الفصل السابع عشر

قواعد بيانات كريستالوجرافية
Crystallographic databases

إن قواعد بيانات الحاسوب هي جمع لبنود معلومات بتركيب وتصميم عام. إن لها عدد من المزايا أكثر من صور حفظ المعلومات ورقياً أو أي صورة أخرى. تشمل هذه السهولة الأكبر في حفظها وتحديثها، إمكانية التقويم الأوتوماتيكي للبنود عند إضافتها، طرق أقوى وأسرع في البحث عنها للبنود ذات الأهمية (بافتراض أن برنامج حاسوب مناسب يكون متاحاً)، تخزين محكم وتوزيع بسيط وإمكانية الحصول على بنود مختارة في صورة إلكترونية لمزيد من التحليل.

إن قاعدة بيانات أي حاسوب له مكونين، محتويات قاعدة البيانات نفسها وبرامج وأنظمة تشغيل الحاسوب software لإجراء مسح وتحليل للبيانات.

إن قواعد البيانات مهمة في مجالات كيميائية عديدة وبعضها معروف جيداً، وهي تشمل قواعد بيانات ببليوجرافية bibliographic database، مثل نسخة حاسوبية من Science Citation Index و Chemical Abstracts؛ قواعد بيانات للكتالوجات الكيميائية التجارية ومعلومات أمان؛ مدى من قواعد بيانات طيفية IR، NMR وغيرها، تجميع بيانات ثرموديناميكية وقواعد بيانات للتفاعلات الكيميائية للاستخدام في التشييد بمساعدة الحاسوب.

إنها بصفة خاصة تكون مناسبة جداً للتطبيقات الكريستالوجرافية بسبب الطبيعة النظامية للمعلومات التركيبية الكريستالوجرافية. توجد خمس قواعد بيانات أساسية تستخدم حالياً على نطاق دولي. إنها تكون محفوظة ومطورة بواسطة هيئات مختلفة، وتكون موزعة للأفراد ومراكز الحسابات الدولية والمحلية، وقد يمكن الوصول إليها إما موقعياً على جهاز حاسوب مكتبي أو من خلال اتصالات الشبكة العنكبوتية. يكون برنامج البحث مفصلاً لقاعدة بيانات معينة. في بعض الأحيان تكون من خلال سطح بيني لبرنامج تصفح موقع شبكي.

(١٧,١) قواعد بيانات تركيبية متاحة Available structural databases

ملف بيانات بلورية للفلزات (CRYSTMET يُعرّف أيضاً على أنه MCDF أو MDF) يكون محتفظاً بها بواسطة National Research Council of Canada. إنه يحتوي على معلومات عن الفلزات والسبائك ويتضمن بعض المركبات الثنائية مثل الهيدريدات والأكاسيد. يبدو هذا كتخصص محدود للغاية، لكن في الحقيقة فإنه يوجد أكثر من 64,000 مدخل بعام 2001. لكل مُدخل بيانات ببليوجرافية (مؤلف، مرجع، اسم مركب أو معدن، صيغة) وبيانات عددية (بارامترات خلية، زمرة فراغية، إحداثيات ذرية وبارامترات إزاحة لو كانت هذه تكون معروفة).

إن قاعدة بيانات التركيب البلوري اللا عضوي (ICSD Inorganic Crystal Structure database هي مسئولية معهد Gmelin institute و Fachinformationszentrum في كارلسروه Karlsruhe، ألمانيا؛ إنها قد أنشأت أساساً بجامعة بون. إنها تتكون من تراكيب مواد لا عضوية ومعادن، لا تحتوي على كربون عضوي، بسبب أنها قد صممت بصفة خاصة لاستكمال قاعدة البيانات الأقدم بكامبريدج Cambridge؛ التعريف المقسِّم لقاعدتي البيانات هاتين هو اختيار مطلق قليل وهناك درجة صغيرة من التداخل في

محتوياتهما. تكون معظم التراكيب في ICSD لا جزيئية والاتصال الكيميائي يكون أكثر صعوبة للتعريف والتخمين لهذه المركبات عنها للجزيئات المترابطة تساهمياً. نجعل هذه وسائل لبحث ICSD أقل مرونة وقوة بشكل عام، لكن إحدى طرق البحث هي عن شكل مألوف للموقع البيني web interface. إن المعلومات المقدمة تكون مشابهة لتلك الموجودة في MCDF. تكون هناك مركبات قريبة الصلة ببعضها تشمل تراكيب متماثلة التبلور (ايزومورفية isomorphous)، محاليل صلبة ومتعددة الصور وأيضاً تحديدات مضاعفة للتراكيب تحت ظروف مختلفة باستخدام تشعيعات مختلفة أو بواسطة مجموعات بحثية مختلفة، بعام 2001، كان هناك ما يقرب من 60,000 إدخال.

