

## **1-INTRODUCCIÓN GENERAL.**

### ***1.1-Concepto de medioambiente.***

Los seres desarrollan su vida en un espacio físico rodeado por otros organismos y el medio físico y socioeconómico. Los factores bióticos y abióticos interactúan entre sí generando un lugar propio y dicho espacio se denomina ambiente. La Conferencia de las Naciones Unidas sobre Medio Ambiente en Estocolmo (1972) lo define como: “Medio ambiente es el conjunto de componentes físicos, químicos, biológicos y sociales capaces de causar efectos directos o indirectos, en un plazo corto o largo, sobre los seres vivos y las actividades humanas”, citado en el libro “Agenda 21” de Foy (1998).

En términos macroscópicos se suele considerar al medioambiente como un sector, una región o un todo (escala global). En cada uno de esos niveles o alcances de estudio hay una interacción entre los factores anteriormente mencionados, especialmente del aire, del agua o del suelo como agentes abióticos y de toda una gran variedad de organismos animales y vegetales, con distinto nivel de organización celular, como integrantes del mundo biótico (Valverde et al., 2007).

La rama de la química que se ocupa del estudio de estos escenarios se denomina Química Ambiental, para ello es necesario tener bien presente el concepto de Medioambiente, y consecuentemente realizar una clasificación de compartimentos ambientales o esferas ambientales, que interactúan entre sí experimentando intercambios continuos de materia y energía (Echarri, 1998; Tyler Miller, 2007). Tradicionalmente se ha considerado al agua, al aire, a la tierra y a la vida, es decir, a la Hidrósfera, a la Atmósfera, la Geósfera y la Biósfera, donde se tenía a la actividad humana como una actividad de perturbación permanente en las distintas esferas. Actualmente se ha incluido una quinta esfera denominada Antropósfera (Manahan, 2007), la que involucra a todos los seres humanos y las actividades relacionadas con su desarrollo y supervivencia (Figura 1.1.1). Esta visión moderna promueve al hombre como parte del medio ambiente, además de incentivar a que las acciones ejecutadas por él, no sólo minimicen el daño hacia el resto de las esferas, sino también que se puedan ejercer acciones beneficiosas para todo el planeta.

En este contexto han surgido ramas de la ciencia dedicadas a estudiar en forma integral al medio ambiente: Ecología, Ecotoxicología, Físicoquímica Ambiental, Biología Aplicada, que en conjunto con la Química Ambiental y las Ciencias Sociales, evalúan permanentemente el estado poblacional, el nivel de contaminación y las metodologías de diagnóstico y remediación ambiental, tal como lo indica Manahan (2007).

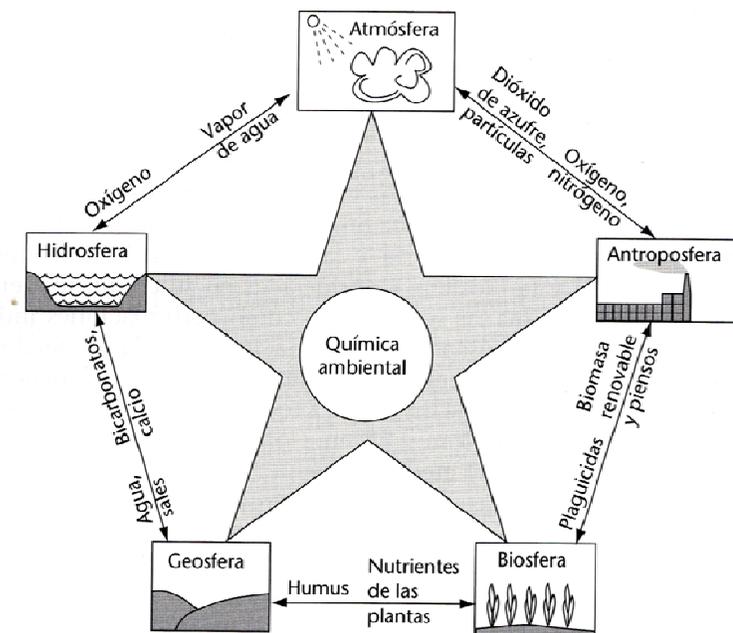


Figura 1.1 Las cinco esferas del ambiente. La materia y la energía se mueven constantemente entre estas esferas, como campo de estudio de la química ambiental. Tomado de Manahan (2007).

## 1.2-Ecosistema.

El clima, los suelos, las bacterias, los hongos, las plantas y los animales de cualquier lugar definido constituyen juntos un ecosistema (Pianka, 1982). Éste comprende la comunidad biológica junto con su ambiente físico (Begon et al., 1999). McNaughton y Wolf (1998) resumen esas ideas definiendo el campo de estudio de la ecología como “La ecología es una ciencia ambiental que trata de descubrir y comprender la forma, cómo actúan los factores ambientales sobre los organismos y cómo los organismos son capaces a la vez de modificar su entorno”

. En un ecosistema se necesita, además de los factores bióticos y abióticos, algo fundamental: una fuente de energía (Odum, 1992). La misma fluye a través de todos los componentes, mantiene la vida y moviliza el agua, los minerales y otros componentes físicos (Tyler Miller, 1998). Esta acción energética, en nuestro planeta, principalmente es provista por el sol. La movilización de estos componentes, se produce en forma permanente en un ecosistema, bajo un sistema cerrado, conservativo, donde se transfieren del aire al agua, al suelo, a los seres vivos y de esta manera la materia se recicla, acompañada de un flujo de energía. Esos procesos se engloban bajo la denominación de “ciclos biogeoquímicos”

(Butcher et al., 1992). “Los ecosistemas se estudian analizando las relaciones alimentarias, los ciclos de la materia y los flujos de energía” (Heinrich, 1990)

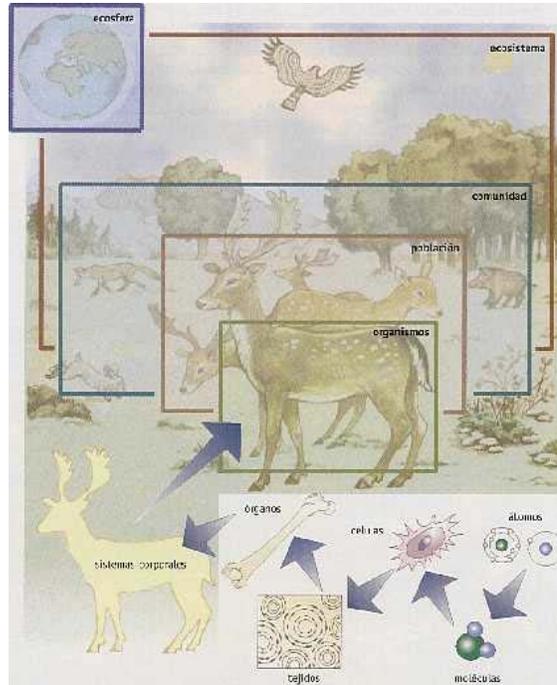


Figura 1.2.1: Esquema de organización de un ecosistema. Desde las átomos hasta los organismos superiores. Tomado de Echarri (1998)

### 1.3-Propiedades fisicoquímicas. Interacciones entre compartimentos.

Todos los compuestos químicos, tanto orgánicos como inorgánicos, están formados por un conjunto de átomos variables en número y tipo. A su vez estas uniones atómicas adoptan una distribución en el espacio según las leyes de la química convencional. Estas características microscópicas se traducen en propiedades macroscópicas medibles, a las cuales se les puede atribuir un número y una unidad. Al conjunto de esas magnitudes físicas medibles, producto de las características químicas microscópicas, se las denominan Propiedades Fisicoquímicas (Chang, 1997).

Desde el punto de vista ambiental, las más relevantes son aquellas que permiten inferir el comportamiento en cada esfera ambiental de las distintas sustancias. Posiblemente las más relevantes resulten el punto de ebullición ( $P_e$ ) y la presión de vapor ( $P_v$ ), ya que estas permiten evaluar el comportamiento de una sustancia en la atmósfera; la solubilidad en agua ( $S$ ), para ponderar la interacción de los compuestos con hidrósfera y el coeficiente de partición octanol-agua ( $K_{ow}$ ) que mide la capacidad de reparto de un compuesto químico entre las fases acuosa y orgánica (Figurello et al., 2004); ayudando a estimar el grado de interacción que puede tener esa sustancia con el tejido vegetal y/o animal a la hora de atravesar membranas biológicas (Connell et al., 2005).

Los distintos estados de agregación de la materia en el medioambiente están en permanente coexistencia, y numerosos equilibrios de reparto están puestos en juego en y entre cada fase de la corteza terrestre, los cuales están íntimamente vinculados con las propiedades moleculares de los distintos compuestos (Domènech y Peral, 2006).

*Figura 1.3.1: Interacciones entre los distintos compartimentos ambientales. Equilibrios involucrados en cada uno de ellos. Construido a partir de Domènech y Peral (2006)*

#### **1.4-Contaminantes ambientales.**

Hasta mediados de los años de 1960, la entrada de nuevas sustancias químicas al mercado se realizaba sin mayores requisitos. Sin embargo, a raíz del número creciente de casos de enfermedades crónicas asociadas con la entrada al ambiente de sustancias cuyos efectos adversos se desconocían, se reconoció que, antes de permitir su comercialización, se debería evaluar las propiedades toxicológicas y ambientales de las sustancias, para disminuir los riesgos en la población y proteger el equilibrio ambiental (Foy, 1998).

La Sociedad Americana de Química (American Chemical Society, ACS) es responsable de llevar un registro de estas nuevas sustancias bajo el número de CAS (Chemical Abstract Service), siendo esta una identificación numérica que se les asigna a los compuestos a medida que hacen su aparición en la literatura. A la fecha se encuentran registrados alrededor de cuarenta y cuatro millones de sustancias, con una tasa de doce mil registros diarios (CAS, database). Asimismo la gran mayoría no tienen ningún uso después de su descubrimiento (Fuente: Sitio web oficial de la CAS: <http://www.cas.org/>).

