J Biomol NMR (2011) 50:411–420
DOI 10.1007/s10858-011-9529-x

ARTICLE

Engineering $[Ln(DPA)_3]^{3-}$ binding sites in proteins: a widely applicable method for tagging proteins with lanthanide ions

Xinying Jia · Hiromasa Yagi · Xun-Cheng Su · Mitchell Stanton-Cook · Thomas Huber · Gottfried Otting

Received: 19 May 2011 / Accepted: 20 June 2011 / Published online: 23 July 2011

Abstract Paramagnetic relaxation enhancements from unpaired electrons observed in nuclear magnetic resonance (NMR) spectra present powerful long-range distance restraints. The most frequently used paramagnetic tags, however, are tethered to the protein via disulfide bonds, requiring proteins with single cysteine residues for covalent attachment. Here we present a straightforward strategy to tag proteins site-specifically with paramagnetic lanthanides without a tether and independent of cysteine residues. It relies on preferential binding of the complex between three dipicolinic acid molecules (DPA) and a lanthanide ion (Ln^{3+}), $[Ln(DPA)_3]^{3-}$, to a pair of positively charged amino acids whose charges are not compensated by negatively charged residues nearby. This situation rarely occurs in wild-type proteins, allowing the creation of specific binding sites simply by introduction of positively charged residues that are positioned far from glutamate or aspartate residues. The concept is demonstrated with the hnRNPLL RRM1 domain. In addition, we show that histidine- and arginine-tags present binding sites for $[Ln(DPA)_3]^{3-}$.

Keywords Lanthanide tag · NMR spectroscopy · Paramagnetic relaxation enhancements · RRM domain · $[Ln(DPA)_3]^{3-}$

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9529-x) contains supplementary material, which is available to authorized users.

X. Jia · H. Yagi · X.-C. Su · M. Stanton-Cook · T. Huber · G. Otting (✉)
Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia
e-mail: gottfried.otting@anu.edu.au

Introduction

Paramagnetic relaxation enhancements (PRE) generated by site-specific paramagnetic centres yield long-range structural information in NMR studies of biological macromolecules and have become a widely used tool to analyse the structure and dynamics of proteins and protein–protein complexes (Clore and Iwahara 2009; Simon et al. 2010; Volkov et al. 2010; Salmon et al. 2010). As paramagnetism depends on the presence of unpaired electrons and most proteins are diamagnetic, they need to be tagged with a paramagnetic reagent to elicit PREs in NMR spectroscopy. The most frequently used tag is the nitroxide radical MTSL [(1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-methanethiosulfonate] which reacts spontaneously with the thiol group of a cysteine residue, forming a disulfide bond. In proteins tagged with MTSL, the nitroxide is mostly located more than 6 Å from the cysteine sulfur, with a flexible tether in between. The position of the paramagnetic centre is thus usually not uniquely defined and the analysis of experimental PREs must take into account the flexibility of the tether, adding considerable uncertainty to any PRE-derived distance restraint to the paramagnetic centre (Battiste and Wagner 2000; Clore and Iwahara 2009). Alternative tags containing paramagnetic metal ions as paramagnetic centres also tend to contain long flexible linkers or require two-armed attachment, e.g., via two cysteine residues (Su and Otting 2010).

Attaching tags via disulfide bonds requires that strategically positioned free cysteine residues are available which often can be achieved only by substantial protein engineering. This poses problems if mutagenesis results in insoluble protein or if cysteine is critical for enzymatic activity. In principle, this problem can be circumvented by incorporating unnatural amino acids in the target protein

which are subsequently derivatized with a paramagnetic compound. Site-specific introduction of nitroxide radicals by this approach has recently been demonstrated and used for distance measurements by EPR (Fleissner et al. 2009). In this case, however, the overall distances between the nitroxide and the backbone atoms of the unnatural amino acid were much greater than in the case of MTSL-cysteine conjugates. Alternatively, unnatural metal-binding amino acids could be used to create site-specific metal binding sites that are much closer to the protein backbone. This approach has been demonstrated for Co^{2+} (Nguyen et al. 2011) but it has not yet worked with strongly paramagnetic metal ions such as Mn^{2+} or Gd^{3+} (Jones et al. 2010).

In an alternative approach, the protein can be mutated to create a metal binding site. It has been shown that lanthanide-binding sites can be created by site-directed mutagenesis that brings several carboxyl groups of acidic amino acid side chains into close proximity (Jones et al. 2008; Li et al. 2008). The attraction of this approach is that it relies exclusively on natural amino acids and no cysteine chemistry is required. In practice, however, the resulting charge concentration leads to a large decrease in thermal stability or even unfolding in the absence of divalent or trivalent metal ions (Li et al. 2008).

Here we show that proteins can be tagged with lanthanide ions by using positively charged amino acids to create specific binding sites for the highly negatively charged complex between three dipicolinic acid molecules (DPA) and a lanthanide ion (Ln^{3+}), $[Ln(DPA)_3]^{3-}$ (Fig. S1). This strategy is based on the observation that $[Ln(DPA)_3]^{3-}$ complexes bind preferentially to positively charged amino acid side chains (Pompidor et al. 2008; Su et al. 2009a). Among many proteins analysed in our laboratory, the N-terminal domain of the *E. coli* arginine repressor (ArgN; Sunnerhagen et al. 1997) bound most tightly with a dissociation constant K_d of 14 μM (Su et al. 2009a). Analysis of the metal position revealed that the complex was in contact with two positively charged residues (Lys45 and Arg48) located in the same α-helix, where both residues point their side chains in similar directions towards the $[Ln(DPA_3)]^{3-}$ complex. A preference for arginine and lysine side chains has also been noted in a crystal structure analysis of several proteins crystallized in the presence of $[Ln(DPA)_3]^{3-}$ complexes (Pompidor et al. 2008, 2010). Although almost every protein contains many positively charged side chains on its surface, clearly not all of them bind the $[Ln(DPA)_3]^{3-}$ complex with significant affinity. We hypothesized that (1) good binding sites involve at least two positively charged residues and (2) that proximity of negatively charged amino acid side chains may prevent binding by reducing the electrostatic attraction of the lanthanide complex.

The present work investigates several aspects of this design principle. (1) We demonstrate creation of $[Ln(DPA)_3]^{3-}$ binding sites by straightforward site-specific mutagenesis for the hnRNPLL RRM1 domain (Wu et al. 2008). (2) The RRM1 domain contains an RNA-binding site which can also bind $[Ln(DPA)_3]^{3-}$ complexes. We show that binding to the new engineered $[Ln(DPA)_3]^{3-}$ binding site can be assessed by calculating the difference in paramagnetic relaxation enhancements (PRE) between mutant and wild-type proteins using $[Gd(DPA)_3]^{3-}$. (3) We compare binding affinities with arginine and lysine. Finally, we show that commonly used histidine tags present binding sites and that a peptide tag with as few as two arginine residues can bind the $[Ln(DPA)_3]^{3-}$ complex.

