Studies of High Molecular Weight Systems by NMR Spectroscopy

Shereen Jabar

Research School of Chemistry

October 2018

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University

Australian National University

Declaration

This thesis is composed of my original work and does not contain material previously published or written by another person, except where due reference has been made in the text. The content of this thesis is the result of work that I have carried out since the commencement of my research degree candidature and has not been previously submitted for another degree or diploma in any university or tertiary institution.

Shereen Jabar

October 2018

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Gottfried Otting. I am really thankful for his excellent guidance, support, patience, enthusiasm and invaluable knowledge, without which this work could have not been completed.

I would like to thank Professor Thomas Huber for his support and guidance regarding computational modeling. I am profoundly grateful to our collaborators Associate Professor Bim Graham and his group at Monash University for providing all the chemical tags and unnatural amino acids used in my projects and Professor Daniella Goldfarb and her group at the Weizmann Institute of Science for the EPR measurements.

My thanks also go to all the present and past members of Otting and Huber groups for their support and friendship.

My sincere thanks to the Research School of Chemistry, the Australian National University and the Australian Research Council for their generous funding.

Last but not least, I would like to acknowledge with gratitude, the support and love of my family and friends.

Abstract

Proteins are essential components for all living cells. For better understanding of proteins and how they work, it is essential to know their 3D structure as well as how they interact with other molecules and their dynamics. Nuclear magnetic resonance (NMR) spectroscopy is a versatile technique to study biomolecules, most notably proteins, in solution. NMR provides structural information at atomic resolution and, compared to the other techniques with similar resolution, the measurements can be carried out under near-physiological conditions that are meaningful for the investigation of dynamics.

A great number of unnatural amino acids can now be incorporated into proteins both *in vivo* and by cell-free protein synthesis at genetically encoded positions. Isotopically labelled unnatural amino acids present outstanding probes for site-specific studies of proteins by NMR. In this thesis, the unnatural amino acid ^{13}C-O-*tert*-butyltyrosine (^{13}C-Tby) was incorporated as an NMR probe to study the proteins IMP-1 and the *E. coli* single-stranded DNA binding protein (SSB).

IMP-1 is a clinically important metallo-β-lactamase which catalyzes the hydrolysis of almost all β-lactam antibiotics. In order to develop inhibitors against this enzyme it is important to understand its structure and dynamics. Pseudocontact shifts (PCSs) provide long-range structural information on biological macromolecules. In Chapter 2, PCSs were measured to study the conformation of an active site loop of IMP-1 and its change upon binding of a ligand. This loop is important as it lines the substrate binding site and contributes to substrate specificity. The results suggest that the binding of an inhibitor induces two different conformations in the loop which are in slow equilibrium between each other.

E. coli SSB is a homotetramer of molecular mass 76 kDa. It binds single-stranded DNA (ssDNA) with high affinity and little sequence specificity in two main binding modes, named $(SSB)_{65}$ and $(SSB)_{35}$. Chapter 3 describes studies of ssDNA-SSB complexes at high and low salt concentrations by solution NMR spectroscopy. Furthermore, experiments were conducted to determine the binding polarity of ssDNA

on SSB. The results obtained are in broad agreement with the unusual binding polarity reported by the crystal structure of a complex between *E. coli* SSB and a poly-deoxycytidine oligomer.

In addition to using the *tert*-butyl group of a Tby residue as an NMR probe, the present thesis introduced a novel chemical tag containing a trimethylsilyl (TMS) group. Chapter 4 demonstrates the ability of the *tert*-butyl and TMS groups to deliver site-specific information in high-molecular weight systems without any isotope labelling. The approach was illustrated by measuring intermolecular nuclear Overhauser effects (NOEs) in the 95 kDa complex between SSB and ssDNA.

List of publications

1. Pearce, B.J.G., **Jabar, S.**, Loh, C.-T., Szabo, M., Graham, B. and Otting, G. (2017) Structure restraints from heteronuclear pseudocontact shifts generated by lanthanide tags at two different sites. *J. Biomol. NMR* 68, 19-32.
   I contributed in preparation of isotope labelled protein samples for NMR measurements.

2. **Jabar, S**., Adams, L., Wang, Y., Aurelio, L., Graham, B. and Otting, G. (2017) Chemical tagging with *tert*-butyl and trimethylsilyl groups for measuring intermolecular nuclear Overhauser effects in a large protein-ligand complex. *Chem. Eur. J.* 23, 13033-13036.
   I carried out all the protein sample preparations including protein expression, purification and ligation with the TMS tag and the NMR measurements and analysis. The chemical synthesis of the tag molecules was performed by the group of A/Prof. Bim Graham at Monash University, including Drs Luke Adams and Luigi Aurelio.

Table of contents

CHAPTER 1

General Introduction

Proteins are essential for life, carrying out many important functions. In order to understand how proteins work, their three-dimensional (3D) structure must be known at atomic resolution. Currently, this can be achieved by techniques like X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (cryoEM). Most 3D structures of proteins and protein-ligand complexes have been solved by X-ray crystallography, while an increasing number of very large complexes are being determined by cryoEM. NMR spectroscopy is unique for delivering high-resolution structural information in solution and at ambient temperature, i.e. under conditions close to the physiological state. This enables detailed studies of protein dynamics and their interactions with binding partners under biologically relevant conditions.

NMR spectroscopy offers a large number of local probes, as each proton on the macromolecule results in an individual resonance in the spectrum, which can be resolved using higher-dimensional (2D, 3D and 4D) techniques. The first step in protein structure determination by NMR involves chemical shift assignment of the ^{1}H, ^{15}N and ^{13}C NMR resonances of the backbone atoms. Nowadays, this process can be carried out quite rapidly and often it is automated (Atreya and Szyperski, 2005; Freeman and Kupce, 2003). The chemical shifts determined carry information on structural features of backbone and side chain conformations and secondary structure elements of the protein (Wagner et al., 1983; Williamson and Asakura, 1993; Case, 1995).

As a second step, stereospecific assignments of prochiral centres and torsion angle restraints need to be determined. These can be derived from three-bond scalar couplings together with nuclear Overhauser effects (NOEs), both of which are short-range effects (< 5 Å; Wüthrich, 1990). Finally, NOEs are used to obtain the long-range restraints that are necessary to define the 3D structure.

More recently, residual dipolar couplings (RDCs) have been introduced which are short-range effects (bond orientation relative to an alignment tensor) containing long-range structure information (as the alignment tensor is a global property of the protein; Tjandra et al., 1997). RDCs depend on some alignment of the molecules in the magnetic field. This can be achieved by an anisotropic magnetic susceptibility, which is usually small for diamagnetic proteins, by the use of dilute liquid crystals that spontaneously orient in a magnetic field (Prestegard and Kishore, 2001), or by labelling the protein with a paramagnetic metal ion endowed with an anisotropic magnetic susceptibility tensor (Otting, 2010). Paramagnetic metal ions also generate direct long-range effects, such as paramagnetic relaxation enhancements (PRE; Furuita et al., 2015) or pseudocontact shifts (PCS; Otting, 2010).

PCSs are unique in not only providing long-range distance restraints but also angular information regarding the polar angles of nuclear spins in the tensor frame of the paramagnetic centre. It has been shown that PCS data alone are sufficient to determine the 3D structure of proteins with the help of structure prediction software (Yagi et al., 2013; Pilla et al., 2016; Pilla et al., 2017).

1.1 Paramagnetism in NMR

Paramagnetism arises from unpaired electrons. Interactions between an unpaired electron and nuclear spins generate pronounced long-range effects in the NMR spectra of biological systems, which provide information about their structure and dynamics with respect to the paramagnetic centre. The main paramagnetic effects are pseudocontact shifts, paramagnetic relaxation enhancements, residual dipolar couplings and cross-correlations between Curie-spin and dipole-dipole relaxation (CCR; Figure 1.1; Otting, 2008).

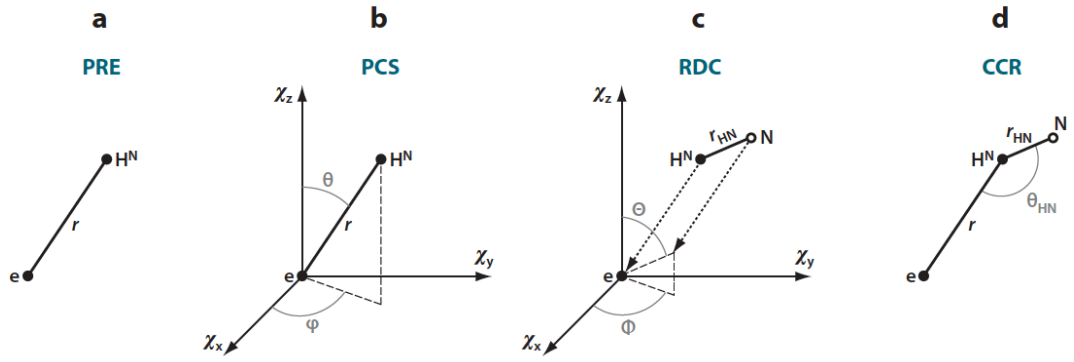


Figure 1.1 Structural information contained in four different paramagnetic effects, illustrated with a nuclear spin, H^N, and a paramagnetic centre denoted e. (a) Paramagnetic relaxation enhancement depends on the distance r between electron and nuclear spin. (b) The pseudocontact shift depends on the polar coordinates of the nuclear spin relative to the magnetic susceptibility tensor c. (c) The residual dipolar coupling between a ^1H and a ^{15}N spin depends on the internuclear distance r_{HN} and the polar angles between the internuclear vector and the c tensor. (d) The cross-correlation effect between Curie spin and ^1H-^{15}N dipole-dipole relaxation depends on the internuclear distance r_{HN}, the distance between nuclear spin and paramagnetic centre (r), and the angle between r_{HN} and r. Figure reproduced from Otting (2010).

All these effects can be measured in heteronuclear single-quantum coherence (HSQC) spectra recorded of the paramagnetic state and the reference diamagnetic state of the biomolecule. The paramagnetic effects can be described by the magnetic susceptibility tensor χ, which is defined by three principal axes (χ_x, χ_y and χ_z). Only the PRE effect is observed for systems with isotropic χ tensors, while all four effects occur in systems with anisotropic χ tensors. The anisotropy of the χ tensor ($\Delta\chi$ tensor) can be characterized by axial and rhombic components (Bertini et al., 2002; Otting, 2010):

$$\Delta\chi_{ax} = \chi_z - (\chi_x + \chi_y)/2 \qquad (1.1)$$

$$\Delta\chi_{rh} = \chi_x - \chi_y \qquad (1.2)$$

Except for RDCs, the magnitude of the paramagnetic effects depends on the distance r between the metal ion and nuclear spins. PREs lead to broadening of the NMR signals of nuclear spins near the paramagnetic centre. As the PRE effect is

proportional to r^{-6}, PREs decay rapidly compared with PCSs and the CCR effect, which show a r^{-3} distance dependence. Therefore, even though nuclear spins near the paramagnetic centre get broadened beyond detection, nuclear spins that are sufficiently far from can give rise to measurable PCSs (Otting, 2010). The present thesis focuses on the measurement of PCSs to study the conformational dynamics of the protein IMP-1 described in Chapter 2.

1.2 Pseudocontact shifts

PCSs originate from the magnetic dipole field generated at the site of a nuclear spin by rapidly relaxing unpaired electrons associated with a metal ion. This is a through-space effect, which can be described by the following equation (Bertini et al., 2002)

$$\Delta\delta^{PCS} = 1/(12\pi r^3)\,[\Delta\chi_{ax}\,(3\cos^2\theta - 1) + 1.5\,\Delta\chi_{rh}\sin^2\theta\cos 2\varphi] \qquad (1.3)$$

where r is the distance between the unpaired electron and the nuclear spin, $\Delta\chi_{ax}$ is the axial component and $\Delta\chi_{rh}$ is the rhombic component of the $\Delta\chi$ tensor, and θ and φ are the polar angles of the nuclear spin with respect to the principal axes of the $\Delta\chi$ tensor. PCSs are independent of the magnetic field when measured in parts per million (ppm) and can be observed as far as 40 Å from the paramagnetic centre (Allegrozzi et al., 2000).

In ^{15}N-HSQC spectra of proteins, PCSs can be observed as chemical shift changes of the cross-peaks along diagonal lines as both the proton and the nitrogen of each ^{1}H-^{15}N group experience similar PCSs. Small deviations from the diagonal shifts occur due to residual anisotropic chemical shifts, which arise for nuclear spins with large chemical shift anisotropy (CSA) in molecules that are subject to weak paramagnetic alignment in the magnetic field (John et al., 2005). In practice, this correction is significant for ^{15}N chemical shifts of backbone amides and ^{13}C chemical shifts of carbonyl groups, but small for other backbone atoms. Therefore PCSs are usually determined as the difference between the chemical shifts of the paramagnetic and diamagnetic states of ^{1}H rather than heteronuclear spins, eliminating the need for this correction.

If the 3D structure of the protein carrying the paramagnetic metal ion is known, the experimentally measured PCSs can be used to fit the $\Delta\chi$ tensor. Eight parameters need to be determined, including the x, y and z coordinates defining the location of the metal ion, three Euler angles α, β and γ defining the orientation of the $\Delta\chi$ tensor with respect to the protein coordinates, and the axial and rhombic components of the $\Delta\chi$ tensor. The $\Delta\chi$ tensor can be depicted in the form of PCS isosurfaces overlaid onto the structure of the protein (Figure 1.2). This gives a graphical representation of the principal axes of the $\Delta\chi$ tensor and its orientation with respect to the protein (Otting, 2010).

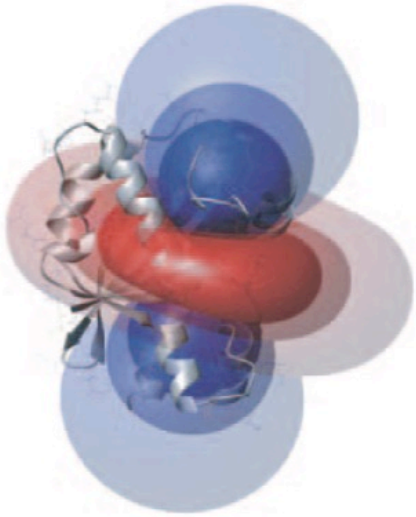


Figure 1.2 Isosurfaces depicting the PCSs induced by a Dy^{3+} ion plotted on a ribbon representation of the crystal structure of the protein ε186. Blue and red surfaces identify the spatial locations of positive and negative PCSs (±3, ±1.5 and ±0.5 ppm), respectively. The PCS isosurfaces constitute a representation of the $\Delta\chi$ tensor. Reproduced from Otting (2010).

1.3 Introducing the paramagnetic metal ion

Among the many paramagnetic transition metal ions, high-spin Co^{2+} stands out for generating relatively large PCSs combined with relatively small PREs. Compared to the transition metals, lanthanide ions can generate much larger PCSs and therefore are popular for measuring PCSs. As the trivalent lanthanide ions contain their unpaired electrons in f-orbitals, which do not interact as strongly with ligand orbitals as the d-

orbitals of transition metals, the delocalization of the unpaired electron density onto the ligand atoms is minimal (Nitsche and Otting, 2017).

The chemical properties of all lanthanide ions are very similar. Therefore, once a lanthanide binding site has been established, it is suitable to bind any lanthanide ion with similar affinity. La^{3+}, Y^{3+} and Lu^{3+} ions are suitable diamagnetic ions to measure the diamagnetic state. The radii of these ions are comparable to those of the paramagnetic lanthanides at the start, centre and end of the lanthanide series, respectively, which makes them good choices for diamagnetic references as required for PCS and PRE measurements. Figure 1.3 gives an overview of the paramagnetic properties of paramagnetic lanthanide ions.

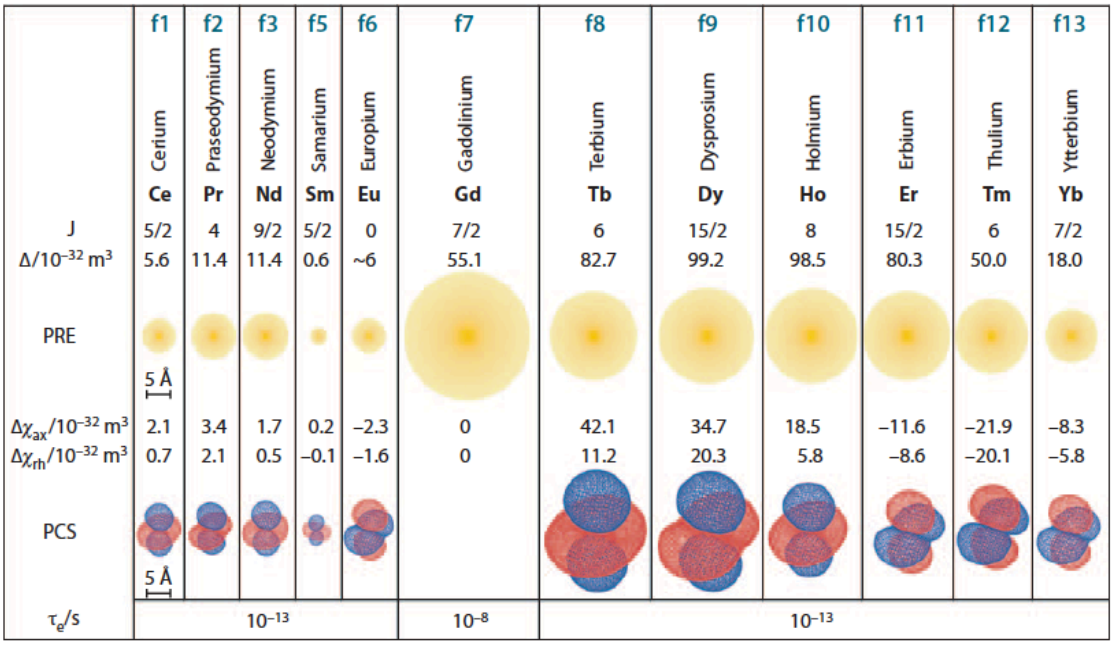


Figure 1.3 Paramagnetic properties of lanthanides. The radii of the yellow spheres indicate the distance from the metal ion where the ^{1}H-NMR signals of a protein with a rotational correlation time of 15 ns are expected to broaden by 80 Hz due to PRE on an 800 MHz NMR spectrometer. Isosurfaces represent PCSs of ±5 ppm based on the tensors reported for calbindin D_{9k} (Bertini et al., 2001). Typical electronic relaxation times for a magnetic field strength of 18.8 T are indicated at the bottom. Reproduced from Pintacuda et al. (2007).

Since lanthanides have no known biological function, natural proteins rarely display strong binding sites for lanthanide ions. It has been shown that lanthanide binding sites can be generated by displacing a native Ca^{2+} ion from calcium-binding proteins. This has been demonstrated, for example, for the Ca^{2+}-binding proteins calbindin D_{9k} (Allegrozzi et al., 2000) and calmodulin (Bertini et al., 2003). Most proteins, however, do not contain natural metal binding sites. Therefore, other methods for introducing lanthanide ions into proteins have been developed.

It has been shown that the EF-hand motif of a calcium-binding protein can be optimized to create a lanthanide binding peptide (LBP; Ma and Opella, 2000; Franz et al., 2003; Nitz et al., 2003; Nitz et al. 2004). LBPs generate PCSs following attachment to a cysteine residue via formation of a disulfide bond (Su et al., 2006; Su et al., 2008), but mobility of the linker between LBP and protein can decrease the PCSs substantially by averaging between PCSs of different magnitude and sign. In order to obtain sizeable paramagnetic effects it is important to have less mobile LBPs attached onto protein. This has been achieved by anchoring the LBP to the protein at two points, which greatly reduces its mobility and leads to much larger PCSs in the protein. Two-point attachments have been achieved by inserting LBPs into loop regions of proteins as, e.g., in the protein interleukin-1β (Barthelmes et al., 2011) or the immunoglobulin G binding protein (Barb et al., 2012; Barb and Subedi, 2016), or by combining fusion to the N- or C-terminus of the target protein with an additional disulfide bond to a cysteine residue as demonstrated for, e.g., the B1 immunoglobulin-binding domain of protein G (GB1; Saio et al., 2009).