Para una sustancia nueva dependerá de su agresividad química/bioquímica la cantidad necesaria para interactuar con el medio, para superar su umbral de tolerancia (Baran, 1995), a estas sustancias se las conoce con el nombre de *xenobióticos* (xeno=extraño y bio =vida). Conjuntamente las actividades antrópicas también pueden generar un aumento en la concentración de algunas especies químicas presentes en el ambiente en forma corriente (*Biogénicos*) (Tyler Miller, 2007). El Programa Nacional de las Naciones Unidas (PNUMA) (Castañón del Valle, 2006) define como sustancias químicas de importancia mundial "*aquellas que se encuentran ampliamente diseminadas en el ambiente, en cantidades de importancia, como resultado de su transporte a través del aire, el agua o las cadenas tróficas, o porque están presentes en las mercancías que son objeto de comercio internacional a gran escala*". También se incluyen en este grupo las sustancias que sólo son importantes en el ámbito local o regional pero preocupan a varios países.

Actualmente se ha incorporado al concepto de contaminantes a los agentes físicos y biológicos, un ejemplo de los primeros puede ser el ruido excesivo (contaminación sonora), la radiación, las ondas electromagnéticas; para el caso de los biológicos pueden citarse a los desechos cloacales, los virus, las bacterias. Manahan (2007) define "En términos generales entonces, puede considerarse a un contaminante como todo agente, físico, químico o biológico que produce alteraciones en los equilibrios establecidos en y entre los compartimentos ambientales o bien de una combinación de varios agentes en lugares, formas y concentraciones tales que sean o puedan ser nocivos para la salud, la seguridad o para el bienestar de la población, o bien, que puedan ser perjudiciales para la vida vegetal o animal, o impidan el uso normal de las propiedades y lugares de recreación y goce de los mismos".

Un caso especial de agentes químicos, de estrecha relación con la contaminación, son aquellos que pueden encuadrarse en la definición de Tóxicos haciendo referencia a toda sustancia que puede producir algún efecto nocivo sobre un ser vivo, alterando sus equilibrios vitales tal como lo indican Bello y López (2001). Estos autores destacan a su vez, la importancia de una rama de la toxicología que se dedica a investigar "lo que puede alterar el equilibrio de un ecosistema y trata de cuantificar los efectos no deseables de los contaminantes, su confirmación experimental y su mecanismo de acción", esta definición es retomada por otros autores bajo la denominación de Ecotoxicología (Capó Martí, 2007), más aún, Repetto y Camean (2005), añaden a la definición que, además de los equilibrios y la cuantificación, también estudia cómo "contribuir a la recuperación del trastorno producido por la llegada de un tóxico al ecosistema".

#### *1.4.1-Biomarcadores.*

Según la definición ya clásica de McCarthy y Shugart (1990), los biomarcadores son "medidas" en los niveles molecular, bioquímico o celular, tanto en poblaciones naturales provenientes de hábitats contaminados, como en organismos expuestos experimentalmente a contaminantes, y que indican que el organismo ha estado expuesto a sustancias tóxicas y la magnitud de la respuesta del organismo al contaminante". Este concepto de biomarcador se

fundamenta en que, para que los efectos de los contaminantes sean evidentes en el ecosistema, el primer nivel de interacción es el molecular-celular. Este nivel de organización es además el ideal para estudiar relaciones causa-efecto y los mecanismos de acción de los contaminantes (Galloway et al., 2004; Depledge y Galloway, 2005).

Así, la propuesta o desarrollo de un biomarcador debe basarse en el conocimiento del mecanismo de respuesta del mismo (Capó Martí, 2007). Otras características que los biomarcadores deben poseer son la reproducibilidad, sensibilidad, especificidad, reversibilidad, aplicabilidad en diversas especies, facilidad de uso y relación costo/efectividad adecuado (Newman y Unger, 2003). Es fundamental que el tiempo de respuesta del biomarcador sea corto (horas, días), de forma que éste pueda ser utilizado como "sistema de alarma temprano" y puedan tomarse medidas de reparación antes de que el ecosistema se vea afectado (Cajaraville et al, 2000). Si, además, el biomarcador muestra valor diagnóstico, éste podrá ser entonces utilizado de forma predictiva (Long et al., 2004)

El análisis de biomarcadores de exposición y de efecto, normalmente, en combinación con otras aproximaciones para la evaluación de la calidad del medio, como los análisis químicos convencionales, los bioensayos y los estudios ecológicos a largo plazo, son un instrumento clave para detectar el impacto de la contaminación sobre la salud de los ecosistemas (Halbrook 1991; Peakall y Kennedy, 1999; Long et al., 2004; Depledge y Galloway, 2005)

### ***1.5-Etapas analíticas involucradas en las mediciones ambientales.***

La gran diversidad química del medio ambiente requiere del desarrollo de metodologías analíticas particulares y la adaptación de herramientas de la química analítica convencional tal que ayuden a conocer tipo y cantidad, tanto de los componentes mayoritarios como los trazas. Esta información es necesaria para comprender el "estado químico" de un lugar, establecer el grado de impacto de la actividad antrópica, presencia y niveles de contaminantes químicos (Baird et al, 2001)

Los contaminantes pueden repartirse entre el aire, el agua o el suelo en función de sus propiedades fisicoquímicas, conforme a las leyes de reparto (Figueruelo et al., 2004). Esta situación hace que deban enfocarse estrategias diferenciadas de análisis según cada matriz ambiental. En términos generales pueden citarse las etapas bien conocidas de la química en general: el muestreo, el análisis químico y la interpretación de resultados (Csuros, 1994; Khopkar, 2004).

#### ***1.5.1-Muestreo.***

En cuanto al muestreo, la esfera que requiere más cuidados a la hora de ser estudiada resulta ser la atmósfera, por el manejo de gases y los niveles de dilución alcanzados en esta matriz para los distintos compuestos (Keith, 1996; Khopkar, 2004), en la mayoría de los

casos los niveles de detección instrumentales no pueden ser aplicados en forma directa por lo que se requiere una concentración de los analitos por alguna operación cuidadosa de concentración (Hess, 1996), esto se logra mediante el uso de soluciones absorbedoras en el caso por ejemplo del SO<sub>2</sub>, NO<sub>2</sub> y de dispositivos de adsorción generales y específicos, llevando un volumen de varias decenas de litros a algunos mililitros, pueden estar involucradas reacciones de derivatización durante el proceso de adsorción sobre un soporte sólido activo, como por ejemplo los aldehídos y las aminas (Colls, 2002; Zhang et al, 2007). Muchas sustancias están formando parte de la atmósfera sorbidas en el material particulado en suspensión lo que necesita del diseño de dispositivos especiales que permitan la toma de muestras minimizando la perturbación el tamaño de las partículas retenidas ni los equilibrios de sorción-desorción existentes entre los analitos y el soporte sólido brindado por las partículas, este es el caso de aplicación de impactadores de partículas de alto volumen, utilizado para el muestreo de Hidrocarburos Aromáticos Policíclicos y sus derivados, Fenoles y algunos metales presentes en la atmósfera (Hesketh, 1996)

En la hidrósfera las estrategias de muestreo principalmente se orientan a la recolección de un volumen de agua transportable desde el recurso natural hasta el laboratorio, tratando de direccionar esa recolección a distintas profundidades, para lo que se cuenta con elementos que van desde dispositivos simples como una botella, o un recipiente que pueda contener líquidos, hasta los más complejos como sondas para líquidos, o toma muestra de fondo, independientemente de los sistemas de monitoreo *in situ* (APHA, AWWA, WEF, 1998). En la geósfera los muestreos están direccionados a los suelos y sedimentos, donde una porción de ellos es recolectada y trasladada al laboratorio para su análisis. Para los suelos se pueden tomar muestras mediante un barrenador, una calicata o simplemente una pala, para los sedimentos la principal herramienta de muestreo es la draga (Keith, 1996).

En la toma de muestras de la biósfera nuevamente dependerá de la esfera química relacionada con el material biológico a investigar y estas estrategias se vinculan íntimamente con la especialidad de la ecología y ecotoxicología (Suter, 2000).

### 1.5.2-Análisis.

Existen numerosas técnicas analíticas con distintos objetivos, algunas de ellas apuntan a caracterizar la muestra, a determinar los componentes mayoritarios que ayudan a una especificación de rutina, de especiación, a la determinación de parámetros fisicoquímicos predeterminados que cuantifiquen los distintos componentes de la matriz. El desafío se presenta a la hora de la determinación de elementos o compuestos trazas, que generalmente apuntan a conocer los niveles de compuestos que generan un impacto ambiental importante aun en concentraciones muy bajas (algunas partes por billón son suficientes) y esto se debe a la sensibilidad de los sistemas biológicos ante algunas sustancias (Zhang, 2007).

Más allá de las técnicas puntuales, a lo largo de la investigación en química analítica ambiental, se han podido sistematizar estas operaciones en una serie de pasos bien definidos que pueden presentarse como (Barceló, 1993; Khopkar, 2004, Smith y Cresser, 2004; Harris, 2007)

- Extracción del analito desde la matriz de origen en función de propiedades químicas generales (orgánicos polares, no polares, por grupos funcionales; inorgánicos, metales y no metales), donde se aplican principios de separaciones básicos como líquido-líquido, líquido-sólido, destilaciones, precipitación.
- Concentración, las fases provenientes de la extracción se someten a procesos de reducción de volumen minimizando las pérdidas del analito, esto generalmente involucra un proceso de evaporación con la ayuda de distintos dispositivos instrumentales según el tipo y cantidad de solvente.
- Purificación o “clean-up”, luego de la concentración se procede a realizar lo que se puede definir como una “segunda extracción”, pero en vez de ser general es específica. En este paso la molécula objetivo se separa del resto de las presentes provenientes de la extracción primaria sometiendo el extracto concentrado a un sistema de interacciones específicas y selectivas, generalmente estas condiciones son reunidas por una técnica cromatográfica a escala semipreparativa o extracciones con solventes activos.
- Análisis instrumental, el extracto purificado y concentrado se somete a alguna metodología de identificación y cuantificación haciendo uso de los instrumentos disponibles para tal fin: espectrofotómetros, voltímetros, cromatógrafos líquidos y gaseosos, etc. Todos ellos permiten tener confirmación de presencia/ausencia y en caso de lo primero determinar la cantidad presente con la mayor precisión y exactitud posible.