Experimental section

Protein preparation: hnRNPLL RRM1

The plasmid used for the expression of hnRNPLL RRM1 has been described previously (Wu et al. 2008; Jia et al. 2009). The amino acid sequence of our construct differs for N-terminal seven and C-terminal six residues from that used for an early structure determination (PDB ID 1WEX). Plasmids encoding the mutant proteins used for the present work were prepared using a one-step site-directed mutagenesis protocol (Zheng et al. 2004; Liu and Naismith 2008). All ^{15}N-labelled samples were expressed by auto-induction methods and purified using a Ni–NTA column and gel filtration (Studier 2005; Jia et al. 2009). Protein yields were about 10 mg of purified protein per litre of cell culture. The NMR buffers used are listed in Table 1.

Preparation of $[Ln(DPA)_3]^{3-}$ complexes

4–10 mM stock solutions of the $[Ln(DPA)_3]^{3-}$ complexes were prepared by adding one equivalent of $LnCl_3$ or $Ln(NO_3)_3$ into an aqueous solution of 3.5 equivalents of DPA. The small excess of DPA was used to ensure the absence of any free lanthanide in solution.

NMR measurements

NMR spectra were recorded on Bruker Avance 600 and 800 MHz NMR spectrometers equipped with TCI cryoprobes. All NMR spectra were acquired at 25°C. PRE data were recorded at 800 MHz.

Binding affinities of $[Ln(DPA)_3]^{3-}$ complexes were determined by monitoring chemical shift changes observed in titrations with $[Tm(DPA)_3]^{3-}$ or $[Tb(DPA)_3]^{3-}$. These binding affinities were used as proxies of the binding affinities of $[Gd(DPA)_3]^{3-}$. The purpose of the $[Tm(DPA)_3]^{3-}$ and $[Tb(DPA)_3]^{3-}$ complexes was to enhance the magnitude of the chemical shift changes by pseudocontact shifts

Table 1 Dissociation constants of complexes between mutant RRM1 domains and $[Ln(DPA)_3]^{3-}$ in different buffers

Mutants	Buffers[a]	K_d/mM[b]
RRM1 D62R	50 mM Tris·HCl, 50 mM NaCl[c]	10 ± 1
RRM1 D62R/E66Q	50 mM Tris·HCl, 50 mM NaCl[c]	4.1 ± 0.4
RRM1 D62R/K65R/E66Q	50 mM Tris·HCl, 50 mM NaCl[c]	3.2 ± 0.3
RRM1 D62R/E66Q	10 mM PIPES	1.3 ± 0.2
RRM1 D62R/K65R/E66Q	15 mM PIPES	2.7 ± 0.7
RRM1 D62R/E66Q	50 mM HEPES	1.4 ± 0.2

[a] All buffers were at pH 7.2 in 90% H_2O/10% D_2O and contained 3 mM β-mercaptoethanol

[b] The dissociation constants were measured by monitoring the chemical shift changes of the amide proton of Ser15 during titration with $[Tm(DPA)_3]^{3-}$ or $[Tb(DPA)_3]^{3-}$. The fits neglected competitive binding by any of the other sites of the protein. Therefore, true K_d values would be somewhat smaller than those listed in the table

[c] Standard NMR buffer used for 3D structure determination (Wu et al. 2008)

(PCSs), making it possible to measure binding affinities even when the chemical shifts remained unchanged upon titration with diamagnetic $[Y(DPA)_3]^{3-}$. In addition, PCSs tend to highlight specific binding sites, as PCSs are efficiently averaged to zero if the DPA complex reorientates due to weak binding or due to binding to a flexible peptide segment at the N- or C-terminus of the protein. In contrast, PREs induced by $[Gd(DPA)_3]^{3-}$ can induce PREs even if the complex is free in solution (Yagi et al. 2010).

Transverse ^1H relaxation rates were measured by two time point (0 and 5.32 ms) measurements (Iwahara et al. 2007). 0.59 mM solutions of the wild-type and mutant RRM1 domains were used. PREs of amides close to the D62R mutation site were measured using a $[Gd(DPA_3)]^{3-}$ to RRM1 ratio of 0.06:1. PREs of amides at greater distance from the mutation site were measured using a $[Gd(DPA_3)]^{3-}$ to RRM1 ratio of 1.25:1 and the data of Fig. 2a, b, are presented for this ratio, extrapolating PREs measured at low $[Gd(DPA)_3]^{3-}$ concentration to fill in data of residues that could not be measured at high $[Gd(DPA)_3]^{3-}$ concentration due to excessive line broadening.

Fitting the metal position in the hnRNPLL RRM1 domain

PREs of amide protons were converted into distance restraints using the equation

$$\Gamma_2 = kr^{-6} \quad (1)$$

where Γ_2 denotes the transverse PRE, k the proportionality factor, and r the distance of the nuclear spin from the paramagnetic centre. The proportionality factors were calculated as described (Yagi et al. 2010), estimating the rotational correlation time τ_R from the molecular weight of the protein ($\tau_R = 6$ ns; Su et al. 2007b), using an electronic relaxation time of 8 ns, and assuming that the residence time of specifically bound $[Ln(DPA)_3]^{3-}$ complex was much longer than τ_R. The resulting k value was 3.5×10^9 Å^6/s. Prior to fitting a metal position, the PREs were scaled to account for limited occupancy of the binding site with the $[Gd(DPA)_3]^{3-}$ complex. The occupancy was calculated from the dissociation constant determined by monitoring chemical shift changes upon titration with a 4.1 mM solution of $[Tb(DPA)_3]^{3-}$. The fitting was performed without considering the possibility of van der Waals violations between protein and lanthanide complex.

The metal position established by PREs from $[Gd(DPA)_3]^{3-}$ was used in fits of magnetic susceptibility tensors generated by other paramagnetic $[Ln(DPA)_3]^{3-}$ complexes, using the program Numbat (Schmitz et al. 2008).

