Special Collection

HERMES – A Software Tool for the Prediction and Analysis of Magnetic-Field-Induced Residual Dipolar Couplings in Nucleic Acids

Ilektra-Chara Giassa,[a] Andrea Vavrinská,[b] Jiří Zelinka,[c] Jakub Šebera,[d] Vladimír Sychrovský,[d] Rolf Boelens,[b] Radovan Fiala,[a] and Lukáš Trantírek*[a]

Field-Induced Residual Dipolar Couplings (fiRDC) are a valuable source of long-range information on structure of nucleic acids (NA) in solution. A web application (HERMES) was developed for structure-based prediction and analysis of the (fiRDCs) in NA. fiRDC prediction is based on input 3D model structure(s) of NA and a built-in library of nucleobase-specific magnetic susceptibility tensors and reference geometries. HERMES allows three basic applications: (i) the prediction of fiRDCs for a given structural model of NAs, (ii) the validation of experimental or modeled NA structures using experimentally derived fiRDCs, and (iii) assessment of the oligomeric state of the NA fragment and/or the identification of a molecular NA model that is consistent with experimentally derived fiRDC data. Additionally, the program's built-in routine for rigid body modeling allows the evaluation of relative orientation of domains within NA that is in agreement with experimental fiRDCs.

Introduction

Direct spin-spin interactions, often referred to as dipolar couplings (DCs), are a valuable source of information about the structure of biomolecules: DCs convey information on the relative orientations of vectors interconnecting interacting nuclear spins with respect to the direction of the external magnetic field.[1,2] However, for diamagnetic molecules in an isotropic solution, DCs are generally averaged to zero due to random molecular tumbling.[3] Thus, to obtain DCs for diamagnetic molecules in solution, the anisotropy of molecular tumbling needs to be externally imposed. This is typically achieved either by supplementation of buffers with so-called alignment media such as bicelles, liquid crystals, nonionic polymers, rod-shaped viruses (including Pf1 bacteriophages), or DNA nanotubes or by (non)covalent attachment of the studied molecules with paramagnetic moieties.[4–14] Under these conditions, even diamagnetic biomolecules adopt a small degree of alignment in a magnetic field, resulting in residual dipolar couplings (RDCs) between nuclear spins. Currently, due to the availability of multiple user-friendly tools for RDC interpretation, RDCs induced by alignment media or paramagnetic tags are routinely used for structure refinement and structure validation[15–26] and/or for deriving information on the relative orientation of rigid domains in both proteins[27–29] and nucleic acids (NAs).[30,31]

However, the use of alignment media for NAs, particularly for DNA, brings about problems connected to the inherent sensitivity of the structure and dynamics of DNA to the physicochemical parameters of the environment.[32,33] The addition of co-solutes (alignment media) might affect the structure and/or dynamics of the studied nucleic acid fragments. As demonstrated by Bryce et al. and Buuren et al.[34,35] this problem can be avoided by the acquisition of so-called field-induced RDCs (fiRDCs): NAs possess sufficiently large anisotropy of magnetic susceptibility (AMS), whose interaction with an external magnetic field results in spontaneous alignment of NA fragments in a supplement-free solution and gives rise to measurable (fi)RDCs (*Theoretical background* – [Equation (1)]). Although fiRDCs have proven to be useful for the characterization of NA oligomeric states[36] and in NA structure determination,[34,35,37] their use for NA characterization has remained scarce. The vast majority of currently available structural information on NAs, including the use of RDCs, has been obtained in buffers supplemented with alignment media. This situation has been mainly due to two reasons. First, for NA systems of moderate size, the fiRDCs at most commonly available magnetic field strengths (11.5–16.5 T) are generally rather small compared to RDCs obtained in alignment media and weighted down by significant experimental errors that compromise their structural interpretation.[34,35] This stems from

[a] *I.-C. Giassa, Prof. R. Fiala, Prof. L. Trantírek*
Central European Institute of Technology
Masaryk University
Brno < postCode 625 00, (Czech Republic)
E-mail: lukas.trantirek@ceitec.muni.cz
[b] *Dr. A. Vavrinská, Prof. R. Boelens*
Bijvoet Centre for Biomolecular Research
Utrecht University
Utrecht 3584 CH, (The Netherlands)
[c] *Dr. J. Zelinka*
Department of Mathematics and Statistics
Faculty of Science
Masaryk University
Brno 611 37, (Czech Republic)
[d] *Dr. J. Šebera, Dr. V. Sychrovský*
Institute of Organic Chemistry and Biochemistry
Czech Academy of Sciences
Prague 166 10, (Czech Republic)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cplu.202000505

This article is part of a Special Collection on "Chemistry in the Czech Republic".

the fact that the size of the fiRDCs scales with the square of the strength of the magnetic field and depends linearly on the number of NA bases (moieties providing a major contribution to molecular AMS) in the NA fragment under investigation.[34,35,38,39] Second, the complex and time-consuming procedure for fiRDC interpretation has not been automated thus far. However, the problem connected with the generally small size of fiRDCs is expected to be (at least partially) resolved with the upcoming generation of NMR spectrometers operating at ultrahigh magnetic fields (up to 28.2 T).[40] With the use of spectrometers operating at ultrahigh magnetic fields, the measured fiRDCs are expected to reach the sizes comparable to those generated by the use of alignment media.[38]

To address the problem connected with the demanding and time-consuming interpretation of fiRDC data, we developed a software tool (HERMES) that, at its core, enables the prediction of field-induced RDCs based on the built-in data in a reference database of base-type-specific anisotropies of magnetic susceptibility and user-defined structural models. The software allows users, including non-experts, to interactively assess information encoded in fiRDCs from nucleic acids. It can be used for planning NMR experiments, the cross-validation and evaluation of NA structural models, the assessment of the oligomeric state of an oligonucleotide, and investigations of the relative orientation of rigid domains in NAs.