Todo analista debe tener presente que en cada una de estas etapas se produce una pérdida del analito a causa de trabajar con operaciones que no son 100% eficientes ya sea por el manejo del operador (evitable o reducible) o por las interacciones químicas intrínsecas en cada etapa (inevitables) (Barceló, 1993). Esto obliga a un manejo cuidadoso de las muestras ambientales en todo su proceso de análisis y a la determinación del factor de eficiencia (o recuperación) en la metodología global para la consiguiente corrección de los resultados.

Existen en la literatura muchas normas y metodologías analíticas internacionales publicadas que abarcan cada una de estas etapas, entre ellas pueden citarse a las normas de la USEPA (United States, Environmental Protection Agency), de la ASTM (Analytical Standard Testing Methods), para ambientes naturales y las de la NIOSH (National Institute for Occupational Safety and Health), de la OSHA (Occupational Safety and Health Administration) para ambientes laborales. Cada día se incorporan a las publicaciones científicas casos particulares para el tratamiento de muestras ambientales, es el caso de Environmental Research, Environmental Pollutions, Atmospheric Environment, Chemosphere, Water Toxicology, Environmental Sciences and Technologies, Sciences of Total Environment, Journal of Environmental Chemistry y tantas otras. Asimismo a nivel nacional se encuentran las Normas Iram de amplia difusión y aplicación en nuestro país.

### 1.5.2.1-Análisis Instrumental Moderno.

De todas las técnicas analíticas de uso corriente en laboratorio, para la medición de residuos en el ambiente, probablemente las más ampliamente difundidas sean las cromatográficas (De Brabander et al., *in press*), desde sus bases conceptuales a cargo de Tswett (1905), Willstatter y Stoll (1913) y Martin y Synge (1941) hasta la cromatografía líquida de ultra-resolución (Ultra Performance Liquid chromatography, UPLC), desarrollada en los últimos años (Castro-Perez et al., 2005), por su versatilidad y rango de aplicaciones..

Conjuntamente a estas técnicas separativas se han desarrollado toda una serie de dispositivos encargados de la detección de las sustancias, desde los más simples como un sensor de conductividad hasta un espectrómetro de masas (Mass Spectrometry Detector, MSD) (Siri, 1947; Eiceman, 2000). Este desarrollo instrumental ha mejorado notablemente los límites de detección sobre muestras de interés ambiental al minimizar el manipuleo y mejorar la selectividad y sensibilidad de los detectores empleados (Khopkar, 2004). El surgimiento de la espectrometría de masas, donde la detección se realiza a través de la masa molecular de la especie química en cuestión (McLafferty, 1993) y su acoplamiento con sistemas de separación de avanzada como un cromatógrafo de gases (Gas Chromatography, GC) o de líquidos de alta resolución (High Performande Liquid Chromatography, HPLC) ha resuelto simultáneamente el problema del aislamiento y la identificación de los distintos componentes de una mezcla (Oliver, 1998).

Los mayores avances tecnológicos en este campo están en la fusión de la cromatografía líquida y la espectrometría de masas, la combinación de estas dos poderosas técnicas de la química analítica dan la capacidad de analizar prácticamente cualquier tipo de especies moleculares, incluyendo, térmicamente lábil, no volátil y de alto peso molecular. Se ha dicho que más del 80% de especies orgánicas conocidas son susceptibles de separación con cromatografía líquida (Snyder y Kirkland, 1991), asimismo la espectrometría de masas es capaz de proporcionar la estructura, el peso molecular, fórmula empírica, y la información cuantitativa acerca de un analito (McLafferty, 1993). El mayor desafío que afrontó esta técnica fue la compatibilización de un espray líquido, proveniente del HPLC, con el alto vacío y la fase gaseosa en la que funciona el MSD, con los inicios de la investigación en los años 70`s, los primeros sistemas confiables surgieron 20 años después (Niessen, 1999)

Existen distintos tipos de interfases, para resolver los dos problemas principales que plantea este tipo de conexión (Niessen, 2003):

a) eliminar la gran cantidad de gas, y vapor procedente de la fase móvil antes de entrar a la región de alto vacío del MS; y

b) transformar las moléculas, en solución de la fase móvil, en iones en fase gaseosa, sin que se produzca degradación térmica.

Siendo, a la fecha, las más comercializadas las denominadas de ionización a presión atmosférica (Atmospheric Pressure Ionization, API), ellas son la electrospray (ES) o

ionización química a presión atmosférica (Atmospheric Pressure Chemical Ionization, APCI) (Niessen, 20036).

La muestra luego de pasar por la interfase/ionización ingresa al espectrómetro de masas propiamente dicho para luego llegar al analizador de iones, donde se obtienen las respuestas de masa/carga ( $m/z$ ) de los iones formados. La energía de los iones es convertida en una señal eléctrica, que es registrada por un sistema gráfico, analógico o digital, obteniéndose: el tipo de ión (masa) e intensidad del mismo, proporcional a la concentración (McLafferty, 1993). Estas respuestas están en relación a la separación analítica previa, por lo que el otro parámetro que caracteriza al compuesto es el tiempo de tención (Kovats, 1958), ésta terna de datos identifica y cuantifica al compuesto de interés con un alto grado de certeza en la identidad y la concentración.

### ***1.5.3-Interpretación de resultados***

Con las determinaciones químicas de los componentes mayoritarios y el barrido de una serie de parámetros relacionados con indicadores de contaminación, comienza un camino de estudio e interpretación que involucra a distintas áreas de las ciencias teórico-experimentales como pueden ser a la ecología, la ecotoxicología, la biología, la ingeniería, la fisicoquímica y las estadísticas aplicadas. Cada una de ellas aporta una visión desde su objetivo de estudio y en conjunto aportan al diseño de estrategias de remediación, de contención y de seguimiento de diversas problemáticas ambientales, en base a:

- ✓ resultados numéricos de las mediciones
- ✓ observaciones generales del tipo descriptivas
- ✓ desarrollo y validación de modelos teóricos y su compatibilidad con los hechos experimentales

#### ***1.5.3.1-Métodos integrados de análisis de datos.***

En la sistemática del análisis ambiental es corriente, que para una determinada muestra, se realicen un conjunto de mediciones orientadas a generar un conocimiento cualitativo y cuantitativo de los distintos componentes en el ambiente de estudio (Baird et al., 2001; Manahan, 2007). Por ejemplo para un curso de agua superficial pueden medirse parámetros generales como pH, conductividad, dureza, alcalinidad, y parámetros específicos de contaminación como pesticidas organoclorados y metales pesados. En este caso, el curso de agua, es el objeto de estudio y los parámetros medidos las variables métricas que lo describen. Desde el formalismo matemático se dice que si se mide un objeto mediante  $n$

medidas se obtiene un perfil métrico (un vector) de  $n$  medidas representándolo. Si luego cada una de estas  $n$  medidas es usada por separado para hallar la media y el desvío estándar poblacional, el vector del objeto es desmembrado y el objeto es perdido. Esta es la limitación fundamental de la estadística univariada, no hay un vector real, o perfil, representando objetos o grupos de objetos (Ramirez, 1999).

En los últimos veinte años, además de los métodos estadísticos convencionales, se han comenzado a aplicar técnicas derivadas del análisis multivariado de datos a todo un conjunto de disciplinas (Chandra et al., 2001; Chiu, 2002; Mirko, 2006), entre ellas, las ciencias ambientales (Ballesteros-Grijalva et al., 1996; King y Jackson, 1999; Pineda-Torres et al., 2002; Serita et al., 2005). La ventaja de la estadística multivariada reside en preservar los objetos como un todo en forma integrada, un vector de medidas, y no un inventario de datos independientes. Cualquier medida está en el contexto de las otras medidas del mismo objeto y de otros objetos formando una matriz coherente. Es por esto que la matriz de covarianza es la base matemática de estos análisis (Howells, 1969). Los análisis multivariados consisten en crear nuevas variables que resumen la mayor parte de la variación del espacio y permiten ver la posición de los objetos en unos pocos ejes (Howells, 1969), abarca las técnicas estadísticas que consideran dos o más variables aleatorias relacionadas como una entidad produciendo un resultado general según las relaciones entre ellas (Jackson, 1991; Hair et al., 1999).

En algunos casos suelen definirse a las técnicas multivariadas como una extensión de las univariadas, por ejemplo la de regresión lineal múltiple como una extensión de la regresión simple (Dillon y Goldstein, 1984). Existen, sin embargo, otra clase de técnicas multivariadas, que si bien pueden ser empleadas para la inferencia estadística, tienen su principal aplicación en el análisis de datos (Jackson, 1991). Sus métodos analizan conjuntamente las  $n$  variables medidas sobre un conjunto de  $p$  individuos u objetos. Las técnicas estadísticas utilizan datos de conjuntos de variables medidas en múltiples objetos y estos datos se escriben en forma de tabla o matriz, siendo procesados por medio de programas computacionales (*XLSTAT*, *STATISTICA*, *R*) leyendo dicha matriz y operando con ella. Sería impensable el desarrollo del Análisis Multivariado sin la ayuda de la Informática, pues dichos métodos procesan, en general, gran cantidad de datos y a pesar de haber sido desarrollado en los años 30, recién a fines de los 70's (Manly, 1994).

