Predicting $^{19}$F NMR Chemical Shifts: A Combined Computational and Experimental Study of a Trypanosomal Oxidoreductase–Inhibitor Complex

Johannes C. B. Dietschreit, Annika Wagner, T. Anh Le, Philipp Klein, Hermann Schindelin, Till Opatz, Bernd Engels, Ute A. Hellmich,* and Christian Ochsenfeld*

Abstract: The absence of fluorine from most biomolecules renders it an excellent probe for NMR spectroscopy to monitor inhibitor–protein interactions. However, predicting the binding mode of a fluorinated ligand from a chemical shift (or vice versa) has been challenging due to the high electron density of the fluorine atom. Nonetheless, reliable $^{19}$F chemical-shift predictions to deduce ligand-binding modes hold great potential for in silico drug design. Herein, we present a systematic QM/MM study to predict the $^{19}$F NMR chemical shifts of a covalently bound fluorinated inhibitor to the essential oxidoreductase tryptaredoxin (Tpx) from African trypanosomes, the causative agent of African sleeping sickness. We include many protein–inhibitor conformations as well as monomeric and dimeric inhibitor–protein complexes, thus rendering it the largest computational study on chemical shifts of $^{19}$F nuclei in a biological context to date. Our predicted shifts agree well with those obtained experimentally and pave the way for future work in this area.

Fluorine is considered a “magic” element in medicinal and agricultural chemistry. It forms strong bonds to carbon, is the smallest biocompatible hydrogen substituent,[1] has the ability to form hydrogen bonds, and possesses a high electronegativity. Its introduction into small molecules can increase metabolic stability and allows the fine-tuning of physicochemical properties.[2] It is therefore not surprising that more than 20% of all FDA-approved drugs and more than 30% of all agrochemicals contain fluorine.[3] Replacing hydrogen by fluorine has been used successfully to, for example, investigate the interaction of inhibitors with proteases, explore their active site properties, and characterize inhibitors for neglected tropical diseases.[3]

With its 100% natural abundance, high gyromagnetic ratio, and the resulting high sensitivity, the spin-1/2 nucleus $^{19}$F is of particular interest for NMR studies.[4] While practical advantages of fluorine for NMR spectroscopy have been exploited for many decades, the performance of corresponding quantum-chemical calculations for complex systems has gained momentum only lately.[5]

Chemical shifts of compounds containing fluorine have been calculated for many decades, from small molecules in the gas phase over biological systems in solution to solids.[6] The two most recent studies focusing on $^{19}$F chemical shifts of biologically relevant molecules investigated crystals of fluorinated tryptophans[7] or monofluorinated phenylalanines in a protein (Brd4).[8] In the case of the tryptophan crystals, four molecules were used as a representation of the entire crystal. For Brd4, a quantum-mechanical/molecular-mechanical (QM/MM) setup was used with a buffer region of 4 Å and Boltzmann weighting of a few conformers. Nonetheless, the calculations differed from the measurements by between one and more than 20 ppm even after improving...
predictions by linear regression to experimental data. Another study benchmarked different levels of quantum-chemical methods for fluorinated amino acids in implicit solvent, achieving at best a mean absolute error of 2.68 ppm with respect to the experiment.[9] Despite the impressive progress in the field, this is not sufficient to explain subtle differences in experimental spectra. Here, we use hundreds of frames from molecular dynamics (MD) simulations to ensure proper sampling of conformers and a significantly larger buffer region in our QM/MM calculations to increase the accuracy of our results.

Methods for computing NMR parameters range from empirical programs, such as SPARTA++,[10] to highly accurate QM calculations.[5,11,12] When using quantum-chemical methods, it has been shown that sufficiently large QM regions are necessary when describing complex systems.[13,14] However, the inclusion of many atoms is computationally very demanding. Thus, a plethora of methods has been devised to reduce the computational effort.[14,15] Here, we employ rigorous linear-scaling formulations that allow us to exploit the locality of the electronic structure within density-matrix-based theories. While this strongly reduces the computational scaling, for example, for the computation of NMR chemical shifts within density-functional theory from cubic to asymptotically linear, the accuracy is numerically unchanged and fully controlled.[5,16]

As a medically relevant test system, we selected the oxidoreductase tryparedoxin (Tpx), an essential enzyme of *Trypanosoma brucei*, the parasite that causes African sleeping sickness.[17] Tpx is inhibited by covalently binding to 2-(chloromethyl)-5-(4-fluorophenyl)thieno[2,3-d]pyrimidine-4-(3H)-one, CFT, which efficiently kills *T. brucei.[18,19]* CFT carries a 4-fluorophenyl moiety (Figure 1A and Supporting Information, Figure S5). The chloride leaving group facilitates the covalent interaction with Cys40 in the active site of Tpx.

