Fast Validation of Static and Dynamic 3D Models of Weak Protein-Ligand Complexes from STD NMR Spectroscopy

Ridvan Nepravishta[a,c], Samuel Walpole[a], Thomas Hicks[a], Juan C. Muñoz-García[b] and Jesús Angulo*[a,b]

Abstract: Saturation transfer difference NMR spectroscopy has revolutionized the study of weak receptor-ligand interactions. Its versatility and popularity are demonstrated by a myriad of approaches developed. Methodologies such as DEEP-STD NMR, K_D determination, SSTD, and determination and validation of protein-ligand complexes are a few elegant examples among them. However, the use of the STD NMR technique together with full relaxation matrix calculations for the determination and structure evaluation of protein-ligand complexes remain a major milestone in the field. In this communication, we present a new approach based on a reduced relaxation matrix that pushes further the boundaries of the relaxation matrix theory applied to structure determination and evaluation, in solution, using STD NMR data and molecular dynamics simulations.

The analysis of the 3D structural determinants for specificity in the molecular recognition of small ligands by proteins in solution is a field of major interest in structural biology and drug design, with NMR spectroscopy playing a key role.[1–4] The NMR approaches towards 3D structure determination in solution fall into two main categories, where the observables can originate either from the spins of the protein or from those of the small ligand. [5,6] Both spectroscopic strategies have a fundamental ground on the same physical principle known as Nuclear Overhauser Effect (NOE).[7] Within the realm of ligand-observed NMR strategies, WaterLOGSY[8,9] and saturation transfer difference STD NMR are powerful tools to probe and characterize weak protein-ligand interactions, being extensively used due to their robustness, short experimental times in comparison to protein-observed NMR approaches, flexibility and relatively low cost.[10] Very conveniently, they rely on the acquisition of ligand ^{1}H NMR spectra in the presence of low amounts of the protein without the need of isotopic enrichment. It should be noted that STD NMR is particularly useful for the structural analysis of weak protein-ligand complexes as it allows to identify the orientation, binding determinants, and ligand conformation within the complex through the use of 3D molecular models that best agree with the experimental data.[11–13]

The STD NMR technique is a difference spectroscopy approach based on the selective saturation of a group of protons of the protein using a low power shaped pulse repeated for long enough time to approach a state of saturation of the whole protein via intramolecular spin diffusion, which is very efficient for large molecules. Upon ligand binding, saturation is first transferred from the protein to the closest protons of the bound ligand, via inter-molecular NOE. This NOE builds up in the complex as a function of both the protein-ligand proton-proton distances and the residence time of the ligand in the bound state. For weak interactions, the ligand residence time is typically much shorter than the relaxation time of the ligand in the complex. The saturation received from the ligand via intermolecular NOE is transferred through chemical exchange in the ligand free state. The resulting NMR spectrum is then compared to a reference spectrum (so-called off-resonance spectrum), which is obtained by positioning the selective saturation pulse in a spectral region far away from both, the ligand, and the protein resonances. The difference between these two spectra results in the STD NMR spectrum. In order to perform the STD NMR experiment, the protein should be present in a very low amount compared to the ligand (usually protein to ligand ratios of 1:30 to 1:100) to minimise fast-rebinding effects.[14] Further, the ligand/receptor complex dissociation constant, K_D, should preferably sits within the high nanomolar to millimolar range (10^{-7}-10^{-3} M) as fast exchange kinetics of binding is needed for efficient accumulation of the saturated ligand in the free state in solution.

A latest development of STD NMR, called DEEP-STD NMR,[15] has pushed further this concept allowing the identification of the types of protein side chains (aromatic, aliphatic, polar, apolar) surrounding the ligand, facilitating the determination of the orientation of the ligand within the architecture of the binding site, and has opened the door to a novel set of STD NMR protocols that are collectively termed multi-frequency STD NMR approaches.[16] Proper analysis of the intensities of the ligand STD NMR signals allows to gain structural information on the molecular recognition of the ligand by the receptor, revealing what particular ligand protons are far or near from the protein-ligand interface in the binding site, identifying what ligand regions are buried in the binding pocket, and what areas are farther apart and consequently more solvent exposed (the so-called binding epitope mapping). Generally speaking, binding epitopes should not be obtained from one saturation time only, as artefacts might

[a] T. Hicks, Dr S. Walpole, Dr R. Nepravishta, Dr J. Angulo
School of Pharmacy, University of East Anglia
Norwich Research Park, NR4 7TJ Norwich, UK
E-mail: j.angulo@uea.ac.uk

[b] Dr Juan C. Muñoz-García, Dr J. Angulo
Instituto de Investigaciones Químicas (IIQ-CSIC)
49 Américo Vespucio Ave. 41092, Sevilla, Spain
E-mail: j.angulo@iiq.csic.es

