Journal of Biomolecular NMR, **18:** 329–336, 2000.
KLUWER/ESCOM
© 2000 *Kluwer Academic Publishers. Printed in the Netherlands.*

Ansig for Windows: An interactive computer program for semiautomatic assignment of protein NMR spectra

Magnus Helgstrand[a], Per Kraulis[b], Peter Allard[a] & Torleif Härd[a,*]
[a]*Department of Biotechnology, Center for Structural Biochemistry, The Royal Institute of Technology (KTH), Novum, SE-141 57 Huddinge, Sweden*
[b]*Stockholm Bioinformatics Center, Department of Biochemistry, Stockholm University, SE-106 91 Stockholm, Sweden*

Received 26 June 2000; Accepted 18 September 2000

Key words: assignment software, interactive graphics, protein NMR, protein structure determination, semiautomatic assignment, sequential assignment

Abstract

Assignment of NMR spectra is a prerequisite for structure determination of proteins using NMR. The time spent on the assignment is comparatively long compared to that spent on other parts in the protein structure determination process, but it can be shortened by using either interactive or fully automated computer programs. To benefit from the advantages of both types of program we have developed a version of the interactive assignment program ANSIG to include automatized, yet user-supervised, routines. The new program includes tools for (i) semiautomatic sequential assignment, (ii) plotting of distances from PDB structure files directly in NMR spectra and (iii) statistical analysis of distance restraint violations with the possibility to directly zoom to violated NOEs in NOESY spectra.

Abbreviations: NOE, nuclear Overhauser enhancement; API, application programming interface; GUI, graphical user interface; HSQC, heteronuclear single quantum coherence; NOESY, NOE spectroscopy; FTP, file transfer protocol.

Introduction

Correct sequence-specific assignments of NMR resonances are crucial for determining protein structures using NMR. Different computer programs are available to assist the NMR spectroscopist during assignment of spectra. The programs can roughly be divided into two sets, one set of programs for graphical display of spectra with databases to keep track of assignments and another set of programs for automatic assignment of resonances and/or NOEs. Both sets of programs have drawbacks. The programs in the first set, e.g. ANSIG (Kraulis, 1989; Kraulis et al., 1994), Pronto (Kjaer et al., 1994), XEASY (Bartels, 1995), NMRView (Johnson and Blevins, 1994), PIPP (Garrett et al., 1991) and AURELIA (Neidig et al., 1995), lack effective routines for computer assistance during the assignment process. The programs in the second set, e.g. AUTOASSIGN (Zimmerman et al., 1994), GARANT (Bartels, 1997) and ARIA (Nilges et al., 1997), are fast but do not utilize the expertise of the user. Hence they might provide erroneous assignments and assignments made are often difficult to validate without retracing data manually, for instance using a program within the first set.

To speed up the assignment process and to utilize the expertise of the user we have chosen to develop one of the available assignment programs in the first set, ANSIG, into an interactive yet semi-automated assignment tool. We have identified three tasks for which computer assistance is of considerable value to speed up the assignment process. These are sequential assignment of resonances, assignment of NOEs when a low-resolution structure is available and ex-

[*]To whom correspondence should be addressed. E-mail: torleif.hard@biochem.kth.se

amination of violated NOEs following initial structure calculations. The goal was to obtain an easy-to-use interactive program, which will assist the user during time-consuming routine tasks.

In this paper a version of ANSIG running on PCs under Windows is presented. This version of the program, 'Ansig for Windows', includes novel features for computer assisted sequential assignment, plotting of distances from PDB structure files directly in NMR spectra, and reading CNS (Brünger et al., 1998) and X-PLOR (Brünger, 1992) output files in order to directly visualize violated NOEs.

Methods

ANSIG was adapted to the Windows operating system using a Fortran compiler from Lahey software solutions. Fortran 77 code from ANSIG (version 3.3), except the code for the graphics interface and system calls, was reused. The graphics interface was rewritten using OpenGL and the Windows API. System-specific operations were re-coded to make the file formats in Ansig for Windows compatible with earlier versions of ANSIG. New features were coded in Fortran 95, with calls to the Windows API for full GUI support. All new features are based on the data structures within ANSIG; no new file formats have been introduced.

