



Program MULDER – A tool for extracting torsion angles from NMR data

Petr Padrta & Vladimír Sklenář*

National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic

Received 5 August 2002; Accepted 4 October 2002

Key words: computer graphics, interatomic distance, NMR, scalar coupling constant, software, torsion angle

Abstract

MULDER (Mostly Universal Dihedral angle ExtractoR) is a program for extraction of torsion angle information from NMR data. Currently, it can analyze two types of input data: *The torsion angle data*, where several 3J -coupling constants and/or interatomic distances are combined in order to reduce the torsion angle ambiguity arising from solving the isolated Karplus (or distance) equation, and *the sugar pucker data*, where the dynamics of five-membered sugar rings is evaluated by postprocessing the results calculated from $^3J_{(HH)}$ coupling constants by program PSEUROT. Program MULDER can be used either as an alternative to r-MD programs in situations where only specific structural features are studied, or as a preparatory tool in connection with full r-MD structure calculation for extraction of unambiguous torsion angle restraints.

Abbreviations: NMR – Nuclear Magnetic Resonance; NOE – Nuclear Overhauser Effect; RDC – Residual Dipolar Coupling; MD – Molecular Dynamics; r-MD – Restrained Molecular Dynamics; GUI – Graphical User Interface; PDF – Portable Document Format.

Introduction

During the past few years nuclear magnetic resonance has become an indispensable tool for the determination of the three-dimensional structure of proteins and nucleic acids in solution. Wealth of structure-related data can be extracted from multi-dimensional NMR spectra of isotopically labeled biopolymers. However, information contained in NMR data describes a structure only indirectly and must be converted into particular structural parameters. For example, vicinal three-bond coupling constants 3J carry information about torsion angles, which can be obtained by solving the empirical Karplus equation in the form $^3J(\phi) = P_1 \cos^2(\phi + \text{phase}) + P_2 \cos(\phi + \text{phase}) + P_3$. The main problem stems from the fact that, due to the \cos^2 function dependence, there may be up to four solutions to the equation. Nevertheless, the degeneracy remains unresolved only in the cases where a single Karplus

equation is used to extract the data for one torsion angle. If a torsion angle can be monitored by more 3J -coupling constants, each one resulting in several solutions, then the correct value(s) should reveal itself by being common to all used 3J -couplings. Assuming that the 3J -couplings and/or torsion angles are used as restraints in a r-MD calculation, the degeneracy may be dealt with in one of the following ways:

- (1) All available 3J -couplings are used directly in a r-MD calculation, along with appropriate Karplus parameterizations, together with other accessible NMR data (e.g., NOEs, RDCs, etc).
- (2) The degeneracy is removed or at least partially reduced outside MD calculations by exploiting several 3J -couplings for each torsion angle and solving a set of Karplus equations. Only the result(s) consistent with all the equations are then used in subsequent r-MD calculations.

The first approach seems to be more convenient since it saves additional data manipulations. However, the procedure is less transparent and the rejec-

*To whom correspondence should be addressed. E-mail: Sklenar@chemi.muni.cz

tion of false solutions in the course of the numerical optimization is difficult to trace. Moreover, r-MD programs supporting direct inclusion of ^3J -couplings (e.g., XPLOR (Brünger, 1992), CNS (Brünger et al., 1998), AMBER (Pearlman et al., 1995), etc.) employ by default only the basic form of the Karplus equation. The use of more sophisticated equations, like those proposed by Haasnoot et al. (1980) or Donders et al. (1989), requires modifications in the source codes of r-MD programs.

Existing implementations of the second approach are usually based on the principle of multi-dimensional grid search, considering whole molecular fragments, with torsion angles as the only degrees of freedom. Thus the programs like ANGLESEARCH (Polshakov et al., 1995), HYPER (Tejero et al., 1999), and algorithm FOUND (Güntert et al., 1998) for program DYANA (Güntert et al., 1997) use a number of J-couplings and/or interatomic distances to reduce the degeneracy of several torsion angles simultaneously. Nevertheless, similarly to a r-MD-based evaluation, the programs based on the multi-dimensional grid search are less transparent, lacking a simple visualization of the obtained results. On the other hand, a simplification to an independent one-dimensional grid search allows for a graphical representation of the process of degeneracy removal. Program MULDER (Mostly Universal Dihedral angle ExtractoR) has been written to provide such a solution. The unambiguous torsion angles calculated by program MULDER can be used as restraints in a r-MD calculation, possibly in combination with those ^3J -couplings which could not be unambiguously converted into torsion angles. However, MULDER can be used also as an alternative to r-MD programs in situations where only specific structural features are studied.

In addition to the degeneracy problem, the extraction of torsion angles from NMR data may be further complicated by a fast (compared to NMR time scale) inter-conversion of conformations, resulting in an averaging of measurable ^3J -coupling constants. The conformational flexibility of (deoxy)ribose rings in nucleic acids represents a typical example. For five-membered rings, program PSEUROT (van Wijk et al., 1995; de Leeuw and Altona, 1983) is most frequently used to assess the dynamic equilibrium between sugar conformers. The program works in the two-state approximation based on the assumption that the conformation, reflected in the experimental sugar $^3\text{J}_{(\text{HH})}$ coupling constants, can be described by dynamical equilibrium between two sugar conformers. PSEU-