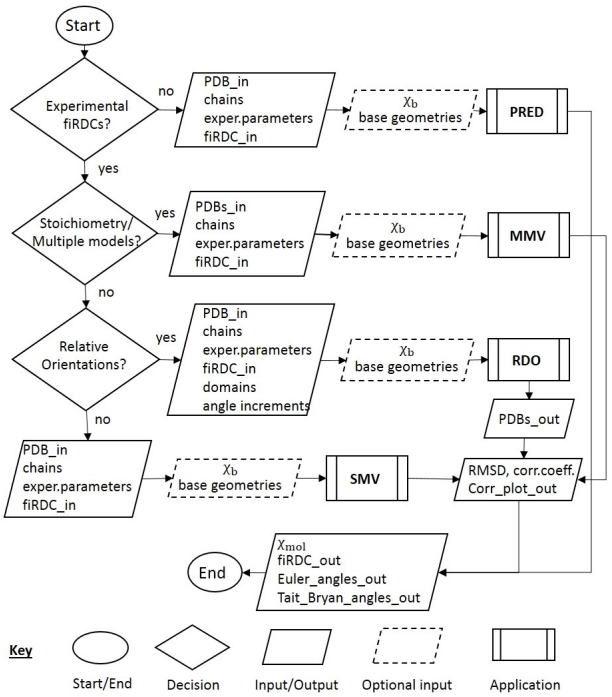

Figure 1. Flowchart of HERMES. PRED, SMV, MMV and RDO stand for prediction, structural model validation, multiple model evaluation, and relative domain evaluation modules, respectively. See text for details.

Results and Discussion

The core routine of HERMES, which is common to all applications the program permits, calculates χ_{mol} from χ_b values (either defined by the user or taken from the built-in database – cf. Documentation section of HERMES (http://hermes.ceitec.muni.cz/Documentation.html and Table S1 in the Supporting Information) and from the user-supplied 3D model/structure of NAs presented in standard PDB format. Subsequently, the theoretical fiRDCs are calculated using [*Eq.* (*2*)] and [*Eq.* (*3*)] *(cf. Theoretical background)*, based on the user-provided information on i] the experimental conditions, i.e., temperature (in Kelvin) and magnetic field strengths (B_0^{low} and B_0^{high} in Tesla) used (or to be used) for fiRDC acquisition, and ii] information specified by the user in the "fiRDC_input" file. This file lists the internuclear A–B vectors and their associated order parameters, S. The core routine returns the list of the predicted fiRDCs for individual internuclear A–B vectors and χ_{mol} expressed in the molecular frame of the user-supplied structure. The overall outline of HERMES with all its applications is presented in Figure 1. The standardized HERMES input interface and the format of the "fiRDC_input" are displayed in Figure 2 and Table 1, respectively.

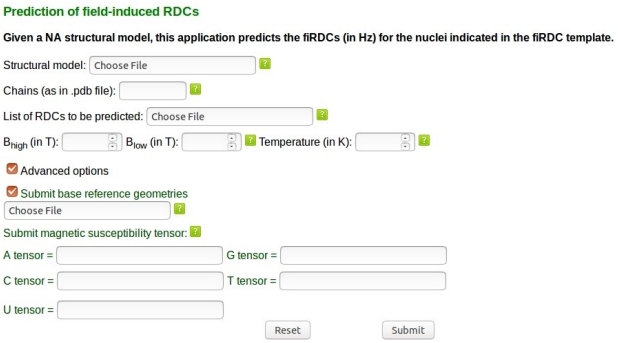

Figure 2. Standardized input interface for HERMES applications. The user is prompted to select and upload structural model(s) of NA molecule(s) in standard PDB format, list chains in the molecule as defined in the PDB file, and upload the "fiRDC_in" file (see below) and experimental parameters (B_0^{high}, B_0^{low} and temperature). If required, the user can provide their own reference base geometries and/or magnetic susceptibility tensors for each type of nucleobase. Otherwise, default geometries and χ_b values from the built-in database[38] are employed.

The HERMES interface allows the user to choose from four basic applications: (i) the Prediction (PRED) of fiRDCs for a given structural model of NA; (ii) Structural Model Validation (SMV) on the basis of the correlation of experimental fiRDCs with those predicted based on the user-supplied structural model; (iii) Multiple Model Evaluation (MMV), the application of which represents an extension of (ii) and allows the user to assess the oligomeric state of NAs and/or to find a molecular model that provides the best match between experimental and predicted fiRDC data; and (iv) the evaluation of the Relative Domain Orientation (RDO) between two user-defined domains in the molecule based on the best fit between the experimental and predicted values for fiRDCs. Next, to the common input (3D structure, the list of internuclear A–B vectors and corresponding S values, B_0^{high}, B_0^{low} and temperature values), the later application requires a user to provide the definition of two domains (rigid segments, whose relative orientation is to be predicted) and the increment for the three angles psi (ψ), phi (ϕ), theta (θ),

Table 1. A description of the "fiRDC_in" file.[a]

Column	Content
1	chain
4–6	nucleotide serial number
9–12	interacting nucleus A
14–17	interacting nucleus B
20–26	experimental fiRDCs
30–35	experimental error
38–42	order parameter S
49–56	predicted fiRDCs

[a] Please note that the file is plain text having a fixed format. The first four columns serve to unambiguously define the pair of interacting nuclei A and B based on information provided in the PDB file. Each of the A–B vectors is linked with the value of the experimental fiRDC (in the 5th column), the associated fiRDC error (in the 6th column), and the order parameter (S) value. The value of S is set to 1 by default (unless specified otherwise). Note: For the structure-based prediction of fiRDCs, the information in the 5th and 6th columns is ignored.

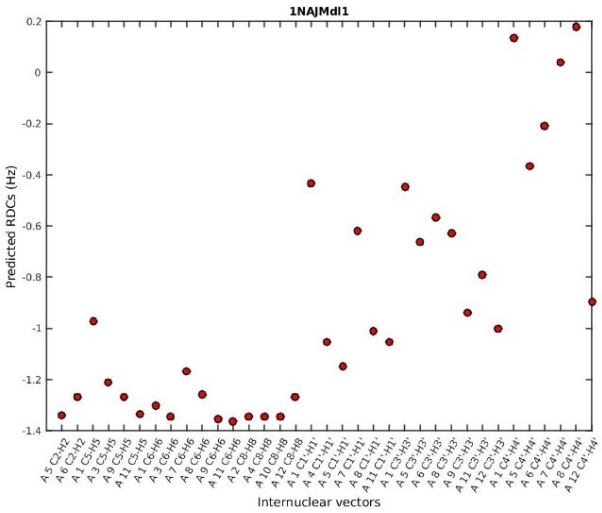

Figure 3. Illustrative example of "plot_out" file displaying predicted fiRDC values (y-axis) for the selected bond vectors (x-axis) in the Dickerson dodecamer (PDB ID: *1naj*).

which describe the rotation around the z, y′ and x′′ axes of the molecular frame, respectively. HERMES uses this information to generate 3D structural models differing by virtue of relative domain orientation. For each of the generated models, HERMES predicts fiRDCs, which are subsequently used to identify the model providing the best fit between the experimental and predicted fiRDCs. Illustrations of individual applications are given below.