Results and discussion

Basic features of ANSIG

The ANSIG program is a graphical tool for display and annotation of 2D, 3D and/or 4D homo- or heteronuclear NMR spectra. It was designed to allow the spectroscopist to inspect and assign the spectra in an interactive fashion. It has the ability to display several spectra simultaneously, either overlaid in different views (projections) or in different windows on the screen. The graphics interface allows for rapid zooming and moving of displayed regions. The user makes new assignments interactively and these are stored in a database consisting of a single file in a custom format. It is possible to configure the program so that the spectra are displayed in a specified set of views, and a user-defined set of macros (groups of commands) is available. There is a computer language embedded in the ANSIG program (the AL language) which provides access to nearly all features within the program and allows implementation of powerful procedures.

Computer assisted assignment

Sequential assignment of protein NMR spectra (Wüthrich, 1986; Roberts, 1993; Cavanagh et al., 1996; and work cited therein) is one of the more time-consuming parts of a protein structure determination using NMR. Comparing cross peaks from different residues in 3D spectra to find sequential connectivities involves searching for matching chemical shifts. This process would be faster if a computer program performed the search.

A set of tools to aid the user during the sequential assignment is implemented in Ansig for Windows. The suggested procedure is divided into four steps as presented in Figure 1. A special tool has been constructed for each step in the procedure. The tools are started from the menu bar and take the form of dialog boxes. The prerequisites are a ^{1}H-^{15}N HSQC spectrum and sequential spectra such as HNCA, HN(CO)CA, CBCANH and CBCA(CO)NH. (The reader is referred to Cavanagh et al. (1996) and related texts for description of information content of various NMR experiments.) Cross peaks must be identified and marked (so-called peak-picking) in all spectra, either manually or with other software, e.g. AZARA (http://www.bio.cam.ac.uk/azara/), prior to the first step in the assignment procedure.

The first step in the assignment procedure is to assign temporary residue identifiers to all peaks in a ^{1}H-^{15}N HSQC. The suggested setup in Ansig for Windows is a spectrum window showing the ^{1}H-^{15}N HSQC together with the 'HSQC' dialog box. The program automatically identifies all non-assigned cross peaks and presents them to the user. Tools are available to look at cross peaks and to assign unique temporary residue identifiers to them. The identifier has the form xN, where N is a number.

In the second step the information gained from the ^{1}H-^{15}N HSQC is used to find cross peaks belonging to temporary residues in 3D spectra containing sequential information such as HNCA, HN(CO)CA, CBCANH and CBCA(CO)NH. The suggested setup in Ansig for Windows during this step is to use two spectrum windows, one showing the ^{1}H-^{13}C plane of the 3D spectrum and the other the ^{1}H-^{15}N planes of the 3D spectrum and the ^{1}H-^{15}N HSQC, together with the '3D HSQC type' dialog box (Figure 2). The program assists the user in the search for cross peaks with amide proton and nitrogen shifts which match the temporary residues assigned in the first step. The result of the search is presented to the user and tools for zooming to, and marking of, found cross peaks are

