Journal of Biomolecular NMR, 9 (1997) 101–104
ESCOM

J-Bio NMR 425

BBReader: A computer program for the combined use of the BioMagResBank and PDB databases

Reinhard Wimmer[a,b], Norbert Müller[a,b] and Steffen B. Petersen[a,c,d,*]

[a]*MR Center, SINTEF-UNIMED, N-7034 Trondheim, Norway*
[b]*Institute of Chemistry, Johannes Kepler University, A-4040 Linz, Austria*
[c]*SINTEF-UNIMED Biotech, N-7034 Trondheim, Norway*
[d]*Department of Biotechnology, Norwegian University of Science and Technology, N-7034 Trondheim, Norway*

Received 4 November 1996
Accepted 14 December 1996

Keywords: BioMagResBank; PDB database; Chemical shift assignment; Protein; Inverse search

Summary

A computer program ('BBReader') was developed which performs an inverse search in the BioMagResBank database. Given (cross) peak positions of a protein, the program searches for atoms with matching chemical shifts and suggests possible assignments for user-specified homo- and heteronuclear one- to three-dimensional COSY- and NOESY-type experiments. It can handle ^{1}H, ^{13}C and ^{15}N spectra. Distance information from PDB files can be utilized for filtering possible NOESY cross peak assignments.

In recent years, a huge amount of spectroscopic and structural data on proteins have been published. In order to facilitate the access to these data, several databases have been established. Most important for structural and NMR spectroscopic work are the PDB database (Bernstein et al., 1977; Abola et al., 1987) for protein structures and the BioMagRes databank (Seavey et al., 1991) for protein NMR data.

The data in each entry of the BioMagResBank are indexed by residue number and atom type. During work on a protein for which the NMR assignments are already known, such as interaction or folding studies, one needs a possibility for an inverse search – given the chemical shift(s) of a (cross) peak, one wants to find the corresponding atoms. We have developed an automated routine for accessing the information in the BioMagResBank in this way.

The program BBReader is designed for quick access to the assignment of protein resonances from peak positions in multidimensional spectra using preselected entries from BioMagResBank and PDB files. (The latter ones must contain the proton positions. Many modelling programs allow for automatic addition of protons to a PDB file.) BBReader basically searches through previously downloaded entries of the BioMagResBank to match chemical shifts. For cross peaks of 2D and 3D spectra, additional checks are performed in order to determine whether a cross peak between the identified atoms is spectroscopically possible. In the case of COSY spectra, they must show scalar coupling; in the case of NOESY spectra, their distance must be within certain limits. The distances are calculated from a PDB file (if available).

Since the matches are not exact, a score is calculated to evaluate the quality of each hit. The score is based on the sum of squared deviations of the chemical shifts from the input, and on the internuclear distance(s) if a PDB file was used. All possible combinations (within a predefined range of shifts) are checked, and a hit list, ranked by the score, is printed.

The program can handle several spectrum types (1D, 2D and 3D spectra with any combination of ^{1}H, ^{13}C and ^{15}N; see Figs. 1 and 2 for examples) and user-imposed restrictions. Coherence transfer between ^{1}H and heteronuclei as well as between heteronuclei is assumed to be of COSY type, i.e. to be based on scalar coupling; for a coherence transfer between two proton dimensions, a choice between COSY type (scalar coupling) and NOESY type (dipolar coupling) is possible.