Results

Potential $[Ln(DPA)_3]^{3-}$ binding sites in natural proteins

Considering that only few if any $[Ln(DPA)_3]^{3-}$ binding sites exist in natural proteins (Su et al. 2009a), single arginine or lysine residues are clearly not sufficient to bind the lanthanide complex. Also the occurrence of two arginine or lysine residues that have their charged groups positioned within 6 Å of each other is not sufficient for binding, as shown by the GCN4 leucine zipper, where this situation occurs several times along the helices without creating binding sites for the $[Ln(DPA)_3]^{3-}$ complex (Yagi et al. 2010). Inspection of the structure of the GCN4 leucine zipper (O'Shea et al. 1991) shows, however, that each lysine and arginine side chain can engage in a salt bridge with a carboxyl group from glutamate or aspartate. This is different from the situation in ArgN, where the $[Ln(DPA)_3]^{3-}$ binding residues Lys45 and Arg48 are not capable of forming any salt bridge with a negatively charged residue.

In a simple analysis we searched protein structures for the presence of a pair of close arginine or lysine side chains (defined as a maximal distance of 6 Å between Lys N^{ζ} or Arg C^{ζ} atoms) without any nearby carboxyl group from aspartate or glutamate (defined by Lys N^{ζ} and Arg C^{ζ} atoms within 6 Å of the carboxyl carbons; Kumar and Nussinov 2002). First, the analysis focused on ten different proteins for which we had NMR results regarding their capacity of binding the $[Ln(DPA)_3]^{3-}$ complex (ArgN; Sunnerhagen et al. 1997), hnRNPLL RRM1 domain (Wu et al. 2008), ERp29 (Liepinsh et al. 2001), ubiquitin (Vijay-Kumar et al. 1987), hen egg-white lysozyme (Cheetham et al. 1992), the C-terminal domain of *E. coli* DnaG (Su et al. 2006), T4 lysozyme (Bell et al. 1991), the intracellular domain (ICD) of the p75 neurotrophin receptor (Liepinsh et al. 1997), the 14 kDa C-terminal domain of the τ subunit of the *E. coli* DNA polymerase III (τ_C14; Su et al. 2007a) and the West Nile virus NS2B-NS3 protease (WNVpro; Aleshin et al. 2007). Only Lys45/Arg48 of ArgN and Lys16/Lys43 of T4 lysozyme were predicted to bind the $[Ln(DPA)_3]^{3-}$ complex.

This result agreed well with experiment (Su et al. 2009a) for ArgN (to which $[Ln(DPA)_3]^{3-}$ binds strongly) and for p75 ICD, τ_C14 and WNVpro (to which $[Ln(DPA)_3]^{3-}$ does not bind). For the other proteins, weak binding of $[Ln(DPA)_3]^{3-}$ has been observed at sites not predicted by the simple analysis above (Su et al. 2009a), showing that additional factors can cause binding. In the case of the RRM1 domain, the search for two positively charged amino-acid side chains did not predict the experimentally observed binding of the $[Ln(DPA)_3]^{3-}$ complex to the RNA binding site which may be explained by the presence of a positively charged histidine residue. In the case of T4 lysozyme, a binding site was predicted between Lys16 and Lys43 which is not supported by experiments. Inspection of the 3D structure shows that the ends of the side chains of these two residues are close only because their side chains point towards each other in the crystal structure.

Importantly, the analysis indicated that the occurrence of two positively charged side chains in close proximity without nearby negatively charged side chains is rare in natural proteins. To confirm this result, we repeated the analysis for 9406 proteins containing at least 80 residues in the non-redundant PDB chain set found at ftp://ftp.ncbi.nih.gov/mmdb/nrtable/nrpdb.latest (version from 03/06/2011). Two thirds of the polypeptide chains (average length 224 residues) had no Arg/Lys pair without Asp/Glu residues in close proximity (by the criterion defined above) and/or only 11% had two or more such Arg/Lys pairs. This opens the door to engineering $[Ln(DPA)_3]^{3-}$ binding sites by strategic mutation of residues to arginine and lysine.

hnRNPLL RRM1 domain

We used the hnRNPLL RRM1 domain (Wu et al. 2008) as a test case for engineering a new $[Ln(DPA)_3]^{3-}$ binding site by site-directed mutagenesis. Binding of $[Gd(DPA)_3]^{3-}$ was monitored by PREs. In the wild-type protein, binding was observed in three different regions of the protein (Fig. 1a, b, e). The first binding site was produced by the N-terminal His_6-tag, resulting in the disappearance of the ^{15}N-HSQC cross-peaks of Thr3, Gly5, and Gly6 in the presence of $[Gd(DPA)_3]^{3-}$. The second binding site was produced by residues of the RNA-binding β-sheet of the RRM1 domain, leading to disappearance of the ^{15}N-HSQC cross-peaks of the side chains of Gln53 and Asn86. The third binding site led to substantial broadening of the cross-peaks of Thr89 and Lys91 (Fig. S2) which could be explained by binding to the side chains of Lys91 and Arg92. Monitoring the ^{15}N-HSQC cross-peak of Thr89 in a titration experiment with $[Tm(DPA)_3]^{3-}$ yielded a dissociation constant K_d of about 8 mM (Fig. S3a).

To introduce a new $[Ln(DPA)_3]^{3-}$ binding site in the hnRNPLL RRM1 domain, we mutated Asp62 to arginine. Asp62 is located in the α-helix and residue 65 is a lysine so that the D62R mutation mimics the situation in ArgN, positioning positively charged side chains in positions i and $i + 3$ of the solvent exposed face of an amphiphilic α-helix. The mutant site in the RRM1(D62R) domain bound $[Ln(DPA)_3]^{3-}$ complexes only weakly (K_d of about 10 mM), as measured by chemical shift changes of Ser15 upon addition of $[Tm(DPA)_3]^{3-}$ (Fig. S3b). To enhance the binding affinity, we removed the charge of Glu66 by preparing the double-mutant RRM1(D62R/E66Q). The double-mutant bound the $[Ln(DPA)_3]^{3-}$ complex much better, with a K_d to the engineered site in the low millimolar range (Fig. S3c and Table 1).

Next we tested the preference of the $[Ln(DPA)_3]^{3-}$ complex to bind to arginine rather than lysine side chains by preparing the triple-mutant RRM1(D62R/K65R/E66Q). Table 1 shows that the binding affinities of $[Ln(DPA)_3]^{3-}$ to the triple and double mutants were very similar, indicating that the binding interactions were predominantly driven by electrostatic attraction rather than, e.g., shape complementarity. This result was confirmed by the observation that buffers with lower ionic strength significantly improved the binding affinity (Table 1).

Despite good binding affinities, however, only small PCSs could be generated by paramagnetic $[Ln(DPA)_3]^{3-}$ complexes. For example, even a 10-fold excess of $[Tm(DPA)_3]^{3-}$ rarely produced PCSs larger than 0.1 ppm in the D62R/E66Q double mutant (Fig. S4). The small size of the PCSs is clear evidence for significant averaging of the PCSs by reorientational motions of the $[Tm(DPA)_3]^{3-}$ complex with respect to the protein.