[c] Dr R. Nepravishta
Cancer Research Horizons, The Beatson Institute
Garscube Estate, Switchback Road, Bearsden, Glasgow
G61 1BD, UK
E-mail: ridvan.nepravishta@cancer.org.uk

occur leading to wrong classification of relative STD intensities within the ligand due to (i) differences in ligand T_1 relaxation among the protons, (ii) kinetics of binding, (iii) fast rebinding effects, and (iv) extent of saturation. In order to minimise the impact of these factors on the determination of the binding epitope mapping from STD NMR, Mayer and James proposed to acquire a set of STD NMR experiments at increasing saturation times. In this procedure, known as the analysis of initial slopes from STD NMR build-up curves, the experimental STD data points are fitted to the mono-exponential equation[17]:

$$\mathrm{STD}(t_{sat}) = \mathrm{STD}^{max}(1 - \exp(-k_{sat}t_{sat})) \quad \text{[eq. 1]}$$

Where $\mathrm{STD}(t_{sat})$ is the STD factor of a particular resonance obtained as $I_0 - I_{sat}/I_0$ at a saturation time t_{sat} , STD^{max} is the max asymptotic value obtained from fitting, k_{sat} is the saturation rate constant while $t_{sat}$is the saturation time. The initial slope, STD_0, is obtained from eq. 2:

$$\mathrm{STD}_0 = k_{sat} \cdot t_{sat} \quad \text{[eq. 2]}$$

The STD_0 values are then normalized to the highest value within the set of ligand protons, allowing to analyse the binding epitope, minimising the possible artefacts described above .[14]

Although the binding epitope mappings derived from STD NMR experiments can inform subsequent optimization of the identified hits for a given protein target, arguably the best aid for a medicinal chemist will come from the availability of experimentally validated 3D molecular models of the protein-ligand complexes. However, during the ligand screening stage of the drug discovery pipeline, fragments identified as hits typically show very weak affinity, precluding in most of the cases the determination of the 3D structure of the receptor-fragment complex by X-ray crystallography. For these cases, the relevance of STD NMR cannot be underappreciated, due to its excellent performance to study protein-ligand complexes under fast chemical exchange or weak affinity, and its capability to access to structural information at atomic detail.

Where crystallization of receptor-fragment complexes has been unsuccessful, the process of ligand optimization ("hit-to-lead") will strongly benefit from availability 3D molecular models validated by STD NMR. Further, it is not only very useful to validate static 3D models, but also trajectories of the protein-ligand complexes from molecular dynamics (MD) simulations.

Simplified algorithms based on protein-ligand distance hierarchies following a distance cutoff have been shown to be efficient to score sets of binding conformations based on minimum deviations with respect to experimental binding epitope mappings.[18,19] However, purely distance-based algorithms suffer from a lack of description of the dipole-dipole relaxation processes during the STD NMR experiment, so that, although flexibility is included by analyzing MD simulations, the impact of internal motions in the cross-relaxation (NOE) efficiency is completely absent, which might compromise the results for protein-ligand complexes with significant side chain mobility for amino acids surrounding the ligand in the binding pocket.

Full relaxation matrix approaches have also been previously applied to validate static and dynamic models of protein-ligand complexes by STD NMR.[20] However, such full matrix approaches can become extremely time-consuming, turning impractical for analyzing very long MD simulations and simultaneously deriving theoretical full STD build-up curves for each frame of the simulation. To reduce the computational burden, the application of these full matrix approaches in the literature has been customarily limited to the case of "one saturation time" analysis of MD trajectories.[21] However, this type of analysis is extremely prone to false optimum structures when scoring is based on a best fitting factor value (NOE R-factor), as the full matrix calculation is not used to predict the whole build-up of saturation. Thus, whereas a particular protein-ligand complex can give a bad (high) R-factor when considering the whole STD build-up curve, a relatively good (low) R-factor might be obtained for the same complex when using a single saturation time, hence misleading the model validation.

In this communication, we present a *reduced matrix* analysis of the STD NMR initial slopes that fully considers both the network of dipole-dipole couplings and the relaxation processes present at the protein-ligand interface. This allows to (i) perform extremely fast calculations of theoretical STD initial slopes (STD_0^{cal}) from a 3D molecular model (or MD trajectory) of the protein-ligand complex, and (ii) explicitly include experimental STD initial slopes (STD_0^{exp}) from complete STD build-up curves in receptor-ligand 3D structure determination and evaluation in solution.