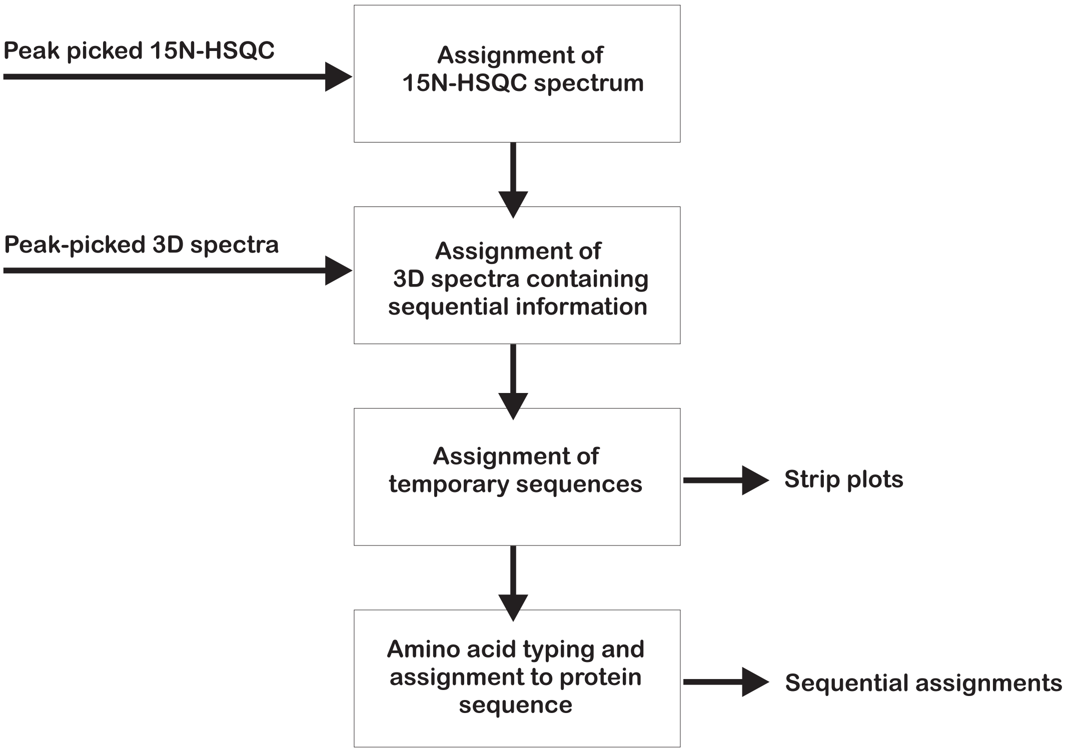
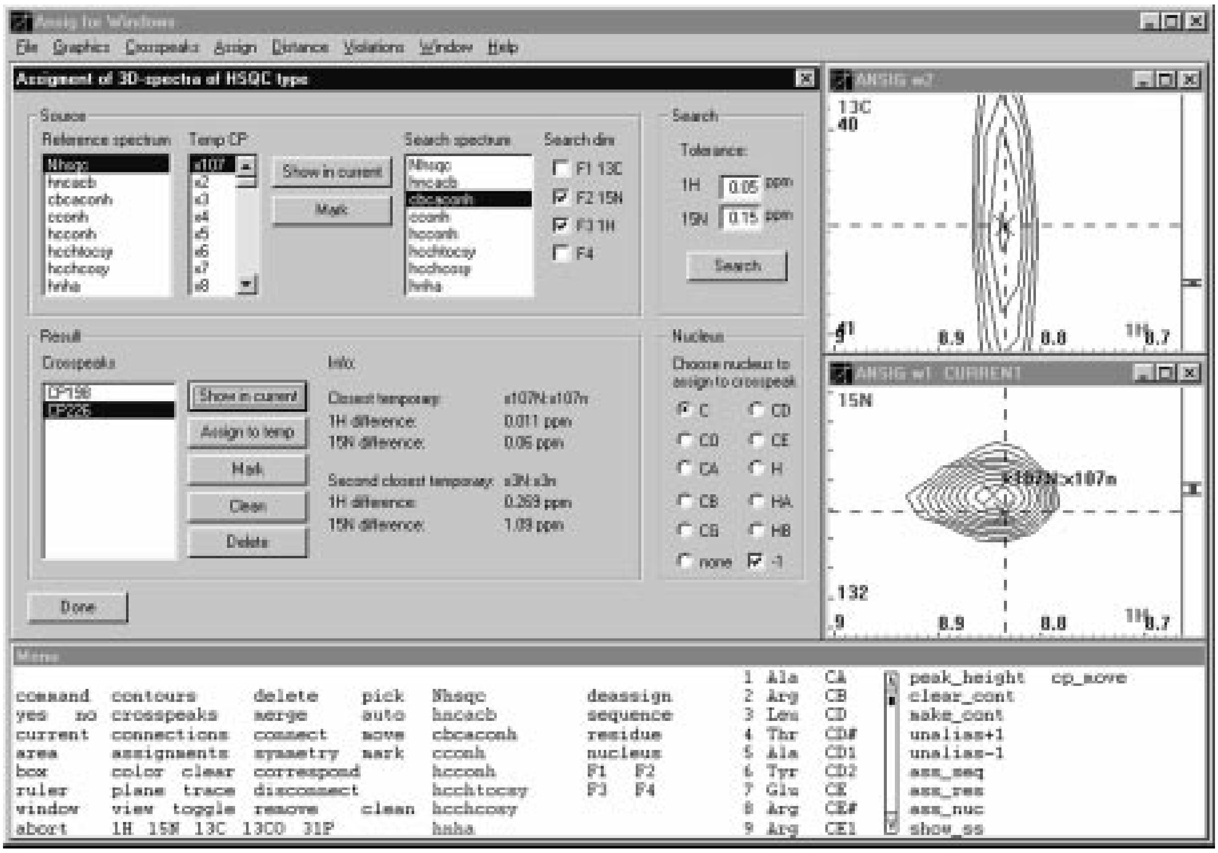