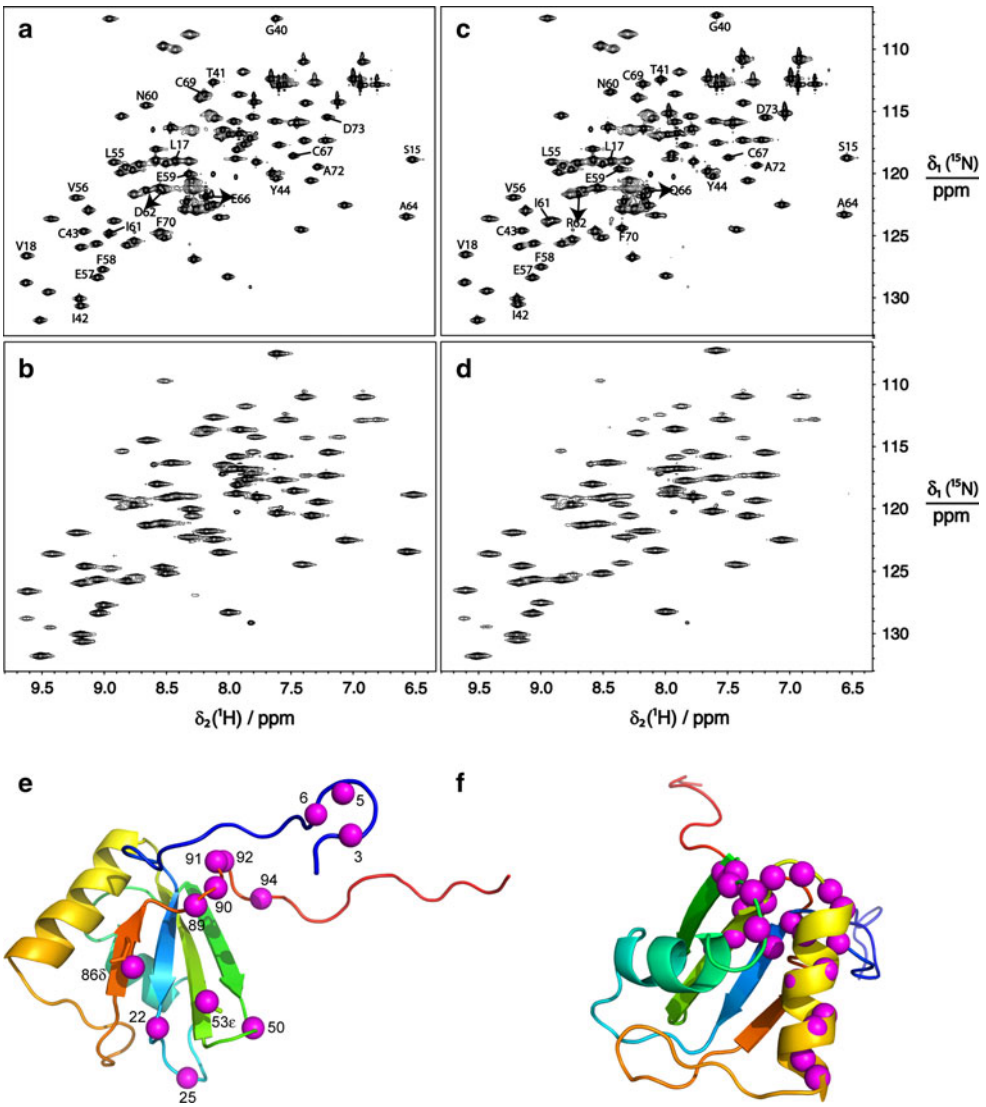

Fig. 1 PREs due to $[Gd(DPA)_3]^{3-}$ binding to the hnRNPLL RRM1 domain. **a** and **b** ^{15}N-HSQC spectra of a 0.59 mM solution of the wild-type RRM1 domain in **a** the absence and **b** the presence of 0.7 mM $[Gd(DPA)_3]^{3-}$. These spectra were recorded with the standard NMR buffer used for 3D structure determination (50 mM TrisHCl, 50 mM NaCl, 3 mM β-mercaptoethanol; Wu et al. 2008). Resonance assignments identify cross-peaks which display significantly larger PREs in the D62R/E66Q mutant than in the wild-type. *Arrows* highlight the cross-peaks of residues 62 and 66 which were targeted by the mutation. **c** and **d** Same as (**a** and **b**), except for the RRM1(D62R/E66Q) double mutant. **e** Structure of the RRM1 domain (PDB ID 1WEX) showing the positions of residues displaying large PREs in the spectrum of the wild-type protein (Fig. S2). The PREs of Gln53 and Asn86 are of side chain amides. **f** Structure of the RRM1 domain showing the positions of residues with larger PREs in the D62R/E66Q mutant than in the wild-type protein

To determine the binding site of the $[Ln(DPA)_3]^{3-}$ complex, we performed quantitative PRE measurements using $[Gd(DPA)_3]^{3-}$. Quantitative measurements are important as PREs are also generated by very weakly bound or unbound $[Gd(DPA)_3]^{3-}$ complexes (Yagi et al. 2010) as well as by $[Gd(DPA)_3]^{3-}$ associated with His-tags (see below). In general, PREs are measured as the relaxation rates measured for the paramagnetic sample minus the relaxation rates measured for a corresponding diamagnetic sample. PREs arising from the mutation site can be singled out by subtracting the PREs measured for the wild-type protein from those measured for the mutant protein. This approach requires a total of eight NMR spectra, even if relaxation rates are measured by no more than two spectra. In a simpler protocol, however, the relaxation rates in the paramagnetic wild-type sample can be subtracted from those of the paramagnetic mutant sample without recording any relaxation rates of the diamagnetic samples, because the relaxation rates of mutant and wild-type proteins would be practically the same in the diamagnetic state. As expected, both methods yielded closely similar results, except that the second approach produced smaller error bars as it involves fewer experimental data. Figure 2b shows the PREs from the engineered binding site determined by subtracting the relaxation rates of the wild-type RRM1 domain from those of the D62R/E66Q mutant (Fig. 2a), both of which were measured in the presence of the same concentration of $[Gd(DPA)_3]^{3-}$. Fitting the PREs to the structure of the RRM1 domain by a least square fit using equation 1 yielded good agreement between experimental and back-calculated PREs (Fig. 2c). The fit showed that the $[Gd(DPA)_3]^{3-}$ complex binds between the side chains of Arg62 and Lys65 as expected (Fig. 3). Using the protein structure and these metal coordinates to fit

