A user-friendly Matlab program and GUI for the pseudorotation analysis of saturated five-membered ring systems based on scalar coupling constants
Pieter MS Hendrickx* and José C Martins*

Address: NMR and Structure Analysis Unit, Ghent University, Krijgslaan 281 S4, B-9000 Ghent, Belgium
Email: Pieter MS Hendrickx* - pieter.hendrickx@ugent.be; José C Martins* - jose.martins@ugent.be
* Corresponding authors

Abstract
Background: The advent of combinatorial chemistry has revived the interest in five-membered heterocyclic rings as scaffolds in pharmaceutical research. They are also the target of modifications in nucleic acid chemistry. Hence, the characterization of their conformational features is of considerable interest. This can be accomplished from the analysis of the $J_{HH}$ scalar coupling constants.

Results: A freely available program including an easy-to-use graphical user interface (GUI) has been developed for the calculation of five-membered ring conformations from scalar coupling constant data. A variety of operational modes and parameterizations can be selected by the user, and the coupling constants and electronegativity parameters can be defined interactively. Furthermore, the possibility of generating high-quality graphical output of the conformational space accessible to the molecule under study facilitates the interpretation of the results. These features are illustrated via the conformational analysis of two 4'-thio-2'-deoxynucleoside analogs. Results are discussed and compared with those obtained using the original PSEUROT program.

Conclusion: A user-friendly Matlab interface has been developed and tested. This should considerably improve the accessibility of this kind of calculations to the chemical community.

Background

Five-membered heterocyclic ring systems constitute an important part of many biologically relevant molecules. They occur in carbohydrates (furanoses), nucleosides and nucleotides, the amino acid proline and their many derivatives. In addition, they often occur as a moiety in complex natural products. Chemical modifications of nucleic acids, often driven by the needs of antisense research, target in part the five-membered cycle or its analogues in order to tailor their conformation towards the desired needs [1,2].

The advent of combinatorial chemistry has also revived the interest in heterocyclic rings and their conformation [3]. As a result, scaffolds containing five-membered heterocycles have received much attention for the rapid generation of potential lead compounds in pharmaceutical research [4-11].

Typically, the chemical and conformational space is explored by introducing a diversity of substituents at varying positions around the cycle. Depending on the position and nature of these substituents, the cycle either
adopts a single conformation or may be in equilibrium between two conformations. These conformations will in turn impact on the conformational space that will be covered by the substituents, making the determination of the cycle’s conformation an issue of considerable interest.

Over the years, NMR has become a well-established technique for this purpose. In particular, $J_{HH}$ scalar coupling constants are well-suited as they are mainly determined by the torsion angle over which they are measured. In the case of ring systems, the vicinal $J_{HH}$ scalar coupling constants are directly correlated to their corresponding exocyclic torsion angles ($\theta_{exo}$). These are related to the corresponding endocyclic torsion angles ($\theta_{endo}$) by a simple equation (1) where $A$ and $B$ are constants determined by the geometry of the atoms linked to the common central bond.

$$\theta_{exo} = A \theta_{endo} + B$$

As the set of all five endocyclic torsion angles in a five-membered ring fully determines its conformation, $J_{HH}$ scalar coupling constants provide a direct measure of the ring’s conformation. The Haasnoot-Altona equation (3) [12] and the Diez-Donders equation (4) [13,14], both based on the well known Karplus equation (2), describe the relation between a $J_{HH}$ Coupling and the corresponding exocyclic torsion angle ($\theta_{exo}$) to a high level of accuracy. In both equations this is mainly achieved by including a set of four parameters $\lambda_i$ ($i = 1, ..., 4$) that account for the influence of electronic effects contributed by the substituents [15,16]. In some studies, the set of experimental $J_{HH}$ scalar coupling constants is further extended by $J_{HH}$ scalar coupling constants [17-19] or interproton distances obtained by NOE NMR experiments. Here however, we assume that only $J_{HH}$ scalar couplings are available for conformational analysis.