Peak picked 15N-HSQC
Assignment of 15N-HSQC spectrum
Peak-picked 3D spectra
Assignment of 3D spectra containing sequential information
Assignment of temporary sequences
Strip plots
Amino acid typing and assignment to protein sequence
Sequential assignments

Figure 1. Diagram showing the sequential assignment procedure in Ansig for Windows. The boxes represent the different steps involved and correspond to a tool in the program. Arrows on the left side represent input and arrows on the right side represent output.

available. Before the found cross peaks are assigned to the temporary residue the user must indicate which nucleus the cross peak belongs to and also specify if the cross peak belongs to the same or previous residue in the protein sequence in the third spectral dimension.

In the third step of the sequential assignment, temporary residues are assigned to temporary sequences. In this step the sequential information obtained in the previous step is used. The tools are available in the ‘Sequential’ dialog box (Figure 3). From a temporary residue given by the user a forward or backward search in the sequence can be done. The result is presented to the user together with statistics over which spectra were used, how many chemical shift matches were found and how found residues fit into other temporary sequence elements. Strips showing the contours, cross peak and assignment of the used cross peaks are presented in an interactive window (Figure 3), which can be translated and zoomed. This window is used to qualitatively judge the proposed sequential assignments. When assignment to a temporary sequence is done, the program updates the database for all cross peaks assigned to the temporary residues involved.

In the fourth step temporary sequences are attached to the protein amino acid sequence. The first part of this step is to identify the amino acid side-chain types of the temporary residues, which is done based on carbon chemical shifts from C(CO)NH-TOCSY (Grzesiek et al., 1993). The second part is to search the amino acid sequence for sites matching the temporary sequences. The tools to use for this step are included in the ‘Sequence editor’ dialog box. All temporary sequences are presented to the user together with information on proposed side chains. Identification is based on Gaussian functions constructed for every specific nucleus using information on average chemical shifts and standard deviations (Kraulis, 1994) taken from the ANSIG dictionary file. All functions are normalized and the score for a nucleus with a certain chemical shift is calculated as the height of the curve at that chemical shift. Given a temporary residue all side chains (including C^{α}) having at least as many different carbon nuclei as the temporary residue and for which the scores for all nuclei are higher than a threshold value, are considered potential side chains. A tool to graphically show cross peaks used for side-


Ansig for Windows
File Graphics Crosspeaks Assign Distance Violations Window Help
Assignment of 3D-spectra of HSQC type
Source
Reference spectrum
Nhsqc
hncacb
cbcaconh
cconh
hcconh
hcchtocsy
hcchcosy
hnha
Temp CP
x107
x2
x3
x4
x5
x6
x7
x8
Show in current
Mark
Search spectrum
Search dim
Search
Show in current
Assign to temp
Mark
Clean
Delete
Done
Nucleus
Choose nucleus to assign to crosspeak
ANSIG w2
13C
40
9
8.9
8.8
1H 8.7
ANSIG w1 CURRENT
15N
x107N:x107n
132
9
8.9
8.8
1H 8.7
Menu
command contours delete pick Nhsqc deassign
yes no crosspeaks merge auto hncacb sequence
current connections connect move cbcaconh residue
area assignments symmetry mark cconh nucleus
box color clear correspond hcconh F1 F2
ruler plane trace disconnect hcchtocsy F3 F4
window view toggle remove clean hcchcosy
abort 1H 15N 13C 13CO 31P hnha
1 Ala CA
2 Arg CB
3 Leu CD
4 Thr CD#
5 Ala CD1
6 Tyr CD2
7 Glu CE
8 Arg CE#
9 Arg CE1
peak_height cp_move
clear_cont
make_cont
unalias+1
unalias-1
ass_seq
ass_res
ass_nuc
show_ss

Reference spectrum: Nhsqc

Temp CP: x107

Search spectrum: cbcaconh

Search dim: ☐ F1 13C ☑ F2 15N ☑ F3 1H ☐ F4

Tolerance:
1H 0.05 ppm
15N 0.15 ppm

Result

Crosspeaks:
CP198
CP226

Info:

Closest temporary:	x107N:x107n
1H difference:	0.011 ppm
15N difference:	0.06 ppm
Second closest temporary:	x3N:x3n
1H difference:	0.269 ppm
15N difference:	1.09 ppm

● C ○ CD
○ CO ○ CE
○ CA ○ H
○ CB ○ HA
○ CG ○ HB
○ none ☑ -1

Figure 2. A typical arrangement of windows for assignment of sequential 3D spectra in the second step of the assignment procedure. The upper left window is the dialog box containing the tools for the assignment of 3D spectra. The upper right window is a spectrum window showing the contours of the CBCA(CO)NH spectrum in a ^{13}C-^{1}H plane. The lower right window is a spectrum window showing the ^{1}H-^{15}N HSQC spectrum together with cross peaks from the CBCA(CO)NH spectrum. The bottom window is the Ansig for Windows menu. The result of a search for cross peaks in the CBCA(CO)NH spectrum belonging to the temporary residue x107 is shown. Dashed rulers in both spectrum windows mark one of the found cross peaks, CP226. The lower spectrum window is used to confirm that the cross peak belongs to x107 and the upper spectrum window is used to confirm that the cross peak does not belong to a spurious peak.

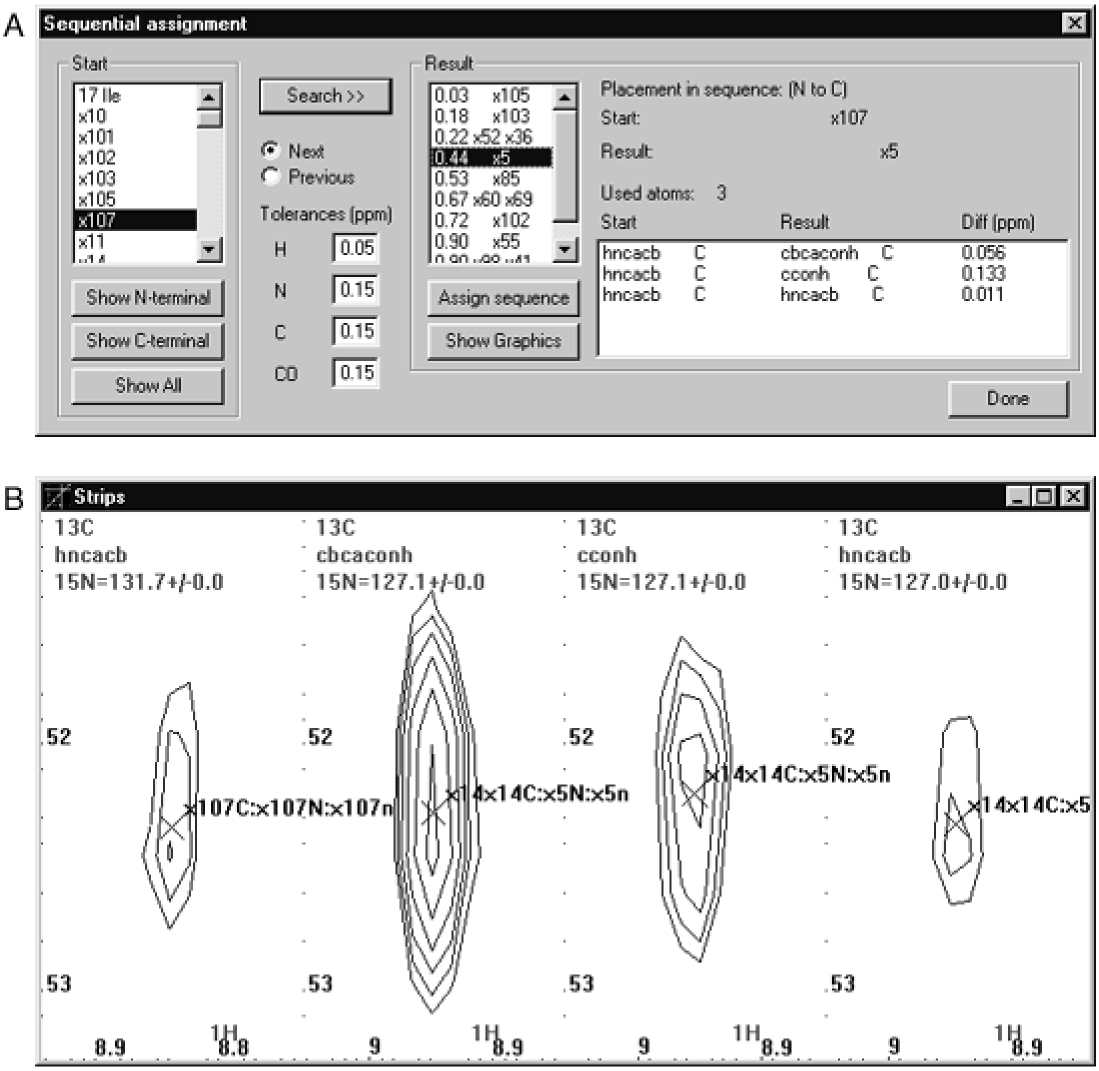
chain identification is available to verify the automatic assignment.