Structure-based PREDiction of fiRDCs (PRED module)

As an illustrative example, the prediction of fiRDC values for the selected set of internuclear A–B vectors in the Dickerson-Drew dodecamer (PDB ID: *1naj*)[41] was performed. A user is prompted to upload the corresponding PDB file (Structural model) and the "fiRDC_input" file (cf. Table 1), listing the internuclear A–B vectors (and their corresponding S values) to the web server, and then enters the experimental conditions. In the present example, values for B_0^{high} and B_0^{low} were set to 18.8 T and 11.75 T, respectively. The temperature was set to 308 K. The calculations were performed under the assumption of a rigid NA structure (S = 1). After clicking on the Submit button, the program returns results in a graphical form (plot the predicted fiRDC values (y-axis) for fiRDCs for each bond indicated in the "fiRDC_input" file (x-axis)) (Figure 3). The user can also download the results in the form of a .zip file: The file contains the list of the predicted fiRDC values ("fiRDCs_out" file), the structural model with the molecular alignment tensor in PDB format ("PDB_out" file), the sets of Euler and Tait-Bryan angles ("Euler_angles_out" and "Tait_Bryan_angles_out" files) and the plot of the predicted fiRDC values ("plot_out" file).

Nucleic Acid Structure/Model Validation (SMV module) with experimental fiRDC data

The SMV application requires essentially identical inputs as those listed above, with a single exception, which is a listing of the experimental fiRDC values (and their errors) for each of the specified internuclear vectors (cf. Table 1).

Following the Submit request, the program returns a simple regression plot and a Deming regression between the experimental and predicted values of fiRDC. Simple linear regression represents the best fit based on the minimization of the sum of squared differences between experimental and predicted data, whereas Deming regression, with the assumption of equal error variances for the two datasets, minimizes the sum of squared perpendicular distances from the data points to the regression line. Along with the plot, the module returns information on the intercept and the slope for each regression line (SlopeL/InterceptL and SlopeD/InterceptD for the linear and Deming regression, respectively), the Pearson correlation coefficient (CorrCoeff), the root-mean-square-deviation (RMSD) in Hz, and the values of R- and Q-factor (Figure 4, Table 2). For definition of CorrCoeff, RMSD, and R- and Q-factors see Experimental

Table 2. A shortened version of the output table from the SMV module.[a]

Model	RMSD	Corr Coeff	SlopeL	InterceptL	SlopeD	InterceptD
1naj	0.52	0.68	0.44	−0.49	0.54	−0.40

[a] SlopeL/InterceptL, SlopeD/InterceptD, CorrCoeff, and RMSD stand for slope/intercept from the linear regression, slope/intercept from the Deming regression, the Pearson correlation coefficient, and the Root-Mean Square Deviation (in Hz), respectively. For complete output table listing all statistical descriptors see Supplementary Information - Table S2. For detailed discussion of outliers from the fit see ref. [34].

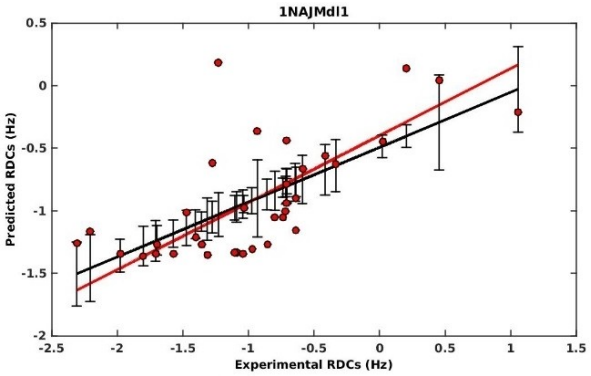

Figure 4. Illustrative example of a graphical output ("corr_plot_out" file) from the SMV module. Simple linear regression plot (in black) and Deming regression (in red) between the experimental[34] and predicted values of fiRDCs for the Dickerson-Drew dodecamer (DDD): d(CGCGAATTCGCG)$_2$ crystal structure (*1naj*). In the present example, values for B_o^{high} and B_o^{low} were set to 18.8 T and 11.75 T, respectively. The temperature was set to 308 K. The calculations were performed under the assumption of a rigid NA structure (S = 1).

Section. The user can download the results in the form of a .zip file, which next to the list of predicted fiRDCs (in the 9th column of the "fiRDC_out" file); a structural model with the molecular alignment tensor transformed into molecular frame in a PDB format ("PDB_out" file); a list of Euler and Tait-Bryan angles used for the transformation ("Euler_angles_out" and "Tait_Bryan_angles_out" files), and the graphical output file ("corr_plot_out") displaying both correlation between the predicted fiRDC and experimental fiRDC values and individual descriptors describing the quality of the fit (Figure 4, Table 2, Table S2).

Multiple Model Validation (MMV module) application

The *Multiple Model Validation* (MMV) application corresponds to an extension of the SMV tool. Analogous to SMV, the MMV application prompts users to enter experimental fiRDCs (in the "fiRDC_input" file) and all experimental details (B_0^{high}, B_0^{low} and temperature). However, in contrast to SMV, the MMV module allows the user to perform an analysis of the experimental fiRDC data using multiple structural models at the same time. There are two main applications of the MMV module: i] the identification of the structural model that best matches the experimental fiRDCs[34,35] and ii] the use of experimental fiRDCs for the assessment of the NA oligomeric state.[42] Examples of both applications using the MMV module are given below.