Una primera diferenciación entre los distintos métodos se basa en los objetivos que persiguen. Genéricamente, se distinguen cuatro posibles objetivos (Dillon y Goldstein, 1984; Mainly, 1994)

a) *Simplificación estructural*: se trata de describir la información original de forma sintética o resumida. El método consiste en condensar las  $n$  variables originales en un número menor de nuevas variables creadas por el propio análisis, que contengan sin embargo gran parte de la información original. A este tipo de objetivo, y de método, se denomina reducción de la dimensión ya que los datos originales se expresan en un espacio de dimensión  $n$  mientras que como resultado del análisis podemos expresarles en otro espacio de menor dimensión  $m$ . Las nuevas variables creadas por el análisis se denominan ejes, factores, o componentes principales. Son variables sintéticas, que estarán en función de las  $n$  variables

originales; generalmente, estos factores tienen una interpretación o significado que el investigador debe descubrir. El Análisis de Componentes Principales (ACP), el Análisis Factorial (AF), y el Análisis de Correspondencias (AC) son métodos de reducción de la dimensión.

b) *Clasificación o agrupación*: estos métodos incluyen los de agrupamiento (Análisis Cluster) y los de Segmentación. La agrupación de individuos consiste en formar grupos de individuos homogéneos en cuanto a las  $n$  variables y heterogéneos respecto a los otros grupos. La agrupación de variables busca la formación de grupos de variables similares en cuanto a su comportamiento en un colectivo de objetos.

c) *Análisis de interdependencia*: se trata de buscar la interdependencia entre grupos de variables, sin que *a priori* se suponga relación de causalidad entre ellas. El método más conocido es el Análisis de Correspondencias, que es una generalización del Análisis de Correspondencias Bivariante.

d) *Análisis de dependencia*: explica las relaciones entre grupos de variables, donde se supone que unas pueden ser causas de otras. El Análisis de Regresión pertenece a este grupo de métodos. Un tipo interesante de análisis de dependencia consiste en buscar un criterio que permita separar o discriminar entre objetos pertenecientes a priori a grupos diferentes. Dicho criterio es una función de las variables originales. En último término, se trata de usar los resultados en el futuro para predecir a qué grupo pertenecen nuevos objetos que no formaban parte de la información original y para los cuales se han medido las  $n$  variables. El Análisis Discriminante (AD) y la Regresión Logística son métodos que persiguen este objetivo.

Estas herramientas, aplicadas al ejemplo de los cuerpos de aguas superficiales, pueden ayudar por ejemplo a decidir cuáles son los parámetros más relevantes en cada región (Via ACP) o si los ríos de una región pueden diferenciarse por lectura de algunos componentes en particular (AD) (Topalian et al., 1999; Parinet et al., 2004; Busso y Santa Cruz, 2005; Lucas y Jauzein, 2008)

Conjuntamente con la expansión de los análisis multivariados, han nacido nuevas metodologías para el tratamiento de datos a partir del desarrollo de los sistemas informáticos. Los esfuerzos estuvieron direccionados en imitar el funcionamiento del sistema nervioso del ser humano, más específicamente en el sistema de aprendizaje, es decir de las neuronas y su funcionamiento (Gurney, 1997; Haykin, 1999) así surgen los sistemas de análisis de datos denominados *Redes Neuronales Artificiales* (RNA o del inglés Artificial Neural Network, ANN).

Desde el punto de vista biológico, una neurona consta de un cuerpo celular más o menos esférico, del que salen una rama principal, el axón, y varias ramas más cortas, llamadas dendritas. Una de las características de las neuronas es su capacidad de comunicarse. En términos generales las dendritas y el cuerpo celular reciben señales de entrada; el cuerpo celular las combina e integra y emite señales de salida y el axón transmite dichas señales a los terminales axónicos, que distribuyen información o un nuevo conjunto de neuronas (Shepherd, 1988). Una neurona se puede considerar como un pequeño procesador

de información, que recibe entradas, las procesa y devuelve salidas a otras neuronas y al estar conectadas unas con otras, la salida de una es la entrada de la otra.

En los sistemas artificiales, las interacciones químicas, son reemplazadas por datos de entrada y salida, que pueden ser bits, números enteros o decimales. Una función procesa estas señales y devuelve el valor resultante a otras neuronas a las que está conectada (Haykin, 1994). Estas neuronas se agrupan formando capas, una de entrada, que recibe la información desde el exterior, una oculta que procesa la información y una de salida, que devuelve la información procesada al exterior ( Figura 1.5.3.1.1). De manera general, un sistema neuronal artificial es un conjunto de funciones ponderadas, es decir que no todas las funciones tienen el mismo peso en el cálculo, donde los resultados que emiten unas son tomados como datos para procesar por las siguientes. La red utiliza datos externos que son cargados a modo de validación sobre los cuales realiza el aprendizaje, este procedimiento es logrado mediante distintos tipos de algoritmos predefinidos y ampliamente investigados (Hornik, 1993; Nicolaos, 1993). Como cuando el ser humano ve una persona y luego la reconoce en una foto, compara el ingreso de la imagen con un patrón de referencia previo, teniendo como herramienta un mecanismo de aprendizaje para la asociación de imágenes).

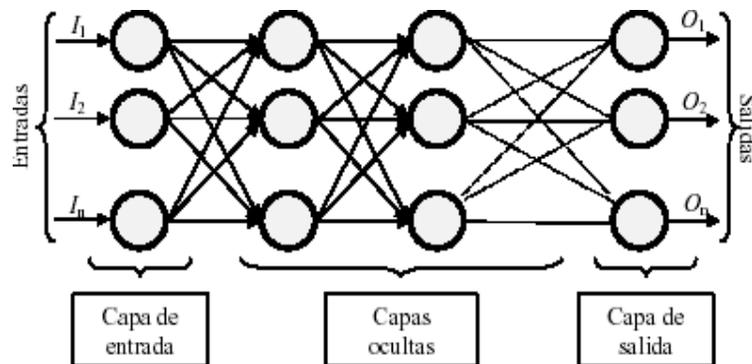


Figura 1.5.3.1.1: esquema de una Red Neuronal, capa de entrada, capa oculta o de entrenamiento y capa de salida.

Estos sistemas de procesamientos de datos han sido ampliamente adoptados por la comunidad científica para predecir la ocurrencia de distintos eventos, para ver cual/es variables son más importantes en estos procesos, para la elaboración de modelos de reconocimiento/identificación. En el campo de las ciencias ambientales en los últimos 10 años han aparecido diversas publicaciones dedicadas a la investigación de contaminantes ambientales (Gardner y Dorling, 1998; Maier y Dandy, 2000; Kolehmainen et al., 2001), la relación con sus fuentes (Reich et al., 1999; Daniel y Sivakumar, 2009) y la elaboración de modelos predictivos (Perez et al., 2000; Chelani et al., 2002) basados en redes neuronales artificiales.

En un estudio ambiental, luego del muestreo, el análisis de laboratorio, la optimización instrumental y el procesamiento de resultados, comienza la etapa de interpretación de las mediciones, las posibles causas y efectos derivados, en base a distintas

propiedades fisicoquímicas de las sustancias y sus equilibrios de reparto, por lo que es necesario apelar a herramientas propias de la fisicoquímica (Figueruelo et al., 2004).

### ***1.6-Fisicoquímica y sus alcances.***

El aporte químico en las evaluaciones ambientales involucra la descripción de los eventos que ocurren en el ambiente, el conocimiento de las concentraciones de algunos compuestos en los distintos compartimentos, la relación con sus propiedades fisicoquímicas y los equilibrios involucrados en estos repartos, confluyen para tal fin áreas de la química analítica y la fisicoquímica. Algunos lineamientos de la química analítica fueron descriptos anteriormente resaltando las principales fases por las que pasa una muestra ambiental desde el momento de la toma hasta el resultado final provisto por un número y una unidad. Los fenómenos de reparto y su vinculación con las propiedades fisicoquímicas necesitan para su descripción estrategias integradas de modelización y formalización matemática tanto de la variante experimental como teórica, macroscópica y microscópica.

#### ***1.6.1-Fisicoquímica moderna. Fisicoquímica Teórica y las computadoras.***

La fisicoquímica no ha quedado al margen de la revolución científica brindada por la mecánica cuántica en los años 1900-1905 de la mano de Planck y Einstein, en las temáticas vinculadas a la radiación del cuerpo negro y al efecto fotoeléctrico respectivamente (Einstein, 2002). En el año 1928 Dirac propone un desarrollo matemático, basado en los conceptos de sus antecesores, que permitió explicar el comportamiento de las partículas descubiertas por ese entonces entre ellas el SPIN (Dirac, 1962), así con 25 años de tradición y menos de tres años de desarrollo nace una nueva ciencia: *la mecánica cuántica* con su formulación definitiva.

Un nuevo mundo de investigación se abre a partir de estas nuevas herramientas, los postulados y los teoremas establecidos permitieron interpretar las propiedades químicas observadas en “primeros principios”, esto implicó conocer la información al máximo punto de detalle. Las partículas se analizaron desde un comportamiento dual onda-partícula, y tras amplios debates de la comunidad científica de la época a cargo de Heisenberg, Born, Jordan, Schrödinger, Planck, Dirac y el mismo Einstein, estas “nuevas ideas” quedaron definitivamente instaladas como una guía de trabajo para muchas ramas de las ciencias, muchas de estas aplicaciones al campo de la química se muestran en la literatura básica del tema (Glasstone, 1948; Levine, 1991; Szabo y Ostlund, 1996; Bort, 2001; Cramer, 2002; Simons, 2003)

La principal dificultad, ampliamente resuelta por los investigadores de la época, de la mecánica cuántica fue la complejidad matemática, no por la falta de ideas, sino por las herramientas de cálculo numérico, el uso de tablas de cálculo, las estimaciones y aproximaciones para facilitar los manejos numéricos. El nacimiento de las computadoras en

el año 1950, con la primera generación comercial de computadoras (aplicadas en el censo de los Estados Unidos de América en el año 1951), han favorecido notablemente las perspectivas de aplicación y manejo de los conceptos cuánticos en todas sus potencialidades. Desde entonces se han direccionado los esfuerzos en desarrollar cada vez equipos de procesamiento de menor tamaño, de mayor capacidad de almacenamiento y respuesta y todo ello a un costo de comercialización aceptable por el mercado. Todos coincidimos que esto está bien logrado al día de hoy, desde la primera computadora con un procesador de 160 m<sup>2</sup> desarrollada por la ENIAC (Electronic Numerical Integrator And Computer) a los actuales 0,06 m<sup>2</sup> de la notebook con la que se redacta este manuscrito, sin tener en cuenta las velocidades de respuesta !.

