

A propos de l'utilisation des bases de données de réactions

Jacques Coste* maître de conférences, Olivier Gien** ingénieur, Andreas Dietz* post-doct,
Claude Laurenço* directeur de recherche CNRS

Summary : *About using reaction databases*

Reaction databases are very useful tools for the synthetic chemist. However, some aspects such as completeness, retrieval of synthetic methods or mapping have to be taken into account in view to use them successfully. This note discusses these problems on the basis of some examples taken from the interrogation of both MDL (ISIS) and Beilstein CrossFireplusReactions databases.

Mots clés : *Bases de données de réactions, exhaustivité, méthodes de synthèse, mapping.*

Key-words : *Reaction databases, completeness, synthetic methods, mapping.*

L'interrogation des bases de données de réactions chimiques (BDR) trouve différentes applications en synthèse organique. Ces bases permettent, en particulier, de retrouver des réactions ou des méthodes de synthèse inconnues de l'utilisateur, de valider une idée en vérifiant que des réactions similaires sont connues ou encore de retrouver des conditions expérimentales pour une réaction particulière. Enfin, faute d'avoir à sa disposition des systèmes informatiques d'aide à la synthèse performants, le chimiste peut se servir de ces bases lors de l'établissement du plan de synthèse d'une molécule cible.

Les BDR sont des outils couramment employés par les chimistes organiciens. Si, en 1994, E. Zass [1] pouvait regretter qu'elles soient peu utilisées dans le cadre universitaire, la situation a considérablement évolué. Grâce à des tarifs académiques, l'interrogation des bases des sociétés MDL (exploitées par le système ISIS) [2] et Beilstein (CrossFireplus Reactions) [2] est possible sur Titane [3], serveur national de bases de données chimiques ; des situations semblables existent dans de nombreux pays [4].

Nous nous intéressons aux BDR pour plusieurs raisons : d'une part, dans le cadre de la construction de la base de connaissance d'un système d'aide à la conception de plans de synthèse (RESYN) [5] ; d'autre part, en tant que laboratoire partenaire du serveur Titane ; enfin, pour l'un d'entre nous, dans le cadre de l'enseignement de la synthèse organique.

Les BDR sont constituées de la description (schéma réactionnel) de réactions particulières auxquelles sont associées des informations (conditions réactionnelles, résultats expérimentaux, références bibliographiques). Notre propos sera limité aux bases de données de réactions ISIS [2b] et Beilstein [2b] qui sont accessibles sur le serveur Titane [3]. ISIS est constituée de réactions sélectionnées (actuellement 750 000) ; Beilstein a une vocation exhaustive [6], cette base contient 10 millions de réactions dont 5 millions sont accessibles par des requêtes par schéma réactionnel (graphical reactions) et 5 millions (nongraphical reactions) ne le sont pas encore par ce type de requête.

Les requêtes peuvent avoir différents objets (réaction, réactant, produit, réactif, catalyseur, conditions, résultats expérimentaux, etc.) et être formulées de différentes façons (schéma réactionnel par structure exacte ou sous-structure, formule brute, nom, valeurs numériques).

Nous n'aborderons ici que le cas des requêtes sous-structurales par schéma réactionnel. Si la puissance des BDR est, en grande partie, due à cette possibilité d'interrogation, la pratique des bases montre l'existence de certaines contraintes et limites dont l'utilisateur doit être averti. La documentation disponible étant peu explicite à ce sujet, notre but est de faire part de quelques problèmes auxquels nous avons été confrontés. A notre connaissance, seul un ouvrage récent aborde, dans un de ses chapitres [7], certains des points qui sont soulignés dans ce manuscrit [8]. Concernant Beilstein, il ne sera pas question de l'utilisation de cette base pour la recherche des composés définis, de leur préparation ou de leurs propriétés : de ce dernier point de vue, Beilstein est une base de données unique.

Nous allons, sur un exemple, mettre en évidence les problèmes qui seront ensuite étudiés plus en détails. Considérons la recherche d'exemples de la réaction figurée sur le schéma 1 [9].

Schéma 1 ^(a)



^(a) Ce schéma ne comporte ni le mapping [10] ni les contraintes qui ont été précisées dans les requêtes [9].

* GDR 1093 - UPR 9023, CCIPE, 141, rue de la Cardonille, 34094 Montpellier Cedex 5.
Fax : 04.67.54.24.32.

E-mail : coste@ficco.ccipe.montp.inserm.fr

** Sanofi Chimie, Route de Gap,
04200 Sisteron.

On obtient pratiquement le même nombre de réponses avec les deux bases (tableau I) et, sur les 5 méthodes de synthèse retrouvées dans chaque base, 4 sont présentes dans les deux listes de réponses. Cependant, alors que 7 références retrouvées avec Beilstein (base à vocation exhaustive) ne le sont pas avec ISIS (réactions sélectionnées), il est plus inattendu de retrouver, avec ISIS, 7 références qui ne le sont pas avec Beilstein. Deux de ces références concernent des « nongraphical reactions » ; une est extraite du journal *Synlett* ; une autre concerne une réaction multi-étapes. Enfin, pour les trois dernières références, les réactions correspondantes sont indexées dans Beilstein sans mapping [10], elles ne sont pas retrouvées car le mapping était précisé dans la requête.

