Molecular Biology Today (2000) 1(1): 23-28.

SwaN-MR from Infancy to Maturity

Giuseppe Balacco

Menarini Ricerche S.p.A., Via Sette Santi 3, I-50131 Firenze, Italy

The NMR software SwaN-MR was described in the scientific literature more than five years ago. During this period the program has met worldwide success, and has been modified many times. The present article corrects and updates the material published in 1994. The task of processing and interpreting NMR spectra will be discussed in general terms. The unique approach of SwaN-MR is stressed. It is centred on the interaction between scientist and spectrum. As an example, the techniques arising from manual phase and baseline correction have been extended to the case of fitting an experimental spectrum with a simulated one.

Introduction

Nuclear Magnetic Resonance, especially at higher magnetic fields, has the advantage, compared to other spectroscopies, that every single signal in the spectrum is easily interpreted or, in other words, assigned to a given nucleus. Given that an n-dimensional spectrum can contain hundreds of signals, the amount of information that can be extracted is substantial. Normal practice suggests that this information must be extracted completely in order to be sure. Any interpretation of a spectrum unable to explain all of its features is probably erroneous. This process of extraction has always been performed by a human brain and can take from a couple of minutes up to months, depending on the complexity of the problem. Although many efforts have been spent teaching computers to perform the task, we cannot say when this will happen. It seems that spectroscopists are quite happy performing the interpretation because it is a challenging and rewarding intellectual activity. In the old days of continuous wave NMR, the spectrum was plotted on a piece of paper during acquisition and studied away from the spectrometer, on the researcher's desk. With the advent of computers and FT-NMR, a second opportunity arose: to study the spectrum interactively on a computer screen. Although this practice began directly on the spectrometer, it is clear that the best place is still the desk. In 1994 the author reported some problems connected with this desktop analysis and proposed his solution in a paper which, after explaining how to transfer spectra onto a Macintosh computer, described the new program SwaN-MR (Balacco, 1994). It is well known that in the six years since then there has been a constant evolution in the world of computers. It is arguable that something changed in NMR software too. SwaN-MR also underwent a constant evolution. Every part of the program has been rewritten at least once, many new functions have been introduced, others were eliminated. In many aspects (speed, stability, compactness) SwaN-MR appears to have reached the top. The program is also much more productive and

useful than it was. The single steps of this evolution were not worth another article. Considering the growing interest in SwaN-MR, it has begun necessary to update the literature and describe the state of SwaN-MR today. This is one of the aims of the present article. The second aim is to explain in more detail the uniqueness of the program, taking advantage of the historical perspective that these six years have given us. Not all things will be said: it is assumed that if one can read this article, he/she can also read the rich information existing on the world wide web (see below). The first paper described in detail the individual functions of the program. In the author's opinion, no other single application can boast the same list of features. This article will examine the program as a whole.

Material and Methods

SwaN-MR version 3.5.2 is a C/C++ program compiled under the Symantec Project Manager version 8.1 with the Symantec C++ for Power Mac and the SPM MrC compilers. It uses the Think Class Library 2.0 plus part of the ANSI C library. The other parts of the program were all specifically written for SwaN-MR. The copyright belongs to Menarini Ricerche S.p.A. The only requirement to run the program is a PowerMacintosh or compatible computer. Installation is unnecessary: the program can in theory be launched from a floppy disk. Many new versions have been released every year, but there are no rules and no plans concerning this. The compressed program (only 319 Kb) can be freely downloaded from several ftp sites, the main one being:

ftp://qobrue.usc.es/nmr/uploads

The updated list of the mirror sites is available at the web address:

http://qobrue.usc.es/jsgroup/swan/index.html

This site also contains other useful information. The documentation consists of a manual that can be read from inside SwaN-MR itself. The first part of the manual, explaining the theory of processing, is published on the internet, at the URL:

http://qobrue.usc.es/jsgroup/swan/theory.html

To obtain a regular license to use the program it is necessary to send an e-mail to the author and to cite SwaN-MR in one's own research work. Users are also invited to report bugs to the author. These conditions are better described on the web site.

The number of registrations is 206 (at November 30, 1999). In most cases one registration was written on behalf on an entire research department, so the number of copies of the program is unknown. Registered users are mainly chemists, but there are also many biologists, engineers, physicians and physicists. There is also a minority of registrations coming from centres not involved in any research activity. Registered users play an important role: they give the ideas. Even when they simply ask for clarifications, they point out which parts of the program are obscure and need to be simplified.

Discussion

What NMR programs have in common is that they are written with the final user clearly in mind. What makes SwaN-MR different from the usual commercial program is the concept of user. This rather ambitious software tries to change the way the scientist works: probably the last thing he wants to do! In fact, writers of commercial software know well that their customers are very conservative: they want their new software to look like something familiar. The logical result is that the classic NMR software is a disk that transforms a computer into a virtual spectrometer. Spectrometers, in turn, are designed and often advertised as machines that everyone can use, even if he/she has no previous experience in NMR. Their language is not the same language of scientific literature. The classic example is the existing confusion about the two dimensions in 2D spectroscopy: which should be considered the first and which the second? Another feature of spectrometers is the central importance they put on printing. In fact a spectrometer is a machine that, given a sample in input, renders a printed spectrum on output. We are back to the beginnings: PC software mimics a FT spectrometer which, in turn, mimics a CW spectrometer! This is, of course, an oversimplification: apart from the fact that it is difficult to generalize, all programs are eventually completed with the same functions and can do the same things.

