

An ^1H NMR Database for Complex Carbohydrate Structures複合糖質糖鎖構造の ^1H -NMRデータベース

J. Albert VAN KUIK and Johannes F. G. VLEGENTHART

Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands, FAX: 31-30-540980

In TIGG(1990) 2, 526-533, Dr. Yamagata made a proposal to facilitate the identification of complex carbohydrates. He suggested to synthesize all known oligosaccharide structures, and with these structures compile databases of NMR spectra and HPLC profiles. Of course with the rapidly growing amount of all sorts of data on oligosaccharides, a systematic approach to handle these data is a must. The building of an HPLC database has the advantage that the apparatus for HPLC analysis is relatively inexpensive, and that a number of compounds can be analyzed in one run. However, as Dr. Lee pointed out in Glycoforum[TIGG(1991) 3, 78-81], the identification of oligosaccharides by an HPLC profile is not so straightforward. At present, NMR spectroscopy provides the best way to unambiguously identify complex carbohydrate structures.

What features should an ideal NMR database for complex carbohydrate structures possess? It should contain all available information, therefore both ^1H -NMR and ^{13}C -NMR spectra, as well as the complete assignments of these spectra, have to be included. Because the data have to be independent of the recording conditions, spectra should be recorded at different spectrometer frequencies and at different temperatures. Powerful tools are needed for fast and easy searches in the database. Of course this database must identify known carbohydrate structures, but it should also be able to assist with the elucidation of unknown structures. The results should be clearly visualized, so that at one glance it should be obvious which part of a spectrum is recognized and to which part of a carbohydrate structure it corresponds. Eventually, this database can be expanded with conformational information of carbohydrates, like NOE's obtained from multidimensional NMR spectra. Some research has already been done to provide us with tools for such a database, e.g. the analysis of (un)assigned lists of ppm values(1-3) and the artificial neural network approach for "instant" recognition of spectra(4).

Unfortunately, it is not feasible to construct such a database at this moment. One problem is that the storage of NMR spectra requires a considerable amount of memory, which would make a large database of NMR spectra difficult to manipulate. This problem is only technical and will be solved in not too far future. The main problem is the unavailability of high quality NMR spectra for most complex oligosaccharide structures. One approach to get spectra is to ask the various

TIGG (1990) 2, 526-533のGlycoforumで、山形博士は複合糖質の同定を容易にするための一つの提案をした。彼の提案は、すべての既知のオリゴ糖を合成してそのNMRスペクトルとHPLCの挙動をデータベースに入れたらどうかというものである。実際、オリゴ糖に関するあらゆる種類のデータが急速に増えているので、このようなデータを取り扱う系統的なアプローチが絶対に必要である。HPLCデータベースを構築することは、HPLC用分析機器が比較的安価であるし、多数の化合物が一回のHPLCで分析できるので、簡単にできそうに思われる。しかし、Lee博士がTIGG(1991)3, 78-81のGlycoforumで指摘したように、オリゴ糖を一回のHPLC分析で同定するのは容易ではない。現在の時点では、NMRによる分析が複合糖質糖鎖構造を正確に同定する最良の方法である。

それでは、複合糖質糖鎖構造のための理想的なNMRデータベースとは何であろうか？このデータベースは入手しうる情報をすべて、つまり ^1H -NMRと ^{13}C -NMRスペクトルを両方とも、そしてそのスペクトルの完全な帰属を含んでいることが必要である。そのデータは測定条件に依存してはならないので、スペクトルはスペクトロメーターの波長と温度を変えて測定した方がいい。このデータベースを速やかにそして簡単に検索するための強力なツールが必要である。勿論このデータベースは既知の糖鎖構造の同定のデータで構築されるが、未知の構造の解析にも役立つものが好ましい。データは明確に視認でき、それゆえ一目でどの部分のスペクトルに糖鎖のどの部分が対応しているか判ることが望ましい。最終的にはこのようなデータベースは、多次元のNMRスペクトルにより得られるNOEのような糖鎖の立体構造に関する情報をも含むように、いずれ拡張することができよう。このようなデータベースを作成するために役立つ研究、例えば、ppm値の帰属および未帰属リスト(1-3)や、スペクトルを一瞥で認識できるような人工的神経回路網の研究(4)が既に始まっている。