In the first example, the MMV module was used to reveal helical parameters in double-stranded DNA. The MMV module was used to confront experimental fiRDCs[34] and five distinct structural models of the Dickerson-Drew dodecamer: an NMR solution structure of the DDD (PDB ID: *1naj*), the NMR solution structure of the DDD carbocyclic analog (*2dau*),[43] the crystal structure of the DDD (PDB ID: *1bna*,[44] note that missing hydrogen atoms in the 1BNA structure were added into the structure using MolProbity web service),[45] and two canonical models of the DDD corresponding to idealized canonical A-DNA (A fiber) and B-DNA (B fiber), both models were generated by 3DNA.[46] For each of the user-provided structural models, the program returns the analogous output as that of the SMV module (Table 3 and S3).

In principle, the structural model providing the best agreement with the experimental fiRDC data can be identified based on the lowest and highest RMSD/R-factor/Q-factor and Pearson correlation coefficient values, respectively. In this particular case, the B-DNA-like solution NMR structure (PDB ID: *1naj*) was revealed as a structure providing the best agreement between the experimental and calculated fiRDCs.

However, considering an approximative nature of the fiRDC calculations and limited accuracy of experimental fiRDCs, it is important to evaluate the goodness of individual fits in terms of their statistical significance to account for a possibility that the used models can be truly distinguished within the limits of the fiRDCs experimental errors. In HERMES, this is achieved in two steps. First, the conventional one-way ANOVA test[47] is performed with null hypothesis that the mean values of residuals, defined as $\left(\overline{\sum|\varepsilon_i|}\right)_k$ (where $(\varepsilon_i)_k = (\boldsymbol{x}_{i,exper} - \boldsymbol{x}_{i,calc})_k$ is the residual i for the user-provided model k), are the same for all tested models.

Note: Prior the assessment of the statistical significance, it is advisable to inspect $(\varepsilon_i)_k$ and remove any potential outlier value(s) from the data set(s) as they can adversely bias both ANOVA and Tukey's tests.

The ANOVA test returns p-value, which, if is less than a user defined significance level S, indicates that there is less than a 100*S% probability the null hypothesis is correct. For the five models under discussion (cf. Table 3), after excluding the outliers (Figure S1 in the Supporting Information), the ANOVA test yields *p-value* of $6.59^{*}10^{-6}$, which even for S = 0.01,

Table 3. A shortened version of the summary table listing statistical descriptors of correlation between experimental fiRDCs acquired for the DDD[34] and those predicted based on five distinct structures of the DDD (*1naj*, *2dau*, *1bna*, A fiber, and B fiber) using the MMV module.[a]

Model	RMSD [Hz]	Corr. Coeff.	Q factor	R factor
1naj	0.52	0.68	0.44	0.38
B fiber	0.55	0.66	0.46	0.42
2dau	0.61	0.64	0.51	0.44
1bna	0.76	0.66	0.64	0.62
A fiber	1.02	−0.08	0.86	0.75

[a] For complete output table – see Table S3. The values for B_o^{high} and B_o^{low} were set to 18.8 T and 11.75 T, respectively. The temperature was set to 308 K. The calculations were performed under the assumption of a rigid structure (S = 1).

indicates that at least one of the models can be excluded from consideration based on statistically significant difference between predicted and experimental fiRDCs. In the second step, a Tukey HSD ("honestly significant difference") test[48] is performed to identify model(s) that provides statistically significantly better fit compared to the other model(s). The test is based on pair-wise comparison of difference in the mean values of residuals. In practical terms, if Tukey's test yields a p-value smaller than a user defined significance level S indicates that the compared models have statistically different means of residuals. Table 2 provides p-values for the pairwise comparisons of mean values of residuals between *1naj*, the model yielding lowest and highest values of RMSD/Q-factor/R-factor and Pearson correlation coefficient, respectively (cf. Table 3), and those corresponding to other structural models (for the complete Tukey's test results see Supplementary Information, Table S4).

As can be seen, *1naj* displays a statistically significant difference of mean residuals (better fit) at 0.01 significance level with both *1bna* and A fiber, whereas there are no statistically significant differences between mean values of residuals calculated for *1naj* and B fiber and *2dau* models (Table 4). In other words, the differences between predicted fiRDCs based on models *1naj*, B fiber, and *2dau* and experimental fiRDCs data are not statistically significant and all these models might be considered as equally representative of experimental DNA structure. Noteworthy, both A fiber and *1bna* models are representatives of A-form for DNA helical geometry, while *1naj*, B fiber, and *2dau* are all models representative of B-form DNA helical geometry. Note: The Tukey's HSD test is run only when the ANOVA test shows an overall statistically significant difference in the means of the residuals. The p-values from Tukey's HSD test are listed in the "Tukey_out" file).

In the second example, the MMV module is used to determine the stoichiometry of (homo)multimeric NA complexes based on measured fiRDC data.[36] To demonstrate this application, the MMV module of HERMES was used to confront experimental fiRDC values acquired for a homodimeric DNA quadruplex (PDB ID: 1f3 s)[49] by Al-Hashimi et al.[36] with fiRDCs predicted for i] the homodimeric DNA quadruplex and ii] a corresponding monomeric G-quadruplex structure (Figure 5).

The experimental parameters B_o^{high}, B_o^{low} and T were set to 18.96 T, 11.85 T, and 293 K, respectively, i. e., to the experimental conditions employed by Al-Hashimi et al.[42] In this setup, the MMV module of HERMES again returns simple linear and Deming regression plots, Pearson correlation coefficient, Q- and R-factor and RMSD values for each of the models (Figure 6, Table 5 and Table S5).

Please note that in the case of the analysis of structures differing only by virtue of their oligomeric state, the Pearson correlation coefficient cannot be directly used as a criterion for model evaluation, as the same linear correlation is expected between the experimental and predicted fiRDCs for both models. Also, Tukey's HSD test is not performed in the case of

Table 4. p-values for the Tukey HSD test for the pairwise comparisons of *1naj* with the remaining structural models.

Models	p-value
1naj-*1bna*	0.0060
1naj-A fiber	0.0000[a]
1naj-B fiber	0.8802
1naj-*2dau*	0.4917

[a] 0.000022.

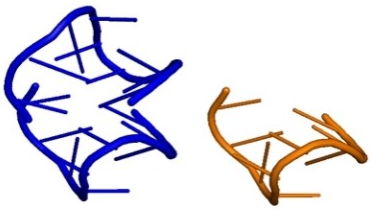

Figure 5. Cartoon representation of the structure of the homodimer (blue, PDB ID: 1f3s) and corresponding monomer (orange).