In order to validate this new approach, we studied three protein-ligand complexes (PDB codes: 6MSY, 4X4A, 6GH2) for which high resolution X-ray and STD_0 values were available. The agreement between theoretical and experimental STD_0 factors was evaluated using the NOE R-factor, defined for *k* ligand protons as:

$$\mathrm{NOE\ R-factor} = \sqrt{\frac{\sum_k W_k\left(\mathrm{STD}_0^{exp,k} - \mathrm{STD}_0^{cal,k}\right)^2}{\sum_k W_k\left(\mathrm{STD}_0^{exp,k}\right)^2}} \quad \text{[eq. 3]}$$

Where W_k is a weighting factor (in our NOE R-factor calculation W_k = 1), $\mathrm{STD}_0^{exp,k}$ is the STD_0 experimental value obtained for the *k-th* proton of the ligand while the $\mathrm{STD}_0^{cal,k}$ is the calculated STD_0 value using the reduced matrix approach presented in this communication.

Theory

The complete relaxation matrix for a two-site exchange system was already described from Jayalakshmi *et al.*[22] In this model, it is proposed that the protein protons can be divided in (i) E1 and E1', protein protons not directly affected by saturation, in the free and bound states, respectively, and (ii) E2 and E2', protein protons directly affected by saturation, in the free and bound states, respectively. Considering the above premises, the theoretical STD NMR build-up curves can be derived using the following equation.

$$\mathbf{I}(\mathbf{t_{sat}}) = \mathbf{I_0} + \left[\frac{\mathbf{M}}{\mathbf{R+K}}\right](\mathbf{1} - \mathbf{exp}[-(\mathbf{R}+\mathbf{K}) \cdot \mathbf{t_{sat}}]) \quad \text{[eq. 4]}$$

Where $\mathbf{I_0}$ is the Intensity of the protons of the ligand at thermal equilibrium while $\mathbf{I(t_{sat})}$ is the intensity of the ligand

protons after protein saturation, **R** is the general relaxation rate matrix for the ligand, **K** is the generalized exchange kinetics matrix and **M** is a general matrix term containing elements of both protein and ligand species in their bound and free form. The equation resembles the mono-exponential equation proposed by Mayer and James for the fitting of STD NMR experimental data.[17] In fact, the sum of the general relaxation rate matrix **R** and the kinetics exchange matrix **K** is equivalent to the saturation rate constant k_{sat}, which is calculated from the mono-exponential fit of the experimental STD NMR data. Moreover, the term **M/(R+K)** is proportional to $\mathrm{STD}^{\max}$, and is also obtained during the mono-exponential fit. On these premises equation 2 can be arranged by substitution:

$$\mathrm{STD}_0 = \mathrm{k_{sat}} \cdot \mathrm{STD}^{\max} \propto (\mathbf{R}+\mathbf{K}) \cdot \left[\frac{\mathbf{M}}{(\mathbf{R}+\mathbf{K})}\right] \qquad \text{[eq. 5]}$$

which simplifies to:

$$\mathrm{STD}_0 \propto \mathbf{M} \qquad \text{[eq. 6]}$$

With the **M** matrix defined in as scheme 1:

$$\mathbf{M} = \begin{vmatrix} \mathbf{0} \\ \mathbf{R}_{\mathbf{E1E2}}\,\mathbf{I}_{\mathbf{E20}} \\ \mathbf{R}_{\mathbf{L'E2'}}\,\mathbf{I}_{\mathbf{E2'0}} \\ \mathbf{R}_{\mathbf{E1'E2'}}\,\mathbf{I}_{\mathbf{E2'0}} \end{vmatrix}$$

Scheme 1. M matrix definition.

In Scheme 1, $\mathbf{R}_{\mathbf{E1E2}}$ is the cross relaxation matrix between the unsaturated (E1) and directly saturated (E2) protein protons in the free state multiplied by the thermal equilibrium intensity of the directly saturated protons of the protein (E2) in the free state $\mathbf{I}_{\mathbf{E20}}$; $\mathbf{R}_{\mathbf{L'E2'}}$ is the cross relaxation matrix between the ligand protons in the bound state (L1') and the directly saturated protons of the protein in the bound state (E2') multiplied by the thermal equilibrium intensity of the saturated bound protons (E2') $\mathbf{I}_{\mathbf{E2'0}}$; $\mathbf{R}_{\mathbf{E1'E2'}}$ is the cross relaxation matrix between the unsaturated and directly saturated protein protons in the bound state (E1' and E2') multiplied by the thermal equilibrium intensity of the directly saturated protein protons in the bound state (E2') $\mathbf{I}_{\mathbf{E2'0}}$.