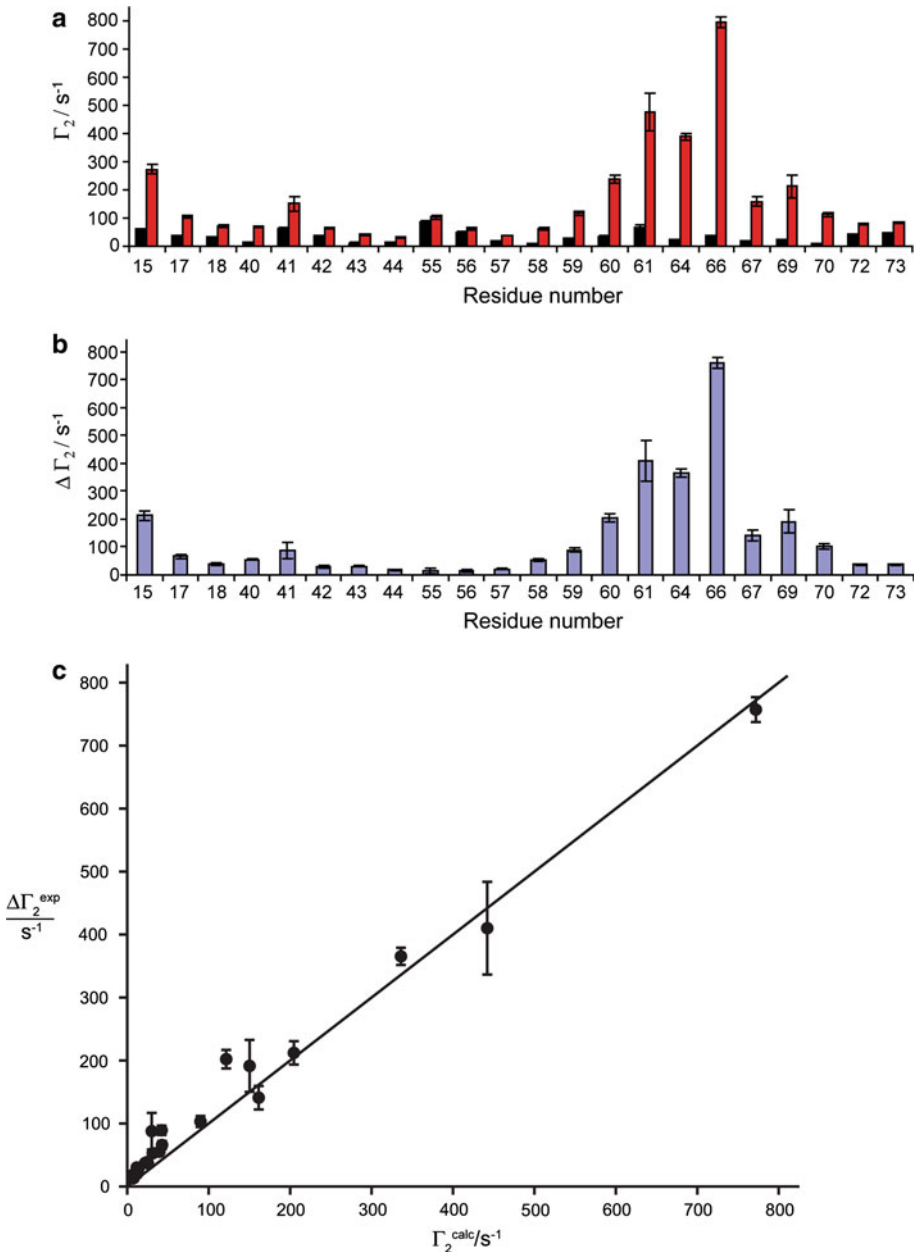

Fig. 2 Paramagnetic enhancements of transverse relaxation rates of amide protons in the RRM1 domain. **a** PREs measured for the wild-type (*black*) and D62R/E66Q mutant (*red*) of the RRM1 domain in the standard NMR buffer. The data pertain to a $[Gd(DPA)_3]^{3-}$ to protein ratio of 1.25:1. Only the PREs of those residues are compiled that were significantly larger in the mutant than in the wild-type protein. **b** Difference in transverse PREs of mutant and wild-type protein shown in (**a**). **c** Plot of experimental versus back-calculated PREs following a least-squares fit based on Eq. 1. Back-calculation used model 12 of the NMR structure (PDB accession code 1WEX) which gave the best fit

$\Delta\chi$-tensors to the PCSs observed with $[Tb(DPA)_3]^{3-}$, $[Tm(DPA)_3]^{3-}$ and $[Yb(DPA)_3]^{3-}$ yielded very small tensors (Table S1), indicating significant reorientational motions of the bound metal complex.

$[Ln(DPA)_3]^{3-}$ binding peptide tags

Our results obtained with the RRM1 domain indicated that His_6 tags are capable of binding $[Ln(DPA)_3]^{3-}$ complexes. Since many proteins are produced with His_6 tags for facilitated purification, we assessed their binding affinity in more detail, using peptides derivatized with an amide group at the C-terminus or an acetyl group at the N-terminus to mimic C- and N-terminal His tags, respectively. The K_d values obtained by titrations with $[Yb(DPA)_3]^{3-}$ ranged between 0.3 mM for a His_6 peptide with a C-terminal amide at pH 6.5 to about 4 mM for a His_6 peptide with an N-terminal acetyl group at pH 7.2 in the presence of 100 mM NaCl (Fig. S5 and Table 2). The presence of salt greatly attenuated the binding affinity. The importance of electrostatics was further underlined by the small K_d value (about 0.2 mM) found for the complex between $[Yb(DPA)_3]^{3-}$ and the peptide MRAAAR with an amide group at the C-terminus.

