



OPEN

Methyl probes in proteins for determining ligand binding mode in weak protein–ligand complexes

Biswaranjan Mohanty^{1,2,3,7}, Julien Orts^{4,7}✉, Geqing Wang⁵, Stefan Nebel¹, Wesam S. Alwan¹, Bradley C. Doak^{1,2}, Martin L. Williams¹, Begoña Heras⁵, Mehdi Mobli⁶ & Martin J. Scanlon^{1,2}✉

Structures of protein–ligand complexes provide critical information for drug design. Most protein–ligand complex structures are determined using X-ray crystallography, but where crystallography is not able to generate a structure for a complex, NMR is often the best alternative. However, the available tools to enable rapid and robust structure determination of protein–ligand complexes by NMR are currently limited. This leads to situations where projects are either discontinued or pursued without structural data, rendering the task more difficult. We previously reported the NMR Molecular Replacement (*NMR*²) approach that allows the structure of a protein–ligand complex to be determined without requiring the cumbersome task of protein resonance assignment. Herein, we describe the *NMR*² approach to determine the binding pose of a small molecule in a weak protein–ligand complex by collecting sparse protein methyl-to-ligand NOEs from a selectively labeled protein sample and an unlabeled ligand. In the selective labeling scheme all methyl containing residues of the protein are protonated in an otherwise deuterated background. This allows measurement of intermolecular NOEs with greater sensitivity using standard NOESY pulse sequences instead of isotope-filtered NMR experiments. This labelling approach is well suited to the *NMR*² approach and extends its utility to include larger protein–ligand complexes.

Structure-based drug design has proven to be very efficient in the hit to lead optimization process¹. X-ray crystallography is the preferred technique to generate structural data². While crystallization generally works well for high affinity ligands, those that bind weakly can prove more challenging. When the complex does not crystallize, NMR can provide an alternative approach to generate structural data. However, the time required to generate a structure by NMR is often too long for medicinal chemistry projects. In recent years, numerous NMR methods have been developed that report on the binding mode of the ligand³. Chemical shift perturbation (CSP)^{4–7}, saturation transfer⁸, partial NMR resonance assignment⁹, and pseudocontact shift approaches^{10,11} are a few methods addressing this issue^{12–14}. We previously reported a method to derive the structure of a protein–ligand binding site called NMR Molecular Replacement, *NMR*². This method can generate accurate structures of a protein–ligand binding site within a few days^{15–17}. Neither backbone nor sidechain assignments are required for *NMR*². Instead, this method uses semi-ambiguous protein–ligand inter-molecular NOEs and intra-ligand NOEs as experimental inputs in conjunction with a starting model of the protein to calculate the bound structure. The structure calculations are based on standard NMR protocols, and the outputs are ranked according to the agreement between the experimental input and the calculated structure. *NMR*² is a systematic method and consequently does not contain any stochastic component. It also does not rely on docking recipes but on a series of robust structure calculation protocols. *NMR*² can be used to solve several different types of complex structures involving strong and weak binders, small molecules, peptides and peptidomimetics^{15–17}. The

¹Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, VIC 3052, Australia. ²ARC Centre for Fragment-Based Design, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, VIC 3052, Australia. ³Sydney Analytical Core Research Facility, The University of Sydney, Sydney, NSW 2006, Australia. ⁴Department of Pharmaceutical Sciences, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. ⁵La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC 3085, Australia. ⁶Centre for Advanced Imaging, The University of Queensland, St Lucia, QLD 4072, Australia. ⁷These authors contributed equally: Biswaranjan Mohanty and Julien Orts. ✉email: julien.orts@univie.ac.at; martin.scanlon@monash.edu

method is also capable of detecting cryptic binding pockets^{15–17}. Herein we detail the use of NMR² in conjunction with specific methyl labelling schemes in a deuterated background (see Supplementary Fig. S1). The advantage of specific methyl labelling schemes in protein NMR, where the methyl groups are protonated in an otherwise deuterated background, allow much higher molecular weight systems to be studied due to the favourable relaxation properties of methyl groups^{18,19}. Additionally, stereospecific labelling of methyl groups is possible, which reduces the ambiguity of protein resonance assignments.

Here, we exemplify the potential of combining the NMR² method with a selective methyl labelling technique on a challenging system comprising phenylthiazole **1**, a flexible low-affinity ligand binding to a shallow and dynamic site on the protein. The model protein used in this study is a 21 kDa DsbA enzyme from *Escherichia coli* (EcDsbA) that catalyzes the formation of disulfide bonds in secreted proteins within the periplasm. EcDsbA is essential for bacterial virulence and therefore represents an attractive target for the development of new antibacterials²⁰.

Results and discussion

We generated structural models for the ligand bound to oxidized EcDsbA using four different approaches: X-ray crystallography, NMR², NMR data driven docking with HADDOCK⁴, and classical NMR structure calculation using CYANA. All four approaches found similar binding poses with small variation in the positioning of phenylthiazole **1** (*vide infra*). EcDsbA was selectively labelled with protonation on the following methyl groups: A β 1 δ 1(LV)^{PRO}R^ME^TY² while the remaining protons were deuterated^{21,22}. Conventional NOESY experiments from [U-²H]-A β 1 δ 1(LV)^{PRO}R^ME^TY²-CH₃ labelled EcDsbA and phenylthiazole **1** may provide ambiguous intermolecular NOE cross-peaks in the amide region due to peak overlap between amide resonances of the protein and aromatic resonances of the ligand, therefore, the NMR experiments were performed in D₂O NMR buffers (see Supplementary Fig. S2 and S3). Labelling all methyl-containing residues provides greater coverage of the protein surface than selective methyl labelling approaches^{23,24}. Moreover, the labelling pattern of the ILV methyl groups in the current scheme reduces the complexity of the NMR spectra because only one methyl group is labelled in each of these residue types, thereby providing unambiguous and stereospecific information (Fig. 1). In addition to reducing peak overlap, spin diffusion is also considerably reduced through this labelling scheme due to the lower proton density. Spin diffusion is a major limitation when measuring direct ¹H-¹H magnetization transfer as it dominates the NOE cross peak intensity at high mixing times. The direct NOE transfer can be converted into a meaningful distance, but the spin diffusion introduces uncertainty into the distance restraints.

Replacing most of the protons in the protein with deuterons improves signal to noise and reduces spin-diffusion, which preserves the quality of the distance information contained in the NOE cross peaks. The NOE cross peaks are shown in Figure 2, and Supplementary Table S1 contains the corresponding estimated NOE distances observed in a sample containing 0.25 mM EcDsbA and 3.5 mM phenylthiazole **1**. Measurement of ligand solubility revealed that phenylthiazole **1** was soluble and free from aggregation at this concentration²⁵ (Supplementary Fig. S4). The ligand affinity (K_D) was estimated at 0.9 mM based on chemical shift perturbations measured in HSQC spectra recorded with increasing ligand concentration. Consequently, 79% of EcDsbA was in the ligand-bound state under the experimental conditions²⁶ (Supplementary Fig. S5). For the HADDOCK protocol as well as for the classical NMR structure calculation, protein assignments were obtained using standard NMR experiments.

