données physiques et chimiques des composés peut directement être calculée par une gamme des méthodes. Les premiers sont les calculs mécaniques de quantum de divers degrés de sophistication. Cependant, des méthodes simples telles que des arrangements d'additivité peuvent également être employées pour estimer une variété de données avec l'exactitude raisonnable.

1.4.6 Calcul des descripteurs de structure

Dans la plupart des cas, cependant, physique, chimique, ou les propriétés biologiques ne peuvent pas être directement calculés à partir de la structure d'un composé. Dans cette situation, une approche indirecte doit être adoptée : la structure du composé est représentée par des descripteurs de structure par la suite un rapport entre les descripteurs de structure et la propriété est établie en analysant une série de paires de descripteurs de structure et les propriétés associées par les méthodes d'étude inductives (Figure 1.5).

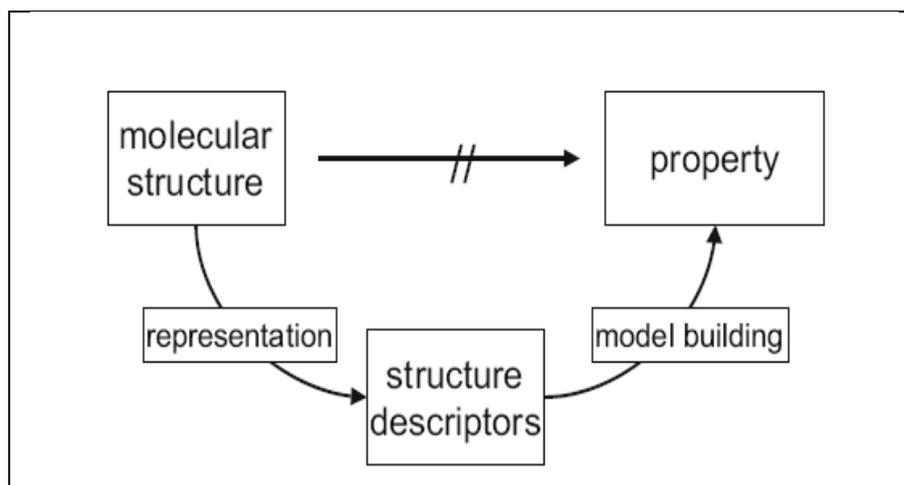


Fig.1.5 : Établissement des relations de structure-propriété/activité.

Une variété de descripteurs de structure ont été développées 1D, 2D, ou information de la structure 3D ou les propriétés de la surface moléculaires.

1.4.7 Méthodes d'analyse de données

Une variété de méthodes pour apprendre des données par des méthodes d'étude inductives sont employées dans la chimie, par exemple, statistiques, méthodes d'identification de modèle, réseaux neurones artificiels, et algorithmes génétiques. Ces méthodes peuvent être classifiées dans des méthodes d'apprentissage supervisé et non supervisé et sont employés pour la classification ou modélisation quantitatif.

1.5 Méthodes et outils

1.5.1 Méthodes de la modélisation moléculaire

La modélisation moléculaire a pour but de prévoir la structure et la réactivité des molécules ou des systèmes de molécules.

Les méthodes de la modélisation moléculaire peuvent être rangées en trois catégories [12] :

1.5.1.1 Méthodes quantiques

Ces méthodes sont basées sur le calcul des orbitales moléculaires (OM). Les principales variantes sont :

1. **La méthode de Hückel**

C'est la plus simple de toutes. Elle ne prend en compte que les électrons p et utilise des approximations assez draconiennes.

2. **Les méthodes de champ auto-cohérent (SCF, Self Consistent Field)**

Ces méthodes prennent en compte les électrons s et reposent sur des calculs plus élaborés que la méthode de Hückel.

3. **Les méthodes basées sur la fonctionnelle de la densité (DFT, Density Functional Theory)**

Ces méthodes utilisent une expression de l'énergie électronique E en fonction de la densité électronique ρ , elle-même fonction de la position r de l'électron : $E = f[\rho(r)]$

1.5.1.2 Mécanique moléculaire

Cette technique calcule l'énergie des atomes au moyen d'approximations. La simplification des calculs permet de travailler sur des molécules de grande taille, ou sur des systèmes comportant un grand nombre de molécules.

1.5.1.3 Dynamique moléculaire

Cette technique a pour but de calculer les mouvements des molécules, le plus souvent à partir des énergies de la mécanique moléculaire. Elle permet de simuler l'évolution des systèmes dans le temps.