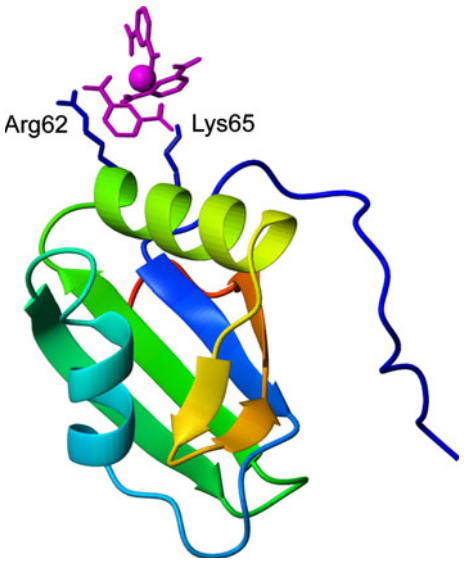

Fig. 3 Structure of the RRM1 domain (PDB accession code 1WEX) with bound $[Gd(DPA)_3]^{3-}$ (*magenta*), showing the position of Gd^{3+} in the RRM1(D62R/E66Q) mutant determined by the fit of Fig. 2b. The side chains of Arg62 and Lys65 are highlighted in *blue*. The orientations of the $[Gd(DPA)_3]^{3-}$ complex and the side chain of Arg62 were modelled to avoid steric clashes

Discussion

Generation of lanthanide binding sites by site directed mutagenesis may be expected to be most easily achieved by introducing a number of negatively charged amino acid residues. This has been tried. It was found that the resulting charge concentration readily leads to unfolded protein (Li et al. 2008).

This disappointing result may be explained by the natural electrostatic potential surrounding the polypeptide backbone. Specifically, carboxyl groups of aspartate and glutamate residues in helical and extended secondary structure are in an area of positive electrostatic potential arising from peptide bond polarity (Spassov et al. 1997). While this stabilizes the negative charges of these residues, it also makes them less available for binding to positively charged metal ions. In contrast, arginine and lysine carry their positive charges at the ends of long side chains, where they escape the positive potential surrounding the polypeptide backbone (Spassov et al. 1997). As the side chains of arginine and lysine are longer than those of aspartate and glutamate, clusters of these positively charged side chains are expected to be less detrimental to protein stability. Therefore, our current approach, where the lanthanide is embedded in a negatively charged complex for binding to positively charged amino-acid side-chains, promises to be less detrimental to protein stability.

Our results show that two positively charged amino acid residues are sufficient to produce a specific binding site for lanthanide tri-DPA complexes, allowing the creation of lanthanide-binding sites with remarkable ease following a rational design. We attribute the success in designing specific binding sites to the fact that the interactions are predominantly driven by a positive electrostatic potential of the protein and not very dependent on the structural details of the interaction. In principle, further stabilization of the interactions may come from cation–π interactions (Ma and Dougherty 1997) aided by induced fits of the amino acid side chains to optimize the interactions. Nonetheless, we find that in designing a binding site for $[Ln(DPA)_3]^{3-}$, it is not necessary to predict the molecular details of the interaction and to go beyond the coarse calculations of electrostatic potentials that are readily performed for 3D protein structures by many existing software packages. This suggests that electrostatic attraction is indeed the overriding determinant of the interaction with $[Ln(DPA)_3]^{3-}$ and that any additional optimization of the interaction is almost always readily achieved by arginine and lysine residues which carry their positive charges at the ends of long side chains that are usually flexible. We note that the binding affinities obtained by provision of two isolated arginine or lysine residues were not as strong as

Table 2 Dissociation constants of complexes between peptides and $[Yb(DPA)_3]^{3-}$

Peptide	Buffers[a]	K_d/mM[b]
His_6-NH_2[c]	20 mM MES, pH 7.2	0.25 ± 0.08
N-acetyl-His_6[d]	20 mM MES, pH 6.5	0.6 ± 0.2
N-acetyl-His_6[d]	20 mM MES, pH 7.2, 100 mM NaCl[c]	4 ± 4
MRAAAR-NH_2[e]	20 mM MES, pH 6.5	0.22 ± 0.06

[a] All buffers were in D_2O

[b] The dissociation constants were measured by monitoring the chemical shift changes of resolved ^1H NMR signals

[c] Containing a C-terminal amide group to model an N-terminal His_6 tag

[d] Containing a N-terminal acetyl amide to model a C-terminal His_6 tag

[e] Containing a C-terminal amide group. Peptide sequence given by single-letter amino acid symbols

those observed previously for ArgN (Su et al. 2009a). This may be explained by additional stabilizing interactions with Gln7 in ArgN which would be difficult to engineer in other proteins.

While the flexibility of amino acid side chains helps to optimize the interactions between protein and $[Ln(DPA)_3]^{3-}$, residual mobility of the lanthanide complex in the adduct tends to average PCSs to zero. This, together with the incomplete population of the designed binding site with $[Ln(DPA)_3]^{3-}$ complex (about 20%), produced only small magnetic susceptibility anisotropy ($\Delta\chi$) tensors. Nonetheless, the binding site proved to be very well suited for PRE measurements using $[Gd(DPA)_3]^{3-}$. As PREs do not vanish by averaging, they are much less sensitive to local flexibility than PCSs.

Generating PREs by binding $[Gd(DPA)_3]^{3-}$ to specifically designed $[Ln(DPA)_3]^{3-}$ binding sites has unique advantages. (1) The fast exchange between bound and free $[Ln(DPA)_3]^{3-}$ complex allows scaling of the PREs by using $[Gd(DPA)_3]^{3-}$ at different concentrations (Yagi et al. 2010). At low concentrations, PREs can be measured for nuclear spins near the paramagnetic centre, enabling accurate determination of the position of the lanthanide. At higher concentrations, long-range paramagnetic effects can be measured for structural investigations. This strategy is not available for covalently bound tags which always broaden the NMR signals of nearby nuclear spins beyond detection, making it hard to determine the position of the paramagnetic centre with high accuracy and allowing quantitative PRE measurements only in a narrow distance range far from the paramagnetic centre.

(2) Increasing the concentration of $[Gd(DPA)_3]^{3-}$ would usually obstruct the PRE measurements by PREs from additional binding sites (Yagi et al. 2010). This problem is circumvented in the present approach by subtracting the relaxation rates of the wild-type protein from those of the mutant, accounting for PREs from alternative binding sites as well as from unbound relaxation agent. Importantly, there is no penalty in the number of NMR spectra that must be recorded, as both relaxation rates are measured in the presence of the relaxation agent and no relaxation rate measurements are necessary in the diamagnetic state, i.e. the diamagnetic reference is simply replaced by the relaxation rates of the wild-type protein in the paramagnetic state. Pre-existing $[Ln(DPA)_3]^{3-}$ binding sites in a protein are thus no obstacle to evaluating PREs from the specifically engineered binding site.