It is important that any lanthanide binding tag (LBT) has a high affinity for lanthanide ions as well as minimal mobility relative to the protein in order to obtain sizeable PCSs. Most LBTs are designed to be attached to proteins by forming a covalent bond with a cysteine residue either via a disulfide or a thioether bond. This includes most of the synthetic LBTs that are based on EDTA, DTPA, TAHA and DOTA chelating groups. Alternatively, the LBT can be attached by a copper-catalyzed click reaction with a site-specifically incorporated 4-azidophenylalanine residue (Nitsche and Otting, 2017). In this case, it is not necessary to mutate the protein to contain only a single cysteine residue at the site targeted for tagging (Loh et al., 2013), which can require multiple mutations of cysteine residues present in the wild-type protein. The C1

tag and its opposite enantiomer, the C2 tag, have successfully been used in different conformational studies of proteins (Graham et al., 2011; de la Cruz et al., 2011, Yagi et al., 2013, Chen at al., 2014; Chen et al., 2016; Pearce et al., 2017; Mahawaththa et al., 2017). In this thesis, the C2 tag has been used in Chapter 2 to study the active-site loop conformation of the metallo-β-lactamase IMP-1 by PCS measurements.

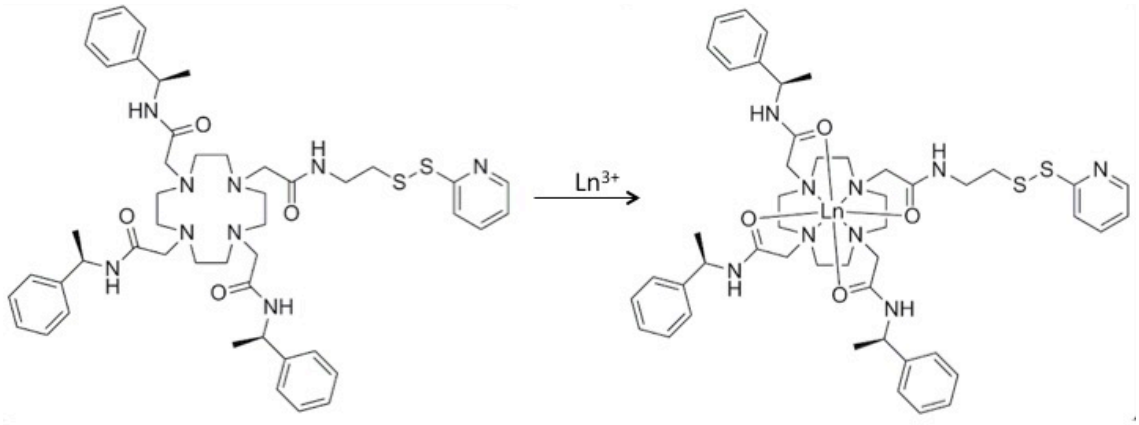


Figure 1.4 Chemical structures of the C2 tag and its lanthanide complex. The tag contains three chiral amide pendants contributing to the lanthanide binding site and a nonchiral pendant for attachment to a cysteine residue on the target protein.

A very elegant solution would be presented by site-specific incorporation of a lanthanide binding unnatural amino acid. To date, a genetic encoding system has been reported only for a single unnatural amino acid, 2-amino-3-(8-hydroxyquinolin-3-yl)propanoic acid (HQ-Ala), that can bind lanthanides with considerable affinity (Lee et al., 2009). Unfortunately, proteins containing HQ-Ala precipitate upon titration with lanthanide ions (Jones et al., 2010).

PCSs produced by lanthanides attached to proteins have been used for 3D structure refinement of proteins as well as 3D structure determination. For example, PCSs and ^{1}H-^{15}N RDCs measured for the calmodulin N60D mutant, where a Ca^{2+} ion can be substituted by a lanthanide ion, have been used to refine the 3D structures of two calmodulin-peptide complexes. The PCS data proved particularly useful for the structural refinement of the overall fold of the protein (Bertini et al., 2003). Using PCSs for determining the 3D structures of proteins is more difficult as $\Delta\chi$ tensor fits rely on the availability of the 3D structure. This problem has been solved in programs such as

PCS-Rosetta and GPS-Rosetta by generating model structures to fit the Δχ tensors and refining only the best-fitting models (Schmitz et al., 2012; Yagi et al., 2013). A particularly attractive application of PCSs is for determining the structures of protein-ligand complexes, where Δχ tensors can be fitted to the known 3D structure of the protein and PCS data of the ligand can be used to determine its binding mode (ligand binding site, relative orientation of the ligand with respect to the protein and ligand conformation; John et al., 2006). In the same way, PCSs can be used to study the dynamics of local parts of proteins.

1.4 Protein production

Structural studies of proteins by NMR spectroscopy require relatively large amounts of pure protein. The required proteins are most frequently produced by over-expression systems in *Escherichia coli*. *In vivo* protein expression is ideal for making large quantities of protein, but is less suitable for producing selectively isotope-labelled samples (due to "isotope scrambling", i.e. the shuffling of isotope-labelled chemical groups between different amino acids effected by the natural amino acid metabolism) or samples containing expensive unnatural amino acids (due to protein yields that are quite low relative to the amino acid provided). For these situations, cell-free protein synthesis can be more suitable.

Cell-free protein synthesis (CFPS) systems employ cell extracts to produce proteins from amino acids. The crude lysates obtained from microbial, plant or animal cells contain the necessary elements for transcription, translation, protein folding and energy metabolism such as ribosomes, aminoacyl-tRNA synthetases, translation initiation and elongation factors, ribosome release factors, nucleotide recycling enzymes, metabolic enzymes, chaperones, foldases, etc. (Nirenberg and Mattaei, 1961). The cell extract can thus be used to synthesize the target protein in a coupled transcription/translation system, where the DNA encoding the target protein is provided together with T7 RNA polymerase and nucleotide triphosphates to produce mRNA, and amino acids and a source of ATP are provided to drive the protein translation reaction (Carison et al., 2012). In this thesis, an *E. coli* CFPS system established in-house (Guignard et al., 2002; Ozawa et al., 2004; Ozawa et al., 2006; Wu et al., 2007;

Apponyi et al., 2008; Jia et al., 2009; Su et al., 2011; Ozawa et al., 2012; Loscha et al., 2012) was used to produce the *E. coli* single-stranded DNA-binding protein samples discussed in Chapters 3 and 4. *E. coli* cell extracts are relatively easy to prepare and high protein yields can be obtained from relatively small amounts of amino acids (Torizawa et al., 2004)

1.5 Isotopic labelling of proteins

The main isotopes usually used in protein NMR spectroscopy includes, ^{1}H, ^{2}H, ^{13}C and ^{15}N. Amongst these isotopes, except for ^{1}H, which has natural abundance of greater than 99.9%, others must be introduced into proteins by other methods.

E. coli is the most commonly used bacterial host for expressing recombinant proteins with isotopic labelling, as *E. coli* cell cultures are readily grown with $^{15}NH_4$ salts as the only nitrogen source and ^{13}C-glucose as the only carbon source, both of which are relatively inexpensive (Markley et al., 1968; Marley et al., 2001). Costs can be saved further by growing the cells initially in an unlabelled rich medium to obtain a high cell density followed by a transfer of the cells to a labelled medium before over-expression (McIntosh and Dahlquist, 1990). In Chapter 2, the metallo-β-lactamase IMP-1 was uniformly labelled with ^{15}N by *in vivo* expression, using a M9 medium with ^{15}N-ammonium chloride as the sole ^{15}N source.

CFPS is an excellent method to produce selectively labelled proteins from a mixture of 20 amino acids, where a specific amino acid type can be supplied in isotopically labelled form. Compared to *in vivo* protein expression, isotope scrambling is strongly suppressed in CFPS because the amino acid metabolism is no longer fully functional in cell extracts.

In the same way as proteins can be produced from amino acid mixtures and their metabolic precursors, a protein can also be site-specifically labelled chemically by incorporating an unnatural amino acid, which may also be labelled isotopically. Chapters 2, 3 and 4 report work performed with a ^{13}C-labelled unnatural amino acid,

which was incorporated into a protein to obtain site-specific information by NMR spectroscopy without having to assign all other NMR resonances of the protein.

1.6 Site-specific incorporation of unnatural amino acids

Site-specific introduction of unique functional groups into proteins presents a powerful tool for structure/function studies. This can be accomplished either by attachment of a chemical probe onto a reactive amino-acid side-chain as in cysteine residues or by incorporating unnatural amino acids (UAAs), which can contain different functional groups. Genetic encoding of UAAs is a particularly elegant way to selectively modify and label proteins without adversely affecting the protein structure.

Genetic encoding of a UAA relies on a pair of tRNA/aminoacyl-tRNA synthetase (aaRS), which is orthogonal to any of the endogenous tRNA/aaRS pairs. Many examples of orthogonal aaRS have been evolved, so that the aaRS can only load the desired UAA onto the orthogonal tRNA without recognizing any of the naturally occurring amino acids (Figure 1.5). The resulting UAA-tRNA complex then incorporates the UAA into proteins in response to a unique codon (the most common codon used for this is the amber stop codon, TAG, but other stop codons and quadruplet codons can also be used; Lin et al, 2008; Liu et al., 2010).

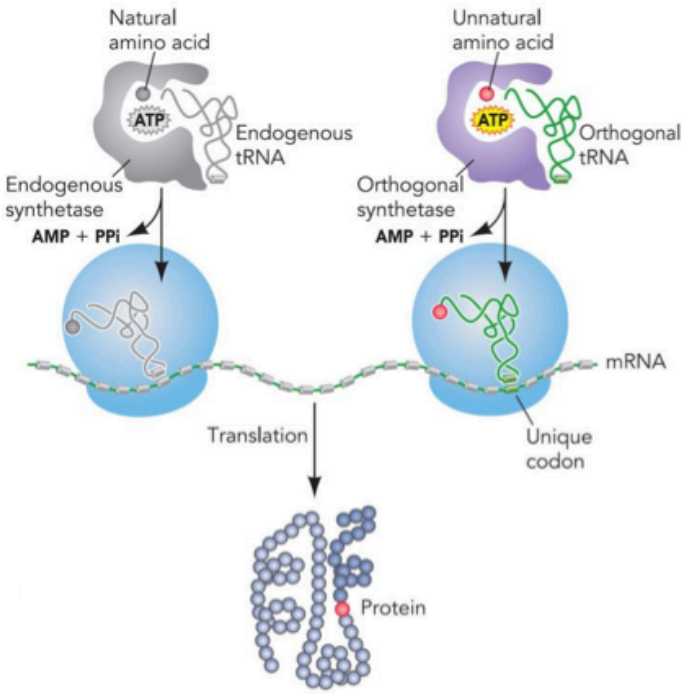


Figure 1.5 Schematic diagram of an orthogonal aminoacyl-tRNA synthetase/tRNA pair system for the incorporation of an unnatural amino acid into a protein. The orthogonal synthetase (purple) aminoacylates the orthogonal tRNA (green) with an unnatural amino acid (red sphere) and does not cross-react with any of the endogenous synthetases and tRNA (grey). The aminoacylated orthogonal tRNA then travels to the ribosome (blue) and incorporates the unnatural amino acid in response to a unique codon. Reproduced from Lin et al. (2008).

Four main aaRS enzymes have been developed for UAA incorporation: (i) the *Methanococcus janaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/$tRNA_{CUA}$ pair, which is orthogonal in *E. coli* (Xie and Schultz, 2005); (ii) pyrrolysyl-tRNA synthetase (PylRS)/$tRNA_{CUA}$ pairs from certain *Methanosarcina* organisms, which are orthogonal in bacteria and eukaryotic cells (Neumann et al., 2008); (iii) the *E. coli* tyrosyl-tRNA synthetase (*Ec*TyrRS)/$tRNA_{CUA}$ pair and the *E. coli* leucyl-tRNA synthetase (*Ec*LeuRS)/$tRNA_{CUA}$ pair, which are orthogonal in eukaryotic cells (Chin, 2003; Chin et al., 2003, Wu et al., 2004).

The TyrRS/$tRNA_{CUA}$ pair derived from the archaebacterium *M. jannaschii* was the first aaRS/$tRNA_{CUA}$ pair reported to be orthogonal with regard to the *E. coli* expression machinery (Wang et al., 2001). This aaRS interacts only little with the

anticodon of its target tRNA, which makes it possible to mutate the anticodon to CUA for recognition of the amber stop codon. A wide range of variants have been developed to adopt the amino-acid binding pocket of this aaRS to accommodate different UAAs (Neumann et al., 2010a, Young et al., 2011).

A major problem associated with reassignment of the amber codon to incorporate UAAs is the presence of release-factor 1 (RF1), which also recognizes the amber stop codon and results in termination of protein translation. The competition between $tRNA_{CUA}$ and RF1 results in mixtures of full-length and prematurely truncated target proteins. It has been reported that amber suppression efficiency can be improved by utilizing *E. coli* strains with deleted RF1 gene. The first *E. coli* RF1 knockout strain was reported by Mukai et al. (2010), where seven essential genes along with the suppressor tRNA were placed on a plasmid. This resulted in an increase in the yield of full-length protein containing UAAs, as translation termination by RF1 was eliminated. Subsequently, the *E. coli* strain JX33 was produced, which is independent of RF1. This strain displayed higher incorporation efficiency of a reporter protein with multiple amber codon sites (Johnson et al., 2011). In a major effort, the *E. coli* genome was re-engineered to replace all 314 naturally occurring amber codons by ochre codons to eliminate the genetic dependence on RF1 and allowed knockout of RF1 (Isaacs et al., 2011). These strains have shown success both *in vivo* (Chatterjee et al. 2014; Schmidt et al. 2014; Lim et al., 2015; Pirman et al., 2015; Wilkins et al., 2015) and in cell-free systems (Hong et al., 2014; Chemla et al., 2015; Li et al., 2015; Sullivan et al., 2015).

Amber codon suppression only allows incorporating a single type of UAA at a time. In order to incorporate multiple different UAAs into a protein, different codons must be used. One possibility is the usage of four-base codons to expand the genetic code (Anderson et al., 2002; Wang et al., 2012). To improve the protein yields with four-base codons, an evolved orthogonal ribosome, Ribo-X, was developed (Rackham and Chin, 2005) and used in combination with the four-base codon AGGA to produce calmodulin with two different unnatural amino acids. The resulting protein yields, however, were still low (not more than 0.5 mg of full-length protein per litre cell culture; Neumann et al., 2010b).

1.7 Chemical probes for protein NMR

From the perspective of NMR spectroscopy, it is attractive to incorporate isotopically labelled UAAs (^{13}C and ^{15}N) or fluorinated UAAs as they allow the recording of greatly simplified NMR spectra, which are ideal for obtaining site-specific information in large proteins and protein–ligand complexes, where complete resonance assignments by established techniques are time consuming or impossible to obtain. The incorporation of an NMR-active tag at a specific site of the target protein and its selective detection in the NMR spectrum instantly provides a site-specific probe for monitoring changes in chemical environment at this site. For example, by monitoring chemical shift changes of this signal it is possible to study ligand binding, conformational changes and dynamics of proteins. This strategy opens a path to study large proteins and protein complexes, which are difficult to study in other ways due to the complexity of their NMR spectra.

In an early example using a UAA, Deiters et al. (2005) incorporated ^{15}N-labelled *p*-methoxyphenylalanine into myoglobin and produced protein amounts (1 mg/L) sufficient for NMR studies. Similarly, ^{13}C-labelled *p*-O-methylphenylalanine was introduced into cytochrome P450 (CYP119) to study local ligand-induced conformational heterogeneity (Lampe et al., 2008). Three NMR active UAAs, $^{13}C/^{15}N$-*p*-methoxyphenylalanine, 2-amino-3-(4-(trifluoromethoxy)phenyl) propanoic acid (4-trifluoromethoxy-phenylalanine) and ^{15}N-O-nitrobenzyltyrosine, were incorporated at 11 sites in fatty acid synthase to study the binding of a target compound (Cellitti et al., 2008).

Previously, ^{19}F-labelled proteins had been synthesised by incorporating 5-fluoro-tryptophan, 3-fluoro-tyrosine, 4-fluoro-phenylalanine and other fluorinated amino acids in place of their natural counterparts (Gerig, 1994; Gakh et al., 2000; Frieden et al., 2004). Unless the respective amino acid only occurs a single time in the entire polypeptide sequence, however, site-specific labelling cannot be achieved in this way. UAAs solve this problem, and fluorinated UAAs such as *p*-trifluoromethyl-phenylalanine and 4-trifluoromethoxy-phenylalanine have been used successfully in proteins to study ligand binding and structural changes by ^{19}F NMR (Hammill et al., 2007; Jackson et al., 2007; Cellitti et al., 2008). Due to the large CSA associated with ^{19}F spins, however, the ^{19}F-NMR signals can be very broad in proteins of high

molecular weight (Chen et al., 2015).

For large proteins, the ^{1}H-NMR signal of a *tert*-butyl group has been proven to deliver an excellent NMR probe that can be observed without any isotopic labelling. It gives a narrow and intense singlet that can be readily identified in 1D ^{1}H-NMR spectra. In particular, the UAA O-*tert*-butyltyrosine has been used successfully for the study of high-molecular-weight systems without any isotope labelling (Chen et al., 2015; Jabar et al., 2017). Similarly, Loh et al. (2018) reported successful incorporation of *p*-(*tert*-butyl)phenylalanine into proteins ranging from 8 to 54 kDa and observed the ^{1}H resonance of the *tert*-butyl group.

The trimethylsilyl (TMS) group is another outstanding NMR probe, as the TMS signal appears near 0 ppm in the ^{1}H-NMR spectrum. At this chemical shift, signal overlap with other protein peaks is minimal, which is an important advantage over the *tert*-butyl group. The TMS group has been site-specifically introduced into proteins by chemical ligation onto a cysteine residue (Jabar et al., 2017; Becker et al., 2018). More recently, a UAA with the TMS group was successfully incorporated by CFPS. Although the ^{1}H-NMR resonance of the TMS group could be observed, the protein yield was low (Loh et al., 2018).

1.8 Outline of the thesis

Chapter 2 documents an NMR study performed on the metallo-β-lactamase IMP-1, where pseudocontact shift measurements were performed to investigate conformational changes of an active-site loop in response to a ligand. Chapter 3 describes a study conducted on the interaction between single-stranded DNA binding protein and single-stranded oligonucleotides by NMR spectroscopy, using a ^{13}C-*tert*-butyl group as NMR probe. Chapter 4 describes an approach for the site-specific introduction of NMR tags to study high-molecular weight systems. The study yielded intermolecular NOEs by employing a novel chemical tag. The publication resulting from this work is reproduced.

Chapter 1. General introduction.

Chapter 2. Study of IMP-1 metallo-β-lactamase active site loop flexibility by NMR spectroscopy.

Chapter 3. Study of oligonucleotide binding by *E. coli* single-stranded DNA binding protein usingz NMR spectroscopy.

Chapter 4. Chemical tagging with *tert*-butyl and trimethylsilyl groups for measuring intermolecular nuclear Overhauser effects in a large protein–ligand complex.

1.9 References

Allegrozzi, M., Bertini, I., Janik, M.B.L., Lee, Y.-M., Liu, G. and Luchinat, C. (2000) Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. *J. Am. Chem. Soc.* **122**, 4154–4161.

Anderson, J.C., Magliery, T.J., and Schultz, P.G. (2002) Exploring the limits of codon and anticodon size. *Chem. Biol*. **9**, 237–244.