Following side-chain identification the protein sequence is searched for sites matching the temporary sequence. Matching sites are color coded and presented in the protein sequence together with already assigned residues. When assignment is performed the program updates the cross-peak database.

Presentation of distances in spectra

Assigning NOEs in NOESY spectra is often a relatively slow process. Our experience is that a preliminary structure with a correct fold usually is available early in the assignment process. This structure can be used to verify NOEs which cannot be unambiguously assigned based on chemical shift information. We have developed tools in Ansig for Windows to show distances from PDB structure files directly in NOESY spectra, thus making it easier to visualize potential NOE assignments.

When PDB files are loaded, the tool for showing distances can be invoked. This tool gives the user the ability to specify which residues and nuclei should be included in the distance calculations, which spectra chemical shift information should be taken from, which structures distance information should be extracted from, and the maximum distance which should be included. Distance information is then plotted in all ^{1}H-^{1}H planes currently available (Figure 4). Distances are presented as circles with a radius which is related to the distance – the shorter the distance the larger the radius (radius $= A - B * \text{distance}$). The circles are labeled with the assignment and distance, including a standard deviation if more than one struc-



17 Ile
x10
x101
x102
x103
x105
x107
x11

● Next
○ Previous

Tolerances (ppm)

H	0.05
N	0.15
C	0.15
CO	0.15

0.03	x105
0.18	x103
0.22	x52 x36
0.44	x5
0.53	x85
0.67	x60 x69
0.72	x102
0.90	x55

Placement in sequence: (N to C)
Start: x107
Result: x5
Used atoms: 3

Start		Result		Diff (ppm)
hncacb	C	cbcaconh	C	0.056
hncacb	C	cconh	C	0.133
hncacb	C	hncacb	C	0.011

Figure 3. Windows used in the sequential assignment. (A) The sequential assignment dialog box containing the tools for the sequential assignment of temporary residues into temporary sequences. (B) The strip window obtained by choosing the 'Show Graphics' button in the dialog box in (A).

ture was used. In the case of non-stereospecifically assigned prochiral pairs, the shortest distance is calculated and displayed. Distances for protons not assigned or partially assigned are presented in a list for further analysis.

Reading violations from CNS and X-PLOR out-files

Following structure calculations, information about violated restraints is available for each structure. This information is important to locate erroneous NOE assignments. Violations turning up in many structures are potential erroneous assignments. Violations just turning up in one or a few structures are potentially due to badly converged structures. Extracting this information from the violation lists requires some effort.

We have implemented functions in Ansig for Windows to directly extract violations from CNS or X-PLOR output files to make it easier to find the desired information. The program adds violations cumulatively to a list so that one single violation never occurs more than once, but statistics based on all occurrences are shown instead. The displayed information includes involved nuclei, number of structures containing the violation, average energy, maximum energy, restraint distance, average distance in structures, the average