إن القاعدة الأكبر والأكثر تطوراً واتساعاً هي قاعدة البيانات التركيبية بكامبريدج (CSD Cambridge Structure Database) الخاصة بمركز البيانات الكريستالوجرافية بكامبريدج Cambridge Crystallographica Data Centre (CCDC). إن نظامها هو مدى كامل من مركبات عضوية، عضو معدنية ومركبات تناسقية لفلز مفهومة بشكل واسع على أصناف جزيئية لكن مع استبعاد الجزيئات البيولوجية الضخمة. مع حوالي 240,000 إدخال بعام 2001، فإن نموها يستمر كوسيلة تحسينات حديثة في الكريستالوجرافية تؤدي إلى جمع بيانات سريعاً وتحليل تركيب سريع. بالإضافة إلى بيانات ببليوجرافية وعددية مشابهة إلى ISCD (لكن بدون بارامترات إزاحة) يكون كل إدخال مشفراً مع ترابطية جزيئية يسمح بتسهيلات بحث قوية معتمدة على شظايا تركيبية.

تقدم الجزيئات البيولوجية الكبيرة بواسطة Protein Data Bank (PDB) الذي يحتوي على أحماض نووية وبروتينات لقد كان مطوراً بواسطة Brookhaven National Laboratory USA لكن تم حديثاً إحرازه بواسطة Research Collaboratory for Structural Bioinformatics. رغم صغر عدد الإدخالات، فإنه أيضاً النمو الأسرع للبيانات التركيبية

مع 14,000 إدخال بعام 2001، ضعف العدد قبل ذلك بثلاث سنوات. إن قاعدة بيانات منفصلة للأحماض النووية NDB توجد الآن مع 800 إدخال تقريباً.

قاعدة البيانات الخامسة هي مختلفة إلى حداً ما: ملف the Crystal Data Identification File (CDIF) بواسطة US National Institute of Science and Technology. تخزن هذه معلومات خلية وحدة التركيب والزمرة الفراغية، حتى حينما لا يكون التركيب الكامل معروفاً، وهي مصدر مفيد جداً لاختبار خلايا وحدة التركيب المتحصل عليها على جهاز قياس الحيود لكي نتجنب إضاعة الوقت لجمع بيانات لمركب معروف (مثل مادة البدء أو ناتج غير متوقع من تفاعل)؛ انظر الفصل الخامس. بعام 1999 كان هناك أكثر من 230,000 إدخال مع احتياطي من إدخال جديد في انتظار للتهيئة.

في UK معظم قواعد البيانات هذه تكون متاحة بالمجان للباحثين الأكاديميين من خلال EPSRC-funded Chemical Databank Service at Daresbury Laboratory (http://cds.dl.ac.uk/cds/cds.html) ترتيبات مشابهة للوصول لكل أو بعض من قواعد البيانات موجودة في دول أخرى.

(١٧,٢) محتويات قاعدة البيانات التركيبية بكامبريدج

Contents of the Cambridge Database

لقواعد البيانات البلورية الخمس بعض السمات العامة، لكن لكل خصائصه المميزة. حيث أن CSD هي الأوسع استخداماً في الكريستالوجرافيا الكيميائية فسوف نركز على بعض المظاهر الخاصة بطبيعتها واستخدامها.

كل تركيب في CSD يكون معرفاً بكود مرجعي منفرد (REFCODE) يصنف عندما يضاف إلى قاعدة البيانات. إن REFCODE هو تسلسل من ست حروف التي لا تحمل بصفة عامة معنى معين؛ يمكن إضافة رقمين عددين في بعض حالات لتدل على

دراسة ثانية أو دراسة أبعد لنفس المادة أو شكل بلوري مختلف (متعدد الأشكال polymorph أو تذاوب solvate مختلف)، إنها تعمل كمرقم مناسب.

توصف المعلومات لكل تركيب بواسطة مُظّهرات CSD's حيث كونها تحت عناوين ثلاثة. تكون البيانات ثلاثية الأبعاد هي بارامترات خلية وحدة التركيب، زمرة فراغية، وإحداثيات ذرة، التي منها يمكن حساب هندسة جزيئية وبين جزيئية مفصلة. تكون المعلومات ثنائية الأبعاد هي ترابطية كيميائية مخزنة، مظهرة أي الذرات تكون متصلة ببعضها، بالإضافة إلى أنواع الذرة والشحنات، شفرات لأنواع مختلفة من روابط ومخطط للصيغة التركيبية الكيميائية مبيناً هذه الترابطية؛ تكون التراكيب البوليمرية معلمة. يشار إلى المعلومات الأخرى على أنها "أحادية-البعد": وهي تتضمن كل البيانات الببليوجرافية بالإضافة إلى بنود عددية منفردة مثل درجة الحرارة، كثافة محسوبة، عامل R؛ بنود نصية مثل مسمى المركب واللون وتعليقات على السمات مثل خلل وأخطاء في البيانات المنشورة.