ROT tries to find these conformers by calculating the theoretical ^3J -couplings from two user-supplied starting conformers and minimizing the differences between the experimental and theoretical ^3J -coupling constants. However, as pointed out by authors themselves (van Wijk et al., 1995), the Newton-Raphson minimization employed generally converges to the nearest local minimum. Consequently, the result depends on the chosen conformations of the starting pair which always leaves certain degree of uncertainty as to whether the real global minimum (in the two-state approximation) was found. Fortunately, program PSEUROT offers a possibility (*via* procedure MANY) to vary automatically the starting conformers in order to scan the hypersurface more completely. Typically, several thousands of sugar conformer pairs per residue (e.g., 2400 pairs for four sugar $^3\text{J}_{(\text{HH})}$ couplings) are optimized, arriving at the same number of resulting pairs. Each resulting pair fulfills the experimental data to a certain extent. Depending on the starting conformations and on the structure of the hypersurface, the result contains also pairs which do not fulfill the experimental J-couplings very well or even conformer pairs with unrealistic sugar geometries (e.g., sugars with a completely flat pucker). To sort the result of the PSEUROT analysis may be a tedious task since PSEUROT does not offer any filtration procedure. In order to facilitate such an analysis, program MULDER has been extended to postprocess the PSEUROT output by filtering the result based on the user-defined criteria. The filtering procedure is facilitated by a graphical display of important characteristics of the PSEUROT output. For conformer pairs matching the experimental results, the exo-/endo-cyclic torsion angles can be calculated from the pseudorotation parameters P and Φ_m (Altona and Sundaralingam, 1972). The calculated sugar torsion angles can be used either in a subsequent r-MD calculation or simply as a qualitative result if the complete three-dimensional structure determination is not a primary goal.

Methods

Equations and parameterizations

Torsion angle data

In order to obtain structure-related parameters, the dependence of NMR data (e.g., ^3J -coupling constants, interatomic distances, chemical shifts) on a torsion angle must be expressed analytically using correctly pa-

parameterized equations. The philosophy used in MULDER is to support the commonly used equations and supply parameterizations for the usual cases. Unlike the equations, which are part of the program code itself, the parameterizations can be changed or modified by the user on the run without the need to recompile the program. Currently, MULDER supports two equations for 3J -couplings (Equations 1 and 2) and one equation for interatomic distances (Equation 3).

The Karplus equation (Karplus, 1959, 1963):

$$^3J(\phi) = P_1 \cos^2(\phi + phase) + P_2 \cos(\phi + phase) + P_3. \quad (1)$$

By default, MULDER contains parameterizations of Equation 1 for torsion angles β , ϵ , and χ in nucleic acids (Wijmenga and van Buuren, 1998), and for torsion angle ϕ in proteins (Hu and Bax, 1997).

The shortened Haasnoot equation (Haasnoot et al., 1980), i.e., without the sum of β -substituent corrections and without the Barfield correction (Marshall et al., 1976; de Leeuw et al., 1983):

$$^3J(\phi) = P_1 \cos^2(\phi + phase) + P_2 \cos(\phi + phase) + P_3 + \sum_{i=1}^4 \Delta\chi_i (P_4 + P_5 \cos^2(\xi_i\phi + P_6|\Delta\chi_i|)). \quad (2)$$

By default, MULDER contains parameterizations of Equation 2 for torsion angle γ in nucleic acids (Wijmenga and van Buuren, 1998).

The interatomic distance equation (derived from geometry)

$$d^2(\phi) = A \cos(\phi + phase) + B \sin(\phi + phase) + C, \quad (3)$$

where d is the distance between the interacting atoms, the parameters A , B , and C describe lengths, angles, and torsion angles (known from an independent analysis) of bonds connecting the interacting atoms.

By default, MULDER contains parameterizations of Equation 3 for the torsion angles γ and χ in nucleic acids. The parameters A , B , and C for these angles were calculated from the corresponding bond lengths and angles in standard guanine and cytosine with the sugar pucker being fixed in the S-region ($P = 192^\circ$, optimized by AMBER 6.0).

Sugar pucker data

The pseudorotation parameters P (phase) and Φ_m (amplitude) contained in sugar pucker data are converted into torsion angle information using the pseudorotation equation (Altona and Sundaralingam, 1972; van Wijk et al., 1995):

$$\phi_i = A_i \Phi_m \cos(P + phase) + B_i. \quad (4)$$

By default, MULDER contains the parameterization for the exocyclic torsions of β -D-deoxyribose (van Wijk et al., 1995) but, similarly to torsion angle data equations, other parameterizations can be added without the recompilation of the program.

Implementation

Program MULDER is a command line program, currently with no graphical user interface, but capable of creating color, vector graphical output in PDF format (Adobe Systems Incorporated, 2001).

The program is written in the ANSI-C programming language, utilizing the ClibPDF graphical library (<http://www.fastio.com>) for the graphical output generation. Due to its command line nature, the program imposes no special requirements on graphical hardware and computational power. The code has been written to be easily portable. So far, the program has been successfully used under various UNIX platforms (*BSD, IRIX, Linux) and in the MS Windows environment (Cygwin port).

The program follows the UNIX philosophy, similarly to NMRPipe (Delaglio et al., 1995), a widely used software for the processing of NMR spectra, relying on concepts of data streams, filters, and pipes. In short, a typical UNIX program behaves like a filter, which takes its input *via* a UNIX pipe from the previous program and prints its output *via* another pipe into the subsequent program. All of the pipe communication is done in memory without a need of additional disk space. MULDER thus concentrates mostly on the calculations, leaving the task of data conversions from/to other data formats than its own to other programs and communicates with them *via* UNIX pipes. A more detailed description of the UNIX pipeline philosophy was provided, for instance, by Delaglio et al. (1995). The examples of MULDER applications and the up-to-date information about the program behavior and command line options can be found in the current version of the program manual, distributed with the program.

| residue | spins | E_{low} | J d | E_{up} | rflag |
|---------|---------|------------------|--------|-----------------|-------|
| G1 | Jh3'p | -0.7 | 5.7380 | +0.6 | R |
| G1 | Jc4'p3 | | large | | U |
| G1 | Jc2'p3 | | small | | U |
| G1 | dh1'h8 | | 3.4 | | U |
| G1 | dh2'h8 | | 2.2 | | U |
| G1 | dh2''h8 | | 3.1 | | U |

Figure 1. Example of the torsion angle data input format (lower rectangle). The labels to the individual data columns shown in the upper rectangle are not part of the actual input.

Program availability

Program MULDER is distributed upon the registration from the program homepage at <http://ncbr.chemi.muni.cz/mulder/>. The program is available for several platforms, including Linux and Windows.