### 1.6.2-Fisicoquímica Computacional.

La química computacional está compuesta de dos grandes áreas basadas en principios físicos distintos. Por una parte, la mecánica molecular (MM) que se basa en la mecánica clásica, considerando los átomos como partículas puntuales (dotadas de masa y carga) interaccionando unos con otros mediante enlaces que se asimilan mayormente a resortes. Por otra parte, los métodos de estructura electrónica (MEE) basados en la aplicación de la mecánica cuántica a los sistemas atómicos y moleculares (Eyring et al., 1944; Cramer, 2002). Tanto la MM como los MEE permiten calcular un conjunto básico de propiedades, por ejemplo la energía de un determinado arreglo de núcleos (conformación de la molécula), la geometría óptima de un sistema molecular (es decir, el arreglo geométrico de los núcleos que brinde la energía más baja) y otras propiedades moleculares como las frecuencias vibracionales y el momento dipolar (Smith y Sutcliffe, 1997; Bort, 2001, Young, 2001).

Los distintos métodos dentro de la MM comparten el hecho de tomar sus fundamentos de las leyes de la física clásica. Se diferencia en: (i) el tipo de ecuaciones que describen como varía la energía potencial de una molécula con la geometría de sus átomos; (ii) el tipo de átomos, es decir el tipo de características que definen a un átomo de número atómico dado en el entorno molecular en el que se encuentre (con lo cual, un oxígeno carboxílico no es equivalente necesariamente a un hidroxílico); y/o (iii) el conjunto de parámetros que ajustan las ecuaciones aplicadas a los átomos definidos por su tipo, a los valores experimentales que se emplean para la parametrización. Estos tres componentes definen lo que se conoce como campo de fuerza y definen completamente un método de MM (Allinger y Burkert, 1982).

Los MEE, emplean la mecánica cuántica para estudiar el comportamiento de núcleos y electrones (y no sólo núcleos como en MM) que se consideran como partículas puntuales con carga y masa fijas e invariables, interaccionando según la ley de Coulomb. Los MEE se dividen en tres grupos principales: *ab initio*, funcionales de la densidad y semi-empíricos en función que no usen más que las constantes atómicas fundamentales o empleen datos experimentales para parametrizar parte de los cálculos. Todos los MEE se basan en el planteo y solución aproximada de la ecuación de Schrödinger (Namboori et al., 2008).

El modelo *ab initio* del Halmiltoniano provee una representación completa de todas las interacciones no relativistas entre el núcleo y los electrones en una molécula. Sin embargo, la solución a la ecuación de Schrödinger es necesariamente una aproximación cuyo tiempo de cálculo es proporcional al número de electrones de la molécula. Muchos de los cálculos *ab initio* se basan en la aproximación de los orbitales moleculares de Hartree-Fock (Szabo y Ostlund, 1989). En general este método provee buenos resultados para la energía electrónica total de la molécula. Otros métodos *ab initio* son ampliamente utilizados y comercializados lo que permite el cálculo de correlaciones electrónicas de las moléculas de una forma fácil. Hay que tener en cuenta que la mayoría de estos métodos no son prácticos y fallan en la convergencia en el caso de moléculas mayores de diez átomos (Smith y Sutcliffe, 1997).

Los métodos *ab initio* Hartree-Fock y post-Hartree-Fock tienen el enorme inconveniente de ser muy costosos desde el punto de vista computacional y, consecuentemente, aplicables sólo a moléculas de tamaño reducido. El factor limitante en este tipo de métodos es el tiempo necesario para calcular el gran número de integrales bielectrónicas sobre las funciones bases y/o para transformarlas en integrales sobre orbitales moleculares. Los métodos de la MM por el contrario, son aplicables a grandes moléculas, pero no tienen en cuenta en forma explícita la estructura electrónica. Es por ello necesario encontrar una metodología intermedia que permita tratar la estructura electrónica de las moléculas de forma aproximada, sin los costosos requisitos de los métodos *ab initio* (Cramer, 2002).

Como métodos alternativos, se han desarrollado métodos semi-empíricos, estos métodos están basados en la teoría de orbitales moleculares y usan simplificaciones que reducen dramáticamente el tiempo de cálculo al eliminar gran parte o una parte de las integrales mencionadas anteriormente (Thiel, 2000). Algunos parámetros se obtienen a partir de los datos experimentales de los átomos y/o de los sistemas moleculares prototipos. Las cargas, dipolos y longitud de los enlaces obtenidas por algunos métodos semi-empíricos son más realistas que aquellas obtenidas con métodos *ab initio* de baja calidad (Janoschek, 1997). Es necesario hacer notar que, los resultados obtenidos por diferentes métodos semi-empíricos no son comparables de manera general, aunque dan tendencias similares (Karelson et al., 1996).

### ***1.7-Relaciones propiedad-actividad-estructura.***

En su momento Dalton planteó: "*Los compuestos químicos están formados por átomos de más de un elemento. En cualquier compuesto, la relación entre el número de átomos está establecida por números enteros o fraccionarios sencillos*". Con el avance del conocimiento, a la unidad que representa a un compuesto se le dio el nombre de molécula y asociada a ellas aparece el concepto de materia. Es cotidiano que la materia se nos presente en tres estados: sólido, líquido y gaseoso. Las magnitudes asociadas a estos estados son de carácter físico:

punto de ebullición, punto de fusión, presión de vapor, índice de refracción, etc., permiten tener una idea del tipo e intensidad de interacciones energéticas entre las moléculas intervinientes (Findlay et al., 1979).

De hecho, ocurre que estos parámetros varían regularmente según el tipo de compuesto que se trate, cuando se estudian los compuestos sólidos formados por átomos de carbono e hidrógeno, se observa un aumento de sus puntos de fusión conforme aumenta el número de átomos de C en la molécula, a estos compuestos se los conoce con el nombre de alcanos, y todo compuesto que responda con esas características químicas se dice que pertenece a la familia de los alcanos (McMurry, 2000). Existen diversidad de familias de compuestos que se agrupan según sus propiedades químicas citándose por ejemplo a los alcoholes (presentan H, C, y obligatoriamente al grupo -OH en su estructura molecular), las aminas (conteniendo el grupo -NH<sub>2</sub>, en su estructura molecular), los éteres (que resultan de la unión de dos grupos alquilo al oxígeno R-O-R') (McMurry, 2000)

Toda la química existente está formada por una conglomeración de familias de sustancias, si averiguamos el tipo de reacciones que da una molécula desconocida y luego cuantificamos sus propiedades macroscópicas podemos llegar a tener idea de al menos su familia de pertenencia (Vogel, 1991). En química eventualmente es necesario conocer la estructura que presenta esta molécula, entendiéndose por estructura a la disposición espacial que ofrecen los átomos. Además al descifrar la estructura estamos encontrando una explicación microscópica a una evidencia macroscópica. Volviendo al ejemplo de los alcanos, experimentalmente se sabe que si bien los puntos de fusión aumentan conforme aumenta el número de átomos de carbono, si se mantiene fijo el contenido de carbono y se permutan las estructuras, las propiedades fisicoquímicas variarán de acuerdo a la interacción espacial de la molécula (Fig. 1.7.1) (McMurry, 2000).

Por lo tanto es imprescindible conocer la estructura de una molécula para poder definirla en su integridad. Estas propiedades asociadas a una estructura molecular, pueden vincularse a través de distinto tipos de funciones matemáticas, que de alguna manera "midan" estas relaciones propiedad-estructura. Esta sistematización del estudio se engloba bajo la denominación de técnicas QSPR (del inglés Quantitative Structure Property Relationships) (Hammet, 1933; Fredenslund et al., 1975; Kier et al., 1976; Joback et al., 1987).



Figura 1.7.1 moléculas isoméricas de C<sub>4</sub>, misma composición, distintas propiedades fisicoquímicas observables.



En los sistemas biológicos los conocimientos básicos indican que los cambios estructurales de las moléculas definen en su acción sobre el organismo vivo, a esta acción se la llama actividad, entonces en vez de establecerse relaciones entre propiedad estructura se las trata como actividad-estructura QSAR.



### 1.7.1-Breve reseña histórica

Este proceso de cuantificación de la información estructural/molecular evolucionó a lo largo de los años en una serie de etapas. La primera de ellas fue en el área de regresión estadística con el desarrollo del análisis de los componentes principales (Hotelling, 1933), le siguió el uso de propiedades físicas para describir una molécula y armar modelos tratando de medir la calidad de la respuesta obtenida a partir de ellos (Hansch y Fujita, 1964). Al mismo tiempo aparece una alternativa paradigmática para el modelado de propiedades físicas: la teoría de orbitales moleculares (Kier, 1971), la cual intentaba combinar cuantificaciones definidas, con estructuras que permitirían medir la calidad del nuevo modelo predictivo, alcanzando su forma final con la innovación del uso de la metodología de los cuadrados mínimos parciales (Wold y Sjostrom, 1974, 1998) que mejoró el peso estadístico de los resultados.

Tomando como base la teoría de grafos se desarrollaron los índices topológicos para la codificación las estructuras moleculares (Kier y Hall, 1976; Randic 1986; Trinajstić, 1992), conjuntamente se adaptaron herramientas del álgebra lineal para la elección de las variables más óptimas, cuyas componentes resulten ortonormales entre sí, las que se llamaron Descriptores Ortogonales (Randic, 1991). Desde ese momento y hasta la fecha los avances son tan veloces que resulta difícil enumerarlos, lo que sí se sabe es que los resultados obtenidos, superan, en calidad y cantidad, a los que se hubieran imaginado los iniciadores del sistema.

