# Protein–fragment complex structures derived by NMR molecular replacement

## RESEARCH ARTICLE

Check for updates

# Protein–fragment complex structures derived by NMR molecular replacement†

Felix Torres, [ID][a] Dhiman Ghosh,[a] Dean Strotz,[a] Celestine N. Chi, [ID][‡][a] Ben Davis [ID][b] and Julien Orts [ID]*[a]

Recently we have established an NMR molecular replacement method, which is capable of solving the structure of the interaction site of protein–ligand complexes in a fully automated manner. While the method was successfully applied for ligands with strong and weak binding affinities, including small molecules and peptides, its applicability on ligand fragments remains to be shown. Structures of fragment–protein complexes are more challenging for the method since fragments contain only few protons. Here we show a successful application of the NMR molecular replacement method in solving structures of complexes between three derivatives of a ligand fragment and the protein receptor PIN1. We anticipate that this approach will find a broad application in fragment-based lead discovery.

## Introduction

Fragment-based lead discovery (FBLD) is a powerful method used to identify chemical starting scaffolds and optimize them to obtain lead molecules. FBLD provides a clear structure activity relationship with simple chemical building blocks that can be relatively easily functionalized and cover a significant amount of chemical diversity.[1–3] This strategy has been successful against several targets which proved difficult for traditional drug discovery process.[1] The hits are generally weak affinity binders, and can subsequently be optimized with the use of structural information.[4] X-ray crystallography is usually the prime method for deriving three-dimensional structures, but this can sometimes be difficult, especially for early fragment hits with low affinity.[5,6]

While NMR is a highly versatile structural technique providing structural as well as dynamic information, its application in the context of structure based drug design remains infrequent.[7] This is mainly due to the laborious process of determining macromolecular three-dimensional structures by NMR. The complexity of spectra acquisition and interpretation as well as considerable time required for resonances assignments, derivation of distance restraints and structure calculations remains a cumbersome and time-

consuming procedure compared to the equivalent protocols used in X-ray crystallography. Recently, we introduced the NMR molecular replacement ($N$MR$^2$) that enables the fast and robust NMR structure calculation of a protein–ligand interaction site with the use of semi-ambiguous NOE distance restraints.[8–11] $N$MR$^2$ structure determination of the complex generally follows a straightforward three steps protocol: (i) preparation of the sample with either the protein or ligand isotopically labelled. (ii) Acquisition of NMR experiments to assign the ligand(s), identify the protein methyl group resonances and derive the intra-ligand and protein–ligand inter-molecular distance restraints. (iii) $N$MR$^2$ structure calculation protocol.[8,9] While this is the standard $N$MR$^2$ process, the method is versatile and can be tailored to the investigated system. $N$MR$^2$ overcomes the barrier of the time-consuming protein assignment step and only requires the interpretation of simple NMR spectra. The method is applicable from hit validation to hit to lead stages, and was recently shown to be suitable for the determination of the structures of protein–ligand complexes with both strong and weak binders.[8,9] However, fragments often contain relatively few observable protons, and thus few intermolecular NOEs can be measured. The small number of intermolecular distance restraints available to derive the structure of a protein–fragment complex is a strong limitation for NMR and in particular $N$MR$^2$ since the inter-molecular distance restraints are not assigned on the protein side and therefore remain ambiguous. Here we show how to optimize the $N$MR$^2$ approach to derive the structure of a set of fragments in complex with their target, the prolyl isomerase PIN1.

PIN1 is a peptidyl-prolyl *cis/trans* isomerase that recognizes phospho-serine/threonine-proline motifs, and a critical modifier of multiple signalling pathways. It is

[a] *Laboratory of Physical Chemistry, ETH, Swiss Federal Institute of Technology, HCI F217, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland.*
*E-mail: julien.orts@phys.chem.ethz.ch*
[b] *Vernalis, Granta Park, Cambridge, UK*

† Electronic supplementary information (ESI) available: ESI Fig. S1–S8, and Table S1 and S2. See DOI:10.1039/d0md00068j
‡ New address: Department of Medical Biochemistry and Microbiology, bmc, Uppsala University, Sweden.

overexpressed in several cancers and its activity contributes to tumour initiation and growth.[12,13] Several studies reported inhibitors of PIN1 but no drug has yet reached the market.[14,15] One of the first developed inhibitor, juglone, did not lead to a drug candidate due to the lack of selectivity.[16] A large number of phenylimidazole fragments were previously identified as binding to PIN1.[17]

## Results and discussion

Compounds **1**, **2** and **3** targeting PIN1 share a common phenylimidazole scaffold substituted with the isostere groups $CH_3$, Cl, and $CF_3$, respectively, Fig. 1.