إن الغالبية الأوسع من الإدخالات CSD تأتي من المسح الأدبي المنشور، شاملة تداخل مع عدد كبير من المجلات. في السابق كان لابد لكل إدخال أن يتم إدراجه بواسطة لوحة مفاتيح ومخطط الترابطية المولد يدوياً. إن ماسح الحاسوب قد جعل هذا أكثر سهولة لكنها كانت بأي حال خالية من الأخطاء. بشكل أكثر حداثة فإن إدخال وتبني واسع الانتشار من CIF قد أدى إلى زيادة التشغيل الآلي للعملية. بعض المجلات تبعث بإرسالات CIF إلى CCDC، الأخرى تطلب من المؤلفين أن يفعلوا هذا قبل إرسال المخطط الكتابي للنشر، بينما لا يكون لمجلات أخرى طريقة محددة. يكون من الممكن للباحثين أيضاً أن يقوموا بإيداع النتائج التركيبية مباشرة مع CCDC لتضمينها في CSD بدون نشر رسمي في مجلة؛ إن قاعدة البيانات في حد ذاتها شكل من أشكال النشر بالطبع. يتم مراجعة كل الإدخالات جيداً للانسجام، على سبيل المثال بحساب الهندسة

الجزيئية من الإحداثيات وبارامترات الخلية واختبارها ضد البارامترات الهندسية المزوّدة في النشر أو في CIF وبمقارنات مثل الصيغة الجزيئية المحددة مع قائمة الذرات. أي أخطاء يتم تصحيحها لو أمكن أو يكون معلماً، وتكون مدخل قاعدة البيانات في بعض الحالات أكثر دقة عن النشر الأصلية.

(١٧,٣) البحث في CSD Searching the CSD

تكون كل البنود المخزنة لكل إدخال متاحة كقاعدة للبحث. هكذا يكون من الممكن أن نجد إدخالات معتمدة على معلومات ببليوجرافية أو كيميائية على بيانات كريستالوجرافية منفردة مثل بارامترات خلية، ظروف معملية، دقة النتائج، وجود خلل، وبيانات عديدة أكثر. الشيء المفيد بوجه خاص هو إمكانية البحث لتراكيب تحتوي على شظية معينة التي تشيد نوعياً في مساحة رسم (مع لوحة ألوان مطبوعة أو مجموعات شظية متاحة لتسهيل هذا التشييد). يمكن وضع حدود على بارامترات هندسية مثل المسافات والزوايا، لكي يتم إدخال أو استبعاد تراكيب من أنواع مختلفة وقوائم من نتائج بحث يمكن أن ترشح وتنقى يدوياً.

من التراكيب المستعادة (يشار إليها على أنها "صدمات" "hits") التي قد تكون عدد قليل أو كثير لأي بحث معطى أو توليفة من بحوث، يمكن توليد قوائم بسيطة، هندسة يمكن حسابها، مخطط يمكن الحصول عليه، يمكن للتراكيب الجزيئية أن تعالج على الشاشة وتحليل إحصائي بسيط أو معقد يمكن عمله للنتائج.

حتى وقت حديث كانت CSD متاحة بشكل أولي على أنظمة حساب معتمد على UNIX. يوجد الآن نسخة متقدمة من برامج البحث التي يمكن تشغيله تحت نسخ متنوعة من Windows على PCs. إن النسخة الحالية من برنامج الحاسوب بالإضافة إلى

قاعدة البيانات تتطلب أكثر من 500 Mbyts تخزين، لا توجد مشكلة على الحاسوبات المتاحة فعلياً.

هناك استخدامات عديدة لقواعد البيانات. إن الأبسط هو اختبار التركيب لمركب معين، لكي نكشف فيما لو أنه بالفعل قد تم تحديده، لكن يمكن بسهولة تمديد هذا للنظر في مركبات ذات علاقة. لمجموعة من تراكيب يمكن فحص العلاقات والاتجاهات. يمكن الحصول على الهندسة لشظية جزيئية خاصة (إما من تركيب منفرد أو متوسط النتائج من تراكيب عديدة)، للاستخدام في الحسابات النظرية، النمذجة الجزيئية، أو اشتقاق نموذج لبحث باترسون أو تنقيح جسم-جاسئ. إن الاتجاهات والنماذج في التراكيب يمكن فحصها، على سبيل المثال في تطابقات حلقات، ترابط هيدروجيني، التأثيرات التركيبية لمستبدلات مختلفة، التداخلات بين الجزيئية...الخ، وهذه قد تكون فحوصات بحث حقيقية لصالحها.

إن قواعد البيانات التركيبية توفر طريقة ملائمة لثراء الأبحاث المنشورة خاصة لو أنها تكون مستخدمة بترافق مع قواعد بيانات ببليوجرافية. إن توافرها ومصداقيتها قد أحدثت صدى ضخم على أبحاث علم البلورات عبر السنين وقد تم تحديدها كأدوات أساسية للمادة.