### 1.7.2-Índices topológicos.

En este modelo se describe a los átomos como puntos en el espacio asumiendo distintas disposiciones relativas. La idea fue útil para describir la existencia de diversas estructuras con

el mismo tipo y relación de átomos (Wiener, 1947; Hosoya, 1971; Katritzky, 1993). Existen modelos matemáticos que permiten cuantificar las relaciones propiedad-estructura, estas representaciones teóricas aproximadas se basan en los índices topológicos, que permiten caracterizar una molécula por un simple número.

La necesidad de representar estructuras moleculares por un número característico surge del hecho de que la mayoría de las propiedades moleculares están descritas por números (números de átomos, longitudes de enlace, ángulo de enlace, etc.). En este paso existe una pérdida considerable de información, acompañada de una ventaja importante: una molécula de tres dimensiones puede ser descrita en una dimensión: su descriptor molecular (Todeschini y Consonni, 2000). Sin embargo es sorprendente que la mayoría de la información estructural relevante sea retenida en los índices topológicos (Wiener, 1947; Castro et al., 2002). Una selección severa nos permite quedarnos con aquellos descriptores que se ajusten con las reglas generales presentadas en la tabla 1.7.2.1 (Randic, 1992; Balaban, 1988).

**Tabla 1.7.2.1:** Características que debe cumplir un descriptor para ser aplicado en un modelo tipo QSAR/QSPR

- 
- 1- Interpretación estructural directa.
  - 2-Buena correlación con las propiedades moleculares
  - 3-Buena discriminación de isómeros.
  - 4-Localmente definidos.
  - 5-Generabilizable.
  - 6-Línealmente independiente.
  - 7-Simplicidad.
  - 8-No deben estar basados sobre propiedades físicas o químicas.
  - 9-No deben ser triviales sobre otros índices.
  - 10-Eficiencia de construcción.
  - 11-Basados sobre conceptos estructurales.
  - 12-Correcto tamaño de dependencia.
  - 13-Cambio gradual con cambio gradual de estructura.
- 

En cuanto al área biológica, puede resultar más conveniente trabajar con descriptores moleculares como alternativa de los índices topológicos, ya que éstos son calculados sobre una base de dos, tres o mayores dimensiones (Oprea, 2002) y tiene incorporado en su construcción algunos efectos electrónicos y estéricos tridimensionales (Randic et al., 1988; Randic, 1997; Randic et al., 2000), que resultan ser determinantes en un mecanismo de acción biológica.

Las propiedades propuestas por Randic que se presentan en la tabla 1.7.2.1, que se exigen que cumplan los índices, representan un muy alto nivel de significación. Los índices

presentes en la tabla 1.7.2.2 se aproximan a este ideal. Su punto más débil es la diferenciación de isómeros a excepción del índice de Balaban.

**Tabla 1.7.2.2:** índice tipológicos más comunes usados en la metodología QSAR/QSPR

| Índice topológico | Símbolo | Interpretación estructural                                  | Autor (año)                         |
|-------------------|---------|---|-------------------------------------|
| de Wiener         | $W$     | Suma de las distancias                                      | Wiener (1947)                       |
| de Hozoya         | $Z$     | Suma de uniones no adyacentes                               | Hosoya (1971)                       |
| de Randic         | $\chi$  | Suma de uniones ponderadas                                  | Randic (1975)                       |
| de Balaban        | $J$     | Suma ponderal de distancias                                 | Balaban (1982)                      |
| de Schultz        | $MTI$   | Suma de los elementos de una fila de una matriz de valencia | Schultz (1989)                      |
| de Hararay        | $H$     | Suma de los cuadrados de las distancias recíprocas          | Plavsic,Nikoliic, Trinajstic (1991) |

### 1.7.3-Conceptos elementales en la teoría de grafos.

La teoría de grafos es una rama discreta de la matemáticas basada sobre fundamentos de combinatoria y topología. Esto habla de la forma en que los objetos son conectados, englobando a su vez todas las consecuencias de tal situación (Harary, 1971; Cvetkovic et al., 1995)

Específicamente la teoría de grafos químicos es un área de la matemáticas química, es decir de la química teórica. Estos grafos representan sistemas químicos, donde teóricamente se evalúan todas sus posibilidades de unión y por ende todas las consecuencias derivadas de ello (Bonchev y Rouvray, 1990; Barysz, 1983). Matemáticamente hablando un grafo es la aplicación de un conjunto sobre sí mismo, esto es, una colección de elementos de un conjunto y sus relaciones binarias, conduciendo a espacios vectoriales de multidimensionales (Randic, 1986). Para un químico el uso de dos dimensiones es suficiente: representado por un conjunto de vértices y sus uniones. Un grafo  $G$  (Fig. 1.7.3.1) puede ser visualizado por un diagrama donde los vértices son caracterizados por pequeños círculos y las uniones por líneas o curvas. Cuando un grafo tiene sus vértices numerados se dice que esta nivelado, como  $G$ .

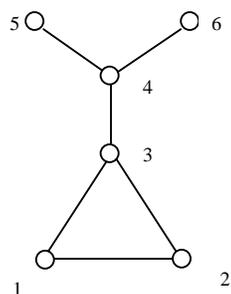


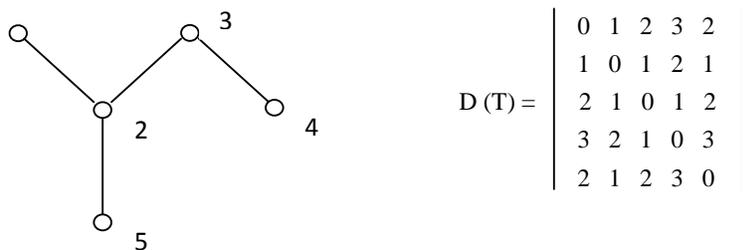
Figura 1.7.3.1: Grafo  $G$  Muestra un diagrama de un grafo en dos dimensiones.

Químicamente un grafo se puede utilizar para representar una variedad de objetos: moléculas, reacciones, cristales, polímeros y "clusters". Sus características comunes son la presencia de determinados sitios y conexiones. Estos sitios pueden ser átomos, electrones, moléculas, fragmentos moleculares, etc., y las conexiones entre ellos representan uniones, pasos de reacción, fuerzas de Van der Waals, etc. Una clase especial de grafos químicos son los grafos moleculares o constitucionales, ellos representan específicamente la construcción de moléculas (Trinajstić, 1992).

#### 1.7.4-Grafos moleculares.

Se pueden aceptar como grafos moleculares a aquellas disposiciones esquemáticas que representan en forma concreta la estructura de una molécula. Distinguiéndose distintos tipos de grafos, uno de los más usados por su simplicidad son los grafos de hidrógenos suprimidos (Fig. 1.7.4.1), donde se representa la molécula de forma convencional exceptuando en su esqueleto a los hidrógenos (Randić y Wilkins., 1979; Randić, 1992).

Dos grafos  $G_1$  y  $G_2$  se definen como isomórficos si existe una correspondencia unitaria entre sus conjuntos de vértices y su conjunto de enlaces. Un invariante de un grafo  $G$  es una cantidad asociada con  $G$  que tiene el mismo valor para cualquier grafo que es isomórfico con  $G$ . Dado un grafo, la valencia de un vértice  $i$  es el número de uniones que lo vinculan a otros vértices, a modo de ejemplo: si todos los vértices en un grafo regular tienen valencia 2, el grafo se vincula con un ciclo.



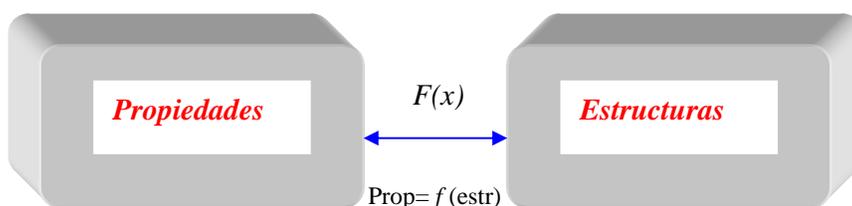
$$W = 1/2 (8 \times 1 + 8 \times 2 + 4 \times 3) = 18$$

Figura 1.7.4.1: grafo de hidrógenos suprimidos y su aplicación al cálculo del índice de Wiener (Wiener et al., 1947)

Los grafos pueden asociarse a su vez con matrices, ya sea de unión o distancia según un espacio vectorial del álgebra lineal, el índice de Wiener consiste en la semisuma de los elementos de una matriz distancia respecto a un punto dado (Wiener, 1947; Castro et al., 2002). Aplicada esta operación a cada molécula de una familia dará origen a uno de los conjuntos numéricos necesarios para el análisis matemático. El hecho de representar una molécula por un grafo y luego buscar relaciones entre los vértices y uniones de ellos es una forma de codificar la información. El uso de la teoría de grafos posee una ventaja potencial

de uso, ya que mientras los índices topológicos no son aplicables a estructuras tridimensionales, la teoría de grafos puede ser extendida para calcular los denominados descriptores 3-D, de importancia esencial en la actualidad para el modelado de actividades biológicas, por ende farmacológicas (Kubinyi et al., 1998; Fontaine et al., 2005; Fernandes et al., 2008).

Finalmente un modelo de QSPR se reduce a una expresión algebraica, funcional, que muestra la correlación entre dos conjuntos de números, a saber: un conjunto representa las propiedades macroscópicas y el otro las estructuras moleculares bajo estudio.



Del análisis matemático de las posibles funciones surge un resultado sumamente importante: a partir del conjunto ordenado de pares de datos (estructura, propiedad) y la función que los relaciona, conociendo la estructura de una molécula de la familia, pueden predecirse las propiedades macroscópicas desconocidas por simple codificación de su estructura (Randic, 1991).