Ces résultats mettent en évidence trois problèmes : (i) l'exhaustivité des bases, (ii) la présence (ou l'absence) des réactions multi-étapes, (iii) les problèmes de mapping.

Exhaustivité des bases

Aucune source de données (informatisée ou papier) n'est exhaustive au sens strict. C'est, par nature, le cas d'ISIS qui est constituée d'une sélection de réactions, mais c'est aussi le cas de Beilstein. En effet, Beilstein contient 7 millions de composés alors que 13 millions de composés organiques sont indexés dans les *Chemical Abstracts* [11]. Par ailleurs, lors d'une requête par schéma réactionnel, on ne pourra accéder qu'aux 5 millions de « graphical reactions ». De plus, s'il est indiqué dans la documentation de Beilstein que depuis 1980 la plupart des composés organométalliques ne sont pas indexés mais le sont dans la base Gmelin, on est par contre mal informé du nombre et des titres des journaux qui sont analysés pour constituer la base. Warr donne les indications suivantes [12] : jusqu'en 1980, Beilstein tentait de couvrir toute la littérature ; dans les années 80, 85 journaux étaient analysés ; ce nombre a augmenté, après 1990, pour atteindre actuellement 180. Sans tenir compte du problème de l'indexation des brevets, il est clair que tous les journaux intéressants la chimie organique ne sont pas analysés, ou ne l'ont pas été dans le passé. Par exemple, il

n'y a pas de référence postérieure à 1979 pour *Int. J. Pept. Protein Res.*, on trouve une seule référence antérieure à 1996 pour *Bioorg. Med. Chem. Lett.* (publié depuis 1990) et il n'y a pas de référence antérieure à 1997 pour *Synlett* (publié depuis 1990) [13].

Types des réactions qui sont décrites dans les bases

Un des intérêts des bases devrait être de permettre de retrouver non seulement des réactions particulières, mais aussi des méthodes de synthèse, une méthode de synthèse étant caractérisée par la transformation structurale générique qu'elle permet d'opérer, indépendamment du nombre d'étapes et du mécanisme des réactions impliquées. La question se pose de savoir comment cette notion de méthode de synthèse est prise en compte dans les BDR : peut-

on, par exemple, accéder au bilan global d'une séquence réactionnelle, ou seulement aux étapes intermédiaires ?

Dans l'exemple ci-dessus (schéma 1, tableau I), une réaction multi-étapes retrouvée dans ISIS ne l'est pas dans Beilstein où elle n'est décrite qu'étape par étape.

La méthode bien connue de Corey-Seebach (schémas 2-a,b) permet le passage d'un aldéhyde à une cétone [14]. La recherche d'exemples de cette méthode (schéma 2-b), formulée par la requête du schéma 2-c, fournit une seule réponse dans ISIS [15a], réponse qu'il faut retrouver dans une liste de 1 130 réponses [15b]. Dans Beilstein, avec une formulation semblable ($R' = \text{alkyl}$), on n'obtient pas de réponse correspondant à la méthode recherchée. Cependant, dans les deux bases, on trouve de nombreux exemples pour chacune des étapes de la méthode.

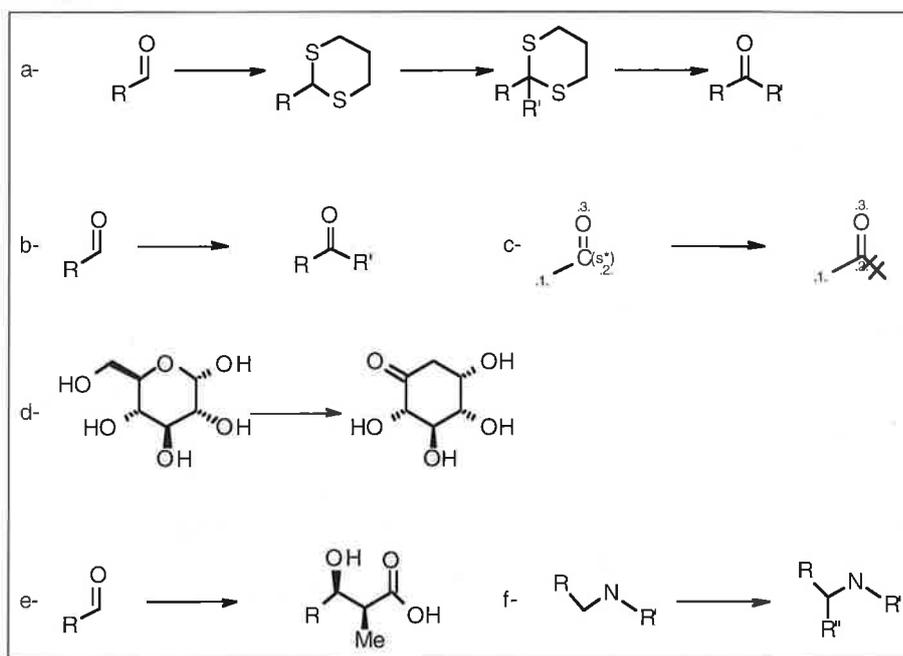
La méthode (multi-étapes) du schéma 2-d, basée sur la réaction de Ferrier

Tableau I.