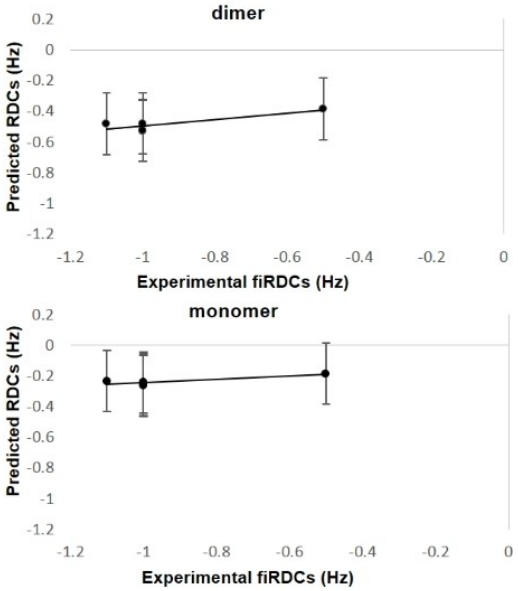

Figure 6. Regression plots between the experimental[36] and predicted fiRDCs for homodimeric DNA quadruplex (left) and monomer (right) (cf. PDB ID: 1f3s).

Table 5. A shortened version of the summary table of statistical descriptors of the fits. For complete output table, see Table S5. Note: In total, there are five experimental data points in the plots. However, three data points have identical value (−1.0 ±0.2 Hz) and therefore appear as a single point along the x-axis.

Model	RMSD	CorrCoeff	Q factor	R factor
dimer	0.47	0.86	0.50	0.48
monomer	0.71	0.86	0.75	0.74

the comparison of only two structural models as it is redundant with respect to one-way ANOVA. Provided that the p-value (in our case 0.015; one outlier value was excluded from the analysis) from one-way ANOVA test is less that user chosen significance level (in our case 0.05), which indicates statistically significant difference between fits from two models, the correct oligomeric state can be unambiguously identified using the criterion of the lowest RMSD value (here, 0.47 Hz was indicated for the dimer and 0.71 Hz for the monomer). Based on that criterion, better agreement between the experimental and predicted fiRDCs is indicated for the homodimer, which is in accordance with the results of a previous study.[36]

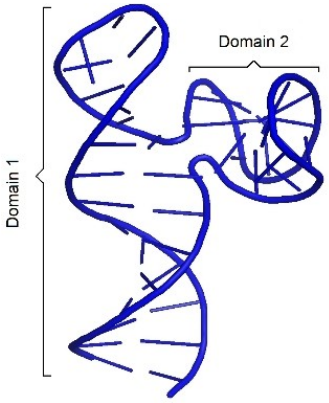

Figure 7. Structural representation of RNA fragment forming a Holliday junction (PDB ID: *2f1q*). For the purpose of our calculations, domain 1 remains in the same position, while successive intrinsic z-y′-z″ rotations are imposed on domain 2.

Relative Domain Orientation (RDO module)

The RDO module of HERMES allows estimation of the relative orientation of two rigid "domains" in NA fragment. The user is prompted to list domain-specific internuclear vectors, corresponding S and experimental fiRDC (including experimental errors) values in the "fiRDC_input" file, and to specify increments (in degrees) for three spatial angles (ψ, ϕ and θ) used by the RDO module to sample the relative orientation between domains 1 and 2. Based on the user-defined increment values, the RDO module automatically generates molecular models differing by the domain orientations. While the orientation of domain 1 in the molecular frame is kept constant, the orientation of domain 2 is sampled by intrinsic Tait-Bryan successive z-y′-x″ rotations (described by the spatial angles ψ, ϕ and θ, respectively) with angle increments defined by the user. For each of the generated models, the RDO module predicted fiRDCs and subsequently confronted the predicted fiRDCs with their experimental counterpart by means of RMSD and the Pearson correlation coefficient. The identification of the structural model revealing (close-to-correct) relative orientation between the two domains is governed by the inspection of RMSD values. (Note: In our experience, the structural model revealing (close-to-correct) relative orientation between the two domains can be identified among 2.5% of structures with the lowest RMSD values).

To illustrate the RDO application, the relative orientation of two helical segments in the RNA fragment forming a Holliday junction (PDB ID: 2f1q) was analyzed using experimental fiRDC data as acquired by van Buuren et al.[35] The Holliday junction structure consists of two stacked rigid helices, herein referred to as domain 1 and domain 2 (Figure 7). In this example, the values for B_o^{high} and B_o^{low} are set to 19 T and 9.5 T, respectively, and the temperature was set to 293 K. The increments for the ψ, ϕ and θ angles were jointly set to 10°. Based on the increment values, the RDO module produced approximately 25 thousand structural models. To facilitate the analysis of the resulting data, it is recommended to visualize the data in the form of a pseudo-4D plot, such as that displayed in Figure 8A. In the plot, the values of ψ, ϕ and θ, and are plotted along the z, y, and × axes, while the corresponding RMSD values are color-coded (cf. Figure 8). In the present case, the plot revealed the existence of four "clusters". While each cluster consists of

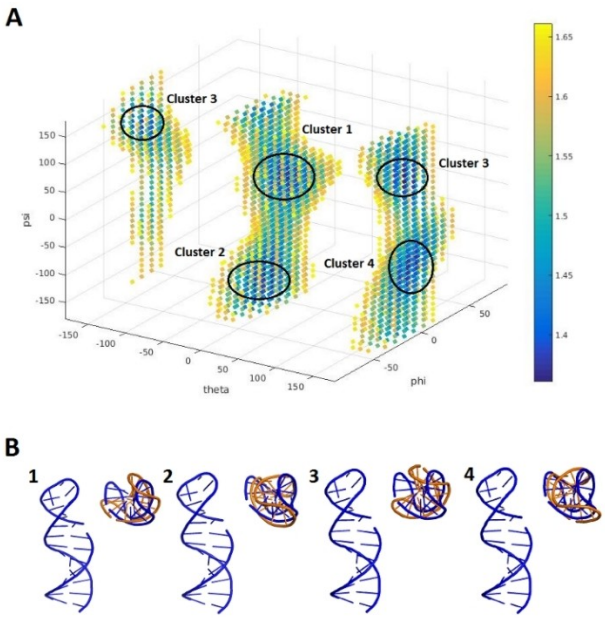

Figure 8. (A) Pseudo-4D plot displaying the distribution of ψ, ϕ, and θ angles as a function of RMSD values. Only data points from the 97,5th-percentile of structures with the lowest RMSD values are displayed. The color bar represents the range of the RMSD values. (B) A superimposition of the structures with the lowest RMSD from clusters (1–4) on the reference structure PDB ID: *2f1q* (in blue).