An X-ray crystal structure of the complex was determined by soaking crystals of EcDsbA with phenylthiazole **1** and refined to 2.0 Å of resolution (Fig. 3; see Supplementary Table S2). The crystals contained two protomers in the asymmetric unit and, as illustrated in Fig. 3, partial electron density was observed for only one phenylthiazole **1** molecule, which was bound to the hydrophobic groove of chain A, but with some additional contacts being made between the ligand and residues of chain B. The structures derived from NMR² were calculated using the intermolecular protein–ligand NOE list without protein assignments (see Supplementary Table S1).

NOESY data were acquired with two mixing times to derive the distance restraints from the initial rates of the NOE build-up curves. NOE build-up curves that did not exhibit an initial linear behavior were not used in the analysis (see Supplementary Fig. S6). Two mixing times are the minimum required since these provide the three data points (two measured data points and the zero-time point) that are needed to check the initial linearity of the build-up curves. The diagonal peak decays were also evaluated because the initial magnetization and the auto-relaxation rates are needed to extract accurate distances²⁷. The intermolecular NOEs were corrected for the protein occupancy since the ligand is a weak binder in fast exchange compared to the chemical shift and relaxation time scales. The distances obtained in this way are more accurate compared to traditional NOE-derived distances but are called semi-ambiguous due to the lack of protein assignment²⁸. Nonetheless, the restraints are used differently compared to the ambiguous restraints previously defined in the literature^{29,30}. The standard ambiguous restraint allows multiple protons, arising from two or more resonances, be used simultaneously to fulfil one NOE. In the NMR² protocol, this is forbidden, and each NOE restraint must explicitly relate to two groups of protons, corresponding to two NMR resonances. This analysis provides more explicit distance information compared to the classical ambiguous restraints but requires that NOEs involving overlapping peaks cannot be used in the NMR² calculation. To minimize the possibility of peak overlap a high-resolution constant time 2D [¹³C,¹H]-HSQC spectrum was recorded. In combination with our labeling scheme this distinguishes between different residue types, and in this case excellent resolution was observed in the 3D carbon resolved NOESY-HMQC spectra (Supplementary Fig. S2, and Fig. 2). Our labelling scheme also avoids the use of isotope-filtered NMR experiments. This affords spectra with greater signal to noise due to the increased sensitivity of standard NOESY pulse sequences. To further increase resolution and sensitivity of intermolecular cross peaks in the indirect NOESY dimension, non-uniform sampling acquisition schemes can be used (see Supplementary Fig. S7). In addition to the protein–ligand NOEs, the non-assigned protein intra methyl–methyl distances derived

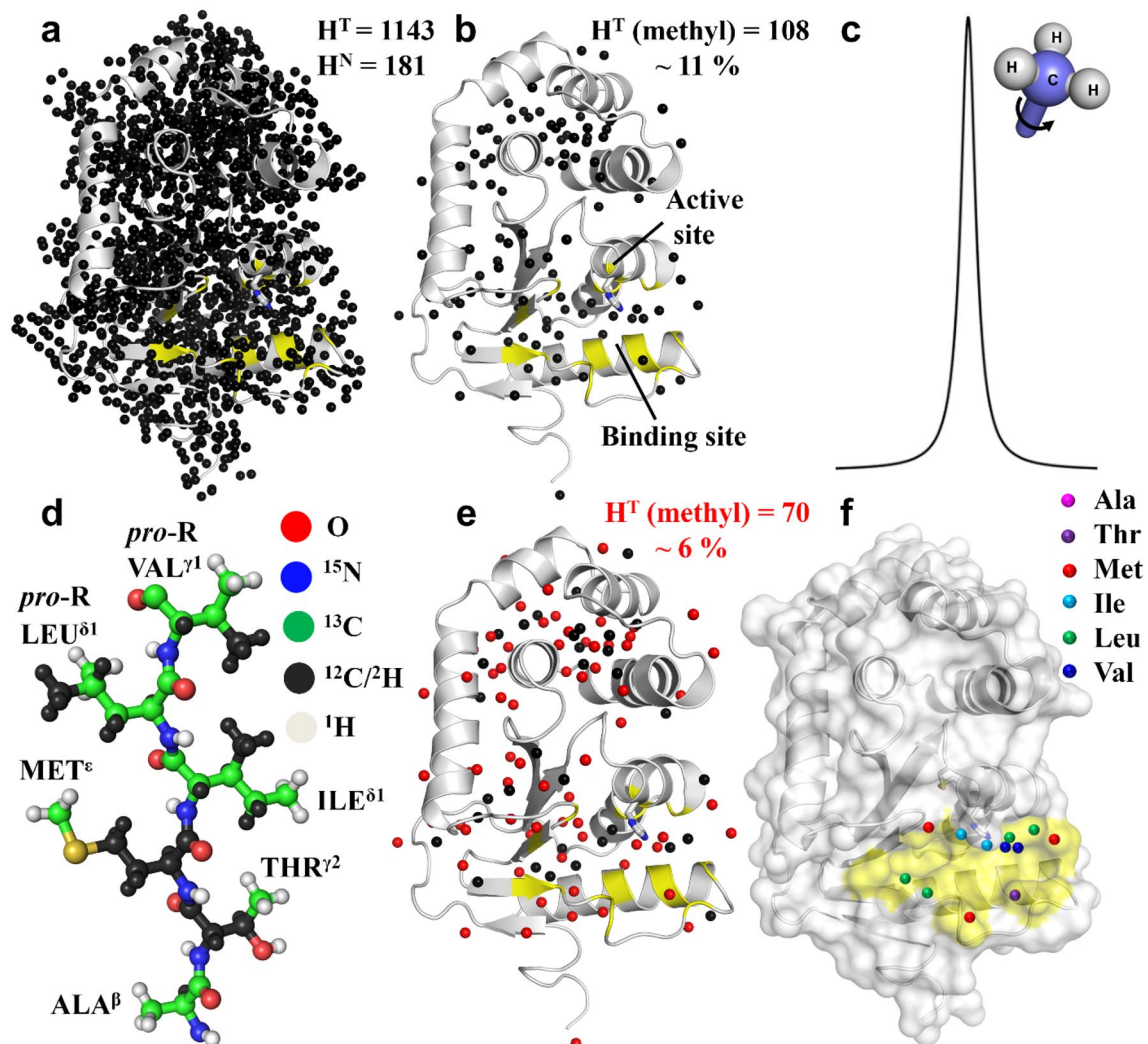


Figure 1. Labelling strategy used in the study. Schematic representation of the proton density in the structure of the bacterial oxidoreductase enzyme *EcDsbA*, showing the position of amide (H^N) and total (H^T) protons (a) when fully protonated, (b) when fully deuterated except for methyl groups, which represent 11% of all protons. (c) Schematic showing that the fast rotation of methyl groups results in narrow linewidths in NMR spectra. (d,e) Labelling scheme used in the triple labelled sample where only one methyl group is protonated in Ile, Leu and Val, decreasing the proton density to 6% versus the fully protonated protein. (e,f) Residues that form the ligand binding site in *EcDsbA* are colored yellow and the positions of methyl-containing residues adjacent to the binding site are highlighted. H^T indicates total number of proton resonances expected in the NMR spectrum of oxidized *EcDsbA* (PDB ID: 1FVK). Total number of expected proton resonances was calculated assuming methyl/methylene/symmetric aromatic protons had the same chemical shift. Analysis was done by MOLMOL.

from the 3D ^{13}C -resolved NOESY-HMQC can also be used by NMR^2 since the structure calculation protocol does not require the methyl assignments (see Supplementary Table S1).