$$3J_{HH} = P_1 \cos^2(\theta) + P_2 \cos(\theta) + P_3$$
$$3J_{HH} = P_1 \cos^2(\theta) + P_2 \cos(\theta) + P_3 + \sum_{i=1}^{4} \lambda_i \left\{ P_4 + P_5 \cos^2((-1)^i \lambda_i \theta + \lambda_0) \right\}$$

$$3J_{HH} = \sum_{i=0}^{3} C_i(\lambda_{1-4}) \cos(i\theta) + \sum_{i=1}^{3} S_i(\lambda_{1-4}) \sin(i\theta)$$

Altona and Sundaralingam showed that the description of a five-membered ring conformation can be reduced to a two-parameter pseudorotation model [20,21] that fully describes its conformation. The first parameter, the pucker phase $P$, represents the phase of the conformation and indicates which ring atoms are positioned out of the ring plane. The second parameter, the pucker amplitude $\nu_{\text{max}}$ corresponds to the amplitude of the conformation and describes the extent to which the atoms determined by $P$ are out of the plane. The relationship with the endocyclic torsion angles $\theta_{endo,i}$ is shown in (5).

$$\theta_{endo,i} = \nu_{\text{max}} \cos(P + \frac{4\pi i}{5})$$

This well-known pseudorotation description, originally described for the furanose ring in nucleosides and nucleotides [20,21], was further generalized to any five-membered heterocycle by Diez et al. [22-24] who introduced two additional parameters $\alpha_i$ and $\epsilon_i$ for each endocyclic bond to cope with differences in bond lengths in various types of five-membered rings (Equation 6). As the phase of the conformation $P$ is a periodic variable, polar plots called pseudorotation wheels are mostly used to depict ring conformations.

$$\theta_{endo,i} = \alpha_i \nu_{\text{max}} \cos(P + \epsilon_i + \frac{4\pi i}{5})$$

Using the above equations, $J_{HH}$ couplings can be used to derive the pseudorotation parameters of the five-membered cycle. As mentioned previously, the cycle may be in equilibrium between two conformations. Thus, generally two sets of pseudorotation parameters ($P$ and $\nu_{\text{max}}$) and the relative population ($\%_1$, i.e. the percentage of the first conformation present with $\%_2 = 1 - \%_1$) need to be fitted to the experimental NMR data. In order to avoid an under-determined model, experimental data measured at different temperatures is generally used. In such cases, the model assumes that only the relative population of the two conformations varies when changing the temperature. Thus $n + 4$ ($n$ being the number of temperatures used) variables will be optimized to fit the experimental data in such cases. To the best of our knowledge, the program PSEUROT [25], originally developed by Altona et al., is still the only generally available program to perform this type of analysis. Written in FORTRAN, its interface as well as its output is purely text-based. In order to facilitate the analysis of the PSEUROT results, a post-processing feature has been included in the independently developed MULDER package [26] to generate a graphical output of the PSEUROT results. In this communication, we propose an integrated, user-friendly Matlab program, including a self-explanatory graphical user interface (GUI), to facilitate the set-up, execution and subsequent analysis of pseudorotation calculations for five-membered ring systems decorated with a variety of substituents. The use of Matlab as high-level programming language enables to create, within a limited time frame, high-quality plots that provide a graphical impression of the conformational space accessible. Furthermore, due to the open-source GNU
GPL license, users have the opportunity to adapt the program to their specific needs.