Printing remains important in SwaN-MR, but it is not central. The result that SwaN-MR tries to obtain is the maximum transfer of information from the spectrum to the brain of the user. This user is also well defined. It is assumed he/she is a scientist or someone who studies to become a scientist. So the first rule in writing the program was to implement as much as possible a scientific language. This is something easy to state but almost impossible to do. Actually the spectroscopic literature, mainly for historical reasons, is often ambiguous or illogical. The first example is the widespread habit of counting the number of real points of a spectrum even when this spectrum is complex. The Fourier transformation is an operation defined on complex numbers, so the natural way to consider a spectrum is as a complex array. This fact is explained in the SwaN-MR manual, which also touches all the other disputed issues that would take too much space if treated here. It suffices to say that, in some cases, more rigorous mathematical terms have been preferred to the spectroscopic conventions.

This first rule alone should make the program user-friendly, assuming that the user is familiar with a scientific language, of course. This is one of the many aspects under which all programs try to be as easy as possible, but implementing opposite recipes. After the language, there is the overall structure. The normal recipe is to combine two programs into one. The simplest program deals with 1D spectroscopy and is the only thing that many users see. The advanced program deals with 2D spectroscopy. The practical result of this policy is that the total number of commands is double the necessary quantity. For example there are two commands to perform a Fourier transformation: the 1D version and the 2D version. SwaN-MR simply contains half of the commands, because the program corresponds to the advanced part alone. In fact, SwaN-MR gives its best in 2D processing. All spectra, including the simplest one-dimensional spectrum, are assumed to have four dimensions; some dimensions, of course, extend over a single point. Such a choice can be

seen as a complication for those who are only interested in simple spectra, but in practice they don't even notice it.

The ideal user exists only in imagination. Many expert spectroscopists find SwaN-MR not professional enough for their standards (the fact that it is not a unix program may be important in this respect), while many chemists use less than half of the program. Surprisingly enough, SwaN-MR is often used by students to embed pictures of NMR spectra into their thesis: here, evidently, the fact the SwaN-MR is a Macintosh application comes into favour. If only the diffusion is considered as a measure of success, price and availability are more important than what the program can do. The reader may wonder, at this point, whether the choice of the platform was as carefully planned as all other things. The plain truth is that in our department we have always had only Macintosh computers (not counting the computers bundled with the instruments), so there was no reason to consider other operative systems. It is known that the most time-consuming part in the creation of a complex program is testing. Almost all the bug reports concerning SwaN-MR have come from inside Menarini: this demonstrates that testing cannot be delegated. Every now and then someone suggests porting SwaN-MR to Windows, but only one group was ready to invest the necessary amount of money and time. Due to technical difficulties and copyrights problems, they opted to write their own program, called MestRe-C (http://qobrue.usc.es/jsgroup/MestRe-C/mestre-c.html). There has always been an informal partnership between the author and the MestRe-C development team, which has helped a lot both of us. This is the first example in which a few of the SwaN-MR concepts were repeated into another program, that has nevertheless its own philosophy and style. On the other hand, SwaN-MR users were patiently waiting for other NMR software running on the Mac. It's clear that a general-purpose program like SwaN-MR cannot cover in depth all aspects of NMR. Some special cases, which are not met every day, are better treated with more specialised software. It was clearly stated in the first paper that special care had been taken to facilitate data exchange between SwaN-MR and other NMR programs. Up to now, no other developer has ever taken this opportunity into consideration, so SwaN-MR is still an island. The author's answer was to attach new features directly into the program. The most important of these was the simulation of complex 1D spectra (crowded spin systems or chemical exchange). During the years the old programs LAOCOON (Castellano et al., 1964), DNMR (Binsch, 1969) and DAVINS (Stephenson et al., 1980) have been integrated into SwaN-MR. This recovery was welcome because the original versions, and sometimes, *e.g.* DNMR6 (Brown et al., 1983), even their portings to DOS, were not compatible with today's computers. To be precise, the actual programs have not been incorporated, only their methods and equations as reported in the cited papers. Merging many programs into a single one not only eliminated the tedious task of transferring data from one application to the other, but created new possibilities. With SwaNMR one can fit a simulated spectrum to the experimental one in a manual and interactive fashion, just like during a phase correction or a baseline correction. This is the perfect complement to the DAVINS principle. The old program found a solution to make the least square fitting as independent as possible from the initial estimates. SwaN-MR, exploiting the power of modern computers, gives the human estimates a previously unknown degree of accuracy. Instead of eliminating the need of brain activity, the computer

gives new eyes to the scientist. It is not a case that SwaN-MR is not scriptable. The program task is to let the spectroscopist interact with the spectra, not to let him/her sleep. Such things are so obvious that the simulation part of SwaN-MR has been an instant success and is often used for teaching purposes.

