

CAPÍTULO I

Objetivos del proyecto

Introducción a la “Cambridge Structural Database”

1.1. OBJETIVOS DEL PROYECTO

Este proyecto se enmarca dentro del estudio de las propiedades supramoleculares de complejos de coordinación y organometálicos con geometría plano-cuadrada. Pretende realizar un análisis estadístico del carácter iónico de los citados complejos y establecer los contraiones más frecuentes utilizados en su cristalización.

Por último pretende establecer las características geométricas de la interacción complejo-contraión como paso previo para un estudio más detallado de dicha interacción.

Los objetivos principales del proyecto son los siguientes:

- I. Revisión bibliográfica sobre el origen, la estructura y las aplicaciones de la *Cambridge Structural Database* (CSD)
- II. Selección de complejos de coordinación homodinucleares con geometría plano-cuadrada contenidos en la CSD y clasificación en función del átomo metálico y del tipo de átomo puente
- III. Clasificación de los complejos más representativos obtenidos en el apartado II en función de su carga eléctrica como neutros, aniónicos y catiónicos
- IV. Determinación de los cationes y aniones más frecuentes utilizados para cristalizar los complejos con carga eléctrica seleccionados en el apartado anterior
- V. Caracterización mediante parámetros geométricos seleccionados (distancias y ángulos de enlace, ángulos de torsión, etc.) de la conformación de alguno de los grupos de complejos cargados más representativos así como de su orientación relativa respecto al contraión correspondiente

1.2. EL CENTRO DE DATOS CRISTALOGRÁFICOS DE CAMBRIDGE

Desde su descubrimiento a principios del siglo XX, una técnica analítica conocida como “cristalografía de rayos X” ha estado descubriéndonos la naturaleza del estado sólido de la materia a escala atómica. Las series ordenadas de átomos y moléculas que componen un cristal, dispersan los haces individuales de rayos X originando miles de haces de diferentes intensidades. Estas intensidades y las posiciones de los haces dispersados dependen de las posiciones relativas de todos los átomos dentro del cristal. Así, midiéndolas con precisión y usando las técnicas matemáticas adecuadas, pueden obtenerse las posiciones atómicas y, a partir de ellas, las formas moleculares dentro del cristal. Con los adelantos que se han producido en computación y en otras tecnologías a lo largo de los años, las técnicas cristalográficas se han ido volviendo cada vez más rutinarias, hasta el punto de que la determinación de la estructura de una molécula relativamente simple, que hace 20 años podía tardar una semana o incluso más, ahora puede lograrse en cuestión de horas.

En los años sesenta, época en la que la obtención de la estructura de una pequeña molécula era rara y dificultosa, en todo el mundo se determinaban unos cuantos cientos de estructuras al año. A finales de los 60 seguían publicándose sólo unos pocos cientos de estructuras al año, y las aproximadamente 2000 estructuras publicadas antes de 1965 (Bruno *y otros*, 2002) fueron incorporadas gradualmente a la base de datos en desarrollo, usando a menudo compendios impresos así como los volúmenes de Informes Estructurales de la IUCr (“International Union of Crystallography”) para localizar las referencias bibliográficas originales. Nace aquí la “Cambridge Structural Database” (CSD).

Pronto se desarrolló un software centrado en los sistemas de validación y almacenamiento de la información acumulada (véase por ejemplo, Allen *y otros*, 1979). Sin embargo, los sistemas de búsqueda, recuperación, análisis numérico y visualización de la base de datos en cuestión, se convirtieron en una de las primeras necesidades, y comenzaron a desarrollarse a finales de los 70 (Allen *y otros*, 1979), reforzándose considerablemente durante los 80 (Allen *y otros*, 1991), para incluir las subestructuras bidimensionales completas e incompletas que mostraron su capacidad y habilidad para localizar contactos intermoleculares sin unión (Allen & Kennard, 1993). En la actualidad, la base de datos, como parte de un sistema de información mayor, continúa reforzándose

para dar respuesta a las necesidades del usuario; los últimos desarrollos en software son descritos por *Bruno y otros* en 2002.

Un amplio espectro de científicos que necesitan acceder a las estructuras cristalinas individuales, o mas comúnmente, que quieren información sistemática sobre estructuras tridimensionales con una resolución atómica, comienzan a usar el sistema de la CSD. Desde finales de los 70, el CCDC comenzó a llevar a cabo una investigación básica usando extensivamente los datos acumulados en la CSD (Bürgi y Dunitz, 1983; Allen *y otros*, 1983; Bürgi y Dunitz, 1994), junto con las herramientas de software auxiliares de las que se empezaba a disponer. De esta manera, durante las dos últimas décadas, el sistema de la CSD ha proporcionado la base esencial para que el CCDC pudiera realizar algunos de los primeros experimentos en el diseño racional de nuevas moléculas bioactivas de interés farmacéutico o agroquímico, o pudiera aplicar a la CSD una gran variedad de metodologías estadísticas, numéricas y computacionales (Taylor y Allen, 1994), dando lugar a un área para la que se han empleado términos como “correlación de estructura” o “adquisición de conocimiento” y el semánticamente cuestionable término “minería de datos”. (Allen *y otros*, 1991)

Esta gran actividad ha planteado a su vez, la posibilidad de generar bibliotecas de información estructural procedente de la CSD basadas en el conocimiento. *IsoStar* (Bruno *y otros*, 1997), una biblioteca de información sobre interacciones sin enlace dada a conocer en 1997, es el primer producto del CCDC basado en el *conocimiento* anterior. En 2004 el CCDC presentó *Mogul*, una base de conocimiento de geometría molecular que, como el anterior, forma parte del actual sistema de la CSD. Recientemente se ha comenzado a llevar a cabo también una investigación fundamental sobre proyectos que han surgido del desarrollo y uso de las bases de conocimiento y de las aplicaciones del software de la CSD. Trabajo éste que se ha dirigido hacia las siguientes áreas:

- Dimensiones moleculares medias
- Estudios de los efectos de los sustituyentes
- Técnicas estadísticas y numéricas de análisis de datos
- Correlación de la estructura y los pasos de reacción
- Geometría del enlace de hidrógeno y direccional
- Enlaces de hidrógeno débiles

- Interacciones sin enlace en las que no media el hidrógeno
- Ingeniería cristalina
- Simetría cristalográfica y simetría molecular

Las referencias a las contribuciones científicas publicadas por los autores del CCDC en éstas y en otras áreas se incluyen en el registro de páginas *CCDC Publications*. Las referencias de los artículos escritos por el personal del CCDC desde 1973, tanto en exclusiva como en colaboración con otros autores no pertenecientes a dicho centro, se incluyen también si son relevantes, en la base de datos *CCDC WebCite* (anteriormente *DBUse*). Esta base de datos de citas que se distribuye como parte del sistema de la CSD desde 1995, cubre las aplicaciones de investigación publicadas de dicho sistema y de otros productos del CCDC. Incluye referencias de artículos para los que el uso sistemático de la CSD ha contribuido sustancialmente a los resultados publicados, llevando un registro de las referencias bibliográficas, el cual incluye el título completo del artículo junto con una corta sinopsis de la investigación llevada a cabo.

La CSD fue una de las primeras bases de datos científicas que empezó a funcionar a nivel mundial, y comenzó a contar con importantes concesiones académicas recibidas a través del Gobierno Británico para poder desarrollar la información científica y técnica. Estos fondos, junto con subvenciones de centros afiliados nacionales, permitieron el desarrollo de la CSD y de su software asociado durante los años 70 y los años 80. Los primeros lanzamientos del sistema de la CSD a Estados Unidos, Italia y Japón se produjeron a principios de los años 70. A principios de los 80, el sistema de la CSD era distribuido en más de 30 países por todo el mundo. Ahora, con la entrada del nuevo milenio, el sistema de la CSD está disponible en más de 55 países.

Durante los años 80, el interés en esta gran base de datos de las compañías farmacéuticas y agrícolas creció considerablemente. Así, la corriente económica favorable que se dio por toda Europa, permitió al CCDC convertirse en una compañía independiente en 1989, con la personalidad jurídica de una institución no lucrativa con sus operaciones supervisadas por un consejo superior internacional. El CCDC instaló su propia sede en el Departamento de Química de la Universidad de Cambridge en 1992, conservando así los lazos de unión con ésta.

Históricamente, el CCDC se originó en el Departamento de Química de la Universidad de Cambridge, y ahora posee su propio edificio en el Campus de Química en Cambridge. Aunque es una organización que se autofinancia y se administra a sí misma, conserva acuerdos con la universidad y se reconoce como institución que pretende enseñar a los estudiantes de ramas técnicas y científicas para acceder a grados más altos de conocimiento.

La entidad creció fuera de las actividades del grupo de la cristalografía, conducido por la doctora Olga Kennard en el entonces Departamento de Química Orgánica, Inorgánica y Teórica de la Universidad de Cambridge. A partir de 1965, el grupo comenzó a recoger los datos bibliográficos, químicos y cristalinos publicados de todas las moléculas estudiadas por difracción de rayos X y neutrones. Con el paso del tiempo, y gracias a los avances tecnológicos, esta *colección* fue codificada en forma electrónica y pasó a conocerse como la *Base de Datos Estructural de Cambridge* (CSD).

La doctora Kennard se retiró de la dirección del CCDC en 1997, y fue sustituida por el doctor David Hartley, que ocupó el cargo de Director Ejecutivo desde 1997 hasta 2002. En octubre de 2002 se nombró al actual Director Ejecutivo, el doctor Frank Allen. El CCDC emplea a 50 personas a jornada completa (16 en el grupo de la base de datos, 12 en el desarrollo del producto de software, 5 en investigación, 6 en ayuda técnica, 3 en ayuda y comercialización científica y 8 en negocio y administración). Su equipo directivo está constituido actualmente por:

Frank Allen *Director ejecutivo*

Robin Taylor *Director de investigación*

Steve Salisbury *Director comercial*

Owen Johnson *Encargado del grupo de investigación*

Sam Motherwell *Encargado de la investigación*

Acoge también a los científicos y estudiantes que visitan ocasionalmente las instalaciones, actuando de anfitriones en los trabajos y proyectos relacionados con la base de datos, y es además uno de los socios del “*Instituto Pfizer para la Ciencia de los Productos*

Farmacéuticos”, junto con los departamentos de Química y Ciencias Materiales de la universidad. El centro organiza visitas científicas y mantiene también estrechos vínculos de colaboración con las universidades y con otras muchas instituciones académicas e industriales, tanto en el Reino Unido como en el resto del mundo.

Esta institución, que hoy en día posee una gran solidez y que, como ya se ha dicho opera sin ánimo de lucro, mantiene la condición social de Organización Benéfica registrada en el Reino Unido, es financieramente autosuficiente y genera su renta de:

- Suscripciones de la academia e industria para mantener el sistema estructural de la base de datos de Cambridge
- Acuerdos de ventas relacionados con el mantenimiento y venta de otros productos de software en las ramas científicas, tecnológicas y de la cristalografía
- Concesiones de la industria y de otras fuentes para apoyar investigaciones de software y proyectos de investigación específicos

Así, más de 600 lugares académicos en más de 40 países reciben copias de la CSD y de su software asociado en CD-ROM dos veces al año. La mayor parte de los ingresos del CCDC, sin embargo, procede más bien de los más de 100 lugares industriales que pagan suscripciones anuales para poder acceder a la CSD.

Durante la década de los 90, los productos de software del CCDC se diversificaron para poder usar al máximo los datos cristalográficos en las aplicaciones de las ciencias de la vida y en la cristalografía. En 1998 se fundó la “*Sociedad Limitada de Software del CCDC*”, una filial de éste que desde entonces se ha encargado de desarrollar la mayor parte del software que requiere la institución, y que se compromete a aportar todos sus beneficios a ésta.

El CCDC mantiene excelentes relaciones con bastantes editores de revistas científicas, y muchos de ellos insisten a los autores que depositen los datos de sus estructuras en la CSD. Como consecuencia de ello, casi todas las estructuras cristalinas orgánicas u organometálicas publicadas a nivel mundial acaban en esta base de datos.

La cristalografía estructural ha cambiado irreconociblemente desde mediados de los 60. Los espectaculares aumentos de potencia en los ordenadores han traído como consecuencia mejoras en la recogida de datos, en la resolución de estructuras y en las técnicas de refinamiento. Así mismo, se ha producido un creciente reconocimiento de la

importancia de la información cristalográfica, por lo que en el año 2001 se archivaron en la CSD más de 24.500 estructuras, lo que representó un aumento cercano a 40 veces la productividad cristalográfica mundial respecto a 1965.

El desorden en las estructuras siempre ha provocado problemas dentro de la CSD. Hasta finales de los 80, la política del CCDC en estos casos era la de guardar sólo algunas coordenadas, llegando incluso a no guardar ninguna en los casos en los que el desorden era muy complejo y afectaba a todos los átomos de una molécula. En la actualidad el software está permitiendo mejorar el análisis y la representación de las estructuras desordenadas. De hecho, el trabajo continuado para mejorar la información contenida en la propia CSD está estrechamente relacionado con la creciente disponibilidad de datos estructurales mediante el formato electrónico CIF (“*Crystallographic Information File*”) desarrollado a principios de los 90. Dicho formato fue impulsado por el *Acta Crystallographica*, para la que la dependencia electrónica basada en el formato CIF (Hall *y otros*, 1991) pasó rápidamente de ser “preferible” a “obligatoria”. Así, la mayor parte de las revistas especializadas, “recomiendan”, “instan” o “requieren” de forma suplementaria que los datos sean presentados electrónicamente en éste formato. Atrás quedan por tanto los días en los que los editores científicos del CCDC necesitaban teclear en largas listas de coordenadas atómicas la información de las estructuras cristalinas a partir de manuscritos impresos, o peor todavía, a partir de documentos depuestos por una calidad variable y a veces organizados de cualquier manera (Bergerhoff *y otros*, 1986).

1.3. **LA “CAMBRIDGE STRUCTURAL DATABASE”: la mayor base de datos cristalográficos del mundo**

La CSD es un archivo totalmente retrospectivo, y casi toda la información que contiene se ha extraído de bibliografía publicada. Se citan más de 800 fuentes bibliográficas primarias, siendo las primeras referencias de 1930. Según F.H.Allen (2002), la CSD contenía a finales de 2002 más de un cuarto de millón de estructuras de moléculas pequeñas. Los editores de la CSD se han esforzado para resumir toda la información química disponible sobre las estructuras, incluyendo los métodos de obtención de los datos, el proceso y la validación. Muchos datos se han recodificado a partir de la bibliografía original y de documentos de deposición suplementarios. Sin embargo, hoy en día, el 80%

de los nuevos datos estructurales se reciben directamente en formato CIF (cifra que va en aumento), formato que al ser normalizado por la Unión Internacional de Cristalografía, ha catalizado el cambio a los esquemas de depósito electrónico. El CCDC actúa como depositario oficial de datos de estructuras cristalinas de las 51 principales revistas especializadas. Mantiene además un archivo de prepublicación de los CIF's depositados previamente a su presentación a una revista especializada con la que el CCDC tenga un acuerdo formal, y un archivo de postpublicación así como un archivo de distribución de la CSD que almacena los datos en formato binario *ASER*, que es exclusivo del CCDC. El completo y creciente archivo *ASER* contiene los datos en forma secuencial, y es suministrado dos veces al año en CD-ROM a todos los subscriptores de la CSD.