*To whom correspondence should be addressed at SINTEF-UNIMED Biotech.
Abbreviations: NOESY, nuclear Overhauser enhancement spectroscopy; COSY, correlation spectroscopy; WWW, world-wide-web.

```
1% BBReader
Program BBReader version 0.9b is running
Input BIOMAG file name :hewlys.dat
Chemical shift data is available for the following nuclei:
        1H
How many dimensions should your spectrum have? (1-3) : 2
Only 1H available from the BMRB-file! Assigning 1H to all dimensions!
Is the coherence transfer between
        dimension 1 and 2 of COSY or NOESY-type? (c/n) : n
Input PDB file name (enter "0", if you wish to work with
       BioMagRes-Files only) : hew.pdb
PDB File hew.pdb read
Do you want a logfile written (will be named 'RBMlog.out')? (y/n) : y
BIOMAG File hewlys.dat read
Do you like: - a list of possible cross-peaks within a shift-range (l)
      or  - an ordered list of the best hits for a certain cross-peak?  (r)
(l/r) : r
Input first dimension (nucleus H) chem shift : 4.05
Input second dimension (nucleus H) chem shift : 7.08

GLY 104 HA2/HA3 4.06 ppm | MET 105       H 7.06 ppm | nearest 2.27 A|sc= 98
GLY 104 HA2/HA3 4.06 ppm | TYR  23 HD1/HD2 7.05 ppm | nearest 5.28 A|sc= 88
ILE  58      HA 4.02 ppm | TRP 108     HD1 7.10 ppm | dist    5.52 A|sc= 83
ARG  61      HA 4.10 ppm | TRP  62     HD1 7.11 ppm | dist    5.30 A|sc= 78
THR  40      HA 4.07 ppm | PHE   3 HD2/HD1 7.03 ppm | nearest 5.51 A|sc= 77
ALA  32      HA 4.02 ppm | TRP 108     HD1 7.10 ppm | dist    6.00 A|sc= 75
GLY   4 HA2/HA3 4.01 ppm | PHE   3 HD1/HD2 7.03 ppm | nearest 5.52 A|sc= 72
ILE  58      HA 4.02 ppm | TYR  53 HD2/HD1 7.12 ppm | nearest 5.97 A|sc= 70
GLU   7      HA 4.12 ppm | PHE   3 HD1/HD2 7.03 ppm | nearest 4.75 A|sc= 67
SER 100 HB2/HB3 4.11 ppm | TRP  63     HH2 7.08 ppm | nearest 6.10 A|sc= 63
GLY 104 HA2/HA3 4.06 ppm | TRP 111     HZ3 7.04 ppm | nearest 6.55 A|sc= 61
GLY   4 HA2/HA3 4.01 ppm | PHE  38 HD1/HD2 6.99 ppm | nearest 4.42 A|sc= 60
TRP 123      HA 4.12 ppm | TRP 123     HZ3 7.14 ppm | dist    5.09 A|sc= 60

Data statistics for 1H (from available residues!):
      The molecule contains 959 1H-nuclei
            of which 263 are exchangeable.
      951 protons are in the PDB-file
      and 683 are in the BioMagRes-file.
```

Fig. 1. Sample session for a NOESY spectrum of hen egg-white lysozyme (assignment of Redfield and Dobson, 1988). User input is printed in bold letters. Note that diastereotopic protons are combined into groups and that the distance information is labelled accordingly.

For NOESY-type spectra the use of 3D structural information is possible. The PDB file is used to select cross peaks between protons that are too far away from each other in the 3D structure to give a NOESY cross peak. If no PDB file is given, the program will output all cross peaks that seem possible according to the chemical shifts. Two modes of operation are possible: in the first mode the user inputs one chemical shift per dimension and BBReader looks for those cross peaks that come closest to the input. These will be printed in an ordered way along with the score. The second mode allows the input of a chemical shift range for each dimension (and

```
1% BBReader  -3rp  HcCcN
Program BBReader version 0.9b is running

Input BIOMAG file name :bmr1657.str

Chemical shift data is available for the following nuclei:
        1H
        13C
        15N
3D-spectrum chosen.
3D H-C-N spectrum chosen.
Coherence transfer between dimension 1 and 2 is of COSY-type.
Coherence transfer between dimension 2 and 3 is of COSY-type.
Logfile will be generated and printed.
BIOMAG File bmr1657.str read

Ordered list of the best hits for a certain cross-peak chosen.
Input first dimension (nucleus H) chem shift : 4.54

Input second dimension (nucleus C) chem shift : 56.7

Input third dimension (nucleus N) chem shift : 115.3

SER  71 HA 4.52 ppm | SER  71 CA 56.80 ppm | SER  71 N 115.50 ppm | sc= 99
SER  36 HA 4.58 ppm | SER  36 CA 56.20 ppm | SER  36 N 113.90 ppm | sc= 84
ASN 130 HA 4.45 ppm | ASN 130 CA 56.60 ppm | ASN 130 N 117.20 ppm | sc= 59
GLN  80 HA 4.47 ppm | GLN  80 CA 55.10 ppm | GLN  80 N 117.30 ppm | sc= 52
ASP 108 HA 4.56 ppm | ASP 108 CA 56.60 ppm | ASP 108 N 118.40 ppm | sc= 48
MET  47 HA 4.44 ppm | MET  47 CA 55.50 ppm | MET  47 N 118.20 ppm | sc= 20

Data statistics for 1H (from available residues!):
      The molecule contains 1215 1H-nuclei
            of which 296 are exchangeable.
      and 317 are in the BioMagRes-file.

Data statistics for 13C (from available residues!):
      The molecule contains 776 13C-nuclei
            and 122 are in the BioMagRes-file.

Data statistics for 15N (from available residues!):
      The molecule contains 227 15N-nuclei
            and 148 are in the BioMagRes-file.
```

