Check for updates

Localization of ligands within human carbonic anhydrase II using ^{19}F pseudocontact shift analysis†

Kaspar Zimmermann, Daniel Joss, [ID] Thomas Müntener, [ID] Elisa S. Nogueira, Marc Schäfer, Livia Knörr, Fabien W. Monnard and Daniel Häussinger [ID] *

Unraveling the native structure of protein–ligand complexes in solution enables rational drug design. We report here the use of ^{19}F pseudocontact shift (PCS) NMR as a method to determine fluorine positions of high affinity ligands bound within the drug target human carbonic anhydrase II with high accuracy. Three different ligands were localized within the protein by analysis of the obtained PCS from simple one-dimensional ^{19}F spectra with an accuracy of up to 0.8 Å. In order to validate the PCS, four to five independent magnetic susceptibility tensors induced by lanthanide chelating tags bound site-specifically to single cysteine mutants were refined. Least-squares minimization and a Monte–Carlo approach allowed the assessment of experimental errors on the intersection of the corresponding four to five PCS isosurfaces. By defining an angle score that reflects the relative isosurface orientation for different tensor combinations, it was established that the ligand can be localized accurately using only three tensors, if the isosurfaces are close to orthogonal. For two out of three ligands, the determined position closely matched the X-ray coordinates. Our results for the third ligand suggest, in accordance with previously reported *ab initio* calculations, a rotated position for the difluorophenyl substituent, enabling a favorable interaction with Phe-131. The lanthanide–fluorine distance varied between 22 and 38 Å and induced ^{19}F PCS ranged from 0.078 to 0.409 ppm, averaging to 0.213 ppm. Accordingly, even longer metal–fluorine distances will lead to meaningful PCS, rendering the investigation of protein–ligand complexes significantly larger than 30 kDa feasible.

Introduction

In order to localize and study ligands bound to proteins, sources of structural restraints are required. In contrast to X-ray crystallography[1] and cryo-electron microscopy,[2] paramagnetic nuclear magnetic resonance (NMR) spectroscopy can deliver valuable structural information in solution.[3–21] Among other effects associated with the introduction of paramagnetic metal centres complexed in a chelating tag, as *e.g.* residual dipolar couplings (RDC)[17,22–24] or paramagnetic relaxation enhancement (PRE),[13,25–27] pseudocontact shifts (PCS) stand out as an ideal tool to deliver long range information on the localization and orientation of ligand molecules within proteins taking into account the 3D structural restraints that can be obtained using this method and the favourable distance dependence of r^{-3}.[28–34]

To unlock the opportunities that PCS of protein-bound ligands offer, the NMR signals of the ligand have to be determined unambiguously. Proton resonances of the ligand usually overlap with the protein signals, thus, rendering a successful

assignment difficult. This holds in particular true for ligands that bind firmly to the protein, since they exhibit the same rotational correlation time as the protein and their lines are therefore broadened to a similar extent. This usually renders a discrimination between protein and ligand signals difficult. Whereas isotope filtered NMR experiments suffer from a low signal-to-noise ratio and long measurement times, isotopic labelling of the ligand carbon- and hetero atoms includes labour intensive chemical modification of the ligand and the effort has to be repeated for each ligand under investigation.

However, due to the 100% natural abundance of NMR active isotopes such as ^{19}F or ^{31}P, no further isotope enrichment is necessary for the acquisition of NMR spectra of molecules containing these elements. In particular fluorine is of great interest, since whereas in 1970 only 2% of drugs incorporated fluorine, the percentage of drugs containing at least one fluorine atom in 2014 amounted to 25%.[35] If protein ligands containing ^{19}F are used, their ^{19}F chemical shift can be obtained directly from one-dimensional ^{19}F NMR spectra, superseding isotopic labelling of ligands and solvent signal.

We selected three sulphonamide ligands for human carbonic anhydrase II (hCA II) in order to test their localization within the protein using PCS NMR spectroscopy (Fig. 1). The tested ligands are all derivatives of the well-established class of

Department of Chemistry, University of Basel, St. Johanns-Ring 19, 4056 Basel, Switzerland. E-mail: daniel.haeussinger@unibas.ch
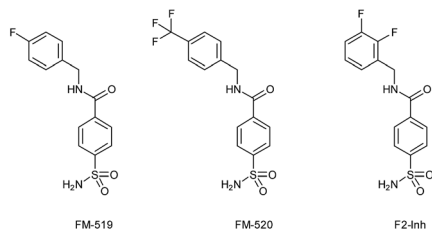
Fig. 1 Selected ligands for our proof-of-principle study.

sulphonamide ligands that bind with their sulphonamide moiety to the zinc ion in the active site of the protein and exhibit nanomolar affinities to the human carbonic anhydrase II.[36–39]

In order to unambiguously localize ligands within a protein using PCS NMR spectroscopy, four different anisotropy tensors need to be determined from four different tagging sites or from less tagging sites by using different lanthanide chelating tags (Fig. 2). Intersection of the isosurfaces of two anisotropy tensors lead to a remaining curve (Fig. 2B). Intersection of this curve with a third isosurface leads to two points as remaining possibilities for the localization of the ligand (Fig. 2C). To determine the position exclusively in a strict mathematic view, a fourth isosurface is needed (Fig. 2D). However, in practice only three tensors suffice, since the remaining other possibility for the position of the ligand can usually be excluded due to chemical and structural reasons, i.e. the remaining "ghost site" lies in a very dense region or completely outside of the investigated protein.