Como se indicó al principio, la CSD es la utilidad principal del CCDC, que también abarca el software para el acceso a base de datos, el análisis de los datos de las estructuras, así como su visualización y las bases de conocimiento estructurales derivadas del CSD. El 1 de enero del presente año, el CCDC hizo público un informe con motivo de su 40º aniversario, en el que se resumían un total de 335.276 estructuras cristalinas, entre las que se encontraban 303.733 complejos diferentes.

En ella se registra la información bibliográfica, química y cristalográfica para:

- moléculas orgánicas
- compuestos organometálicos

Para determinar las estructuras tridimensionales, los diferentes autores han utilizado:

- **difracción por rayos X**.- al incidir un haz de rayos **X** sobre un átomo, esta radiación interacciona solamente con los electrones del átomo, no interacciona con el núcleo. Ahora bien, mientras más electrones tenga el átomo, mayor será la intensidad de la interacción. Para átomos de muy pocos electrones, la radiación de rayos **X** casi no experimenta interacción alguna. Así, por ejemplo, si la muestra contiene átomos de hidrógeno, que tienen solamente un electrón alrededor de su núcleo, prácticamente no "siente" a los rayos **X**. En consecuencia, usar los rayos **X** para determinar estructuras de sustancias que contengan muchos átomos de hidrógeno es muy difícil porque la intensidad de los haces difractados resulta ser muy baja. Sin embargo, determinar estructuras que contengan átomos de muchos

electrones, por ejemplo cloro, es relativamente fácil porque la interacción entre los rayos **X** y el átomo es muy fuerte, y el patrón de difracción correspondiente es muy intenso.

- **difracción por neutrones.**- al incidir un neutrón sobre un átomo, en virtud de su neutralidad eléctrica no interacciona con los electrones. Debido a ello, el neutrón puede penetrar con toda facilidad dentro del átomo, cruzar su nube electrónica y llegar al núcleo. Cuando el neutrón se encuentra a distancias muy pequeñas del núcleo se provoca una reacción entre el núcleo del átomo y el neutrón, interacción que es de origen nuclear. En consecuencia, un neutrón solamente interacciona con el núcleo de los átomos. La intensidad de la interacción depende del tipo de núcleo; de hecho esta interacción depende de cuántos protones y neutrones se encuentren dentro del núcleo. Así resulta que, por ejemplo, al incidir un neutrón sobre un núcleo de hidrógeno, que contiene solamente un protón, interacciona muy intensamente. Esto significa que el neutrón "siente" fuertemente al hidrógeno y puede dar información acerca de sus características dentro del cristal. En general, la interacción que ocurra entre el neutrón y el núcleo dependerá de qué isótopo se trate. Distintos isótopos de un mismo átomo interaccionan de maneras diferentes con un neutrón.

Esta base de datos registra resultados de:

- estudios cristalográficos
- estudios de difracción del polvo

En algunos casos el CCDC no puede obtener datos fiables o completos, y las entradas que así se encuentran están archivadas también en el CSD, aunque cita expresamente que esas estructuras podrían no ser del todo correctas.

El CSD incluye los datos de la estructura cristalina que se presentan de:

- publicaciones abiertas
- comunicaciones privadas al CSD (previa deposición directa de los datos)

Y no almacena:

- Polipéptidos y polisacáridos que tienen más de 24 unidades. Éstos se registran en el banco de datos de proteínas: <http://www.rcsb.org/pdb/>
- Oligonucleótidos. Éstos se almacenan en el banco de datos de los ácidos nucleicos: <http://ndbserver.rutgers.edu/>
- Estructuras inorgánicas, que se almacenan en la base de datos ICSI: <http://www.fiz-informationsdienste.de/en/DB/icsd/>
- Metales y aleaciones, que se almacenan en CRYSTMET®: <http://www.tothcanada.com>

1.4. **PROGRAMAS AUXILIARES DE LA CSD: *ConQuest*, *Mercury*, *Vista*, *PreQuest* y *Mogul***

Durante el período 1965-1975, el desarrollo del software del CCDC se centró en programas para la organización, validación y clasificación de las entradas de la CSD. También se desarrolló software para transformar el texto bibliográfico y químico introducido en formato de guía de referencia cada año en la CSD, en *series de Dimensiones y Estructuras Moleculares*.

Desde 1975 y en adelante, se llevaron a cabo también esfuerzos encaminados a generar un software de acceso a la CSD, con programas de búsqueda, recuperación, análisis y visualización de la información de la CSD. Originalmente, dos programas, *Bibser* y *Connser*, localizaban las entradas correspondientes a partir de los archivos bibliográficos y de enlazabilidad, que estaban separados, y las funciones geométricas y de visualización de *Geom78* y *Pluto78* usaban también archivos de datos de la CSD separados. *Geom78* y su sucesor *GSTAT* estaban entre los primeros programas que proporcionaban servicios de búsqueda para estructuras moleculares en 3D, y con toda certeza fueron los primeros en proporcionar servicios de búsqueda para interacciones moleculares.

Durante los años 80, los archivos separados se fusionaron en un solo fichero de archivos, *ASER* (mencionado anteriormente), y un solo programa, *QUEST* (más tarde, *QUEST3D*, y actualmente *ConQuest*), reunió a todos los servicios de búsqueda y de

recuperación de geometrías moleculares de la CSD. La información recuperada se visualizaba usando el programa *Pluto* (reemplazado en la actualidad por *Mercury*), y el análisis de los datos geométricos lo proporcionaba *VISTA* (que aun se usa).

La CSD constituye desde hace mucho tiempo una valiosa herramienta de investigación en cristalografía, en química estructural y en el diseño de fármacos (Allen *y otros*, 1990) como confirma el gran número de estudios publicados en los que ha jugado un importante papel (Allen y Motherwell, 2002; Orpen, 2002; Taylor, 2002) y la alta tasa de citas de algunos de estos estudios (Redman *y otros*, 2001). Sin embargo, a mediados de los años 90 quedó claro que el software para investigar y visualizar la base de datos necesitaba una importante revisión. Aunque la funcionalidad de los programas era buena, éstos se habían quedado atrás en cuanto a las normas contemporáneas de usabilidad y no eran operativos en el soporte Windows, cuyo dominio iba aumentando. Estas dos consideraciones obligaron a realizar un esfuerzo para proporcionar programas mejorados a los usuarios de la CSD.

De entrada, se puso especial énfasis en el uso de interfaces gráficas de ayuda al usuario (*GUIs*) en lugar de aumentar la funcionalidad de los programas. Las interfaces intuitivas son especialmente importantes para los programas de la CSD, debido a su naturaleza *usuario-base*. Aunque la CSD tiene un grupo de usuarios estable, muchos otros son usuarios ocasionales y acceden a la base de datos eventualmente, por lo que no poseen la especialización suficiente y precisan de un acceso fácil

Actualmente el sistema de la CSD se distribuye como un paquete que comprende la base de datos en sí misma, el software que la acompaña (con los programas *ConQuest*, *PreQuest*, *Mercury* y *VISTA*), y las bases de datos de conocimiento *IsoStar* y *Mogul* (T. V. Davis, M. S. Zaver, M. Zimmer, 2002). A continuación se expone lo más significativo de cada uno de estos componentes:

- **recuperación de la búsqueda y de datos (*ConQuest*).**- es el programa base para buscar y recuperar la información de la CSD. *ConQuest* proporciona una amplia gama de opciones para la búsqueda en la base de datos, además de una funcionalidad más compleja aunque intuitiva de la búsqueda.

Contiene, entre otras opciones:

- panel de dibujo intuitivo para dibujar subestructuras químicas y definir parámetros geométricos
- copia y pega estructuras de otros programas como *ChemDraw* e *ISIS/Draw*
- mecanismos directos para combinar preguntas y generar búsquedas complejas
- vistazo previo interactivo de los resultados de la búsqueda mientras que las búsquedas están funcionando
- extensa gama de formatos de archivo (CIF, PDB, MOL2, SHELX, pdf, SDfile etc.)
- su interface se acopla perfectamente con la del visualizador gráfico *Mercury*
- documentación completa de cada estructura

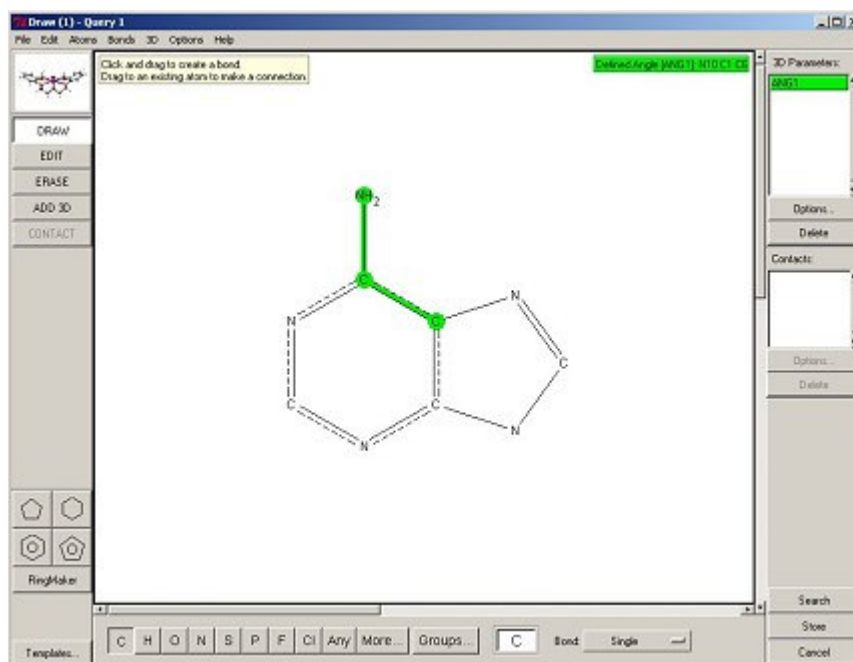


Figura 1.1. La ventana principal (“DRAW”) de *ConQuest* se utiliza para dibujar las subestructuras químicas y definir parámetros geométricos

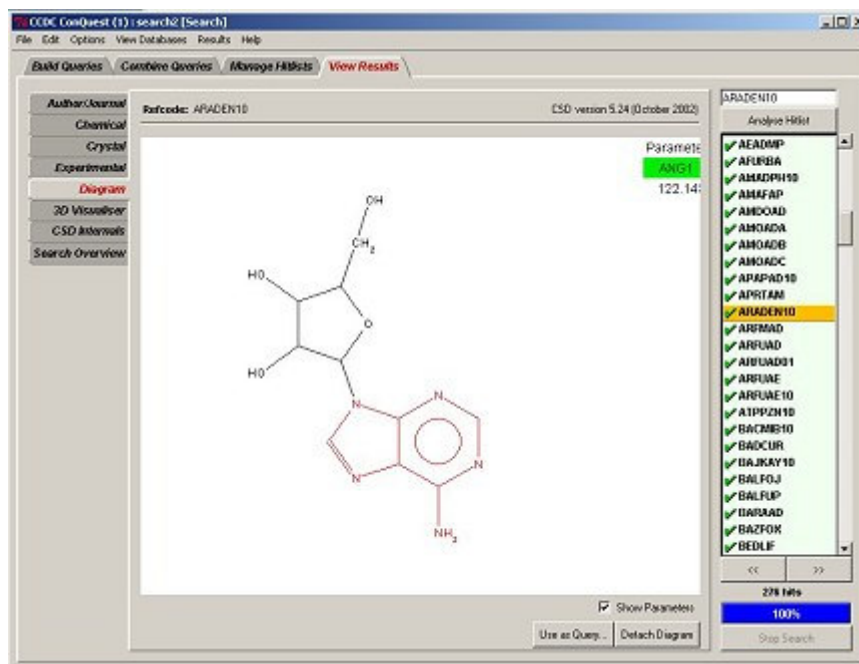


Figura 1.2. Los resultados obtenidos en *ConQuest* se muestran en la columna de la derecha, destacando en color naranja la que se ha seleccionado. A la izquierda, panel con toda la información del complejo seleccionado

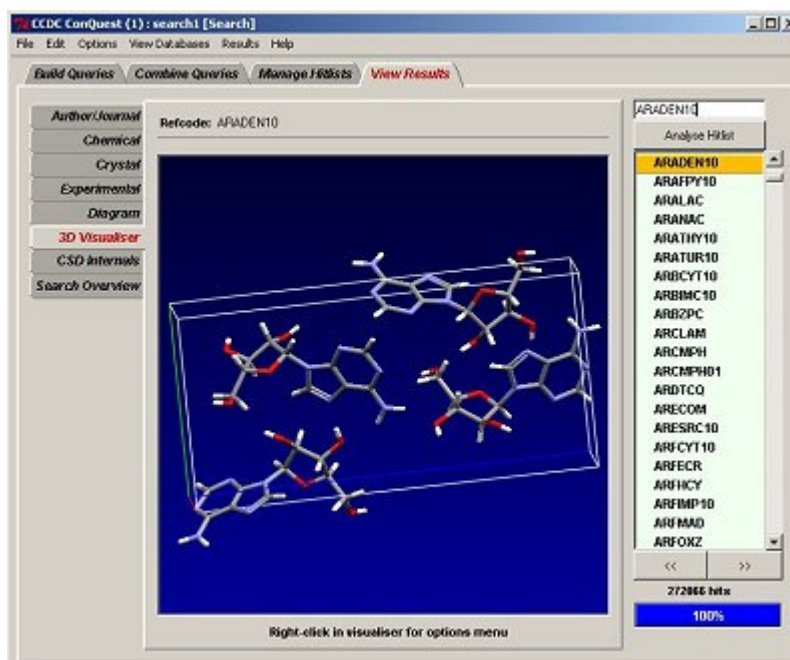


Figura 1.3. *ConQuest* ofrece también distintas posibilidades de visionado de la estructura seleccionada, aparte de los típicos 2D y 3D