Apponyi, M., Ozawa, K., Dixon, N.E. and Otting, G. (2008) Cell-free protein synthesis for analysis by NMR spectroscopy. *Methods in Molecular Biology 426, Structural proteomics: high-throughput methods B. Kobe, M. Guss, T. Huber Eds., Humana Press, Totowa, USA* pp. 257-268.

Atreya, H.S. and Szyperski, T. (2005) Rapid NMR data collection. *Methods Enzymol.* **394**, 78–108.

Barb, A.W., Ho, T.G., Flanagan-Steet, H., Prestegard, J.H., (2012) Lanthanide binding and IgG affinity construct: potential applications in solution NMR, MRI, and luminescence microscopy. *Protein Sci.* **21**, 1456–1466.

Barb, A.W. and Subedi, G.P. (2016) An encodable lanthanide binding tag with reduced size and flexibility for measuring residual dipolar couplings and pseudocontact shifts in large proteins. *J. Biomol. NMR* **64**, 75–85.

Barthelmes, K., Reynolds, A.M., Peisach, E., Jonker, H.R.A., DeNunzio, N.J., Allen, K.N., Imperiali, B. and Schwalbe, H. (2011) Engineering encodable lanthanide-binding tags into loop regions of proteins. *J. Am. Chem. Soc.* **133**, 808–819.

Becker, W., Adams, L.A., Graham, B., Wagner, G.E., Zangger, K., Otting, G. and Nitsche, C. (2018) Trimethylsilyl tag for probing protein-ligand interactions by NMR *J. Biomol. NMR* **70**, 211-218.

Bertini, I., Janik, M.B.L., Lee, Y.M., Luchinat, C. and Rosato, A. (2001) Magnetic susceptibility tensor anisotropies for a lanthanide ion series in a fixed protein matrix. *J. Am. Chem. Soc.* **123**, 4181–4188.

Bertini, I., Luchinat, C. and Parigi, G. (2002) Magnetic susceptibility in paramagnetic NMR. *Prog. Nucl. Magn. Reson. Spec.* **40**, 249–273.

Bertini, I., Gelis, I., Katsaros, N., Luchinat, C. and Provenzani, A. (2003) Tuning the affinity for lanthanides of calcium binding proteins. *Biochemistry* **42**, 8011–8021.

Carison, E.D., Gan, R., Hodgman, C.E. and Jewett, M.C. (2012) Cell-free protein synthesis: applications come of age. *Biotechnol. Adv.* **30**, 1185-1194.

Case, D.A. (1995) Calibration of ring-current effects in proteins and nucleic acids. *J. Biomol. NMR* **6**, 341–346.

Cellitti, S.E., Jones, D.H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S.M., Brinker, A., Caldwell, J., Bursulaya, B., Spraggon, G., Brock, A., Ryu, Y., Uno, T., Schultz, P.G. and Geierstanger, B.H. (2008) *In vivo* incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **130**, 9268-9281.

Chatterjee, A., Lajoie, M.J., Xiao, H., Church, G.M. and Schultz, P.G. (2014) A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *ChemBioChem* **15**, 1782–1786.

Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V. and Alfonta, L. (2015) Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotechnol. Bioeng.* **112**, 1663–1672.

Chen, W.-N., Loscha, K.V., Nitsche, C., Graham, B. and Otting, G. (2014) The dengue virus NS2B-NS3 protease retains the closed conformation in the complex with BPTI. *FEBS Lett.* **588**, 2206-2211.

Chen, W.-N., Kuppan, K.V., Lee, M.D., Jaudzems, K., Huber, T. and Otting, G. (2015) O-*tert*-butyltyrosine, an NMR tag for high-molecular-weight systems and measurements of submicromolar ligand binding affinities. *J. Am. Chem. Soc.* **137**, 4581-4586.

Chen, W.-N., Nitsche, C., Pilla, K.B., Graham, B., Huber, T., Klein, C.D. and Otting, G. (2016) Sensitive NMR approach for determining the binding mode of tightly binding ligand molecules to protein targets. *J. Am. Chem. Soc.* **138**, 4539-4546.

Chin, J.W. (2003) An expanded eukaryotic genetic code. *Science* **301**, 964-967.

Chin, J.W., Cropp, T.A., Chu, S., Meggers, E. and Schultz, P.G. (2003) Progress towards and expanded eukaryotic genetic code. *Chem. Biol.* **10**, 511-519.

de la Cruz, L., Nguyen, T.H.D., Ozawa, K., Shin, J., Graham, B., Huber, T. and Otting, G. (2011) Binding of low-molecular weight inhibitors promotes large conformational changes in the dengue virus NS2B-NS3 protease: fold analysis by pseudocontact shifts. *J. Am. Chem. Soc.* **133**, 19205-19215.

Deiters, A., Geierstanger, B.H., and Schultz, P.G. (2005) Site-specific *in vivo* labeling of proteins for NMR studies. *ChemBioChem* **6**, 55–58.

Franz, K.J., Nitz, M. and Imperiali, B. (2003) Lanthanide-binding tags as versatile protein coexpression probes. *ChemBioChem* **4**, 265–271.

Freeman, R. and Kupce, E. (2003) New methods for fast multidimensional NMR. *J. Biomol. NMR* **27**, 101–113.

Frieden, C., Hoeltzli, S.D. and Bann, J.G. (2004) The preparation of ^{19}F-labeled proteins for NMR studies. *Methods Enzymol.* **380**, 400–415.

Furuita, K., Kataoka, S., Sugiki, T., Hattori, Y., Kobayashi, N., Ikegami, T., Shiozaki, K., Fujiwara, T. and Kojima, C. (2015) Utilization of paramagnetic relaxation enhancements for high-resolution NMR structure determination of a soluble loop-rich protein with sparse NOE distance restraints. *J. Biomol. NMR* **61**, 55-64.

Gakh, Y.G., Gakh, A.A. and Gronenborn, A.M. (2000) Fluorine as an NMR probe for structural studies of chemical and biological systems. *Magn. Reson. Chem.* **38**, 551–558.

Gerig, J.T. (1994) Fluorine NMR of proteins. *Prog. Nucl. Magn. Reson. Spectrosco.* **26**, 293–370.

Graham, B., Loh, C.-T., Swarbrick, J.D., Ung, P., Shin, J., Yagi, H., Jia, X., Chhabra, S., Pintacuda, G., Huber, T. and Otting, G. (2011) A DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. *Bioconjugate Chem.* **22**, 2118-2125.

Guignard, L., Ozawa, K., Pursglove, S.E., Otting, G. and Dixon, N.E. (2002) NMR analysis of in vitro-synthesized proteins without purification: a high-throughput approach. *FEBS Lett.* **524**, 159-162.

Hammill, J.T., Miyake-Stoner, S., Hazen, J.L., Jackson, J.C. and Mehl, R.A. (2007) Preparation of site-specifically labeled fluorinated proteins for ^{19}F-NMR structural characterization. *Nat. Protoc.* **2**, 2601–2607.

Hong, S.H., Ntai, I., Haimovich, A.D., Kelleher, N.L., Isaacs, F.J. and Jewett, M.C. (2014) Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS Synth. Biol*. **3**, 398–409.

Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., et al. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353.

Jabar, S., Adams, L., Wang, Y., Aurelio, L., Graham, B., and Otting, G. (2017) Chemical tagging with *tert*-butyl and trimethylsilyl groups for measuring intermolecular nuclear Overhauser effects in a large protein-ligand complex. *Chem. Eur. J.* **23**, 13033-13036.

Jackson, J.C., Hammill, J.T. and Mehl, R.A. (2007) Site-specific incorporation of a ^{19}F amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* **129**, 1160–1166.

Jia, X., Ozawa, K., Loscha, K. and 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.

John, M., Park, A.Y., Pintacuda, G., Dixon, N.E. and Otting, G. (2005) Weak alignment of paramagnetic proteins warrants correction for residual CSA effects in measurements of pseudocontact shifts. *J. Am. Chem. Soc.* **127**, 17190–17191.

John, M., Pintacuda, G., Park, A.Y., Dixon, N.E. and Otting, G. (2006) Structure determination of protein-ligand complexes by transferred paramagnetic shifts. *J. Am. Chem. Soc.* **128**, 12910–12916.

Johnson, D.B., Xu, J., Shen, Z., Takimoto, J.K., Schultz, M.D., Schmitz, R.J., Xiang, Z., Ecker, J.R., Briggs, S.P. and Wang, L. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779–786.

Jones, D.H., Cellitti, S.E., Hao, X., Zhang, Q., Jahnz, M., Summerer, D., Schultz, P.G., Uno, T. and Geierstanger, B.H. (2010) Site-specific labeling of proteins with NMR-active unnatural amino acids. *J. Biomol. NMR* **46**, 89–100.

Lampe, J.N., Floor, S.N., Gross, J.D., Nishida, C.R., Jiang, Y., Trnka, M.J. and de Montellano, P.R.O. (2008) Ligand-induced conformational heterogeneity of cytochrome P450 CYP119 identified by 2D NMR spectroscopy with the unnatural amino acid ^{13}C-*p*-methoxyphenylalanine. *J. Am. Chem. Soc.* **130**, 16168–16169.

Lee, H.S., Spraggon, G., Schultz, P.G. and Wang, F. (2009) Genetic incorporation of a metal-ion chelating amino acid into proteins as biophysical probe. *J. Am. Chem. Soc.* **131**, 2481-2483.

Li, J., Lawton, T.J., Kostecki, J.S., Nisthal, A., Fang, J., Mayo, S.L., Rosenzweig, A.C. and Jewett, M.C. (2015) Cell-free protein synthesis enables high yielding synthesis of an active multicopper oxidase. *Biotechnol. J.* **11**, 212-218.

Lim, S.I., Hahn, Y.S. and Kwon, I. (2015) Site-specific albumination of a therapeutic protein with multi-subunit to prolong activity *in vivo*. *J. Control Rel*. **207**, 93–100.

Lin, M.Z. and Wang, L. (2008) Selective labelling of proteins with chemical probes in living cells. *Physiology* **23**, 131-141.

Liu, C.C. and Schultz, P.G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413-444.

Loh, C.-T., Ozawa, K., Tuck, K.L., Barlow, N., Huber, T., Otting, G. and Graham, B. (2013) Lanthanide tags for site-specific ligation to an unnatural amino acid and generation of pseudocontact shifts in proteins. *Bioconjugate Chem.* **24**, 260–268.

Loh, C.-T., Adams, L.A., Graham, B., Otting, G. (2018) Genetically encoded amino acids with *tert*-butyl and trimethylsilyl groups for site-selective studies of proteins by NMR spectroscopy. *J. Biomol. NMR* **71**, 287-293.

Loscha, K.V., Herlt, A.J., Qi, R., Huber, T., Ozawa, K. and Otting, G. (2012) Multiple-site labeling of proteins with unnatural amino acids. *Angew. Chem. Int. Ed.* **51**, 2243-2246.

Ma, C. and Opella, S.J. (2000) Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy. *J. Magn. Reson.* **146**, 381–384.

Mahawaththa, M.C., Pearce, B.J.G., Szabo, M., Graham, B., Klein, C.D., Nitsche, C. and Otting, G. (2017) Solution conformation of a linked construct of the Zika virus NS2B-NS3 protease. *Antiviral Res.* **142**, 141-147.

Markley, J.L., Putter, I. and Jardetzky, O. (1968) High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science* **161**, 1249–1251.

Marley, J., Lu, M. and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–77.

McIntosh, L.P. and Dahlquist, F.W. (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q. Rev. Biophys.* **23**, 1–38.

Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S. and Sakamoto, K. (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* **38**, 8188–8195.

Neumann, H., Peak-Chew, S.Y. and Chin, J.W. (2008) Genetically encoding N^{ε}-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4**, 232–234.

Neumann, H., Slusarczyk, A.L., and Chin, J.W. (2010a) De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J. Am. Chem. Soc.* **132**, 2142-2144.

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. and Chin, J.W. (2010b) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444.

Nirenberg, M.W. and Matthaei, J.H. (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* **47,** 1588–1602.

Nitsche, C. and Otting, G. (2017) Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. *Prog. Nucl. Magn. Reson. Spec.* **98-99**, 20-49.

Nitz, M., Franz, K.J., Maglathlin, R.L. and Imperiali, B. (2003) A powerful combinatorial screen to identify high-affinity terbium(III)-binding peptides. *ChemBioChem* **4**, 272–276.

Nitz, M., Sherawat, M., Franz, K.J., Peisach, E., Allen, K.N. and Imperiali, B. (2004) Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide. *Angew. Chem. Int. Ed.* **43**, 3682–3685.

Otting, G. (2008) Prospects for lanthanides in structural biology by NMR. *J. Biomol. NMR* **42**, 1–9.

Otting, G. (2010) Protein NMR using paramagnetic ions. *Annu. Rev. Biophys*. **39**, 387–405.

Ozawa, K., Headlam, M.J., Schaeffer, P.M., Henderson, B.R., Dixon, N.E. and Otting, G. (2004) Optimization of an Escherichia coli system for cell-free synthesis of selectively ^{15}N-labelled proteins for rapid analysis by NMR spectroscopy. *Eur. J. Biochem.* **271**, 4084-4093.

Ozawa, K., Wu, P.S., Dixon, N.E. and Otting, G. (2006) ^{15}N-labelled proteins by cell-free protein synthesis: strategies for high-throughput NMR studies of proteins and protein-ligand complexes. *FEBS J.* **273**, 4154–4159.

Ozawa, K., Loscha, K.V., Kuppan, K.V., Loh, C.T., Dixon, N.E. and Otting, G. (2012) High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochem. Biophys. Res. Commun.* **418**, 652-656.

Pearce, B.J.G., Jabar, S., Loh, C.-T., Szabo, M., Graham, B. and Otting, G. (2017) Structure restraints from heteronuclear pseudocontact shifts generated by lanthanide tags at two different sites. *J. Biomol. NMR* 68, 19-32.

Pilla, K.B., Otting, G. and Huber, T. (2016) Pseudocontact shift-driven iterative resampling for 3D structure determinations of large proteins. *J. Mol. Biol*. **428**, 522-532.

Pilla, K.B., Otting, G. and Huber, T. (2017) Protein structure determination by assembling super-secondary structure motifs using pseudocontact shifts. *Structure* **25**, 559-568.

Pirman, N.L., Barber, K.W., Aerni, H.R., Ma, N.J., Haimovich, A.D., Rogulina, S., Isaacs, F.J. and Rinehart, J. (2015) A flexible codon in genomically recoded *Escherichia coli* permits programmable protein phosphorylation. *Nat. Commun.* **6**, 8130.

Pintacuda, G., John, M., Su, X.-C. and Otting, G. (2007) NMR structure determination of protein-ligand complexes by lanthanide labeling. *Acc. Chem. Res.* **40**, 206-212.

Prestegard, J.H. and Kishore, A.I. (2001) Partial alignment of biomolecules: an aid to NMR characterization. *Curr. Opin. Chem. Biol.* **5**, 584–590.

Rackham, O. and Chin, J.W. (2005). A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol*. **1**, 159–166.

Saio, T., Ogura, K., Yokochi, M., Kobashigawa, Y. and Inagaki, F. (2009) Two-point anchoring of a lanthanide-binding peptide to a target protein enhances the paramagnetic anisotropic effect. *J. Biomol. NMR* **44**, 157–166.

Schmitz, C., Vernon, R., Otting, G., Baker, D. and Huber, T. (2012) Protein structure determination from pseudocontact shifts using ROSETTA. *J. Mol. Biol.* **416**, 668–677.

Schmidt, M.J., Borbas, J., Drescher, M. and Summerer, D. (2014) A genetically encoded spin label for electron paramagnetic resonance distance measurements. *J. Am. Chem. Soc.* **136**, 1238–1241.

Su, X.-C., Huber, T., Dixon, N.E. and Otting, G. (2006) Site-specific labelling of proteins with a rigid lanthanide-binding tag. *ChemBioChem* **7**, 1599–1604.

Su, X.-C., McAndrew, K., Huber, T. and Otting, G. (2008) Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles. *J. Am. Chem. Soc.* **130**, 1681–1687.

Su, X.-C., Loh, C.-T., Qi, R. and Otting, G. (2011) Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enzymes for selective ^{15}N-labelling and production of perdeuterated proteins in H_2O. *J. Biomol. NMR* **50**, 35-42.

Sullivan, C.J., Pendleton, E.D., Sasmor, H.H., Hicks, W.L., Farnum, J.B., Muto, M., Amendt, E.M., Schoborg, J.A., Martin, R.W., Clark, L.G., Anderson, M.J., Choudhury, A., Fior, R., Lo, Y.H., Griffey, R.H., Chappell, S.A., Jewett, M.C., Mauro, V.P. and Dresios, J. (2015) A cell-free expression and purification process for rapid production of protein biologics. *Biotechnol. J.* **11**, 238-248.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) Use of dipolar ^{1}H–^{15}N and ^{1}H–^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* **4**, 732–738.

Torizawa, T., Shimizu, M., Taoka, M., Miyano, H. and Kainosho, M. (2004) Efficient production of isotopically labeled proteins by cell-free synthesis: A practical protocol. *J. Biomol. NMR* **30**, 311-325.

Wagner, G., Pardi, A. and Wuthrich, K. (1983) Hydrogen-bond length and ^{1}H NMR chemical shifts in proteins. *J. Am. Chem. Soc*. **105**, 5948–5949.

Wang, L., Brock, A., Herberich, B. and Schultz, P.G. (2001) Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500.

Wang, K., Schmied, W.H., and Chin, J.W. (2012) Reprogramming the genetic code: from triplet to quadruplet codes. *Angew. Chem. Int. Ed. Engl*. **51**, 2288–2297.

Wilkins, B.J., Hahn, L.E., Heitmüller, S., Frauendorf, H., Valerius, O., Braus, G.H. and Neumann, H. (2015) Genetically encoding lysine modifications on histone H4. *ACS Chem. Biol.* **10**, 939–944.

Williamson, M.P. and Asakura, T. (1993) Empirical comparisons of models for chemical-shift calculation in proteins. *J. Magn. Reson. B.* **101**, 63–71.

Wu, N., Deiters, A., Cropp, T.A., King, D. and Schultz, P.G. (2004) A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 14306-14307.

Wu, P.S.C., Ozawa, K., Lim, S.P., Vasudevan, S., Dixon, N.E., and Otting, G. (2007) Cell-free transcription/translation from PCR amplified DNA for high-throughput NMR studies. *Angew. Chemie Int. Ed.* **46**, 3356-3358.

Wüthrich, K. (1990) Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* **265**, 22059-22062.

Xie, J. and Schultz, P.G. (2005) An expanding genetic code. *Methods* **36,** 227-238.

Yagi, H., Pilla, K.B., Maleckis, A., Graham, B., Huber, T. and Otting, G. (2013) Three dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites. *Structure* **21**, 883–890.

Young, D.D., Young, T.S., Jahnz, M., Ahmad, I., Spraggon, G. and Schultz, P.G. (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894-1900.

CHAPTER 2

Study of IMP-1 metallo-β-lactamase active site loop flexibility by NMR spectroscopy

2.1 Introduction

2.1.1 β-lactam antibiotics

Antibiotics are substances used for the treatment and prevention against bacterial infections. They act by killing or inhibiting the growth of bacteria. There are different classes of antibiotics, yet β-lactam antibiotics are the most widely used group among all commercially available antibiotics. β-lactam antibiotics include penicillin derivatives (penams), cephalosporins (cephems), monobactams and carbapenems, which contain a β-lactam ring in their core structure.