With the information about the anisotropy tensors in hand, compounds can then be screened and their position localized by measuring only one-dimensional ^{19}F NMR spectra. We envisioned to introduce five serine to cysteine mutations suitable to serve as tagging sites into hCA II and express the different mutants in different labelling schemes, i.e. uniformly ^{15}N labelled, selectively ^{15}N leucine labelled, as well as one

mutant [2]H [13]C [15]N. Whereas the ^{15}N labelled sample can be used to record ^{1}H–^{15}N HSQC spectra for determination of the pseudocontact shifts needed for the determination of the anisotropy tensors, selectively ^{15}N labelled samples are a convenient starting point for the evaluation of the anisotropy tensors caused by lanthanide chelating tags. Leucine stands out as an ideal amino acid type for selective labelling, since it is the most abundant amino acid in proteins and the selective labelling of recombinant proteins is straightforward and inexpensive. For hCA II, the 26 Leu residues (10%) show a favourable distribution in the primary sequence, as well as in [1]H–[15]N spectra. The triple labelled mutant was expressed for backbone assignment and a diamagnetic lutetium tag was attached to the protein in order to avoid ambiguities in the assignment caused by a small number of residues that shift upon tagging. Upon expression, tagging and determination of the anisotropy tensors, fluorine pseudocontact shifts can then be analysed in order to localize the selected ligands within the protein.

In order to obtain sufficiently large PCS for the localization of ligands in a protein with a molecular weight and size in the range of hCA II (diameter: 40–56 Å, molecular weight: 30 kDa),[39] sufficiently rigid tags have to be used. Lanthanide chelating tags based on a sterically overcrowded DOTA-M8 scaffold[11] as well as the recently reported M7PyThiazole-DOTA[21] provided suitable tools for our intent.

By application of the envisioned methodology, we demonstrate that the position of ligands within a protein can be determined solely by measurement of one-dimensional ^{19}F NMR experiments and analysis of the obtained ^{19}F PCS over a distance range of 22–38 Å. Human carbonic anhydrase II is expressed in three different labelling schemes for five different single-cysteine constructs and three ligands are tested. During the drug candidate screening process, the presented method significantly reduces NMR measurement times compared to NOE-based approaches and eliminates the need for isotope labelled protein as well as chemical modification of the ligand. Furthermore, we observe differences in the solution structure of one ligand (F2-Inh) when compared to the crystal structure and establish an angle score as a measure of the suitability of chosen anisotropy tensors.

Materials and methods

Extensive protocols of the expression of uniformly ^{15}N labelled hCA II mutants, selectively ^{15}N leucine labelled hCA II mutants, uniformly [2]H [13]C ^{15}N labelled hCA II mutants, tagging reactions, NMR experiments and pseudocontact shift analysis are available in the ESI.†

Sample preparation

To a protein sample, 1.1 eq. of ligand dissolved in deuterated dimethyl sulphoxide (DMSO-d_6) were added and the sample was stirred well. The excess of ligand was removed by ultrafiltration using 4 ml Amicon ultrafiltration tubes (MW cut-off 10 kDa). Trifluoroacetate as internal standard for fluorine chemical shift
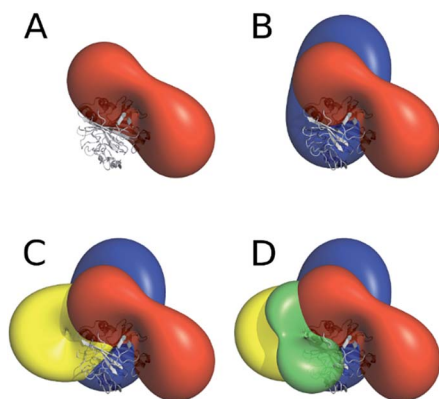

Fig. 2 Tensors required for unambiguous localization of a given atom by using PCS as sole source of structural restraints (A) isosurface (red) and hCA II (grey), (B) intersection of two isosurfaces, (C): intersection of three isosurfaces resulting in two intersection points (one visible, one on the other side of the protein scaffold), (D) intersection of four isosurfaces resulting in only one intersection point (visible in center of (D)).

calibration as well as 5% v/v D_2O were added and the pH was adjusted to 6.80.

NMR experiments

A ^1H–^{15}N-HSQC experiment reported by Grzesiek et al.[40] with a pulse sequence including a water suppression scheme using selective water flip-back pulses and gradients to suppress radiation damping was used for the determination of the PCS, since it does not cause the small additional shift due to dipolar coupling (rdc) introduced in TROSY-$^1H$$^{15}N$-HSQC spectra. For backbone assignment of uniformly ^2H ^{13}C ^{15}N labelled hCA II S50C Lu-DOTA-M8 and all spectra recorded for protein characterization, TROSY[41] variants of HSQC, HNCO, HNCA, HN(CO)CA, HN(CA)CO and HNCACB experiments incorporating a WATERGATE scheme were used.[42–44] NMR spectra were recorded on Bruker Avance III or Avance III HD spectrometers operating at 600.13 MHz proton frequency, equipped with 5 mm BBFO, ^1H/^{13}C/^{15}N TXI or ^1H/^{19}F–^{13}C/^{15}N-D QCI-F probe heads, all fitted with z-axis pulsed field gradients.

Results and discussion

In order to ensure a rigid and stable attachment point for the Ln-DOTA-M8-SSPy and Ln-M7PyThiazole-DOTA complexes without compromising the global fold of hCA II, five solvent exposed serine residues in regions of the protein with rigid secondary structure were selected and mutated to cysteine (Fig. 3). The native cysteine residue 206 was replaced by a serine for all constructs to avoid any interference. Double mutants containing this modification, e.g. hCA II S50C C206S, are abbreviated in the following, e.g. as hCA II S50C.

Although an NMR assignment for the wild type protein is available,[45] a backbone assignment of uniformly ^2H ^{13}C ^{15}N labelled hCA II S50C Lu-DOTA-M8 was performed to have an unambiguous assignment including the shifted residues close to the tagging site. In order to provide a convenient starting point for the PCS assignment, selectively ^{15}N leucine labelled hCA II mutants were expressed. Having selectively ^{15}N leucine labelled protein and an assignment of uniformly labelled hCA II tagged with Lu-DOTA-M8 for our purpose in hand, more than eight shifted peaks were readily assigned for each mutant from an overlay of the ^1H–^{15}N HSQC spectra of diamagnetic Lu-

DOTA-M8-SSPy and paramagnetic Tm-DOTA-M8-SSPy attached to selectively ^{15}N leucine labelled hCA II (Fig. 4). Based on the assigned PCS for the leucine residues, the remaining PCS can be back-calculated using the initially fitted anisotropy tensor. Although it is possible to start the analysis of the PCS from the fully labelled protein (in the less crowded regions of the spectrum), this method simplifies considerably the analysis of the spectra. The mutant hCA II S173C was excluded from further analysis due to the appearance of two sets of signals in the HSQC spectrum of the thulium tagged construct. These two sets arise most likely from two equally populated positions of the tag molecule when attached to the protein, since residue 173 is located on the edge of a beta sheet (Fig. 9 and 10 in ESI†).