In the asymmetric unit of our monoclinic crystals, three protein chains with two different inhibitor orientations are present (PDB: 6GXY, binding pose 1 for chains A and B, binding pose 2 for chain C, Figure 1B).[19] In binding pose 1, the covalently bound CFT features extensive intramolecular interactions with the protein, including T-shaped π-stacking interactions with Trp70 and a weak hydrogen bond of the CFT fluorine with the backbone-Hα of Glu107. In binding pose 2, CFT is not in contact with the protein beyond the covalent bond to Cys40, and its fluorine atom is solvent exposed instead. In both, crystal and solution, CFT binding to the wild-type protein in pose 1 leads to Tpx dimerization mediated by extensive intermolecular inhibitor–inhibitor stacking and inhibitor–protein interactions.[19] The dissociation constant for the CFT-induced Tpx dimer is approximately 5 μM. In binding pose 2, dimerization is structurally not possible. We
identified residue Trp39 to be crucial for dimerization. Mutation of this active-site residue to alanine (Tpx-W39A) yields a protein that can still covalently interact with CFT, but dimerization upon inhibitor binding is extremely weak (Supporting Information, Figure S1).[19]

The $^{19}$F signals for free CFT and its unreactive analogue 2-methyl-5-(4-fluorophenyl)thieno[2,3-d]pyrimidin-4(3H)-one, MFT, which is missing the chlorine leaving group (Figure 1A), were measured in solution at 298 K. Both $^{19}$F chemical shifts were found to be very similar (−114.79 and −114.77 ppm, respectively). Upon binding of CFT to Tpx and the subsequent dimerization, a downfield chemical shift of approximately 0.3 ppm for the $^{19}$F signal (−114.47 ppm) and substantial line broadening are observed, in agreement with incorporation of CFT into a high molecular weight, dimeric complex (Figure 2). Hence, the simultaneous availability of $^{19}$F NMR and X-ray data for the Tpx–inhibitor system renders it exceptionally well-suited for systematic $^{19}$F chemical shift studies. At high concentrations (greater than 500 μM), the $^{19}$F chemical shift for the CFT–Tpx complex does not depend on the concentration of the protein–inhibitor complex. This suggests that under these conditions the $^{19}$F chemical shift is not influenced by the monomer/dimer equilibrium since these concentrations are far above the $K_d$ for dimerization. Only after significantly lowering the concentration, the $^{19}$F signal starts to shift further downfield, indicating an increasing population of the monomer in exchange with the dimer (Supporting Information, Figure S2). However, due to limitations in the signal-to-noise ratio for the $^{19}$F NMR measurements at high dilution, the concentration of the complex could not be reduced far enough below the $K_d$ (≪ 5 μM) to reach conditions where the monomer is exclusively observed. In agreement with the above observation, the measured chemical shift for CFT bound to the W39A-mutant is shifted further downfield by an additional approximately 0.8 ppm compared to the WT complex (−113.7 ppm). This mutant does not significantly dimerize, and the distances from the inhibitor’s fluoride group to the W39 C× atom and to the W39 indole group are more than 16 Å and 11.5 Å, respectively, and thus should not affect the inhibitor’s $^{19}$F chemical shift. Ergo, this mutant can be used as a reference point for the $^{19}$F chemical shift of CFT bound to a Tpx monomer.

In contrast to small organic molecules, the free energy landscape of solvated biomolecules typically does not possess one deep minimum, but rather a multitude of energetically close configurations that are thermodynamically accessible at physiological temperatures. In the NMR experiment, this implies the recording of an ensemble-averaged NMR chemical shift. Therefore, many different configurations of the system have to be taken into account to compute the observed chemical shift.[20] It has been demonstrated that inclusion of bond-length and bond-angle vibrations is often crucial for reliable chemical-shift computations.[21,22] Furthermore, the different relative orientations of molecules have to be accounted for as well. Here, we present a robust methodology based on MM-MD sampling and subsequent QM/MM calculations of $^{19}$F NMR chemical shifts to identify CFT’s binding poses relevant in solution.

Based on the crystal structure of the Tpx–CFT complex (PDB: 6GXY),[19] we computed separate MM-MD trajectories for free CFT and MFT in solution, for the monomeric complex with CFT bound in either pose 1 or 2, and the Tpx-dimer with both protomers binding to CFT in pose 1 (see the Supporting Information for setup details). An analysis of the complex stability during the simulations and inhibitor flexibility is given in Figures S4–S8 and Table S2 in the Supporting Information.

Subsequently, the $^{19}$F chemical shifts of the inhibitor were calculated with our linear-scaling methods[5,16] for 200 snapshots taken evenly spaced in time from each MD trajectory. All interactions of either the protein or solvent atoms with the inhibitor were included explicitly. The aim was to perform high quality $^{19}$F chemical-shift calculations on a large scale, based on an adequate description of the accessible phase space. The configurational ensembles included, for example, different solvation patterns or orientations of the inhibitor towards Tpx.