Fig. 2. Sample session for a 3D HCN spectrum of ribonuclease H from *Escherichia coli* (assignment of Yamazaki et al., 1991), demonstrating the simplifications for users by the introduction of command-line arguments. User input is printed in bold letters.

a distance threshold, if a PDB file is specified) and will print all (possible) cross peaks that fall within this range (in all dimensions simultaneously). Along with the output of a potential NOESY cross peak, distance information will be given; along with COSY-type cross peaks all non-vicinal couplings between protons will be labelled. (Allylic couplings are only available between protons in aromatic systems.)

Diastereotopic protons are given special treatment, because the stereospecific assignment is not known in the majority of cases. They are combined into a group, which is indicated in the output (see Fig. 1). The closest distance in the PDB structure between any proton belonging to a group and its potential NOE partner (which can be a single atom as well as another group) will be given in the NOESY mode and this will also be marked clearly in the output (Fig. 1). Isochronous nuclei like $H^{\delta 1}/H^{\delta 2}$ and $C^{\delta 1}/C^{\delta 2}$ in phenylalanine and tyrosine are treated in a similar way as diastereotopic protons, as are methyl groups.

Currently the program is limited to the 20 'classical' L-amino acids, but it is written in such a way that it can be easily extended. It can work with BioMagRes files in the STAR flat-file format (Hall, 1991; Hall and Spadaccini, 1994; Hall and Cook, 1995), as can be obtained from the WWW site (http://www.bmrb.wisc.edu), and with PDB files in the current format available also via WWW (http://www.pdb.bnl.gov).

The program is written in C and is designed to run under UNIX operating systems; it does not provide any graphical user interface. The source code and the manual can be obtained from the BioMagResBank server (http://www.bmrb.wisc.edu/bbreader/BBReader.html). We have tested the program on several SGI workstations running under IRIX5.3 and 6.01 and on an HP9000/735. BBReader is copyrighted, but freely available; more details are given in the manual.

Acknowledgements

This work was in part supported by the FWF (Austrian Science Foundation, project P10571-CHE) and by the NFR (Norges forskningsråd, KAS scholarship to R.W.). We are also grateful to LIZENS (Linz Center for Supercomputing) for support and access to computing resources.

References

Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J. (1987) In *Crystallographic Databases - Information Content, Software Systems, Scientific Applications: Protein Data Bank* (Eds., Allen, F.H., Berghoff, G. and Sievers, R.), Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, pp. 107–132.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Hall, S.R. (1991) *J. Chem. Inf. Comput. Sci.*, **31**, 326–333.

Hall, S.R. and Spadaccini, N. (1994) *J. Chem. Inf. Comput. Sci.*, **34**, 505–508.

Hall, S.R. and Cook, A.P.F. (1995) *J. Chem. Inf. Comput. Sci.*, **35**, 819–825.

Redfield, C. and Dobson, C.M. (1988) *Biochemistry*, **27**, 122–136.

Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, L. (1991) *J. Biomol. NMR*, **1**, 217–236.

Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H. and Nagayama, K. (1991) *Biochemistry*, **30**, 6036–6047.