



# NMR refinement and peptide folding using the GROMACS software

Anna Sinelnikova<sup>1</sup> · David van der Spoel<sup>2</sup>

Received: 10 September 2020 / Accepted: 18 March 2021 / Published online: 28 March 2021  
© The Author(s) 2021

## Abstract

Nuclear magnetic resonance spectroscopy is used routinely for studying the three-dimensional structures and dynamics of proteins and nucleic acids. Structure determination is usually done by adding restraints based upon NMR data to a classical energy function and performing restrained molecular simulations. Here we report on the implementation of a script to extract NMR restraints from a NMR-STAR file and export it to the GROMACS software. With this package it is possible to model distance restraints, dihedral restraints and orientation restraints. The output from the script is validated by performing simulations with and without restraints, including the *ab initio* refinement of one peptide.

**Keywords** Python · NMR-STAR · Force Field · Amber · Charmm

## Introduction

Nuclear Magnetic Resonance spectroscopy is a powerful technique to study structure and dynamics of biologically relevant molecules in solution (Palmer 2004; Kay 2016). Due to steady methodological progress, membrane proteins (Opella and Marassi 2017) as well as disordered proteins (Gibbs et al. 2017) and even macromolecules *in vivo* (Inomata et al. 2009; Sakakibara et al. 2009; Luchinat and Banci 2017) can now be studied using NMR spectroscopy techniques. Molecular dynamics simulations have been used for over thirty years as a tool to supplement the sometimes limited amounts of data, and to allow determination and refinement of structures or aid the interpretation of experimental data (Torda and Van Gunsteren 1991; Torda et al. 1993). In addition, NMR data can be used to validate simulation results giving detailed insights when simulated structures deviate from experimental data (van der Spoel and Lindahl 2003; Lange et al. 2010) or to validate force fields (Hornak et al. 2006; Huang and MacKerell 2013).

Determination of biomolecular structures is to a large extent automated these days (Wrz et al. 2017). Nevertheless, it may be advantageous to both the NMR and the simulations communities to have a variety of tools to analyze biomolecules using NMR data. Therefore we have implemented a script to include restraints from NMR into the GROMACS software suite for classical molecular dynamics simulations (Berendsen et al. 1995; Lindahl et al. 2001; van der Spoel et al. 2005; Hess et al. 2008; Pronk et al. 2013; Páll et al. 2015). The script is validated by performing restrained as well as unrestrained molecular dynamics simulations of peptides from the Protein Data Bank (Westbrook et al. 2002) and by performing *ab initio* refinement of a short peptide.

## Theory

### Background

Here, we briefly recap relevant equations that are implemented in the GROMACS software suite. Within classical molecular simulation software packages, trajectories of molecules can be simulated by numerically solving of Newton's equations of motion (Allen and Tildesley 1987). To do so, the force at every atom is calculated as the negative gradient of the potential function. The potential functions in turn, are divided into three different categories:

✉ David van der Spoel  
david.vanderspoel@icm.uu.se

Anna Sinelnikova  
anna.sinelnikova@physics.uu.se

<sup>1</sup> Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden

<sup>2</sup> Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

- bonded forces, including chemical bonds, angles and dihedrals,
- Van der Waals and Coulomb forces,
- different kind of restraints.

In this paper we are interested in the last group, using the restraint information that can be obtained from NMR experiments. We consider the three types of restraints that are implemented in GROMACS: distance, dihedral and orientation restraints.

Distance restraints introduce a lower and upper limit for the distance for a particular atom pair. In GROMACS this is implemented as a flat-bottom harmonic oscillator potential:

$$V_{dr}(r_{ij}) = \begin{cases} \frac{1}{2}k_{dr}(r_{ij} - r_0)^2, & \text{for } r_{ij} < r_0 \\ 0, & \text{for } r_0 \leq r_{ij} < r_1 \\ \frac{1}{2}k_{dr}(r_{ij} - r_1)^2, & \text{for } r_1 \leq r_{ij} < r_2 \\ \frac{1}{2}k_{dr}(r_2 - r_1)(2r_{ij} - r_2 - r_1), & \text{for } r_2 \leq r_{ij} \end{cases} \quad (1)$$

where  $r_{ij}$  is a distance between atoms  $i$  and  $j$ ,  $k_{dr}$  is a distance restraint force constant,  $r_0$  is a lower bound of the restraint and  $r_1$  and  $r_2$  are two upper bounds. The second upper bound is introduced to prevent extremely large forces in case an atom pair is far from the target distance. In addition, GROMACS implements time averaging (Torda et al. 1989) as well as ensemble averaging of distance restraints.