Input data formats

The MULDER input is a standard ASCII file or stream (for the pipe communication) with a support for comments (any line beginning with the symbol '#') and with no limitation on the total number of data lines. Currently, two basic types of data are recognized, i.e., the torsion angle data and the sugar pucker data. Although handled separately (see the following two sections), both types may be freely combined into one input file/stream.

Torsion angle data

The torsion angle data represents either ^3J -coupling constants or interatomic distances (e.g., NOE derived H-H distances), both with their respective experimental errors. An example of the basic format is shown in Figure 1. The first column (residue) typically contains both the residue name and the residue number, the second column (spins) specifies the interacting nuclei and has to match one of the parameterizations known to the program. The following columns contain the value of a ^3J -coupling or a distance (J|d) and the errors in a format compatible with program CHEOPS (Macaya et al., 1992). Note that both the errors are required for the program to work. If the errors are not supplied, MULDER uses the default values (0.5 Hz for ^3J -couplings and 0.5 Å for distances). The program also offers the possibility to specify the J|d value qualitatively based on the assumption that only an estimate of the value is available. In such a case, the symbols 'L' and 'S' may be used for large and small values,

respectively, or the symbols '>' (greater than) and '<' (less than) may be used for more accurate estimates. For example, when a ^3J -coupling is estimated to be larger than 8 Hz it is sufficient to write '>8' into the J|d column. Finally, the column rflag contains the reliability flag, which controls the way the data is used in the search for the common solutions (see section Input data processing).

Sugar pucker data

The sugar pucker data is the output of procedure MANY of program PSEUROT. It represents all combinations of possible pairs of sugar conformers, each combination (input line) fulfilling, to a various degree, the experimental H-H couplings. A detailed description of the PSEUROT approach and of the PSEUROT output can be found in the original paper (de Leeuw and Altona, 1983) and/or in the PSEUROT manual (van Wijk et al., 1995). A snippet from a sugar pucker data input file is shown in Figure 2. The first two columns of the input (counter and residue), which are not part of PSEUROT MANY output, can be generated by MULDER auxiliary program SCONV. The columns P_1 , $\Phi_{m,1}$, P_2 , and $\Phi_{m,2}$ are the pseudorotation parameters of conformers 1 and 2, X_2 is the fraction of the second conformer. The columns $\Delta J_1 \dots \Delta J_4$ are the differences between the experimental coupling values $J_{i,\text{exp}}$ and their calculated counterparts $J_{i,\text{calc}}$. The column RMS is the root mean square deviation of the calculated couplings $J_{i,\text{calc}}$ from the experimental couplings $J_{i,\text{exp}}$. The column cflag denotes the type of the PSEUROT convergence.

Input data processing

The program treats each input line separately. Based on the input data format, MULDER analyses the data either as the torsion angle data or the sugar pucker data.

Torsion angle data

MULDER starts the analysis of torsion angle data by calculating the lower and the upper limits from the supplied values and their errors ($J_{\text{low}} = J - |E_{\text{low}}|$, $J_{\text{up}} = J + |E_{\text{up}}|$). The string in the column spins is used to find an appropriate equation and its parameterization from the MULDER database. The equation is numerically solved, both as a discrete problem (the value in column J|d gives the torsion angle ϕ_i) and an interval problem (the lower and the upper bounds of J|d give the solution interval $\phi_{i,\text{low}} - \phi_{i,\text{up}}$).

| counter | residue | RMS | P ₁ | Φ _{m,1} | P ₂ | Φ _{m,2} | cflag | X ₂ | ΔJ ₁ | ΔJ ₂ | ΔJ ₃ | ΔJ ₄ |
|---------|---------|-------|----------------|------------------|----------------|------------------|-------|----------------|-----------------|-----------------|-----------------|-----------------|
| 1870 | A4 | 0.350 | 14.0 | 36.0 | 161.5 | 36.2 | #i | 1.00 | 0.22 | -0.52 | 0.12 | -0.39 |
| 1871 | A4 | 5.251 | 14.0 | 36.0 | 271.0 | 36.3 | *R* | 0.94 | 9.50 | -2.00 | -2.52 | -3.10 |
| 1872 | A4 | 5.752 | 14.0 | 36.0 | 286.9 | 38.0 | c | 0.99 | 10.00 | -1.78 | -2.58 | -4.74 |
| 1873 | A4 | 6.111 | 14.0 | 36.0 | 298.3 | 39.3 | *R* | 1.00 | 10.11 | -1.58 | -2.30 | -6.28 |
| 1874 | A4 | 5.504 | 14.0 | 36.0 | 275.0 | 48.1 | *R* | 0.98 | 10.04 | -1.27 | -2.63 | -3.45 |
| 1875 | A4 | 6.324 | 14.0 | 36.0 | 301.4 | 44.0 | *R* | 0.90 | 10.01 | -1.24 | -1.96 | -7.38 |
| 1876 | A4 | 6.726 | 14.0 | 36.0 | 317.8 | 41.5 | c | 0.80 | 9.95 | -1.60 | -1.38 | -8.80 |
| 1877 | A4 | 6.897 | 14.0 | 36.0 | 327.2 | 41.6 | c | 0.80 | 9.94 | -1.59 | -1.06 | -9.37 |
| 1878 | A4 | 7.040 | 14.0 | 36.0 | 336.5 | 42.1 | c | 0.80 | 9.95 | -1.64 | -0.75 | -9.79 |
| 1879 | A4 | 7.158 | 14.0 | 36.0 | 346.2 | 42.9 | c | 0.80 | 9.98 | -1.78 | -0.49 | -10.10 |
| 1880 | A4 | 7.256 | 14.0 | 36.0 | 356.0 | 44.4 | c | 0.80 | 10.00 | -1.98 | -0.24 | -10.32 |

Figure 2. Example of the sugar pucker data input format (lower rectangle). The labels to the individual data columns shown in the upper rectangle are not part of the actual input.