Base	Réponses	Références	Recouvrement des références ^(a)	Méthodes de synthèse
ISIS	13	12	5	5
Beilstein	12	12		5

^(a) Nombre des références identiques qui sont présentes dans les deux bases.

Schéma 2^(a).



^(a) Dans le schéma 2-c, toutes les substitutions sont permises, sauf pour l'atome de carbone noté C(s*) qui porte un atome d'hydrogène. Les atomes qui sont mappés portent le même numéro.

⊕ : indique que la liaison doit être formée.

[16], n'est décrite, dans les deux bases, qu'étape par étape. La méthode (multi-étapes) d'Evans (*schéma 2-e*) [17] n'est présente, en tant que telle, que dans ISIS. A une requête concernant l'alkylation en α d'amines secondaires (*schéma 2-f*), Beilstein donne 2 méthodes multi-étapes au lieu de 7 retrouvées avec ISIS.

ISIS est qualifiée de « database of synthetic methodology » [18] et est réputée pour contenir des « synthetic methods » [18]. La documentation de Beilstein est peu explicite. Pour les données jusqu'en 1959, la construction de la base [19] a été faite à partir du Beilstein Handbuch dont l'objet principal est la molécule ; on conçoit que, jusqu'à cette date, la base ne contienne, principalement ou exclusivement, que des réactions mono-étape. Après cette date, et plus particulièrement depuis 1980, date à partir de laquelle les informations ont été directement tirées de la littérature primaire [19], on peut par contre concevoir que des réactions multi-étapes aient été indexées : effectivement, c'est parfois le cas.

Les exemples ci-dessus montrent que, lors de la recherche de méthodes de synthèse dans les BDR, il faut tenir compte de ce que les méthodes multi-étapes peuvent ne pas être indexées, en tant que telles, en particulier dans Beilstein, mais aussi dans ISIS.

Problème du mapping

Dans les deux bases, lorsqu'on indique dans le schéma de la requête un (des) réactant(s) et un (des) produit(s) de réaction, il faut, en principe, mapper [10]. Dans le cas d'ISIS, on peut utiliser la fonction de mapping automatique (Automap) ou faire un mapping manuel total ou partiel. Pour Beilstein, le mapping est uniquement manuel et est laissé au choix de l'utilisateur.

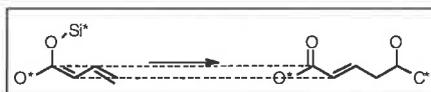
Faut-il mapper ou ne pas mapper ? Les deux modes sont possibles, cependant, dans le cas d'une requête par sous-structure, ne pas mapper conduit le plus souvent à l'obtention d'un ensemble de réponses très important et globalement non pertinent (*cf.* ci-dessous). Toutefois, la question du mapping se pose car, comme nous allons le voir, les réactions sont parfois indexées sans mapping et parfois avec un mapping chimiquement non pertinent.

Absence de mapping

Dans l'exemple donné ci-dessus (*schéma 1, tableau 1*), 3 des réactions trouvées avec ISIS ne l'étaient pas avec Beilstein faute d'être indexées dans la base avec indication du mapping. Une interrogation sans mapping ne résout pas le problème car on obtient alors une liste de 311 réponses dont la plupart ne correspondent pas à la question que l'on souhaitait poser. Il est fastidieux de rechercher les bonnes réponses dans cette longue liste : dans un tel cas, l'utilisateur fera probablement une requête avec mapping et il perdra de l'information.

Dans d'autres cas, une requête sans mapping s'impose. Par exemple, dans Beilstein, on n'obtient pas de réponse à la requête correspondant au *schéma 3*, si elle est posée en précisant le mapping tel qu'il est indiqué. Par contre, sans le mapping, on obtient 6 réponses pertinentes. Dans la base, aucune de ces 6 réactions n'est mappée.

Schéma 3^(a).



^(a) Les atomes qui sont mappés sont joints par un trait pointillé. Sur les atomes affectés d'une étoile (*), toutes les substitutions sont permises.

Le problème d'absence de mapping se retrouve également dans ISIS. Une exploration des bases donne les résultats suivants [20]. Dans ISIS, l'absence de mapping n'apparaît que dans la base ChemInform (la plus importante des sous-bases d'ISIS) où elle peut être estimée à 14,5 % ; cela conduit à un taux de 9 % pour l'ensemble de la base. Dans le cas de Beilstein, on peut estimer ce taux à 30 %.