Parameters describing the magnetic susceptibility tensor of the lanthanide metal were then determined from this initial set of PCS using the program Numbat and the X-ray structure of hCA II (PDB code: 3KS3).[46,47] The initially obtained tensor set allowed to back-calculate the expected PCS for the remaining leucine residues. Upon unambiguous assignment of all remaining shifted peaks apart from those leucine residues in a distance smaller than 10.9 Å from the tag, where the signals were either broadened beyond the limit of detection due to PRE or shifted outside of the spectral window, refined tensor parameters were then calculated from this larger set of PCS.

Subsequently, a complete assignment and evaluation of the magnetic susceptibility tensors using all detectable resonances was performed (Fig. 5). The refined tensors for the uniformly labelled hCA II coincided within 20% with the initial tensors of the ^{15}N selectively leucine labelled protein constructs.

Table 1 lists the refined tensor parameters obtained for the different protein mutants and used Monte–Carlo methods. All metal coordinates were located in a distance range of 6–8 Å to the γ-oxygen of the serine residue at the position where the
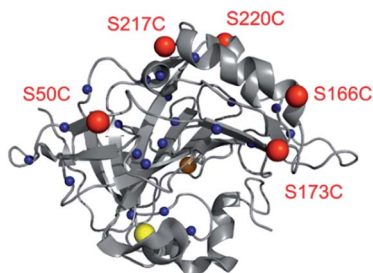

Fig. 3 X-ray structure of hCA II (3KS3). Red: selected serine to cysteine mutation sites, yellow: native Cys-206, blue: leucine residues and orange: zinc ion.
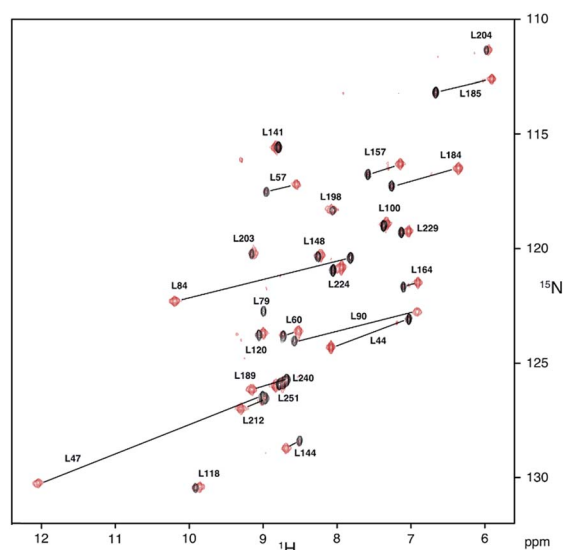

Fig. 4 Overlay of the ^1H–^{15}N-HSQC spectra (600 MHz, 298 K) of selectively ^{15}N leucine labeled hCA II S50C Lu-DOTA-M8-SSPy (black) and hCA II S50C Tm-DOTA-M8-SSPy (red). Assignment and PCS are indicated.
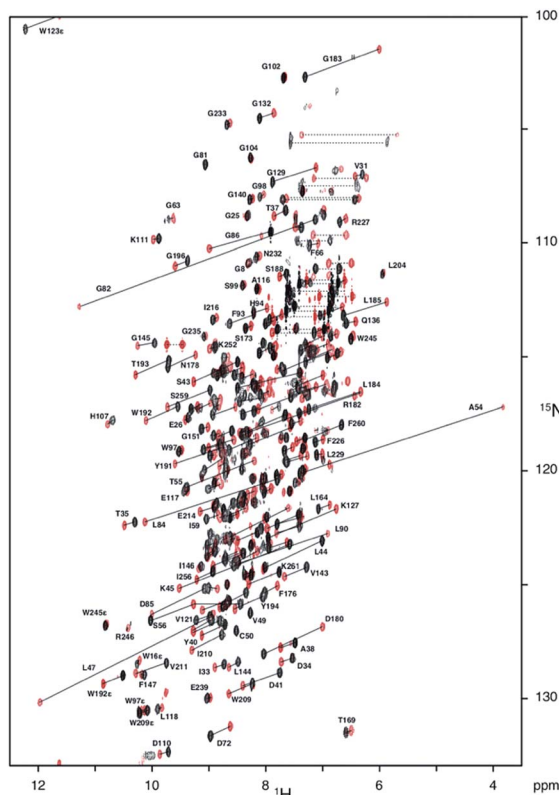
Fig. 5 Overlay of the ^1H–^{15}N HSQC spectra (600 MHz, 298 K) of uniformly ^{15}N labeled hCA II S50C Lu-DOTA-M8-SSPy (black) and hCA II S50C Tm-DOTA-M8-SSPy (red). Aliased signals with negative intensity are shown with dashed contour lines. PCS are indicated with solid and NH$_2$ groups with dashed lines. Water signal was omitted for clarity.

corresponding sulphur atom of the protein mutant is expected and the obtained Q-factors for the fits of 3.5–10.6% were excellent. For the mutants S50C, S217C and S220C similar values for $\Delta\chi_{ax}$ were found, whereas this value was almost twice as large for the S166C mutant indicating a lower flexibility of the tag when attached to the protein.[48] This was also reflected in the obtained pseudocontact shifts, where significantly larger shifts were found for the S166C mutant. An extensive protocol of the PCS assignment and fitting procedure applied is given in the ESI.† In order to have a further anisotropy tensor at hand, Lu- and Tm-M7PyThiazole-DOTA were attached to selectively ^{15}N leucine labelled hCA II S166C and the anisotropy tensor of the magnetic susceptibility was determined (Table 1). The additional lanthanide chelating tag was introduced, since it was expected to be orthogonal to the already investigated constructs and we wanted to demonstrate that successful localization of the ligand is possible with a combination of tensors from uniformly and selectively leucine labelled constructs.