We have developed a new algorithm for the quantitative analysis and fast validation of 3D models of protein-ligand complexes against experimental STD NMR data, based on the determination of the reduced relaxation matrix **M** (Eq. 6). The protocol has been implemented in the form of a web application, called *RedMat*, which takes as input data (i) the Cartesian coordinates of the protein-ligand complex, (ii) the experimental STD initial slopes (STD_0 values), in relative scale, *i.e.* normalised to the highest intensity STD_0 (arbitrarily assigned a value of 100%), (iii) the protein protons initially saturated in the STD NMR experiment, and (iv) a number of additional experimental parameters. These parameters include: the NMR spectrometer frequency (in MHz), the rotational correlation time of the bound protein-ligand complex (in ns), the concentrations of ligand and protein (both in μM units), the dissociation constant of the protein ligand complex (in μM units), and the cut-off distance (in Å units) for protein protons from the ligand to be considered in the calculation.

RedMat accepts both 3D structures of complexes and dynamic ensembles from molecular dynamics (MD) simulations. The former takes the sets of coordinates in the form of Protein Data Bank (PDB) files, one for the protein and one for the ligand. Predicted initial slope values ($\mathrm{STD}_0^{\mathrm{cal,k}}$) and the NOE R-factor (Eq. 3) are outputted for each ligand proton. In our experience, as a rule of thumb, an NOE R-factor of less than 0.3 indicates a good agreement between the calculated and experimental STD_0 values. If multiple models are present in the ligand file (*i.e.* different ligand poses from docking calculations), separate calculations will be performed for each model, making this mode useful for quantitatively assessing the agreement of molecular docking models with the experimental STD NMR data. The second mode accepts AMBER[23] topology and trajectory files, allowing MD trajectories to be used as input files. This mode gives the evolution of the NOE R-factor as a function of the simulation time for each MD trajectory frame, allowing to elucidate to what extent changes in the conformation and/or orientation of the ligand and the protein binding site along the MD simulation agree or disagree with the experimental data. The application additionally allows individual frames to be downloaded for further analysis. For flexible receptors and ligands, the dynamic mode significantly improves the analysis over static models from rigid or semi-rigid docking calculations, where the dynamics of both molecular partners is not fully considered.

In order to verify the validity of our reduced relaxation matrix algorithm, three protein-ligand systems, for which we have previously determined accurate experimental STD_0 values,[15,24,25] were tested. In all cases we tested both application modes, by first calculating the reduced relaxation matrix from the experimentally determined X-ray crystal structures of the protein-ligand complexes, and, secondly, performing the calculation on 100 ns MD trajectories with frames saved at 100 ps intervals (1000 frames). In all cases, for the purpose of STD_0 calculation from trajectories, all methyl protein protons were assumed to be saturated and a relatively long cut-off distance of 12 Å from the ligand was considered. This allowed to include all the protein protons that mostly contribute to the saturation transfer at the protein-ligand interface in the STD NMR experiment.

Human anti-HIV-1 antibody 2G12 – tetramannoside complex.

The first system considered was the human monoclonal antibody 2G12 in complex with a tetramannoside (PDB code: 6MSY), which is therapeutically of interest due to its role as a broadly neutralising antibody to human immunodeficiency virus (HIV)[26]. For the RedMat calculation, we used a rotational correlation time of the protein of 43.1 ns, estimated with HYDRONMR[27], and a dissociation constant of 1000 μM. The concentrations of ligand and protein were 8000 μM and 25 μM, respectively, as per experimental conditions.[24] Figure 1 shows the X-ray crystal structure of the 2G12-tetramannoside complex, and the comparison of experimental STD_0 values (blue bars) with those calculated using RedMat (red bars). A NOE R-factor of 0.27 was obtained, indicating a good fit between the experimental and theoretical STD_0 values. The MD trajectory displayed an average NOE R-factor