Mapping « faux »

Un autre problème lié au mapping concerne l'existence de réactions qui ont été saisies dans les bases avec un mapping qui ne reflète pas le mécanisme de la réaction qui est décrite.

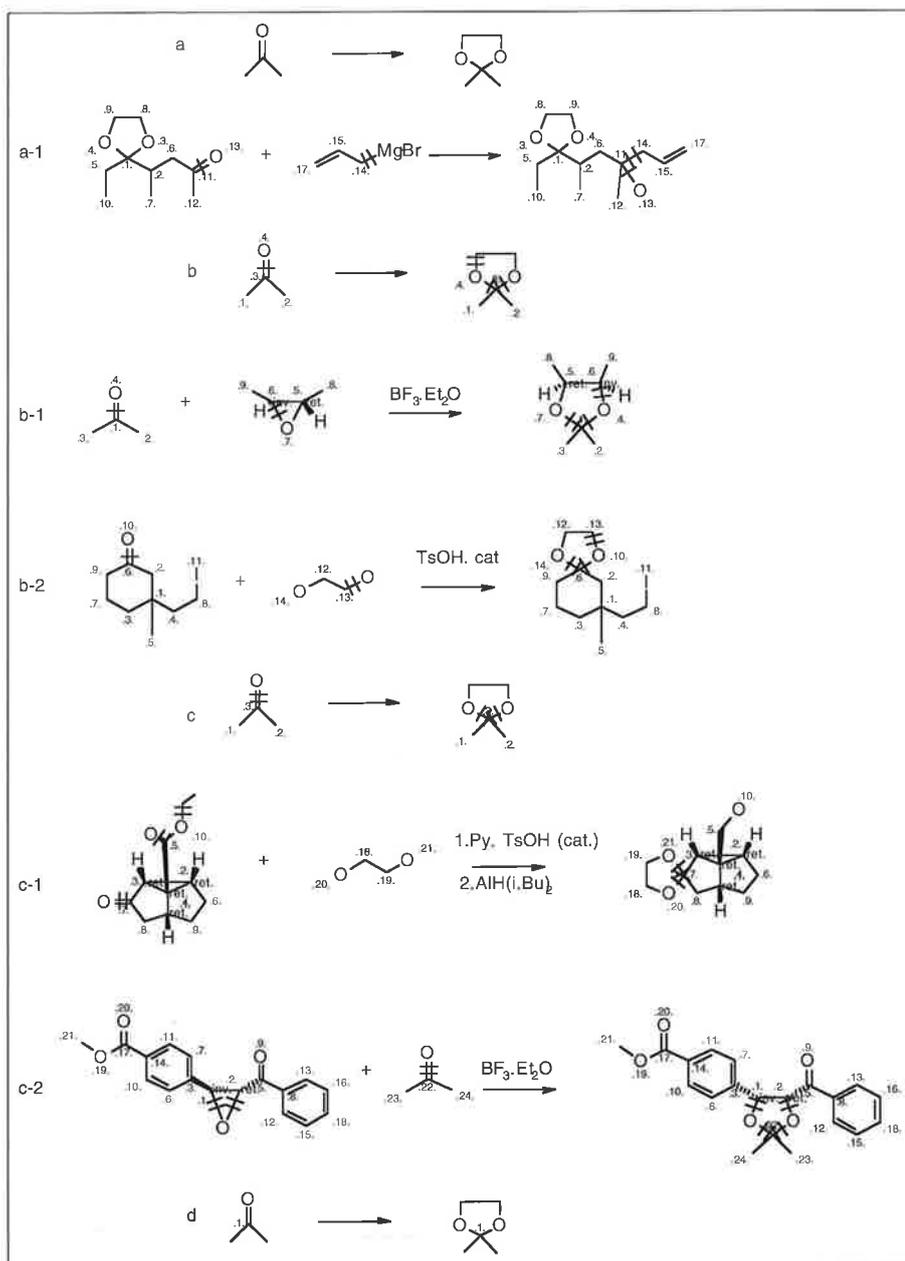
Considérons la recherche de méthodes de synthèse de cétals à partir d'une cétone. Si, dans ISIS, la requête est faite sans mapping, comme sur le *schéma 4-a*, on obtient 3 104 réponses dont un grand nombre, comme celle du

schéma 4-a-1, ne concerne pas la formation d'un cétal à partir d'une cétone. Il est normal d'obtenir cette réponse : un des réactants contient une cétone, le produit de la réaction une fonction cétal et la requête ne précisait pas que le cétal devait provenir de la cétone. Il convient donc de mapper.