For the NMR^2 calculations, the bound conformation of the ligand was calculated using tr-NOEs derived from short 2D ω_1 , ω_2 - ^{13}C , ^{15}N -filtered [^1H , ^1H]-NOESY spectra (data not shown)^{3,31–33}. Using the top 20 bound conformations of the ligand, an input structure of the protein (PDB code 1FVK), the set of semi-ambiguous intermolecular distances, and the unassigned protein methyl-methyl distances, NMR^2 was employed to determine the structure of phenylthiazole **1** in the binding site of *EcDsbA*. Figure 4a shows the best and the second-best NMR^2 models.

NMR^2 provides both the structure of the complex and a partial assignment for the methyl groups. The assignments produced in this way may contain errors since NMR^2 is designed to calculate the structures that best satisfy the experimental distance constraints. In cases where two methyl groups are near to one another, or two methyl groups are equidistant from a ligand proton, it is possible to swap the methyl assignments without impacting the structure. For example, in the structure of *EcDsbA* bound to phenylthiazole **1**, the methyl protons of Met153 and Met166 are located at similar distances from the ligand. In this case their assignments can be swapped. This may account for the variation in the positioning of phenyl carboxylic acid group between two NMR^2 models (Fig. 4a). However, the assignments for the other methyl resonances that were used in the NMR^2

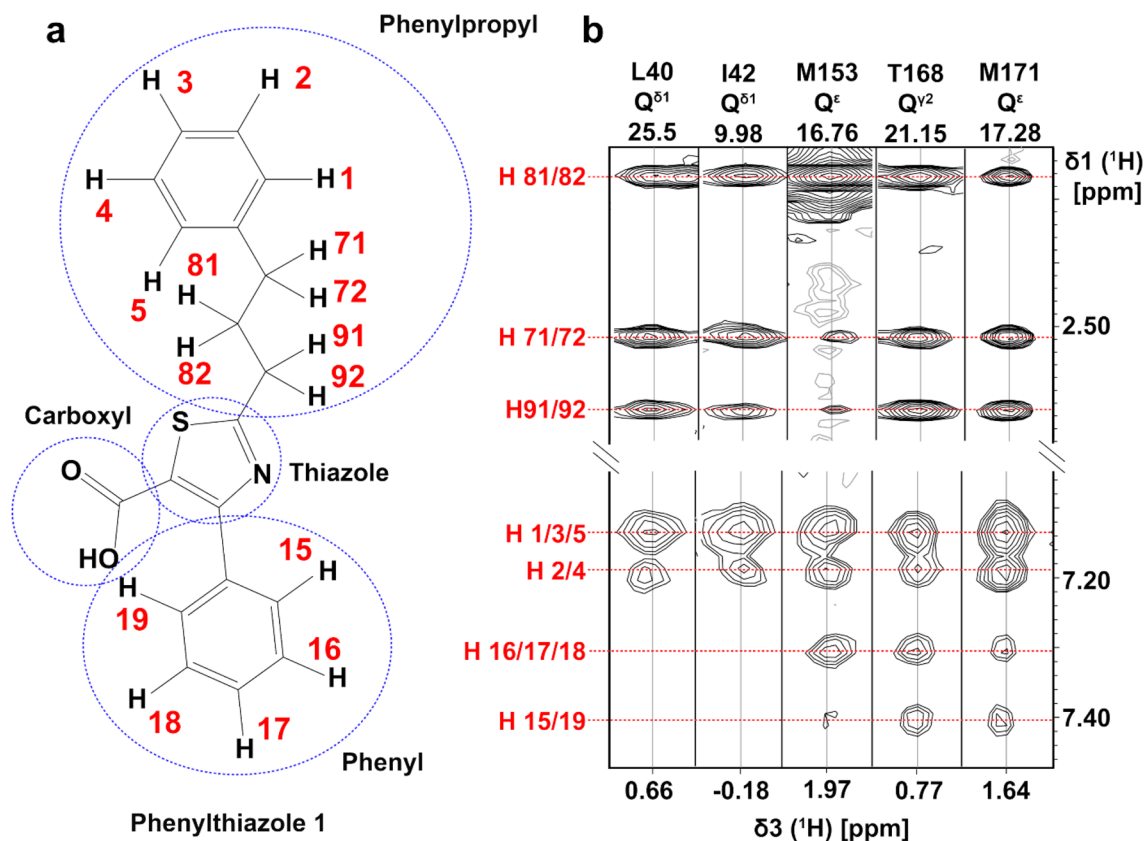


Figure 2. Intermolecular NOEs between oxidized EcDsbA and phenylthiazole 1. (a) Chemical structure of phenylthiazole 1. (b) ^1H - ^1H strips from 3D ^{13}C -edited ^1H - ^1H -NOESY HMQC spectrum of 0.25 mM EcDsbA with 3.5 mM phenylthiazole 1 recorded at 298 K with an NOE mixing time of 400 ms.

calculations were consistent with assignments obtained from conventional methods. Eleven high-quality NOE build-up curves were observed from the 3-phenylpropyl moiety to Leu40 $^{\delta 1}$, Ile42 $^{\delta 1}$, Thr168 $^{\gamma 2}$ and Met171 $^{\epsilon}$ and one NOE was observed between Met153 $^{\epsilon}$ and the second phenyl moiety of phenylthiazole 1. A comparison of intermolecular ^1H - ^1H distances derived from the crystal structure and NOE data is shown in Supplementary Table S3. Structure calculations were also performed using two classical NMR-based approaches: HADDOCK and CYANA, where the protein resonance assignments were utilized. The classical structure calculation was performed with CYANA using the NOE constraints from Supplementary Table S1 with known methyl assignments. A model with lowest target function (TF) from the classical CYANA structure calculation is shown in Fig. 4b. The binding mode of phenylthiazole 1 from CYANA is similar to the best NMR 2 model despite one difference in the methyl assignments. This discrepancy is not structurally relevant because the NOE derived distance involved in two different assignments (M5—Gln59 corresponds to Met153/M166-Q $^{\epsilon}$ —Gln59 in Supplementary Table S1) is large and cannot discriminate between the two 34 . Furthermore, both NOE-derived distance restraints and CSP were used to generate constraints in calculating the structure of the complex using HADDOCK (see Supplementary Table S3). An ensemble of the top five lowest energy HADDOCK models is shown in Supplementary Fig. S8. A unique binding mode of phenylthiazole 1 was observed in all HADDOCK models which is similar to the second best NMR 2 model. The best HADDOCK model (with the lowest HADDOCK score among the five models) is shown in Fig. 4c.