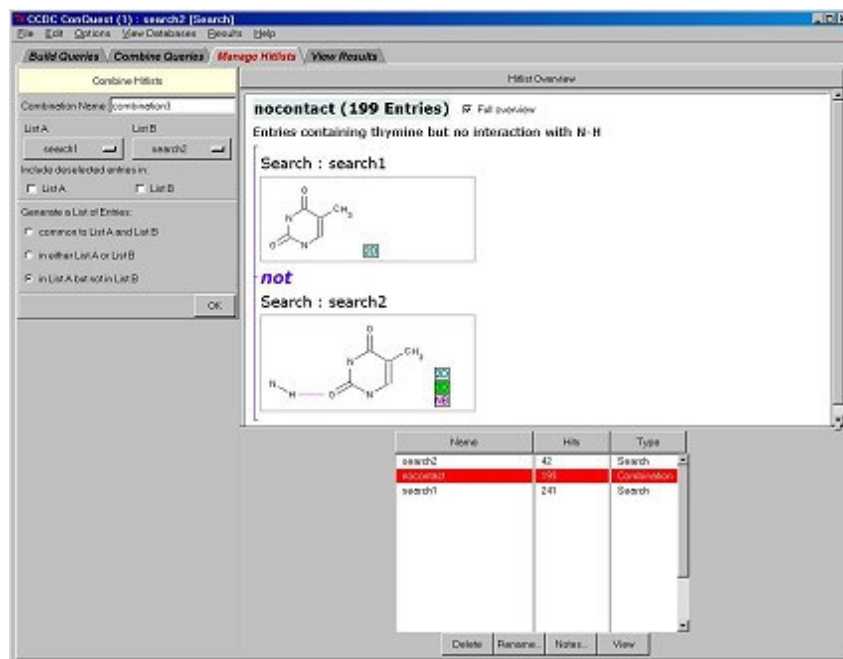


Figura 1.4. *ConQuest* permite combinar y seleccionar búsquedas independientes de forma que el resultado final obtenido será la combinación de ambas

- **visualización de la estructura (*Mercury*).**- ofrece gran cantidad de herramientas para la visualización de la estructura y la exploración de las celdas cristalográficas. Sus características incluyen:
 - Una gama completa de los estilos de visionado de la estructura
 - Herramientas de medición y visualización de distancias, ángulos y ángulos de torsión, centroides y planos
 - Creación y visionado de centroides y planos
 - Visualización de cualquier perspectiva de las celdas unidad, así como acercamiento a las mismas
 - Localización y visionado de los enlaces intermoleculares e intramoleculares
 - Construcción y visualización de una red de enlaces intermoleculares
 - Información adicional sobre la estructura durante el visionado 3D, tal como el diagrama químico (si está disponible) y las coordenadas atómicas
 - Cálculo y visionado del patrón de difracción del polvo para la estructura
 - Visionado simultáneo de estructuras

- Manejo de las estructuras incluyendo la adición de los átomos del hidrógeno, la adición de enlaces, corrección de tipos en enlace y eliminación de átomos, de enlaces y de moléculas

Mercury se distribuye como una parte más del sistema de la CSD, pero también está disponible por separado y gratuitamente en la web del CCDC.

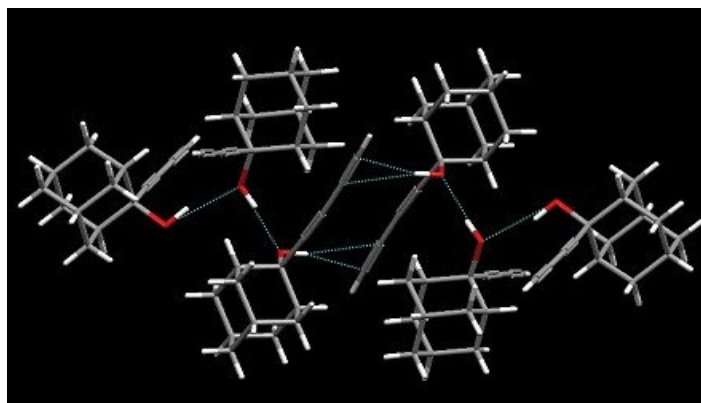


Figura 1.5. Dos unidades asimétricas de la entrada con *Refcode* NAMLAO en la CSD. Obsérvese cómo se enlazan las tres moléculas independientes

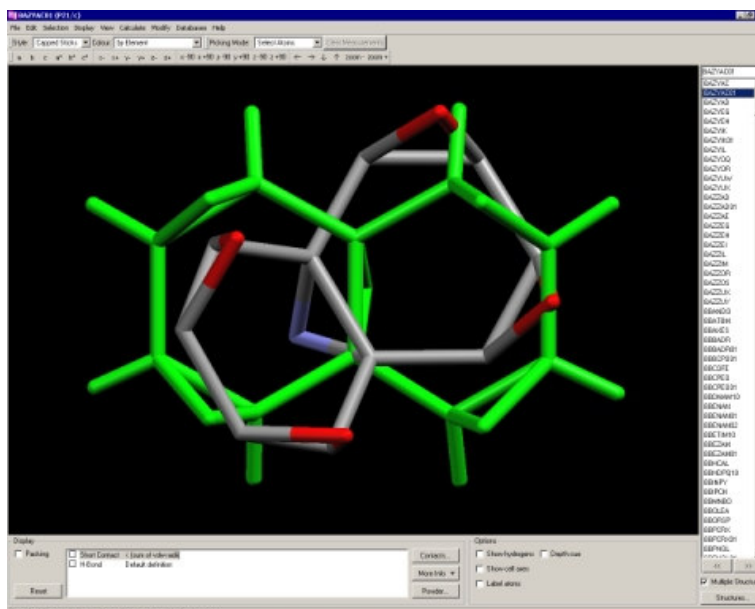
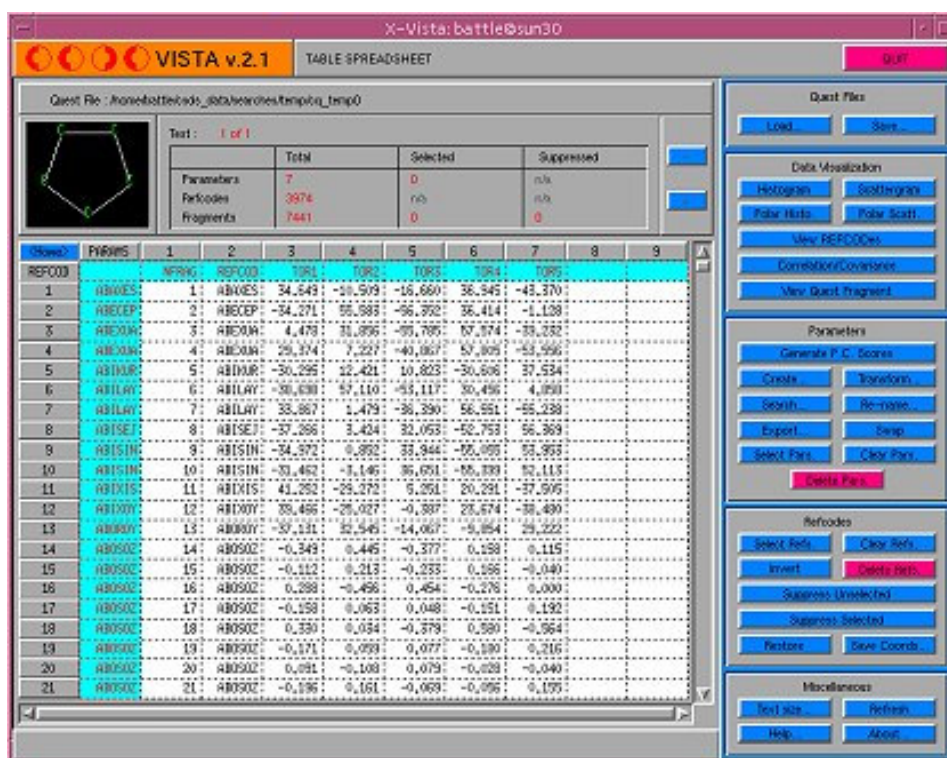


Figura 1.6. *Refcode* BAZYAC de la CSD en *Mercury*

- **análisis numérico (VISTA).**- *VISTA* es un programa analítico y estadístico interactivo. Procesa los archivos generados automáticamente por *ConQuest* cuando los parámetros 3D se han definido para una búsqueda de la CSD (o de una base de datos interna). Los datos obtenidos se pueden visualizar directamente en *VISTA*, que muestra la geometría molecular y otros parámetros referentes a una subestructura molecular o supramolecular.

VISTA realiza gran variedad de funciones analíticas y de visualización, incluyendo la generación de:

- Estadística descriptiva simple para las distribuciones del parámetro
- Histogramas y diagramas polares
- Análisis estadísticos, incluyendo análisis de la regresión lineal y de componentes principales
- *Hyperlinking* de datos geométricos a la entrada original de CSD
- Preparación de los diagramas para los informes y las publicaciones
- Acceso directo a *ConQuest*



The screenshot shows the VISTA v.2.1 software interface. At the top, it displays 'X-Vista: battle@sun30' and 'VISTA v.2.1 TABLE SPREADSHEET'. Below the title bar, there is a 'Quest File' section with a small molecular structure diagram and a 'Text: 1 of 1' field. A summary table shows: Parameters: 7, Refcodes: 3974, Fragments: 7441, Selected: 0, Suppressed: 0. The main data table has columns: REFCOD, PARAMS, 1, 2, 3, 4, 5, 6, 7, 8, 9. The rows represent different carbon rings (1-21) with their corresponding torsion angles (TOR1-TOR9). On the right side, there are several control panels: 'Quest File' (Load, Save), 'Data Visualization' (Histogram, Scatter, Polar Plots, Polar Scat), 'Parameters' (Generate P.C. Scores, Create, Transform, Search, Remove, Export, Temp, Select Para, Clear Para, Delete Para), 'Refcodes' (Select Refs, Clear Refs, Invert, Delete Refs, Suppress Unselected, Suppress Selected, Restore, Save Coord), and 'Miscellaneous' (Text List, Print, Help, About).

REFCOD	PARAMS	1	2	3	4	5	6	7	8	9
1	ABOVES	1	ABOVES	34.549	-20.509	-16.660	36.345	-43.370		
2	ABOCEP	2	ABOCEP	-34.271	25.589	-26.352	36.414	-1.138		
3	ABCOXA	3	ABCOXA	4.478	31.896	-99.785	17.374	-33.232		
4	ABCOXA	4	ABCOXA	29.374	7.227	-40.067	57.305	-53.596		
5	ABIMUR	5	ABIMUR	-30.295	22.421	10.823	-30.506	37.534		
6	ABILAY	6	ABILAY	-36.630	57.110	-45.117	50.496	4.090		
7	ABILAY	7	ABILAY	33.867	1.479	-36.390	56.951	-66.238		
8	ABISEJ	8	ABISEJ	-37.266	3.424	32.053	-52.753	56.369		
9	ABISIN	9	ABISIN	-34.372	0.892	33.944	-55.995	53.363		
10	ABISIN	10	ABISIN	-33.462	-3.146	36.051	-55.333	52.113		
11	ABISIS	11	ABISIS	41.252	-29.372	5.251	20.291	-37.595		
12	ABITON	12	ABITON	33.456	-25.027	-0.387	25.574	-38.480		
13	ABOONV	13	ABOONV	-37.131	33.546	-14.067	-0.854	23.222		
14	ABOSOC	14	ABOSOC	-0.349	0.445	-0.377	0.138	0.115		
15	ABOSOC	15	ABOSOC	-0.112	0.213	-0.253	0.156	-0.040		
16	ABOSOC	16	ABOSOC	0.288	-0.496	0.454	-0.276	0.000		
17	ABOSOC	17	ABOSOC	-0.158	0.063	0.048	-0.151	0.132		
18	ABOSOC	18	ABOSOC	0.330	0.034	-0.379	0.380	-0.564		
19	ABOSOC	19	ABOSOC	-0.171	0.099	0.077	-0.180	0.216		
20	ABOSOC	20	ABOSOC	0.091	-0.108	0.079	-0.029	-0.040		
21	ABOSOC	21	ABOSOC	-0.196	0.161	-0.069	-0.096	0.155		

Figura 1.7. Hoja de datos en *VISTA* que muestra los resultados de una búsqueda en la CSD para los ángulos de torsión en anillos de carbono

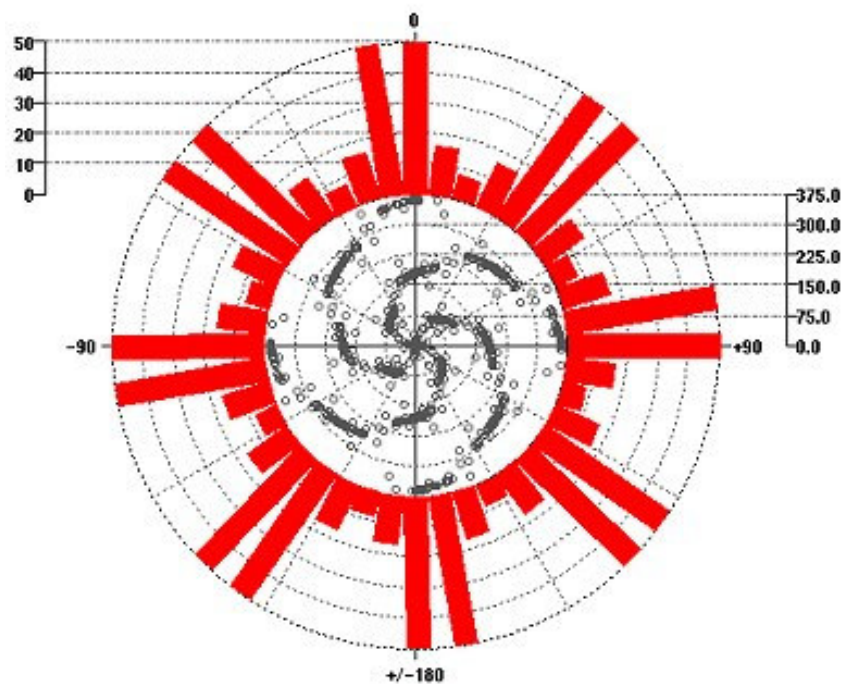


Figura 1.8. Scatterplot generado por VISTA

- **creación bases de datos (*PreQuest*).**- *PreQuest* es un programa para la construcción de bases de datos que tiene como objetivo primario crear ficheros de datos estructurales de alta calidad en un formato compatible con *ConQuest*.

PreQuest permite a los usuarios elaborar sus propias bases de datos con las estructuras objeto de investigación, conjuntamente con la CSD. Cada entrada en el CSD se comprueba y se valida usando *PreQuest*.