A partir de 4-a, par utilisation de l'Automap, on obtient le mapping indiqué sur le *schéma 4-b*. Avec cette requête, on accède à une liste de 495 réponses (liste 1) contenant des réactions de formation d'un cétal par réaction entre une cétone et un époxyde (39 réponses, par ex., *schéma 4-b-1*) et, principalement, des réactions entre une cétone et un diol (*schéma 4-b-2*). Si les réactions cétone + époxyde sont indexées avec un mapping en accord avec le mécanisme de réaction, cet accord n'existe pas pour les réactions cétone + diol. Avec un mapping manuel correspondant au mécanisme de la réaction d'acétalisation d'une cétone par un diol (*schéma 4-c*) [21], on obtient une liste de 401 réponses (liste 2). Cette liste contient principalement des réactions cétone + diol (*schéma 4-c-1*), indexées dans la base avec un « bon mapping » mais aussi 6 réactions cétal + époxyde (*schéma 4-c-2*) qui ont été indexées avec un mapping *a priori* « faux » (d'après ce mapping, les quatre liaisons carbone-oxygène du cétal seraient formées et les atomes d'oxygène du cétal ne proviendraient ni de la cétone ni de l'époxyde).

En utilisant la fonction Logic, on constate qu'une seule réponse est commune aux deux listes. Pour avoir toutes les réponses, il faut donc interroger successivement avec les deux types de mapping (4-b et 4-c), puis faire la somme des réponses des deux listes. Appelons liste 3 (895 réponses) la liste ainsi obtenue.

On pourrait penser accéder à la liste 3 en ne précisant qu'un « mapping minimum ». Avec une requête selon le *schéma 4-d*, où seul le carbone du carbonyle est mappé, on obtient une liste de 1 282 réponses. Cette dernière liste contient, outre les réponses de la liste 3, 388 nouvelles réactions dont 16 sont mappées et 372 ne le sont pas ; il s'agit de réactions d'acétalisation non mappées, de réactions d'acétalisation avec mapping erroné et, principalement, de

Schéma 4^(a).

^(a) Les atomes qui sont mappés portent le même numéro. Pour les schémas a, b, c et d, toutes les substitutions sont permises sur les atomes de carbone. Les schémas a-1, b-1, b-2, c-1 et c-2 correspondent à ceux qui ont été obtenus dans les réponses ; en outre, les réactifs et catalyseurs ont été rajoutés sur les flèches de réaction. Les notations ret. et inv. indiquent une rétention ou une inversion de configuration.

⊕ : liaison rompue ou formée ⊖ : liaison transformée

réactions non mappées qui ne concernent pas la formation d'un cétal.

Ainsi, concernant l'obtention d'un cétal à partir d'une cétone, il ne semble pas y avoir de façon directe d'accéder à une liste exhaustive de réponses pertinentes.

Une remarque supplémentaire paraît s'imposer : lors de la recherche de méthodes permettant le passage d'une cétone à un cétal, on trouve, en particulier, les deux méthodes : cétone + diol (schémas 4-b-2 et 4-c-1) et cétone +

époxyde (schémas 4-b-1 et 4-c-2). Il semble satisfaisant que ces méthodes soient retrouvées quel que soit le mapping utilisé pour la requête (mapping proposé par l'Automap (schéma 4-b) ou mapping manuel (schéma 4-c)). Toutefois, ce succès est paradoxalement dû à la présence de réactions indexées dans la base avec des erreurs de mapping. Inversement, dans le cas où le mapping reflète bien le mécanisme de la réaction, il faudra connaître le mapping, donc la réaction elle-même, pour la

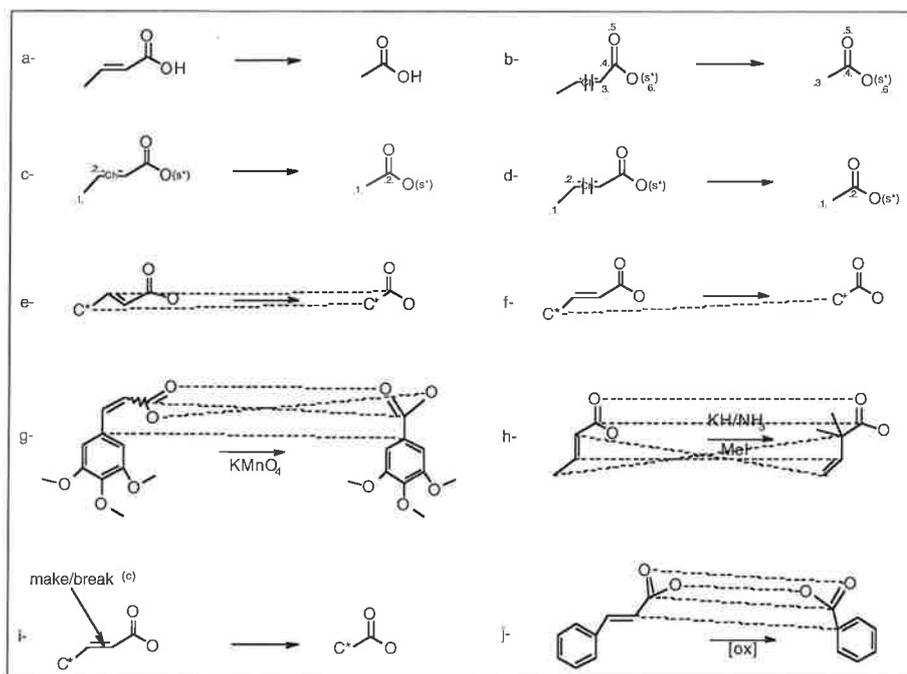
retrouver. Ce problème structurel des BDR doit être pris en compte par les utilisateurs.