The four structural models derived from X-ray crystallography and NMR calculations all exhibit binding of phenylthiazole 1 to the same binding site on a shallow groove of oxidized EcDsbA (Fig. 4a–c). In each case the 3-phenylpropyl moiety is located between the two α -helices that define the hydrophobic groove and close to Met171 while the second phenyl group is either pointing away from the protein, where it is more solvent exposed (Fig. 4b) or towards the protein (Fig. 4c). The 3-phenylpropyl moiety in all structures makes hydrophobic contacts to Leu40, Met171 and Ile42, while the second phenyl ring orients towards either Met166 (Fig. 4b) or Met153 (Fig. 4c). The thiazole ring also shares a similar location in the binding site but with different orientations of the carboxylate group. The X-ray structure, the best NMR 2 and the CYANA models have the carboxylate oriented towards the protein, whereas the second best NMR 2 , and the HADDOCK models orient the carboxylate towards the solvent. In addition, where the carboxylic acid group is oriented towards the protein, it binds deeper into the binding site in the NMR models compared to the X-ray structure. This discrepancy is difficult to interpret due in part to the fact that only partial electron density was observed for phenylthiazole 1 in the crystal structure (Fig. 3). We hypothesize that the binding mode in which the carboxylate is buried and the phenyl ring points towards the solvent is less likely in solution (Fig. 4b) because no intermolecular NOEs were observed between

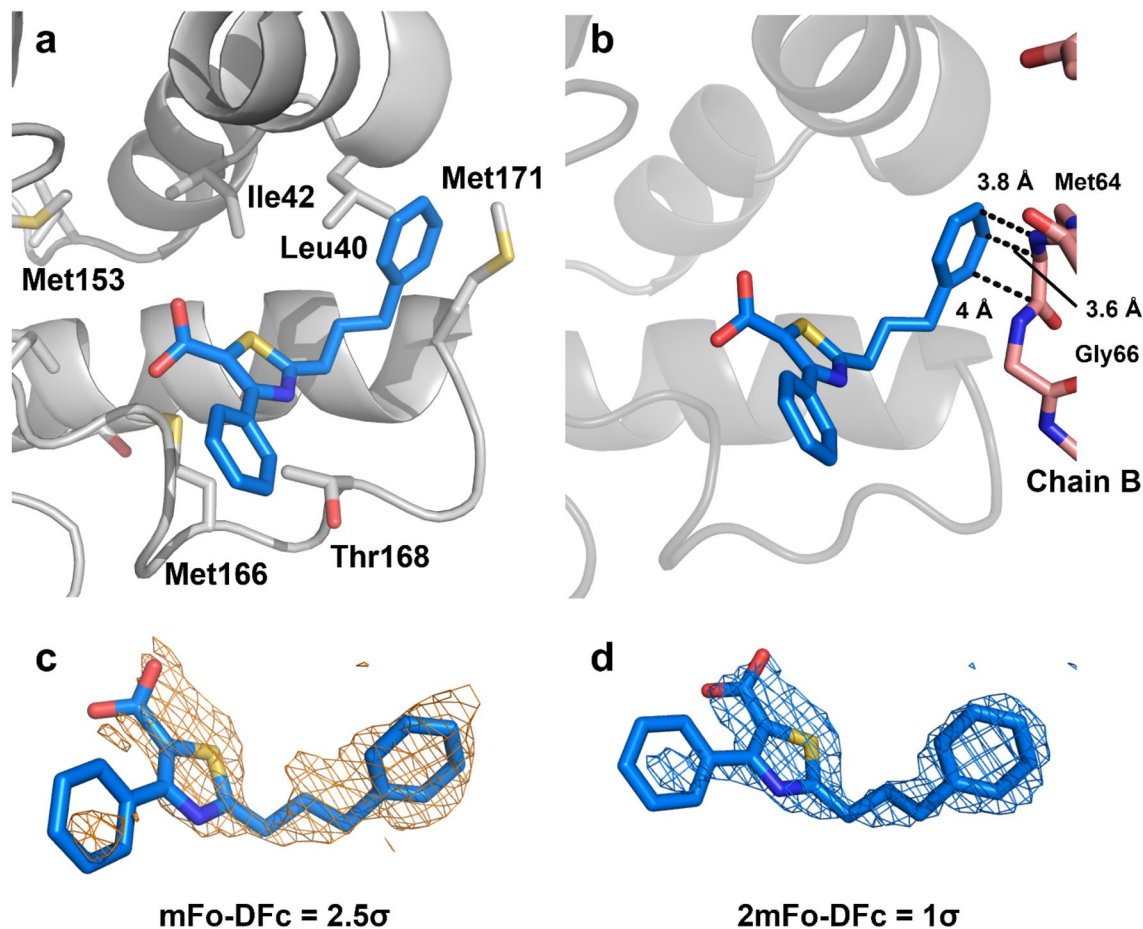


Figure 3. X-ray structure of *EcDsbA*–phenylthiazole **1** complex at 2.0 Å resolution. **(a)** Ligand is shown in blue sticks and the protein is in grey cartoon. Several side chain residues of *EcDsbA* at the binding site are shown in sticks. Only the binding site region is shown for clarity. **(b)** Crystal contacts between the heavy atoms of Chain B (depicted in salmon sticks) and the ligand are shown as dotted black lines. **(c)** Simulated annealing omit σ_A -weighted mFo-DFc electron density map of phenylthiazole **1** contoured at 2.5 σ and displayed in yellow mesh. **(d)** σ_A -weighted 2mFo-DFc electron density map of phenylthiazole **1** contoured at 1 σ and displayed in blue mesh.

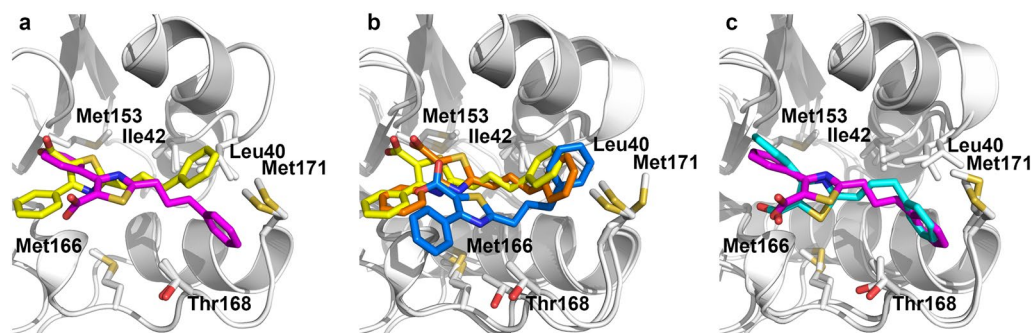


Figure 4. *EcDsbA*–phenylthiazole **1** complex structures derived by NMR², X-ray crystallography and classical NMR structure calculation. **(a)** Overlay of best and second-best NMR² models depicted in yellow and magenta, respectively. **(b)** Comparison of best NMR² model against classical CYANA and crystal structures, depicted in yellow, orange, and blue respectively. **(c)** Comparison of the second-best pose of NMR² against the best HADDOCK model, depicted in cyan. All models were globally aligned to the crystal structure in PyMOL. Only the binding site of all models is shown for clarity.