PIN1 was expressed as doubly [$^{15}N$,$^{13}C$]-labelled protein in order to discriminate between the NMR signal of the protons from the protein and the protons from the fragments. Three NMR samples containing 1.3–1.5 mM PIN1 and 3 mM ligands were prepared for subsequent NMR measurements (see Experimental section). The protein methyl group resonances were identified by collecting $^{13}C$-ctHSQC spectra

for each complex (see below). The fragment assignments were straightforward using 1D $^1H$ NMR spectra in the free form and in complex with the PIN1, as well as the $F_1$-[$^{15}N$,$^{13}C$]-filtered [$^1H$,$^1H$]-NOESY experiments (see below), Table S1.†

A series of $F_1$-[$^{15}N$,$^{13}C$]-filtered [$^1H$,$^1H$]-NOESY spectra were recorded on a 900 MHz spectrometer for each PIN1–fragment complex.[18] A total of 17, 14, and 18 inter-molecular NOEs, could be measured for the fragment **1**, **2**, and **3** respectively. Build-up curves of poor quality or showing quadratic behaviours characteristic of spin diffusion, were discarded. The cross-relaxation rates derived from the NOE build-up curves were converted to distances using the effective correlation time of the complexes. The effective correlation times, defined from the population averaged correlation times of the free and bound fragments, were derived from the apo- and holo-populations calculated from affinity measurements and when possible, from the sterically known distances from the fragments (Table S2†).[19]
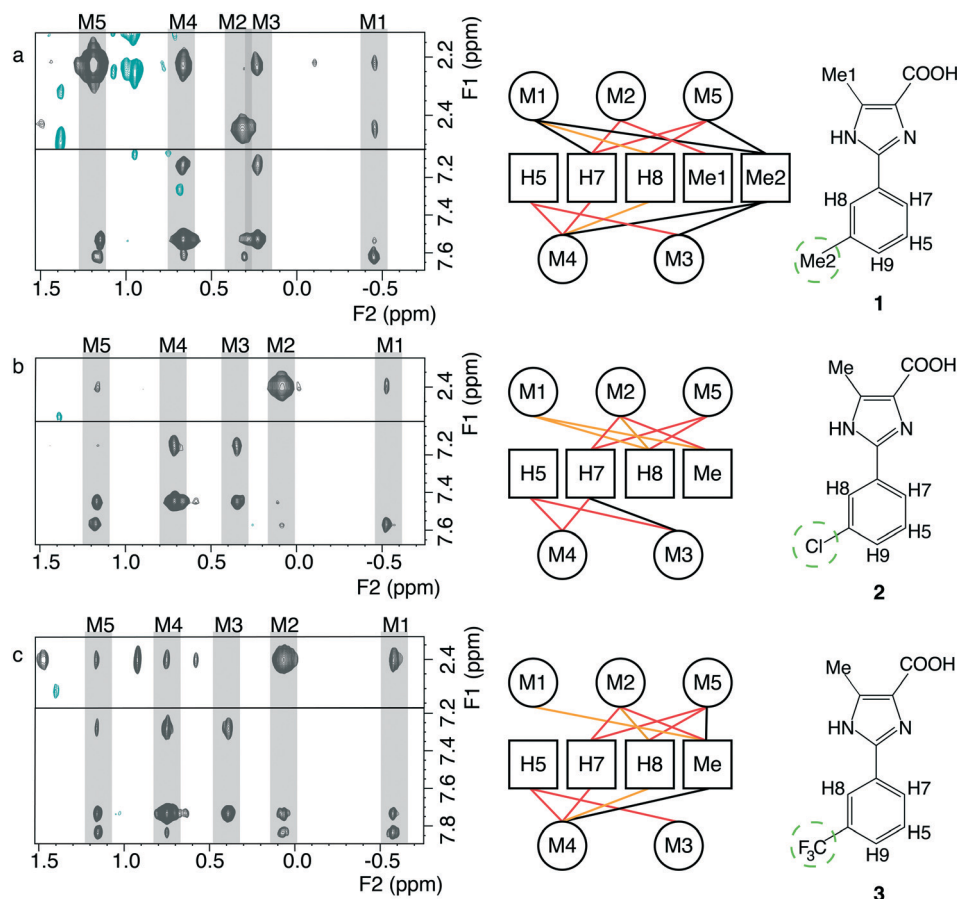