Avec Beilstein, on rencontre des problèmes semblables lors de la recherche de réactions de cétylation des cétones. En outre, dans cette base, il semble que les réactions cétone + époxyde soient indexées sans mapping. Si l'utilisateur ne connaît pas cette méthode, il n'en trouvera des exemples que par exploration de la longue liste de réponses (17 960) obtenue par une requête sans mapping.

La recherche de méthodes permettant d'effectuer la réaction du schéma 5-a (oxydation de la double liaison) donne une autre illustration du problème des mappings chimiquement non pertinents. Dans ISIS, à partir de 5-a, l'Automap fournit 5-b. Ce mapping ne correspond pas à celui qui aurait été fait par un chimiste ; cependant, avec cette requête, on obtient une réponse pertinente parmi les 6 réponses (liste A) qui sont obtenues. Avec un mapping manuel selon 5-c, on obtient une liste (liste B) de 58 réponses dont 52 ne sont ni pertinentes ni mappées. Les 6 réponses pertinentes de la liste B (où l'on retrouve la réponse pertinente de la liste A) sont les seules qui sont obtenues lorsque la requête précise que la double liaison est rompue (schéma 5-d). C'est donc le mapping manuel complet 5-d qu'il convient d'utiliser.

Dans Beilstein, on peut vérifier que la requête sans mapping (schéma 5-a) conduit à des réponses (4 936) dont la plupart ne sont pas pertinentes. Avec une requête selon le schéma 5-e, on obtient seulement 2 réponses, d'ailleurs pertinentes. En précisant le mapping pour un seul atome, schéma 5-f, on obtient une liste de 99 réponses. Certaines (49 réactions) sont pertinentes, comme celle du schéma 5-g ; elles n'ont pas été trouvées précédemment car elles sont indexées avec un mapping faux qui est reproduit sur le schéma 5-g tel qu'il apparaît dans la réponse. D'autres réponses ne sont pas pertinentes et correspondent, soit à des réactions semblables à celle du schéma 5-h (21 réactions), soit à d'autres types de réactions. Pour les 21 réactions semblables à celle du schéma 5-h, le mapping est évidemment faux. Ces réactions viennent encombrer la liste des

Schéma 5 (a, b, c).



(a) Dans les schémas a-d (ISIS), les substitutions sont permises sur tous les atomes de carbone. O(s*) = OH. Les atomes qui sont mappés portent le même numéro, la notation Ch précise que la double liaison doit être dans une chaîne.

⊥ : liaison rompue ou formée ⊕ : liaison transformée

(b) Dans les schémas e-j (Beilstein), les atomes qui sont mappés sont joints par une ligne pointillée ; en l'absence de la notation C* les substitutions ne sont pas permises. Pour les schémas g, h et j, le mapping est la reproduction, totale ou partielle, de celui qui est indiqué dans les réponses ; les réactifs ont été rajoutés sur les flèches de réaction.

(c) Dans la requête correspondant au schéma i, make/break a été indiqué avec la notation ad hoc.

réponses et, surtout, elles ne seront pas retrouvées lorsqu'on recherchera des réactions de leur type en formulant une requête avec un mapping en accord avec le mécanisme réactionnel. On constate enfin qu'avec les requêtes 5-e et 5-f, on n'a pas retrouvé toutes les réactions correspondant au schéma 5-a qui sont présentes dans la base. En effet, avec la requête 5-i, c'est-à-dire en précisant seulement la liaison qui est affectée par la réaction, on accède à une liste de 61 réponses dont 6 répondent à notre problème mais ont été indexées avec un autre type de mapping faux (cf. schéma 5-j).

En conséquence, concernant la réaction du schéma 5-a, dans le cas d'ISIS, il faut faire un mapping manuel des atomes et indiquer la liaison qui est modifiée. Dans le cas de Beilstein, il faudrait faire plusieurs requêtes avec, en particulier, des formulations basées sur les intuitions que l'on peut avoir quant aux erreurs de mapping dans la base, ce qui n'est évidemment pas une solution raisonnable.

Zass [7] a noté le problème de l'absence de mapping et recommande, en

règle générale, de mapper dans le cas d'ISIS et de ne pas mapper dans celui de Beilstein. Au moins dans ce dernier cas, cette proposition ne semble pas satisfaisante.