In the next step, the results for a particular torsion angle are grouped together. In each group, the common solutions are found. When a common intersection of solution intervals exists in the group, the upper and lower bounds of the final torsion angle interval are determined. A torsion angle interval is given by the errors of all solution intervals marked with the reliability flag 'R' in the column rflag of the input data. Solution intervals marked with the flag 'U' are not used to define the upper and the lower bounds of the final torsion angle, however, their overlap with the defined torsion angle interval is still required. The column rflag thus allows to distinguish the reliable experimental data, used to define the torsion angle bounds, from the data estimated qualitatively, which may be still used to exclude false solutions. As the final step, the text (not shown) and graphical (Figures 3 and 4) outputs are generated.

Sugar pucker data

If the line contains sugar the pucker data, MULDER applies a set of user-defined filters and accepts only those lines which satisfy the selection criteria. The following filtering criteria are available:

ΔJ-filter – Sets the limits for the differences between the measured and the calculated sugar J-couplings (columns ΔJ₁...ΔJ₄ in Figure 2).

RMS-filter – Defines the maximum allowed root-mean square deviation between the experimental and the calculated J-couplings (column RMS).

Φ_m-filter – Specifies the ranges of sugar amplitudes Φ_m of both conformers (columns Φ_{m,1} and Φ_{m,2}). This filter is used mainly to ensure that the sugar

pucker calculated by PSEUROT is realistic. The flat (Φ_m ~ 0°) as well as too distorted (e.g., Φ_m > 45°) conformations can be readily rejected.

X₂-filter – Sets the minimum fraction of the second conformer (column X₂).

Flag-filter – PSEUROT convergence flag (column cflag) must be equal to specified string to accept the record.

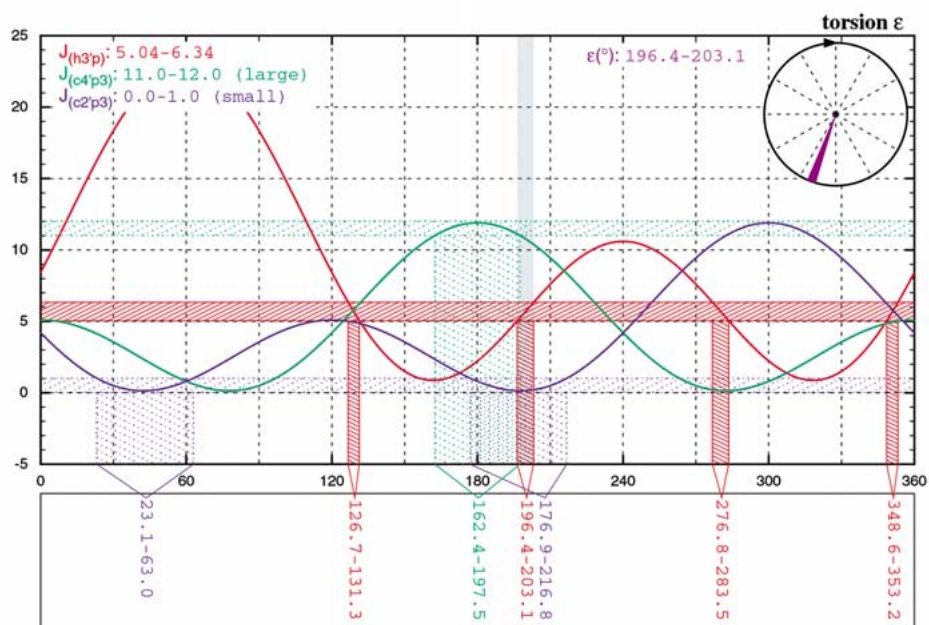
N-filter – Sets the number of accepted records. This filter is based on the idea that a large number of records (for instance 500 out of 2400 starting pairs) found in the same conformational region indicates the most favorable solution.

As the second step, the exo- and/or endo-cyclic sugar torsions are calculated from the pseudorotation phase P and the amplitude Φ_m of both sugar conformers using the Equation 4. Subsequently, all accepted records are grouped according to the residue column and in each group, the final intervals of P₁, Φ_{m,1}, P₂, Φ_{m,2}, and of all sugar torsions are calculated. In the final step, the text (not shown) and graphical outputs (Figure 5) are generated.

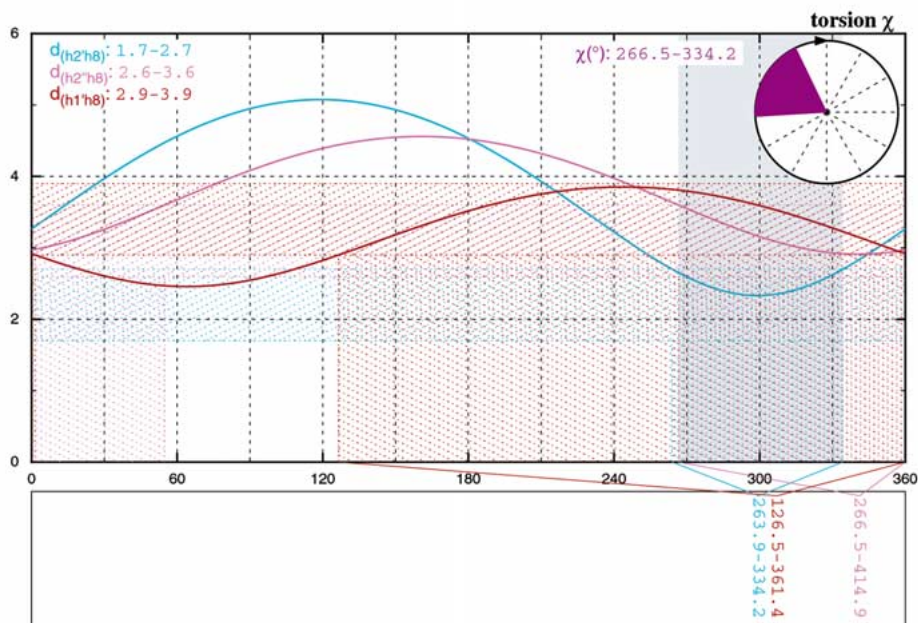
Results and discussion

Program MULDER was tested on the data obtained for various DNA fragments in our laboratory, including d(CCCCGGG) (Trantírek et al., 2000) and d(GGGCCCC) (Štefl et al., 2001). The most extensive testing was performed during the structure refinement of the d(GCGAAGC) hairpin (Padrta et al., 2002) where MULDER was used to prepare restraints

residue: G1



(a)