残念ながら、このようなデータベースを現時点で構築することは簡単ではない。問題点の一つは、NMRスペクトルを貯えるとコンピューターのメモリーを相当喰ってしまい、しかも巨大化したNMRスペクトルのデータベースの取り扱いが簡単ではないということである。しかしこれは、純然たる技術的問題なので、そう遠くない将来解決するだろう。最大の問題は、殆どどの複合糖質糖鎖構造に対して質の高いNMRスペクトルが得られないことである。スペクトルを入手する一方法は、そのような研究をしている研究室にもとのデータを呉れるよう依頼し、これらのすべてのデータを統一標準フォーマットに変換することである。すべての研究室がこの要請に応じて大事なデータ

carbohydrate laboratories for their original data and convert all these data in one standardized format. It is doubtful if all laboratories are able and willing to donate their data to such a project. Furthermore, a large number of compounds published show bad quality $^1\text{H-NMR}$ spectra, due to a low amount of material, or the presence of impurities. In other cases carbohydrate chains could not be separated completely from each other, consequently resulting in spectra of mixtures. For the majority of published oligosaccharides, the low amount of carbohydrate material inhibits the recording of $^{13}\text{C-NMR}$ spectra. Another approach would be to record new spectra from all known complex carbohydrates. This would by no means be easy as a lot of compounds, due to a variety of reasons, are no longer available. Dr. Yamagata suggested that all compounds should be synthesized, in order to provide the carbohydrate community with a set of reference compounds. Even if it would be possible to synthesize the larger mass oligosaccharides, it would be an extremely expensive and time-consuming project. In both cases the construction of such a database is an enormous task, which cannot be undertaken by a single laboratory.

Currently, the most practical approach is to collect NMR tables from literature, as the NMR features of hundreds of carbohydrate structures have been published. Fortunately, the published $^1\text{H-NMR}$ tables contain all the information that is needed for the unambiguous assignment of complex carbohydrate structures. The tables have to be normalized, *e.g.* give all ppm values relative to one reference, and combined with carbohydrate structures and literature references. The logical way to do this is to utilize the efforts that have already been made in this field by CarbBank. In fact, our laboratory is undertaking this project at this moment and we hope to make the result available in the near future.

References

1. Bot, D. S. M., Cleij, P., van 't Klooster, H. J., van Halbeek, H., Veldink, G. A. and Vliegenthart, J. F. G. (1988), *J. Chemometrics* **2**, 11-27
2. Jansson, P. E., Kenne, L., and Widmalm, G. (1989) *Carbohydr. Res.* **193**, 322-325
3. Hounsell, E. F., and Wright, D. J. (1990) *Carbohydr. Res.* **205**, 19-29.
4. Meyer, B., Hansen, T., Nute, D., Albersheim, P., Darvill, A., York, W., and Sellers, J. (1991) *Science* **251**, 542-544.

を提供してくれるかどうか不明である。更に重要なことは、今迄に発表されているNMRデータの多くは、測定する物質が少なすぎるかあるいは不純物を含む為に、質が悪いということである。ある場合には、二種以上の糖鎖が完全に分離されないまま、従って混合物のNMRスペクトルであったりする。今までに発表された多くのオリゴ糖では、測定に供する試料が少なすぎる為に $^{13}\text{C-NMR}$ を測定することができない。他の方法は、既に報告された糖鎖を用いてNMRスペクトルを記録し直すことであろう。しかしこれは、既に報告された糖鎖の多くが様々な理由でもう入手できないので、現実的ではない。山形博士は、糖鎖の研究社会のために標準物質となるよう既知のすべての糖鎖を合成したらどうかと提案している。多数のオリゴ糖を合成することができたとしても、それは大変高価なそして時間のかかるプロジェクトとなろう。どちらの場合もこのようなデータベースの構築は大変な手間であり、一研究室では遂行できるものではない。

現在のところ最も現実的なアプローチは、数百に及ぶ糖鎖構造のNMRデータが発表されているので文献からNMRデータ表を集めることである。幸いに、公表された $^1\text{H-NMR}$ データ表は糖鎖構造を絶対確実に同定するのに必要なすべての情報を含んでいる。このデータ表は、標準化のためにすべてのppmを1つの標準物質に対する相対値で表わし、それを糖鎖構造のどの部分と対応するか、そしてさらに出典も示す必要がある。理性的に考えれば、これを実現するには複合糖質研究の領域ではCarbBankにより既に払われている努力を利用することである。実際、私の研究室ではこのプロジェクトを開始していて、もうすぐその結果を諸研究者が利用できるようになることを願っている。

三菱化成生命科学研究所 細胞認識研究部
山形 達也 訳