Dihedral angles can be restrained using a similar flat-bottom potential:

$$V_{dihr}(\phi') = \begin{cases} \frac{1}{2}k_{dihr}(\phi' - \phi_0 - \Delta\phi)^2, & \text{for } \phi' > \Delta\phi \\ 0, & \text{for } \phi' \leq \Delta\phi \end{cases} \quad (2)$$

where

$$\phi' = (\phi - \phi_0) \text{MOD } 2\pi \quad (3)$$

with  $\phi_0$  the reference angle, typically derived from J-coupling constants using a Karplus relation (Karplus 1959). Time averaging can be applied for dihedral restraints (Torda et al. 1993) in GROMACS as well (Lindahl et al. 2020).

Orientation restraints can be obtained from e.g. residual dipolar couplings. They have been implemented in GROMACS previously, including time and ensemble averaging (Hess and Scheek 2003). We refer to that paper or the GROMACS manual (Lindahl et al. 2020) for more information because the mathematics is rather extensive.

## Implementation details

Here we briefly describe the script “nmr2gmx.py” used to convert a NMR-STAR file (Ulrich et al. 2019) to GROMACS inputs. The NMR-STAR file format is supported by a

number of software packages and is the standard for storing processed NMR data. For the purpose of this project it is important to note that there is a Python library that can be used to read and process the content of the files (Wedell and Baskaran 2020).

The scripts use two layers of conversion of inputs. First, from NMR notation for degenerate groups to actual atoms, and second, to convert to atom names matching the force fields. The latter is needed since, unfortunately, at least three different naming schemes for hydrogen atoms are in use today despite that standard nomenclature (IUPAC-IUB 1970) predates biomolecular force fields. Table 1 lists the effective translations. The script expands, for instance, the interaction between an Ala MB and an Ile MD to 9 distances that are however treated as one restraint using  $r^{-6}$  averaging of the distances. Logical OR statements in the input for distance restraints are honored. Both dihedral restraints and orientation restraints apply the renaming conventions in Table 1. In the case of dihedral restraints, the lower and upper bounds are extracted from the data and the average angle is computed, taking periodicity into account. A harmonic potential is applied starting from the bounds (Eqn. 2). For orientation restraints the chemical shift anisotropy  $\delta$  is read from the NMR data and output to GROMACS format (Lindahl et al. 2020). Multiple chains are supported as well. More documentation for the script is at the GitHub repository (Sinelnikova et al. 2020).

## Methods

### Technical validation

A test set is part of the package. In short, 44 PDB entries are downloaded and processed and the resulting output files compared to reference tables (that is, the GROMACS input files). If input files are indeed correct, a GROMACS energy minimization is run and the output structure compared to the PDB structure. Since the energy minimization is performed *in vacuo* some conformational changes does occur but in all cases the root mean square deviation remains within 0.02 nm. By applying the scripts to a few dozen different entries, it was possible to detect potential errors. If the script is updated or extended in the future the test set can be used to make sure functionality remains intact. The test set includes systems containing proteins, RNA and DNA and those biomolecules supported by the Amber force field should work with the script as well. In systems where GROMACS does not recognize e.g. the protonation state of Histidine residues, a warning is printed and one or more restraints may be skipped.