Características de *PreQuest*:

- Gran gama de formatos de fichero de entrada (CIF, SHELX, SD-archivos, MOL2 y BCCAB (un formato de archivo del CCDC))
- Generación automática de un diagrama químico en 2D con información de la conectividad en 3D
- Modificación de la información en forma de texto y numérica, usando un editor de textos
- Visualización y corrección de la información estructural 3D
- Exportación de entradas corregidas en el formato de CSD (*ASER*) para formar bases de datos individuales

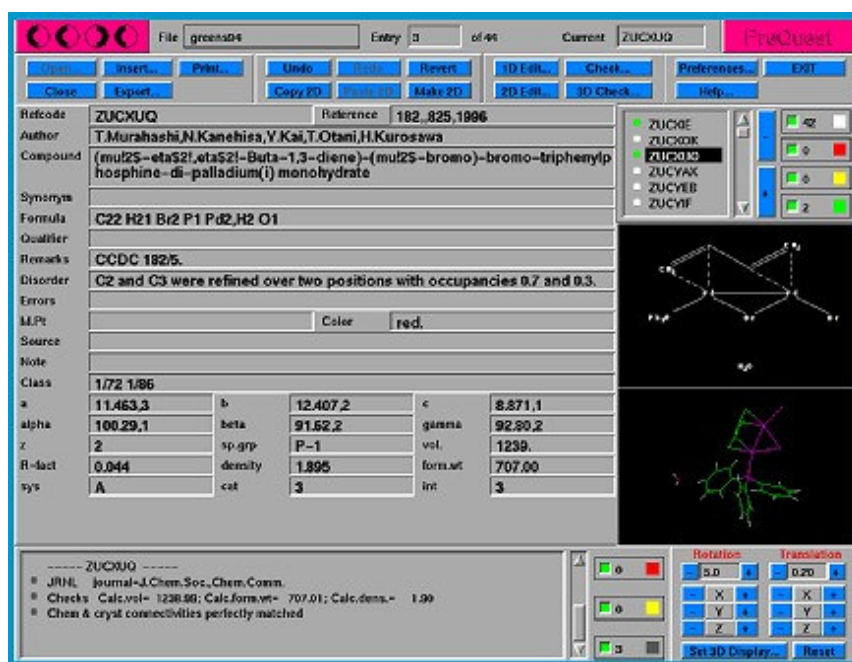


Figura 1.9. Menú principal de *PreQuest*

- **la base de conocimiento de la geometría molecular (*Mogul*).**- *Mogul* se ha presentado en 2004, y todavía se está probando. Esta base de conocimiento proporciona un fácil acceso a información sobre los valores más habituales de las longitudes de enlace, de los ángulos de valencia y de los ángulos de torsión acíclicos en un gran número de fragmentos estructurales, usando datos obtenidos de la CSD. Es fácilmente ampliable para incluir, por ejemplo, las geometrías de las esferas de coordinación de los metales.

Una vez que se le ha dado la instrucción de recuperar los datos para un rasgo geométrico particular en una molécula, como por ejemplo un ángulo de valencia, *Mogul* genera automáticamente unas claves basadas en los átomos y en los enlaces que describen el entorno del rasgo. Estas claves se usan entonces para recuperar todas las entradas de la CSD que contienen el rasgo de interés.

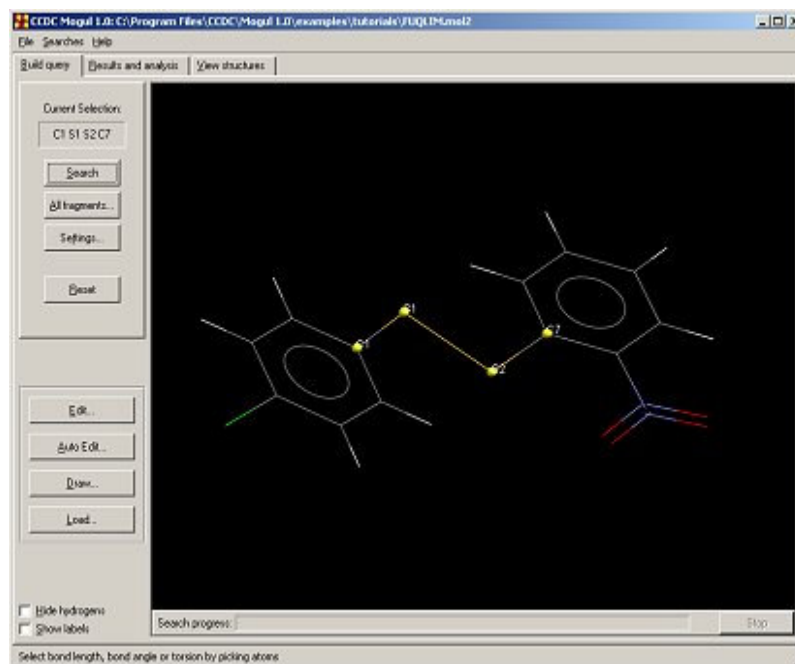


Figura 1.10. A través del interface gráfico de *Mogul* se puede establecer una distancia fija, el ángulo geométrico o el de torsión seleccionando de los átomos requeridos

Las estadísticas resultantes, como los valores medios del ángulo de valencia, pueden visualizarse entonces interactivamente en la interfaz gráfica, o pasarse a otros programas a través de la interfaz de un archivo ASCII (tipo texto, con caracteres reconocibles).

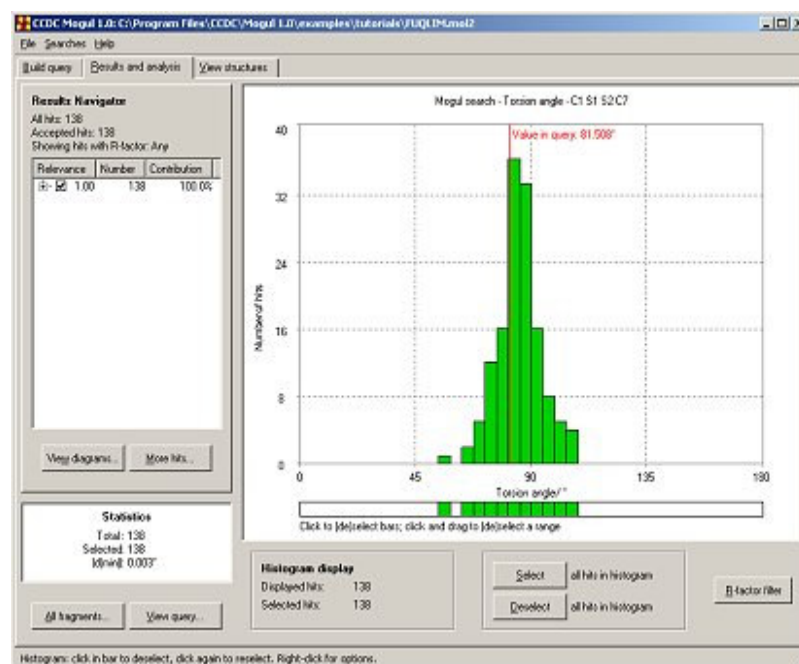


Figura 1.11. Los resultados de las búsquedas se representan a través de histogramas y sus valores estadísticos

Entre las características más destacables de Mogul figuran las siguientes:

- Interfaces gráficas o tipo mando-línea (o comando)
- Fácil definición de una condición: importación de moléculas en una variedad de formatos de archivo o bosquejando una estructura en 2D. Con un simple clic se inicia la búsqueda de cualquier enlace, ángulo de valencia o ángulo de torsión seleccionado
- La búsqueda rápida incluye la capacidad de buscar entradas estructuralmente relacionadas si las encontradas son insuficientes
- La interfaz gráfica permite la inspección de los histogramas de las longitudes de enlace, de los ángulos de valencia y de los ángulos de torsión, la selección/deselección de las barras de los histogramas, la visualización de las estructuras cristalinas que contribuyen al histograma, y el filtrado de los resultados en base al factor-R

El sistema de la CSD también incorpora **IsoStar**, una base de datos que contiene interacciones intermoleculares, conteniendo además los datos derivados de la CSD y el PDB.

1.5. **PUBLICACIÓN Y VALIDACIÓN DE ESTRUCTURAS: requisitos para la publicación en la CSD**

Para tener una idea general de la forma de trabajo de la CSD en lo referente a la inclusión de nuevos complejos, destacar que para publicar en la CSD un nuevo complejo es necesario enviar los datos del mismo en formato CIF (archivo de información cristalográfico, que es el medio estándar para el intercambio de información en cristalografía y que está avalado por la “*Unión internacional de la cristalografía (IUCr)*”), incluyendo autor y los detalles del diario al CCDC.

Los **datos bibliográficos y químicos** obligatorios que se deben adjuntar son:

- Nombre del autor o autores
- Nombre de la revista o libro donde se ha publicado
- Nombre del compuesto químico, sistemático y trivial

- Secuencia de aminoácidos en los péptidos
- Texto en el que se indican condiciones experimentales o resultados especiales (estudio de neutrones, temperatura o presión distintos de las del ambiente, etc.)
- Clase química (alcaloide, esteroide, etc.)
- Menciones de textos y comentarios acerca del desorden, errores localizados durante la validación y características estructurales especiales
- Información cristalográfica numérica básica (factor R y otros indicadores de la precisión)
- Fórmula química molecular

En cuanto a la **tabla de conexión química** se precisa:

- Diagrama de la estructura química bidimensional formal en términos de las propiedades de los átomos y de los enlaces, totalmente localizable por fragmentos subestructurales o por estructuras complejas
- Aparte de los tipos de enlace covalente comunes, la CSD asegura sus tablas de conexión la provisión de las situaciones de enlace aromático, doble, deslocalizado, cuádruple, pi, y de catena que se producen en los compuestos organometálicos

En cuanto a la **estructura cristalina tridimensional** del complejo, se precisa:

- Dimensiones de la celda e incertidumbres estándar
- Grupo espacial y operadores de simetría
- Celda reducida (de Niggli)
- Z (Nº de unidades estructurales por cada celda)
- Z' (Nº de unidades estructurales por unidad asimétrica)
- Dx (densidad calculada)
- R-factor cristalográfico
- Sistema cristalino
- Grupo espacial (notación de Herman-Mauguin)
- Coordenadas atómicas y sus incertidumbres estándares
- Radios covalentes atómicos

Para la validación de los datos, destacar que todos los datos que se incorporan a la CSD (directamente de un CIF, o como entrada a través los redactores científicos del CCDC) se comprueba y se evalúa usando *enCIFer* (Smith y Johnson, 2002), *PreQuest* y demás software interno del CCDC. De vez en cuando, se pueden presentar pequeñas diferencias entre los datos que se pretenden publicar y una entrada de la CSD. Estas diferencias se presentan debido a los propios procesos de la publicación y revisiones del CIF.

Cualquier error localizado se resuelve en la misma sede del CCDC o se devuelve de nuevo al autor original para su corrección. Los redactores científicos suelen insertar observaciones referentes a la resolución de errores, así como observaciones referentes a las características especiales de la estructura, y registran la naturaleza de cualquier desorden cristalográfico.

Cada entrada en la CSD corresponde a la determinación individual de una estructura cristalina específica, y es identificada por un código de la referencia de CSD, abarcando:

- seis caracteres alfabéticos (por ejemplo *BENZEN*) para identificar el compuesto del producto químico
- un número de dos dígitos (como en *BENZEN05*, *FERROC14*, etc.) para identificar una determinación experimental específica de la estructura cristalina de *BENZEN* o de *FERROC*, que quizás por la determinación experimental llevada a cabo por cada autor y dependiendo de las radiaciones o condiciones de temperatura, puede haberse obtenido una información sensiblemente distinta.

Un ejemplo de una entrada típica en la CSD (se ilustra el complejo con referencia **TUSJIA** en la CSD en la Figura 1.12.), podría darnos la información bibliográfica, química, estructural en dos dimensiones y estructural en tres dimensiones.

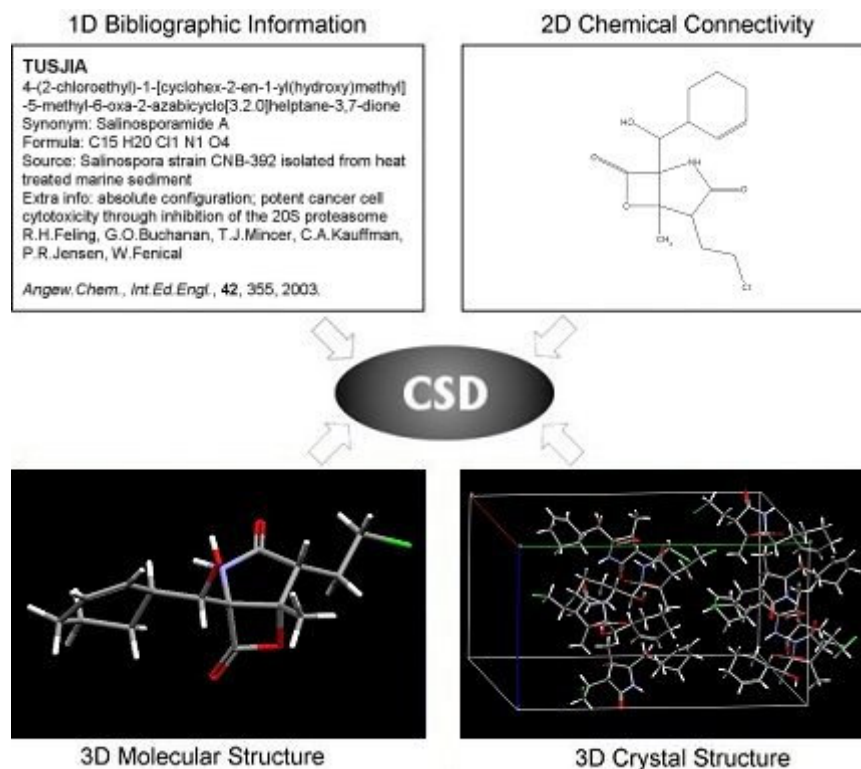


Figura 1.12. Complejo con *Refcode* TUSJIA en la CSD

1.6. APLICACIONES DE LA CSD

1.6.1. Introducción

El análisis correlativo de datos y de estructuras químicas es un área cada vez más importante en la ciencia, y sobre todo en biología, donde los estudios sobre el genoma y las proteínas han generado una enorme cantidad de datos, convirtiéndose en fuente inagotable de investigaciones privadas y subencionadas. También en la industria farmacéutica y química se empieza a dar gran importancia a las bases de datos de compuestos, con el objetivo de rentabilizar las investigaciones y convertir su fabricación en algo rutinario. El CCDC fija como uno de sus objetivos más importantes, la promoción y aplicación del uso de los datos cristalográficos en la industria.