phenylthiazole 1-H15/H19 and Thr168-Q², which is what would be expected from the orientation observed in the X-ray crystallography structure (see Supplementary Table S3). In addition, several crystal contacts between the ligand and a neighbouring protein are observed and could explain the discrepancies between the NMR and X-ray structures (Fig. 3b). While the classical NMR models are similar to the NMR² structures, the time taken to generate models of the complex structure is orders of magnitude greater due to the requirement that sequence specific assignment of the resonances must be obtained. In essence, the NMR², classical CYANA, HADDOCK and X-ray structures are all reporting binding at the same position, which was proposed to be the substrate-binding site of EcDsbA³⁵. In each case the model indicates that the phenylthiazole core binds in a similar orientation along the hydrophobic groove, while the largest differences in the structural models are located at the solvent-exposed groups. Essentially, NMR² generates models in a reasonable time frame, which are sufficient to define vectors on the fragment molecules where the fragment can be further expanded or optimized through medicinal chemistry. NMR² does not overcome the challenges that result from the low affinity of the ligand, and so relatively high protein concentrations are required to afford the necessary sensitivity. Nonetheless, suitable amounts of the selectively labelled protein sample can be produced to run the experiments successfully.

In summary, we have shown that selective labeling of all methyl groups of a protein in a deuterated background yields useful probes for utilizing NMR² in the process of structure-based drug design with large complexes. We also report the binding mode of phenylthiazole 1 at the substrate binding site of EcDsbA. We anticipate a broad application of the proposed NMR² spectroscopy approach in the future of medicinal chemistry projects that are not amenable to X-ray crystallography.

Methods

Protein expression and purification of selective methyl labelled sample. *E. coli* BL21 (DE3) cells carrying the plasmid B0013-(5644bb) coding for EcDsbA were first adapted to D₂O by repeated subculturing with an increasing percentage of D₂O in the medium followed by double colony-selection³⁶. The colony showing the strongest growth was selected for preparation of glycerol stocks, which were stored at -80 °C. To produce a labelling pattern of A^βI^{δ1}(LV)^{proR}M^εT^{γ2}-¹³CH₃ in a deuterated background (hereafter referred to as [U-²H]-A^βI^{δ1}(LV)^{proR}M^εT^{γ2}-¹³CH₃), we used a HLAM-A^βI^{δ1}/LV^{proR}, U-¹³C, M^ε/T^γ kit from NMR-Bio (<http://www.nmr-bio.com/>) according to the manufacturer's instructions. In this labelling scheme, the alanine β, isoleucine δ1, methionine ε, and threonine γ methyl positions are selectively protonated. The methyl groups of leucine and valine residues are stereospecifically labelled at pro-R positions. The methionine and threonine methyls are selectively ¹³C-labelled, with the remainder of the side chain being ¹²C. Alanine, isoleucine, leucine and valine residues are labelled with ¹³C as depicted to generate a linear spin system (Fig. 1d). ¹⁵NH₄Cl was added into the expression media to allow for ¹⁵N incorporation at each amino acid site. To suppress scrambling from threonine to glycine, unlabeled glycine was also included in the expression media. As a result, glycine backbone amide resonances are weak in the [¹⁵N, ¹H]-HSQC. The methionine and threonine precursors did not contain ¹⁵N so the weak cross-peaks for methionine and threonine residues in the [¹⁵N, ¹H]-HSQC are possibly due to a transamination effect (Supplementary Fig. 3a). Pre-expression cultures were prepared from 10 mL M9/D₂O medium supplemented with 1 g/L ¹⁵NH₄Cl and 2 g/L D-glucose-¹³C₆-1,2,3,4,5,6,6-d₇, inoculated from the glycerol stock and incubated for 16 h at 37 °C with constant agitation at 200 rpm. ¹⁵NH₄Cl was added into the expression media to assess the efficiency of ¹⁵N incorporation at each amino acid site. 250 mL M9/D₂O medium supplemented with 1 g/L ¹⁵NH₄Cl and 2 g/L D-glucose-¹³C₆-1,2,3,4,5,6,6-d₇ was inoculated with 5 mL pre-expression culture and incubated at 37 °C with constant agitation at 200 rpm. The D₂O solutions of 2-(D₃)-methyl-1,2,3,4-(¹³C₄)-acetolactate and 2,3,3,4,4-(D₅)-(L)-methyl-L-methionine were added to the culture 1 h before induction (OD₆₀₀ ~ 0.6). 15 min prior to induction, the D₂O solutions of 2-hydroxy-2-[2-(¹³C),1-(D₂)]-ethyl-3-oxo-4-(D₃)-butanoic acid, U-(¹³C)-2-(D)-L-alanine, 2,3-(D₂),4-(¹³C)-L-threonine and deuterated glycine were added. When the optical density at 600 nm of the growth culture reached 0.8, the culture was induced with 1 mM IPTG and incubated for 16 h at 18 °C with constant agitation at 200 rpm. The cells were harvested at 3200×g at 4 °C for 20 min and stored at -80 °C. The protein was subsequently purified and oxidized as previously described³⁷. To achieve complete backbone amide deuterium-to-proton exchange, the protein was partially unfolded in 2 M guanidine hydrochloride for 1 h at room temperature and refolded by rapid dilution in a tenfold volume of buffer (50 mM HEPES (pH 6.8), 50 mM NaCl). After a further desalting step, the refolding yield was approximately 70%. The total cost of the selective methyl labelling precursors used in preparation of the sample was ~€400, which yielded ~1.6 mg of refolded protein.

Binding affinity of phenylthiazole 1 and oxidized EcDsbA. To determine the equilibrium dissociation constant (*K_D*) for phenylthiazole 1 binding to EcDsbA, [U-¹⁵N]-labelled oxidized EcDsbA was prepared in HEPES buffer (50 mM HEPES (pH 6.8), 50 mM NaCl) with 90% H₂O and 10% D₂O. A series of NMR samples (each at a total volume of 160 μL) was prepared containing 50 μM protein and different concentrations of phenylthiazole 1 and transferred into 3 mm thin-walled NMR tubes. d₆-DMSO concentration was 2% (v/v) in each sample. For NOESY experiments, 350 μL of 0.25 mM [U-²H]-A^βI^{δ1}(LV)^{proR}M^εT^{γ2}-¹³CH₃-labelled oxidized EcDsbA and 3.5 mM unlabeled phenylthiazole 1 was prepared in 50 mM NaPi (pH 6.8), 25 mM NaCl, with 98% D₂O and 2% d₆-DMSO and transferred to a Shigemi NMR tube. The reported pH of buffers in D₂O are uncorrected meter reading. Spectra were processed using Topspin. Peak assignments in the individual spectra were manually adjusted by CARA. Chemical shift perturbations (CSP) were calculated using the equation as described previously⁶. *K_D* was calculated with GraphPad Prism[©] using the ligand depletion model and the equation as previously described.