Le mapping étant généralement indispensable lors de recherches sous-structurales, la question se pose de savoir quel type de mapping il convient d'utiliser. Dans l'aide d'ISIS, il est indiqué que : « Atom-atom maps in MDL reaction databases are usually assigned based on the apparent change in the reaction, rather than the actual reaction mechanism » et Zass [7] indique que le mapping ne doit pas être confondu avec un mécanisme de réaction. Le problème est que l'utilisateur ignore quelles sont les réactions pour lesquelles le changement apparent correspond au mécanisme et celles où ce n'est pas le cas. La documentation de Beilstein ne donne pas d'indication sur la façon dont le mapping des réactions est saisi ; dans le cas de cette base, le problème est très mal résolu : on est bien loin des critères de qualité dont se prévalent les concepteurs de la base [22].

En conclusion, pour une même question, on est conduit à faire des requêtes sans mapping et des requêtes avec différents mappings. Cette situation n'est pas satisfaisante, il conviendrait que toutes les réactions soient mappées et qu'elles le soient avec un mapping chimiquement pertinent.

Conclusion

L'intérêt des BDR ISIS et Beilstein en synthèse organique est indéniable, il n'est besoin, pour s'en convaincre, que de considérer le nombre des utilisateurs du serveur Titane et celui des interrogations qu'ils font quotidiennement. Il est notamment clair que l'interrogation par sous-structure permet un accès rapide à des informations qu'il peut être très long, voire impossible, d'obtenir par la documentation papier.

Si les utilisateurs ne peuvent être que convaincus de l'intérêt de ces outils, ils doivent aussi être avertis de l'existence de différents problèmes qu'ils sont susceptibles de rencontrer en les employant. C'est ce dernier point qui était l'objet de notre propos.

Nos observations montrent que ces BDR devraient être améliorées en vue de mieux répondre aux besoins des chimistes. Ainsi, il serait nécessaire (et facile), de mieux informer les utilisateurs quant à la littérature qui est couverte (Beilstein) ; il serait utile de présenter plus explicitement la philosophie qui préside au choix des réactions (ISIS) et à l'indexation (ou non) des méthodes de synthèse (ISIS, Beilstein) ; il serait enfin important de résoudre les problèmes de mapping. Par ailleurs, on peut noter, sans le développer ici [8b], le fait que, avec les BDR, on accède directement à des réactions particulières alors que le chimiste raisonne aussi (d'abord) à des niveaux généraux de description des réactions ; cela suppose une hiérarchisation et un classement des réactions contenues dans les bases. Cette catégorisation permettrait d'éviter certains des problèmes rencontrés, en particulier, de restreindre le volume des réponses [23].

Certains utilisateurs du serveur national Titane se posent la question de savoir s'il est utile de disposer des deux bases ISIS et Beilstein. Il nous semble, en accord avec Zass [7], que ces deux

bases sont complémentaires et doivent être utilisées conjointement.

Enfin, dans le cadre d'un accès des chercheurs et des étudiants français à des moyens d'information efficaces, d'autres bases de données chimiques pourraient être évaluées, en particulier, des bases de données de réactions parmi lesquelles on peut citer celles qui sont proposées par CAS [11], Infochem [24a], ISI [24b], et Synopsys [24c].