Como ya se dijo, la CSD cubre sólo las partes orgánica y organometálica del conjunto de los compuestos químicos. Otras bases de datos, gemelas de la CSD cubren el

resto. Éstas son el *Banco de Datos de las Proteínas* (PDB), la *Base de Datos de las Estructuras Cristalinas Inorgánicas* (ICSD), y la *Base de Datos de los Metales* (CRYSMET). Juntas, las cuatro bases de datos completan el rango de información sobre la estructura cristalina y molecular obtenida mediante técnicas cristalográficas y relacionadas. La obtención y el uso del conocimiento estructural fundamental y el desarrollo de herramientas para habilitar y usar éste es mayor en los casos de la CSD y del PDB, ya que estas bases de datos se benefician del impulso y de los considerables recursos de la industria farmacéutica y de otros intereses de la investigación terapéutica.

El valor real del conocimiento obtenido de la CSD es el uso que se le puede dar. Está claro que la información obtenida sobre las conformaciones “preferidas” por las moléculas y sobre cómo les gusta a éstas interactuar es útil en cualquier procedimiento que implique la comprensión y el control de estos parámetros. En las compañías químicas y farmacéuticas, esto es lo que se pretende en los procesos de investigación y desarrollo y, de hecho, es la industria farmacéutica la que constituye la mayor parte de la clientela del CCDC. Así, los investigadores que estudian los fármacos o realizan otras actividades de diseño molecular, usan rutinariamente la CSD como una guía para comprobar, por ejemplo, cómo una pequeña molécula de un fármaco preferirá interactuar con una gran molécula proteínica. De esta manera, en este campo de los fármacos y del diseño molecular se han dado grandes pasos en los últimos años. El conocimiento de las tendencias observadas en las interacciones moleculares, extraído de la CSD, puede llevar a una mejor comprensión de por qué ciertas sustancias químicas son farmacológicamente activas mientras que otras no lo son. Una información como ésta, basada en la investigación de la CSD, puede llevar incluso a la obtención de nuevas entidades químicas con una actividad farmacológica reforzada (fármacos mejores y más seguros). El CCDC está desempeñando un importante papel entre las compañías e instituciones académicas en el ámbito farmacológico y de otras aplicaciones, y cada vez se reconoce más su valor y se está comenzando a comprender el potencial de esta aproximación basada en el conocimiento. Ésto lo está logrando a través de desarrollos internos, y tomando parte en la investigación colaboradora de nuevas metodologías y productos para aprovechar los recursos de la informática química y molecular de los datos cristalográficos. Así, por ejemplo, se están usando los datos cristalográficos en la determinación de estructuras cristalinas a partir de difracción de polvo, o de otros métodos analíticos.

Es posible vislumbrar un futuro en el que el conocimiento obtenido a partir de los datos cristalográficos sea una parte vital de un amplio rango de herramientas de aplicaciones reales, resolviendo problemas de relevancia tanto académica como industrial. El mantenimiento continuado y el desarrollo de la CSD y de sus sistemas asociados son una piedra angular vital para el desarrollo de estos objetivos.

No podemos cerrar esta introducción sin hacer una breve mención de las redes neurales, consideradas como uno de los desarrollos más excitantes de los últimos años en la ciencia computacional (Bearle y Jackson, 1990). Este software “inteligente” se ha aplicado a una amplia variedad de problemas que implican el reconocimiento de patrones y de relaciones no lineales. Las aplicaciones químicas (Gasteiger y Zupan, 1993) incluyen la predicción de la estructura secundaria de las proteínas a partir de la secuencia primaria (Qian y Sejowski, 1990), la interpretación de los espectros de RMN (Mayer *y otros*, 1991), la racionalización de las relaciones estructura cuantitativa–actividad (Aoyama *y otros*, 1990), y el análisis conformacional, así como la predicción de las distancias interatómicas mínimas y máximas (Jordan y Leach, 1995)

1.6.2. Aplicaciones de la CSD a la química inorgánica molecular

Entre las aplicaciones de la CSD a la química inorgánica molecular, podemos destacar las siguientes:

1.6.2.1. Dimensiones moleculares obtenidas a partir de los datos de rayos X en monocristales y su fiabilidad

Un uso obvio de la CSD es la obtención de valores estándar para las longitudes de los enlaces covalentes en un rango muy amplio de entornos químicos. Tal información tiene un interés intrínseco por derecho propio y proporciona también valiosos criterios frente a los que comparar nueva información estructural. La CSD proporciona una base de datos de dimensiones moleculares que puede ser valiosa en situaciones en las que la determinación directa de la geometría de una especie molecular no es fiable o es imposible. Cuando se procesa el conocimiento (esto es, se obtienen los datos apropiados, se corrigen los

errores, se validan los datos, se determinan las estadísticas, etc.), entonces aparece un rango de aplicaciones del modelado molecular. A finales de los 80, el CCDC y sus colaboradores de investigación publicaron dos estudios importantes (ver Orpen, 1998) en los que se hacían largas recopilaciones (citadas a menudo) que contenían estadísticas descriptivas (medias, medianas, valores de los cuartiles superior e inferior, etc.) de las dimensiones intramoleculares típicas correspondientes a varios cientos de tipos de enlaces químicos diferentes contenidos en las aproximadamente 70000 estructuras cristalinas de los complejos orgánicos (Allen *y otros*, 1987) y metálicos (esto es, compuestos de los elementos del bloque-*d* y del bloque-*f*) -Orpen *y otros*, 1989- disponibles por aquel entonces. Estos análisis se llevaron a cabo caso por caso en las primeras versiones de la CSD, que eran mucho más pequeñas, conteniendo alrededor del 20% de los datos actualmente disponibles. No tardó en plantearse la posibilidad de incluir tablas de actualización de estas recopilaciones dentro del marco de las bibliotecas basadas en el conocimiento. Los más recientes esfuerzos por seleccionar los contenidos de la CSD y llevarlos a bases de conocimiento útiles se han centrado en la generación automática de la geometría intermolecular y de archivos de geometría intramolecular: *IsoStar* (Bruno *y otros*, 1997) y *Mogul* (Taylor *y otros*, 2001), respectivamente. Aquí el objetivo es construir dentro de la base de conocimiento disponible subconjuntos adecuados de los datos presentes en la CSD organizados por su contexto químico.

Estas bases de conocimiento permiten realizar funciones vitales como las que se citan a continuación:

- Validación de datos mediante la comparación de la geometría molecular en una estructura cristalina parcial o recientemente determinada, o determinación de una estructura obtenida por absorción de rayos X, espectroscopía de RMN o computacionalmente. Muchos de estos estudios se han publicado, siendo particularmente interesantes cuando se desarrollan nuevos campos de fuerzas para procedimientos de MM o cuando se evalúa la exactitud de los métodos de optimización de la geometría de la mecánica cuántica (ver Cundari *y otros*, 2000; Burton y Deeth, 1995; Burton *y otros*, 1995). En ocasiones, ocurre lo contrario, y los datos erróneos determinados

cristalográficamente se corrigen usando información computacional (Kahr y otros, 1995).

- Construcción de un modelo: la previsión de la geometría para su incorporación en un modelo estructural, como por ejemplo en el refinamiento de la estructura de una proteína constreñida basado en la geometría de un fragmento “ideal”, o en la solución y refinamiento de estructuras a partir de datos de difracción de polvo (“*ab initio*”, pero basado en componentes moleculares o submoleculares conocidos y esencialmente rígidos). El desarrollo de modelos fiables para el refinamiento de la estructura de las proteínas sobre la base de la relativamente alta resolución de los datos de difracción de monocristales en pequeñas moléculas de la CSD ha sido el principal enfoque de tales estudios.
- Información que permita refinar la geometría restringida para el análisis de la estructura macromolecular o cuasimacromolecular por métodos de rayos X. Aquí el objetivo es agregar cierta información al proceso de refinamiento a la vez que se permite una mayor flexibilidad en los modelos de refinamiento de la que es posible con un modelo constreñido (ver Harding, 1999, 2000, 2001).
- Generar modelos para la optimización usando orbitales moleculares (MO) *ab initio* u otras teorías que parten de una geometría molecular razonable basada en datos observados en la CSD.
- Mejorar la parametrización de los procedimientos de modelado empíricos o semiempíricos mediante la incorporación de datos estructurales en el proceso de optimización (ver por ejemplo, Halgren, 1996).

Todas estas funciones se basan en la suposición de que una geometría molecular (parcial) sacada de una estructura es de alguna manera transferible a una estructura cristalina o molécula diferente.

1.6.2.2. Modelos de estructura y enlace

La CSD ofrece un gran número de evidencias estructurales frente a las que probar una hipótesis estructural dada o en las que inspirarse para desarrollar nuevos modelos estructurales. Estos modelos pueden adoptar una forma cualitativa (la hipótesis sería del tipo: “*la longitud de enlace x aumentará conforme decrezca el ángulo de enlace y* ” o una forma cuantitativa, en la que se postula una correlación más explícita de dimensiones moleculares o en la que se prueba un modelo computacional basado en, por ejemplo, los métodos de la mecánica cuántica frente a los datos experimentales disponibles. Con en este espíritu se ha tratado una gran variedad de áreas.

1.6.2.2.1 Enlace metal-ligando

Una familia importante de estudios en esta área hace referencia a los efectos de un metal (y de sus coligandos) sobre la geometría de un ligando o familia de ligandos. Esto ha llevado al desarrollo de modelos de la naturaleza del enlace metal-ligando que pueden probarse frente a los patrones de comportamiento observados en los datos extraídos de la CSD. La tesis es que se puede probar la naturaleza de la interacción $L'_n M-L$ variando el metal M (y también el ligando L' , en principio). Los efectos de M (y de sus estados de oxidación, número de electrones, coligandos, etc.) sobre la geometría del ligando L y sobre la longitud del enlace $M-L$ pueden proporcionar una información útil sobre la naturaleza de la interacción $M-L$. En algunos casos, puede ser útil también incluir un rango más amplio de átomos o de grupos de átomos en lugar de M , y no suponer sólo que M = metal de transición, con lo que se podrían inspeccionar los casos en los que M pertenece al amplio rango de los ácidos de Lewis, como H^+ , O , etc. Los modelos cualitativos que se usan en esta clase de estudio se basan típicamente en los dos siguientes aspectos de la interacción $M-L$:

- Donación $\sigma L'_n M-L$. Esta es típicamente la principal interacción de enlace, y es probable que provoque la mayor influencia sobre la geometría de L . Sin embargo, hay que señalar que los efectos sobre la geometría del ligando pueden no estar directamente relacionados sobre la energía del enlace $M-L$.

- Enlace $\pi L'_n M-L$ (a menudo, donación inversa). Este es típicamente un efecto secundario, pero puede ser importante tanto en términos de la interacción $M-L$ como por sus efectos sobre la geometría de L .

Esta clase de estudio tiene implicaciones para el diseño de ligandos, y es que al irse comprendiendo mejor la naturaleza del enlace $M-L$, es posible que se pueda diseñar el ligando L para aumentar, por ejemplo, la fortaleza del enlace $M-L$ a expensas de una o más interacciones $M-L'$, lo que haría que aumentara la reactividad del enlace $M-L'$. Estos estudios se han enfocado hacia metalfosfinas y enlaces relacionados (ver, por ejemplo, Garner y Orpen, 1993; Crispini *y otros*, 1996). Otros han estudiado el comportamiento de los ligandos que tienen estados de oxidación variables, que pueden determinarse a partir de datos estructurales sobre la base de la geometría del ligando (Carugo, 1994). Cundari ha informado recientemente de nuevas aproximaciones al estudio de enlaces múltiples metal-ligando (Cundari y Russo, 2001).

1.6.2.2.2 Efectos del ligando sobre las geometrías de ácidos de Lewis

La aproximación inversa a la anteriormente descrita puede basarse en el mismo modelo de unión (las dos interacciones citadas en el apartado anterior), pero ahora la información sobre la primera interacción se refiere a la geometría del fragmento $L'_n M$ y a su orbital aceptor σ , mientras que la segunda proporciona información sobre los orbitales de simetría $\pi M-L$ del fragmento $L'_n M$.

El estudio de los complejos $L-SbCl_5$ llevó a la observación de que las longitudes de enlace $Sb-Cl$ *cis* y *trans* variaban casi lo mismo al variar L . Esto contrastaba claramente con las normas sobre el papel dominante de la influencia *trans* sobre la influencia *cis* en la química de los complejos metálicos. Recientemente, se ha presentado un modelo cualitativo simple para interpretar la variación en las magnitudes relativas de la influencia *trans* y *cis* observadas en función de la configuración electrónica en un rango de complejos metálicos de los bloques *p*- y *d*- (Anderson y Orpen, 2001). Dicho modelo continúa manteniendo el punto de vista tradicional de la influencia *trans*, que se basa principalmente en la experiencia obtenida de las especies planas-cuadradas d^8-ML_4 (Bugarcic *y otros*,

1993) y de las no tan estudiadas especies octaédricas d^6-M_4 (Coe y Glenwright, 2000).

En una importante serie de estudios, Álvarez y sus colaboradores se han centrado en una variedad de aspectos de las interacciones metal-metal y metal-ligando para desarrollar y probar modelos basados en la teoría de OM tanto cualitativa como cuantitativa (Aullon y Álvarez, 1993, 1996, 1997; Álvarez y Aullon, 1999; Aullon *y otros*, 2000; Liu y Álvarez, 1997; Mota *y otros*, 1993; Palacios *y otros*, 2000). En general, el enfoque ha sido ofrecer una comprensión cualitativa, así como probar niveles de teoría más cuantitativos frente a los datos cristalográficos presentados en la CSD.

1.6.2.3. Modelos cuantitativos

Varios grupos de investigadores han usado el *Modelo del Enlace de Valencia* semiempírico de Pauling para analizar el enlace y la estructura electrónica. Esto permite primero, la racionalización de la variación de las longitudes de enlace en complejos de metales de transición y de otros metales (See *y otros*, 1998; Palenik, 1997; Wood *y otros*, 2000; Wood y Palenik, 1998,1999; Jensen *y otros*, 2001) y en segundo lugar la asignación de estados de oxidación en los complejos metálicos (lo que es destacable en los casos en los que la identidad del ligando es ambigua o inútil en la asignación del estado de oxidación del metal en una estructura).

Como ya se ha dicho, la CSD es una acumulación de datos de una inestimable utilidad para la parametrización, calibración o validación de los modelos computacionales cuantitativos de las estructuras, tanto si éstos se basan en la mecánica molecular o en la mecánica cuántica, o incluso si se basan en combinaciones de estos dos métodos (Burton y Deeth, 1995; Burton *y otros*, 1995; Rappé *y otros*, 1992).