(b)

Figure 3. Example of the torsion angle data graphical output – d(GCGAAGC) hairpin, residue G1, the torsion angles ϵ (panel a) and χ (panel b), the data by Padrta et al. (2002). The items of input data, shown in the upper left hand corner, are represented by the curves of the same color in the graph. The partial solutions corresponding to the individual curves appear in the graph as bands of their respective colors and the numerical values are given in the rectangle below the graph. The final solutions are given in the upper right hand corner in both numerical and graphical forms and also shown in the graph as gray bands. The reliable data ($J_{(h3'p)}$), as defined by the rflag value in the input, are depicted by solid hatching, the unreliable data by dashed hatching.

residue: Q2

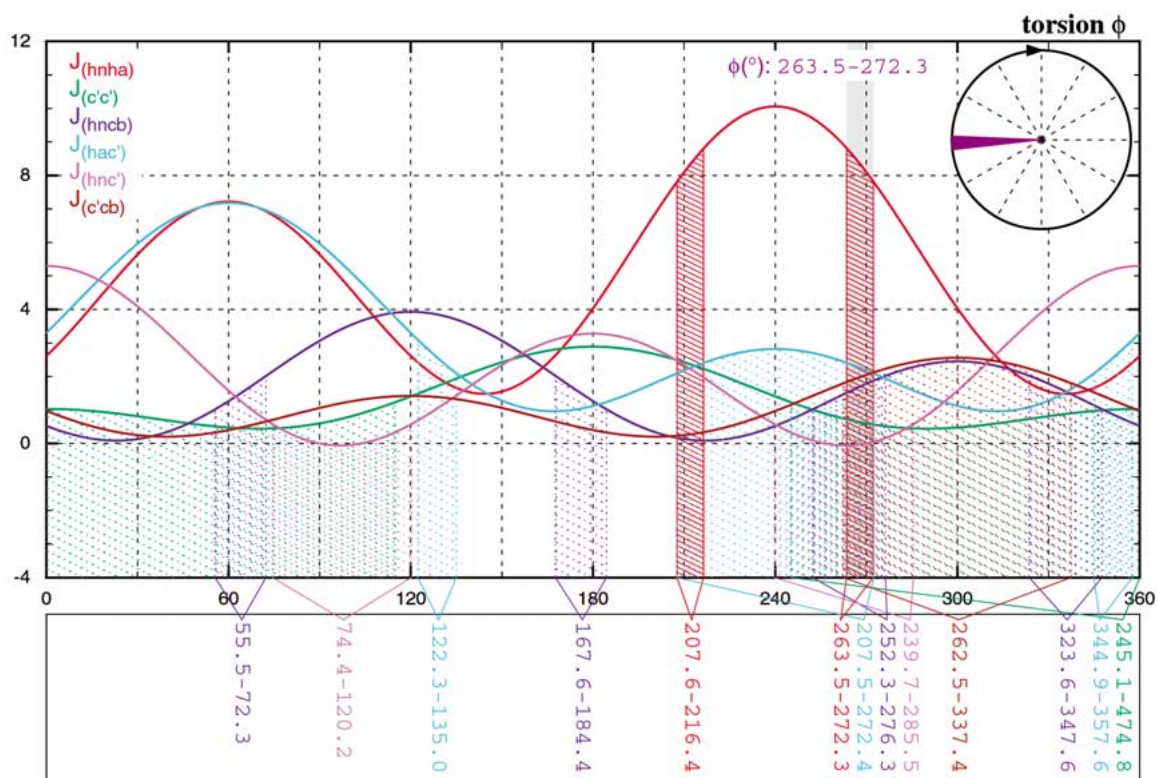


Figure 4. Example of the torsion angle data graphical output – human ubiquitin, residue Q2, the torsion angle ϕ , the data are from Bax and coworkers (Hu and Bax, 1996, 1997; Wang and Bax, 1996). The torsion angle ϕ was determined from six 3J -couplings: $^3J_{(HNH\alpha)}$, $^3J_{(C'C')}$, $^3J_{(HNC\beta)}$, $^3J_{(H\alpha C')}$, $^3J_{(HNC')}$, $^3J_{(C'CB)}$, only the first one was treated as reliable (depicted by solid hatching). To reduce the clutter, an option in program MULDER was used to leave out the horizontal bands representing the input data.

for the backbone torsion angles β , γ , and ϵ , and for the sugar torsion angles. We have found the MULDER graphical output especially useful for excluding the erroneous data in situations where the normal procedure, represented by several repetitive runs of r-MD with different sets of restraints to monitor the restraints violations, turned out to be lengthy and unreliable. As an example, the MULDER graphical output, demonstrating the evaluation of the torsion angles ϵ and χ for the residue G1 of the d(GCGAAGC) hairpin, is shown in Figure 3a and Figure 3b, respectively. The torsion angle ϵ was determined from $^3J_{(H3'P)} \sim 5.0 - 6.3$ Hz, which was obtained quantitatively by the line shape simulation of the H3'-P crosspeaks in the 2D J-resolved 1H - ^{31}P correlation spectra. The values of $^3J_{(C4'P3)}$ and $^3J_{(C2'P3)}$ were estimated from the E.COSY patterns of H4'-C4' and H2'-C2' crosspeaks in the 2D HSQC spectra as 'large' and 'small', respectively. Therefore, $^3J_{(H3'P)}$ was treated as

reliable (depicted by solid hashing in Figure 3a), and $^3J_{(C4'P3)}$ and $^3J_{(C2'P3)}$ as unreliable (dashed hashing). Both of the C-P couplings thus served only as pointers to find the correct solution and were not used to estimate its error interval. The result $\epsilon \sim 196.4 - 203.1^\circ$, consistent with all of the input data, is shown in the circle in the upper right hand corner of Figure 3a. MULDER also shows the error intervals for individual input data on the respective Karplus curves along the horizontal axis.