NMR data acquisition, processing and data analysis. All NMR spectra were collected at 298 K on Bruker 600 MHz or 800 MHz spectrometers equipped with CryoProbes. For tracking side-chain methyl CSP in the presence of phenylthiazole **1**, constant time [$^{13}\text{C}, ^1\text{H}$]-HSQC spectra of [$\text{U-}^2\text{H}$]- $\text{A}^{\beta\delta 1}(\text{LV})^{\text{proRMeT}^{\text{Y}2-13}\text{CH}_3}$ -labelled *EcDsbA* with and without phenylthiazole **1** were collected. 3D $^{13}\text{C}_{\text{ali}}$ -edited [$^1\text{H}, ^1\text{H}$]-NOESY-HMQC (NOE mixing times of 100 ms and 400 ms) experiments of 0.25 mM [$\text{U-}^2\text{H}$]- $\text{A}^{\beta\delta 1}(\text{LV})^{\text{proRMeT}^{\text{Y}2-13}\text{CH}_3}$ -labelled *EcDsbA* and 3.5 mM phenylthiazole **1** were acquired, either by uniform sampling or by non-uniform samples drawn from a probability density function with an exponential decay³⁸. To acquire 3D NUS $^{13}\text{C}^{\text{methyl}}$ -edited [$^1\text{H}, ^1\text{H}$]-NOESY-HMQC, 25% of 30×512 complex points were acquired for the indirect carbon and NOESY dimensions respectively in ~ 27 h. 16 transients per FID were acquired and NOE mixing time was set to 400 ms. Uniformly sampled 3D $^{13}\text{C}^{\text{methyl}}$ -edited [$^1\text{H}, ^1\text{H}$]-NOESY-HMQC data were collected with the same acquisition parameters but with 8 transients per FID in 54 h. Uniformly sampled NOESY data were processed by Topspin. A series of 2D $\omega 1, \omega 2$ - $^{13}\text{C}, ^{15}\text{N}$ -filtered [$^1\text{H}, ^1\text{H}$]-NOESY spectra was collected with NOE mixing of 10 ms, 20 ms, 40 ms and 70 ms, to calculate the bioactive ligand conformation in solution. Linearly sampled datasets were processed using Topspin 3.2 (pl6) and non-uniformly sampled datasets were processed either by compressed sensing in Topspin or using hmsIST³⁹. The spectra were converted to XEASY format for analysis in CARA. Previously obtained backbone and side-chain methyl assignments were used as the input of CARA however prochiral methyl assignments of Leucine and Valine residues were established from the constant time [$^{13}\text{C}, ^1\text{H}$]-HSQC spectrum of [$\text{U-}^2\text{H}$]- $\text{A}^{\beta\delta 1}(\text{LV})^{\text{proRMeT}^{\text{Y}2-13}\text{CH}_3}$ -labelled *EcDsbA*²³.

The methyl-specific isotope labelling used here allows the assignment of the type of amino acid residue based on the chemical shift of the ^{13}C methyl resonance. Moreover, because of the labelling strategy, it is possible to distinguish between residue types depending on whether the ^{13}C -methyl carbon is directly attached to another ^{13}C atom or not (Fig. 1). In the case of methionine, the methyl carbon is attached to a sulfur atom and for threonine the methyl is attached to a $^{12}\text{C}^{\beta}$ carbon. This results in peaks in the constant time HSQC spectrum having different phases, with the negative peaks belonging to the methyl group resonances of the threonine and the methionine (see Supplementary Fig. S2) and the positive peaks corresponding to the other methyl resonances. Methyl peaks corresponding to the amino acid residue types Met and Thr can therefore directly be identified by their phase and chemical shift and this assignment information can be used by NMR².

Solubility of phenylthiazole **1 in D_2O NMR buffer.** A 2D [$^1\text{H}, ^1\text{H}$]-NOESY spectrum at 4 mM phenylthiazole **1** in D_2O NMR buffer (50 mM NaPi (pH 6.8), 25 mM NaCl, 2% d_6 -DMSO) was acquired at 298 K. The NOE mixing time was set to 800 ms. The NOE cross peak signals are of opposite sign to the diagonal peaks, suggesting the ligand is soluble at a concentration of 4 mM (see Supplementary Fig. S4).

Proton chemical shift assignments of phenylthiazole **1 in its free and bound states.** Proton chemical shifts of free phenylthiazole **1** in D_2O NMR buffer (phenylthiazole **1** concentration 4 mM; 50 mM NaPi (pH 6.8), 25 mM NaCl, 2% d_6 -DMSO) were assigned using 1D ^1H , 2D [$^{13}\text{C}, ^1\text{H}$]-HMBC and 2D [$^1\text{H}, ^1\text{H}$]-NOESY spectra. ^1H chemical shift assignments of phenylthiazole **1** in the protein bound state were made using 1D ^1H and 2D $\omega 1, \omega 2$ - $^{13}\text{C}, ^{15}\text{N}$ -filtered [$^1\text{H}, ^1\text{H}$]-NOESY spectra.

NMR structure calculation with HADDOCK. The data-driven HADDOCK docking calculation was performed using the protocol as described previously by Mohanty et al.²³. The topology and parameter files of phenylthiazole **1** for HADDOCK were created using ATB server⁴⁰. 40 protein conformers sampled from both chain A and B (PDB ID:1FVK) were provided as the input of HADDOCK. To account for ligand flexibility, 26 discrete sets of ligand conformers were generated using the ConfGen advanced panel in Schrödinger Maestro. The residues that exhibited CSP > 0.02 ppm and had a calculated solvent accessibility above 15% were selected for use as ambiguous intermolecular distances in HADDOCK. NMR² input NOE lists with protein assignments were used for HADDOCK. The intermolecular NOE distance constraints were tabulated in CNS format with the lower distance limit (target distance in Å - 1.8 Å) and the upper distance limit (target distance in Å + 15% of target distance in Å) as described²³. 10,000 models were generated in it0 (rigid body docking stage) and 400 models were calculated in it1 (semi-flexible simulated annealing) followed by water refinement. Of the final 400 water refined HADDOCK models, the 10 lowest energy structures with minimal distance restraint violations (< 0.5 Å) were selected for the analysis. The electrostatic flag was switched on during the HADDOCK calculation. The segment “Q164-D172” was defined as fully flexible throughout HADDOCK docking as reported previously²³.

X-ray structure determination. Soaking of phenylthiazole **1** into *EcDsbA* crystals was performed using the protocol as described previously³⁷. *EcDsbA*-phenylthiazole **1** complex structure coordinates have been deposited into PDB with accession number 7TTV.