En muchas áreas, una meta importante es el desarrollo de una relación entre la estructura de un complejo y sus propiedades. Quizás las dos áreas de aplicación más llamativas en química inorgánica están en el contexto del comportamiento magnético de los complejos de los metales de transición y del

comportamiento catalítico de los complejos fosfina de los últimos elementos de transición.

1.6.2.3.1 Magnetismo

Hace mucho tiempo que se estableció una correlación en general entre el comportamiento magnético y el acoplamiento magnético (acoplamiento entre *espines* a nivel molecular), y varios estudios han tratado de fundamentarla bien, examinando la correlación existente entre ciertos aspectos de la geometría molecular de los complejos de cobre (II) (y de otros metales de transición) y las propiedades magnéticas que se han observado en los mismos (Ruiz *y otros*, 1997; Cano *y otros*, 2000).

1.6.2.3.2 Catalizadores fosfina

Los ligandos fosfina tienen una enorme importancia en la catálisis homogénea, sobre todo en los procesos comerciales de hidrogenación, carbonilación e hidrocianación. Su éxito se basa en el hecho de que las propiedades electrónicas y estéricas de la fosfina pueden manipularse fácilmente mediante el ajuste de los sustituyentes del átomo de P. El rango de los tipos de sustituyentes es enorme, y la consecuente amplitud de la química desarrollada para la fosfina (y para sus compuestos relacionados, como la fosfito, la fosfinito, etc.) es considerable. Aunque está claro que hay relativamente pocos catalizadores “activos” que se hayan sometido por ellos mismos a un análisis de sus estructuras cristalinas, en muchos casos se ha estudiado el ligando que es responsable del control del comportamiento catalítico de las especies metálicas. Esto prepara el camino de estudios con la meta a largo plazo de desarrollar modelos que permitan el diseño de sistemas mejorados de ligandos para ayudar a la catálisis.

Algunos parámetros clave que se consideran importantes en la determinación de la naturaleza del enlace fosfina-metal y de las propiedades catalíticas de los complejos caracterizados por cristalografía, pueden establecerse a partir de los datos de la CSD. Estos incluyen los siguientes:

- Los ángulos de cono del ligando (desarrollados por Tolman, 1977, como una medida de la influencia estérica del ligando), que pueden determinarse fácilmente a partir de los datos de la CSD (Mueller y Mingos, 1995; Smith *y otros*, 1997).
- Los ángulos de mordedura “naturales” de los ligandos quelato. Van Leeuwen *y otros* (Dierkes y van Leeuwen, 1999) han estudiado este concepto como factor determinante de la selectividad en las reacciones de hidroformilación. Los ángulos de mordedura reales y sus rangos de variabilidad pueden obtenerse directamente de la CSD y compararse con los modelos computacionales de los mismos sistemas de ligandos para proporcionar evidencias que corroboren el modelo o modelos.
- El rango de conformaciones de los ligandos flexibles. La gran mayoría de los ligandos fosfina que son de interés en la catálisis homogénea son flexibles en virtud del rango de conformaciones que los sustituyentes del P pueden adoptar o de la flexibilidad de los anillos quelato saturados. En la CSD se pueden estudiar las conformaciones probables y su rango de disponibilidad y, de esta manera, tener alguna idea del “núcleo” dentro del cual se unen los sustratos en los ciclos de reacción catalítica de los complejos metal-fosfina (ver, por ejemplo, Kadyrov *y otros*, 1999; Brunner *y otros*, 1998; Harris *y otros*, 2001; Beyreuther *y otros*, 1996; Morton y Orpen, 1992).

El resultado de estos estudios es el progreso hacia modelos cualitativos (y, eventualmente, cuantitativos) que relacionan la actividad y otros parámetros indicativos de la utilidad de un sistema catalizador con su estructura. Tales modelos, como por ejemplo el análisis cuantitativo de los efectos del ligando (QALE) de Prock, Giering y colaboradores (ver, por ejemplo, Fernandez *y otros*, 2000) son estimulantes, debido sobre todo a la naturaleza fugaz de algunos ciclos catalíticos y de sus estados de transición.

1.6.2.4. Análisis del camino de reacción

El núcleo de la hipótesis de la correlación de la estructura expuesta por Bürgi y Dunitz en algunos artículos clásicos de 1973 en adelante (Bürgi y Dunitz, 1983), es que para un fragmento submolecular, el entorno cristalino o molecular local actúa como una perturbación sobre el fragmento o sobre la geometría molecular y, por tanto, sobre el rango de geometrías energéticamente accesibles que el fragmento podría adoptar. Esta hipótesis se ha desarrollado para permitir hacerse una idea de la naturaleza de la correlación estructura-energía, y se ha aplicado a un rango de sistemas, tanto de naturaleza orgánica como organometálica.

Un ramal importante de estos estudios ha sido la variación de la geometría dentro de la esfera de coordinación de los metales. Muchas de las técnicas usadas para estudiar las conformaciones de los fragmentos orgánicos se han aplicado también a la investigación de las geometrías de las esferas de coordinación de los metales en las especies organometálicas contenidas en la CSD. Así, se han llevado a cabo estudios de grandes recopilaciones de metales de coordinaciones 4-, 5- y mayores. Los principales objetivos han sido: primero investigar los mecanismos S_N1 , S_N2 y otros mecanismos de reacción de sustitución e intercambio de ligandos (ver, por ejemplo, Bürgi y Dunitz, 1983; Yao *y otros*, 2001) y segundo hacer un seguimiento de las reacciones unimoleculares que implican, como por ejemplo, la interconversión entre geometrías de coordinación alternativas (ver, por ejemplo, Alvarez y Llundell, 2000; Alvarez, Pinsky, Llundell y Avnir, 2001; Alvarez, Pinsky y Avnir, 2001; Raithby *y otros*, 2000).

En unos estudios más tempranos, Bürgi, Dunitz *y otros* aprovecharon las medidas de la distorsión adaptada a la simetría, el análisis de componentes principales y otras técnicas estadísticas, para extraer información del gran conjunto de datos. Nørskov-Lauritsen y Bürgi usaron análisis de cluster para examinar la flexibilidad conformacional de 62 fragmentos $M(PPh_3)_2$ obtenidos de la CSD. Estos investigadores identificaron un camino de interconversión conformacional que implicaba un movimiento de engranaje de los dos fragmentos PPh_3 . Así mismo, Auf der Heyde y Bürgi (Auf der Heyde y Bürgi, 1989) analizaron complejos

metálicos pentacoordinados por comparación de las geometrías de coordinación experimentales con los poliedros de referencia usando coordenadas de deformación de la geometría. Se obtuvieron de la CSD las geometrías de coordinación de 196 complejos de metales de transición pentacoordinados y se expresaron en términos de las coordenadas de deformación de la simetría obtenidas de las geometrías ideales bipiramidal trigonal (TBP) o piramidal planacuadrada (SQP). Tras la expansión de la simetría se llevó a cabo el análisis de cluster y el análisis del componente principal (PCA). Los resultados mostraron que el camino de interconversión entre las geometrías TBP y SQP estaba dominado por las distorsiones: la transición S_N2 desde la tetracoordinación hasta la pentacoordinación y la deformación de la pseudo-rotación de Berry. Klebe y Weber (Klebe y Weber, 1994) han usado métodos similares para investigar la interconversión entre la tetra-coordinación tetraédrica y la planacuadrada.

En ejemplos más recientes, Avnir y Álvarez y sus colaboradores han usado técnicas continuas de simetría para cuantificar el rango de distorsiones con respecto a las geometrías ideales (Álvarez y Llundell, 2000; Álvarez, Pinsky, Llundell y Avnir, 2001; Álvarez, Pinsky y Avnir, 2001; Zabrodsky y Avnir, 1995; Keinan y Avnir, 2001*a,b*). Allen, Howard y colaboradores han usado métodos alternativos para cuantificar las distorsiones de los complejos de coordinación a partir de geometrías prototipo (Yao *y otros*, 2001).

En un trabajo que no usó explícitamente la CSD, Crabtree y Lavin (1986) estudiaron la trayectoria del intercambio de CO en complejos de hierro dinucleares usando el método de la correlación de la estructura, y examinando la geometría de un solo ligando carbonilo en una variedad de especies dihierro. En contraste, Johnson, Mann *y otros* han usado métodos de correlación de la estructura para investigar el mecanismo de los procesos de intercambio de CO en especies carbonilo tipo cluster metálico en las que múltiples ligandos CO experimentan simultáneamente intercambios de lugar (ver Mann, 1997; Johnson, 1997, y las referencias allí citadas).

Los estudios de reacciones intermoleculares en especies inorgánicas han sido menos comunes e incluyen el trabajo de Crabtree (Crabtree *y otros*, 1985) sobre

la activación del C–H en los alcanos mediante interacción directa con metales de transición y el de Brammer (Brammer *y otros*, 2000) sobre la transferencia de protones en las especies N–H \cdots Co y N \cdots H–Co.

1.6.2.5. Análisis conformacional

El análisis de la estructura cristalina de un complejo individual sólo determina la geometría de equilibrio de esa especie en un campo cristalino particular. No existe información sobre las posibles conformaciones lejos de esa posición de equilibrio. Como se dijo anteriormente, los estudios de las recopilaciones de estructuras cristalinas que contienen los mismos o similares fragmentos (sub)moleculares ofrecen la oportunidad de examinar cómo perturba el campo cristalino la geometría del fragmento. En el caso de las moléculas conformacionalmente flexibles (que son habituales en química inorgánica), la variación del entorno cristalino permitirá un muestreo más completo de la hipersuperficie de energía que se apuntó con anterioridad en muchos de los análisis del “paso de reacción”. Bürgi y Dunitz (1983) se fijaron pronto en esto y estudiaron algunas interconversiones claves entre conformaciones. Los objetivos en esta clase de estudios incluyen: primero la identificación de los confórmeros más comunes (los preferidos); segundo la interpretación del patrón de los confórmeros más comunes y de los caminos entre las estructuras observadas en términos de las características de la hipersuperficie de energía potencial; tercero la comparación entre la distribución de estructuras y la hipersuperficie de energía potencial computada; cuarto las relaciones entre las conformaciones observadas y otros aspectos de la geometría molecular, como la estereoquímica de la coordinación del metal al que se coordina un ligando flexible.

La mayoría de los estudios de esta clase se han enfocado hacia la variación de la conformación en los complejos metálicos con ligandos flexibles. Para la mayor parte de ellos, los sistemas de ligandos estudiados han sido cualquier especie de quelatos (saturados), incluyendo a los macrociclos (ver Harris *y otros*, 2001; Beyreuther *y otros*, 1996; Morton y Orpen, 1992; Donnelly y Zimmer, 1999; Zimmer, 2001; Raithby *y otros*, 1997, *a,b*; Leuwerink *y otros*, 1993) o las rotaciones acopladas de los sustituyentes arilo unidos a los átomos de contacto del ligando

(Barker y Orpen, 1999; Costello y Davies, 1998; Costello *y otros*, 1999; Hunger *y otros*, 1998). Se han llevado a cabo relativamente pocos estudios de sistemas de ligandos con sustituyentes alquilo (Smith y Coville, 2001; Smith *y otros*, 1998) o de sistemas con cierta rigidez, como las porfirinas (ver Cullen *y otros*, 2001).

El rango de métodos estadísticos empleado es notable: se han usado análisis de componentes principales, redes neurales y análisis de cluster de varios tipos, además de métodos más tradicionales, como los gráficos (representaciones) y las estadísticas descriptivas basadas en los ángulos de torsión. Se han usado otras aproximaciones obtenidas especialmente para anillos de cinco miembros y para otros sistemas cíclicos (parámetros de plegamiento de Cremer-Pople; Cremer y Pople, 1975). En todos los casos, debe considerarse la simetría del espacio conformacional, como describieron Longuet-Higgins, Mislow *y otros*, y aplicaron a este campo de estudio Bürgi, Dunitz *y otros*. El objetivo de usar el análisis del componente principal es identificar la “mejor” manera de ver el conjunto de datos, mediante la proyección del generalmente multidimensional conjunto de datos en un más eficaz conjunto de ejes mutuamente ortogonales, que son combinaciones lineales de los ejes originales (ángulo de torsión). En los análisis de cluster, se buscan agrupaciones objetivas de estructuras, normalmente para identificar conjuntos de conformaciones similares correspondientes a los distintos conformeros. Se han usado métodos de redes neurales para trazar las conexiones entre grupos de estructuras semejantes (Beyreuther *y otros*, 1996).

1.6.2.6. Interacciones intermoleculares entre cristales

La cristalografía proporciona la información de mejor calidad disponible sobre la geometría y la frecuencia de las interacciones entre grupos funcionales. Esto es consecuencia de que se hayan llevado a cabo muchos estudios de la CSD sobre la geometría y el predominio de varias interacciones intermoleculares en los cristales, entre las que destaca el enlace de hidrógeno en sus varias formas. El rápido crecimiento de la química supramolecular (Bruno *y otros*, 2002) y de los campos relacionados, como por ejemplo el diseño farmacológico o la ingeniería cristalina (o más específicamente la síntesis cristalina basada en las especies moleculares), ha reforzado el interés por este bien establecido tipo de estudio de las

bases de datos. En muchos casos, estos estudios han tratado de cuantificar la importancia y las características (principalmente geométricas) de estas interacciones. El tipo de interacción más estudiado con diferencia es el enlace de hidrógeno en sus muchas manifestaciones, pero también ha llamado la atención un rango de tipos de interacciones más exóticas y peculiares en química inorgánica.

1.6.2.6.1 Enlaces de hidrógeno

Aunque existe una inmensa bibliografía sobre el enlace de hidrógeno basada en la química orgánica en el estado sólido, hay un cuerpo sustancial que tiene sus orígenes en la química inorgánica molecular, particularmente en los sistemas organometálicos. El contexto inorgánico ha demostrado ser, como cabía esperar, un entorno rico para explorar tipos exóticos de enlaces de hidrógeno. Entre los que se han estudiado están: primero los metales como aceptores del enlace de hidrógeno [$M\cdots HD$; Brammer *y otros* (Brammer *y otros*, 1995) han estudiado los enlaces de hidrógeno implicados en los centros de metales de transición; Braga *y sus colaboradores* (Braga *y otros*, 1996; Braga *y otros*, 1997*a,b*) han dirigido una serie de estudios para determinar las frecuencias y las geometrías medias de un rango de interacciones de enlace de hidrógeno descubiertas en complejos metálicos]; segundo los hidruros metálicos y otros hidruros como aceptores del enlace de hidrógeno (llamados también enlaces dihidrógeno; $MH\cdots HD$, $BH\cdots HN$, etc.); tercero los ligandos unidos a un metal como aceptores del enlace de hidrógeno ($M-X\cdots HD$, $X = \text{haluro, CO, NCS, etc.}$); cuarto las especies organometálicas como dadores del enlace de hidrógeno ($M-CH\cdots A$, etc.); quinto los enlaces de hidrógeno “orgánicos”, en los que los grupos funcionales ($-CO_2H$, etc.) son familiares en química orgánica y están presentes como sustituyentes en ligandos unidos a complejos organometálicos o de coordinación de elementos inorgánicos.