I have just said that, despite the abundance of many FORTRAN algorithms, like those described above, all algorithms inside SwaN-MR were written specifically for the occasion. They were patiently optimized for the MC68040 processor (often working in assembler) and then rewritten to give optimal performance under the PowerPC processor. Considering that the whole SwaN-MR project has been carried out with minimal resources of money and time, this may sound strange. In fact the unlimited optimization was both necessary and possible. Necessary because an interactive program must, by definition, give the quickest responses. Possible because when time is little and discontinuous, the details can be cured, not the overall structure. Another time SwaN-MR demonstrates to be heretic. The common practice today is to write programs in high level, object-oriented languages; to heavily exploit existing libraries and to reuse old pieces of code. Only in this way it is possible to release new programs in a reasonable amount of time and with a reasonable amount of money. As a result programs are huge, slow, always hungry for computer resources and rarely stable enough. It is neglected that any complex program inevitably has some bug and that complete debugging inevitably takes years.

The old program accumulates and concentrates the experience of its users. A strange question arises now: is it possible to repeat the process with the single spectrum? After the spectroscopist has extracted the information from the spectrum (whether this process is finished or unfinished, correct or incorrect), can the software help in organizing and storing this information? SwaN-MR helps because it concentrates everything into a single computer file: the FID with the experimental conditions, the spectroscopist's doubts with the proven assignments, the processing parameters with the structural formula. There are two mechanisms for this purpose. First come the free annotations that can contain any text or picture and are normally associated with single peaks. Second come the assignments, that are organized as a small database into the single spectrum. Though necessary, these methods are not completely satisfactory. They also seem to be unpromising for further development.

Many algorithms taken from the literature have been first introduced into SwaN-MR and, after a trial period whose length ranged from days to years, abandoned. They simply didn't work. Other parts of the program have become obsolete for other reasons, like the ability to draw on a pen plotter: an item of industrial archaeology. Other algorithms have been limited in scope and tuned to the users' need. The best example in this class is Linear Prediction. It is a better alternative to zero-filling or to baseline correction, provided that some special conditions are met. Parametric LP has, instead, been removed because it is clearly uninteresting for the totality of users. There are also many simple methods, like polynomial baseline correction, which are so effective, despite what has often being said, that they discourage trying newer alternatives.

What can always be improved and simplified is the human interface, in other words the appearance of the program. There is always room for simplification and improvements. It would be very dangerous, however, to give the false impression that a scientific program

can be used without reading or, better, studying its manual. Since the beginning constant care has been exercised to keep the number of menu commands to a minimum. Most importantly, the menus and the commands are always the same: in computer terminology, the program has a single "mode". Despite all these effort to make the user's life simpler and simpler, the program appears to be aimed at initiated people: it has been shown that the casual user can't even locate the simplest commands if he doesn't read the manual first. The simplicity of the interface becomes an advantage only after he/she has used the program routinely for, let's say, one month. The author likes to say that SwaN-MR is long-time-user friendly.

We have seen that this program goes against many rules. There are two last examples of which the author is very proud. First, although the program is mainly used inside universities, it has been written in an industry. Second, in the field of NMR, this is the first Italian invention to be exported in so large quantities, to the best of my knowledge.

Conclusion

We have seen that all NMR programs do the same things and use the same algorithms, yet they are strikingly different in style. This is unfortunate for those who have to learn more than one program, but is a good thing for the scientific community as a whole. While other programs tend to hide the mathematical complexities from the user, SwaN-MR tries to make processing simple and understandable. The choice of a software is normally based on marginal details. This can be discouraging, but actually gives the programmer a great freedom (users won't blame him for having a bad style). In the confusing situation that results, SwaN-MR is a valid option for the seriously intentioned scientist. If he/she doesn't already master the art of processing, he/she probably will, after having explored all the program possibilities.

We have no idea about how many years we have to wait for an artificial intelligence to take the burden of interpreting NMR spectra. What we know is that a lot of natural intelligence is already available, which only needs the right tools to become productive. The fact that a computer is intrinsically stupid becomes an advantage. It is the ideal tool to complement natural intelligence.

References

Balacco, G. 1994. SwaN-MR: a Complete and Expansible NMR Software for the Macintosh. J. Chem. Inf. Comput. Sci. 34:1235-1241.

Binsch, G. 1969. A Unified Theory of Exchange Effects on Nuclear Magnetic Resonance Line Shapes. J. Am. Chem. Soc. 91: 1304-1309.

Brown, J.H., and Hackett Bushweller, C. 1983. DNMR6: Calculation of NMR Spectra (I=1/2) Subject to the Effects of Chemical Exchange. Quant. Chem. Prog. Exch. 3: 103. Program no. QCPE 633.

Castellano, S., and Bothner-By, A.A. 1964. Analysis of NMR Spectra by Least Squares. J. Chem. Phys. 41: 3863-3869.

Stephenson, S.D., and Binsch, G. 1980. Automated Analysis of High-Resolution NMR Spectra. I. Principles and Computational Strategy. J. Magn. Reson. 37: 395-407.