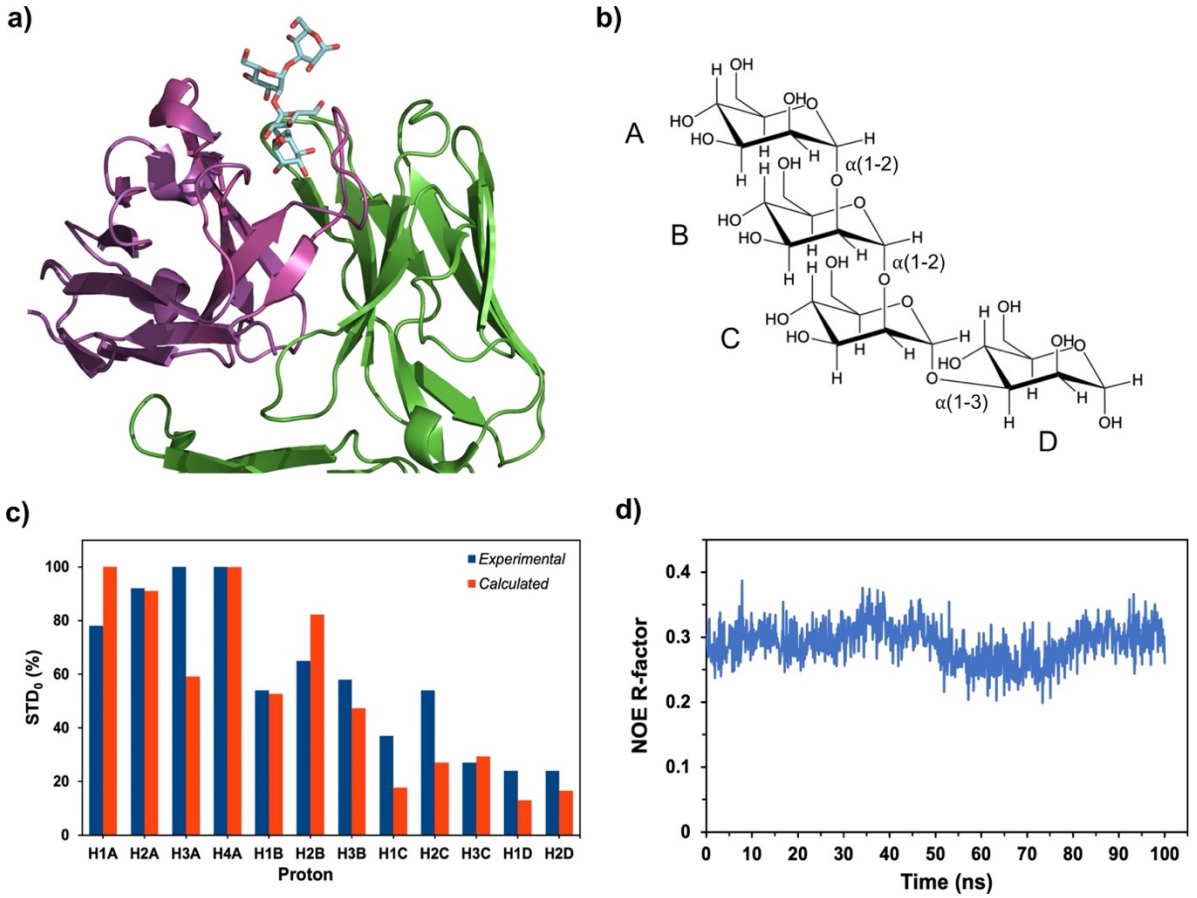

Figure 1. a) X-ray structure of the anti-HIV 2G12 monoclonal antibody (heavy and light chain in magenta and green, respectively) in complex with tetramannoside (in cyan; PDB 6MSY). The protein residues (at 12 Å distance from the ligand) that were considered for the calculation of theoretical STD_0 factors are shown in sticks. **b)** 2D sketch of the ligand. The labels associated to each sugar ring are shown next to it. **c)** Comparison between calculated (red bars) and experimental (blue bars) STD_0 factors for the protons of tetramannoside ligand. A NOE R-factor of 0.27 was obtained, showing a very good agreement between the crystal and the solution state structures of the complex. **d)** Evolution of the NOE R-factor of the tetramannoside ligand over 100 ns of MD simulation.

of 0.29 (with a standard deviation of 0.03), indicating that the complex remained relatively stable throughout the simulation (Fig. 1).

Gut intramolecular trans-sialidase *Rg*NanH-GH33 - 2,7-anhydro-Neu5Ac complex.

The second system considered was the complex of the catalytic domain of the intramolecular trans-sialidase from *Ruminococcus gnavus* RgNanH-GH33 with the ligand 2,7-anhydro-Neu5Ac (PDB code: 4X4A).[28] Unravelling the binding determinants of 2,7-anhydro-Neu5Ac in solution by sialidases is of major interest for the understanding of the fundamental mechanisms of gut microbiota adaptation. For the RedMat calculation, we used a rotational correlation time of the protein of 34.5 ns, estimated with HYDRONMR[27], and a dissociation constant of 2000 μM. The concentrations of ligand and protein were 1000 μM and 20 μM, respectively, as per experimental conditions.[15] Figure 2 shows the X-ray crystal structure of the RgNanH-GH33-2,7-anhydro-Neu5Ac complex, and the comparison of the experimental STD_0 values (blue bars) with those calculated using RedMat (red bars). The NOE R-factor was 0.13, indicating an extremely good fit between the STD NMR data and the X-ray crystal structure. When running the MD simulations of the RgNanH-GH33-2,7-anhydro-Neu5Ac complex, we observed that the model was in extremely good agreement with the experimental data for the first 45 ns, with an average NOE R-factor of 0.10 (with a standard deviation of 0.01). However, a conformational change event occurred at this point, resulting in a ligand conformation that did not agree that well with the experimental data. Since this effect might be due to the limitation of the forcefield parameters used, we introduced some experimental restraints to the model based on information known from DEEP-STD NMR.[15] These were restraints between H9/H9'/H8 of the ligand with Trp 698, H3a atom of the ligand with Ile 258, and the ligand methyl group with the methyl groups of Val 502 and Ile 338. In all cases the restraints allowed each pair to have a distance range between 2-6 Å, based on the NOE being detectable within these ranges. With these restraints, the trajectory showed extremely good

agreement (except for the last 3 ns), with an average NOE R-factor of 0.11 (with a standard deviation of 0.02) (Fig. 2). This example highlights the power of combining the latest advanced multi-frequency STD NMR techniques alongside the proposed reduced relaxation matrix approach to develop a 3D model that strongly agrees with experimental data in solution.