structurally similar models, the existence of four clusters stems from the existence of degenerate solutions of [Eq. (1)]. Superpositions of cluster-specific structural models with lowest RMDS values over the "reference" structure of a Holliday junction are displayed in Figure 8B. Note: Although it is, in principle, impossible to distinguish among individual clusters (degenerate solutions of [Eq. (1)]) purely based on RMSD values, the most plausible orientation can be estimated based on the knowledge of fundamental principles of NA stereochemistry: The most plausible orientation between two helices can be assigned to that presented by the minimum RMSD model representing cluster 1 as it does not, in contrast to the structures

representing cluster 2, 3, and 4, presume an existence of unnatural kinks and breaks in the molecular structure. Indeed, in contrast to the structure models representing cluster 2 ($\psi = -145°$, $\phi = -5°$ and $\theta = -25°$), 3 ($\psi = 105°$, $\phi = -10°$ and $\theta = 170°$), and 4 ($\psi = -75°$, $\phi = 5°$ and $\theta = 155°$), the minimum RMSD structure model representing cluster 1 and being described by $\psi = 30°$, $\phi = 5°$ and $\theta = -10°$ is reasonably close to the reference (correct) structure ($\psi = 0°$, $\phi = 0°$ and $\theta = 0°$).

A word of caution: The individual modules of HERMES can be regarded as the software implementations of the original methods developed by others.[34–36] It needs to be remembered that these methods have several inherent limitations[34–36] and that uncritical and/or inappropriate use of HERMES might lead to a structure misidentification. Next to the χ_b values and used nucleobase geometries, the assessment of nucleic acid structure from fiRDC data using HERMES will critically dependent on a choice of stereo-chemically meaningful input structure(s) and good estimates of internuclear vector-associated order parameters. (Note: For detail discussion of sensitivity of the predicted fiRDCs to χ_b values see ref. [34]). If possible, it is highly advisable to cross-validate the HERMES output with an independent set of other experimental structural restraints, such as NOEs. This is particularly the case for the RDO module of HERMES, which needs to be regarded only as an experimental-data-guided "coarse-grained" modeling tool.

All examples discussed above are accessible in the Examples section of HERMES. More detailed information is provided in the documentation section of the web application.

Conclusion

The web application HERMES for fast and reliable structural interpretation of fiRDC was implemented, developed and tested. HERMES allows users to perform complex analysis of fiRDC data in a few minutes. It is our hope that the availability of our automated tool will facilitate the use of fiRDCs in the characterization of the structure of nucleic acids and open new horizons for their applications. As fiRDCs arise as a result of spontaneous alignment of NAs in the static magnetic field, the applications of fiRDCs are suited to situations where the use of alignment media or paramagnetic tags adversely interferes with the structure of the studied nucleic acid fragments or in situations where the use of alignment media is inherently precluded, such as in NMR studies conducted in complex environments of crude cellular homogenates[50] or in the intracellular space of living cells.[49,51–55]

Theoretical background

Field-induced residual dipolar couplings

The field-induced residual dipolar couplings (fiRDCs) between nuclei A and B can be expressed as a function of the anisotropic part of the molecular magnetic susceptibility (AMMS) tensor (χ_{mol}) and A–B vector orientation using the following equation:[34]

$$fiRDC = -\left[\frac{\mu_0 (B_0)^2 \Delta\chi S \gamma_A \gamma_B h}{240\pi^3 kTr_{AB}^3}\right]\left[(3\cos^2\theta - 1) + \frac{3}{2}R\sin^2\theta\cos 2\phi\right] \quad (1)$$

where S is the generalized order parameter; γ_A and γ_B are the magnetogyric ratios of nuclei A and B, respectively; $\Delta\chi$ and R are the anisotropy and rhombicity, respectively, of the AMMS tensor; r_{AB} is the distance between nuclei A and B; θ (colatitude) and ϕ (longitude) are polar coordinates describing the orientation of the internuclear A–B vector in the principal axis system of the molecular magnetic susceptibility tensor; T is the temperature in Kelvin; k is the Boltzmann constant; and μ_0 is the vacuum permeability.

Molecular magnetic susceptibility tensor (χ_{mol})

As demonstrated by Buuren et al.,[35] χ_{mol} is caused by magnetic susceptibilities of individual nucleobases (χ_b). χ_{mol} can be approximated as a sum of individual χ_b values given that the structure of the NA molecule is known.[34,35]

$$\chi_{mol} = \sum_{i=1}^{n} (\overrightarrow{v_{b,i}}^T \times \overrightarrow{v_{b,i}})\chi_{b,i} \quad (2)$$

where $\overrightarrow{v_{b,i}}$ is a unit row vector perpendicular to the plane of base *i*, *n* is the number of bases in the NA molecule, $\chi_{b,i}$ is the magnetic susceptibility of nucleobase i and the symbol $\times$ represents the direct vector product.

Experimental fiRDCs

It needs to be noted that the Hamiltonians for both the indirect (J) and the direct (D) spin-spin interactions have the same functional form.[56] As a result, the apparent coupling constant that is experimentally measured in the case of molecular alignment is $J_{AB} + D_{AB}$. Therefore, the fiRDC values are typically determined from measurements at two magnetic fields[34] as follows:

$$fiRDC = \{({}^nJ_{AB} + {}^nD_{AB})^{high} - ({}^nJ_{AB} + {}^nD_{AB})^{low}\} = -\left[\frac{\mu_0 (B_0^{high})^2 \Delta\chi S \gamma_A \gamma_B h}{240\pi^3 kTr_{AB}^3}\right]\left[(3\cos^2\theta - 1) + \frac{3}{2}R\sin^2\theta\cos 2\phi\right] \frac{(B_0^{high})^2 - (B_0^{low})^2}{(B_0^{high})^2} \quad (3)$$

where B_0^{high} and B_0^{low} are the high and low magnetic field strengths, respectively.