The glycosidic torsion angle χ for the residue G1 was determined from the interproton distances $d_{(H2',H8)}$, $d_{(H2'',H8)}$, and $d_{(H1',H8)}$. In this example, all of the data were treated as equally reliable. As a result, all of them contributed to the error interval of the final result. As pointed out earlier (Wijmenga et al., 1993; Wijmenga and van Buuren, 1998), these three distances are relatively little sensitive to the torsion angle χ (Figure 3b). Despite this obstacle, MULDER

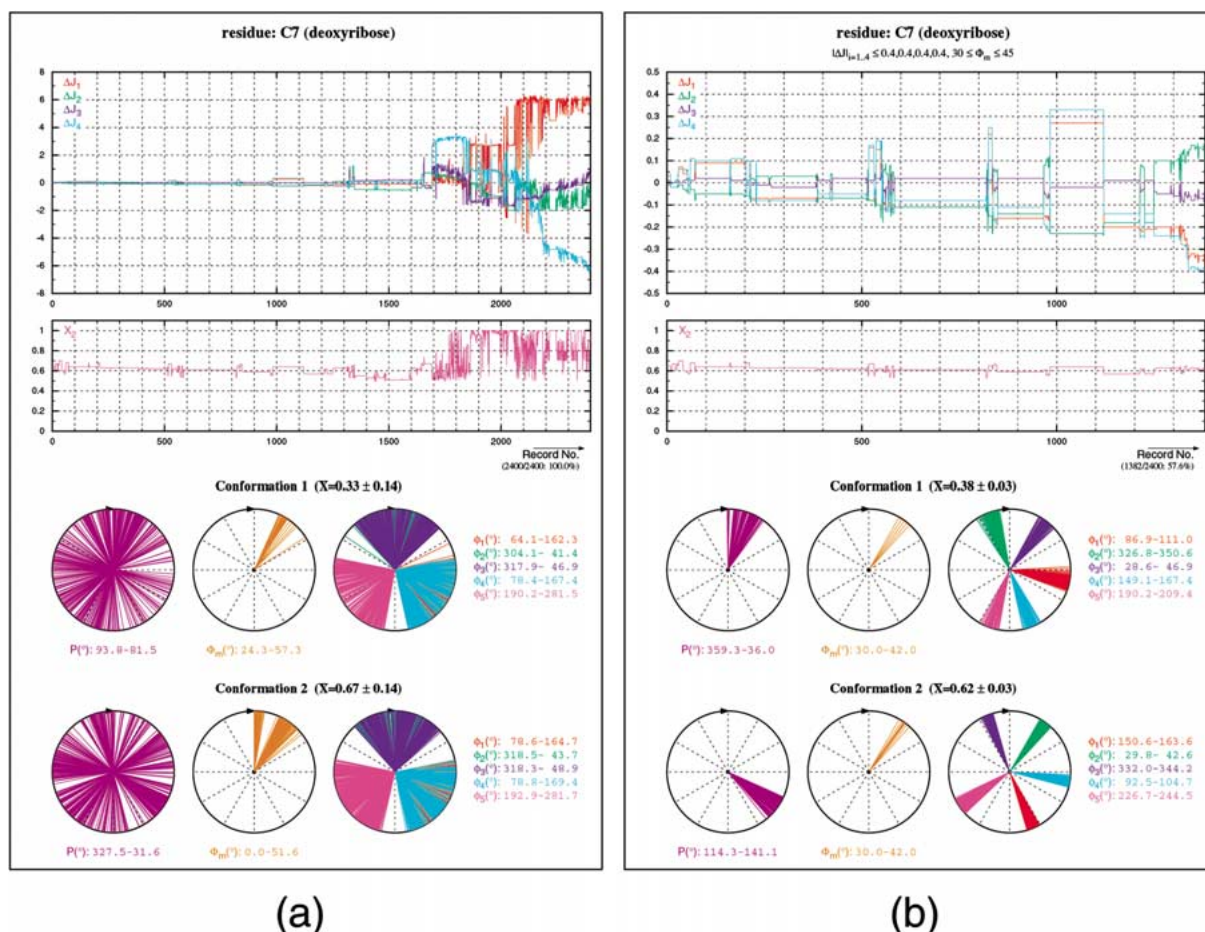


Figure 5. Example of the sugar pucker data graphical output – d(GCGAAGC) hairpin, the residue C7, the data by Padrta et al. (Padrta et al., 2002). The displayed data were calculated from the ^1H - ^1H ^3J -coupling constants by the program PSEUROT 6.3 MANY procedure, then converted into the MULDER acceptable format by auxiliary program SCONV, and finally processed by MULDER itself. Panel a shows the data before any filter was applied, panel b is the result after several filters were used. The titles specify the residue and the filters used. The two graphs on the top of each panel show the dependence of individual ΔJ_i on record number (ΔJ -plot) and the dependence of X_2 on record number (X_2 -plot), respectively. Then there are two rows of radial graphs, one row for each of the two sugar conformers, showing the distributions of pseudorotation phase P , amplitude Φ_m , and of the calculated exocyclic torsion angles $\phi_1 \dots \phi_5$.

found only one common solution and determined the range of torsion angles consistent with the set of input data $\chi \sim 266.5-334.2^\circ$. Similar results were obtained also for the remaining residues of the d(GCGAAGC) hairpin (data not shown). Our experience with MULDER indicates that the distance information contained in the three relevant NOEs ($\text{H}1'-\text{H}6/\text{H}8$, $\text{H}2'-\text{H}6/\text{H}8$, and $\text{H}2''-\text{H}6/\text{H}8$) is sufficient for direct determination of the torsion angle χ provided that all of the distances are used jointly and their values are known with a sufficiently good precision. We found such an exploitation of the interatomic distances especially useful in studies of unlabeled DNA samples where the χ -related J-coupling constants ($^3\text{J}_{(\text{H}1'/\text{C}4/\text{C}2)}$, $^3\text{J}_{(\text{H}1'/\text{C}6/\text{C}8)}$) are not

accessible. The simplicity of the MULDER approach, in comparison with r-MD, might come in particularly handy if one is interested just in orientation of the glycosidic bond and not in the determination of a complete 3D structure.