In order to test the newly developed approach for the localization of ligands within a protein of interest, three sulphona-mide ligands were added to the protein as a dimethyl sulphoxide (DMSO) solution in small excess of 1.1 eq. to ensure saturation of the protein with ligand. Complete loading of the protein with ligand was then confirmed by ^1H–^{15}N HSQC spectra. Due to the high affinity of phenyl-sulphonamide based ligands,[38,39] the excess ligand could then be removed by ultra-filtration. Chemical shift changes upon ligand binding were only observed for the residues in the binding pocket of the protein, which indicates that the overall structure of the protein did not change upon binding to the ligand.

^{19}F PCS in the range of −0.078 to −0.409 could be observed for the five investigated hCA II constructs (Fig. 6, 6 and 7 in

Table 1 Refined magnetic susceptibility tensor parameters for the different hCA II constructs

Parameter	Unit	S50C Tm-DOTA	S166C Tm-DOTA	S217C Tm-DOTA	S220C Tm-DOTA	S166C Tm-Thiazole
No. of PCS	—	366	397	364	366	44
$\Delta\chi_{ax}$	[10^{-32} m^3]	21.6 ± 1.2	38.5 ± 2.0	25.7 ± 1.0	23.6 ± 0.9	34.7 ± 0.6
$\Delta\chi_{rh}$	[10^{-32} m^3]	8.5 ± 0.7	8.0 ± 1.0	13.2 ± 0.6	4.3 ± 0.3	13.3 ± 1.1
x	[Å]	−27.8 ± 0.3	−16.3 ± 0.4	−24.9 ± 0.2	−13.1 ± 0.3	−11.8 ± 0.5
y	[Å]	13.7 ± 0.3	−3.6 ± 0.4	−17.7 ± 0.3	−26.7 ± 0.3	−1.7 ± 0.1
z	[Å]	18.1 ± 0.3	−11.2 ± 0.4	19.6 ± 0.2	3.2 ± 0.2	−11.0 ± 0.2
α	[°]	104.1 ± 1.6	52.2 ± 1.8	143.7 ± 0.8	14.9 ± 1.4	119.4 ± 4.4
β	[°]	142.3 ± 1.1	123.6 ± 1.4	70.9 ± 0.5	153.6 ± 0.6	162.1 ± 0.5
γ	[°]	116.2 ± 1.7	140.3 ± 5.6	125.5 ± 1.0	1.0 ± 2.6	44.8 ± 3.7
Q		0.072	0.037	0.064	0.106	0.061
Monte–Carlo structure variation with σ = 0.5 Å						σ = 0.05 Å
No. of PCS	—	366	397	364	366	44
$\Delta\chi_{ax}$	[10^{-32} m^3]	21.1 ± 0.9	37.4 ± 0.9	25.5 ± 1.1	23.0 ± 2.1	34.3 ± 0.8
$\Delta\chi_{rh}$	[10^{-32} m^3]	8.5 ± 0.6	7.8 ± 0.6	13.1 ± 0.6	4.4 ± 0.8	12.7 ± 1.6
x	[Å]	−27.5 ± 0.3	−16.2 ± 0.2	−24.8 ± 0.2	−13.0 ± 0.6	−12.0 ± 0.7
y	[Å]	13.6 ± 0.2	−3.6 ± 0.2	−17.5 ± 0.3	−26.4 ± 0.6	−1.8 ± 0.2
z	[Å]	18.2 ± 0.2	−11.0 ± 0.2	19.6 ± 0.2	3.2 ± 0.5	−10.8 ± 0.4
α	[°]	104.0 ± 1.8	52.7 ± 1.2	143.3 ± 0.8	16.2 ± 3.5	121.6 ± 6.8
β	[°]	141.8 ± 0.8	123.1 ± 0.7	71.2 ± 0.7	153.7 ± 1.5	161.6 ± 1.1
γ	[°]	115.9 ± 1.7	141.1 ± 2.7	125.1 ± 1.1	3.9 ± 6.6	45.1 ± 5.2
Q		0.071	0.035	0.068	0.106	0.059

Monte–Carlo protocol where random subsets consisting of 20% of the available PCS were used
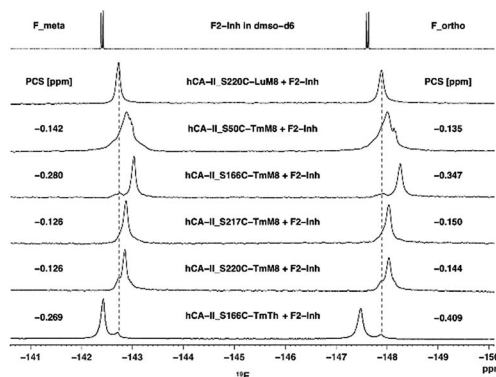
Fig. 6 One-dimensional ^{19}F NMR spectra (565 MHz, 298 K) of F2-Inh in the presence of different hCA II constructs. Spectra of diamagnetic hCA II constructs with F2-Inh did not show any differences, therefore only one diamagnetic reference spectrum is displayed. ^{19}F chemical shift was calibrated to an internal signal of trifluoroacetate at −79.0 ppm.