Experimental Section

Nucleobase-specific geometries and χ_b values

The reference nucleobase-specific geometries and χ_b values used by default in all HERMES calculations are listed in Table S1 (Supporting Information) and in the Documentation section of the web application. Please note that the reference nucleobase-specific geometries are used to replace the corresponding nucleobase geometries provided by the user in the user-supplied 3D model/structure of NA presented in the PDB file. This serves two purposes: (i) to associate individual nucleobases in the PDB structure with the corresponding χ_b values,[38] in particular the relative orientation of the χ_b tensor with the nucleobase structure, and (ii) to correct/account for the nonplanarity of nucleobases in the user-supplied 3D NA model/structure.[57]

Statistical analysis

HERMES, by default, calculates four statistical descriptors that can be used by the user to evaluate the quality of a structural model(s) on the basis of correlation/association between the calculated and the experimental fiRDCs. These include: i) the Pearson product-moment correlation coefficient (r) (cf. [Eq. (4)]), ii) root-mean-square deviation (RMDS – cf. [Eq. (5)]), iii) R-factor (cf. [Eq. (6)]), and iii) Q-factor (cf. [Eq. (7)]).[1,58]

$$r_{pred,exper} = \frac{\sum_{i=1}^{n}(x_{i,calc} - \overline{x_{calc}})(x_{i,exper} - \overline{x_{exper}})}{\sqrt{\sum_{i=1}^{n}(x_{i,calc} - \overline{x_{calc}})^2 (x_{i,exper} - \overline{x_{exper}})^2}} \quad (4)$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}(x_{i,calc} - x_{i,exper})^2}{n}} \quad (5)$$

$$R\ factor = \frac{\sum_{i=1}^{n}|x_{i,exper} - x_{i,calc}|}{\sum_{i=1}^{n}|x_{i,exper}|} \quad (6)$$

$$Q\ factor = \frac{rms(x_{exper} - x_{calc})}{rms(x_{exper})} \quad (7)$$

where rms stands for root-mean square.

Complementary information on the numerical proportions between the two sets of fiRDCs is obtained by regression analysis. HERMES provides simple linear regression analysis and Deming regression analysis.[59] The former determines the best-line fit between the calculated and the experimental fiRDCs on the basis of minimizing the sum of absolute values of residuals and assumes that only the calculated fiRDCs are associated with random measurement errors, while the latter accounts for errors in observations in both calculated and experimental fiRDCs.

Program availability

With PHP and MATLAB binaries on the back-end and HTML/CSS and JavaScript on the front-end, the HERMES web application is publicly accessible at http://hermes.ceitec.muni.cz. The fully documented source code of the program along with the detailed user manual can be downloaded from the Additional Material section of the website. Part of the MATLAB code is based on modules developed by Vavrinská et al.[38] The binary (compiled) versions of the modules for the use of the procedure on a local PC can also be downloaded from the HERMES web site (http://hermes.ceitec.muni.cz/Binaries).

Acknowledgements

This work was supported by a grant from the Ministry of Health of the Czech Republic (NV19-08-00450 to LT) and a grant from the Czech Science Foundation (19-13436S to VS).

Conflict of Interest

The authors declare no conflict of interest.

Keywords: field-induced residual dipolar couplings · magnetic susceptibility anisotropy · NMR · nucleic acids · prediction software

[1] A. Bax, *Protein Sci.* **2003**, *12*, 1–16.
[2] J. H. Prestegard, H. M. Al-Hashimi, J. R. Tolman, *Q. Rev. Biophys.* **2000**, *33*, 371–424.
[3] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9279–9283.
[4] J. L. Lorieau, A. S. Maltsev, J. M. Louis, A. Bax, *J. Biomol. NMR* **2013**, *55*, 369–377.
[5] P. Thiagarajan-Rosenkranz, A. W. Draney, S. T. Smrt, J. L. Lorieau, *J. Am. Chem. Soc.* **2015**, *137*, 11932–11934.
[6] S. Meier, D. Häussinger, S. Grzesiek, *J. Biomol. NMR* **2002**, *24*, 351–356.
[7] R. Tycko, F. J. Blanco, Y. Ishii, *J. Am. Chem. Soc.* **2000**, *122*, 9340–9341.
[8] M. Zweckstetter, A. Bax, *J. Biomol. NMR* **2001**, *20*, 365–377.
[9] X.-C. Su, K. McAndrew, T. Huber, G. Otting, *J. Am. Chem. Soc.* **2008**, *130*, 1681–1687.
[10] J. Wöhnert, K. J. Franz, M. Nitz, B. Imperiali, H. Schwalbe, *J. Am. Chem. Soc.* **2003**, *125*, 13338–13339.
[11] M. R. Hansen, L. Mueller, A. Pardi, *Nat. Struct. Biol.* **1998**, *5*, 1065–1074.
[12] G. M. Clore, M. R. Starich, A. M. Gronenborn, *J. Am. Chem. Soc.* **1998**, *120*, 10571–10572.
[13] G. Metz, K. P. Howard, W. B. S. van Liemt, J. H. Prestegard, J. Lugtenburg, S. O. Smith, *J. Am. Chem. Soc.* **1995**, *117*, 564–565.
[14] S. M. Douglas, J. J. Chou, W. M. Shih, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6644–6648.
[15] A. Bax, A. Grishaev, *Curr. Opin. Struct. Biol.* **2005**, *15*, 563–570.
[16] N. Tjandra, A. Bax, *Science* **1997**, *278*, 1111–1114.
[17] F. Delaglio, G. Kontaxis, A. Bax, *J. Am. Chem. Soc.* **2000**, *122*, 2142–2143.
[18] G. M. Clore, A. M. Gronenborn, N. Tjandra, *J. Magn. Reson.* **1998**, *131*, 159–162.
[19] J. H. Prestegard, C. M. Bougault, A. I. Kishore, *Chem. Rev.* **2004**, *104*, 3519–3540.
[20] J. R. Tolman, H. M. Al-Hashimi, L. E. Kay, J. H. Prestegard, *J. Am. Chem. Soc.* **2001**, *123*, 1416–1424.
[21] J.-C. Hus, D. Marion, M. Blackledge, *J. Am. Chem. Soc.* **2001**, *123*, 1541–1542.
[22] H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, L. J. Smith, *Protein Sci.* **2001**, *10*, 677–688.
[23] P. Bayer, L. Varani, G. Varani, *J. Biomol. NMR* **1999**, *14*, 149–155.
[24] E. T. Mollova, M. R. Hansen, A. Pardi, *J. Am. Chem. Soc.* **2000**, *122*, 11561–11562.
[25] E. C. Murphy, V. B. Zhurkin, J. M. Louis, G. Cornilescu, G. M. Clore, *J. Mol. Biol.* **2001**, *312*, 481–499.
[26] R. Stefl, H. Wu, S. Ravindranathan, V. Sklenár, J. Feigon, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1177–1182.
[27] G. M. Clore, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9021–9025.
[28] C. Tang, D. C. Williams, R. Ghirlando, G. M. Clore, *J. Biol. Chem.* **2005**, *280*, 11770–11780.
[29] J. L. Weaver, J. H. Prestegard, *Biochemistry* **1998**, *37*, 116–128.
[30] L. Trantírek, R. Štefl, M. Vorlíčková, J. Koča, V. Sklenář, J. Kypr, *J. Mol. Biol.* **2000**, *297*, 907–922.
[31] R. J. Richards, H. Wu, L. Trantirek, C. M. O'Connor, K. Collins, J. Feigon, *RNA* **2006**, *12*, 1475–1485.
[32] R. Fiala, N. Špačková, S. Foldynová-Trantírková, J. Šponer, V. Sklenář, L. Trantírek, *J. Am. Chem. Soc.* **2011**, *133*, 13790–13793.
[33] S. Nakano, N. Sugimoto, *Mol. BioSyst.* **2017**, *13*, 32–41.