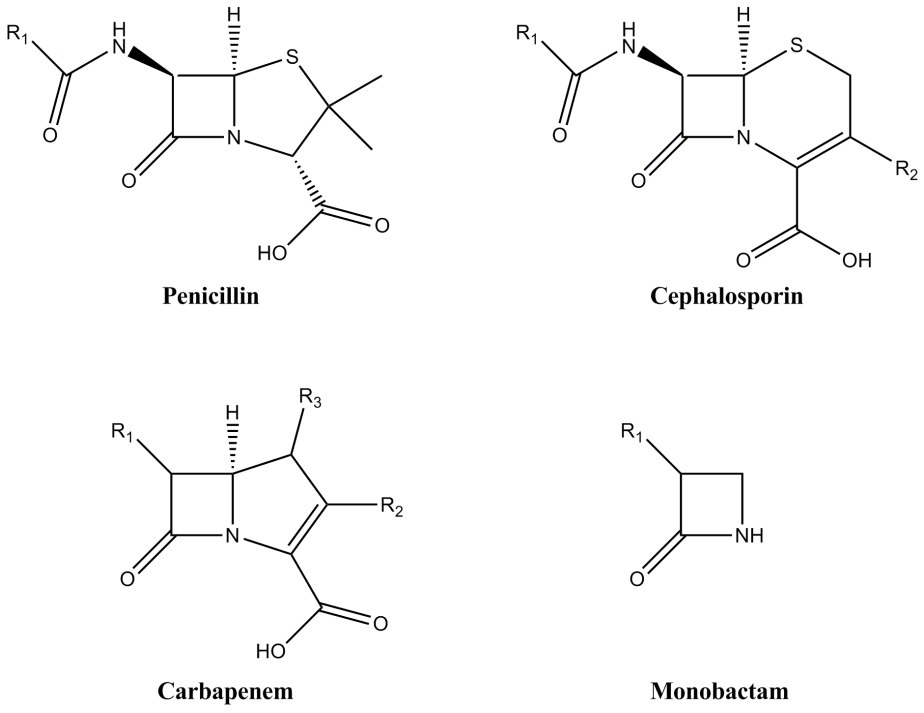


Figure 2.1 Core structures of β-lactams.

β-lactam antibiotics kill the bacterial cells by inhibiting peptidoglycan synthesis, which constitutes a major portion of bacterial cell wall synthesis. As a consequence of inhibiting cell wall synthesis, the autolytic system will be activated, which causes bacterial cell death. However, some bacteria can survive against these antibiotics by various mechanisms. These include a) production of another enzyme (β-lactamases) which can inactivate the antibiotic before reaching the target, b) mutation of the active site of the target enzyme, preventing binding of the antibiotic and c) modification of the cell wall to prevent the entry of the drug into the cell or to actively remove the antibiotic via enhanced expression of efflux pumps (Levy and Marshall, 2004).

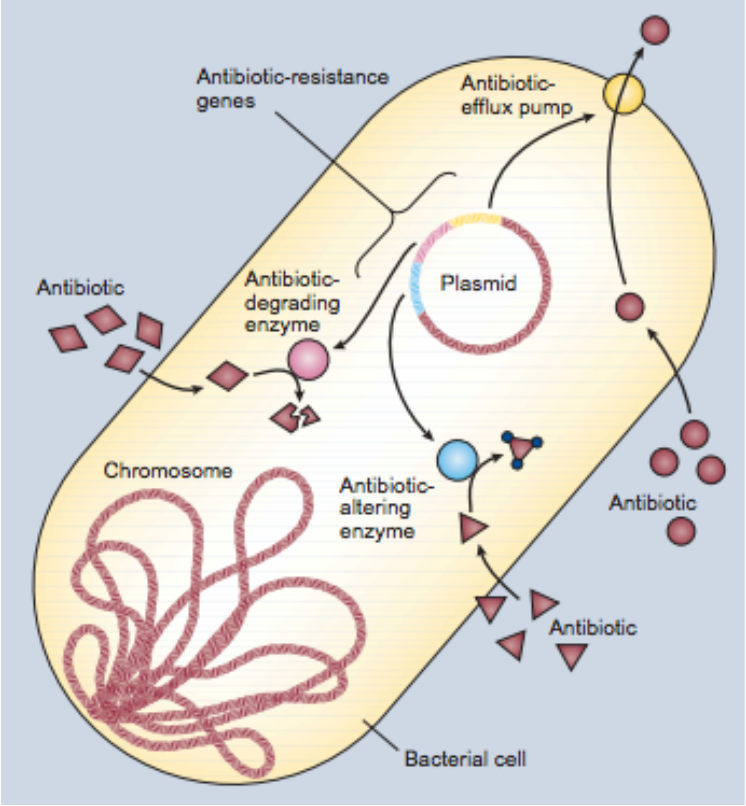


Figure 2.2 An overall representation of bacterial resistance mechanism. Reproduced from Levy and Marshall (2004).

2.1.2 β-lactamase enzymes

β-lactamases are enzymes that inactivate β-lactam antibiotics by hydrolysing the amide bond of the β-lactam rings (Essack, 2001; Siu, 2002). Abraham and Chain identified the first β-lactamase in *E. coli* already in 1940, even before the widespread use of penicillin in the treatment of bacterial infections (Abraham and Chain, 1940). Since then more

than 1000 of these enzymes have been found in Gram-positive bacteria, Gram-negative bacteria and mycobacteria (Davies and Davies, 2010; Livermore, 1995; Majiduddin et al., 2002). β-lactamases are either chromosomally encoded or acquired through mobile genetic elements such as plasmids or transposons (Walsh et al., 2005). As a result, bacteria are able to acquire β-lactamase genes and become resistant to β-lactams. A few β-lactamases utilize zinc ions to disrupt the β-lactam ring, but many form an acyl-enzyme intermediate through an active site serine.

2.1.3 Classification of β-lactamases

β-lactamases are classified according to either their amino acid sequence similarities (Ambler, 1980) or by their functional and biochemical properties (Richmond and Sykes, 1973; Sykes, 2010). According to the Ambler classification, β-lactamases are divided into metallo-β-lactamases (class B) and serine β-lactamases (classes A, C and D) based on amino acid sequence (Ambler, 1980). The Bush-Jacoby classification is based on substrate selectivity and inhibitor profile and separates the enzymes again into four classes, 1 to 4 (Bush et al., 1995).

2.1.4 Metallo-β-lactamases

Unlike serine β-lactamases, which require an active site serine residue for enzymatic activity, metallo-β-lactamases (MBLs) are metallo-enzymes which require coordination of bivalent metal ions, usually zinc, in their active site.

Subclasses of MBLs

MBLs are further divided into three subclasses, B1, B2 and B3, which are grouped according to substrate selectivity, amino acid sequence (particularly the residues that chelate the Zn^{2+} ions) and the Zn^{2+} ion requirement for catalysis (Rasmussen and Bush, 1997).

The MBLs in subclass B1 require the presence of two Zn^{2+} ions at the active site for activity and they show broad-spectrum activity towards most β-lactam compounds including carbapenems. A large number of 3D structures have been solved for subclass

B1 compared to subclasses B2 and B3. This subclass harbours MBLs, which are both chromosomally encoded and acquired

Subclass B2 MBLs harbour only a single Zn^{2+} ion in their active site, which is required for β-lactam hydrolysis. However, these MBLs are similar to the MBLs in subclass B1 in size and amino acid sequence (Hall et al., 2003).

Subclass B3 is the least studied of the three subclasses. MBLs in this subclass harbor two Zn^{2+} ions at their active site and show a similar substrate profile to subclass B1 MBLs. Nevertheless, the structure and the amino acid sequence of these enzymes differ significantly from both subclasses B1 and B2. In the present work, the metal binding properties of a B1 MBL were investigated as well as conformational changes of an active site loop in response to ligand binding. In the following, the structural features of the metal binding sites and polypeptide loops near the active site are reviewed.

Three dimensional structure of MBLs

Carfi and co-workers solved the X-ray crystal structure of the mono-zinc form of BcII in 1995, which was the first MBL structure determined (Carfi et al., 1995). Since then crystal structures of MBLs have been determined from all three subclasses. These structures revealed a common characteristic fold, a αβ/βα quaternary structure, and catalytic function, indicating that all MBLs evolved from a common ancestor (Daiyasu et al., 2001). The overall structure comprises two central β-sheets flanked by five solvent-exposed α-helices with variable loops which connect the α-helices and β-strands. The Zn^{2+} ion(s) binding active site is located at the interface of two αβ domains.

Despite common structural features, there are also notable differences between the 3D structures of each subclass, such as the number of Zn^{2+} ions in the active site, different Zn^{2+} ion(s) coordinating residues and mobile β-hairpin loop regions contributing to substrate binding. In addition, MBLs in subclass B3 contain intramolecular disulfide bridges, which play an important role in protein folding (Ullah et al., 1998).

Subclass B1 contains a flexible loop referred to as loop 1 in close proximity to

the active site of the enzyme. The crystal structures suggest that this loop interacts with bound substrates or inhibitors, and deletion of loop 1 negatively affects the enzyme activity (Moali et al., 2003). The loop conformation becomes more restricted when an inhibitor is bound (Bebrone, 2007; Concha et al., 1996). In addition, there is a flexible loop 2, which also influences the substrate or inhibitor interactions in subclass B1 MBLs (Dal Peraro et al., 2004).

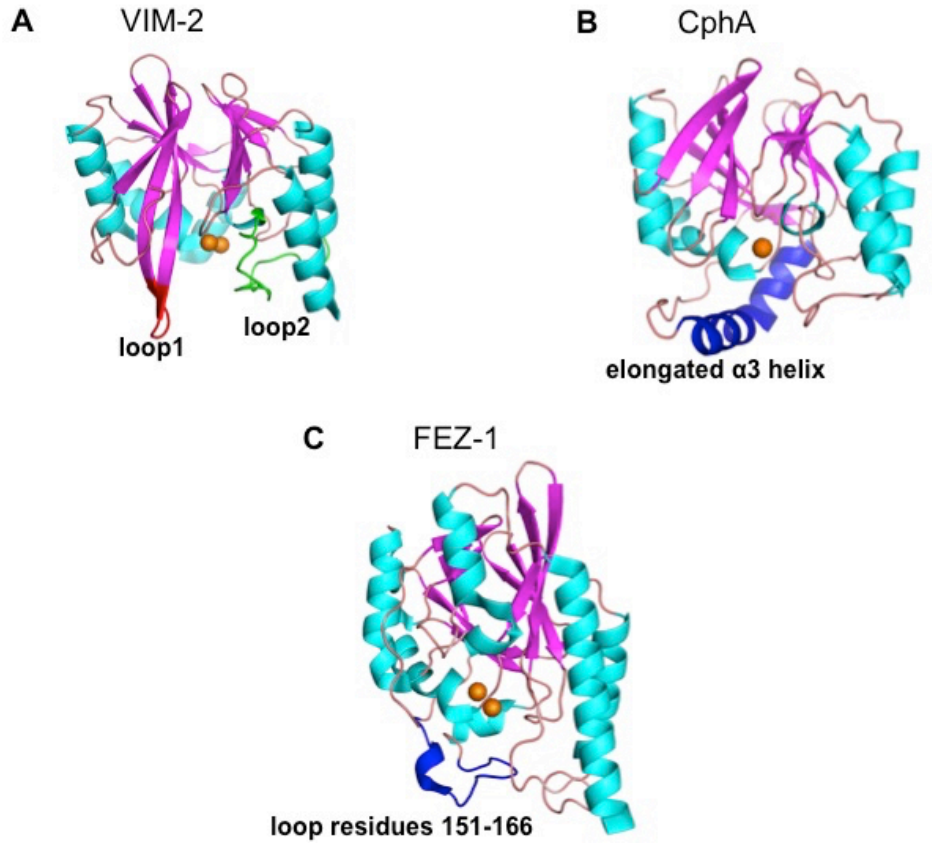


Figure 2.3 Three-dimensional structures of MBLs. (A) VIM-2, loop 1 and loop 2 are highlighted in red and green respectively (PDB ID: 1KO3, B1 subclass; Garcia-Saez et al., 2008), (B) CphA, with the elongated α3 helix highlighted in blue (PDB id: 1X8G, B2 subclass; Garau et al., 2005b) and (C) FEZ-1, where a flexible loop region near the active site is highlighted in blue. (PDB id: 1K07, B3 subclass; Garcia-Saez et al., 2003). Zn^{2+} ion(s) are represented as orange spheres.

The loop 1 region is absent in subclass B2 and B3 MBLs (Bebrone, 2007). Subclass B2 contains an elongated α3 helix, which forms a part of a hydrophobic wall lining the substrate binding groove, which endows the subclass B2 MBLs with

specificity for binding carbapenems (Bebrone, 2007; Garau et al., 2005a). The subclass B3 contains a loop region, loop 2, which is formed between helix α3 and β-strand β7 and is in close proximity to the active site (Garcia-Saez et al., 2003).

Active site and Zn binding residues

The active site of MBLs contains two potential catalytic Zn^{2+} ion(s) coordinating sites known as Zn1 and Zn2. Nevertheless, the number of Zn^{2+} ions at the active site and the residues that coordinate Zn^{2+} ion(s) vary between the subclasses B1, B2 and B3. Subclass B1and B3 contains two catalytic Zn^{2+} ions, whereas subclass B2 acts as a mono-zinc enzyme.

In subclasse B1, the Zn1 site is formed by three histidine residues, His116, His118 and His 196 ("histidine site") and a water molecule in a tetrahedral geometry. The Zn2 site is formed by Asp120, Cys221, His263 ("cysteine site") and two water molecules in a trigonal bipyramidal arrangement. The same water molecule that links both Zn^{2+} ions also acts as a nucleophile during hydrolysis (Concha et al., 1996; Wang et al.,1999).

The crystal structure of the mono-zinc subclass B2 MBL CphA suggests that the catalytic Zn^{2+} ion resides in the cysteine site. His116 in the histidine site is replaced by Asn116 and shows no direct interaction with the Zn1 site, which explains why MBLs in subclass B2 bind to a single Zn^{2+} ion via Asp120, Cys221 and His263 (Garau et al., 2005a).

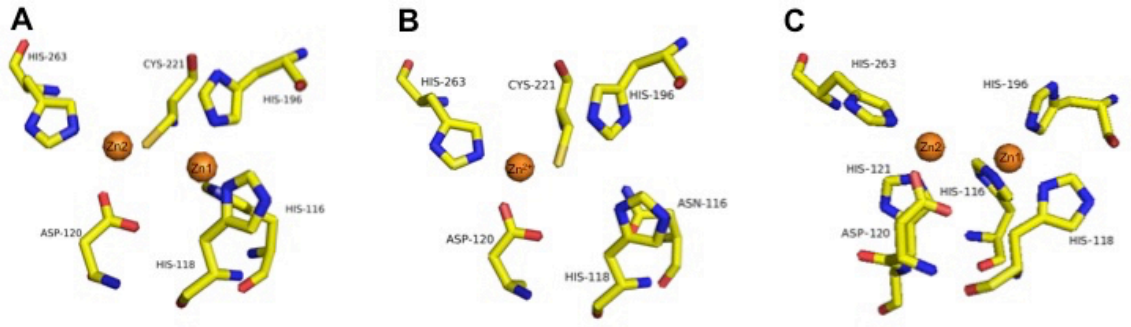


Figure 2.4 Zinc sites and metal ligand architecture at the active site. (A) VIM-2 (PDB ID: 1KO3, B1 subclass; Garcia-Saez et al., 2008), (B) CphA (PDB ID: 1X8G, B2 subclass; Garau et al., 2005b) and (C) FEZ-1 (PDB ID: 1K07, B3 subclass; Garcia-Saez et al., 2003). All figures generated with PyMol.

The Zn1 site in the subclass B3 is a histidine site similar to that of subclass B1. The Zn2 site is formed with His121, which replaces a cysteine residue in subclasses B1 and B2. Hence, the Zn2 site is formed by Asp120, His121, His263 and two water molecules in trigonal bipyramidal geometry (Ullah et al., 1998; Garcia-Saez et al., 2003).

2.1.5 IMP-1 metallo-β-lactamase

Among the families of MBLs, the IMP-type (for "active on imipenem") β-lactamases were the first to be identified. They have been detected in many clinically important Gram-negative bacteria like *Pseudomonas* spp., *Acinetobacter* spp. and members of the *Enterobacteriaceae* family (Zhao and Hu, 2011).

The MBL IMP-1, which is a subclass B1 β-lactamase, was first identified in a *Pseudomonas aeruginosa* isolate in Japan in 1990 (Watanabe et al., 1991). Subsequently, IMP-1 was found to be prevalent also in *Serratia marcescens* and in numerous members of the *Enterobacteriaceae* family in Japan (Queenan and Bush, 2007). IMP-1 is encoded by the gene bla_{IMP}, which is integrated as a gene cassette into integrons carried by transferrable plasmids (Arakawa et al., 1995). Hence, the bla_{IMP} gene can be spread amongst different nosocomial pathogens via horizontal gene transfer.

IMP-1 has the common 3D structure of a MBL. Published IMP-1 crystal structures show the typical αβ/βα quaternary structural fold and two Zn^{2+} ions (Concha et al., 2000). The two Zn^{2+} ions in the active site of IMP-1 are bound in two distinct sites ("histidine site" and "cysteine site"). Strong binding of divalent metal ions to the protein has been demonstrated by retention of the metal ions even after prolonged dialysis and exposure to chelators (Siemann et al., 2002; Laraki et al., 1999).

The loop 1 on IMP-1 comprises residues 60–66 (standard numbering defined by Galleni et al., 2001). None of these residues is fully conserved among subclass B1 β-lactamase (Moali et al., 2003). The X-ray structure of the native enzyme shows an open, disordered conformation of the flap. In the enzyme-substrate complexes, however, the flap is in a closed conformation (Concha et al., 2000). Based on kinetic data by Moali et al. (2003), Trp64 accounts for approximately 50% of the loop effect, probably through hydrophobic interactions.

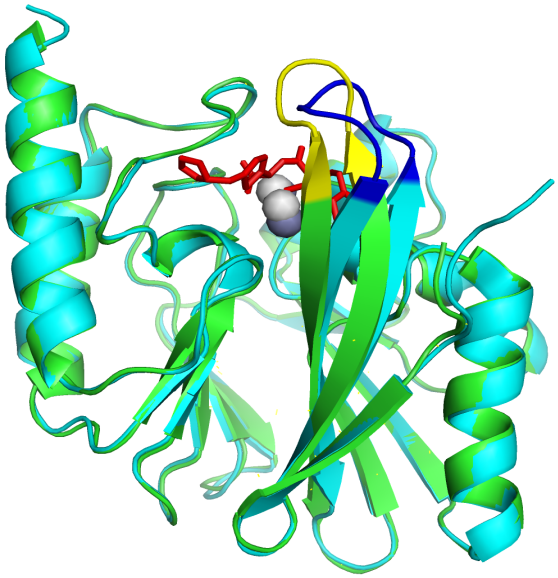


Figure 2.5 Overlay of crystal structures of native IMP-1 (cyan with the loop 1 highlighted in blue; PDB ID: 1DDK; Concha et al., 2000) and its complex with a mercaptocarboxylate inhibitor (green with loop 1 highlighted in yellow; PDB ID: 1DD6; Concha et al., 2000). Zn atoms are shown as spheres and the inhibitor molecule is shown in red.

In 2014, Carruthers et al. reported the observation of a heterodinuclear form of IMP-1, [FeZn]-IMP-1. It was shown that when IMP-1 is overexpressed *in vivo,* the presence of small amounts (25 μM) of $FeCl_3$ in the expression medium results in [FeZn]-IMP-1. The formation of [FeZn]-IMP-1 could not be suppressed even with the addition of ten-fold excess of zinc. Moreover, it was determined that the iron bound to IMP-1 was present as Fe^{3+} in the Zn1 site (Carruthers et al., 2014). Interestingly, the purified [FeZn]-IMP-1 protein samples were purple in colour.

2.1.6 Research objectives

In the present work, the clinically important MBL IMP-1 was studied by NMR spectroscopy. PCS measurements were used to explore the flexibility of the active site loop 1 upon binding of an inhibitor, using Tby as the NMR probe. Furthermore, a preliminary analysis was carried out to determine whether the active site cysteine residue plays a role in the purple colouration of [FeZn]-IMP-1.