ESI†). Interestingly, the fluorine atoms of the ligand in hCA II S166C-Tm-Thiazole shift in the opposite direction when compared to the signals for hCA II tagged with Tm-DOTA-M8-SSPy. This shift behaviour originates in the different Euler angles (α, β, γ) observed for the individual tags. The linewidths of the signals correspond to 20–30 Hz for the ligands FM-519 and FM-520, 40–50 Hz for the signal of the fluorine atom of F2-Inh in *meta* position and 60–70 Hz for the signal of the fluorine atom of F2-Inh in ortho position. When comparing these values with linewidths obtained by Eddy *et al.* for the β_2-AR(^{TET}C265)-carazolol complex in the range of 220–230 Hz (ref. 49) and by considering the thermal displacement parameters in the PDB 1G54,[50] it can be concluded that the aromatic ring of the ligands used in this study pointing out of the enzyme's pocket still exhibits residual mobility. The largest PCS using Tm-DOTA-M8-SSPy was observed for the hCA II S166C mutant, caused by the lower mobility of the tag with respect to the protein.[48] The observed PCSs of a given value are found on an isosurface defined as follows:

$$\delta^{PCS} = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh} \sin^2\theta \cos 2\phi \right]$$

whereas δ^{PCS} is an experimentally determined PCS, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic components of the magnetic susceptibility tensor. The solutions for a given PCS and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ of this equation for the polar coordinates r, θ and φ, deliver the mentioned isosurface. The position and orientation of this isosurface within a reference coordinate system is determined by three Euclidean coordinates of the corresponding lanthanide metal and three Euler angles (α, β, γ) given in the ZYZ' convention.[46]

According to these considerations, when four back-calculated PCSs of different tensors are available, the position of a specific nucleus can be determined by the method of least squares. A sum of square residuals was defined as follows:

$$s(x, y, z) = \sum_{i=1}^{n} \left(\delta_i^{PCS}\left(x_i', y_i', z_i' \right) - \delta_i^{PCS}(\text{experiment}) \right)^2$$

where i is the index of a corresponding tensor and n is the total number of individual tensors available. δ_i^{PCS} (experiment) is the experimental PCS for the analyzed tensors and $\delta_i^{PCS}(x_i', y_i', z_i')$ is defined as follows:

$$\delta_i^{PCS}\left(x_i', y_i', z_i' \right) = \frac{1}{12\pi r_i^3} \left[\Delta\chi_{ax,i} \frac{2(z_i')^2 - (x_i')^2 - (y_i')^2}{r_i^2} \right.$$
$$\left. + \frac{3}{2}\Delta\chi_{rh,i} \frac{(x_i')^2 - (y_i')^2}{r_i^2} \right]$$

(x_i', y_i', z_i') are the coordinates x, y, z transformed into the according tensor frame. This transformation is carried out by translation of the reference coordinate system to the position of the corresponding metal center and rotation of the coordinate system by the three Euler angles (α, β, γ) which are given for each tensor. $\Delta\chi_{ax,i}$, $\Delta\chi_{rh,i}$ and r_i correspond to the axial and rhombic tensor components and the distance between lanthanide metal and the observed position (x, y, z) respectively.

The optimized position of the fluorine atoms of the used ligands was then determined by minimization of the target function $s(x, y, z)$ defined above using the SciPy library.[51] Error analysis of the determined fluorine position was performed applying a Monte–Carlo protocol where the tensor parameters were varied for every iteration according to the uncertainties determined in Numbat. For each fluorine position, 10 000 iterations were carried out with a random seed in order to ensure comparability. The resulting values and uncertainties were the average and standard deviation of these 10 000 iterations.

Interestingly, upon analysis of the obtained pseudocontact shifts and calculation of the position of the ligand, we unambiguously and successfully localized all three inhibitors within the protein using fluorine pseudocontact shift restraints over a distance range of 22–38 Å. Since the achieved distances for the localization of a ligand within a protein using lanthanide
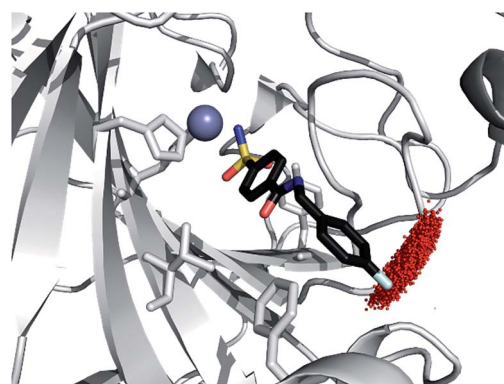

Fig. 7 Point cloud of the Monte–Carlo fluorine position calculation for FM-519 ⊂ hCA-II (modified PDB 1G54,[50] procedure in ESI†); light blue stick: fluorine atom, blue sphere: Zn^{2+} ion.
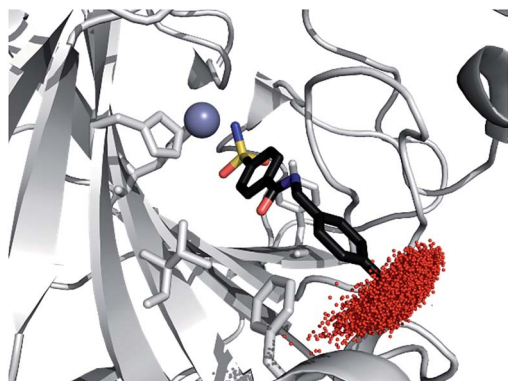
Fig. 8 Point cloud of the Monte–Carlo fluorine position calculation for FM-520 ⊂ hCA-II (modified PDB 1G54,[50] procedure in ESI†); light blue sticks: fluorine atoms, blue sphere: Zn^{2+} ion.

chelating tags are in the range of 9.9–25 Å (Xu *et al.*: 14.8–19.4 Å,[34] Saio *et al.*: 9.9 Å,[28] Guan *et al.*: 15–25 Å (ref. 29)), this result constitutes an unprecedented distance range.

The back-calculated positions for the fluorine atoms of FM-519 and FM-520 show a deviation of 3.3 and 0.8 Å when compared to the closely related X-ray structure of the penta-fluoro derivate (PDB 1G54 (ref. 50)). Both the graphical analysis using the tensor isosurfaces for a given PCS value as well as the performed Monte–Carlo simulations yield similar positions for the fluorine atoms (Fig. 7 and 8, Table 2).