#### 1.7.5-Descriptores cuánticos.

Las propiedades fisicoquímicas de una molécula y aún su comportamiento en reacciones o interacciones intermoleculares, son dependientes del contenido energético y de las características tridimensionales de su estructura molecular, forma y tamaño. El rápido avance en la tecnología de los ordenadores y el desarrollo de algoritmos eficientes, ha facilitado la aplicación de nuevos métodos denominados semi-empíricos en el área de química cuántica. Estos métodos intentan proporcionar información molecular con un coste de cálculo reducido. Los métodos semi-empíricos han dado lugar a una fuente atractiva de nuevos descriptores moleculares, los cuales pueden en principio, expresar todas las propiedades geométricas y electrónicas de las moléculas y sus interacciones. Es por ello que en trabajos recientes se emplean este tipo de descriptores, ya sea, solos o conjuntamente con los descriptores convencionales

El procedimiento implica los siguientes pasos: i) se calculan las geometrías moleculares usando un método computacional conocido, generalmente se utilizan los semi-

empíricos por la relación existente entre el tiempo de cálculo y el grado de ajuste necesario para este fin; ii) se barren las distintas posibilidades energéticas de las moléculas, obteniéndose energías del HOMO, del LUMO; iii) estas energías se utilizan como indicadores de propiedades internas de la molécula. También se suelen barrer las energías de algún intermediario moléculas frente a distintas reacciones químicas. Existen metodologías más complejas como las que necesitan de una medida de similitud cuántica molecular (MQS) definida por Carbó-Dorca y colaboradores (Carbó-Dorca et al., 1988; 1998), es útil para establecer medidas cuantitativas de semejanza entre las estructuras moleculares por medio de la proyección de sus funciones de densidad. Las matrices de similitud cuántica molecular (MQSM) están basadas en postulados mecánico-cuánticos y usan a la función de densidad de primer orden como descriptor molecular, el cual provee una representación 3D coherente (Amat et al., 1997; Amat et al., 1998; Carbó-Dorca et al., 1998).

Estos valores son usados como descriptores en estudios de QSAR (Amat et al., 1997; Amat et al., 1998; Carbó-Dorca et al., 1998). Los descriptores moleculares para cada compuesto se obtienen de la estructura molecular. Dichas estructuras son dibujadas con ayuda de un software comercial, Hyperchem, Molecular Modeling Pro 3.01 (ChemSW Software Inc.) y convertidas a estructuras 3-D cuando los cálculos así lo requieren usando el Software CAChe (Oxford Molecular Ltd.) y Spartan Pro (Wavefunction, Inc.).

#### *1.7.6-Estrategia general de trabajo en el desarrollo de un modelo QSPR/QSAR.*

En QSAR/QSPR no existe una metodología “dura” de procedimiento, si bien existen una serie de pasos con los que el operador se puede orientar. Estos no son determinantes en el cálculo y permiten que los resultados obtenidos respondan a cierto perfil, sin necesidad de ser elementales (Testa y Kier, 1964).

Se pueden citar como pasos iniciales de trabajo, la búsqueda de propiedades fisicoquímicas o actividades biológicas y la elección del descriptor que se usará para codificar la estructura, en principio la propiedad resulta ser la variable dependiente (sobre la que no se puede tomar decisiones) y la variable independiente es el descriptor, de los cuales existen en la literatura alrededor de 3000 aceptados y validados (Todeschini y Consonni, 2000). Para obtener buenos resultados deben separarse el conjunto de moléculas en dos conjuntos de datos: conjunto de trabajo y conjunto de prueba, el primero será útil para construir el modelo y el segundo para chequearlo, validarlo.

El paso siguiente consiste en encontrar una función matemática que vincule los datos del conjunto de trabajo con los descriptores, la validez del mismo queda ratificada por los valores estadísticos del  $r$ ,  $S$  y el  $F$ , a continuación se aplica la ecuación al conjunto de prueba. Esta etapa es decisiva ya que si los parámetros estadísticos clásicos  $S$ ,  $F$ ,  $r$  son aceptables, debe cumplirse, como condición excluyente, que las diferencias entre los valores predichos y los verdaderos estén dentro de un valor aceptable (en el caso de una propiedad puede ser la incertidumbre instrumental o un múltiplo de la misma), si se cumplen con estas condiciones se valida el modelo. De lo contrario hay que probar con diferentes funcionalidades

(exponentes enteros y fraccionarios, logaritmo natural o decimal del descriptor, inversa del descriptor, etc.). Si repetido el proceso los resultados no son aun satisfactorios, se opta por cambiar el descriptor, no la propiedades, ya que son el objetivo de trabajo (Randic, 1992).

En la actualidad existen diversos programas computacionales que permiten realizar regresiones lineales, potenciales, exponenciales, logarítmicas, etc.; también pueden realizarse regresiones multilineales donde un conjunto de variables métricas son correlacionadas linealmente con una misma variable dependiente sin mayor dificultad (Ecuación 1.7.6),

$$Y = \sum_{i=1}^n a_i X_i + b \quad \text{Ecuación 1.7.6.1}$$

Resultaría poco útil detenerse a describir las operaciones anteriores, más aún cuando la intervención del operador está limitada al ingreso de valores numéricos y las operaciones de cálculo están facilitadas por la capacidad computacional. No ocurre esto con metodologías matemático estadísticas con un mayor formalismo de manejo como pueden ser, la ortogonalización de variables (Randic, 1991), los análisis multivariados (Eriksson et al., 2002) y la aplicación de redes neuronales (Claycamp et al., 1999).

#### 1.7.7-Estudios por Fragmentos moleculares.

Si se elije un grupo de moléculas pertenecientes a una familia química, ellas en principio estarán conformadas por una estructura base, o un grupo funcional propio, variando en el primer caso los sustituyentes y en el segundo los esqueletos químicos que acompañan al grupo funcional. Estas variaciones repercuten en el valor de la propiedad/actividad de la molécula y en la variación regular a lo largo de la familia (Hansch y Fujita, 1964). Otra componente microscópica que influye en el valor de la magnitud observable resultan los efectos estéricos, es decir la distribución en el espacio de los átomos como consecuencia de las repulsiones electrónicas presentes y los volúmenes ocupados en el espacio en función directa con el tamaño atómico (Unger y Hansh, 1976).

A partir de un importante banco de datos integrado por un gran número de moléculas (Pomona Project) fueron Rekker y colaboradores (1977, 1979) los primeros en investigar este tipo de influencias microscópicas sobre las propiedades macroscópicas y lo aplicaron al cálculo de coeficientes de partición Octanol-Agua. A partir de datos experimentales emplearon métodos estadísticos para ponderar la contribución de fragmentos simples como C, CH-, CH<sub>2</sub>-, CH<sub>3</sub>-, OH- y NH<sub>2</sub>-, también las contribuciones de distintos tipos de esqueletos carbonados C<sub>6</sub>H<sub>5</sub>-, C<sub>6</sub>H<sub>4</sub>-, etc , como una forma de considerar los sistemas aromáticos. De esta forma puede plantearse la fórmula de Rekker's como Ecuación 1.7.7.1

$$\log P = \sum a_n f_n + \sum b_m F_m \quad \text{Ecuación 1.7.7.1}$$

Donde  $\log P$ , representa la hidrofobicidad;  $a_n$  es el número de ocurrencia del fragmento  $f$  de tipo  $n$  y  $b$  es el número de ocurrencia de correlación  $F$  de tipo  $m$ . Desde el punto de vista formal este tipo de prácticas resulta ser una metodología semienpírica, tal que a partir de medidas experimentales, se parametrizan las correlaciones con los distintos fragmentos para la evaluación del segundo término sumatoria de la expresión.

A partir de este concepto se han desarrollado mejoras considerables al incluir formalmente factores estructurales relacionados con: el tamaño (Kamlet et al, 1983; Leo y Hansch, 1976), la unión, las interacciones intraatómicas (por Ej. Un grupo amino y un alcohol en la misma molécula), los sistemas conjugados y los equilibrios tautoméricos, estas inclusiones expandió la aplicación de este tipo de modelizaciones a distintos procesos predictivos (Hansch y Leo, 1995).

### ***1.8-Aplicaciones de los modelos QSAR/QSPR/QSTR.***

Los primeros trabajos desarrollados en el campo de las relaciones propiedad estructura fueron desarrollados por Hammett en el intento de establecer relaciones entre los cambios estructurales de la molécula del ácido benzoico y su reactividad, en el año 1935; en el año 1964 Hansch y Fujita, realizan la primera aproximación en lo que se posteriormente se dio en llamar *toxicidad no específica* al trabajar sobre los coeficientes de partición octanol-agua.