Références et notes

- [1] Zass E., *Chimia*, **1994**, *48*, p. 109.
- [2] a) Par commodité, nous utiliserons « ISIS » pour désigner les bases MDL exploitées par le système ISIS et « Beilstein » pour la base Beilstein Cross FireplusReactions ; b) Par la suite, ISIS désignera la base RXL132.DB (constituée des sous-bases ChemInform 98-92, Orgsyn, REACCS-JSM, CHC, RefLib) et Beilstein désignera Beilstein CrossFire plusReactions BS 9801 PR.
- [3] Titane (<http://vulcain.u-3mrs.fr>) est un serveur national (CNRS/MENRT) qui permet aux chercheurs académiques et étudiants français d'accéder aux bases suivantes : MDL Information System (Chem Inform RX, CHC, ORGSYN, REACCS-JSM, RefLib, SPORE, METBRO, ACD, NCI, MDDR, CMC) ; Beilstein Cross FireplusReactions ; Cambridge Structural Database ; Chirbase. Renseignements : A. Baldy (baldy@titane.u-3mrs.fr).
- [4] C'est le cas depuis de nombreuses années en Grande-Bretagne (<http://www.dl.ac.uk/CDS/cds.html>), aux Pays-Bas (<http://www.caos.kun.nl>), en Suisse, en particulier à l'ETH Zurich (<http://www.infochem.ethz.ch>), et dans un nombre de plus en plus grand de pays : voir Beilstein Brief II/1996 et I/1998 et le site <http://www.beilstein.com>
- [5] Vismara P., Régis J.-C., Quinqueton J., Py M., Laurenço C., Lapied L., *Actes des 12^e journées internationales sur les systèmes experts et leurs applications*, **1992**, p. 305 ; Vismara P., Jambaud P., Laurenço C., Quinqueton J., *Langages et modèles à objets : états et perspectives de la recherche*, Ducourneau R., Euzenat J., Masini G., Napoli A. (Eds), Collection didactique de l'INRIA, **1998**, vol. 19, p. 397.
- [6] La documentation de Beilstein (<http://www.beilstein.com/products/xfire/xfplus/xns.html>) qualifie CrossFireplusReactions de « Fully integrated reaction database : the performance of CrossFire, with the comprehensive coverage and quality only Beilstein can deliver »
- [7] Zass E., *The Beilstein System Strategies for Effective Searching*, S. R. Heller, Ed., ACS, Washington DC, **1998**, p. 99.
- [8] a) L'essentiel de ces résultats a été présenté par l'un d'entre nous (J.C.), en février 1998, lors d'une formation sur les bases de données en chimie organisée par P. Gerbier à l'UMII (Montpellier) ; b) Un article complet va être soumis pour publication.
- [9] Dans la requête, faite dans ISIS et dans Beilstein, on a indiqué le mapping [10] et les contraintes suivantes qui n'apparaissent pas sur le schéma 1 : la double liaison de configuration E se trouve dans une chaîne et la partie R de l'aldéhyde contient une fonction ester qui ne doit pas réagir dans les conditions de la réaction.
- [10] Lors de l'interrogation des BDR, le terme mapping est utilisé pour indiquer une relation de correspondance entre les réactants et les produits de la réaction. Cette relation concerne : l'appariement des atomes, l'indication des centres réactionnels et les changements de statut des liaisons. Dans ce manuscrit, afin d'éviter l'utilisation de périphrases (et faute de termes français adaptés), nous utiliserons les termes « mapping » et « mapper ».
- [11] <http://www.cas.org>
- [12] Warr W., *The Beilstein System, Strategies for Effective Searching*, S. R. Heller, Ed., ACS, Washington DC, **1998**, p. 149.
- [13] On trouve ces renseignements en questionnant par le mode « Facts » dans les champs jt et py.
- [14] March J., *Advanced Organic Chemistry : Reactions, Mechanisms, and Structure*, 4^e Edition, John Wiley & Sons, New York, **1992**, p. 474.
- [15] a) On retrouve cette réponse pertinente (R' = allyl) en faisant une requête (mode By Form) dans laquelle figure, outre le schéma 2-c, la précision que du propane dithiol doit être utilisé en tant que réactant ou en tant que réactif ; b) On pourrait réduire ce nombre de réponses en précisant mieux la requête quelle ne l'est sur le schéma 2-c ; cependant, le problème demeurerait.
- [16] Ferrier R.J., Middleton S., *Chem. Rev.*, **1993**, *93*, p. 2779.
- [17] Carey F.A., Sundberg R.J., *Advanced Organic Chemistry, Part B : Reactions and Synthesis*, 3^e Edition, Plenum Press, New York, **1993**, p. 73.
- [18] <http://www.mdli.com/dats/synthdb.html>
- [19] Jochum C., *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, p. 71.
- [20] Dans les deux bases, nous avons exploré un certain nombre de réactions afin de constituer un échantillonnage représentatif. Dans le cas d'ISIS, chaque sous-base a été explorée ; pour la sous-base Chem Inform, dans chacune des actualisations annuelles, nous avons exploré 3 échantillons de 50 réactions, à savoir, les réactions 1-50, 30001-30050 et 60001-60050 (sauf dans le cas de ChemInform 98.4.1.3 qui ne contient que 56 231 réactions et pour laquelle le troisième échantillon a été constitué par les réactions 50001-50050). Le nombre de réactions non-mappées est le suivant : ChemInform 98.1 (respectivement, pour les 3 échantillons, 8, 1 et 10 réactions non mappées), Chem Inform 97.1.1 (4, 9, 11), ChemInform 96.1.1 (0, 6, 9), ChemInform 95.1.2 (3, 9, 13), ChemInform 94.1.3 (0, 12, 12), ChemInform 93.1.3 (7, 16, 5), Chem Inform 9 (4, 0, 12) ; pour les autres sous-bases nous n'avons pas observé d'absence de mapping. Dans le cas de Beilstein, en interrogeant par le Fact Editor dans le champ rx.id (numéro de réaction), nous avons exploré des échantillons de 50 réactions dans les domaines suivants (et obtenu les résultats notés entre parenthèses) : 101-150 (5 réactions indexées sans mapping), 1000001-1000050 (16), 2000001-2000050 (29), 3000001-3000050 (15), 4000001-4000050 (11), 4810001-4810050 (15).
- [21] Carey F.A., Sundberg R.J., *Advanced Organic Chemistry, Part A : Structure and Mechanisms*, 3^e Edition, Plenum Press, New York, **1993**, p. 442.
- [22] Luckenbach, R., *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, p. 923.
- [23] Dans Beilstein, il n'y a pas de fonction permettant le classement des listes de réponses. Dans ISIS, la fonction « clusterisation » peut être utile mais elle ne répond souvent que mal à ce problème.
- [24] a) <http://www.infochem.de> ; b) <http://www.isinet.com> ; c) <http://www.synopsys.co.uk>