The distance equation (Equation 3), though differently parameterized, is also applicable for the determination of the torsion angle γ in nucleic acids. If the signals of protons $\text{H}4'$, $\text{H}5'$ and $\text{H}5''$ are well separated, the distances $d_{(\text{H}4',\text{H}5')}$, $d_{(\text{H}4',\text{H}5'')}$, $d_{(\text{H}3',\text{H}5')}$, and $d_{(\text{H}3',\text{H}5'')}$ can be employed. In such a case, the distance curves can be used even for the stereospecific assignments of the $\text{H}5'/\text{H}5''$ resonances. Program MULDER can perform the assignment semi-automatically.

The output for both stereospecific variants is produced (data not shown). Visual comparison of the graphical output for both assignment options allows the user to make the decision which assignment is the correct one.

To test program MULDER on proteins, the data for the torsion angle ϕ in human ubiquitin published by Bax and coworkers (Hu and Bax, 1996, 1997; Wang and Bax, 1996) were used. Using up to six ^3J -couplings per torsion angle, one unambiguous solution was found for 61 out of 73 residues (more than 80%). Figure 4 shows the MULDER graphical output for the residue Q2 of human ubiquitin to demonstrate the contributions and information value of the individual ^3J -couplings. In all cases, $^3\text{J}_{(\text{HNH}\alpha)}$, as the most sensitive coupling, was used to determine the error interval of the result. This was achieved by treating that J-coupling as reliable and the other J-couplings as unreliable.

However, the program application with respect to the torsion angle data is not limited to nucleic acids and proteins. The MULDER implementation is general enough and theoretically can be applied to any system where more than one J-coupling or distance is available for a studied torsion angle. In that case, the only other requirement is that the appropriate parameterizations are known. The MULDER ability to change the parameterizations on the run can also be utilized for studying the dependence of a specific torsion angle on a particular interatomic distance, or for a comparison of different parameterizations of the same scalar interaction, etc. A limitation of the current version of the program is that it does not allow to add new types of torsion angle dependent interactions, i.e., new equations (see section Equations and parameterizations), without the program recompilation. Expectably, the program usefulness will increase with the number of supported equations.

The MULDER limitations in the treatment of sugar pucker data are given by the fact that MULDER relies on the results calculated by program PSEUROT. Since MULDER works only as a postprocessor of PSEUROT procedure MANY, data for any five-membered sugar ring supported by PSEUROT (van Wijk et al., 1995) can be processed by MULDER. Generally, the process of filtering the PSEUROT calculated data by removing the unacceptable conformer pairs is easily controlled and understood when displayed in a graphical form. For that purpose, MULDER graphical output contains the ΔJ -plot, displaying the individual ΔJ_i vs. the record number, the X_2 -plot, displaying X_2 vs. the record number, and the angular values plot,

containing the distributions of pseudorotation phase P, amplitude Φ_m and of the calculated exo-/endo-cyclic torsion angles for both sugar conformers.

Figure 5 shows two examples of MULDER graphical output for the data of C7 residue in the d(GCGAAGC) hairpin, demonstrating the first and the last stages of sugar pucker data processing (panel a and b, respectively). To remove the records with unrealistic sugar pucker it was sufficient to set the Φ_m -filter to $30^\circ < \Phi_m < 45^\circ$. At the same time, only records with the calculated ^3J -coupling constants close to the experimental values, selected by setting the ΔJ -filter to $|\Delta\text{J}|_{i=1\dots4} < 0.4$, were accepted. The result shown in Figure 5b indicates the following: in the two-state approximation, two sugar conformers and their ratio were found that reproduce the experimental values of proton-proton coupling constants with the precision better than 0.4 Hz. As seen from the X_2 -plot in Figure 5b, an equilibrium of two sugar conformations describes the experimental data for residue C7 very well. The sugar conformer 1 is the N-conformer ($\text{P}_1 : 359.3^\circ - 36.0^\circ$) with the fraction $\text{X}_1 = 0.38 \pm 0.03$, the conformer 2 is the S-conformer ($\text{P}_2 : 114.3^\circ - 141.1^\circ$) with fraction $\text{X}_2 = 0.62 \pm 0.03$.

It should be emphasized, however, that the presented example represents a favorable case where the two conformers can be clearly identified. In certain situations, the PSEUROT data are of such nature that no combination of MULDER filters leads to a distinct conformer pair. Occasionally, the conformer pairs can be revealed only at the expense of lowering the demands to fulfill the experimental data, for example, by relaxing the ΔJ -filter. These unfavorable cases may either indicate inconsistency of the experimental J-couplings or alternatively, sugar dynamics which cannot be expressed by the two-state model approximation. Repetitive runs with various sets of filter parameters can help to gain an insight into the behavior of a particular sugar residue. To begin with, the geometrically unrealistic conformers can be removed by setting the Φ_m -filter. The result can be assessed from the Φ_m -plot (middle two circles in Figure 5a). In the next step, the ΔJ -plot can be studied, estimating the number of the records (conformer pairs) with low ΔJ_i differences. A larger proportion of such records with respect to the total number of records indicates better PSEUROT convergence. For example, about 1500 out of 2400 conformer pairs shown in Figure 5a had ΔJ_i values lower than 1 Hz, demonstrating very good convergence of the PSEUROT minimization. Records with ΔJ_i values that are too large are

removed by setting the ΔJ -filter. By gradually tightening the ΔJ -filter in subsequent steps, MULDER is performing a zoom into the ΔJ -plot by removing more conformer pairs that got trapped in various local minima during PSEUROT minimization. Using this approach, more detailed map of conformer pairs close to global minimum of the PSEUROT hypersurface is obtained. During this procedure, a direct examination of the remaining conformer pairs is possible by studying the X_2 -plot and the angular values plot. In this respect, the most informative is probably the P-plot (leftmost two circles in Figure 5a) where the first hints of distinct regions of pseudorotation phases gradually emerge from originally totally crowded circles (compare panels a and b in Figure 5). The impact of tightening the ΔJ -filter is also visible in the X_2 -plot where the removal of 'bad' conformers is represented by a disappearance of the unsteady (oscillating) part of the X_2 curve. Using the values of X_2 and the graphical representation of the pseudorotation phase, two situations can be easily distinguished. One, with a single major sugar conformation, either C2'- or C3'-endo ($X_2 \sim 0.8 - 1.0$), and the other, described more appropriately by the equilibrium between these two ($X_2 \sim 0.3 - 0.7$) (see Figure 5).