**Fig. 1** Overview of the NMR restraints for the compounds **1**, **2** and **3**, displayed in panel a–c respectively. Left) Intermolecular cross-peaks from F1-[$^{15}N$,$^{13}C$]-filtered [$^1H$,$^1H$]-NOESY spectra of PIN1 in complex with the corresponding fragments. The protein methyl groups involved in inter-molecular NOE(s) are arbitrarily named M1 to M5 and the ligand resonance assignments are reported in Table S1.† Middle) Intermolecular distance restraint network between the assigned fragment's protons and the unknown protein's methyl groups. The red lines show the restraints that are shared between all the fragments, the orange lines represent restraints that are shared between two fragments, and the black lines represent restraints that are specific to the considered fragment. Right) Structures of the fragments. The isosteric groups are emphasized by the green dashed circles.

The binding affinity constants were derived using protein methyl group chemical shift perturbations upon fragment titration and found to be $K_{D,1}$ = 260 μM, $K_{D,2}$ = 670 μM, and $K_{D,3}$ = 6700 μM (Table S2, Fig. S1–S3†). A set of 14, 11, and 12 distances for the fragment **1**, **2**, and **3** respectively were retained for the $N$MR$^2$ structure calculations (see ESI† Fig. S4). The fragment conformations were found to be planar in agreement with the intra-NOE derived distances measured in their bound states (see Experimental section), as well as *ab initio* calculations using the software Gaussian. The $N$MR$^2$ calculations were performed at the known catalytic site of PIN1 (Fig. S5†) using the experimentally derived distances restraints, the derived ligand structures and a starting structure of the protein arbitrarily taken from a previously determined crystal structure.[17] Each side chains dihedral angle of PIN1 could rotate by 20° providing a large degree of flexibility to accommodate the ligand. Additional useful information for the structure determination could be easily determined from the recorded set of NMR spectra. One protein methyl group assignment was readily derived from the $^{13}$C-ctHSQC, and used in the $N$MR$^2$ calculations, since the methionine resonance peaks are negative and only one methionine, M130, is present in the binding site (Fig. S5 and S6†). The alignment of the F$_1$-[$^{15}$N,$^{13}$C]-filtered [$^1$H,$^1$H]-NOESY spectra with the $^{13}$C-ctHSQC spectrum of each complex enables the identification of which NOE restraints involve the M130 methyl group and the protons of the fragments (Fig. S6 and S7†).

Initially, the $N$MR$^2$ structure determination for two out of three PIN1–fragment complexes involving compounds **2** and **3**, failed to converge. The 10 best structures exhibited null target functions expressed as the sum of the squares of residuals, suggesting that the calculations are underdetermined and that several complex structures fulfilled equally well the NOE restraints. On the other hand, fragment **1**, for which 14 inter-molecular distances could be determined, converged and therefore also provided the protein methyl assignments involved in the inter-molecular NOEs. These results suggest that the poor convergence of the structure calculation was due to the paucity of restraints, 11 and 12 inter-molecular distance restraints for the complex involving fragment **2** and **3**, respectively. We therefore investigated the complex PIN1–**2** *in silico* with the aim to find what would be the minimal amount of restraints sufficient to calculate an $N$MR$^2$ PIN1–fragment structure. The distance restraints were taken from the X-ray crystallography structure of the complex (PDB 2XP6) with the visualization program Chimera and randomized by 20% with white noise. The number of restraints used in $N$MR$^2$ was decreased incrementally by removing large distances first. We observed that the true positive structures among the 10 best structures, exhibiting the lowest target function, systematically decreased to reach random selection when only 13 distances were used, suggesting that the complex PIN1–**2** cannot be derived with fewer than 13 distance restraints.