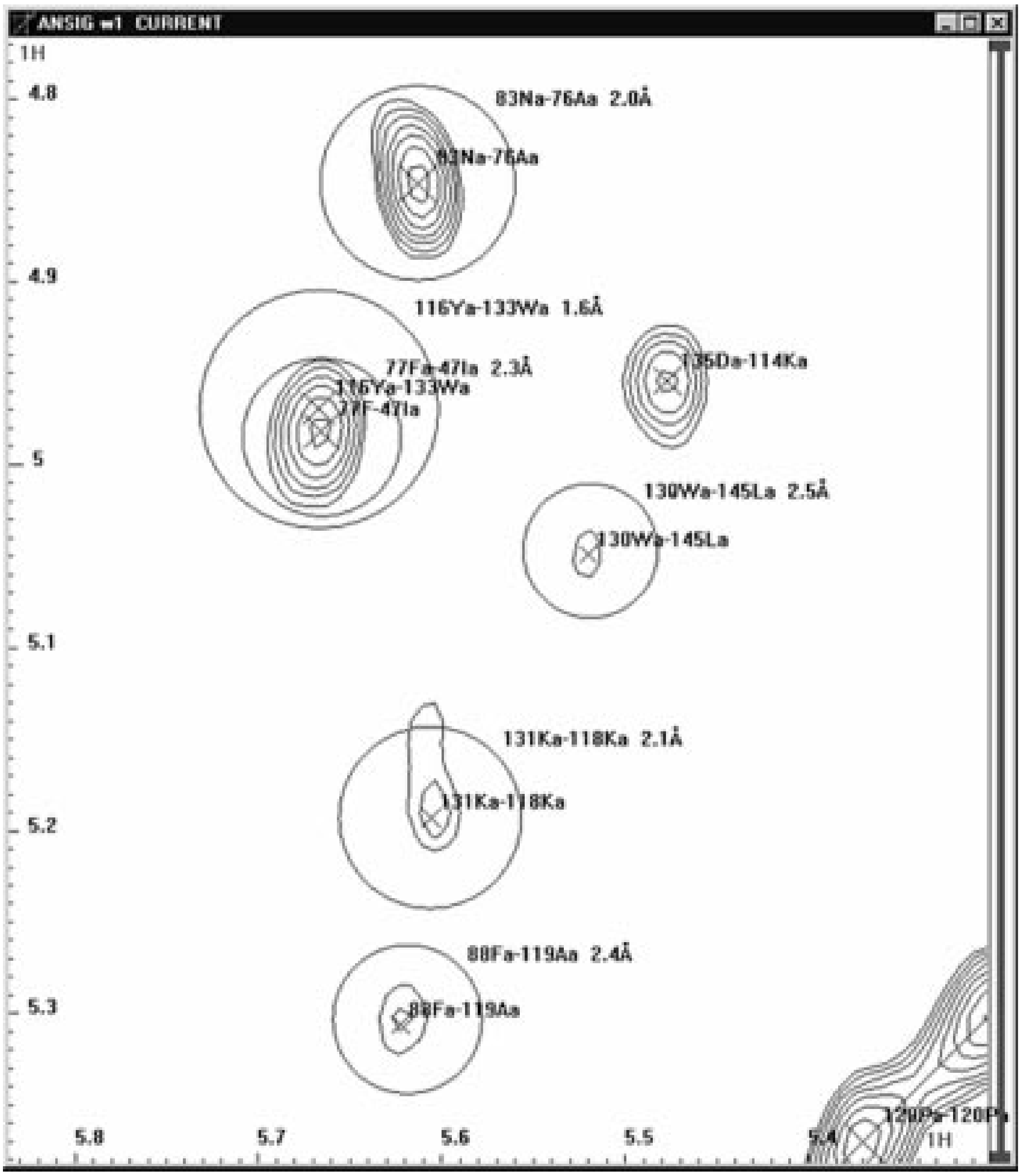


Figure 4. A spectrum window showing the H^{α}-H^{α} region of a 2D homonuclear NOESY spectrum in which distances calculated from a PDB structure file are illustrated as circles. Both distance circles and cross peaks are labeled with their assignments. The radii of the distance circles are related to the distances as described in the text. Distance information is included in the labels.

deviation from the restraint distance and the maximum deviation from the restraint distance. The user can then easily zoom to a violation in a NOESY spectrum by selecting it from a list.

Other features in the program

Ansig for Windows includes other features not available in earlier versions of ANSIG. Some examples are printing of spectrum windows and searching for and zooming to cross peaks. Ansig for Windows also includes options for turning off parts of the graphics during translation and zooming, which speeds up the graphics on computers with slow processors and/or graphics cards.

Testing

Ansig for Windows has been tested with complete sets of data files originating from ANSIG version 3.3 used in our laboratory. All files were transferred directly to the PC using FTP and did, after adjustment of file paths in control, spectrum and initiation files, work without any other alterations. The program then looks and behaves as earlier versions of ANSIG. Sequential assignment procedures, plotting of distances and functions for reading and displaying violations were tested on different ongoing projects. The program is currently being used in our laboratory for the assignment of several new proteins.