**Table 1** Atom and identifier name translations applied in the script for amino acids and nucleobases

Residue	Identifier	Amber	Charmm
Backbone	H	H	HN
Ala	MB	HB <sub>1,2,3</sub>	HB <sub>1,2,3</sub>
Thr	MG	HG <sub>1,2,3</sub>	HG <sub>1,2,3</sub>
Ile	MG	HG2 <sub>1,2,3</sub>	HG2 <sub>1,2,3</sub>
Ile	HG1 <sub>2</sub>	HG1 <sub>1</sub>	HG1 <sub>1</sub>
Ile	HG1 <sub>3</sub>	HG1 <sub>2</sub>	HG1 <sub>2</sub>
Ile	MD	HD <sub>1,2,3</sub>	HD <sub>1,2,3</sub>
Ile	CD1	CD	CD
Ile	HD1 <sub>x</sub>	HD <sub>x</sub>	HD <sub>x</sub>
Gly	HA2	HA1	HA1
Gly	HA3	HA2	HA2
Ser, Thr, Cys	HG	HG	HG <sub>1</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met, Trp, Tyr, Phe, His, Leu	HB <sub>2</sub>	HB <sub>1</sub>	HB <sub>1</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met, Trp, Tyr, Phe, His, Leu	HB <sub>3</sub>	HB <sub>2</sub>	HB <sub>2</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met, Val	HG <sub>3</sub>	HG <sub>2</sub>	HG <sub>2</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met, Val	HG <sub>3</sub>	HG <sub>2</sub>	HG <sub>2</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met	HD <sub>3</sub>	HD <sub>2</sub>	HD <sub>2</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met	HD <sub>3</sub>	HD <sub>2</sub>	HD <sub>2</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met	HE <sub>3</sub>	HE <sub>2</sub>	HE <sub>2</sub>
Lys, Asn, Ser, Asp, Glu, Pro, Gln, Arg, Met	HE <sub>3</sub>	HE <sub>2</sub>	HE <sub>2</sub>
Met	ME	HE <sub>1,2,3</sub>	HE <sub>1,2,3</sub>
Lys	QZ	HZ <sub>1,2,3</sub>	HZ <sub>1,2,3</sub>
Arg	QH1	HH1 <sub>1,2</sub>	HH1 <sub>1,2</sub>
Arg	QH2	HH2 <sub>1,2</sub>	HH2 <sub>1,2</sub>
Leu	MD1	HD1 <sub>1,2,3</sub>	HD1 <sub>1,2,3</sub>
Leu	MD2	HD2 <sub>1,2,3</sub>	HD2 <sub>1,2,3</sub>
Tyr, Phe	QD	HD <sub>1,2</sub>	HD <sub>1,2</sub>
Tyr, Phe	QE	HE <sub>1,2</sub>	HE <sub>1,2</sub>
G, A, U, C, DG, DA, DT, DC	HO2'	HO'2	HO'2
G, A, U, C, DG, DA, DT, DC	H5'	H5'1	H5'1
G, A, U, C, DG, DA, DT, DC	H5''	H5'2	H5'2
G, A, U, C, DG, DA, DT, DC	H2'	H2'1	H2'1
G, A, U, C, DG, DA, DT, DC	H2''	H2'2	H2'2
DT	M7	H7 <sub>1,2,3</sub>	H7 <sub>1,2,3</sub>

## Simulation details

Several short polypeptides were taken from the Protein Data Bank to make a full MD run with and without restraints and thus verify the compatibility of the output from the program with the GROMACS software. The peptides are of different lengths and have different types of restraints. Table 2 lists the polypeptides used, their lengths in number of residues and what type of restraints were obtained from NMR data file for each. All proteins were simulated for 20 ns in a cubic water box with periodic boundary conditions at temperature equals 300 K. Particle mesh-Ewald summation (Darden et al. 1993; Essmann et al. 1995) was used to treat long-range Coulomb interactions, while Lennard-Jones interactions were cut-off at 1 nm with analytical tail corrections for the long range

dispersion (Allen and Tildesley 1987). Whether or not such approximations will be acceptable in the future is under scrutiny right now (van der Spoel et al. 2020). Temperature was controlled using the Bussi thermostat (Bussi et al. 2007) with a time constant of 0.5 ps, while pressure was maintained at 1 bar using the Parrinello-Rahman algorithm (Parrinello and Rahman 1981) with a time constant of 2 ps. The Amber99SB-ILDN (Cornell et al. 1995; Lindorff-Larsen et al. 2011) force field was used in combination with the tip3p water model (Jorgensen et al. 1983) to perform MD simulations.

The Charmm force field version 27 (MacKerell et al. 1998; Foloppe and MacKerell 2000) as implemented in GROMACS is supported as well in v1.0 of the program although the support in GROMACS is somewhat more

**Table 2** Proteins used for the MD simulations with number of residues and number of distance (#disres), dihedral (#dihres) respectively orientation (#orires) restraints obtained from corresponding NMR data files

PDB ID	#residues	#disres	#dihres	#orires	#models
6cj8 (Yang et al. 2018)	17	577			20
2luf (Bathula et al. 2013)	20	398			10
2leu (Fregeau Gallagher et al. 1997)	37	420			18
1lb0 (Biron et al. 2002)	13	202	9		1
2md6 (Lebbe et al. 2014)	18	59	10		15
1qqv (Vardar et al. 1999)	67	327	19		1
1d3z (Cornilescu et al. 1998)	78	2727	98	62	10
1lvz (Koenig et al. 2002)	11	121	12	8	20

Number of experimental models

limited than for Amber and therefore there are only 28 test cases. Other force fields can readily be implemented in the script and guidelines for this can be found in README file on the GitHub page (Sinelnikova et al. 2020).