2.2 Materials and Methods

2.2.1 Plasmid constructs

The wild-type gene of *bla*$_{IMP-1}$ cloned into the pET-47b(+) vector was used in this study. The gene was inserted between *Nde*I and *Xho*I restriction sites, removing the N-terminal His$_6$-tag from the vector.

```
MESLPDLKIE KLDEGVYVHT SFEEVNGWGV VPKHGLVVLV NAEAYLIDTP  50
FTAKDTEKLV TWFVERGYKI KGSISSHFHS DSTGGIEWLN SRSIPTYASE 100
LTNELLKKDG KVQATNSFSG VNYWLVKNKI EVFYPGPGHT PDNVVVWLPE 150
RKILFGGCFI KPYGLGNLGD ANIEAWPKSA KLLKSKYGKA KLVVPSHSEV 200
GDASLLKLTL EQAVKGLNES KKPSKPSN                         228
```

Figure 2.6 Amino acid sequence of the metallo-β-lactamase IMP-1. The active site cysteine residue is indicated in yellow, the sites of the Val30Tby and Y163Tby mutations are indicated in green and the sites of the introduced cysteine mutations for attaching the lanthanide tag are indicated in red.

The active site cysteine residue was mutated to glycine, alanine and threonine (mutants C158G, C158A and C158T) for ICP-OES analysis. In addition, the *bla*$_{IMP-1}$ gene was mutated to produce the amber stop mutants V30U and Y163U and three different cysteine mutants (A53C, N172C and S204C) by PCR. Additionally, three double mutants V30U/A53C, V30U/N172C and V30U/S204C were produced. The mutation primers are listed in Table 2.1.

Table 2.1 Nucleotide sequences of the mutation primers used in this study.

Mutation site	Primer Sequence	
C158G	Forward	GTTTGGCGGCGGCTTTATTAAACCGTATGGCCTG
	Reverse	GTTTATTAAAGCCGCCGCCAAACAGAATTTTACG
C158A	Forward	GTTTGGCGGCGCGTTTATTAAACCGTATGGCCTG
	Reverse	GTTTATTAAACGCGCCGCCAAACAGAATTTTACG
C158T	Forward	GTTTGGCGGCACCTTTATTAAACCGTATGGCCTG
	Reverse	GTTTATTAAAGGTGCCGCCAAACAGAATTTTACG

V30TAG	Forward	TGGGGCTAGGTGCCGAAACATGGCCTGGTGGTG
	Reverse	CGGCACCTAGCCCCAGCCGTTCACTTCTTCAAAGC
A53C	Forward	TTTACCTGCAAAGATACCGAAAAACTGG
	Reverse	ATCTTTGCAGGTAAACGGGGTATCAATC
N172C	Forward	ATGCGTGCATTGAAGCGTGGCCGAAAAG
	Reverse	TTCAATGCACGCATCACCCAGGTTGC
S204C	Forward	ATGCGTGCCTGCTGAAACTGACCCTG
	Reverse	AGCAGGCACGCATCACCAACTTCGC
Y163TAG	Forward	TATTAAACCGTAGGGCCTGGGCAACCTGGGTGATG
	Reverse	TGCCCAGGCCCTACGGTTTAATAAAGCAGCCGCCAAAC

2.2.2 Protein expression

Expression of C158*- IMP-1 mutants (*- G, A and T)

For *in vivo* protein expression, the pET-47b(+) plasmid containing the respective IMP-1 mutant was transformed into *E. coli* BL21(DE3) cells and grown at 37 °C in the presence of 50 µg/mL kanamycin. 10 mL of an overnight culture was used to inoculate 1 L of Luria-Bertani medium containing 50 µg/mL kanamycin and was incubated at 37 °C. When the OD_{600} value of the culture reached 0.6-0.8, protein overexpression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 500 µM $FeCl_3$ was added to the medium to favour the production of [FeZn]-IMP-1. The culture was then incubated overnight at 25 °C. The cells were harvested by centrifugation and the pellet was resuspended in 20 mM MES (pH 6.5) buffer. The pellet was lysed by sonication, centrifuged at 30,000 *g* for one hour and clarified by filtration through a 0.45 µm syringe filter unit (Millipore) prior to purification.

Expression of ^{15}N-labelled- IMP-1 single cysteine mutants

pET-47b(+) plasmid containing the respective IMP-1 cysteine mutant was transformed into *E. coli* BL21(DE3) cells and grown at 37 °C in the presence of 50 µg/mL kanamycin. 10 mL of an overnight culture was used to inoculate 1 L of Luria-Bertani

medium containing 50 μg/mL kanamycin and was incubated at 37 °C until the OD_{600} value reached 0.6-0.8. The cells were then harvested by centrifugation and resuspended in 300 mL of M9 minimal medium containing 0.1% $^{15}NH_4Cl$ as the sole nitrogen source for synthesizing uniformly ^{15}N-labelled samples. 500 μM $ZnSO_4$ was also added to the medium to favour the production of [ZnZn]-IMP-1. Following incubation at 37 °C for another one to two hours, protein overexpression was induced with 1 mM IPTG. The cells were harvested after overnight expression (~16 hours) at 25 °C by centrifugation. The cell pellet was resuspended in 50 mL of 50 mM HEPES (pH 7.5) buffer with 100 μM $ZnSO_4$ and lysed by sonication. The lysate was centrifuged at 30,000 *g* for one hour and then clarified by filtration through a 0.45 μm syringe filter unit (Millipore) prior to purification.

Expression of the IMP-1 double mutants V30Tby and Y163Tby with site-specific incorporation of ^{13}C-Tby

The pET-47b(+) plasmid containing the respective IMP-1 double mutant was co-transformed into *E. coli* BL21(DE3) cells with the pEvol plasmid encoding the *p*-cyanophenylalanyl-tRNA synthetase and the requisite $tRNA_{CUA}$ (Young et al., 2011) for recognition of the TAG stop codon. Cells were grown at 37 °C in the presence of 50 μg/mL kanamycin. 10 mL (0.5 mL for double mutants) of an overnight culture was used to inoculate 1 L (50 mL for double mutants) of Luria-Bertani medium supplemented with 50 μg/mL kanamycin and 0.2% arabinose and was incubated at 37 °C until the OD_{600} value was 0.6-0.8. The cells were harvested by centrifugation and resuspended in 500 mL (50 mL for double mutants) of M9 minimal medium containing 0.2% arabinose, 1 mM ^{13}C-Tby and 500 μM $ZnSO_4$. After switching the medium, the cells were grown at 37 °C for one to two hours and induced with 1 mM IPTG followed by overnight incubation at 25 °C. The cells were then harvested by centrifugation and the pellet was resuspended in 50 mM HEPES (pH 7.5) buffer with 100 μM $ZnSO_4$. The pellet was lysed by sonication following by centrifugation at 30,000 *g* for one hour and clarification by filtration through a 0.45 μm syringe filter unit (Millipore) prior to purification.

Expression of wild-type IMP-1

The pET-47b(+) plasmid containing the gene for wild-type IMP-1 was transformed into *E. coli* BL21(DE3) cells and grown at 37 °C in the presence of 50 μg/mL kanamycin. 10 mL of an overnight culture was used to inoculate 1 L of Luria-Bertani medium containing 50 μg/mL kanamycin and was incubated at 37 °C until the OD_{600} value reached 0.6-0.8. The cells were then harvested by centrifugation and resuspended in 500 mL of M9 minimal medium containing 500 μM $ZnSO_4$ to favour the production of [ZnZn]-IMP-1. Following incubation at 37 °C for another one to two hours, protein overexpression was induced with 1 mM IPTG. The cells were harvested after overnight expression (~16 hours) at 25 °C by centrifugation. The cell pellet was resuspended in 50 mL of 50 mM HEPES (pH 7.5) buffer with 100 μM $ZnSO_4$ and lysed by sonication. The lysate was centrifuged at 30,000 *g* for one hour and then clarified by filtration through a 0.45 μm syringe filter unit (Millipore) prior to purification.

2.2.3 Protein purification

Clarified lysate was loaded onto a 30 mL (or 5 mL for small scale protein expression) SP Sepharose cation exchange column pre-equilibrated with 50 mM HEPES, pH 7.5 with 100 μM $ZnSO_4$ (or 20 mM MES, pH 6.5). The protein was washed with one column volume of buffer and the bound protein was eluted using a linear gradient from 0 to 250 mM NaCl over five column volumes. The eluted fractions were analyzed by 12% SDS-PAGE and the fractions containing protein were pooled. The pure protein was concentrated using an Amicon ultrafiltration centrifugal tube with a molecular weight cut-off of 10 kDa. The protein was stored in 20 mM MES pH 6.5 and 100 mM NaCl at 4 °C.

2.2.4 Sample preparation for ICP-OES analysis

Purified C158* mutant protein samples were exchanged into analysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl) by passing through a PD-10 size exclusion column. The samples were then acidified with concentrated nitric acid (final concentration 4%) and passed through a 0.45 μm filter to remove insoluble denatured protein. The resulting solution was used for ICP-OES analysis.

2.2.5 Protein ligation with C2-Ln^{3+} tag

Protein samples in 20 mM MES pH 6.5 and 100 mM NaCl were reduced with 2 mM dithiothreitol (DTT) for 2 hours and the DTT was washed out using an Amicon ultrafiltration centrifugal tube with a molecular weight cut-off of 10 kDa. The reduced protein samples were then reacted with five-fold excess of C2 tag loaded with either Y^{3+}, Tb^{3+} or $Tm^{3+}$and incubated overnight at room temperature with shaking. Following the tagging reaction, the samples were washed with NMR buffer (20 mM MES, pH 6.5, 100 mM NaCl) using an Amicon centrifugal filter unit to remove unbound tag.

2.2.6 NMR spectroscopy

All 2D NMR data were acquired at 37 °C on a Bruker 800 MHz NMR spectrometer and 1D NMR data were acquired at 25 °C on a Bruker 600 MHz NMR spectrometer using 3 mm NMR tubes. Both spectrometers were equipped with TCI cryoprobes. ^{13}C-HSQC spectra were recorded with protein concentrations of 80-120 μM. Parameters for the ^{13}C-HSQC spectra: t_{1max} = 42.4 ms, t_{2max} = 150.7 ms, total recording time per spectrum between 20–80 min. $^{1}J_{CC}$ couplings were refocused in the t_1 dimension by a selective 180°(^{13}C) pulse on the resonance of the quaternary carbon. Pseudocontact shifts of ^{1}H-NMR signals of amide protons were measured in ^{15}N-HSQC spectra with protein concentrations of 250 μM. Parameters for the ^{15}N-HSQC spectra: t_{1max} = 39.5 ms, t_{2max} = 170.4 ms, total recording time per spectrum 80 min. 1D ^{1}H-NMR spectra were recorded with protein concentrations of 100-200 μM and the total recording time per spectrum was about 10 minutes.

2.2.7 Δχ-tensor fitting

The experimental PCSs ($\Delta\delta^{PCS}$) were measured in ppm as the amide proton chemical shift observed in NMR spectra recorded for the IMP-1 mutants A53C, N172C and S204C tagged with C2-Tm^{3+} or C2-Tb^{3+} minus the corresponding chemical shift measured of samples made with C2-Y^{3+}. The PCS values were used to fit the Δχ tensors to the crystal structure 5EV6 in the absence of the inhibitor and crystal structure 4C1F in the presence of the inhibitor captopril using the program Numbat (Schmitz et al., 2008).

2.2.8 Modelling*

Based on the Δχ tensors obtained with the backbone amides, a computer simulation was performed to optimize the bond angles of the Tby residue to fit the experimental average PCS values of hydrogens of *tert*-butyl group.

Coordinates of the *tert*-butyl group were calculated based on the Δχ tensors determined for the backbone amides, by randomly varying the angles χ_1 (0°–360°), χ_2 (90° and 270°) and the dihedral angle between the phenyl group and the O-*tert*-butyl group of Tby (90° and 270°) and the O-C$(CH)_3$ bond (25°–85°). For each conformation, average PCS values were calculated by averaging over the PCSs of the individual methyl protons of the *tert*-butyl group and six average PCS values were obtained in this way (i.e. PCSs originating from Tb^{3+} and Tm^{3+} tags at the three different tagging sites A53C, N172C and S204C). Calculation of the root mean square deviation (RMSD) values between the calculated and experimental PCS values of the *tert*-butyl group identified the best fitting model.

The above computations were performed three times, once with the set of PCSs obtained for IMP-1 without ligand and twice with the two sets of PCSs obtained in the presence of captopril.

2.2.9 Isothermal calorimetry measurements

Isothermal calorimetry (ITC) measurements were carried out at 25 °C using a Nano ITC titration calorimeter. The inhibitor was used as the titrant (1.64 mM captopril). The protein concentration was 0.15 mM. All the solutions used in the experiment were made with buffer containing 20 mM MES pH 6.5 and 100 mM NaCl and were thoroughly degassed in vacuum for at least 10 minutes before each measurement. During the measurements, the solution in the cell was mixed by rapid stirring with the syringe at 250 rpm. The inhibitor was injected in 22 injections of 2 μL each, with intervals of 250 seconds between injections to allow complete equilibration. The titration was carried out until the inhibitor was present in 3.5-fold molar excess. While analysing the data,

* The calculations were performed by Mr Soumya Behera (RSC, ANU).

the heat of dilution was accounted for by subtracting background titration data obtained with the identical titrant solution but with only buffer solution in the sample cell. The data were analysed with the assumption of 1:1 binding using the NanoAnalyze software v3.7.5 supplied by the manufacturer.

2.3 Results and discussion

2.3.1 ICP-OES analysis of IMP-1 C158* mutants

Charge-transfer complexes often show intense colours arising from electron transitions in the visible region of the electromagnetic spectrum. As samples of [FeZn]-IMP-1 were reported to be purple (Carruthers et al., 2014), exploratory experiments were conducted to determine whether the purple colour arises from a charge transfer interaction between Fe^{3+} and the active site cysteine.

A previous study, which attempted to understand the requirement for cysteine at position 158 for the structure and function of IMP-1, reported that only Asp, Gly, Ser, Thr, Val and Ala substitutions allowed successful protein expression (Horton et al., 2012). The dizinc subclass B3 MBLs contain either an aspartic acid or a serine residue at position 158. However the residue 158 does not form part of the Zn2 site. To explore the role of cysteine in colour formation in the present study, the active site cysteine residue was mutated to glycine, alanine and threonine.

The mutants (C158G, C158A and C158T) were produced with the addition of 500 μM $FeCl_3$ in the culture medium during protein expression to favour the formation of [FeZn]-IMP-1. After purification, the mutant protein samples were visually compared with the wild-type [FeZn]-IMP-1 protein in similar protein concentrations (Figure 2.7). While wild-type [FeZn]-IMP-1 displayed purple colouration, all mutants of Cys158 were colourless.

Figure 2.7 Purified [FeZn]-IMP1 samples. Left: wild-type [FeZn]-IMP1. Right: C158G [FeZn]-IMP1 mutant.

To investigate whether the mutations compromised the uptake of metal ions, the metal content of each mutant protein sample was analysed by ICP-OES. The results showed that the C158G mutant contained both Fe and Zn, whereas C158A and C158T contained only Zn. Haruta et al. (2000) reported that the replacement of cysteine with alanine produces a mononuclear form of the protein, which is consistent with our result (Table 2.2). Moreover, alanine and threonine are not good ligands for coordinating Zn^{2+} ions. Carruthers et al. (2014) demonstrated that Fe^{3+} occupies the Zn1 site. Since the Zn2 site appears to be deficient in metal binding in the C158A and C158T mutants, it can be suggested that the Zn1 site prefers a Zn^{2+} ion over a Fe^{3+} ion.

Table 2.2 ICP-OES results obtained with mutants at position 158.

Protein Sample	Fe:Zn ratio
Cys158Gly	1:0.6
Cys158Ala	0.11:1
Cys158Thr	0.14:1

In light of the results obtained for the C158A and C158T mutants, it was surprising to observe metal ions in both sites of the C158G mutant, as glycine also would not be a Zn binding ligand. Notably, however, Horton et al. (2012) reported that IMP-1 C158G displays considerable catalytic activity compared to the wild-type protein. Furthermore, molecular modelling studies suggest that substitution of glycine provides the space for a water molecule to replace the thiol group of cysteine and coordinate zinc in the Zn2 site to restore the function of the enzyme (Horton et al., 2012). Therefore, the IMP-1 C158G mutant may very well contain Fe^{3+} in the Zn1 site and Zn^{2+} in the Zn2 sites, but with reduced binding affinity. Furthermore, as the sample was colourless even in the presence of bound Fe^{3+} ions, these experiments support the hypothesis that the observed purple colouration of [FeZn]-IMP1 arises from a charge transfer complex involving the active site cysteine residue and the Fe^{3+} ion.

2.3.2 Study of active site loop flexibility by NMR spectroscopy

Flexibility of active site loops in subclass B1 MBLs is well documented by published crystal structures (Concha et al., 2000; Ullah et al., 1998; Fitzgerald et al., 1998; Toney et al., 1998; Scrofani et al., 1999). Using PCSs, it should be possible to observe the corresponding conformational changes in solution, which may be important for enzymatic activity and the development of clinically useful MBL inhibitors with a broad spectrum of activity.

According to the crystal structures of IMP-1, it is clear that the active site loop is involved in substrate binding (Figure 2.8). Using paramagnetic tags, its conformational changes upon binding of the inhibitor captopril were studied in solution by NMR spectroscopy.

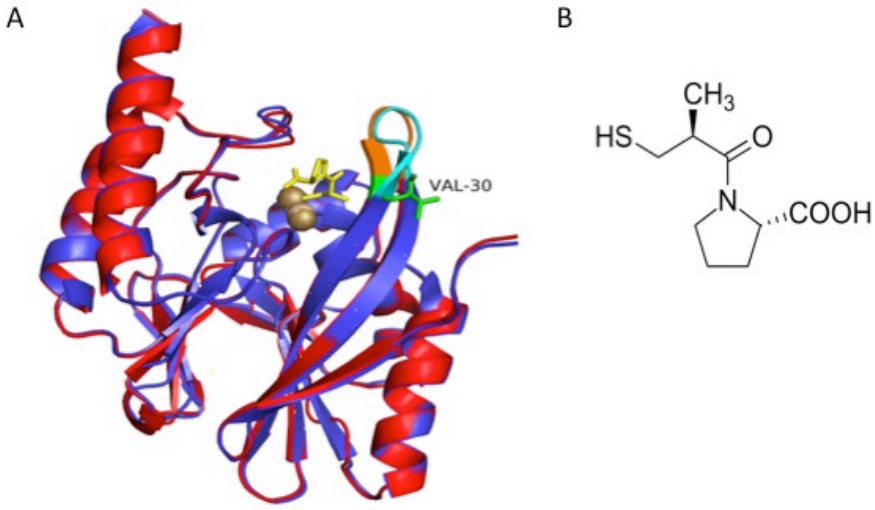


Figure 2.8 (A) Overlay of crystal structures of IMP-1 determined in the presence (blue with the loop 1 highlighted in cyan; PDB ID: 4C1F; Brem et al., 2015) and absence of captopril (red with the loop highlighted in orange; PDB ID: 5EV6; Hinchliffe et al., 2016). Zn atoms are shown as spheres and the captopril molecule as yellow sticks. The side chain of Val30 in the active site loop, which was mutated to Tby, is displayed by green sticks. (B) Chemical structure of captopril.

Three sites were selected on IMP-1 (A53, N172 and S204; Figure 2.9) to generate PCSs in the protein, by mutation to cysteine and attachment of a lanthanide binding tag. This

study used the C2 tag (Figure 1.4; de la Cruz et al., 2011) loaded with paramagnetic Tb^{3+} or Tm^{3+}, and the C2 tag with Y^{3+} was used as the diamagnetic reference.