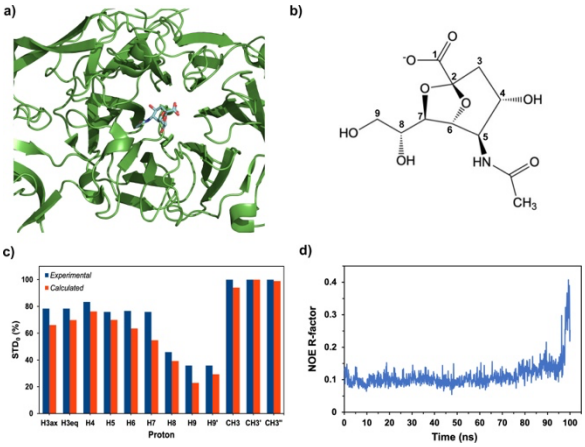

Figure 2. a) X-ray structure of the RgNanH-GH33 (coloured in green) in complex with 2,7-anhydro-Neu5Ac (in cyan; PDB 4X4A). The protein residues (at 12 Å distance from the ligand considered for the calculation of theoretical STD_0 factors are shown in sticks. **b)** 2D sketch of the ligand. The ligand numbering is shown next to each carbon atom. **c)** Comparison between calculated (red bars) and experimental (blue bars) STD_0 factors for the protons of the ligand. The NOE R-factor was 0.13 showing a very good agreement between the crystal and the solution state structure of the complex. In the insert, the naming scheme used is shown to identify the protons of the ligands. **d)** Evolution of the NOE R-factor of the 2,7-anhydro-Neu5Ac ligand over 100 ns of MD simulation.

Laminaribiose phosphorylase - α-glucopyronase-1-P complex.

The third system considered was the complex of the enzyme GH94 laminaribiose phosphorylase from *Paenibacillus sp.* YM-1 (PsLBP) with the ligand α-glucopyronase-1-phosphate (PDB code: 6GH2).[25] The PsLBP reverse reaction (synthesis of glycosidic linkages) is of high significance nowadays as it can be used as an alternative way of enzymatic glycosylation using sugar 1-phosphates as donor substrates.[21,29] For the RedMat calculation, we used a rotational correlation time of the protein of 68.5 ns, estimated with HYDRONMR[27] (GH94 has a molecular weight of ≈ 102 kDa), and a dissociation constant of 2000 μM. The concentrations of ligand and protein were 5000 μM and 50 μM, respectively, as per experimental conditions.[25]

Figure 3 shows the X-ray crystal structure of the α-Glc-1-phosphate-GH94 laminaribiose phosphorylase complex, and the comparison of experimental STD_0 values (blue bars) with those calculated using RedMat (red bars). The NOE R-factor based on the X-ray crystal structure was 0.17, again indicating a very good fit. Similarly, the MD simulation remained stable for the first 75 ns, with an average NOE R-factor of 0.23 (with a standard deviation of 0.03) (Fig. 3).

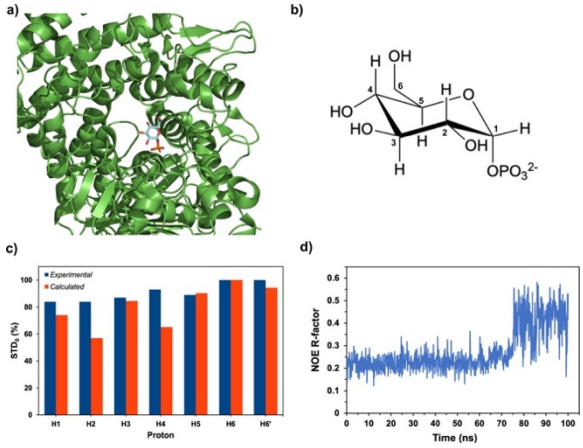

Figure 3. a) X-ray structure of the *PsLBP* (coloured in green) in complex with α-Glc-1-phosphate (in cyan; PDB 6GH2). The protein residues (at 12 Å distance from the ligand) considered for the calculation of theoretical STD_0 factors are shown in sticks. **b)** 2D sketch of the ligand. The ligand numbering is shown next to each carbon atom. **c)** Comparison between calculated (red bars) and experimental (red bars) STD_0 factors for the protons of the ligand. The NOE R-factor was 0.17 showing an excellent agreement between the crystal and the solution state structure of the complex. **d)** Evolution of the NOE R-factor of the α-Glc-1-phosphate ligand over 100 ns of MD simulation.