Implementation
The program consists of a computational core that is accessed and controlled through a GUI (Figure 1), both written in Matlab. Its goal is to search pseudorotation parameters \( \{P, v_{\text{max}}\} \) for at most two conformations as well as their relative population \( (\phi_1, n) \) at \( n \) temperatures that fit a series of experimental NMR scalar coupling constants. The initial pseudorotation parameters are set by the user at the start of the computational procedure. The user is also given the choice to define a subset of pseudorotation parameters that have to be optimized. Using equations 1, 4 and 6, the scalar coupling constants relating to conformation defined by the initial pseudorotation values are calculated and a root-mean-square-deviation (RMSD) is determined with respect to the experimental coupling data. Next, the pseudorotation parameters are adapted so as to minimize this RMSD. The Matlab fmincon function from the Optimization Toolbox, which is based on a Sequential Quadratic Programming (SQP) algorithm [27], was used for this purpose. To restrict the optimization to physically sensible solutions, values of partition coefficients were restrained to the \([0, 1]\) interval. Furthermore, puckering amplitudes \( (v_{\text{max}}) \) were restricted to the physically relevant \([0, \pi/3]\) interval.

A choice is provided between two operational modes. In the first mode, an optimization of the chosen parameters is performed as is, yielding the pseudorotation parameters that best fit the experimental data. The output generated in this operation mode is purely text-based and contains the optimized variables, their corresponding endocyclic torsion angles and a tabular comparison between experimental and fitted scalar coupling data (See accompanying manual). In the second mode of the program, the complete pseudorotation space of the cycle of interest is explored via 3600 combinations of pseudorotation parameters \( \{P, v_{\text{max}}\} \) as described hereafter. The pseudorotation space of the cycle is covered by sampling 120 values of \( P \) in the \([0, 2\pi]\) interval and 30 values of \( v_{\text{max}} \) in the \([0, \pi/3]\) interval. For each \( \{P, v_{\text{max}}\} \) thus sampled, the minimized RMSD between the experimental and calculated scalar couplings is determined. As for the output, a pseudorotation wheel is plotted and the RMSD is represented through 20 contour lines in the \([\Delta_{\text{min}}, 2\Delta_{\text{min}} + 0.1]\) interval where \( \Delta_{\text{min}} \) is the lowest RMSD obtained over the 3600 fittings.

Whether the parameters of a second conformation and the relative populations at each temperature are optimized during each of the 3600 runs is again determined by the user’s choice. This results in two possible outputs for this operating mode. For the case where the user chooses not to optimize any parameters of the second conformation, one can easily assess from the pseudorotation wheel if the experimental data can be fitted by a single five-membered ring conformation. For the case where the second conformation’s parameters (and the relative population of both conformations) are also to be optimized, the pseudorotation wheel represents the conformational space accessible to the cycle when two five-membered ring conformations are fit to the experimental data. In this latter case, one additional pseudorotation wheel is created for each temperature depicting the relative population of the two fitted conformations for each set of \( \{P, v_{\text{max}}\} \) using contour levels.

This second mode of calculation has some interesting advantages over the first mode. First, a better appreciation of the conformational space of the cycle that is in agreement with the experimental data is obtained. Furthermore, the extent of the contour levels in the pseudorotation plot allow to establish whether the model is under-determined or not. Even when under-determined, such calculation can already indicate those conformations that can be excluded from further investigation.

Results
Program features
Only the main features of the program, fully described in the manual available together with the program, are discussed. The GUI is easy to use and handles the input as well as the various operational modes and data output (Figure 1). Two modes of calculation for fitting five-membered ring conformations to \( ^3J_{HH} \) scalar coupling data.
obtained at up to five temperatures can be selected. Both the coupling data and additional parameters required by equations 1 and 6 are defined interactively, using a simplified representation of the cycle (Figure 1, upper panel). Both the atom types in the cycle and the substituents can be defined interactively. The GUI includes an 'electronegativity editor' (Figure 1, bottom panel) which provides a convenient graphical aid to set up the group-electronegativities ($\lambda_i$) for many common substituents as required by equation 4.

Also, the program is tolerant for sparse input data. When only a sum of two or more $J_{HH}$ scalar couplings can be measured, e.g. due to overlap, this sum can be included in the model instead of the individual couplings. Contrary to the PSEUROT program, a coupling does not necessarily have to be measured at each temperature to be included in the calculation.