## Analysis

Apart from inspecting the restraints, we compute the root mean square deviation of distances (RMSD) from the simulation trajectories as follows. All the atom-atom distances  $r_{ij}^{MD}$  in a protein are computed at each time in the simulation and the distances are compared to the corresponding  $r_{ij}^{NMR}$  in the experimental references structure. The RMSD is then computed as the root mean square difference between  $r_{ij}^{MD}$  and  $r_{ij}^{NMR}$ . The advantage of this method over positional RMSD is that the superposition step is omitted, which may lead to arbitrary jumps in RMSD due to small changes in coordinates if the protein structures differ a lot. Since multiple experimental models are available for all the proteins (Table 2), we compute the RMSD to each of the models at each time point in the simulation and then take the lowest value. The rationale behind this is that the experimental structures are equally likely, and if the simulated protein is close to any of the structures, the deviation is low.

## Results and discussion

### Evaluation of distance restraint parameters

Restrained simulations require a number of parameters like the force constant  $k_{dr}$  (Eqn. 1) and the constant  $\tau_{dr}$  used for time averaging (Torda et al. 1989). A number of different combinations of these parameters were evaluated to find values that work well in most cases. Table 3 lists the distance violations averaged over 20 ns simulations of first 6 proteins in Table 2. Based on this result we recommend a force constant  $k_{dr}$  of 1000 kJ mol<sup>-1</sup> nm<sup>-2</sup> and an averaging time  $\tau_{dr}$  of 500 ps. It should be noted that the optimal values for these

**Table 3** Evaluation of the effect of force constant  $k_{dr}$  (kJ mol<sup>-1</sup> nm<sup>-2</sup>) and averaging time  $\tau_{dr}$  (ps) on average distance violations (nm) averaged over 6 proteins

$\tau_{dr}$	$k_{dr}$			
	10	100	500	1000
0	0.020 ± 0.010	0.015 ± 0.009	0.010 ± 0.007	0.008 ± 0.006
10	0.026 ± 0.009	0.014 ± 0.009	0.010 ± 0.006	0.008 ± 0.005
100	0.021 ± 0.010	0.015 ± 0.009	0.008 ± 0.005	0.005 ± 0.002
500	0.020 ± 0.010	0.014 ± 0.008	0.006 ± 0.003	0.004 ± 0.002
1000	0.022 ± 0.010	0.014 ± 0.008	0.007 ± 0.003	0.004 ± 0.002

parameters depend on a number of factors, such as peptide length and flexibility and indeed how well folded the peptide is to start with.

### Validation

Table 4 presents a comparison of the distance, dihedral and orientation violations together with distance RMSD (see section 3.3), in simulations with and without restraints using the recommended set of distance restraints parameters according Table 3:  $k_{dr} = 1000$  kJ mol<sup>-1</sup> nm<sup>-2</sup> and  $\tau_{dr} = 500$  ps. For all types of restraints the average violations are quite a bit lower with restraints turned on, showing that the potentials are effective. The same tendency can be seen for the distance RMSD in some simulations (2leu, 1lb0, 1lvz): without restraints the deviations are higher than with restraints. For the other proteins the difference in RMSD is within the uncertainty.