1.6.2.6.2 Enlace secundario

El trabajo germinal de Alcock (1972) indicaba que lo que él calificó de enlace secundario, a saber, las interacciones intermoleculares hipervalentes de los elementos más pesados del bloque-*p*, eran el equivalente funcional del enlace de

hidrógeno en la química de esos elementos. Esta idea se ha investigado tanto en estudios de bases de datos como en estudios experimentales, buscando establecer su validez. Estas interacciones tienen mucho en común con los enlaces de hidrógeno, tanto conceptualmente, esto es, como sistemas de cuatro electrones en tres centros (Landrum y Hoffmann, 1998; Starbuck *y otros*, 1999), como en la práctica, por su capacidad para formar redes con una alta probabilidad de existencia.

1.6.2.6.3 Interacciones aurofílicas

El trabajo de Schmidbaur (2001) y otros ha hecho hincapié en las interacciones atractivas que existen entre los núcleos de oro (I) en química molecular, y Pyykko (1997) y *otros* han explicado los orígenes de este efecto. Desiraju ha investigado la influencia de tales interacciones en las estructuras cristalinas de las especies de oro(I), mientras que se han nombrado interacciones análogas $\text{Ag}^1 \cdots \text{Ag}^1$ y $\text{Cu}^1 \cdots \text{Cu}^1$ (Liu *y otros*, 1998).

1.6.2.6.4 Interacciones interligandos

Se ha señalado la importancia de las interacciones dipolo-dipolo en especies carbonilo tanto orgánicas como inorgánicas (Allen *y otros*, 1998). Dance y sus colaboradores han investigado la química intermolecular de un rango de sistemas polifenilados, que incluyen normalmente complejos metálicos. Las estructuras cristalinas de estas especies muestran motivos notablemente persistentes en los que están presentes interacciones atractivas (obtenidas por computación) filo-cara entre conjuntos de fenilos. Algunos investigadores podrían calificar estas interacciones como enlaces de hidrógeno, pero independientemente de la terminología, éstas parecen influir en la determinación de la disposición del empaquetamiento molecular en la estructura cristalina.

1.6.2.6.5 Ligando-metales del bloque s

Ha habido un interés considerable por las interacciones “intermoleculares” de los metales biológicamente importantes, muchos de los cuales pertenecen al

bloque-s (Ca, Mg, y también el Zn y los metales del Grupo 1; ver Bock *y otros*, 1995; Katz *y otros*, 1996; Pidcock y Moore, 2001)). Estas interacciones están probablemente dentro de la esfera de coordinación de los metales del bloque-s y por consiguiente puede decirse propiamente que tienen un carácter intramolecular. Sin embargo, tanto por razones históricas como por razones prácticas (la labilidad de estos enlaces en disolución, por ejemplo), estas interacciones se han considerado intermoleculares y se han clasificado como tales.

1.6.2.6.6 Interacciones agósticas

En contraste con el predominio del enlace de hidrógeno rico en electrones (cuatro electrones en tres centros), existe relativamente poca información disponible sobre el enlace agóstico intermolecular deficiente en electrones (dos electrones en tres centros) ($M\cdots HC$, etc.). No obstante, se han llevado a cabo estudios para establecer patrones de geometría usando los datos que hay disponibles (Braga *y otros*, 1996).

1.6.2.6.7 Nuevas aproximaciones

Aunque muchos estudios se han enfocado a tipos de interacciones individuales, el CCDC posee una base de conocimiento que se ha recopilado de forma automática, *IsoStar* (Bruno *y otros*, 1997), en la que se registran las interacciones entre los grupos funcionales (orgánicos) obtenidos tanto de la CSD como de algunos datos computacionales. Estas parcelas, junto con la información geométrica, pueden generarse para recopilaciones más ortodoxas de datos estructurales obtenidos de la CSD.

1.6.2.7 Grupos espaciales y empaquetamiento molecular

La predicción de la manera en la que las moléculas se empaquetan en las estructuras cristalinas es un problema importante aún sin solucionar en la cristalografía química moderna. La comprensión del empaquetamiento de los complejos moleculares organometálicos e inorgánicos no es menos problemática que en la química orgánica en estado sólido y, de hecho, puede ser más compleja

debido a la mayor suavidad de los compuestos organometálicos y a los problemas de modelado tanto de los aspectos intramoleculares como de los intermoleculares de la geometría. Aunque se han llevado a cabo con cierto éxito aproximaciones computacionales a la predicción de estructuras cristalinas y estudios de empaquetamientos alternativos (Braga *y otros*, 1994), se ha dedicado más esfuerzo a la racionalización y al análisis de los empaquetamientos observados. Mingos (ver Mingos y Rohl, 1991), Braga (ver Braga *y otros*, 1993) y colaboradores han estudiado la manera en la que se relacionan la forma y el tamaño global de los iones en las sales complejas con sus motivos de empaquetamiento.

1.6.3. Aplicaciones de la CSD a la química orgánica y a la química cristalina

Entre las aplicaciones de la CSD a la química orgánica y a la química cristalina, podemos destacar las siguientes:

1.6.3.1 Modelos estructurales sistemáticos y estadísticos

Los datos estructurales obtenidos de los primeros análisis cristalográficos jugaron un papel fundamental en la sistematización de la química estructural, en la que fue pionero Pauling en la década de los años 30, elaborando tablas de radios covalentes, de van der Waals, iónicos, etc. (Pauling, 1939), y tablas de distancias interatómicas estándar (Sutton, 1956). La rápida disponibilidad de cantidades crecientes de datos de estructuras cristalinas de forma electrónica junto con el software de búsqueda y de análisis apropiado ha hecho de la obtención de las dimensiones moleculares medias una aplicación obvia del desarrollo del sistema de la CSD (Allen, 2002; Bruno *y otros*, 2002). Dos recopilaciones principales, obtenidas de la CSD, y que contienen unas 50.000 estructuras, proporcionan tablas de distancias interatómicas estándar actualizadas para estructuras orgánicas (Allen, Kennard *y otros*, 1987: 682 tipos de enlaces químicos que incluyen 65 elementos) y para estructuras metaloorgánicas (Orpen *y otros*, 1989: 325 tipos de enlaces que incluyen a los metales de los bloques *d*- y *f*-). Estos dos artículos han recibido más de 3000 citas (Redman *y otros*, 2001). Engh y Huber (1991) han obtenido longitudes de enlace y ángulos de valencia medios para estructuras peptídicas basándose en una CSD que en aquella época contenía unas 80000 entradas, y

basando su clasificación en los 31 tipos de átomos que son más apropiados para el entorno proteínico. Estos datos se usan ampliamente en la determinación, refinamiento y validación de nuevas estructuras macromoleculares, y los autores originales han sugerido su actualización, ya que la actual CSD es más de tres veces mayor. Más recientemente, Clowney *y otros* (1996) y Gelbin *y otros* (1996) han usado los datos de la CSD para generar diccionarios estructurales que permitan verificar las nuevas estructuras que entran en la base de datos de los ácidos nucleicos (Berman, Westbrook *y otros*, 2002).

En las recopilaciones basadas en la CSD y en otros estudios pioneros, puede verse cómo han surgido tratamientos estadísticos apropiados para los resultados, como por ejemplo:

- en la presentación de estadísticas descriptivas apropiadas para las distribuciones de datos que dan lugar a valores medios específicos
- el uso tanto de pruebas paramétricas como no-paramétricas evalúa la importancia de las diferencias entre medias
- el uso de la covarianza, de la correlación y de la regresión para determinar hasta qué punto están relacionados dos parámetros.

Estas técnicas, junto con una variedad de métodos estadísticos multivariable (análisis estadísticos de variables múltiples), se incluyeron en el archivo del subprograma *CAMAL* (Taylor, 1986*b*) y fueron revisados por Taylor y Allen (1994).

1.6.3.2 Interacciones intermoleculares

En su conferencia del Premio Nobel, Lehn (1988) definió la química supramolecular como “la química del enlace intermolecular”. Una estructura cristalina es la supermolécula arquetipo y proporciona un conocimiento experimental directo sobre las muchas y variadas interacciones intermoleculares que controlan la agregación molecular, esto es, los tipos de interacción que se producen, sus características geométricas, las preferencias direccionales y los motivos supramoleculares comunes que las envuelven. Este conocimiento es

fundamental para nuestra comprensión y para el desarrollo de todas las cuestiones supramoleculares, incluyendo la síntesis supramolecular, la validación de la estructura cristalina, la ingeniería cristalina, el acoplamiento proteína-ligando y la predicción de la estructura cristalina. La CSD es por consiguiente una fuente primaria de conocimiento que puede aplicarse a todas estas áreas.

Los estudios clásicos de los caminos de reacción en una serie de aminoquetonas de Dunitz, Bürgi y colaboradores (Dunitz, 1979; Bürgi y Dunitz, 1983, 1994) han estado en la vanguardia de los estudios de interacciones intermoleculares haciendo uso de la CSD. Estos investigadores relacionaron los cambios en la geometría de un centro carbonilo (alargamiento del enlace C=O, piramidal del C en >C=O) con la distancia sin enlace entre el átomo de carbono y los nucleófilos (Nu) oxígeno y nitrógeno al acercarse estos últimos. Los científicos constataron la constancia del ángulo Nu...C=O, que estaba en torno a 110°, y modelaron el sistema completo, representando las etapas iniciales del ataque del nucleófilo sobre el C=O, que desembocaban finalmente en la conversión del centro del átomo de carbono sp^2 en una geometría tetraédrica sp^3 .

Los primeros estudios de interacciones intermoleculares en la CSD fueron bastante dificultosos, ya que se basaban en un proceso de búsqueda en dos fases en cada una de las cuales se usaba un programa diferente:

- búsqueda de subestructuras bidimensionales para localizar las entradas de la CSD que contenían los grupos funcionales químicos de interés
- búsqueda geométrica tridimensional para localizar interacciones intermoleculares concretas.

En 1992, dentro del marco del programa *QUEST3D*, se integraron y se mejoraron unos servicios para la búsqueda de subestructuras moleculares (bidimensionales) y supramoleculares (tridimensionales), cuya filosofía de búsqueda fundamental está ahora disponible dentro del software *ConQuest* del CCDC (Bruno *y otros*, 2002). Los servicios de análisis de datos están disponibles en la actualidad a través del programa *Vista*, y el programa *Mercury* proporciona una visualización muy mejorada de las estructuras (Taylor y Macrae, 2001; Bruno *y otros*, 2002).

A pesar de las dificultades operacionales, antes de 1992 se publicaron muchos estudios significativos de la CSD sobre sistemas con enlaces de hidrógeno, incluyendo varias mediciones y revisiones importantes (Kroon *y otros*, 1975; Jeffrey y Maluszynska, 1982, 1986; Murray-Rust y Glusker, 1984; Taylor y Kennard, 1982, 1984; Desiraju, 1991*a*). Y es que los estudios de las interacciones del enlace de hidrógeno han dado lugar a la mayor parte de las aplicaciones intermoleculares de la CSD. Un éxito particular del análisis de la CSD ha sido la clara identificación (Taylor y Kennard, 1982) de la interacción C–H \cdots O como un enlace de cierta importancia en la estabilización tanto de las grandes estructuras biológicas como de las disposiciones del empaquetamiento de pequeñas moléculas. Desde entonces, se han publicado muchos estudios sobre interacciones más débiles, como la O–H \cdots π y la N–H \cdots π (Viswamitra *y otros*, 1993), y la C–H \cdots X, donde X es O, N o incluso algunos sistemas π (Steiner, 1995*b*); por su parte, Desiraju (Desiraju, 1991*a*) ha presentado una revisión completa de los sistemas C–H \cdots O. En su revisión de 1984, Taylor y Kennard han resumido los primeros trabajos, discutiendo los motivos de la direccionalidad de los pares solitarios, particularmente en el grupo >C=O. Jeffrey y Saenger (1991) y Jeffrey (1997) proporcionaron también en sus monográficos una apreciación global comprensiva de gran parte del material inicial y muchas referencias importantes sobre la investigación basada en la CSD en relación con esta área.

1.6.3.3 Enlaces de hidrógeno

Los enlaces de hidrógeno son las interacciones intermoleculares más importantes que determinan el reconocimiento y la agregación molecular en biología y en química, y están implicados en las estructuras de muchos materiales nuevos. La estructura cristalina ha sido la principal fuente de conocimiento sobre los enlaces de hidrógeno durante muchos años, como muestra un examen de las referencias contenidas en la base de datos *DBU_{se}* (actualmente, *WebCite*). Muchos estudios recientes han usado el sistema de software de la CSD para cuantificar las propiedades métricas y direccionales de los enlaces de hidrógeno fuertes comunes que incluyen a los dadores N–H y O–H y a una variedad de aceptores de N y de O (ver por ejemplo, Steiner y Saenger, 1992*b*; Pirard *y otros*, 1995; Chertanova y Pascard, 1996). También se han estudiado de esta manera otros enlaces de

hidrógeno, como los $N,O-H\cdots haluro$ (Steiner, 1998a).

Sin embargo, estos estudios básicos se han ampliado de varias maneras. Así, Gavezzotti y Filippini (1994) han usado los datos de la CSD sobre los sistemas $N,O-H\cdots N,O$ para calibrar los potenciales empíricos del enlace de hidrógeno. Steiner y Saenger (1994) y Steiner (1995) han estudiado, respectivamente, los sistemas $O-H\cdots O$ y $N-H\cdots N$, y han demostrado que existe una ligera correlación entre el alargamiento del enlace $O,N-H$ y el acortamiento de las separaciones $O\cdots O$ y $N\cdots N$. Allen, Cole y Howard (1995a,b) han estudiado la geometría de los enlaces de hidrógeno $N,O-H\cdots NH_2$ en las anilinas y han establecido el camino de reacción para la protonación del N de la amina: como el grupo amina rota fuera del plano del anillo fenilo, la interacción conjugativa desaparece, generándose un par solitario en el crecientemente piramidal átomo de N, que actúa entonces como un aceptor del enlace de hidrógeno (que es cada vez más fuerte).