It is possible that $N$MR$^2$ encounters convergence problem with small fragments due to the lack of sufficient protons

and their reduced size. Fragments have a low molecular weight and contain generally few protons. Consequently, the number of protein–ligand inter-molecular NOEs is reduced as well as their interaction surface. However, we found that it was possible to determine structures of the two complexes which did not converge using the methyl assignments from the successfully completed PIN1–**1** $N$MR$^2$ calculation. The alignment of the NOE patterns from the F$_1$-[$^{15}$N,$^{13}$C]-filtered [$^1$H,$^1$H]-NOESY spectra enable the transfer of protein methyl assignments from the $N$MR$^2$ derived PIN1–**1** complex structure to the distance restraints of the other fragments whose complex structures could not be derived. The transfer of assignments is greatly facilitated by following the chemical shift perturbations of the PIN1 methyl groups during the fragment titration, previously recorded to determine their binding affinities.

The newly calculated and converged structures are overlaid with the X-ray crystal structure of PIN1–**2** on Fig. 2, and exhibit similar poses to the crystallographic reference with RMSDs of 1.1 Å, 2.5 Å, and 1.4 Å for the fragments **1**, **2**, and **3** respectively. The relatively high RMSD for **2** is due to a slight translation of the fragment toward the outside of the binding pocket. The high similarity in the fragment binding modes to PIN1 could be anticipated since a common NOE pattern can be identified for the three fragments suggesting beforehand a similar binding pose, Fig. 1. Furthermore, the fragments **1**, **2** and **3** are structurally very similar and only differ by an isostere substitution, therefore the binding mode is expected to remain similar. The substituted phenyl group makes hydrophobic interactions with L122, M130, F134, T152, while the embedded histidines, H59 and H157, are candidates for possible π–π or cation–π interactions. The carboxylic acid is involved in salt bridges with the cationic pocket formed by R68, R69 and K63. The imidazole moiety makes hydrogen bonds to S154 and C113, while the trifluoromethyl groups may engage in hydrogen bonding to Q131 and T152.[20] However steric factors may be the primary reason for the decrease of affinity with CF$_3$ ($K_D$ = 6700 μM, Van der Waals radius ∼ 2.5 Å), compared to CH$_3$ ($K_D$ = 260 μM, radius ∼ 2.0 Å) and Cl ($K_D$ = 670 μM, radius ∼ 1.8 Å).[21] Overall the three $N$MR$^2$ structures exhibit similar binding modes and are consistent with the reported structure derived by X-ray crystallography.

## Conclusion

Fragment-based drug discovery is a widely used approach in both pharmaceutical companies and academic laboratories. Fragment-based methods need significantly fewer compounds to be screened, synthesized and the fragment hits show usually high ligand efficiencies. Yet fragments contain usually only few protons, rendering the structure elucidation by NMR more difficult. We have shown that the cooperativity between several $N$MR$^2$ datasets can overcome the experimental ambiguity due the lack of distance restraints that would otherwise prevent structure elucidation
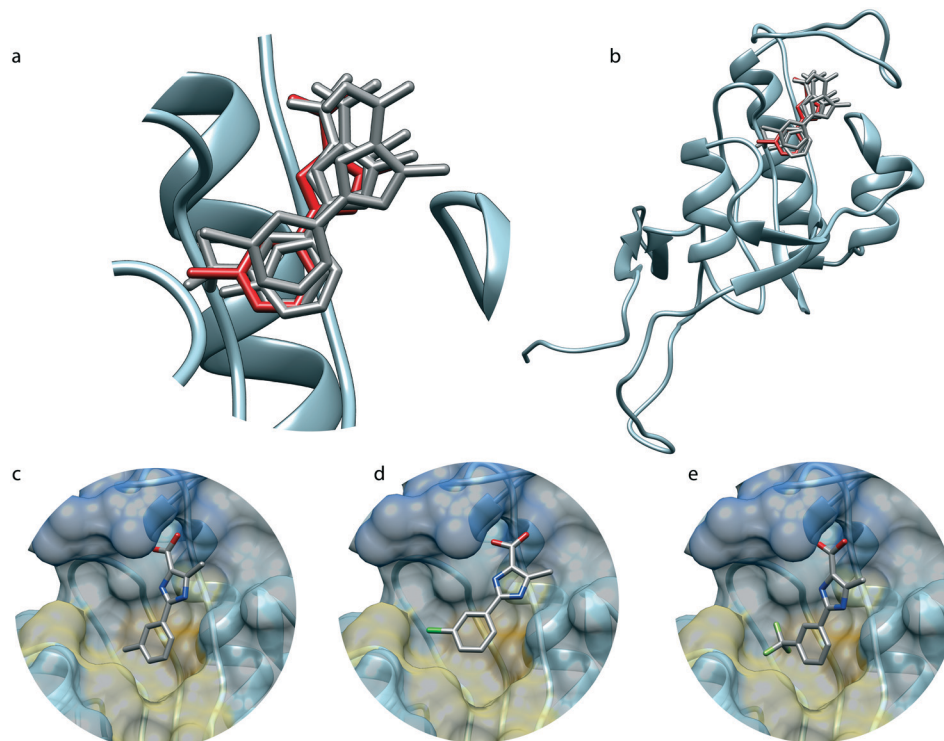
**Fig. 2** Structure of the PIN1–fragment complexes derived by *N*MR[2]. a and b, Overlap of the *N*MR[2] structures, depicted in grey, with the X-ray structure of PIN1 in complex with **2** (PDB 2XP6), depicted in red sticks and blue ribbon. c–e Complex binding-site structures of **1**, **2** and **3**, respectively, with PIN1 derived by *N*MR[2]. The dark blue surface represents a positively charged region (R68, R69, K63), the yellow colour shows a hydrophobic region (M130, F134, L122), and the gold part, located between the yellow and the blue regions, corresponds to two embedded histidines (H157, H59).