Desde esta publicación a la fecha han transcurrido más de cuatro décadas, con una gran expansión en las herramientas complementarias de trabajo: nuevas estrategias en el cálculo de los descriptores moleculares, los algoritmos matemático-estadísticos (análisis multivariados (Malinowski, 1991), redes neuronales artificiales (Zupan, 1998), programas computacionales específicos y confiables y computadoras más rápidas y con mayor capacidad operativa. La disciplina se aplicó a prácticamente toda la química de síntesis (Di Santo et al., 2002; Mor et al., 2004), la bioquímica (Hansch y Leo, 1995; Benigni y Passerini, 2002; Ekins, 2007), en desarrollos tecnológicos e industriales (Durham y Pearl, 2001) y en el campo de las ciencias ambientales (Hansch y Leo, 1995, Cai et al., 2008). Se encontró en estas metodologías mucha potencialidad desde el punto de vista económico, al incluir un paso previo en el diseño de productos según la predicción de sus propiedades/actividades por vía computacional, para luego en una segunda etapa aplicar los procedimientos experimentales sólo a las moléculas con mayores posibilidades de éxito. En otros casos se aplicó la metodología QSAR/QSTR (Quantitative Structure Toxicology Relationships) para elaborar bancos de datos toxicológicos sobre un gran número de moléculas y distintas especies, que luego son utilizados a la hora de registrar nuevos productos (Ectoc, 2003)

Así mismo son de especial interés, los efectos que puedan predecirse en el caso de las mezclas de varios componentes, por los posibles efectos sinérgicos, aditivos o antagónicos que puedan ocasionarse, en este aspecto recién en los últimos 5 años han comenzado a

aparecer publicaciones científicas dedicadas a este tema, con estudios sobre mezclas binarias y ternarias (Young et al., 2002; Ekins, 2007)

### *1.8.1- Aplicaciones oficiales.*

En los ítems anteriores se han detallado temas vinculados al medio ambiente y las metodologías del cálculo computacional aplicadas a la teoría QSPR/QSAR/QSTR. El crecimiento de las problemáticas ambientales y las obligaciones reglamentarias sitúan a la química en un contexto global y ecosistémico. Los organismos oficiales de los Estados Unidos como la EPA, la FDA (Food and Drug Administration), el CDER (Center for Drug Evaluation and Research), la OPS (Office of Pharmaceutical Science), el ICSAS (Informatics and Computational Safety Analysis Staff) y europeos como la ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals) han desarrollado o aprobado la aplicación de programas computacionales con aplicación directa en la predicción de propiedades, que van desde las fisicoquímicas generales, toxicidad en distintos organismos vegetales y animales, con distintos niveles de organización, efectos sobre salud humana, aplicaciones a carcinogénesis, mutagénesis y teratogénesis. En la Tabla 1.8.1.1 se presenta un resumen de las aplicaciones desarrolladas por este tipo de programas computacionales y su capacidad predictiva (Fuente: ECETOC, 2003).

Desde el punto de vista ambiental es de especial interés el conocimiento de la velocidad de degradación y la distribución en los distintos compartimentos ambientales de un compuesto químico, para evaluar la persistencia química y las especies biológicas más afectadas en función de la biodisponibilidad (Richardson, 1996). En las tablas Tabla 1.8.1.2 y 1.8.1.3 se detallan los programas informáticos desarrollados para las aplicaciones ambientales y de amplia aceptación tanto por los organismos oficiales como por la comunidad científica, según ECETOC (2003).

**Tabla 1.8.1.1:** distintos tipos de predicciones realizadas por los programas computacionales que actualmente se comercializan. Calidad de la predicción realizada ya sea sobre una propiedad fisicoquímica o una actividad biológica.

| <b>Propiedades Fisicoquímicas</b> | <b>Predicción</b> | <b>Ecotoxicidad y degradación Ambiental</b> | <b>Predicción</b>   | <b>Salud Humana</b>          | <b>Predicción</b>       |
|-----------------------------------|-------------------|---|---------------------|------------------------------|-------------------------|
| -Punto de Fusión                  | <i>Buena</i>      | -Fotodegradación                            | <i>Limitado</i>     | -Toxicidad oral Aguda        | <i>Limitado</i>         |
| -Punto de Ebullición              | <i>Buena</i>      | -Hidrólisis                                 | <i>Limitado</i>     | -Irritación de la piel       | <i>Limitado</i>         |
| -Presión de Vapor                 | <i>Muy Buena</i>  | -Biodegradación                             | <i>Buena</i>        | -Irritación de la vista      | <i>Limitado</i>         |
| -Coeficientes de Partición        | <i>Muy Buena</i>  | -Bioacumulación                             | <i>Buena</i>        | -Toxicidad mamaria crónica   | <i>Muy Limitado</i>     |
| -Solubilidad en Agua              | <i>Muy Buena</i>  | -Toxicidad aguda en peces                   | <i>Buena</i>        | -Mutagénesis <i>in vitro</i> | <i>Limitado a bueno</i> |
|                                   |                   | -Toxicidad aguda en invertebrados           | <i>Limitado</i>     | -Carcinogénesis              | <i>Muy Limitado</i>     |
|                                   |                   | -Toxicidad aguda para algas                 | <i>Limitado</i>     | -Teratogénesis               | <i>Muy Limitado</i>     |
|                                   |                   | -Toxicidad aguda para bacterias             | <i>Limitado</i>     | -Sensibilización de a piel   | <i>Limitado</i>         |
|                                   |                   | -Toxicidad Crónica                          | <i>Muy Limitado</i> |                              |                         |

**Tabla 1.8.1.2:** Propiedades fisicoquímicas predichas por distintos programas computacionales existentes en el mercado.

|                     | PROGRAMAS ACEPTADOS |       |       |     |         |
|---------------------|---------------------|-------|-------|-----|---------|
|                     | EPIWIN              | ASTER | SPARC | ACD | PREDICT |
| Punto de Fusión     | X                   |       |       |     |         |
| Punto de Ebullición | X                   | X     | X     | X   | X       |
| Densidad            |                     |       | X     | X   | X       |
| Presión de Vapor    | X                   | X     | X     |     | X       |
| Solubilidad en Agua | X                   | X     | X     | X   |         |
| Kow                 | X                   | X     | X     | X   |         |
| Constante de Henry  | X                   | X     | X     |     |         |
| Ka                  |                     | X     | X     | X   |         |
| Tensión Superficial |                     |       |       | X   | X       |

**Tabla 1.8.1.3:** Propiedades ecotoxicológicas predichas por distintos programas computacionales existentes en el mercado

| PROGRAMAS ACEPTADOS           |     |           |        |        |       |               |
|-------------------------------|-----|-----------|--------|--------|-------|---------------|
|                               | SRC | MultiCASE | TOPKAT | ECOSAR | ASTER | CATABOL-OASIS |
| DESTINO y DISTRIBUCIÓN        |     |           |        |        |       |               |
| Degradación Abiótica          |     |           |        |        |       |               |
| *Fotólisis                    |     |           |        |        |       |               |
| *Hidrólisis                   | X   |           |        |        | X     |               |
| *Oxidación Atmosférica        | X   |           |        |        |       |               |
| Degradación Biótica           |     |           |        |        |       |               |
| *Aeróbica                     | X   | X         | X      |        | X     | X             |
| *Anaeróbica                   |     | X         |        |        |       |               |
| Sorción                       | X   |           |        |        | X     |               |
| Bioacumulación                |     |           |        |        |       |               |
| Bioconcentración              | X   |           |        |        | X     |               |
| TOXICIDAD : Especies/Ambiente |     |           |        |        |       |               |
| Peces                         |     |           |        |        |       |               |
| *Aguda                        | X   | X         | X      | X      | X     | X             |
| *Crónica                      |     |           |        | X      |       |               |
| Dafnia                        |     |           |        |        |       |               |
| *Aguda                        |     |           | X      | X      |       |               |
| *Crónica                      |     |           |        | X      |       |               |
| Algas                         |     |           |        |        |       |               |
| *Aguda/Crónica                |     |           |        | X      |       |               |
| Terrestres                    |     |           |        |        |       |               |
|                               |     |           |        | X      |       |               |
| Marinos                       |     |           |        |        |       |               |
|                               |     |           |        | X      |       |               |

La EPA ha desarrollado y difundido un programa que abarca tanto los aspectos fisicoquímicos como ecotoxicológicos: EPI Swite (Version 4.0, actualizado Diciembre 2008), que resulta ser un ejecutor de toda una serie de módulos *ad on`s*, entre los cuales se encuentra el ECOSAR y el EPIWIN citados anteriormente, junto a otros auxiliares para el cálculo de todas las propiedades citadas en las tablas anteriores. El Software dispone de una ventaja importante, como el cargado manual de propiedades fisicoquímicas medidas en el laboratorio del compuesto para realizar optimizaciones de correlación y los valores predichos resulten ser más consistentes, ya que de alguna manera se acotan los parámetros de estimación por parte del programa y los resultados son parametrizados en función de las aobservaciones experimentales. Una ventana del programa se muestra en la figura 1.8.1.

A la hora de evaluar el estado de un ambiente es necesario poder describir sus componentes y las relaciones existentes entre ellos, tanto desde la química como desde la biología (Manahan, 2007). Esto incluye un relevamiento detallado de la flora y la fauna del lugar, las características del suelo, el agua y el aire, la presencia/ausencia de posibles

sustancias contaminantes, la acción de las mismas sobre organismos blancos y su influencia sobre la cadena trófica (Levin et al., 1989).

The Estimation Programs Interface (EPI) Suite™ was developed by the US Environmental Protection Agency's Office of Pollution Prevention and Toxics and Syracuse Research Corporation (SRC). It is a screening-level tool, intended for use in applications such as to quickly screen chemicals for release potential and "bin" chemicals by priority for future work. Estimated values should not be used when experimental (measured) values are available.

EPI Suite™ cannot be used for all chemical substances. The intended application domain is organic chemicals. Inorganic and organometallic chemicals generally are outside the domain.

Important information on the performance, development and application of EPI Suite™ and the individual programs within it can be found under the Help tab. Copyright 2000-2008 United States Environmental Protection Agency for EPI Suite™ and all component programs except BioHCwin and KOWWIN.

Figura 1.8.1: Ventana de manejo de los módulos ad on en el programa EPI Swite para el ingreso de la molécula sobre la que se quiere saber sus propiedades químicas y toxicológicas.

### Metodologías integradas para el relevamiento ambiental.

#### 1.9-Objetivos e Hipótesis Generales de Trabajo.

En cada capítulo se presentarán los distintos objetivos específicos y las hipótesis que sustentan el trabajo. En términos generales todo el trabajo de la siguiente tesis persigue el objetivo general de “*aplicar en forma integrada metodologías experimentales y herramientas derivadas de la química teórico-computacional a compuestos orgánicos de relevancia ambiental, para interpretar su comportamiento fisicoquímico y las posibles respuestas biológicas derivadas de su presencia en el ambiente.*”

Basado en la hipótesis fundamental de que “**el análisis multivariado de datos, conjuntamente con la modelización molecular, aplicados a observaciones experimentales, contribuyen al desarrollo u optimización de metodologías para el monitoreo y evaluación de impacto ambiental de contaminantes químicos.**”