1.6.3.4 Interacciones en las que no media el hidrógeno

En una revisión, Desiraju (1995) ilustró la importancia estructural de un amplio rango de interacciones atractivas sin enlace en las que no mediaba el hidrógeno y destacó el valor de la CSD para la identificación y caracterización de tales interacciones. En este contexto, la combinación del análisis de la CSD (para indicar las geometrías de interacción preferidas) y de los cálculos *ab initio* (para explorar las energías de las interacciones en estas regiones preferidas), tiene un valor considerable. Cuantificando las energías de interacción, la aproximación combinada no sólo ayuda a predecir los patrones de interacción más probables, sino también su robustez relativa por comparación con, por ejemplo, enlaces de hidrógeno análogos.

La fuerte tendencia de los halógenos $X = Cl, Br, I$ a formar contactos cortos con otros halógenos y, en especial, con átomos electronegativos de O y de N (Nyburg y Faerman, 1985) es bien conocida (Price *y otros*, 1994). Recientes estudios combinados CSD/IMPT de sistemas $C-X\cdots O=C<$ (Lommerse *y otros*, 1997) y de sistemas $C-X\cdots O(\text{nitro})$ (Allen, Lommerse *y otros*, 1997) han mostrado una marcada preferencia porque se produzca la interacción $X\cdots O$ a lo largo de la

extensión del enlace C–X, con energías de interacción en el rango de -7 a -10 kJmol^{-1} . Estas energías son comparables con las de los enlaces de hidrógeno C–H \cdots O, y Desiraju y colaboradores (ver Desiraju, 1995) las han usado para diseñar varias estructuras cristalinas correspondientes a nuevas moléculas pequeñas. En contraste, las interacciones C–Cl \cdots Cl–C tienden a formar con las C–Cl \cdots Cl ángulos próximos a 90° (Price *y otros*, 1994). Juntos, estos estudios proporcionan una imagen del Cl en un entorno electrostático tetrapolar con potenciales especies interactuantes, con un área de densidad positiva apuntando hacia fuera a lo largo de la extensión del enlace C–Cl [con lo que interactúa con el O(δ^-)], y una región de densidad negativa en las direcciones perpendiculares al enlace C–Cl [que interactúa con el Cl(δ^+) y con el H(δ^+), como se señaló en los enlaces M–Cl \cdots H anteriormente discutidos]. Se han usado métodos computacionales similares para estudiar las interacciones cortas S \cdots O y S \cdots N que aparecen en los nucleósidos tiazol (Burling y Goldstein, 1992) y se ha comprobado el predominio general de tales interacciones usando el Sistema de la CSD (Burling y Goldstein, 1993).

Taylor *y otros* han estudiado la importancia de las interacciones multicentro entre grupos funcionales, construyendo entornos compuestos de campo cristalino para los grupos carbonilo y nitro durante una búsqueda de reemplazamientos isómeros (en los que se mantienen estructuras electrónicas similares en las moléculas) en el modelado de las interacciones proteína-ligando. Sobre la base de este trabajo, Allen *y otros* (1998) han comprobado que las interacciones dipolares antiparalelas que incluyen grupos $>\text{C}=\text{O}$ tienen energías de interacción (de unos -20 kJmol^{-1}) que son comparables con las de los enlaces de hidrógeno de fuerza media, y que están claramente implicadas en la estabilización del empaquetamiento cristalino de, por ejemplo, los 3-oxoesteroides. También se ha comprobado que las interacciones carbonilo-carbonilo son significativas en las proteínas y que son responsables de la estabilización de las estructuras secundarias de ciertas proteínas (Maccallum *y otros*, 1995*a,b*), y de la estabilización de las conformaciones de Ramachandran parcialmente permitidas de la asparagina y del ácido aspártico (Deane *y otros*, 1999).

1.6.3.5 Ingeniería cristalina

El campo de la ingeniería cristalina (Desiraju, 1989), que se ha desarrollado con rapidez desde mediados de los años 80, es una de las principales áreas de aplicación para la obtención de nuevos conocimientos de la CSD. Temas de particular interés en esta área son las topologías y las frecuencias relativas de formación de los motivos de interacción intermolecular más comunes observados en las estructuras cristalinas. Etter *y otros* (1990) propusieron una descripción por conjunto gráfico de los motivos, un concepto que adoptaron desde entonces Bernstein *y otros* (1995), mientras que Desiraju (1995), consciente de la importancia de estos robustos y reproducibles motivos supramoleculares en el diseño estructural, los ha descrito como “sintones” supramoleculares (esto es, influencias para la dirección de una estructura). Esta terminología reconoce la base sintética de la ingeniería cristalina, tomando prestado el término “sintón” del mundo de la química orgánica molecular (Corey, 1967). En su artículo, Desiraju (1995) cita el uso de la CSD para llevar a cabo retrosíntesis supramoleculares, esto es, usar estructuras existentes para reconocer y emplear entonces los más frecuentes (y por tanto más robustos) patrones de determinación de estructuras sin enlaces. Para aportar cierta automatización al análisis retrosintético, Allen *y otros* (1999) modificaron el sistema de software de la CSD para reconocer y generar un ranking de probabilidad de motivos en anillos bimoleculares con enlaces de hidrógeno en las estructuras orgánicas publicadas, restringiendo el análisis a los enlaces de hidrógeno más fuertes que incluían a los dadores N,O–H y a los aceptores N,O. El concepto de motivo ha demostrado también ser importante en un estudio de enlaces de hidrógeno intramoleculares presentado por Bilton *y otros* (2000), y que tiene implicaciones significativas para la organización supramolecular.

Aunque los enlaces de hidrógeno están implicados en la formación de muchos sintones robustos, hay una considerable evidencia (ver por ejemplo, Desiraju, 1995) de que los sintones implican otras interacciones sin enlace que también se pueden predecir. Una de las primeras menciones en ingeniería cristalina fue la del sintón C–Cl··Cl–C, que llevó al diseño de un eje de celda corto de unos 4 Å (ver Sarma y Desiraju, 1986). Se ha observado que los grandes grupos funcionales, como los cationes tetrafenil–P⁺, forman múltiples motivos fenilo

entrelazados (Dance y Scudder, 1996). Otro tipo de interacción se ejemplifica en las interacciones C–halógeno $\cdots\pi$ con los sistemas aromáticos, que parecen jugar también un papel importante en el empaquetamiento cristalino (Prasanna y Row, 2000). Nangia (2002) ha presentado una reciente revisión sobre la investigación de las bases de datos en la ingeniería cristalina.

1.6.3.6 Polimorfismo y pseudopolimorfismo

La CSD dada a conocer en octubre de 2001 (con 245392 entradas) sólo contenía 4335 compuestos en los que se declarara polimorfismo, lo que representaba menos del 2% de las entradas. Casi con total certeza, esta cifra no representa la verdadera frecuencia con la que aparecen las formas polimórficas en las moléculas orgánicas en las condiciones normales de laboratorio. La razón más común para el uso de los análisis de rayos X es la de determinar inequívocamente una estructura molecular, por lo que la mayoría de los investigadores eligen la muestra cristalina más conveniente disponible y no están interesados en la búsqueda de otros polimorfos. Sin embargo, a pesar de la frecuentemente expresada opinión de que el número de polimorfos declarados es proporcional al tiempo empleado en su búsqueda (McCrone, 1965), las intensas investigaciones en ciertos compuestos farmacéuticos normalmente revelan sólo unos pocos polimorfos, a pesar de que las moléculas son flexibles y tienen muchas posibilidades de formar enlaces de hidrógeno. Además, muchos compuestos comunes que se están estudiando desde hace casi 100 años muestran sólo un polimorfo, como es el caso del naftaleno, del ácido benzoico, de la sacarosa, etc. Los cálculos de la energía reticular de la estructura cristalina de los polimorfos revelan un estrecho rango de energía entre las formas; el mayor estudio realizado hasta la fecha, llevado a cabo para 204 pares de polimorfos de la CSD (Gavezzotti y Filippini, 1995), discute las diferencias en la densidad, en la energía de empaquetamiento estático, en la entropía vibracional de la red y en otras propiedades cristalinas.

El término “pseudopolimorfismo” se ha usado para describir los casos en los que un compuesto existe en una forma cristalina pura, y también cocrystalizado con moléculas del disolvente. Un análisis de la hidratación en cristales orgánicos

(Desiraju, 1991*b*) mostró que la aparición de una forma hidratada aumentaba conforme se incrementaba la proporción dador/aceptor en el enlace de hidrógeno. Un estudio de aproximadamente 5000 estructuras de la CSD que contenían moléculas de disolventes comunes (Nangia y Desiraju, 1999) mostró la importancia del reconocimiento multipunto de los disolventes para enlaces de hidrógeno tanto fuertes como débiles. Un estudio comprensivo de las estructuras de la CSD que contenían cualquier tipo de molécula de disolvente (Görbitz y Hersleth, 2000) identificó más de 300 tipos de disolventes, dando las frecuencias relativas con las que aparecían.

1.6.3.7 Predicción de la estructura cristalina

Durante muchos años, un grupo relativamente pequeño de investigadores ha estado persiguiendo la meta de la predicción *ab initio* de la estructura cristalina de una molécula partiendo sólo de su diagrama químico (ver Gavezzotti, 1994, 1998; Verwer y Leusen, 1998). El método de aproximación predominante consiste en calcular la energía reticular para un rango de grupos espaciales con una molécula en la unidad asimétrica, usando normalmente un modelo molecular rígido. Se generan entonces muchos miles de estructuras cristalinas posibles y se usan varios métodos de optimización matemática para buscar el mínimo global en la superficie de energía. Esta aproximación está mostrando señales de éxito y, recientemente, el CCDC ha organizado pruebas *a ciegas* para la predicción de pequeñas estructuras orgánicas inéditas (Lommerse *y otros*, 2000; Motherwell *y otros*, 2002). Sin embargo, aunque a veces pueden predecirse las estructuras con una exactitud razonable, ni una sola de estas metodologías es fiable. El problema básico es que hay muchos más mínimos a menos de 5 kJmol^{-1} del mínimo global de los que cabía esperar incluso hace 10 años. Esto lleva a un conjunto típico de unos 30 posibles polimorfos para una molécula de 25 átomos, y las decisiones finales para elegir una sola energía a una temperatura de cero se toman a menudo sobre la base de unas diferencias energéticas muy pequeñas (en torno a 0.5 kJmol^{-1}). Se han realizado esfuerzos para incluir la entropía debida al movimiento vibracional en el cristal, lo que mejora a veces la predicción (van Eijck *y otros*, 2001). Estas pequeñas diferencias en la energía dependen críticamente de los campos de fuerzas usados, y estos campos de fuerzas se han construido y validado a lo largo de los años

tomando como base la CSD (ver, por ejemplo, Filippini y Gavezzotti, 1992*b*).

El cálculo del componente electrostático de la energía ha mostrado a menudo que una descripción multipolo es más efectiva en la reproducción de estructuras de la CSD (Beyer *y otros*, 2001). Otros métodos han tratado de usar:

- potenciales estadísticos basados directamente en las distancias interatómicas entre las moléculas de la CSD (Hofmann y Lengauer, 1997)
- un ajuste directo de las curvas de la frecuencia de la distancia (Motherwell, 2001)
- cuasipotenciales basados en el ajuste de los diagramas de dispersión de la interacción entre los grupos químicos de la base de conocimiento *IsoStar* del CCDC (Lommerse *y otros*, 2000)

Pero ninguno de estos métodos tuvo éxito en las pruebas *a ciegas*. Sin embargo, como la cristalización es un proceso cinético, se acepta que la única estructura de menor energía libre predicha puede no ser siempre el polimorfo observado experimentalmente. Así, el futuro papel de la CSD puede ser el de proporcionar mayores muestras de estructuras cristalinas para moléculas similares a la molécula designada y así argumentar, a partir de patrones y de analogías entre estructuras, que ciertos miembros de un conjunto de posibles estructuras de baja energía es más probable que existan bajo unas condiciones dadas.

1.6.3.8 Determinación de la estructura a partir de los datos de polvo

La determinación de estructuras cristalinas a partir de datos experimentales de difracción de rayos X de polvo se ha vuelto rutinaria para moléculas pequeñas, para las que puede construirse un modelo molecular razonable. Se generan las estructuras de muchos miles de cristales de prueba, y el patrón de polvo simulado se ajusta frente a los datos observados con técnicas de optimización global similares a las usadas en la predicción de la estructura cristalina, con el fin de encontrar el mejor ajuste (David *y otros*, 1998; Engel *y otros*, 1999). La CSD proporciona una valiosa comprobación de la geometría del modelo molecular y, a veces, como ocurre por ejemplo en el caso de los sistemas cíclicos y acíclicos flexibles con energías internas casi idénticas, las configuraciones observadas en la

CSD pueden proporcionar una guía esencial sobre las áreas del espacio paramétrico que deben considerarse. Debido a la pérdida de datos que se produce en el patrón de polvo monodimensional por el solapamiento de los máximos, hay un límite de complejidad en el máximo número de parámetros independientes (en torno a 20) que pueden ajustarse inequívocamente a los datos. Así, para el caso de una sola molécula en la unidad asimétrica con 6 grados de libertad, normalmente pueden tratarse como parámetros cerca de 14 ángulos de torsión. Un ejemplo en el que sólo era posible llegar a la solución restringiendo ciertos ángulos de torsión a los rangos observados en la CSD es el hidrocloreuro de tetracaína (Nowell *y otros*, 2002), con un total de 18 parámetros, de los que nueve eran ángulos de torsión, dos de los cuales se restringieron.

1.6.3.9 Análisis de la precisión estructural

Taylor y Kennard (1986) realizaron el primer trabajo que usó datos de la CSD para evaluar la precisión de las estructuras cristalinas. Estos investigadores realizaron un análisis estadístico de 100 estructuras que habían sido determinadas independientemente por dos grupos de investigación diferentes, y en las que se había descubierto que las incertidumbres estándar de los parámetros de la celda estaban muy infravalorados. Allen *y otros* (1995*a,b*) volvieron a este tema con la exploración de la relación entre la incertidumbre isotrópica media para diferentes tipos de elementos y varios factores experimentales y estructurales como, por ejemplo, el factor R , el número atómico del elemento, el número atómico del elemento más pesado de la estructura, el límite del $\sin \theta/\lambda$ para la recopilación de datos, etc. y se usaron múltiples técnicas de regresión lineal para generar ecuaciones predictivas para los valores de la incertidumbre isotrópica. Aunque las predicciones eran generalmente aceptables para un amplio rango de estructuras, los descubrimientos de Taylor y Kennard (1986) se vieron reforzados.