The results obtained for FM-519 and FM-520 show that the method is applicable for the localization of ligands of a protein with a reasonable precision by only acquiring one-dimensional ^{19}F spectra and analysis of the obtained pseudocontact shifts.

Compared to the X-ray structure (PDB 1G52 (ref. 50)), the determined fluorine positions for F2-Inh differed by 1.6 Å for the fluorine atom in *meta* position of the difluorophenyl substituent and by 2.6 Å for the fluorine atom in *ortho* position (Fig. 9). Interestingly, this result is only obtained when the difluorophenyl substituent of the ligand is rotated by 157° in a way that it aligns in a nearly perpendicular fashion with the neighbouring phenylalanine ring of the residue F131. Supported by MP2

Table 2 Positional accuracy derived by the intersection of different combinations of isosurfaces, their associated angle score and comparison with the X-ray structure (procedures in ESI)

Ligand	S50C Tm-DOTA	S166C Tm-DOTA	S217C Tm-DOTA	S220C Tm-DOTA	S166C-Tm-Thiazole	Positional deviation from X-ray structure (Å)	Pos. dev. obtained using combinations with angle score < 30° (Å)	Angle score (°)
FM-520	✓	✓	✓	x	✓	0.8	—	—
	✓	✓	✓	x	x	1.8	1.2	14.3
	✓	✓	x	x	✓	2.1	2.7	29.8
	✓	x	✓	x	✓	2.8	2.1	19.2
	x	✓	✓	x	✓	9.3	9.6	40.0
FM-519	✓	✓	✓	x	✓	3.3	—	—
	✓	✓	✓	x	x	4.1	1.3	10.3
	✓	✓	x	x	✓	2.5	2.8	28.5
	✓	x	✓	x	✓	4.7	1.9	18.1
	x	✓	✓	x	✓	10.9	11.2	34.0
F2-Inh (*ortho*-F)	✓	✓	✓	✓	✓	2.7	—	—
	✓	✓	✓	x	x	4.3	1.3	23.2
	✓	✓	x	✓	x	3.2	1.2	25.7
	✓	✓	x	x	✓	1.9	5.3	24.0
	✓	x	✓	✓	x	4.3	1.0	14.4
	✓	x	✓	x	✓	5.2	2.0	21.8
	✓	x	x	✓	✓	13.0	9.2	38.2
	x	✓	✓	✓	x	4.3	1.1	29.0
	x	✓	✓	x	✓	8.1	4.8	29.4
	x	✓	x	✓	✓	9.8	6.6	49.6
	x	x	✓	✓	✓	11.3	7.6	45.3
F2-Inh (*meta*-F)	✓	✓	✓	✓	✓	1.7	—	—
	✓	✓	✓	x	x	3.9	1.0	14.7
	✓	✓	x	✓	x	6.7	3.4	28.8
	✓	✓	x	x	✓	3.3	5.9	24.8
	✓	x	✓	✓	x	4.0	1.1	14.5
	✓	x	✓	x	✓	5.3	3.8	30.5
	✓	x	x	✓	✓	5.2	3.5	30.3
	x	✓	✓	✓	x	4.6	1.3	29.4
	x	✓	✓	x	✓	9.1	6.2	32.2
	x	✓	x	✓	✓	13.3	10.2	42.4
	x	x	✓	✓	✓	16.3	12.9	47.4

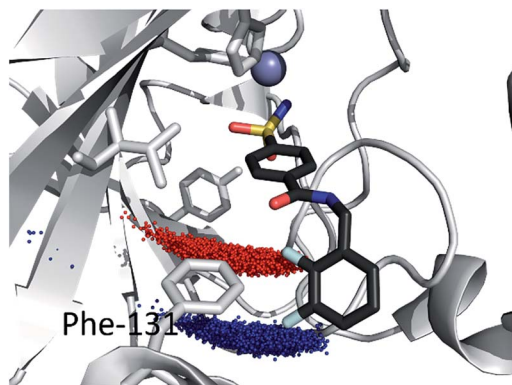**Fig. 9** Point clouds of the Monte−Carlo fluorine position calculation for F2-Inh ⊂ hCA- II. Red points are obtained from the PCS of the fluorine in *ortho* position and the blue points from the fluorine in *meta* position respectively (difluorophenyl substituent of the ligand rotated by 157° (modified PDB 1G52 (ref. 50)); light blue sticks: fluorine atoms, blue sphere: Zn^{2+} ion.

Fig. 10 Intersection points of different isosurfaces for FM-520 ⊂ hCA-II (cyan: center of gravity of the CF_3 fluorine atoms of FM-520 (modified PDB 1G54,[50] procedure in ESI†), gold: position obtained by least square minimization using all four tensors; orange and red: positions obtained by intersecting three isosurfaces in all possible combinations; orange: angle score below 30°, red: angle score above 30°); light blue sticks: fluorine atoms.

calculations that investigated the influence of the fluorine substitution pattern of the difluorophenyl substituent on the interactions with a benzene ring and were published in a previous study[52] (ref. 52, page 2, motif 1b), we propose that in the solution state structure of F2-Inh the difluorophenyl substituent adopts the described position (Fig. 9).

Notably, the intramolecular fluorine–fluorine distance in F2-Inh was determined as 3.5 Å, reproducing the distance obtained in the X-ray structure (2.8 Å) with an accuracy of 0.7 Å. This result corroborates the high accuracy of the method presented in this work. Upon primary localization of the fluorine-containing ligand within the protein, further optimization of the position could be performed using protein–ligand docking software.

In order to investigate the precision and accuracy of the graphical analysis of the presented method by using three iso-surfaces, we iterated through all possible combinations of three isosurfaces for the measured PCS for FM-520. Using a Python script, the normal vectors at the intersection point were deter-mined and their intersection angle was extracted.