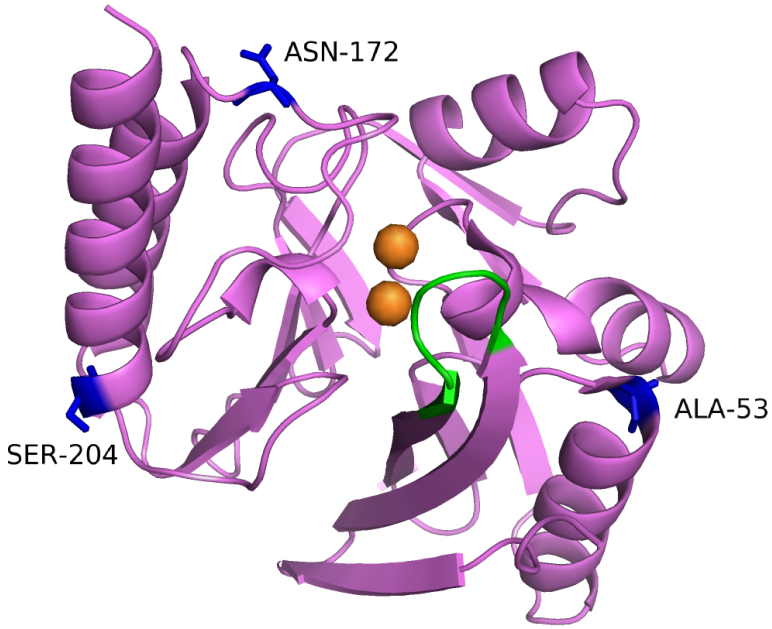


Figure 2.9 Crystal structure of IMP-1 β-lactamase (PDB ID: 5EV6; Hinchliffe et al., 2016). Orange spheres indicate zinc atoms. Side chains shown in blue are the sites mutated to cysteines to attach the C2 tag for PCS measurements. The loop 1 is highlighted in green.

NMR spectroscopy

In order to measure PCSs, ^{15}N-HSQC spectra were measured of uniformly ^{15}N-labelled IMP-1 with the C2 tag loaded with Tb^{3+}, Tm^{3+} or Y^{3+} attached to the sites A53C, N172C or S204C (Figure 2.10). The PCSs measured are listed in Table 2.3 The measurements were repeated in the presence of the inhibitor captopril (Figure 2.8B). The ^{15}N-HSQC spectra are shown in Figure 2.11 and the PCSs extracted are listed in Table 2.4.

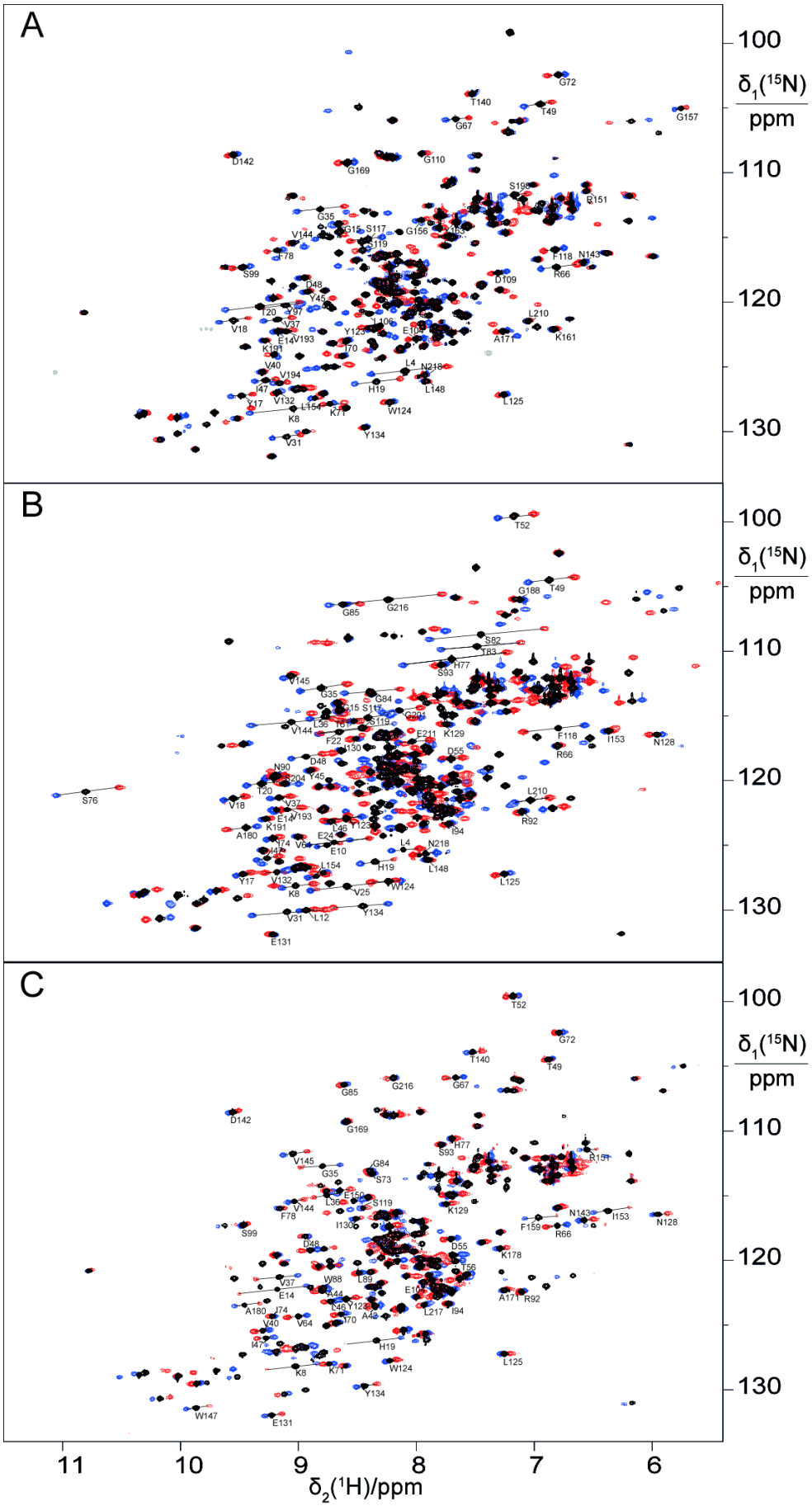

A
B
C
100
110
120
130
δ1(15N)
ppm
11
10
9
8
7
6
δ2(1H)/ppm

Figure 2.10 Superimpositions of [^{15}N,^{1}H]-HSQC spectra recorded of uniformly ^{15}N-labelled [ZnZn]-IMP-1 tagged with C2-Ln^{3+} at three different sites of [ZnZn]-IMP-1. A-C show the spectra with tags at sites A53C, N172C and S204C, respectively. Spectra with diamagnetic tag (C2-Y^{3+}) are shown in black and the corresponding spectra with paramagnetic tags are shown in red (C2-Tb^{3+}) and blue (C2-Tm^{3+}). All spectra were measured at 310 K in buffer containing 20 mM MES, pH 6.5, 100 mM NaCl.

Table 2.3 Pseudocontact shifts measured in ppm for backbone amide protons of IMP-1 cysteine mutants ligated with C2 tag loaded with Tb^{3+} and Tm^{3+}.

		A53C-C2		**N172C-C2**		**S204C-C2**	
Residue		**Tb^{3+}**	**Tm^{3+}**	**Tb^{3+}**	**Tm^{3+}**	**Tb^{3+}**	**Tm^{3+}**
4	Leu	-0.344	0.289	-0.135	0.102		-0.044
8	Lys	-0.428	0.365	-0.148	0.120	0.234	-0.188
10	Glu			-0.106	0.083		
12	Leu			-0.069	0.061		
14	Glu	-0.029	0.036	-0.042	0.033	0.321	-0.244
15	Gly	-0.049	0.054	-0.030	0.025		-0.183
17	Tyr	-0.097	0.091	-0.066	0.058		
18	Val	-0.117	0.120	-0.103	0.081		
19	His	-0.199	0.181	-0.155	0.125	0.245	-0.207
20	Thr	-0.330	0.291	-0.162	0.127		
22	Phe			-0.231	0.178		-0.089
24	Glu			-0.294	0.219		-0.027
25	Val			-0.417	0.310		-0.009
31	Val	-0.111	0.125	-0.389	0.291		-0.043
35	Gly	-0.204	0.195	-0.222	0.178	0.195	-0.147
36	Leu					0.119	-0.082
37	Val	-0.116	0.116	-0.158	0.135	0.206	-0.156
40	Val	-0.008	0.015	-0.094	0.078	0.069	-0.054
42	Asn					0.061	-0.045
44	Ala					0.066	-0.040
45	Tyr	-0.017	0.034	-0.025	0.023		
46	Leu			-0.031	0.034	0.056	-0.046
47	Ile	-0.083	0.093	-0.098	0.088	0.080	-0.064
48	Asp	-0.021	0.056	-0.143	0.125	0.031	-0.024
49	Thr	-0.099	0.137	-0.210	0.177	0.029	-0.034
52	Thr			-0.170	0.136	0.058	-0.048
55	Asp			-0.099	0.079	0.064	-0.052
56	Thr					0.069	-0.057
59	Leu					0.106	-0.083
61	Thr			-0.041	0.033	0.090	-0.069

64	Val	-0.229	0.179	-0.026	0.022	0.093	-0.070
66	Arg	-0.172	0.136	-0.026	0.017	0.099	-0.077
67	Gly	-0.108	0.083			0.083	-0.070
70	Ile	0.040	-0.020			0.056	-0.044
71	Lys	0.061	-0.033			0.047	-0.037
72	Gly	0.095	-0.055			0.037	-0.033
73	Ser					0.018	-0.019
74	Ile			-0.051	0.054	0.015	-0.017
76	Ser			-0.283	0.247		
77	His	0.050		-0.465	0.409	-0.033	0.014
78	Phe	0.083	-0.054			-0.028	0.022
82	Ser			-0.537	0.435		
83	Thr			-0.373	0.302		
84	Gly			-0.256	0.212	0.024	-0.023
85	Gly			-0.138	0.119	0.032	-0.028
88	Trp					0.029	-0.025
89	Leu					0.034	-0.026
90	Asn			0.042	-0.030		
92	Arg			0.037	-0.028	0.022	-0.017
93	Ser			0.051	-0.039	0.014	-0.013
94	Ile			0.038	-0.025	0.022	-0.021
97	Tyr	0.201	-0.139				
99	Ser	0.090	-0.058			-0.030	0.018
104	Glu	0.067	-0.061			-0.019	0.010
106	Leu	0.094	-0.087			-0.012	0.006
109	Asp	0.054	-0.074				
110	Gly	-0.047					
114	Ala			0.174	-0.123		
117	Ser	0.154	-0.119	0.234	-0.182		
118	Phe	0.099	-0.072	0.297	-0.225		
119	Ser	0.058	-0.050	0.354	-0.266	-0.031	0.021
120	Gly					-0.051	0.031
123	Tyr	0.036	-0.019	0.190	-0.131	-0.081	0.055
124	Trp	0.047	-0.033	0.145	-0.106	-0.055	0.039
125	Leu	0.047	-0.028	0.076	-0.052	-0.047	0.029
128	Asn			0.067	-0.048	-0.069	0.053
129	Lys			0.044	-0.029	-0.051	0.033
130	Ile			0.040	-0.026	-0.035	0.028
131	Glu			0.030	-0.016	-0.083	0.059
132	Val	0.030	-0.018	0.102	-0.069		
133	Phe			0.070	-0.028	-0.116	0.077
134	Tyr	0.022	0.014	0.320	-0.217	-0.103	0.072
140	Thr	0.034	-0.018			-0.077	0.050
142	Asp	0.049	-0.031			-0.045	0.030
143	Asn	0.041	-0.023			-0.076	0.046
144	Val	0.029	-0.008	-0.351	0.336	-0.084	0.055

145	Val			-0.025	0.059	-0.131	0.084
147	Trp					-0.103	0.079
148	Leu	-0.015	0.023	-0.018	0.025		
150	Glu					-0.089	0.069
151	Arg	-0.009	0.014			-0.066	0.053
153	Ile			-0.032	0.030	-0.182	0.120
154	Leu	-0.029	0.036	-0.080	0.073		
156	Gly	-0.048	0.047				
157	Gly	-0.050	0.051				
159	Phe					-0.155	0.121
160	Ile	-0.023	0.030				0.190
161	Lys	-0.025	0.018				
163	Tyr	-0.041	0.028				
169	Gly	0.079	-0.057			-0.010	0.017
171	Ala	0.056	-0.039			-0.025	0.015
178	Lys					-0.045	0.031
180	Ala			0.167	-0.092	-0.119	0.085
181	Lys					-0.127	0.074
188	Gly			0.053	-0.035		
189	Lys			0.031	-0.016		
191	Lys	-0.020	0.020	-0.036	0.027		
192	Leu			-0.062	0.053		
193	Val	-0.060	0.052	-0.122			
194	Val	-0.066	0.056	-0.174	0.145		
198	Ser	-0.108	0.091				
201	Gly			-0.176	0.141		
204	Ser			-0.089	0.073		
210	Leu	-0.010	0.011	-0.158	0.134		
211	Glu			-0.148	0.118		
213	Ala			-0.355	0.281		0.088
216	Gly			-0.454	0.354	0.058	-0.022
217	Leu					0.044	-0.027

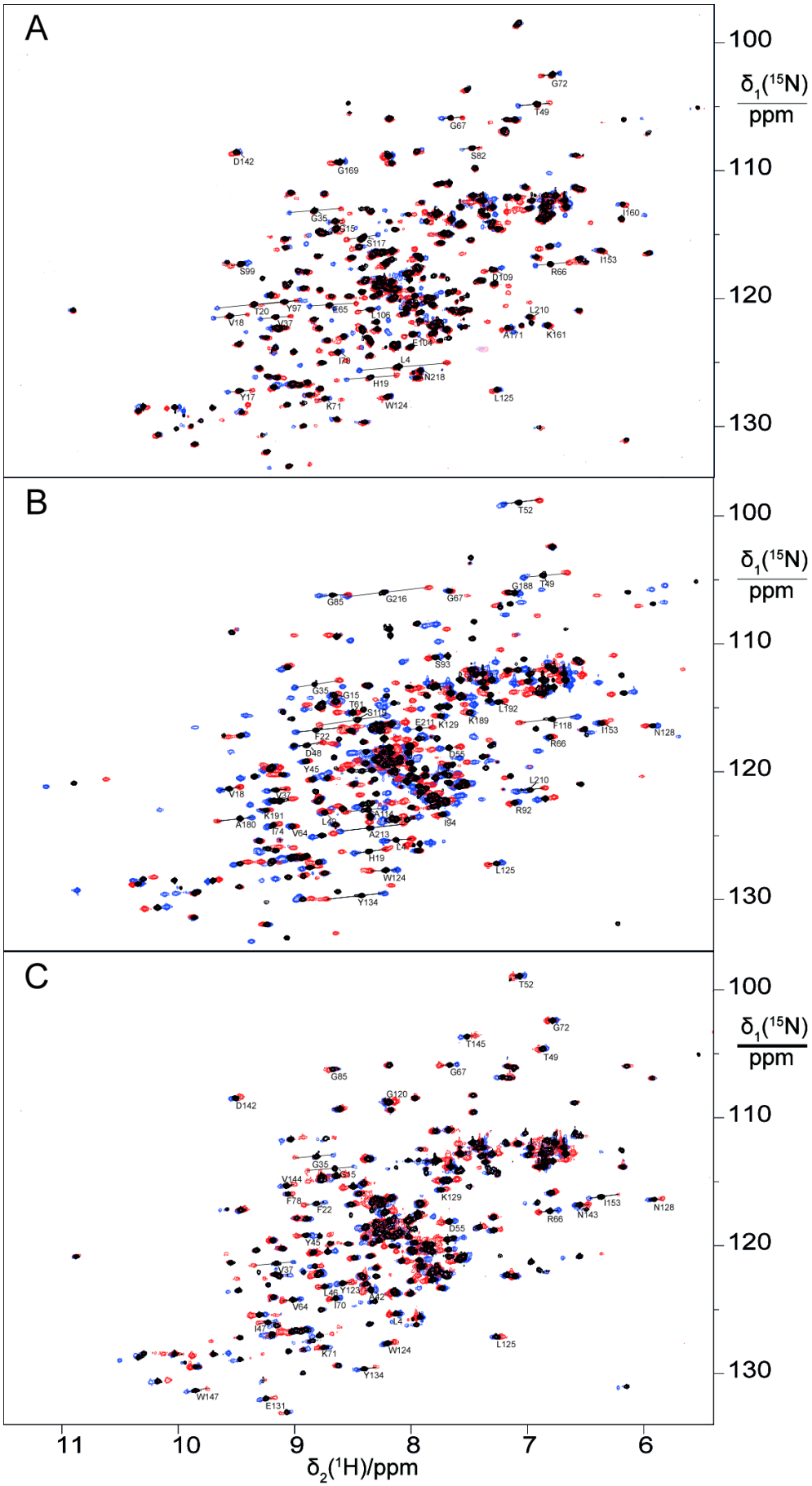

A
B
C
100
110
120
130
δ1(15N)
ppm
11
10
9
8
7
6
δ2(1H)/ppm

Figure 2.11 Same as Figure 2.10, except that the spectra were recorded following addition of 1.5-fold excess captopril.

Table 2.4 Pseudocontact shifts measured in ppm for backbone amide protons of IMP-1 cysteine mutants ligated with C2 tag loaded with Tb^{3+} and Tm^{3+} in the presence of 1.5-fold excess of the inhibitor captopril.

		A53C-C2		N172C-C2		S204C-C2	
Residue		Tb^{3+}	Tm^{3+}	Tb^{3+}	Tm^{3+}	Tb^{3+}	Tm^{3+}
4	Leu	-0.414	0.341	-0.138	0.104	0.054	-0.032
15	Gly	-0.062	0.052	-0.033	0.026	0.236	-0.162
17	Tyr	-0.112	0.104				
18	Val	-0.149	0.140	-0.094	0.081		
19	His	-0.221	0.210	-0.151	0.121		
20	Thr	-0.363	0.316				
22	Phe			-0.229	0.170	0.098	-0.080
35	Gly	-0.227	0.214	-0.209	0.173	0.178	-0.137
37	Val	-0.128	0.132	-0.089	0.077	0.192	-0.150
40	Val					0.070	-0.052
42	Asn					0.064	-0.043
45	Tyr			-0.023	0.021	0.075	
46	Leu			-0.031	0.031	0.051	-0.044
47	Ile					0.085	-0.068
48	Asp			-0.137	0.118		
49	Thr	-0.102	0.166	-0.208	0.169	0.031	-0.022
52	Thr			-0.175	0.137	0.052	-0.040
55	Asp			-0.101	0.082	0.068	-0.042
61	Thr			-0.036	0.031		
64	Val			-0.026	0.022	0.086	-0.064
65	Glu	-0.228	0.167				
66	Arg	-0.164	0.129	-0.025	0.019	0.099	-0.073
67	Gly	-0.101	0.078	-0.015	0.011	0.086	-0.059
70	Ile	0.055	-0.032			0.053	-0.041
71	Lys	0.064	-0.043			0.041	-0.032
72	Gly	0.095	-0.063			0.040	-0.031
74	Ile			-0.053	0.049		
76	Ser			-0.274	0.241		
78	Phe					-0.031	0.037
82	Ser	-0.068	0.092				
85	Gly			-0.134	0.111	0.038	-0.026
92	Arg			0.042	-0.024		
93	Ser			0.053	-0.041		
94	Ile			0.039	-0.029		

97	Tyr	0.205	-0.140				
99	Ser	0.087	-0.052				
104	Glu	0.072	-0.051				
106	Leu	0.107	-0.067				
109	Asp	0.086	-0.061				
114	Ala			0.196	-0.140		
117	Ser	0.158	-0.117				
118	Phe	0.095	-0.067	0.281	-0.218		
119	Ser	0.062	-0.035	0.322	-0.246		
120	Gly					-0.050	0.025
123	Tyr					-0.080	0.047
124	Trp	0.051	-0.024	0.138	-0.098	-0.050	0.030
125	Leu	0.037	-0.025	0.071	-0.047	-0.039	0.022
128	Asn			0.062	-0.039	-0.069	0.039
129	Lys			0.045	-0.024	-0.039	0.026
131	Glu					-0.074	0.049
134	Tyr			0.306	-0.193	-0.084	0.066
140	Thr					-0.074	0.049
142	Asp	0.039	-0.020			-0.047	0.031
143	Asn					-0.069	0.042
144	Val					-0.064	0.048
145	Val					-0.103	0.077
147	Trp					-0.098	0.068
153	Ile	-0.023	0.037	-0.023	0.028	-0.150	0.101
160	Ile	-0.026	0.043				
161	Lys	-0.011	0.024				
169	Gly	0.069	-0.048				
171	Ala	0.050	-0.035				
180	Ala			0.199	-0.100		
188	Gly			0.063	-0.035		
189	Lys			0.034	-0.018		
191	Lys			-0.023	0.032		
192	Leu			-0.055	0.049		
210	Leu	-0.008	0.014	-0.125	0.115		
211	Glu			-0.112	0.102		
213	Ala			-0.292	0.250		
216	Gly			-0.383	0.310		
217	Leu	0.018	-0.005				

Figure 2.12 shows that the PCSs of the backbone amides were similar in the presence or absence of the inhibitor. This shows that the presence of the inhibitor has very little influence on the $\Delta\chi$ tensors, indicating conservation of the protein structure and the conformations of the tags. In general, more PCSs could be assigned in the free protein than in the complex, as the spectra of the free protein were better resolved.