### De novo refinement and folding

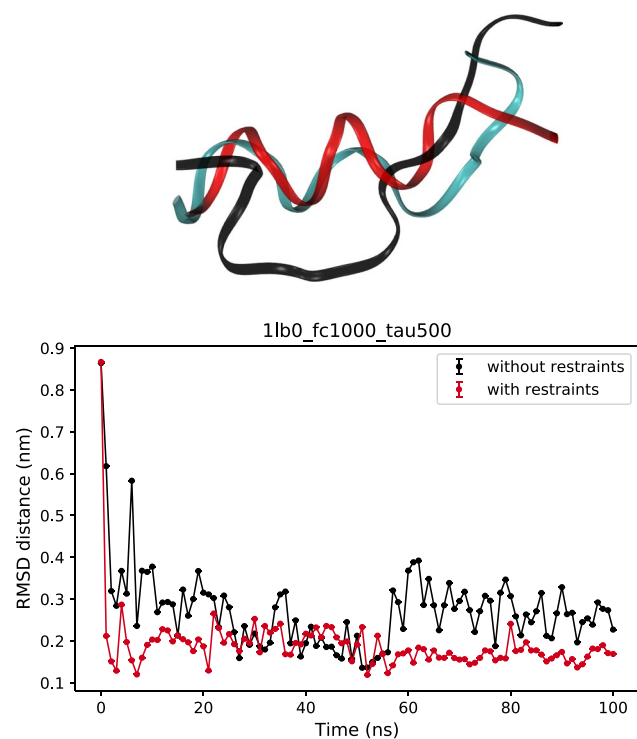
For one of the peptides a *de novo* refinement was attempted where the initial conformation for the simulation is completely extended. The folding of 1lb0 in simulations with and without restraints is shown in Figure 1 using the distance RMSD as a function of time. The reference frame for RMSD calculation is the original PDB structure. The largest change

**Table 4** Comparison table for violation of distance, dihedral and orientation restraints together with RMSD distance average for all simulated proteins

PDB ID	Distance (nm)		Dihedral ( $^{\circ}$ )		Orientation (Hz)		RMSDist (nm)	
	With	Without	With	Without	With	Without*	With	Without
6cj8	0.014	0.071	–	–	–	–	$0.14 \pm 0.05$	$0.14 \pm 0.02$
2luf	0.001	0.016	–	–	–	–	$0.14 \pm 0.05$	$0.17 \pm 0.02$
2leu	0.001	0.012	–	–	–	–	$0.29 \pm 0.08$	$0.71 \pm 0.02$
1lb0	0.002	0.002	0	5.03	–	–	$0.17 \pm 0.04$	$0.24 \pm 0.05$
2md6	0.002	0.012	0	0.20	–	–	$0.11 \pm 0.02$	$0.115 \pm 0.009$
1qqv	0.005	0.007	0	10.7	–	–	$0.18 \pm 0.02$	$0.20 \pm 0.03$
1d3z	0.001	0.002	6.12	15.0	0.41	0.94	$0.085 \pm 0.009$	$0.08 \pm 0.02$
1lvz	0.001	0.011	0.01	8.40	0.28	0.43	$0.07 \pm 0.02$	$0.15 \pm 0.04$

“With” indicates that all restraints are taken into account, while “without” means no restraints were used. The parameters for restraints are the following: Distance restraints  $\{\kappa_{dr} = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}; \tau_{dr} = 500 \text{ ps}\}$ , dihedral restraints  $\kappa_{dih} = 1000 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ , and orientation restraints  $\kappa_{or} = 10 \text{ kJ mol}^{-1} \text{ Hz}^{-2}$

\* To be able to analyse the orientation restraints violation we had to include them into the simulations, but the corresponding force constant was 3 order of magnitude lower than the one we usually use



**Fig. 1** Folding of 1lb0. Top: comparison of the structures after 100 ns of simulations. The native conformation stated in the original PDB structure (Biron et al. 2002) (cyan), with (red) and without (black) restraints. We used VMD software for the visualisation (Humphrey et al. 1996). Bottom: The distance RMSD as a function of time for simulations with (red) and without (black) restraints. RMSD is calculated against the native conformation shown at the top

in structure occurs at the beginning both simulation, where the protein is fully denatured. It can be concluded that taking into account restraints provide a faster and more robust approach to obtaining the native conformation as well as

more stable structure. Nevertheless, the final structure of the restrained simulation still differs somewhat from the reference structure. This could be due to the difference in temperature, the NMR structure was derived at 277K, whereas our simulations were done at room temperature. Indeed, circular dichroism measurements show that the peptide is somewhat less structured at room temperature (Biron et al. 2002). The average distance restraint violation is 0.001 nm for simulations with restraints and 0.010 nm for “without restraints” simulations.

## Conclusion

In this contribution, we present a Python package for importing data from nuclear magnetic resonance files NMR-STAR files into the GROMACS software. We have examined 8 different polypeptides with distance, dihedral and orientation restraints (Table 2). From a comparison of the values of corresponding restraint violation from GROMACS simulations with the restraints and without them (Table 4), we conclude that the package treat the restraints correctly.