The closer the obtained angle matches 90°, the more precisely the position of a fluorine atom is defined at the intersection point. We then added up the obtained normalized differences to 90° of the different pairs of the normal vectors to get an angle score. An angle score of 0° means perfectly orthogonal intersections, while 90° results from parallel iso-surfaces. The positions determined from three isosurfaces with an angle score below 30° closely matches the determined position using Monte−Carlo protocols with 4 tensors (orange spheres in Fig. 10). For three isosurfaces with an unfavourable angle score of 40° we found a significant deviation of 10 Å from the above position (red sphere in Fig. 10). However, the ach-ieved accuracy would still be high enough to discriminate *e.g.* two distant binding sites in a protein.

In summary, the results for FM-519, FM-520 and F2-Inh show that it is possible to unambiguously localize a ligand approaching an accuracy of 0.8 Å based solely on the fluorine pseudocontact shift caused by lanthanide chelating tags by
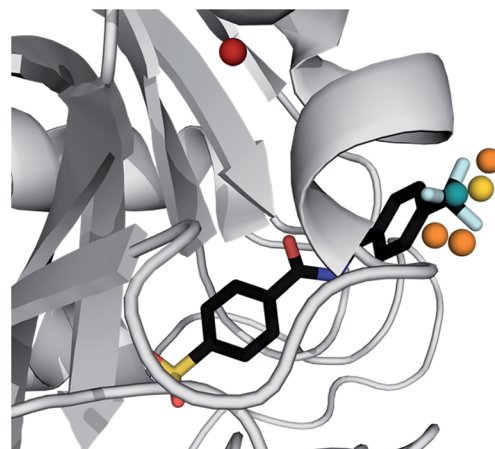
measuring 4–5 data points per ligand obtained from one-dimensional ^{19}F spectra, which provide one unique solution for the position the ligand. Interestingly, for F2-Inh the results obtained with our method suggest differences of the solution state and crystal state structure. The graphical analysis using only three tensors shows a precision that is sufficient to discriminate two binding sites within a protein. When only combinations are taken into account showing an angle score below 30°, the obtained accuracy allows a clear localization of the fluorine atom and the position matches the one determined in Monte−Carlo protocols using 4 tensors. By omitting the need for solvent suppression, chemical modification of the ligand and extensive measurement times during the screening process, the method constitutes a fast, reliable and convenient approach to screen a high number of fluorine-containing ligands for a specific protein of interest.

Conclusions

To conclude, it was demonstrated for the first time that the position of ligands within a protein can be determined solely by measurement of one-dimensional ^{19}F NMR experiments and analysis of the obtained ^{19}F PCS over a distance range of 22–38 Å in a protein of a molecular weight of 30 kDa. Human carbonic anhydrase II was expressed in different labelling schemes for five different single-cysteine constructs and three ligands were tested. During the drug candidate screening process, the presented method drastically reduces NMR measurement times compared to NOE experiments and eliminates the need for isotope labelled protein as well as labour intensive chemical modification of the ligand. Due to the large PCS observed over a distance range of 22–38 Å, the method could be extended to proteins significantly larger than 30 kDa, allows for the investigation of deeply buried ligand binding sites, avoids interference of the tag with the ligand

interaction site and renders monitoring of ternary or quaternary protein ligand complexes feasible.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips, *Nature*, 1958, **181**, 662–666.

2 J. Vonck, D. N. Parcej and D. J. Mills, *PLoS One*, 2016, **11**, e0159476.

3 C. Nitsche and G. Otting, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2017, **98–99**, 20–49.

4 B. B. Pan, F. Yang, Y. Ye, Q. Wu, C. Li, T. Huber and X. C. Su, *Chem. Commun.*, 2016, **52**, 10237–10240.

5 T. Müntener, D. Häussinger, P. Selenko and F. X. Theillet, *J. Phys. Chem. Lett.*, 2016, **7**, 2821–2825.

6 Y. Hikone, G. Hirai, M. Mishima, K. Inomata, T. Ikeya, S. Arai, M. Shirakawa, M. Sodeoka and Y. Ito, *J. Biomol. NMR*, 2016, **66**, 99–110.

7 W. M. Liu, M. Overhand and M. Ubbink, *Coord. Chem. Rev.*, 2014, **273**, 2–12.

8 C. T. Loh, K. Ozawa, K. L. Tuck, N. Barlow, T. Huber, G. Otting and B. Graham, *Bioconjugate Chem.*, 2013, **24**, 260–268.

9 G. Otting, *Annu. Rev. Biophys.*, 2010, **39**, 387–405.

10 X. C. Su, H. Liang, K. V. Loscha and G. Otting, *J. Am. Chem. Soc.*, 2009, **131**, 10352–10353.

11 D. Häussinger, J. R. Huang and S. Grzesiek, *J. Am. Chem. Soc.*, 2009, **131**, 14761–14767.

12 X. C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle and G. Otting, *J. Am. Chem. Soc.*, 2008, **130**, 10486–10487.

13 P. H. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand and M. Ubbink, *J. Am. Chem. Soc.*, 2008, **130**, 14802–14812.

14 G. Pintacuda, A. Y. Park, M. A. Keniry, N. E. Dixon and G. Otting, *J. Am. Chem. Soc.*, 2006, **128**, 3696–3702.

15 K. D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J. D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A. B. Seven, E. A. Prinslow, R. Voleti, D. Haussinger, A. M. Bonvin, D. R. Tomchick, M. Vendruscolo, B. Graham, T. C. Sudhof and J. Rizo, *Nat. Struct. Mol. Biol.*, 2015, **22**, 555–564.

16 V. Gaponenko, A. S. Altieri, J. Li and R. A. Byrd, *J. Biomol. NMR*, 2002, **24**, 143–148.

17 T. Ikegami, L. Verdier, P. Sakhaii, S. Grimme, B. Pescatore, K. Saxena, K. M. Fiebig and C. Griesinger, *J. Biomol. NMR*, 2004, **29**, 339–349.

18 F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovacic, S. Becker, R. Boelens and C. Griesinger, *J. Biomol. NMR*, 2011, **51**, 329–337.