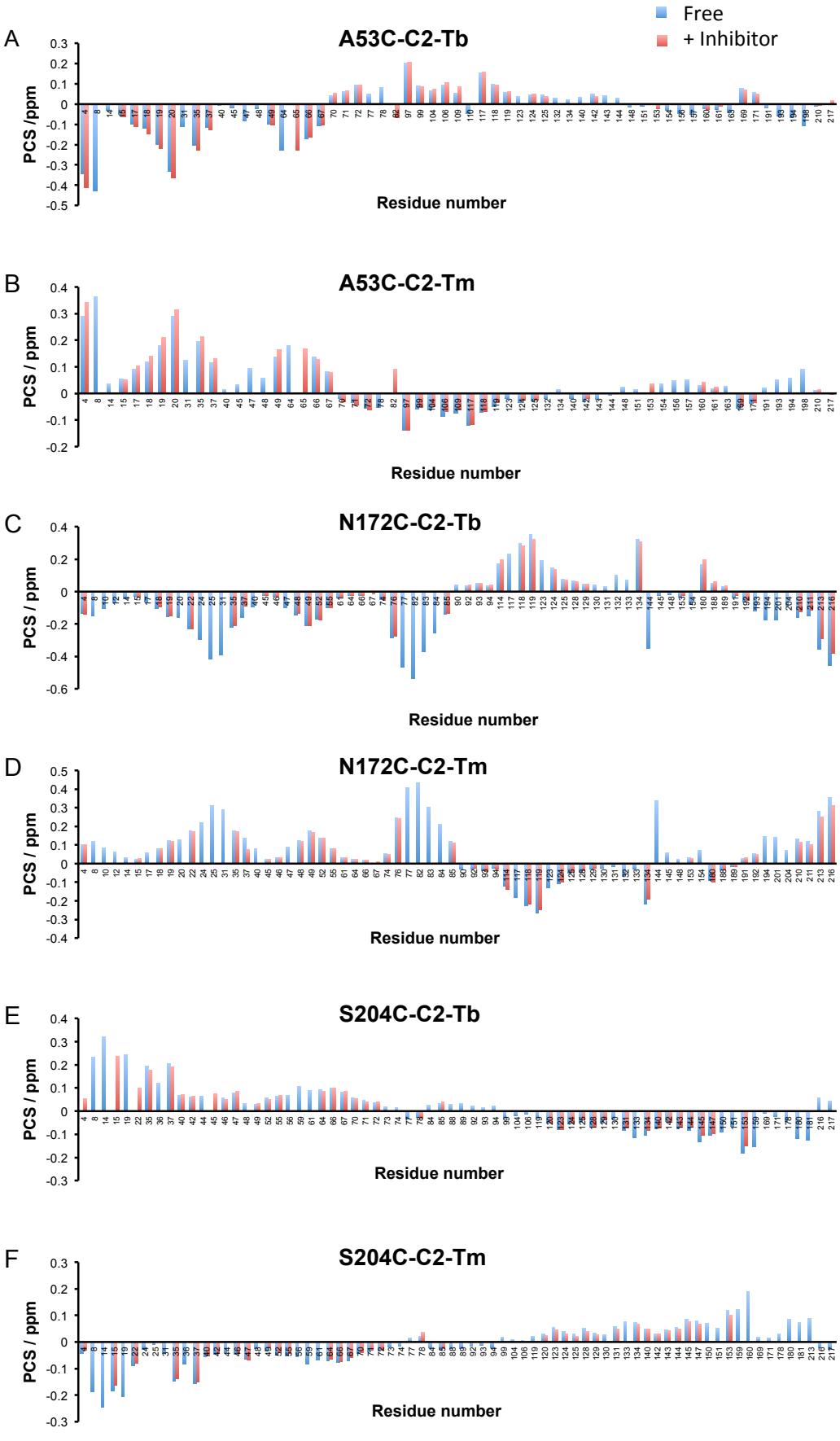

A
A53C-C2-Tb
Free
+ Inhibitor
PCS /ppm
Residue number
B
A53C-C2-Tm
PCS / ppm
Residue number
C
N172C-C2-Tb
PCS / ppm
Residue number
D
N172C-C2-Tm
PCS / ppm
Residue number
E
S204C-C2-Tb
PCS / ppm
Residue number
F
S204C-C2-Tm
PCS / ppm
Residue number

Figure 2.12 Comparison of experimental PCSs obtained for the backbone amide protons of IMP-1 in the presence or absence of captopril. Blue bars represent PCSs of free IMP-1and red bars represent PCSs of the IMP-1-inhibitor complex. (A)–(F) Data for C2 Tb^{3+} and Tm^{3+} tags attached at three different sites as indicated.

Tby residues in proteins present outstanding probes that are readily detected by NMR (Chen et al., 2015). By substituting the solvent-exposed residue Val30 in loop 1 (see Figure 2.8) by ^{13}C-Tby, the signal of the *tert*-butyl group was monitored by ^{13}C-HSQC spectra and PCSs generated by C2 lanthanide tags attached to cysteine in the mutants A53C, N172C and S204C were measured. Figure 2.13 shows the ^{13}C-HSQC spectra acquired with and without the inhibitor.

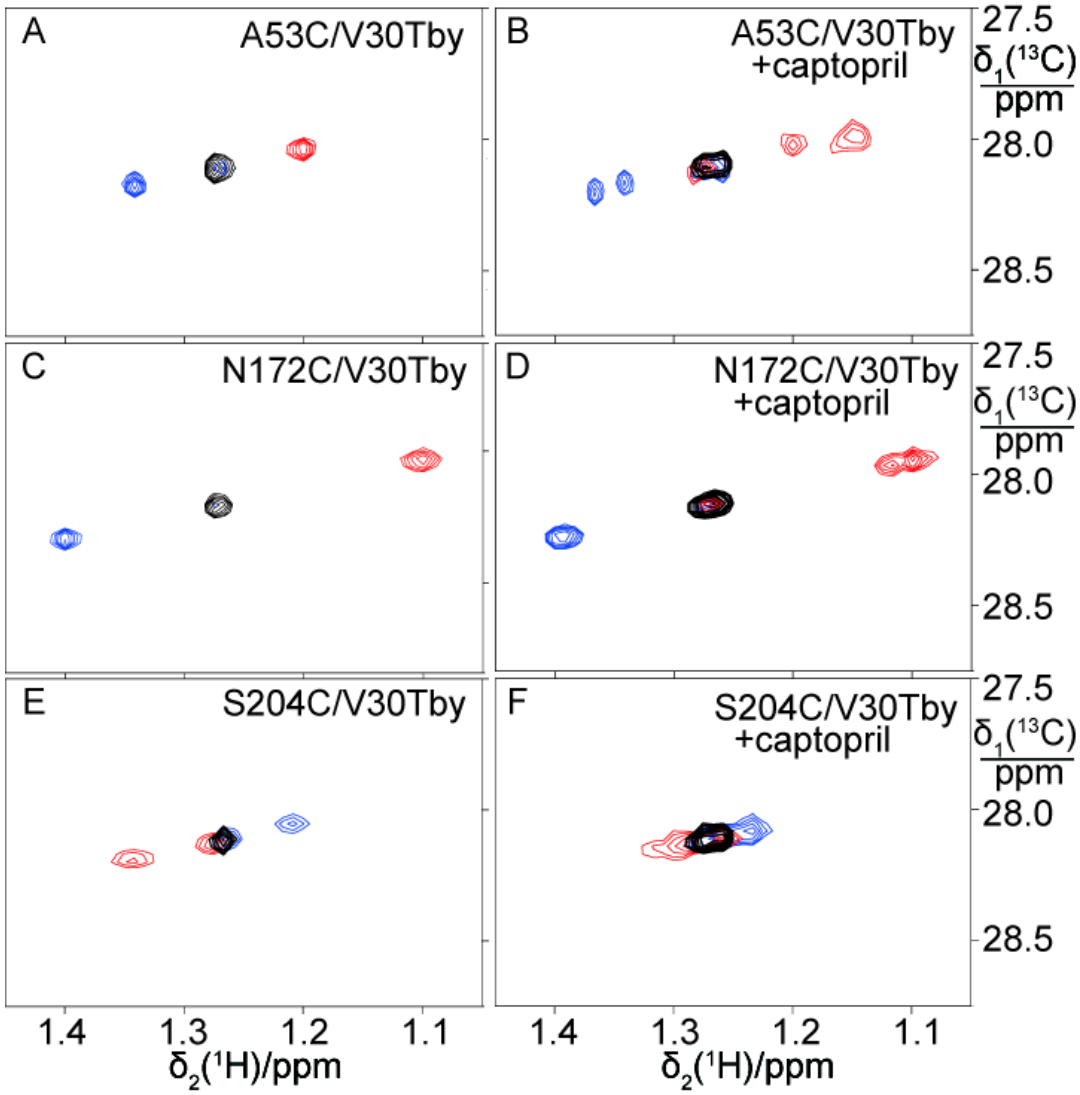


Figure 2.13 Superimpositions of the spectral region of [^{13}C,^{1}H]-HSQC spectra of [ZnZn]-IMP-1 V30^{13}C-Tby showing the cross-peaks of the *tert*-butyl group. The samples were tagged with C2-Ln^{3+} tags at three different sites and measured in the presence or absence of captopril. The *tert*-butyl group of Tby was uniformly labelled with ^{13}C. Spectra with diamagnetic tag (C2-Y^{3+}) are shown in black, and the corresponding spectra with paramagnetic tags are shown in red (C2-Tb^{3+}) and blue (C2-Tm^{3+}). The spectra were recorded at 310 K in 20 mM MES, pH 6.5, 100 mM NaCl. (A) A53C mutant. (B) A53C mutant with captopril. (C) N172C mutant. (D) N172C mutant with captopril. (E) S204C mutant. (F) S204C mutant with captopril.

For all three sites, a single *tert*-butyl resonance was observed in the absence of the inhibitor, which implies that the *tert*-butyl group has only a single conformation or that any different conformations exchange rapidly with each other. Upon addition of the inhibitor, however, two cross-peaks were observed for the *tert*-butyl group. For most of the samples the peak separation is unambiguous. This result suggests that the loop assumes two different conformations upon binding to the inhibitor.

Additionally, overlap of low intensity *tert*-butyl cross-peaks arising from paramagnetic samples with the cross-peaks resulting from diamagnetic samples were also observed in all spectra. This undesirable observation was due to incomplete attachment of the C2 tag on to the protein.

Δχ-tensor determination

Δχ tensors were determined using the software Numbat (Schmitz et al., 2008) using the the PCSs measured for the backbone amide protons (Table 2.3 and 2.4). The Δχ-tensor parameters are presented in Tables 2.5 and 2.6, and the correlations between back-calculated and experimental PCSs are shown in Figures 2.14 and 2.15 Good correlations were obtained between back-calculated and experimental PCSs for all twelve tensors (three mutants, two different paramagnetic metal ions, with and without captopril).

Table 2.5 $\Delta\chi$-tensor parameters determined for IMP-1 in the absence of inhibitor using the crystal structure 5EV6.

	A53C		N172C		S204C	
	Tb^{3+}	Tm^{3+}	Tb^{3+}	Tm^{3+}	Tb^{3+}	Tm^{3+}
$\Delta\chi_{ax}/(10^{-32}\ m^3)$	-12.6	9.8	-18.2	13.9	8.9	-6.3
$\Delta\chi_{rh}/(10^{-32}\ m^3)$	-5.9	4.6	-4.7	3.6	4.0	-2.7
x/Å	57.389	57.389	40.869	40.869	67.761	67.761
y/Å	106.352	106.352	77.901	77.901	68.476	68.476
z/Å	30.945	30.945	25.202	25.202	49.839	49.839
$\alpha/^o$	156	150	22	20	47	52
$\beta/^o$	131	130	29	31	103	104
$\gamma/^o$	79	75	172	172	65	71

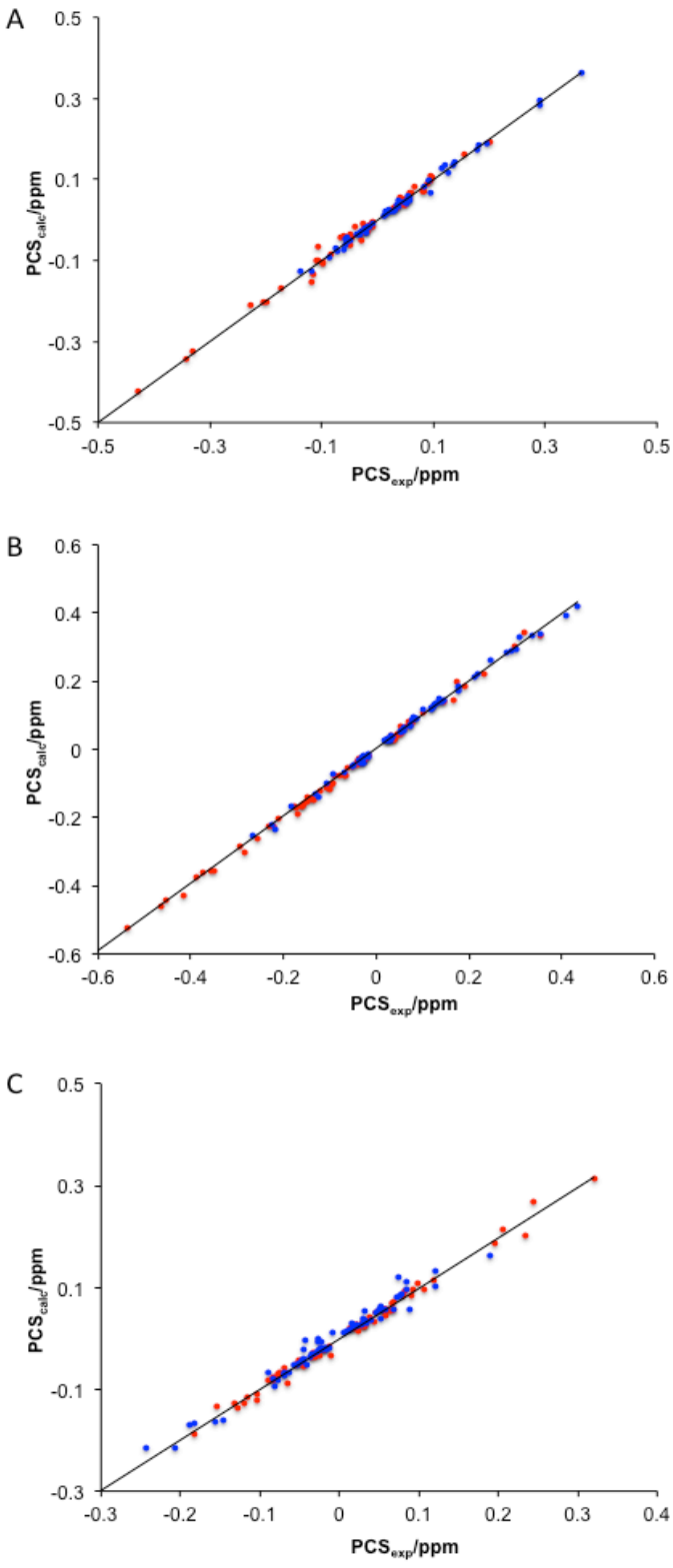


Figure 2.14 Correlation between back-calculated and experimental pseudocontact shifts of backbone amide protons of IMP-1 in the absence of inhibitor for the mutants (A) A53C-C2, (B) N172C-C2 and (C) S204C-C2. Red and blue points mark the PCSs obtained with tags loaded with Tb^{3+} and Tm^{3+}, respectively.

Table 2.6 $\Delta\chi$-tensor parameters determined for IMP-1 in the presence of inhibitor using the crystal structure 4C1F.

	A53C		N172C		S204C	
	Tb^{3+}	Tm^{3+}	Tb^{3+}	Tm^{3+}	Tb^{3+}	Tm^{3+}
$\Delta\chi_{ax}/(10^{-32}\ m^3)$	-12.8	10.4	-18.4	13.8	9.0	-6.7
$\Delta\chi_{rh}/(10^{-32}\ m^3)$	-6.2	5.4	-5.3	4.0	1.8	-0.7
x/Å	2.689	2.689	10.398	10.398	20.790	20.790
y/Å	31.224	31.224	1.058	1.058	0.217	0.217
z/Å	13.786	13.786	23.902	23.902	-15.929	-15.929
$\alpha/^o$	175	167	43	43	121	119
$\beta/^o$	17	18	133	135	92	91
$\gamma/^o$	109	114	41	40	22	15

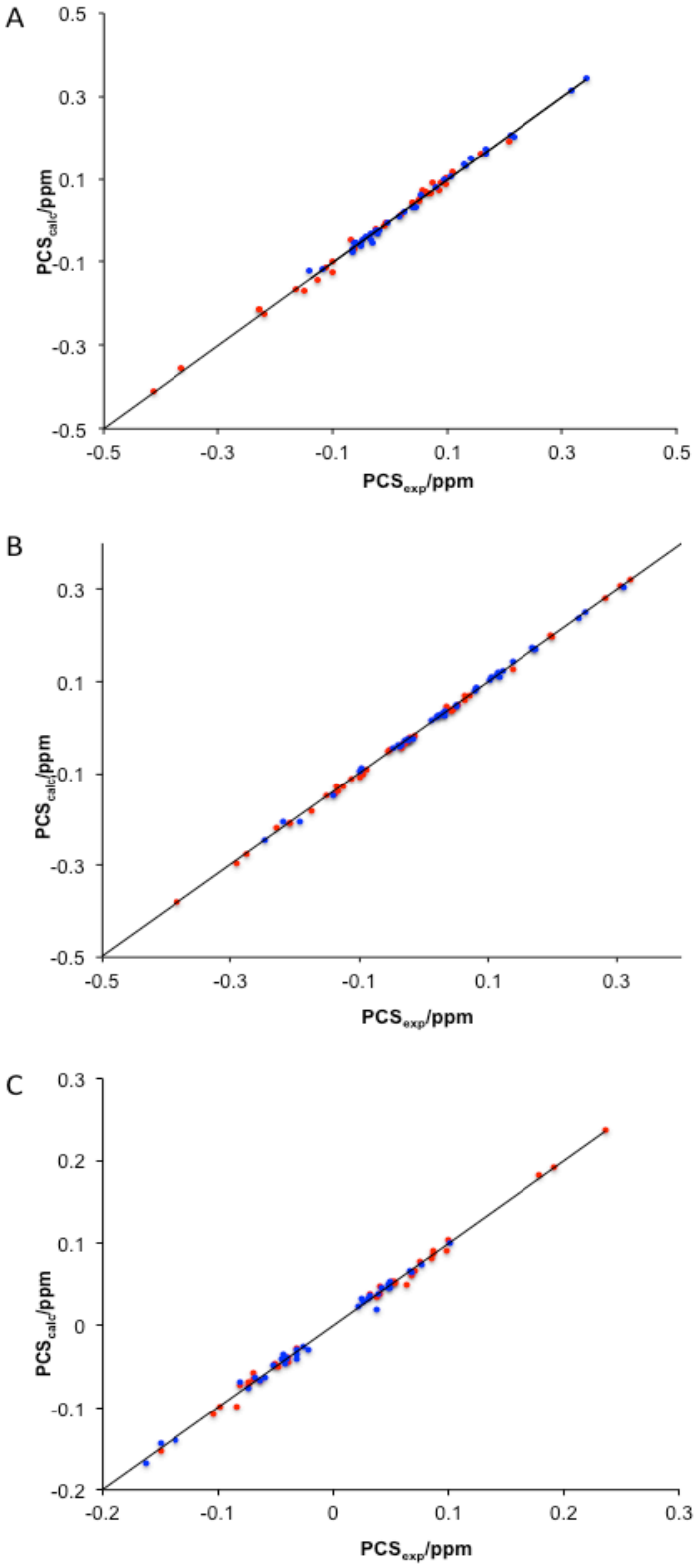


Figure 2.15 Correlation between back-calculated and experimental pseudocontact shifts of backbone amide protons of IMP-1 in the presence of captopril for the mutants (A) A53C-C2, (B) N172C-C2 and (C) S204C-C2. Red and blue points mark the PCSs obtained with tags loaded with Tb^{3+} and Tm^{3+}, respectively.