Conclusions

In conclusion, we have developed a reduced relaxation matrix theoretical approach that allows fast validation of static and dynamic 3D models of weak protein-ligand complexes based on experimentally determined STD NMR binding epitopes. The new algorithm allows the calculation of theoretical binding epitopes from STD_0 factors using the Cartesian coordinates of the receptor-ligand 3D structure, in the form of either a PDB structure or a MD trajectory. The practical implementation of this theoretical approach, in the form of a web application called RedMat, has been tested on different protein-ligand complexes of biological or biotechnological relevance. We demonstrate that RedMat is very robust, precise (as demonstrated by the low NOE R-factors obtained) and, remarkably, very fast (within seconds time scale per complex on a desktop computer). The development of such fast method for the calculation of theoretical STD_0 factors and for the validation of experimental STD_0 data obtained in solution is of major interest for both academic research and the pharmaceutical industry; this is because it can be readily used for rapidly screening a large range of protein-ligand complexes obtained from long MD simulations or docking calculations, as part of drug discovery pipelines. We foresee that RedMat will have a significant impact in the fields of structural biology of weak protein-ligand interactions and drug development.

Material and Methods

MD simulations

The initial atomic coordinates of each of the three protein-ligand complexes were obtained from their X-ray crystal structures deposited in the Protein Data Bank: intramolecular trans-sialidase from *R. gnavus* in complex with 2,7-anhydro-Neu5Ac (4X4A), laminaribiose phosphorylase from *Paenibacillus sp.* in complex with α-Glc-1-phosphate

(6GH2), and the broadly-neutralising anti-HIV-1 antibody 2G12 in complex with Man4 (6MSY).

Each system was parametrised using the AMBER ff14SB forcefield[30] for the protein and GAFF[31] (4X4A, 6GH2)/GLYCAM_06j-1[32] (6MSY) forcefields for the ligand. Ligand charges were determined using the *antechamber* software[33] using AM1-BCC level of theory. The systems were solvated with the TIP3P water model within a truncated octahedron bounding box buffered from the complex by 10 Å. Each system was neutralised with either Na+ or Cl- ions.

Conjugate gradient minimisation was run with 20 $kcal \cdot mol^{-1} \cdot Å^{-2}$ restraints on solute atoms, before repeating with no restraints. Each system was heated to 300 K over a period of 500 ps at constant volume, before equilibrating at constant pressure (1 atm) for a period of 2 ns. Production dynamics simulations were run for 100 ns each, saving a frame every 100 ps. In all cases periodic boundary conditions and the particle mesh Ewald method were applied. A Langevin thermostat with a collision frequency of 5 ps^{-1} and a Berendsen barostat with a relaxation time of 2 ps were used. The SHAKE algorithm was used to restrain all bonds involving hydrogen, allowing a timestep of 2 fs. A cutoff of 8 Å was used for all non-bonded interactions.

In the case of 4X4A, we observed excessive movement of the ligand within the protein binding site. In order to improve the simulations, NOE-based restraints were applied between protons for which experimental STDs had been observed and protein sidechains known from the X-ray crystal structure to be in close proximity. The restraints applied a 20 $kcal \cdot mol^{-1} \cdot Å^{-2}$ penalty for interatomic distances outside the 2 – 6 Å range, in agreement with the observed NOEs.

Availability: the RedMat app is freely available for academic users upon request to the authors for granting access to our server-based application.

Acknowledgements

This research was funded by the Ministerio de Ciencia e Innovación through the grant PID2019-109395GB-I00 (J.A.) and the Junta de Andalucia (Proyectos I+D+i) grant PY20_01176 (J.A.). J.A. and R.N. also acknowledge support of BBSRC, grant BB/P010660/1. S.W. and T.H. were funded by Biotechnology and Biological Sciences Research Council (BBSRC) Norwich Research Park Doctoral Training Grant BB/M011216/1. J.C.M-G. thanks the European Commission for an HORIZON-TMA-MSCA-PF-EF grant (Sweet2Gel, ID 101064251).

Keywords: Protein-ligand complex NMR validation • Weak affinity biomolecular interactions • STD NMR • Fragment-based drug discovery

[1] J. Orts, M. A. Wälti, M. Marsh, L. Vera, A. D. Gossert, P. Güntert, R. Riek, *J. Am. Chem. Soc.* **2016**, *138*, 4393–4400.

[2] L. Codutti, M. Grimaldi, T. Carlomagno, *J. Chem. Inf. Model.* **2017**, *57*, 1488–1498.