[34] D. L. Bryce, J. Boisbouvier, A. Bax, *J. Am. Chem. Soc.* **2004**, *126*, 10820–10821.
[35] B. N. M. van Buuren, J. Schleucher, V. Wittmann, C. Griesinger, H. Schwalbe, S. S. Wijmenga, *Angew. Chem. Int. Ed.* **2004**, *43*, 187–192; *Angew. Chem.* **2004**, *116*, 189–194.
[36] H. M. Al-Hashimi, J. R. Tolman, A. Majumdar, A. Gorin, D. J. Patel, *J. Am. Chem. Soc.* **2001**, *123*, 5806–5807.
[37] J. Romanuka, G. E. Folkers, N. Biris, E. Tishchenko, H. Wienk, A. M. J. J. Bonvin, R. Kaptein, R. Boelens, *J. Mol. Biol.* **2009**, *390*, 478–489.
[38] A. Vavrinská, J. Zelinka, J. Šebera, V. Sychrovský, R. Fiala, R. Boelens, V. Sklenář, L. Trantírek, *J. Biomol. NMR* **2016**, *64*, 53–62.
[39] X. Sun, Q. Zhang, H. M. Al-Hashimi, *Nucleic Acids Res.* **2007**, *35*, 1698–1713.
[40] H. Schwalbe, *Angew. Chem. Int. Ed.* **2017**, *56*, 10252–10253; *Angew. Chem.* **2017**, *129*, 10386–10387.
[41] Z. Wu, F. Delaglio, N. Tjandra, V. B. Zhurkin, A. Bax, *J. Biomol. NMR* **2003**, *26*, 297–315.
[42] H. M. Al-Hashimi, A. Majumdar, A. Gorin, A. Kettani, E. Skripkin, D. J. Patel, *J. Am. Chem. Soc.* **2001**, *123*, 633–640.
[43] A. Y. Denisov, E. V. Zamaratski, T. V. Maltseva, A. Sandström, S. Bekiroglu, K. H. Altmann, M. Egli, J. Chattopadhyaya, *J. Biomol. Struct. Dyn.* **1998**, *16*, 547–568.
[44] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2179–2183.
[45] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. Sect. D* **2010**, *66*, 12–21.
[46] X.-J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
[47] G. E. P. Box, *Biometrika* **1953**, *40*, 318–335.
[48] J. W. Tukey, *Biometrics* **1949**, *5*, 99–114.
[49] A. Kettani, G. Basu, A. Gorin, A. Majumdar, E. Skripkin, D. J. Patel, *J. Mol. Biol.* **2000**, *301*, 129–146.
[50] R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek, V. Dötsch, *Nucleic Acids Res.* **2011**, *39*, 5768–5775.
[51] R. Hänsel, S. Foldynová-Trantírková, F. Löhr, J. Buck, E. Bongartz, E. Bamberg, H. Schwalbe, V. Dötsch, L. Trantírek, *J. Am. Chem. Soc.* **2009**, *131*, 15761–15768.
[52] S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J. Mergny, S. Foldynova-Trantirkova, L. Trantirek, *Angew. Chem. Int. Ed.* **2018**, *57*, 2165–2169; *Angew. Chem.* **2018**, *130*, 2187–2191.
[53] Y. Yamaoki, A. Kiyoishi, M. Miyake, F. Kano, M. Murata, T. Nagata, M. Katahira, *Phys. Chem. Chem. Phys.* **2018**, *20*, 2982–2985.
[54] H.-L. Bao, T. Ishizuka, T. Sakamoto, K. Fujimoto, T. Uechi, N. Kenmochi, Y. Xu, *Nucleic Acids Res.* **2017**, *45*, 5501–5511.
[55] I.-C. Giassa, J. Rynes, T. Fessl, S. Foldynova-Trantirkova, L. Trantirek, *FEBS Lett.* **2018**, *592*, 1997–2011.
[56] P. Diehl, C. L. Khetrapal, R. G. Jones, *NMR: Basic Princ. Prog.. Grundlagen Und Fortschritte*, Springer Science & Business Media, **2013**.
[57] V. Sychrovsky, S. Foldynova-Trantirkova, N. Spackova, K. Robeyns, L. Van Meervelt, W. Blankenfeldt, Z. Vokacova, J. Sponer, L. Trantirek, *Nucleic Acids Res.* **2009**, *37*, 7321–7331.
[58] P. D. Thomas, V. J. Basus, T. L. James, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 1237–1241.
[59] Deming W. Edwards, *Statistical Adjustment Of Data*, **1938**.

Manuscript received: June 30, 2020
Revised manuscript received: August 31, 2020
Accepted manuscript online: September 4, 2020