Four different parameterizations of equation 4 are implemented in the program. These include the 20-parameter and 12-parameter parameterization by Donders [14] and two slightly different 9-parameter parameterizations [28,29] one of which is used in PSEUROT 6.2. The use of Diez-parameters into the calculations is fully implemented by using equation 6 for the calculation of exocyclic torsion angles out of the puckering parameters. The current version does not include the use of the Barfield-correction [30,31] used in some studies.

Depending on the calculation mode selected, textual or graphical output is provided. As the program is written in Matlab, the latter plots can be easily post-processed to comply with the user’s needs. In addition, several parameters involving the construction of the plots can be set manually. These parameters include the resolution of the search grid used for the pseudorotation scan (i.e. the increment in $P$ and $\nu_{\text{max}}$) and the minimal and maximal contour levels used in the plots. Furthermore, interpolation of these contour plots using cubic spline functions can be selected for obtaining smoothed graphs.

Standard data sets for $\beta$-D-ribose and $\beta$-D-deoxyribose as well as input data of 1 and 2 are available with the distribution of the software.

Comparison to PSEUROT
To test the performance of the computational core and the input/output handling using the GUI and compare the results with respect to PSEUROT (version 6.2), a set of two 4’-thio-2’-deoxynucleoside analogs [32] was used. For both molecules, depicted in Figure 2, literature [32] provides a full set of five coupling constants measured at five different temperatures as well as Diez-parameters for each endocyclic bond. For both compounds, two conformations were fit to the experimental data using our Matlab GUI and PSEUROT 6.2. The most recent parameterization of equation 4 was used. Details of these optimal fittings, as well as a third fit, taken from literature [32] and referred to as LHK, using the Haasnoot equation 3 [12] are presented in Table 1. This table includes the optimized pseudo-

![Chemical structure of compounds 1 and 2](image-url)
dorotation parameters for the North and South conformations (\(P_N, \varphi_N, P_S, \varphi_S\)) in degrees and the population of the North-conformer at each temperature for which couplings were provided. Finally, the maximal deviation and RMSD between the fitted couplings and the experimental ones is indicated.

Quasi identical results are obtained when comparing the Matlab GUI (ML) to PSEUROT (PS). Differences between the two programs are most likely introduced by the difference in optimization method and convergence criteria. Looking at the residual deviations with respect to the experimental data, the Matlab GUI finds a slightly better fit. When comparing to the literature data (LHK), one can appreciate the influence of different Karplus equations on the fitting. Differences of 10 and 3 degrees are found for \(P_N\) and \(\varphi_N\) respectively. Also slight differences in populations are observed.

In order to get a better appreciation of the precision of these optimal values, the full pseudorotation wheel of 1 and 2 has been scanned by the Matlab program resulting in Figures 3 and 4. In Figure 3, contour plots indicate the root-mean-square-deviation between the fitted and the experimental scalar coupling couplings for 1 (left) and 2 (right). Looking at the extent of the contour lines, it is clear that the South-conformation is better defined than the North-conformation. Taking into account errors originating from measuring scalar couplings, deficiencies of the model and the Karplus equation, the optimal fit does not necessarily correspond to the real situation. Therefore, interpreting the regions that have a RMSD lower than a certain threshold (e.g. 0.2 Hz), as representative of the cycle's conformation seems more reasonable than taking the optimal fit for granted, especially when comparing different compounds. In addition to Figure 3, Figure 4 can