19 J. Wohnert, K. J. Franz, M. Nitz, B. Imperiali and H. Schwalbe, *J. Am. Chem. Soc.*, 2003, **125**, 13338–13339.

20 D. Joss, R. M. Walliser, K. Zimmermann and D. Häussinger, *J. Biomol. NMR*, 2018, **72**, 29–38.

21 T. Müntener, J. Kottelat, A. Huber and D. Häussinger, *Bioconjugate Chem.*, 2018, **29**, 3344–3351.

22 I. Bertini, P. Kursula, C. Luchinat, G. Parigi, J. Vahokoski, M. Wilmanns and J. Yuan, *J. Am. Chem. Soc.*, 2009, **131**, 5134–5144.

23 J. Feeney, B. Birdsall, A. F. Bradbury, R. R. Biekofsky and P. M. Bayley, *J. Biomol. NMR*, 2001, **21**, 41–48.

24 R. Barbieri, I. Bertini, G. Cavallaro, Y. M. Lee, C. Luchinat and A. Rosato, *J. Am. Chem. Soc.*, 2002, **124**, 5581–5587.

25 I. Bertini, A. Donaire, B. Jimenez, C. Luchinat, G. Parigi, M. Piccioli and L. Poggi, *J. Biomol. NMR*, 2001, **21**, 85–98.

26 H. Yagi, K. V. Loscha, X. C. Su, M. Stanton-Cook, T. Huber and G. Otting, *J. Biomol. NMR*, 2010, **47**, 143–153.

27 H. W. Orton and G. Otting, *J. Am. Chem. Soc.*, 2018, **140**, 7688–7697.

28 T. Saio, K. Ogura, K. Shimizu, M. Yokochi, T. R. Burke Jr and F. Inagaki, *J. Biomol. NMR*, 2011, **51**, 395–408.

29 J. Y. Guan, P. H. Keizers, W. M. Liu, F. Lohr, S. P. Skinner, E. A. Heeneman, H. Schwalbe, M. Ubbink and G. Siegal, *J. Am. Chem. Soc.*, 2013, **135**, 5859–5868.

30 G. Pintacuda, M. John, X. C. Su and G. Otting, *Acc. Chem. Res.*, 2007, **40**, 206–212.

31 M. John, G. Pintacuda, A. Y. Park, N. E. Dixon and G. Otting, *J. Am. Chem. Soc.*, 2006, **128**, 12910–12916.

32 U. Brath, S. I. Swamy, A. X. Veiga, C. C. Tung, F. Van Petegem and M. Erdelyi, *J. Am. Chem. Soc.*, 2015, **137**, 11391–11398.

33 J. Gao, E. Liang, R. Ma, F. Li, Y. Liu, J. Liu, L. Jiang, C. Li, H. Dai, J. Wu, X. Su, W. He and K. Ruan, *Angew. Chem., Int. Ed. Engl.*, 2017, **56**, 12982–12986.

34 D. Xu, B. Li, J. Gao, Z. Liu, X. Niu, G. Nshogoza, J. Zhang, J. Wu, X. C. Su, W. He, R. Ma, D. Yang and K. Ruan, *J. Phys. Chem. Lett.*, 2018, **9**, 3361–3367.

35 J. Wang, M. Sanchez-Rosello, J. L. Acena, C. del Pozo, A. E. Sorochinsky, S. Fustero, V. A. Soloshonok and H. Liu, *Chem. Rev.*, 2014, **114**, 2432–2506.

36 M. Schmid, E. S. Nogueira, F. W. Monnard, T. R. Ward and M. Meuwly, *Chem. Sci.*, 2012, **3**, 690–700.

37 C. T. Supuran, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 3467–3474.

38 J. B. Doyon, E. A. Hansen, C. Y. Kim, J. S. Chang, D. W. Christianson, R. D. Madder, J. G. Voet, T. A. Baird Jr, C. A. Fierke and A. Jain, *Org. Lett.*, 2000, **2**, 1189–1192.

39 V. M. Krishnamurthy, G. K. Kaufman, A. R. Urbach, I. Gitlin, K. L. Gudiksen, D. B. Weibel and G. M. Whitesides, *Chem. Rev.*, 2008, **108**, 946–1051.

40 S. Grzesiek and A. Bax, *J. Am. Chem. Soc.*, 1993, **115**, 12593–12594.

41 K. Pervushin, R. Riek, G. Wider and K. Wuthrich, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 12366–12371.

42 M. Salzmann, K. Pervushin, G. Wider, H. Senn and K. Wuthrich, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 13585–13590.

43 M. Salzmann, G. Wider, K. Pervushin and K. Wuthrich, *J. Biomol. NMR*, 1999, **15**, 181–184.

44 M. Salzmann, G. Wider, K. Pervushin, H. Senn and K. Wuthrich, *J. Am. Chem. Soc.*, 1999, **121**, 844–848.

45 R. A. Venters, B. T. Farmer 2nd, C. A. Fierke and L. D. Spicer, *J. Mol. Biol.*, 1996, **264**, 1101–1116.

46 C. Schmitz, M. J. Stanton-Cook, X. C. Su, G. Otting and T. Huber, *J. Biomol. NMR*, 2008, **41**, 179–189.

47 B. S. Avvaru, C. U. Kim, K. H. Sippel, S. M. Gruner, M. Agbandje-McKenna, D. N. Silverman and R. McKenna, *Biochemistry*, 2010, **49**, 249–251.

48 E. A. Suturina, D. Häussinger, K. Zimmermann, L. Garbuio, M. Yulikov, G. Jeschke and I. Kuprov, *Chem. Sci.*, 2017, **8**, 2751–2757.

49 M. T. Eddy, T. Didenko, R. C. Stevens and K. Wuthrich, *Structure*, 2016, **24**, 2190–2197.

50 C. Y. Kim, J. S. Chang, J. B. Doyon, T. T. Baird, C. A. Fierke, A. Jain and D. W. Christianson, *J. Am. Chem. Soc.*, 2000, **122**, 12125–12134.

51 T. E. Oliphant, *Comput. Sci. Eng.*, 2007, **9**, 10–20.

52 P. P. Chandra, A. Jain and A. M. Sapse, *J. Mol. Model.*, 2004, **10**, 1–5.