(3) The relatively weak binding affinity of $[Ln(DPA)_3]^{3-}$ complexes to the engineered protein sites has several advantages. (a) Protein structure and solubility are barely affected by weakly binding ligands present at low concentration. (b) PREs scale linearly with the concentration of bound $[Gd(DPA)_3]^{3-}$ only if the exchange rate between bound and free lanthanide complex is fast compared with the PRE rate (Iwahara and Clore 2006; Yagi et al. 2010). (c) As the population of bound $[Gd(DPA)_3]^{3-}$ must be known to translate PREs into distances from the metal, the binding affinity must be determined. This is most easily achieved in the fast exchange regime by monitoring the gradual chemical shift changes in response to titration with a PCS inducing $[Ln(DPA)_3]^{3-}$ complex. (d) A weak binding affinity ensures that the concentration of free $[Gd(DPA)_3]^{3-}$ is similar in mutant and wild-type samples. Using the same concentration of protein and $[Gd(DPA)_3]^{3-}$ complex for the relaxation measurements of mutant and wild-type samples, the engineered binding site in the mutant slightly reduces the amount of $[Gd(DPA)_3]^{3-}$ available for binding at competing sites. While this reduces the chances of false positives (PREs at unrelated binding sites), data analysis is easier if the concentrations remain similar. This is the case when the concentration of free paramagnetic complex is only little perturbed by the engineered site. (e) Measuring the distance r between a nuclear spin and the paramagnetic centre is robust, as the metal position can be accurately determined from short-range PREs measured at low $[Gd(DPA)_3]^{3-}$ concentration and errors in PRE measurements translate into only small distance errors due to the r^{-6} distance dependence of the PREs.

Mutations in positions i and $i + 3$ of an α-helix present a straightforward way to place two positively charged residues near each other with their side chains projecting into similar directions. To test the design principles of $[Ln(DPA)_3]^{3-}$ binding sites also for non-helical secondary structure, we engineered a $[Ln(DPA)_3]^{3-}$ binding site into the West Nile virus NS2B-NS3 protease (WNVpro) by preparing the double mutant E101R/K107R. NMR spectra were recorded in the presence of an inhibitor (carbamimidothioic acid (2,5-dimethyl-1,4-phenylene)bis-(methylene) ester; Ekonomiuk et al. 2009). Titration with $[Tm(DPA)_3]^{3-}$ indicated a dissociation constant of about 2.9 mM, whereas the wild-type protein showed no evidence of binding other than at the N-terminal His_6 tag. PREs measured with the $[Gd(DPA)_3]^{3-}$ complex clearly highlighted the new binding site (which is located in a solvent exposed β-hairpin of NS3), but also the amides of Gly32, Ser33 and Tyr34 which showed about one quarter of the PREs of Arg101 and Lys107, although they are about 17 Å from the engineered metal binding site in all crystal structures of WNVpro (data not shown; Erbel et al. 2006; Aleshin et al. 2007; Robin et al. 2009). An implausibly large structural change would be required to explain the PREs by a single metal binding site. At present, we have no explanation for this result. We cannot rule out the possibility that altering the electrostatics of a protein produces unexpected long-range effects that are not manifested in chemical shift changes. Alternatively, $[Ln(DPA)_3]^{3-}$ may have mediated

protein dimerization which is always possible when proteins are chemically modified. Intermolecular PREs have previously been reported in a protein that showed no evidence of self-association until it was tagged with a nitroxide radical (Su et al. 2009b).

The finding that histidine tags are $[Ln(DPA)_3]^{3-}$ binding sites means that proteins with a histidine tag can be used for PRE measurements without any further modification. This applies to a very large group of proteins, as His_6 tags are commonly used to facilitate protein purification. In situations where histidine tags are not an option, fusion with an arginine-rich peptide tag presents an alternative way of generating a $[Ln(DPA)_3]^{3-}$ binding site. Clearly, the information content of PREs from a bound $[Ln(DPA)_3]^{3-}$ complex is greatest, if the peptide tag is fused to the protein with a minimal number of intervening flexible residues, although this would also have the greatest potential to affect the electrostatics of the protein.

Conclusion

Engineering site-specific lanthanide binding sites on proteins at strategically chosen positions is an important prerequisite for accessing the power of paramagnetic NMR for structure analysis, but none of the currently available strategies of chemical protein modification with lanthanide tags are universally applicable. Generating specific binding sites for $[Ln(DPA)_3]^{3-}$ complexes by a small number of point mutations is attractive as, in contrast to the majority of lanthanide tags that are designed for covalent binding to cysteine residues, the $[Ln(DPA)_3]^{3-}$ complex is compatible with any number of cysteine residues or disulfide bonds in the target protein. In our hands, mutating proteins to display only a single accessible cysteine residue on the surface frequently leads to insoluble or poorly folded proteins, whereas introduction of one or two arginine residues always maintained the structure and solubility characteristics of the protein. In addition, the possibility of scaling PREs simply by changing the concentration of $[Gd(DPA)_3]^{3-}$ allows measurement of PREs of nuclear spins near the metal and assists in accurate determination of the metal position (Yagi et al. 2010). This is an advantage over covalently bound tags that bind lanthanides tightly, where any bound Gd^{3+} ion becomes a non-scalable relaxation source that causes excessive line broadening for nuclear spins in the vicinity of the metal ion. Engineering $[Ln(DPA)_3]^{3-}$ binding sites by mutation thus presents a most promising tool for generating paramagnetic sites. Even in the presence of multiple binding sites, PREs from bound $[Gd(DPA)_3]^{3-}$ will generate useful structural restraints, for example on the binding mode of a bound ligand molecule.

Acknowledgments We thank Ms Inca Hutchinson for titration measurements of peptides and the Biomolecular Resource Facility at the ANU for peptide synthesis. Financial support by the Australian Research Council, including a Future Fellowship for T. H., is gratefully acknowledged.

References

Aleshin AE, Shiryaev SA, Strongin AY, Liddington RC (2007) Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. Protein Sci 16:795–806

Battiste JL, Wagner G (2000) Utilization of site-directed spin labelling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry 39: 5355–5365

Bell JA, Wilson K, Zhang XJ, Faber HR, Nicholson H, Matthews BW (1991) Comparison of the crystal structure of bactiophage T4 lysozyme at low, medium, and high ionic strengths. Proteins Struct Funct Genet 10:10–21

Cheetham JC, Artymiuk PJ, Phillips DC (1992) Refinement of an enzyme complex with inhibitor bound at partial occupancy. Hen egg-white lysozyme and tri-N-acetylchitotriose at 1.75 Å resolution. J Mol Biol 224:613–628

Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev 109:4108–4139

Ekonomiuk D, Su XC, Ozawa K, Bodenreider C, Lim SP, Yin Z, Keller TH, Beer D, Patel V, Otting G, Caflisch A, Huang D (2009) Discovery of a non-peptidic inhibitor of West Nile virus NS3 protease by high-throughput docking. PLoS Negl Trop Dis 3:e356