The distributions of pseudorotation parameters P and Φ_m in combination with the X_2 -plot represent the main result of the PSEUROT-MULDER tandem. For a convenience, the intervals of pseudorotation parameters are also converted into intervals of the exocyclic torsion angles (using Equation 4), which may be directly employed as restraints in r-MD calculations. Similarly to torsion angle data, the parameterizations of Equation 4 may be added or changed without the program recompilation, enabling the user, for instance, to calculate the endo-cyclic torsions alternatively.

Conclusions

Program MULDER is a software tool for extracting the torsion angle information from three-bond scalar coupling constants or interatomic distances. MULDER reduces the torsion angle ambiguity arising from solving the isolated Karplus or distance equation by using several 3J -couplings and/or distances. In addition, MULDER can be applied to postprocess the data calculated by program PSEUROT to assess the dynamics of five-membered sugar rings. Great emphasis is put on the MD independence and the transparency

of calculations, the latter particularly enhanced by the possibility to create publication quality graphical representations in PDF format. The program is useful for preparing the torsion angle restraints for r-MD calculations, as a quick test to verify the precision and consistency of the experimental data, to evaluate the impact of accuracy and precision of the data on a particular torsion angle, and last but not least, as a structural probe for studying a limited number of important torsions without the need to perform a time-consuming 3D structure determination.

Acknowledgements

We are grateful to L. Židek for deriving the generic distance equation, for encouraging comments and useful suggestions during the manuscript preparation. Stimulating discussion and testing of the program by L. Trantírek and J. Chmelík, Jr. is greatly appreciated. We thank also R. Fiala and M. L. Munzarová for critical reading of the manuscript. This work was supported by grant LN00A016 from the Ministry of Education of the Czech Republic.

References

- Adobe Systems Incorporated (2001) *PDF Reference: Adobe Portable Document Format Version 1.4*, 3rd edn. Addison-Wesley.
- Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc.*, **15**, 8205–8212.
- Brünger, A.T. (1992) *X-PLOR Version 3.1, A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven and London.
- 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. Sect. D-Biol. Crystallogr.*, **54**, 905–921.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.
- de Leeuw, F.A.A.M. and Altona, C. (1983) *J. Comput. Chem.*, **4**, 428–438.
- de Leeuw, F.A.A.M., van Beuzekom, A.A. and Altona, C. (1983) *J. Comput. Chem.*, **4**, 438–448.
- Donders, L.A., Deleeuw, F. and Altona, C. (1989) *Magn. Reson. Chem.*, **27**, 556–563.
- Güntert, P., Billeter, M., Ohlenschläger, O., Brown, L.R. and Wüthrich, K. (1998) *J. Biomol. NMR*, **12**, 543–548.
- Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.
- Haasnoot, C.A.G., Deleeuw, F. and Altona, C. (1980) *Tetrahedron*, **36**, 2783–2792.
- Hu, J.S. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 8170–8171.
- Hu, J.S. and Bax, A. (1997) *J. Am. Chem. Soc.*, **119**, 6360–6368.

- Karplus, M. (1959) *J. Chem. Phys.*, **30**, 11–15.
- Karplus, M. (1963) *J. Am. Chem. Soc.*, **85**, 2870–2871.
- Macaya, R.F., Schultze, P. and Feigon, J. (1992) *J. Am. Chem. Soc.*, **114**, 781–783.
- Marshall, J.L., Walter, S.R., Barfield, M., Marchand, A.P. and Segre, A.L. (1976) *Tetrahedron*, **32**, 537–542.
- Padrta, P., Štefl, R., Králík, L., Žídek, L. and Sklenář, V. (2002) *J. Biomol. NMR*, **24**, 1–14.
- Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham III, T.E., DeBolt, S., Ferguson, D.M., Seibel, G.L. and Kollman, P.A. (1995) *Comp. Phys. Commun.*, **91**, 1–41.
- Polshakov, V.I., Frenkiel, T.A., Birdsall, B., Soteriou, A. and Feeney, J. (1995) *J. Magn. Reson. Ser.*, **B108**, 31–43.
- Štefl, R., Trantírek, L., Vorlíčková, M., Koča, J., Sklenář, V. and Kypr, J. (2001) *J. Mol. Biol.*, **307**, 513–524.
- Tejero, R., Monleon, D., Celda, B., Powers, R. and Montelione, G.T. (1999) *J. Biomol. NMR*, **15**, 251–264.
- Trantírek, L., Štefl, R., Vorlíčková, M., Koča, J., Sklenář, V. and Kypr, J. (2000) *J. Mol. Biol.*, **297**, 907–922.
- van Wijk, J., Haasnoot, K., de Leeuw, F., Huckriede, D. and Altona, C. (1995) *PSEUROT 6.2 – A Program for the Conformational Analysis of Five Membered Rings*, Gorlaeus Laboratories, University of Leiden.
- Wang, A.C. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 2483–2494.
- Wijmenga, S.S. and van Buuren, B.N.M. (1998) *Prog. Nucl. Magn. Reson. Spectrosc.*, **32**, 287–387.
- Wijmenga, S.S., Mooren, M.M.W. and Hilbers, C.W. (1993) In *NMR of Macromolecules*, Vol. 134, Roberts, G.C.K. (Ed.), Oxford University Press, New York, pp. 217–288.