NMR² structure determination. The NMR² method does not use any force field to calculate the structures, while the structures derived from HADDOCK and X-ray crystallography refinement software are optimized with explicit water molecules. All spectra were processed with Topspin 3.1 (Bruker). The spectra were evaluated with ccpNMR analysis 2.3⁴¹. Distances were derived from NOE build-up curves using a simple two-spin system model (i,j) and following the established protocol (see Supplementary Table S1)^{3,15,28,42,43}. The auto-relaxation rates, ρ_i , and initial magnetizations, $\Delta M_{ii}(0)$, were extracted using a mono-exponential decay function, $\Delta M_{ii}(t) = \Delta M_{ii}(0) \exp(-\rho_i t)$. Missing auto-relaxation rates were set to the median of the experimentally derived ones. Cross-relaxation rates, σ_{ij} , were fitted using a two-spin system approximation model for the protein-ligand

NOEs, $\Delta M_{ij}(t)$, Eq. (1)⁴⁴. The corresponding distances, r_{ij} , are derived from the cross-relaxation rates, σ_{ij} , defined in Eq. (3),

$$\frac{\Delta M_{ij}(t)}{\Delta M_{ii}(0)} = -\frac{\sigma_{ij}}{\lambda_+ - \lambda_-} (e^{-\lambda_- t} - e^{-\lambda_+ t}) \quad (1)$$

$$\lambda_{\pm} = \frac{\rho_i + \rho_j}{2} \pm \sqrt{\left(\frac{\rho_i - \rho_j}{2}\right)^2 + \sigma_{ij}^2} \quad (2)$$

$$\sigma_{ij} = \frac{b^2}{r_{ij}^6} (6J(2\omega) - J(0)) \quad (3)$$

$$J(\omega) = \frac{2}{5} \left(\frac{\tau_c}{1 + (\omega\tau_c)^2} \right) \quad (4)$$

$$b = \frac{1}{2} \frac{\mu_0}{4\pi} \hbar \gamma_H^2 \quad (5)$$

where μ_0 is the permeability of free space, \hbar the reduced Planck constant and γ_H the gyromagnetic ratio of the nucleus and τ_c the rotational correlation time of the protein–ligand complex. 37 intermolecular and 49 intraligand NOEs could be measured. Excluding build-up curves with large spin diffusion contributions, Supplementary Fig. S6 reduced the amount of distances restraints to 12 for the protein–ligand and 19 for the ligand alone. Because the bound conformation of the ligand was poorly converged, we used the 20 best ligand conformations as input for the NMR² calculations. Then the ligand conformation was restrained during the NMR² calculations while the protein side chains were allowed to adjust. The NMR² structure is the one with the lowest CYANA target function (target function $\sim 1.0 \text{ \AA}^2$). As expected from the NOE data set, the NMR² second-best binding mode (target function $\sim 1.5 \text{ \AA}^2$) is the mirror image along the long axis of the ligand of the best structure. This is due to the symmetry of the NMR restraints from the ligand aromatic rings. The structures were then minimized (100 steepest descent minimization steps) in Chimera using the amberff14sb force field and Gasteiger partial charges for the ligand.

Data availability

The datasets and analysis generated in the current study are available upon request (correspondence to martin.scanlon@monash.edu and julien.orts@univie.ac.at).

Received: 29 March 2022; Accepted: 25 May 2022

Published online: 04 July 2022

References

- Carr, R. A., Congreve, M., Murray, C. W. & Rees, D. C. Fragment-based lead discovery: Leads by design. *Drug Discov. Today* **10**, 987–992. [https://doi.org/10.1016/S1359-6446\(05\)03511-7](https://doi.org/10.1016/S1359-6446(05)03511-7) (2005).
- Carvalho, A. L., Trincão, J. & Romão, M. J. X-ray crystallography in drug discovery. *Methods Mol. Biol.* **572**, 31–56. https://doi.org/10.1007/978-1-60761-244-5_3 (2009).
- Ni, F. Recent Developments in transferred Noe methods. *Prog. Nucl. Mag. Res. Sp.* **26**, 517–606. [https://doi.org/10.1016/0079-6565\(94\)80015-4](https://doi.org/10.1016/0079-6565(94)80015-4) (1994).
- Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737. <https://doi.org/10.1021/ja026939x> (2003).
- McCoy & Wyss. Structures of protein–protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations. *J. Am. Chem. Soc.* **124**, 2104–2105. <https://doi.org/10.1021/ja017242z> (2002).
- Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16. <https://doi.org/10.1016/j.pnmrs.2013.02.001> (2013).
- van Zundert, G. C. P. *et al.* The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725. <https://doi.org/10.1016/j.jmb.2015.09.014> (2016).
- Hajduk, P. J. *et al.* SOS-NMR: A saturation transfer NMR-based method for determining the structures of protein–ligand complexes. *J. Am. Chem. Soc.* **126**, 2390–2398. <https://doi.org/10.1021/ja039480v> (2004).
- Madrid, M., Llinas, E. & Llinas, M. Model-independent refinement of interproton distances generated from 1H-NMR overhauser intensities. *J. Magn. Reson.* **93**, 329 (1991).
- Guan, J. Y. *et al.* Small-molecule binding sites on proteins established by paramagnetic NMR spectroscopy. *J. Am. Chem. Soc.* **135**, 5859–5868. <https://doi.org/10.1021/ja401323m> (2013).
- Chen, W. N. *et al.* Sensitive NMR approach for determining the binding mode of tightly binding ligand molecules to protein targets. *J. Am. Chem. Soc.* **138**, 4539–4546. <https://doi.org/10.1021/jacs.6b00416> (2016).
- Orts, J. *et al.* Specific methyl group protonation for the measurement of pharmacophore-specific interligand NOE interactions. *Chem. Eur. J.* **14**, 7517–7520. <https://doi.org/10.1002/chem.200800880> (2008).
- Orts, J. *et al.* Crystallography-independent determination of ligand binding modes. *Angew. Chem. Int. Ed.* **47**, 7736–7740. <https://doi.org/10.1002/anie.200801792> (2008).
- Totrov, M. & Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **18**, 178–184. <https://doi.org/10.1016/j.sbi.2008.01.004> (2008).
- Orts, J. *et al.* NMR-based determination of the 3D structure of the ligand–protein interaction site without protein resonance assignment. *J. Am. Chem. Soc.* **138**, 4393–4400. <https://doi.org/10.1021/jacs.5b12391> (2016).