|        | ML | PS | LHK | ML | PS | LHK |
|--------|----|----|-----|----|----|-----|
| \(P_N\)| 3.9| 3.6| 13  | 1.8| 0.1| 9   |
| \(\varphi_N\)| 41.7| 42.6| 45 | 42.3| 44.0| 45  |
| \(P_S\)| 184.3| 182.5| 177| 184.2| 182.5| 177 |
| \(\varphi_S\)| 41.9| 41.0| 43 | 42.4| 40.9| 44  |
| \%N_{295K} | 29 | 29 | 25 | 36 | 36 | 33  |
| \%N_{100K} | 31 | 31 | -  | 37 | 37 | -   |
| \%N_{113K} | 32 | 32 | -  | 38 | 38 | -   |
| \%N_{333K} | 35 | 35 | -  | 39 | 39 | -   |
| \%N_{333K} | 37 | 37 | 33 | 40 | 40 | 37  |
| \(\Delta J_{\text{max}}\)| 0.14| 0.18| 0.21| 0.18| 0.21| 0.29|
| \(\Delta J_{\text{rms}}\)| 0.055| 0.060| 0.14| 0.075| 0.083| 0.12|

Numbers are indicated in degrees, percentages and Hz.

Table 1: Optimal puckering coordinates, partition coefficients and residual errors found for compounds 1 and 2 using this Matlab GUI (ML), PSEUROT (PS) and literature data (LHK) [32].

Figure 3
Pseudorotation wheel for compounds 1 and 2. Contour lines indicate the total RMSD with the experimental data (Hz). Figures are directly obtained from the GUI. No post-processing has been performed.
be used to interpret the allowed population range for each conformation. These graphs represent the population of both conformations of 1 at 285K (left) and 353K (right). The thick black contour is the highest contour level of Figure 3. Within this black contour, a variety of populations can be found, ranging from 27 to 40 percent of the North-conformation at 285K and from 35 to 49 percent at 353K. Again, this gives a more realistic view of the conformational behavior of compound 1.

**Conclusion**

A matlab program with easy to use graphical user interface for the calculation of five-membered ring conformations has been presented. This program has been made freely available online under a GNU GPL license. The performance of the program was tested on a set of two 4'-thio-2'-deoxynucleoside analogs. It has been shown that identical results can be obtained as in PSEUROT. Furthermore, high-quality graphical output can be generated, facilitating the interpretation of the calculations.

**Availability and requirements**

Project name: Matlab Pseudorotation GUI

Project home page: [http://sourceforge.net/projects/matlabpseudorot](http://sourceforge.net/projects/matlabpseudorot)

Operating system(s): Platform independent (tested in WinXP)

Programming language: Matlab (tested in release R2007a)

Other requirements: Matlab Optimization Toolbox

License: GNU GPL

**Authors’ contributions**

PMSH is the main author of this manuscript. His contributions to this article include the programming of the Matlab program and GUI, performing extensive tests to ensure its stability and functionality. Furthermore, the comparison of the obtained results between the GUI, PSEUROT and literature data were contributions from this author. JCM has provided intellectual, technical and financial support for the development of the Matlab GUI. Also, he contributed in testing phase of the GUI.

**Acknowledgements**

The IWT Flanders is thanked for a PhD SBO-grant to P.M.S.H. J.C.M. thanks the FWO for financial support and several NMR equipment grants (G.0036.00N, G.0365.03). P.M.S.H. acknowledges Duane Hanselman for the ‘Fully Customizable Polar Plot’ (mmpolar) Matlab script used in the GUI. This article is dedicated to Prof. P. De Clercq at the occasion of his 60th birthday.

**References**

1. Cobb AJA: Recent Highlights In Modified Oligonucleotide Chemistry. *Org Biomol Chem* 2007, 5:3260-3275.
2. Leumann CJ: DNA Analogs: From Supramolecular Principles To Biological Properties. *Bioorg Med Chem* 2002, 10:841-854.
3. Geyssen HM, Schoenen F, Wagner D, Wagner R: Combinatorial Compound Libraries For Drug Discovery: An Ongoing Challenge. Nature Drug Discov 2003, 2(3):222-230.

4. Saotome C, Wong CH, Kanie O: Combinatorial Library Of Five-Membered Iminocyclitol And The Inhibitory Activities Against Glyco-Enzymes. Chem Biol 2001, 8(11):1061-1070.