and consequently the corresponding structure-based drug design. The *N*MR[2] structure calculations only took a couple of minutes of computer time and only a few hours were needed to process and analyse the NMR data so that the complete process perfectly matches the time line required for medicinal chemistry. The time-consuming part of the structure elucidation is no longer the analysis of the NMR spectra series, such as the protein resonances assignment, but rather on the acquisition of NMR experiments. The presented *N*MR[2] approach may open a new avenue in NMR drug design by enabling fast and robust fragment structure-based drug discovery. Having access to the structure of the binding site for each (weak) binder allows investigating chemical scaffolds that would otherwise be discarded, which broadens the chemical knowledge as well as enables the druggability of the protein targets.

## Experimental section

### Protein expression, compounds and NMR samples

[$^{15}$N,$^{13}$C]-Labelled recombinant PIN1 was produced as previously reported.[22] PIN1 was buffer exchanged to 100% D$_2$O, phosphate 20 mM, sodium chloride 50 mM at pH 6.6 and concentrated to 1.3 mM for **1** and **3**, and to 1.5 mM for **2**. The ligand concentration in the samples used for the NOESY experiments were 2.5, 3, and 3.5 mM for **1**, **2**, and **3**,

respectively. Compounds **1**, **2**, and **3** were ordered from Enamine Ltd.

### NMR experiments

All NMR experiments were recorded at 298 K. F$_1$-[$^{13}$C,$^{15}$N]-Filtered [$^1$H,$^1$H]-NOESY spectra were measured with [$^{13}$C,$^{15}$N]-labelled PIN1 and unlabelled ligands on a Bruker Avance 900 MHz spectrometer with cryoprobe, with 2048 complex points in the direct time dimension ($t_2$) and 241, 256, and 225 complex points in the indirect dimension ($t_1$) for **1**, **2**, and **3** respectively. The interscan delay was set to 1.5 s, the scans per increment was 160, $t_{2,max}$ = 190.05 ms, and $t_{1,max}$ = 22.44 ms for **1**, 23.75 ms for **2**, and 20.87 ms for **3**. The mixing times used for the NOE build-ups of **1** were 30, 70 ms, **2**: 40, 60, 90, 120 ms, and **3**: 30, 60, 90 ms. Constant time $^{13}$C-HSQC ($^{13}$C-ctHSQC) for titration, $T_{1\rho}$ and $T_1$ $^{15}$N-HSQC for relaxation experiment and F$_1$-[$^{13}$C,$^{15}$N]-filtered [$^1$H,$^1$H]-TOCSY for fragment proton assignments were recorded on 600 MHz Bruker Avance spectrometer. PIN1 $^{15}$N-$T_1$ and $^{15}$N-$T_{1\rho}$ experiments were recorded at 0, 20, 40, 60, 80, 100, 120, 140, 180, 200 ms and 5, 10, 20, 30, 40, 50, 70, 90, 100, 120 ms relaxation delays, respectively.[23] 64 ($t_1$) × 2048 ($t_2$) complex points were used with an interscan delay of 3 s (2 s for $T_1$), 32 scans per increment, $t_{1,max}$ = 138.35 ms, and $t_{2,max}$ = 121.65 ms (243 ms for $T_1$). Spectra were processed with NMRPipe.[24]