Erbel P, Schiering N, D'Arcy A, Renatus M, Kroemer M, Lim SP, Yin Z, Keller TH, Vasudevan SG, Hommel U (2006) Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. Nat Struct Mol Biol 13:372–373

Fleissner MR, Brustad EM, Kalai T, Altenbach C, Cascio D, Peters FB, Hideg K, Peuker S, Schultz PG, Hubbell WL (2009) Site-directed spin labelling of a genetically encoded unnatural amino acid. Proc Natl Acad Sci USA 106:21637–21642

Iwahara J, Clore GM (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. Nature 440: 1227–1230

Iwahara J, Tang C, Clore GM (2007) Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. J Magn Reson 184:185–195

Jia X, Ozawa K, Loscha K, Otting G (2009) Glutarate and N-acetyl-L-glutamate buffers for cell-free synthesis of selectively ^{15}N-labelled proteins. J Biomol NMR 44:59–67

Jones LM, Yang W, Manniccia AW, Harrison A, van der Merwe A, Yang JJ (2008) Rational design of a novel calcium-binding site adjacent to the ligand-binding site on CD2 increases its CD48 affinity. Protein Sci 17:439–449

Jones DH, Cellitti SE, Hao X, Zhang Q, Jahnz M, Summerer D, Schultz PG, Uno T, Geierstanger BH (2010) Site-specific labelling of proteins with NMR-active unnatural amino acids. J Biomol NMR 46:89–100

Kumar S, Nussinov R (2002) Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys J 83:1595–1612

Li S, Yang W, Maniccia AW, Barrow D Jr, Tjong H, Zhou HX, Yang JJ (2008) Rational design of a conformation-switchable Ca^{2+}- and Tb^{3+}-binding protein without the use of multiple coupled metal-binding sites. FEBS J 275:5048–5061

Liepinsh E, Ilag LL, Otting G, Ibáñez C (1997) NMR structure of the death domain of the p75 neurotrophin receptor. EMBO J 16:4999–5005

Liepinsh E, Baryshev M, Sharipo A, Ingelman-Sundberg M, Otting G, Mkrtchian S (2001) Thioredoxin-fold as a homodimerization module in the putative chaperone ERp29: NMR structures of the domains and experimental model of the 51 kDa homodimer. Structure 9:457–471

Liu H, Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol 8:91–100

Ma JC, Dougherty DA (1997) The cation–π interaction. Chem Rev 97:1303–1324

Nguyen THD, Ozawa K, Stanton-Cook M, Barrow R, Huber T, Otting G (2011) Generation of pseudocontact shifts in protein NMR spectra with a genetically encoded Co^{2+}-binding amino acid. Angew Chem Int Ed 50:692–694

O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the GCN4 leucine zipper, a 2-stranded, parallel coiled coil. Science 254:539–544

Pompidor G, D'Aléo A, Vicat J, Toupet L, Giraud N, Kahn R, Maury O (2008) Protein crystallography through supramolecular interactions between a lanthanide complex and arginine. Angew Chem Int Ed 47:3388–3391

Pompidor G, Maury O, Vicat J, Kahn R (2010) A dipicolinate lanthanide complex for solving protein structures using anomalous diffraction. Acta Cryst D Biol Crystallogr 66:762–769

Robin G, Chappell K, Stoermer MJ, Hu S, Young PR, Fairlie DP, Martin JL (2009) Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. J Mol Biol 385:1568–1577

Salmon L, Nodet G, Ozenne V, Yin GW, Jensen MR, Zweckstetter M, Blackledge MJ (2010) NMR characterization of long-range order in intrinsically disordered proteins. J Am Chem Soc 132:8407–8418

Schmitz C, Stanton-Cook MJ, Su XC, Otting G, Huber T (2008) Numbat: an interactive software tool for fitting $\Delta\chi$-tensors to molecular coordinates using pseudocontact shifts. J Biomol NMR 41:179–189

Simon B, Madl T, Mackereth CD, Nilges M, Sattler M (2010) An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. Angew Chemie Int Ed 49:1967–1970

Spassov VZ, Ladenstein R, Karshikoff AD (1997) Optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins. Protein Sci 6:1190–1196

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207–234

Su XC, Otting G (2010) Paramagnetic labelling of proteins and oligonucleotides for NMR. J Biomol NMR 46:101–112

Su XC, Schaeffer PM, Loscha KV, Gan PHP, Dixon NE, Otting G (2006) Monomeric solution structure of the helicase binding domain of *Escherichia coli* DnaG primase. FEBS J 273:4997–5009

Su XC, Jergic S, Keniry MA, Dixon NE, Otting G (2007a) Solution structure of domains IVa and V of the τ subunit of *Escherichia coli* DNA polymerase III and interaction with the α subunit. Nucl Acids Res 35:2825–2832

Su XC, Jergic S, Ozawa K, Burns ND, Dixon NE, Otting G (2007b) Measurement of dissociation constants of high-molecular weight protein–protein complexes by transferred ^{15}N-relaxation. J Biomol NMR 38:65–72

Su XC, Liang H, Loscha KV, Otting G (2009a) $[Ln(DPA)_3]^{3-}$ is a convenient paramagnetic shift reagent for protein NMR studies. J Am Chem Soc 131:10352–10353

Su XC, Ozawa K, Qi R, Vasudevan SG, Lim SP, Otting G (2009b) NMR analysis of the dynamic exchange of the NS2B cofactor between open and closed conformations of the West Nile virus protease. PLoS Negl Trop Dis 3:e561

Sunnerhagen M, Nilges M, Otting G, Carey J (1997) Solution structure of the DNA-binding domain and model for the binding of multifunctional hexameric arginine repressor to DNA. Nat Struct Biol 4:819–826

Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 194:531–544

Volkov AN, Bashir Q, Worrall JAR, Ullmann GM, Ubbink M (2010) Shifting the equilibrium between the encounter state and the specific form of a protein complex by interfacial point mutations. J Am Chem Soc 132:11487–11495

Wu Z, Jia X, de la Cruz L, Su XC, Marzolf B, Troisch P, Zak D, Hamilton A, Whittle B, Yu D, Sheahan D, Bertram E, Aderem A, Otting G, Goodnow CC, Hoyne GF (2008) Memory T cell RNA rearrangement programmed by heterogeneous nuclear ribonucleoprotein hnRNPLL. Immunity 29:863–875

Yagi H, Loscha KV, Su XC, Stanton-Cook M, Huber T, Otting G (2010) Tunable paramagnetic relaxation enhancements by $[Gd(DPA)_3]^{3-}$ for protein structure analysis. J Biomol NMR 47:143–153

Zheng L, Baumann U, Reymond JL (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucl Acids Res 32:e115