5. Steger M, Hubschwerlen C, Schmid G: Solid-And Solution-Phase Synthesis Of Highly-Substituted-Pyrrolidine Libraries. Biorg Med Chem Lett 2001, 11(18):2537-2540.

6. Boldi AM, Dener JM, Hopkins TP: Solid-Phase Library Synthesis Of Alkoxylprolines. J Comb Chem 2001, 3(4):367-373.

7. Jackson RW, Tabone JC, Howbert J: Identification Of TNF-α Inhibitors From A Split-Pool Library Based On A Tyrosine-Proline Peptidomimetic Scaffold. Biorg Med Chem Lett 2003, 13(2):205-208.

8. Ohtake N, Imamura H, Jona H, Kiyonaga H, Nakagawa S: Novel Dithiocarbamate Carbapenems With Anti-MRSA Activity. Biorg Med Chem 1998, 6(7):1089-1101.

9. Bray AM, Chieffi DS, Valerio RM, Maeji NJ: Rapid Optimization Of Organic Reactions On Solid Phase Using The Multipin Approach: Synthesis Of 4-Aminoproline Analogues By Reductive Amination. Tetrahedron Lett 1995, 36:5081-5084.

10. Curran TP, Chandler NM, Kennedy RJ, Keaney MT: N-α-Benzoyl-4-α-mono-L-proline: 2′-turn Mimetic. Tetrahedron Lett 1996, 37:1933-1936.

11. Klein SI, Dener JM, Molino BF, Gardner CJ, D’Alisa R, Dunwiddie CT, Kasiewski C, Leadley RJ: O-Benzyl Hydroxyproline As A Biosostere For Phe-Pro: Novel Dipeptide Thrombin Inhibitors. Biorg Med Chem Lett 1996, 6:2225-2230.

12. Haasnot CAG, Deleeuw FAAM, Deleeuw HPM, Altona C: The Relationship Between Proton-Proton NMR Coupling Constants And Substituent Electronegativities I. Formulation Of An Equation Including Interactions Between Substituents And Its Application To Coupling-Constant Data Calculated By The Extended Huckel Method. Magn Reson Chem 1989, 27(6):556-563.

13. Altona C, Ippel JH, Hoekzema AJAWs, Erkelens C, Groesebeek M, Donders LA: The Relationship Between Proton-Proton NMR Coupling Constants And Substituent Electronegativities. I. Empirical Substituent Constants Deduced From Ethanes And Propanes. Magn Reson Chem 1989, 27(6):564-576.

14. Altona C, Francke R, Denhaan R, Ippel JH, Dehaisman GJ, Hoekzema AJAW, van Wijk J: Empirical Group Electronegativities For Vicinal NMR-Proton Couplings Along A C-C Bond: Solvent Effects And Reparameterization Of The Haasnoot Equation. Magn Reson Chem 1994, 32(11):670-678.

15. San Fabian J, Guilleme J: Vicinal Fluorine-Proton Coupling Constants. Ab Initio Calculations Of Angular Dependence And Substituent Effects. Chem Phys 1996, 206(3):325-337.

16. San Fabian J, Guilleme J, Diez E: Vicinal Fluorine-Proton Coupling Constants II. Individual Substituent Effects. J Magn Reson 1998, 132(2):253-257.

17. Thibaudeau C, Plavec J, Chattopadhyaya J: A New Generalized Karplus-Type Equation Relating Vicinal Proton-Proline Coupling Constants To H-C-F Torsion Angles. J Org Chem 1999, 64:4967-4984.

18. Altona C, Sundaralingan M: Conformational Analysis Of The Sugar Ring In Nucleosides And Nucleotides. A New Description Using The Concept Of Pseudorotation. J Am Chem Soc 1972, 94:8205.

19. Altona C, Sundaralingan M: Conformational Analysis Of The Sugar Ring In Nucleosides And Nucleotides. An Improved Method For The Interpretation Of Proton-proton Magnetic Resonance Coupling Constants. J Am Chem Soc 1973, 95:2233.