For the assignment of the fragments, filtered $F_1$-[$^{13}$C,$^{15}$N]-filtered [$^1$H,$^1$H]-TOCSY spectra were recorded with $1024(t_1) \times 2048(t_2)$ complex points, an interscan delay of 1.2 s, 128 scans per increment, $t_{1,max} = 13.30$ ms, and $t_{2,max} = 121.65$ ms. For the titration of **2** and **3**, $^{13}$C-ctHSQC spectra were recorded with $256(t_1) \times 2048(t_2)$ complex points, an interscan delay of 1.2 s, 8 scans per increment, a $t_{1,max} = 71.08$ ms, and a $t_{2,max} = 141.99$ ms. The fragment concentrations used for the titration were 60, 156, 252, 348, 444, 540, 636, 732, 828, 924, 1020, 1116, 1308, 1444 µM for **2** and 296, 592, 988, 1183, 1479, 1775, 2071, 2367, 2663, 2959, 3255, 3551, 3847, 4143, 4439, 4735, 5024, 5320, 5616 µM for **3**.

### Data analysis and structure calculations

The effective correlation time of each complex was derived using the population averaged correlation times with $\tau_{c,PIN1} = 8.9$ ns and $\tau_{c,fragments} = 0.1$ ns. The PIN1 correlation time was calculated using the software Tensor2 and the relaxation rates of the amide groups.[25] The effective correlation time of each complex was validated using sterically known distances and ligand populations derived from binding affinity constants (see ESI†) and subsequently used to derived the NOE distance restraints. The spectra were processed with topspin 3.1 (Bruker) and analysed with ccpNMR Analysis 2.3.

Distances were computed from NOE build-up curves, assuming an isolated two spin system. First the ligand protons diagonal peaks were fitted to obtain auto-relaxation rates, $\rho_i$, with a mono-exponential decay $\Delta M_{ii}(t) = \Delta M_{ii}(0)e^{-\rho_i t}$. The auto-relaxation rates of the protein methyl groups were assumed to be 4 Hz. The cross-peaks intensities, $\Delta M_{ij}(t)$, were normalized to the ligand protons' diagonal peak intensities, $\Delta M_{ii}(t)$. The cross-relaxation rate, $\sigma_{ij}$, was then computed from these normalized intensities, $\dfrac{\Delta M_{ij}(t)}{\Delta M_{ii}(0)} = -\dfrac{\sigma_{ij}}{(\lambda_+ - \lambda_-)}(e^{-\lambda_+ t} - e^{-\lambda_- t})$, where $\lambda_\pm = \dfrac{\rho_i + \rho_j}{2} \pm \sqrt{\left(\dfrac{\rho_i + \rho_j}{2}\right)^2 + \sigma_{ij}^2}$. The auto-relaxation rates of the ligand's protons were obtained from the fitted decay plots of the diagonal peaks. The distances, $r_{ij}$, could be extracted from the cross-relaxation rates from of the equation, $\sigma_{ij} = \dfrac{1}{2}\dfrac{\mu_0}{4\pi}\dfrac{\hbar\gamma_H^2}{r_{ij}^6}(6J(2\omega) - J(0))$ where $J(\omega) = \dfrac{2}{5}\left(\dfrac{\tau_{c,eff}}{1 + (\omega\tau_{c,eff})^2}\right)$ is the spectral density, $\mu_0$ is the permeability of the vacuum, $\hbar$ the reduced Planck constant, $\gamma_H$ is the proton gyromagnetic ratio, and $\tau_{c,eff}$ is the effective correlation time of the protein–ligand complex (Table S2†).