[3] S. Monaco, J. Ramírez-Cárdenas, A. T. Carmona, I. Robina, J. Angulo, *Pharmaceuticals* **2022**, *15*, DOI 10.3390/ph15081030.

[4] M. Pellecchia, D. S. Sem, K. Wüthrich, *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219.

[5] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.

[6] M. Mayer, B. Meyer, *Angew. Chemie Int. Ed.* **1999**, *38*, 1784–1788.

[7] A. W. Overhauser, *Phys. Rev.* **1953**, *92*, 411–415.

[8] C. Dalvit, P. Pevarello, M. Tatò, M. Veronesi, A. Vulpetti, M. Sundström, *J. Biomol. NMR* **2000**, *18*, 65–68.

[9] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *J. Biomol. NMR* **2001**, *21*, 349–359.

[10] S. Walpole, S. Monaco, R. Nepravishta, J. Angulo, in *Biol. NMR Part B* (Ed.: A.J.B.T.-M. in E. Wand), Academic Press, **2019**, pp. 423–451.

[11] P. Latorre-Muro, J. Baeza, R. Hurtado-Guerrero, T. Hicks, I. Delso, C. Hernández-Ruiz, A. Velázquez-Campoy, A. J. Lawton, J. Angulo, J. M. Denu, J. A. Carrodeguas, *J. Biol. Chem.* **2021**, *296*, 100205.

[12] V. Gabrielli, J. C. Muñoz-García, G. Pergolizzi, P. de Andrade, Y. Z. Khimyak, R. A. Field, J. Angulo, *Chem. – A Eur. J.* **2021**, *27*, 15688–15698.

[13] J. E. Watt, G. R. Hughes, S. Walpole, S. Monaco, G. R. Stephenson, P. C. Bulman Page, A. M. Hemmings, J. Angulo, A. Chantry, *Chem. – A Eur. J.* **2018**, *24*, 17677–17680.

[14] J. Angulo, P. M. Enríquez-Navas, P. M. Nieto, *Chem. – A Eur. J.* **2010**, *16*, 7803–7812.

[15] S. Monaco, L. E. Tailford, N. Juge, J. Angulo, *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 15289–15293.

[16] R. Nepravishta, S. Monaco, M. Distefano, R. Rizzo, P. Cescutti, J. Angulo, *Front. Mol. Biosci.* **2021**, *8*.

[17] M. Mayer, T. L. James, *J. Am. Chem. Soc.* **2004**, *126*, 4453–4460.

[18] O. Korb, H. M. Möller, T. E. Exner, *ChemMedChem* **2010**, *5*, 1001–1006.

[19] J. Magalhães, G. Annunziato, N. Franko, M. Pieroni, B. Campanini, A. Bruno, G. Costantino, *J. Chem. Inf. Model.* **2018**, *58*, 710–723.

[20] Y. Yuan, D. W. Bleile, X. Wen, D. A. R. Sanders, K. Itoh, H. Liu, B. M. Pinto, *J. Am. Chem. Soc.* **2008**, *130*, 3157–3168.

[21] J. D. Martínez, A. S. Infantino, P. Valverde, T. Diercks, S. Delgado, N.-C. Reichardt, A. Ardá, F. J. Cañada, S. Oscarson, J. Jiménez-Barbero, *Pharmaceuticals* **2020**, *13*, 179.

[22] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2002**, *155*, 106–118.

[23] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668–1688.

[24] P. M. Enríquez-Navas, M. Marradi, D. Padro, J. Angulo, S. Penadés, *Chem. – A Eur. J.* **2011**, *17*, 1547–1560.

[25] S. Kuhaudomlarp, S. Walpole, C. E. M. Stevenson, S. A. Nepogodiev, D. M. Lawson, J. Angulo, R. A. Field, *ChemBioChem* **2019**, *20*, 181–192.

[26] D. A. Calarese, H.-K. Lee, C.-Y. Huang, M. D. Best, R. D. Astronomo, R. L. Stanfield, H. Katinger, D. R. Burton, C.-H. Wong, I. A. Wilson, *Proc. Natl. Acad. Sci.* **2005**, *102*, 13372–13377.

[27] J. G. de l. Torre, M. L. Huertas, B. Carrasco, *J Magn Reson* **2000**, *147*, 138–146.

[28] L. E. Tailford, C. D. Owen, J. Walshaw, E. H. Crost, J. Hardy-Goddard, G. Le Gall, W. M. de Vos, G. L. Taylor, N. Juge, *Nat. Commun.* **2015**, *6*, 7624.

[29] M. KITAOKA, Y. MATSUOKA, K. MORI, M. NISHIMOTO, K. HAYASHI, *Biosci. Biotechnol. Biochem.* **2012**, *76*, 343–348.

[30] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

[31] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[32] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J.

Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.
[33] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.