Computer requirements and performance

The recommended computer configuration for Ansig for Windows is at least a 400 MHz Pentium II CPU, 128 MB of RAM and a graphics card with full support for OpenGL (e.g. NVIDIA TNT2). Ansig for Windows has been tested under Windows 98 and Windows NT but should also run under Windows 95 and Windows 2000. On a computer with the above configuration Ansig for Windows updates the graphics approximately at the same rate as ANSIG on a Silicon Graphics O2 (200 MHz). With an even faster graphics card (e.g. NVIDIA GeForce) and a faster CPU (e.g. Pentium III, 600 MHz), Ansig for Windows runs much faster on a PC than on a Silicon Graphics O2.

Availability

Information on how to obtain Ansig for Windows is available on http://www.csb.ki.se/nmr/AFW.html. Ansig for Windows is distributed together with the source code under a GNU license (see http://www.gnu.org/copyleft/gpl.html for more information on the license).

Conclusions

We have taken a powerful and commonly used computer program, ANSIG, and updated it to include tools for semiautomatic analysis of NMR data for protein structure determination. We have also ported the program to Windows to make it available on a more common and cheaper computer platform.

The sequential assignment scheme built into Ansig for Windows is based on the procedure one usually follows for manual assignment. This should make the program easy to use for the experienced user and it will at the same time guide the first-time user through the sequential assignment process. The complete functionality of the original program has been kept intact in Ansig for Windows, giving the experienced user the option to use manual assignment methods when these are preferred.

The function for showing distances graphically will decrease the time spent on assigning NOEs by indicating potential new assignments in two ways. First the lists of pairs of closely spaced nuclei found in PDB structure files, but for which complete chemical shift information is not available, give the user an indication on where to look for new assignments. Second, the function will help the user to assign NOEs including nuclei with almost identical chemical shifts.

The ability to use Ansig for Windows to quickly browse through violations extracted from output files has been much appreciated by the test users. The function dramatically reduces time spent on cross-peak validation and thus allows the user to proceed quickly and efficiently to obtain final refined structures.

We believe that interactive, but semiautomatic, analysis of NMR spectra is going to be the solution to the time-consuming assignment bottleneck in biomolecular NMR assignment. The method combines the expertise of the user with the capability of the computer to quickly search through large amounts of data; making it faster than traditional assignment and more rigorous than fully automated assignment. We believe Ansig for Windows to be a step in this direction.

Acknowledgements

The authors wish to thank Esmeralda Woestenenk and Magnus Wolf-Watz for providing spectra and ANSIG set-ups and for valuable comments on the program.

References

Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR,* **6**, 1–10.

Bartels, C., Güntert, P., Billeter, M. and Wüthrich, K. (1997) *J. Comput. Chem.*, **18**, 139–149.

Brünger, A.T. (1992) *X-PLOR (Version 3.1) A system for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr.*, **D54**, 905–921.

Cavanagh, J., Fairbrother, W.J., Palmer III, A.G. and Skelton, N.J. (1996) *Protein NMR Spectroscopy,* Academic Press, San Diego, CA.

Garrett, D.S., Powers, R., Gronenborn, A.M. and Clore, G.M. (1991) *J. Magn. Reson.*, **95**, 214–220.

Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson.*, **101**, 114–119.

Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603–614.

Kjaer, M., Andersen, K.V. and Poulsen, F.M. (1994) *Methods Enzymol.*, **239**, 288–307.

Kraulis, P.J. (1989) *J. Magn. Reson.*, **84**, 627–633.

Kraulis, P.J. (1994) *J. Mol. Biol.*, **243**, 696–718.

Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., Van Aken, T. and Laue, E.D. (1994) *Biochemistry,* **33**, 3515–3531.

Neidig, K.-P., Geyer, M., Görler, A., Antz, C., Saffrich, R., Beneicke, W. and Kalbitzer, H.R. (1995) *J. Biomol. NMR,* **6**, 255–270.

Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269,** 408–422.

Roberts, G.C.K. (Ed.) (1993) *NMR of Macromolecules,* IRL Press, Oxford.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids,* John Wiley & Sons, New York, NY.

Zimmerman, D., Kulikowski, C., Wang, L., Lyons, B. and Montelione, G.T. (1994) *J. Biomol. NMR*, **4**, 241–256.