The high electron density of $^{19}$F renders its spectroscopic properties, such as NMR chemical shifts, particularly challenging to calculate.[23] Motivated by previous studies,[24,25] we tested KT2[26] and B97-2[27] together with the NMR-specific basis set pcS-1[28] and the necessary QM buffer size (Supporting Information, Figure S9). Both functionals show identical QM size convergence. As the radius of the QM region is increased from 4 to 5 Å, the calculated chemical shift changes by more than 0.5 ppm, highlighting the importance of
a sufficiently large QM sphere. A 7 Å QM buffer region around the inhibitor was found to be necessary to obtain size-converged shifts. Hence, we performed QM/MM-NMR calculations with KT2/pcS-1 for a 7 Å QM region. For the dimeric wild-type Tpx–CFT complex, this region (Figure 1C) includes more than 1000 atoms. In combination with the 200 sampling points taken from each of the individual MD trajectories, this makes it one of the largest QM-based 19F NMR chemical-shift calculations reported so far.[5,8,27] The QM size and the combined number of calculations in this context are unprecedented and are at the frontier of what is currently possible for such large biological systems. The embedded QM regions are shown in Figure 1C–F. A detailed description is given in the Supporting Information (Supporting Information, Table S3).

It is important to note that the values of the calculated 19F shifts for individual MD frames are scattered over a large shift range of about 60 ppm (Supporting Information, Figure S10). This is, however, not unexpected due to bond-length and bond-angle vibrations in the MD simulations.[25] The vibrations causing these distributions, especially those of the C-F bond, are fast processes on the spectroscopic time scale, and thus not observed experimentally. We use the experimentally measured 19F signal of free CFT in solution as reference for our calculated values, as relative shieldings are much more accurate.

As expected, our calculated 19F chemical shifts correctly predict that the chemical shifts for free CFT and MFT are very similar to each other. Importantly, the predicted chemical shift for the inhibitor bound to Tpx is calculated to be downfield shifted compared to free CFT. An even more pronounced downfield shift compared to free CFT is predicted for CFT bound to monomeric Tpx in binding pose 1. This is in excellent agreement with what is observed experimentally for CFT bound to the monomeric W39A Tpx-mutant and in our dilution experiments (Supporting Information, Figures S1 and S2). In contrast, for the monomeric complex in pose 2, a chemical shift is calculated that lies between free CFT and the dimeric complex. Thus, our calculations qualitatively predict the correct order of the 19F chemical shifts for CFT in the different states, as well as the true direction of chemical-shift changes induced by protein binding and complex dissociation. They further suggest that true direction of chemical-shift changes induced by protein

Acknowledgements

C.O. acknowledges funding by the “Deutsche Forschungsgemeinschaft” (DFG, German Research Foundation)—SFB 1309-325871075 and support as a Max-Planck Fellow at the Max-Planck Institute for Solid-State Research in Stuttgart. U.A.H. acknowledges support by the Carl Zeiss Foundation and the JGU Mainz Innenuniversitäre Forschungsförderung. This work was supported by the Rhineland-Palatinate Natural Products Research Center and the Center for Biomolecular Magnetic Resonance (BMRZ), Frankfurt University which is funded by the state of Hesse. H.S. acknowledges support from the Rudolf Virchow Center for Experimental Biomedicine. We thank Elke Duchardt-Ferner, Benedikt Goretzki, and Luise Krauth-Siegel for stimulating discussions.

Conflict of interest

The authors declare no conflict of interest.

Keywords: African sleeping sickness - covalent inhibitors - NMR spectroscopy - quantum chemistry - structural biology

[1] E. Neil, G. Marsh, Chem. Biol. 2000, 7, R153–R157.
[2] A. Strunecáň, J. Patočka, P. Connett, J. Appl. Biomed. 2004, 2, 141–150; K. Müller, C. Fieh, F. Diederich, Science 2007, 317, 1881–1886.
[3] M. Giroud, M. Harder, B. Kuhn, W. Haap, N. Trapp, W. B. Schweitzer, T. Schirmeister, F. Diederich, ChemMedChem 2016, 11, 1042–1047; J. A. Olsen, D. W. Banner, P. Seiler, U. O. Sander, A. D’Arcy, M. Stihle, K. Müller, F. Diederich, Angew. Chem. Int. Ed. 2003, 42, 2507–2511; Angew. Chem. 2003, 115, 2611–2615; M. Berninger, C. Erk, A. Füll, J. Skaf, E. Al-Momani, I. Israel, M. Raschig, P. Güntzel, S. Samnick, U. Holzgrabe, Eur. J. Med. Chem. 2018, 152, 377–391.
[4] C. Kang, Curr. Med. Chem. 2019, 26, 4964; D. Rose-Spering, M. A. Tran, L. M. Lauth, B. Goretzki, U. A. Hellmich, Biol. Chem. 2019, 400, 1277.
[5] C. Ochsenfeld, J. Kussmann, F. Koziol, Angew. Chem. Int. Ed. 2004, 43, 4485–4489; Angew. Chem. 2004, 116, 4585–4589; M. Beer, J. Kussmann, C. Ochsenfeld, J. Chem. Phys. 2011, 134, 074102.
[6] U. Sternberg, M. Klipfel, S. L. Grage, R. Witter, A. S. Ulrich, Phys. Chem. Chem. Phys. 2009, 11, 7048–7060; M. E. Harding,