For distance restraints we suggest the following parameters for force constant and the averaging time:  $k_{dr} = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  and  $\tau_{dr} = 500 \text{ ps}$ , based on the evaluation presented in Table 3. Another step in this research should be to find the optimal parameters for dihedral and orientation restraints in the same way as we have done for distance restraints, however it should be kept in mind that these parameters may be system dependent.

Finally, we have used GROMACS to refine the 1lb0 peptide structure from an extended conformation. Simulations with and without restraints were run and it was found (Figure 1 bottom) that simulations with restraints converge to the experimental structure faster and end up with lower



violations. The better converge can also be seen in the 3D representation of the structures at the top of the Figure 1. The red protein was simulated with the restraints and it fits the original 11b0 (cyan protein) much better than the black one which was simulated without the restraints. However, one can see that after 100 ns of the simulations even for simulations with restraints, the folding is not perfect.

**Acknowledgements** The Swedish research council is acknowledged for a grant of computer time (SNIC2019-2-32 and SNIC2020-15-67) through the National Supercomputing Centre in Linköping, Sweden. Funding from eSENCE - The e-Science Collaboration (Uppsala-Lund-Umeå, Sweden) is gratefully acknowledged. The authors would like to thank Snehal Patel for stimulating discussions. A.S. acknowledges funding from the Knut and Alice Wallenberg foundation through the Wallenberg Academy Fellow grant of J. Nilsson.

**Funding** Open access funding provided by Uppsala University.

**Data availability** The software described here, including test data, is available free of charge under the Apache License 2.0 from GitHub (Sinelnikova et al. 2020).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Allen MP, Tildesley DJ (1987) Computer simulation of liquids. Oxford Science Publications, Oxford
- Bathula S, Sklenar V, Zidek L, Vondrasek J, Vymetal J (2013) Retro trap-cage peptide to be published. <https://doi.org/10.2210/pdb21uf/pdb>
- Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Comm* 91:43–56
- Biron Z, Khare S, Samson AO, Hayek Y, Naider F, Anglister J (2002) A monomeric  $3_{10}$ -helix is formed in water by a 13-residue peptide representing the neutralizing determinant of HIV-1 on gp41. *Biochemistry* 41(42):12687–12696
- Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz-Jr KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins and nucleic acids. *J Amer Chem Soc* 117:5179–5197
- Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Amer Chem Soc* 120(27):6836–6837
- Darden T, York D, Pedersen L (1993) Particle mesh ewald: an N-log(N) method for ewald sums in large systems. *J Chem Phys* 98:10089–10092
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8592
- Foloppe N, MacKerell Jr AD (2000) All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* 21, 86–104 (2000). [https://doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2<86::AID-JCC2>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G)
- Fregeau Gallagher NL, Sailer M, Niemczura WP, Nakashima TT, Stiles ME, Vederas JC (1997) Three-dimensional structure of leucocin A in trifluoroethanol and dodecylphosphocholine micelles: spatial location of residues critical for biological activity in type IIa bacteriocins from lactic acid bacteria. *Biochemistry* 36(49):15062–15072
- Gibbs EB, Cook EC, Showalter SA (2017) Application of NMR to studies of intrinsically disordered proteins. *Arch Biochem Biophys* 628:57–70
- Hess B, Kutzner C, Van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4(3):435–447
- Hess B, Scheek RM (2003) Orientation restraints in molecular dynamics simulations using time and ensemble averaging. *J Magnet Reson* 164:19–27
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct Funct Gen* 65:712–725
- Huang J, MacKerell AD Jr (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* 34(25):2135–2145
- Humphrey W, Dalke A, Schulten K (1996) VMD - visual molecular dynamics. *J Mol Graph* 14:33–38
- Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H et al (2009) High-resolution multi-dimensional nmr spectroscopy of proteins in human cells. *Nature* 458(7234):106–109
- IUPAC-IUB Commission on biochemical nomenclature: abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969). *Biochemistry-US* 9, 3471–3478 (1970)
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935
- Karplus M (1959) Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys* 30:11–15
- Kay LE (2016) New views of functionally dynamic proteins by solution NMR spectroscopy. *J Mol Biol* 428:323–331
- Koenig BW, Kontaxis G, Mitchell DC, Louis JM, Litman BJ, Bax A (2002) Structure and orientation of a G protein fragment in the receptor bound state from residual dipolar couplings. *J Mol Biol* 322(2):441–461
- Lange OF, van der Spoel D, de Groot BL (2010) Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys J* 99:647–655
- Lebbe EK, Peigneur S, Maiti M, Devi P, Ravichandran S, Lescrier E, Ulens C, Waelkens E, D'Souza L, Herdewijn P et al (2014) Structure-function elucidation of a new  $\alpha$ -conotoxin, Lo1a, from *Conus longurionis*. *J Biol Chem* 289(14):9573–9583
- Lindahl E, Abraham M, Hess B, van der Spoel D (2020) Gromacs 2020.3 manual. <https://doi.org/10.5281/zenodo.3923644>
- Lindahl E, Hess B, Van Der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7(8), 306–317 (2001)

- Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520
- Luchinat E, Banci L (2017) In-cell NMR: a topical review. *IUCrJ* 4:110–118
- MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
- Opella SJ, Marassi FM (2017) Applications of NMR to membrane proteins. *Arch Biochem Biophys* 628:92–101
- Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E (2015) Tackling exascale software challenges in molecular dynamics simulations with gromacs. *Lect Notes Comput Sci* 8759:3–27
- Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104:3623–3640
- Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52:7182–7190
- Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854
- Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Wälchli M et al (2009) Protein structure determination in living cells by in-cell nmr spectroscopy. *Nature* 458(7234):102–105
- Sinelnikova A, Patel S, van der Spoel D (2020) Read NMR data files for proteins and generate gromacs input files. <https://doi.org/10.5281/zenodo.4019826>
- van der Spoel D, Henschel H, van Maaren PJ, Ghahremanpour MM, Costa LT (2020) A potential for molecular simulation of compounds with linear moieties. *J Chem Phys* 153(8):084503
- van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible and free. *J Comput Chem* 26:1701–1718
- Torda A, Brunne R, Huber T, Kessler H, Van Gunsteren W (1993) Structure refinement using time-averaged J-coupling constant restraints. *J Biomol NMR* 3(1):55–66. <https://doi.org/10.1007/BF00242475>
- Torda A, Van Gunsteren W (1991) The refinement of NMR structures by molecular-dynamics simulation. *Comput Phys Comm* 62(2–3):289–296. [https://doi.org/10.1016/0010-4655\(91\)90101-P](https://doi.org/10.1016/0010-4655(91)90101-P)
- Torda AE, Scheek RM, van Gunsteren WF (1989) Time-dependent distance restraints in molecular dynamics simulations. *Chem Phys Lett* 157:289–294
- Ulrich EL, Baskaran K, Dashti H, Ioannidis YE, Livny M, Romero PR, Maziuk D, Wedell JR, Yao H, Eghbalian HR, Hoch JC, Markley JL (2019) NMR-STAR: comprehensive ontology for representing, archiving and exchanging data from nuclear magnetic resonance spectroscopic experiments. *J Biomol NMR* 73:5–9
- van der Spoel D, Lindahl E (2003) Brute-force molecular dynamics simulations of villin headpiece: comparison with NMR parameters. *J Phys Chem B* 107(40):11178–11187
- Vardar D, Buckley DA, Frank BS, McKnight CJ (1999) NMR structure of an F-actin-binding headpiece motif from villin. *J Mol Biol* 294(5):1299–1310
- Wedell J, Baskaran K (2020) A python module for reading, writing, and manipulating NMR-STAR files. <https://github.com/uwbnmr/PyNMRSTAR>. Accessed 17 July 2020
- Westbrook J, Feng Z, Jain S, Bhat TN, Thanki N, Ravichandran V, Gilliland GL, Bluhm W, Weissig H, Greer DS, Bourne PE, Helen M (2002) Berman: the protein data bank: unifying the archive. *Nucleic Acids Res* 30:245–248
- Wrz JM, Kazemi S, Schmidt E, Bagaria A, Gntert P (2017) NMR-based automated protein structure determination. *Arch Biochem Biophys* 628:24–32
- Yang Y, Cornilescu G, Tal-Gan Y (2018) Structural characterization of competence-stimulating peptide analogues reveals key features for ComD1 and ComD2 receptor binding in *Streptococcus pneumoniae*. *Biochemistry* 57(36):5359–5369

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.