The $NMR^2$ structure calculations were performed according to the published protocol for the complex PIN1–**1**, and the complexes PIN1–**2** and PIN1–**3** were calculated knowing the protein methyl group assignments.[8,9,11] The methyl group assignments determined for the PIN1–**1** complex were cross-validated using a 3D HCCH-TOCSY experiment (Fig. S8†). The fragment geometries were kept fixed during the calculation and optimized in a two steps process: first a molecular mechanics energy minimization was run with Avogadro (UFF

force field, steepest descent algorithm); then an *ab initio* energy minimization was run with Gaussian (6-31G(d,p) basis set, DFT B3LYP method). Frequency calculations were run afterwards and the hessians were all positives.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## References

1  D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discovery*, 2016, **15**, 605–619.

2  B. J. Davis and S. D. Roughley, in *Annual Reports in Medicinal Chemistry*, ed. R. A. Goodnow, Academic Press, 2017, vol. 50, pp. 371–439.

3  B. Lamoree and R. E. Hubbard, *Essays Biochem.*, 2017, **61**, 453.

4  A. L. Hopkins, C. R. Groom and A. Alex, *Drug Discovery Today*, 2004, **9**, 430–431.

5  M. J. Hartshorn, C. W. Murray, A. Cleasby, M. Frederickson, I. J. Tickle and H. Jhoti, *J. Med. Chem.*, 2005, **48**, 403–413.

6  D. Patel, J. D. Bauman and E. Arnold, *Prog. Biophys. Mol. Biol.*, 2014, **116**, 92–100.

7  D. A. Erlanson, B. J. Davis and W. Jahnke, *Cell Chem. Biol.*, 2019, **26**, 9–15.

8  J. Orts, M. A. Walti, M. Marsh, L. Vera, A. D. Gossert, P. Guntert and R. Riek, *J. Am. Chem. Soc.*, 2016, **138**, 4393–4400.

9  M. A. Walti, R. Riek and J. Orts, *Angew. Chem., Int. Ed.*, 2017, **56**, 5208–5211.

10  F. Torres and J. Orts, *Future Med. Chem.*, 2018, **10**, 2373–2376.

11  M. Wälti and J. Orts, *Magnetochemistry*, 2018, **4**, 12.

12  X. Z. Zhou and K. P. Lu, *Nat. Rev. Cancer*, 2016, **16**, 463.

13  A. Ryo, Y. C. Liou, G. Wulf, M. Nakamura, S. W. Lee and K. P. Lu, *Mol. Cell. Biol.*, 2002, **22**, 5281–5295.

14  A. J. Potter, S. Ray, L. Gueritz, C. L. Nunns, C. J. Bryant, S. F. Scrace, N. Matassova, L. Baker, P. Dokurno and D. A. Robinson, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 586–590.

15  B. M. Dunyak and J. E. Gestwicki, *J. Med. Chem.*, 2016, **59**, 9622–9644.

16  J. D. Moore and A. Potter, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 4283–4291.

17  A. Potter, V. Oldfield, C. Nunns, C. Fromont, S. Ray, C. J. Northfield, C. J. Bryant, S. F. Scrace, D. Robinson and N. Matossova, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 6483–6488.

18  C. Zwahlen, P. Legault, S. J. F. Vincent, J. Greenblatt, R. Konrat and L. E. Kay, *J. Am. Chem. Soc.*, 1997, **119**, 6711–6721.

19 G. M. Lippens, C. Cerf and K. Hallenga, *J. Magn. Reson.*, 1992, **99**, 268–281.

20 C. Dalvit and A. Vulpetti, *Chemistry*, 2016, **22**, 7592–7601.

21 M. Dekker, *The STERIMOL approach to drug design*, Academic Press, New York, 1987.

22 I. Landrieu, J. M. Wieruszeski, R. Wintjens, D. Inze and G. Lippens, *J. Mol. Biol.*, 2002, **320**, 321–332.

23 N. A. Lakomek, J. F. Ying and A. Bax, *J. Biomol. NMR*, 2012, **53**, 209–221.

24 F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277–293.

25 F. Cordier, M. Caffrey, B. Brutscher, M. A. Cusanovich, D. Marion and M. Blackledge, *J. Mol. Biol.*, 1998, **281**, 341–361.