16. Wälti, M. & Orts, J. The NMR2 method to determine rapidly the structure of the binding pocket of a protein-ligand complex with high accuracy. *Magnetochemistry* **4**, 12 (2018).
17. Wälti, M. A., Riek, R. & Orts, J. Fast NMR-based determination of the 3D structure of the binding site of protein-ligand complexes with weak affinity binders. *Angew. Chem. Int. Ed.* **56**, 5208–5211. <https://doi.org/10.1002/anie.201612304> (2017).
18. Sprangers, R., Velyvis, A. & Kay, L. E. Solution NMR of supramolecular complexes: Providing new insights into function. *Nat. Methods* **4**, 697–703. <https://doi.org/10.1038/nmeth1080> (2007).
19. Wiesner, S. & Sprangers, R. Methyl groups as NMR probes for biomolecular interactions. *Curr. Opin. Struct. Biol.* **35**, 60–67. <https://doi.org/10.1016/j.sbi.2015.08.010> (2015).
20. Heras, B. *et al.* DSB proteins and bacterial pathogenicity. *Nat. Rev. Microbiol.* **7**, 215–225. <https://doi.org/10.1038/nrmicro2087> (2009).
21. Kerfah, R., Plevin, M. J., Sounier, R., Gans, P. & Boisbouvier, J. Methyl-specific isotopic labeling: A molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **32**, 113–122. <https://doi.org/10.1016/j.sbi.2015.03.009> (2015).
22. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* **1**, 749–754. <https://doi.org/10.1038/nprot.2006.101> (2006).
23. Mohanty, B. *et al.* Determination of ligand binding modes in weak protein-ligand complexes using sparse NMR data. *J. Biomol. NMR* **66**, 195–208. <https://doi.org/10.1007/s10858-016-0067-4> (2016).
24. Proudfoot, A., Frank, A. O., Ruggiu, F., Mamo, M. & Lingel, A. Facilitating unambiguous NMR assignments and enabling higher probe density through selective labeling of all methyl containing amino acids. *J. Biomol. NMR* **65**, 15–27. <https://doi.org/10.1007/s10858-016-0032-2> (2016).
25. LaPlante, S. R. *et al.* Monitoring drug self-aggregation and potential for promiscuity in off-target in vitro pharmacology screens by a practical NMR strategy. *J. Med. Chem.* **56**, 7073–7083. <https://doi.org/10.1021/jm4008714> (2013).
26. Orts, J. & Gossert, A. D. Structure determination of protein-ligand complexes by NMR in solution. *Methods* **138–139**, 3–25. <https://doi.org/10.1016/j.ymeth.2018.01.019> (2018).
27. Vogeli, B. The nuclear overhauser effect from a quantitative perspective. *Prog. Nucl. Magn. Reson. Spectrosc.* **78**, 1–46. <https://doi.org/10.1016/j.pnmrs.2013.11.001> (2014).
28. Orts, J., Vogeli, B. & Riek, R. Relaxation matrix analysis of spin diffusion for the NMR structure calculation with eNOEs. *J. Chem. Theory Comput.* **8**, 3483–3492. <https://doi.org/10.1021/ct3002249> (2012).
29. Mareuil, F., Malliavin, T. E., Nilges, M. & Bardiaux, B. Improved reliability, accuracy and quality in automated NMR structure calculation with ARIA. *J. Biomol. NMR* **62**, 425–438. <https://doi.org/10.1007/s10858-015-9928-5> (2015).
30. Nilges, M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE cross-peaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660. <https://doi.org/10.1006/jmbi.1994.0053> (1995).
31. Lippens, G. M., Cerf, C. & Hallenga, K. Theory and experimental results of transfer-Noe experiments. 1. The influence of the off rate versus cross-relaxation rates. *J. Magn. Reson.* **99**, 268–281 (1992).
32. Campbell, A. P. & Sykes, B. D. Theoretical evaluation of the 2-dimensional transferred nuclear overhauser effect. *J. Magn. Reson.* **93**, 77–92. [https://doi.org/10.1016/0022-2364\(91\)90033-P](https://doi.org/10.1016/0022-2364(91)90033-P) (1991).
33. Breeze, A. L. Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. Nucl. Magn. Reson. Spectrosc.* **36**, 323–372. [https://doi.org/10.1016/S0079-6565\(00\)00020-0](https://doi.org/10.1016/S0079-6565(00)00020-0) (2000).
34. Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378. <https://doi.org/10.1385/1-59259-809-9:353> (2004).
35. Paxman, J. J. *et al.* The structure of the bacterial oxidoreductase enzyme DsbA in complex with a peptide reveals a basis for substrate specificity in the catalytic cycle of DsbA enzymes. *J. Biol. Chem.* **284**, 17835–17845. <https://doi.org/10.1074/jbc.M109.011502> (2009).
36. Sivashanmugam, A. *et al.* Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* **18**, 936–948. <https://doi.org/10.1002/pro.102> (2009).
37. Adams, L. A. *et al.* Application of fragment-based screening to the design of inhibitors of *Escherichia coli* DsbA. *Angew. Chem. Int. Ed. Engl.* **54**, 2179–2184. <https://doi.org/10.1002/anie.201410341> (2015).
38. Mobli, M., Stern, A. S., Bermel, W., King, G. F. & Hoch, J. C. A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. *J. Magn. Reson.* **204**, 160–164. <https://doi.org/10.1016/j.jmr.2010.02.012> (2010).
39. Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H. & Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. *J. Biomol. NMR* **52**, 315–327. <https://doi.org/10.1007/s10858-012-9611-z> (2012).
40. Malde, A. K. *et al.* An automated force field topology builder (ATB) and repository: Version 1.0. *J. Chem. Theory Comput.* **7**, 4026–4037. <https://doi.org/10.1021/ct200196m> (2011).
41. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinform.* **59**, 687–696. <https://doi.org/10.1002/Prot.20449> (2005).
42. Vogeli, B. *et al.* Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. *J. Am. Chem. Soc.* **131**, 17215–17225. <https://doi.org/10.1021/ja905366h> (2009).
43. Noggle, J. H. & Schirmer, R. E. *The Nuclear Overhauser Effect* (Academic Press, 1971).
44. John Cavanagh, N. S., Fairbrother, W., Rance, M. & Palmer III, A. *Protein NMR Spectroscopy, Principles and Practice* (Academic Press, 2006).

Acknowledgements

Research support for this project was provided by SNF grant 310030_192646 to JO, NHMRC grant 1099151 to MS and BH, and ARC grant IC180100021 to MS. The study was financially supported by the ETH Zurich, the University of Vienna, and Monash University. The authors acknowledge the use of the Bio21 NMR Facility and the Sydney Analytical Magnetic Resonance Facility at the University of Sydney for providing access to infrastructure. We thank Scott Anthony Robson for helping us to process the 3D NUS-NOESY-HMQC spectra using hmsIST. We also thank to Rime Kerfah from NMR-Bio for the valuable discussion on implementing the specific methyl labelling scheme.

Author contributions

Conceptualization, M.J.S., B.M. and J.O.; Methodology, B.M., J.O., G.W., S.N., W.S.A., B.C.D. and M.L.W.; Investigation, B.M., J.O. and M.J.S.; Writing—Original Draft, J.O. and B.M.; Writing—Review, B.M., J.O., B.C.D. and M.J.S.; Editing, all authors; Resources and Funding Acquisition, M.J.S. and J.O.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13561-y>.

Correspondence and requests for materials should be addressed to J.O. or M.J.S.

Reprints and permissions information is available at www.nature.com/reprints.

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



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/>.

© The Author(s) 2022