Figure 2.16 compares the positions of the fitted lanthanide positions determined from the PCSs measured in the presence or absence of the inhibitor. The metal positions are very similar for both states of the protein.

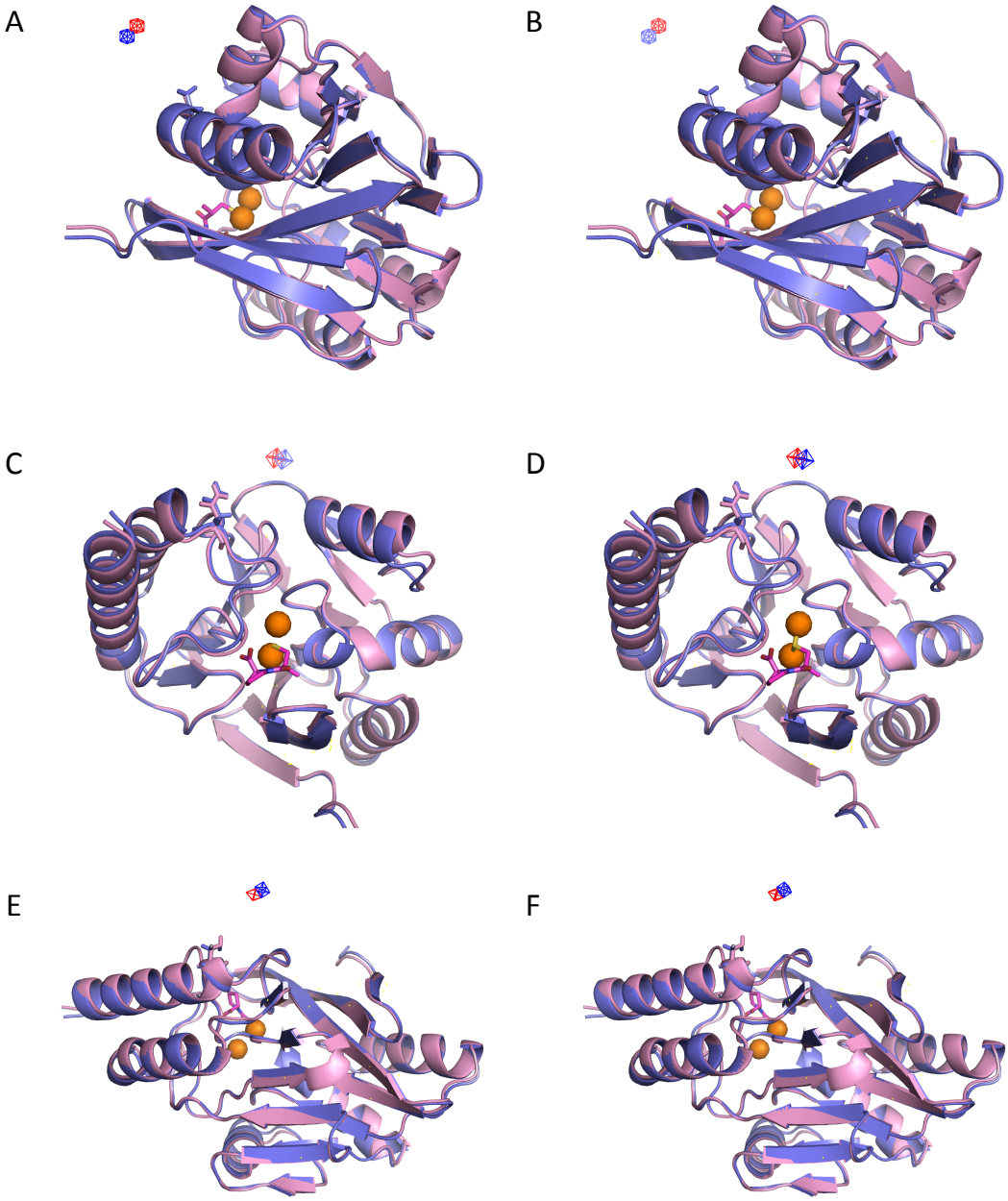


Figure 2.16 Superimposition of IMP-1 crystal structures determined in the presence or absence of captopril (PDB ID 4C1F and 5EV6 respectively), displaying the metal positions determined by the $\Delta\chi$-tensor fits. The structure in the absence of captopril is shown in pink (lanthanide position in red) and the structure in the presence of captopril

is shown in blue (lanthanide position in blue). The zinc ions in the active site are shown as orange spheres and the inhibitor is shown by magenta sticks. Results are shown separately for the different mutants tagged with Tb^{3+} or Tm^{3+} ions: (A) A53C-C2-Tb^{3+}; (B) A53C-C2-Tm^{3+}; (C) N172C-C2-Tb^{3+}; (D) N172C-C2-Tm^{3+}; (E) S204C-C2-Tb^{3+}; (F) S204C-C2-Tm^{3+}. The metal coordinates reported in Table 2.5 and 2.6 were determined by $\Delta\chi$-tensor fits, which simultaneously used the PCSs generated by Tb^{3+} or Tm^{3+} ions.

Modelling *

Using the $\Delta\chi$ tensors determined from the PCSs of the backbone amide protons, the location of the *tert*-butyl group of the Tby residue in position 30 was determined from its PCSs. The PCSs used for the calculations are listed in Table 2.7. As an additional restraint, the search was limited to sites accessible to the *tert*-butyl group in a library of conformations of the Tby residue modeled onto the protein structure.

Table 2.7 ^{1}H-pseudocontact shifts in ppm of the *tert*-butyl group of IMP-1 Val30Tby cysteine mutants ligated with the C2 tag loaded with Tb^{3+} and Tm^{3+} in the absence or presence of 1.5-fold excess of captopril.

	Free IMP-1		**IMP-1-captopril complex**			
			Set 1		**Set 2**	
	Tb^{3+}	**Tm^{3+}**	**Tb^{3+}**	**Tm^{3+}**	**Tb^{3+}**	**Tm^{3+}**
A53C	-0.073	0.067	-0.072	0.069	-0.115	0.105
N172C	-0.172	0.126	-0.150	0.124	-0.249	0.124
S204C	0.076	-0.058	0	-0.035	0	-0.035

- For the PCSs obtained in the presence of captopril, smaller PCS values were taken as set 1 and the larger PCS values were taken as set 2.

* The calculations were performed by Mr Soumya Behera (RSC, ANU).

Figure 2.17 displays the result obtained based on the data measured in the absence of the inhibitor. The volume shown in blue depicts the space, where the RMSD between back-calculated and experimental PCSs is sufficiently small to be within the accuracy of the PCS measurement. It can be seen that the side chain of the Tby residue contacts the side chain of Leu4 at the N-terminus of the protein.

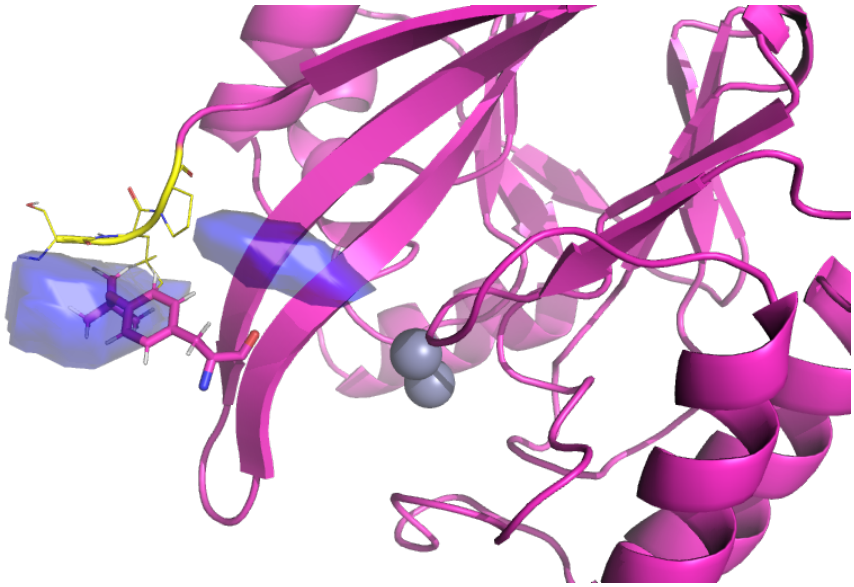


Figure 2.17 Model of IMP-1 V30^{13}C-Tby in the absence of the inhibitor. The protein backbone presents the crystal structure 5EV6. The Tby residue is shown in a stick representation. The cloud plotted in blue shows the localization space for the *tert*-butyl group determined from its PCSs. The residue Leu4 is shown by a line representation in yellow.

In the presence of the inhibitor, two sets of PCS values were obtained for the *tert*-butyl group. Figure 2.18 depicts the localization spaces of the *tert*-butyl group calculated from these two sets.

The blue localization space shown in Figure 2.18 is similar to that found for the *tert*-butyl group in the absence of the inhibitor. The Tby residue interacts with the residue Leu4 near the N-terminus of the protein in a similar way as observed in the absence of the inhibitor.

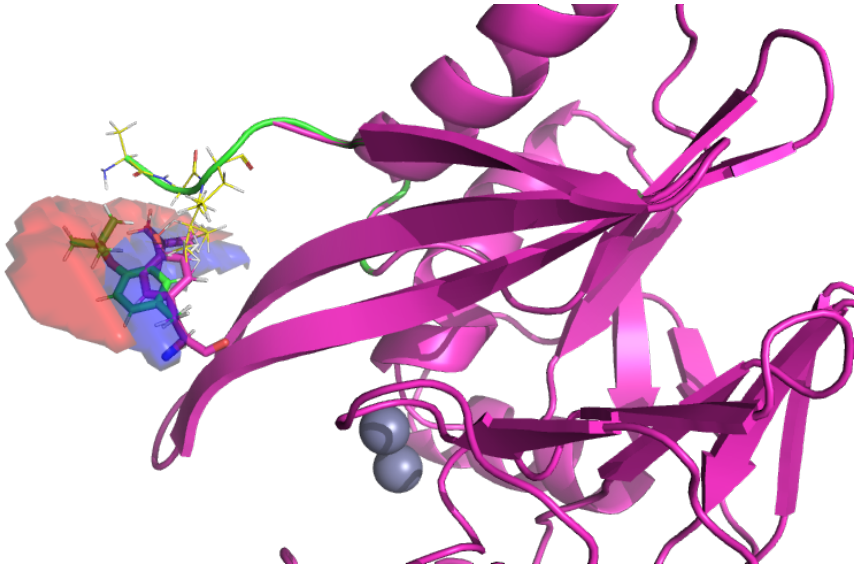


Figure 2.18 Overlay of the two models generated for IMP-1 V30^{13}C-Tby in the presence of the inhibitor captopril. The backbone conformation corresponds to the crystal structure 4C1F. The conformations of the Tby residues delivering the smallest RMSD between back-calculated and experimental PCSs of the *tert*-butyl group are shown in stick representation. The blue and red clouds depict the localization spaces for the *tert*-butyl group derived from the two sets of experimental PCSs.

The red localization space indicates that the *tert*-butyl group has moved further away from Leu4. This shift may reflect a movement of loop 1 from an open conformation to a closed conformation upon binding to the ligand. It is somewhat unexpected, that the exchange between open and closed conformation would be sufficiently slow to observe two separate cross-peaks in the ^{13}C-HSQC spectrum. At this stage it cannot be excluded that the mutation of valine to *tert*-butyltyrosine affects the dynamics and the structure of the loop in the presence of captopril.

2.3.3 Captopril binding to IMP-1

To assess the binding of captopril to IMP-1, the solvent-exposed residue Val30 in loop 1 was substituted by Tby and the *tert*-butyl signal was monitored by measuring 1D ^{1}H-NMR spectra (Figure 2.19) before and after adding the inhibitor in 1:0.5 and 1:1 protein-to-inhibitor ratios.

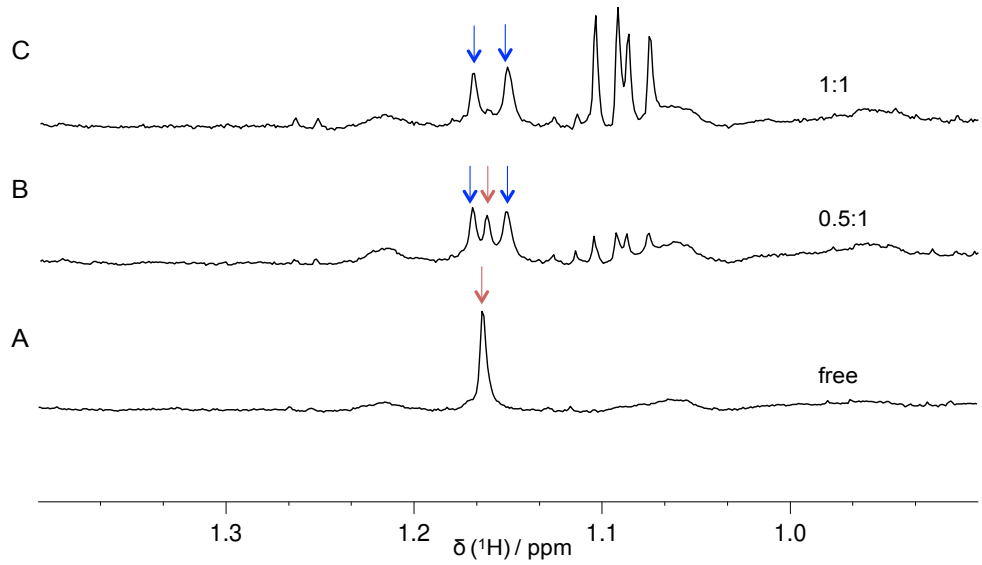


Figure 2.19 Selected spectral region from 1D ^{1}H-NMR spectra recorded of the 114 μM Val30Tby mutant of IMP-1 in the presence or absence of captopril. The spectra were recorded at 298 K in 20 mM MES, pH 6.5, 100 mM NaCl. The peaks of the *tert*-butyl signal are marked by arrows. (A) Free IMP-1 Val30Tby. (B) Same as (A), but in the presence of captopril in 0.5:1 ratio. (C) Same as (A), but in the presence of captopril in 1:1 ratio.

The free protein gave a single *tert*-butyl signal in the 1D ^{1}H-NMR spectrum. Following addition of the inhibitor in 0.5:1 ratio, three peaks were observed (Figure 2.19B), where one peak corresponds to the free form (red arrow) and two peaks correspond to the inhibitor-bound form as observed when the inhibitor was present in 1:1 ratio (Figure 2.19C). As distinct *tert*-butyl signals were observed for the free and bound states, it is clear that the exchange between bound and free IMP-1 is slow on the NMR time scale.

To determine whether the Tby substitution at position 30 affects the affinity of captopril to IMP-1, ITC measurements were carried out to measure the dissociation constants of captopril to the wild-type protein and the Val30Tby mutant. The ITC results are shown in Figure 2.20. The dissociation constants derived from these data using a 1:1 binding model were 3.4 and 7 μM for the wild-type protein and Val30Tby mutant, respectively, indicating little perturbation by the mutation. As the NMR data were all recorded at much higher concentrations than the K_d values (e.g. the ^{13}C-HSQC spectra of Figure 2.13 used protein concentrations of at least 80 μM), the appearance of

two signals in the presence of captopril indicates the co-existence of two different conformational species of the complex rather than IMP-1 with and without inhibitor.

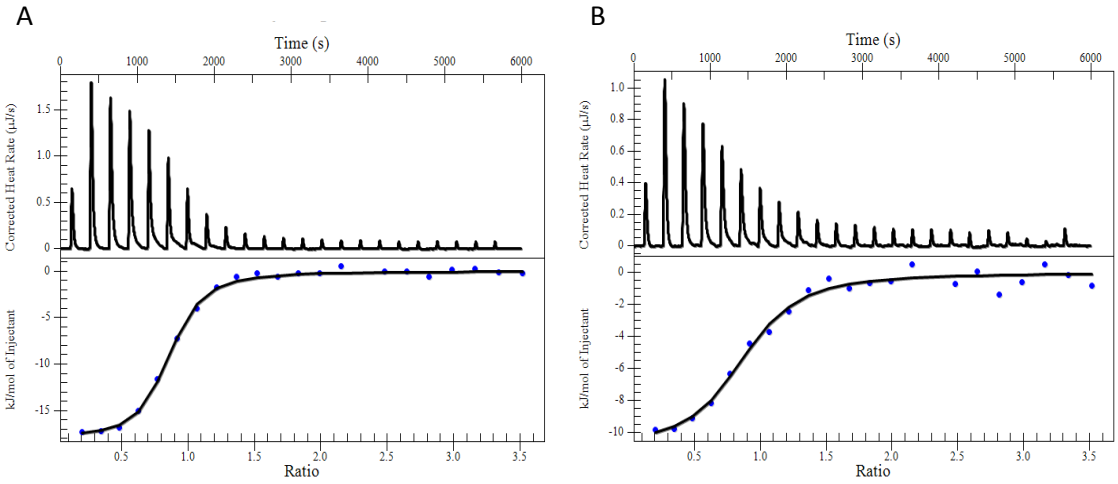


Figure 2.20 ITC data of captopril binding to wild-type IMP-1 and its Val30Tby mutant. Thermograms are shown at the top of each panel and binding isotherms at the bottom. All measurements were performed at 25 °C in buffer containing 20 mM MES pH 6.5 and 100 mM NaCl. The solid black lines of the isotherms represent the best fits assuming 1:1 binding. (A) ITC results for wild-type IMP-1. (B) ITC results for IMP-1 Val30Tby.

2.3.4 Study of the flexibility of loop 2 by NMR spectroscopy

The active site pocket of IMP-1 is surrounded by several loop regions. The above experiments demonstrated that loop 1 is involved in substrate binding as has previously been indicated by published crystal structures.

Although the conformation of loop 2 could plausibly also respond to substrate binding, the IMP-1 crystal structures in the native and inhibitor bound forms show no evidence for this (Figure 2.21). Carruthers et al. (2014), however, were unable to assign the resonances from this region by solution NMR, which suggests that this loop may be dynamic on an intermediate timescale, where its NMR resonances are broadened beyond detection.

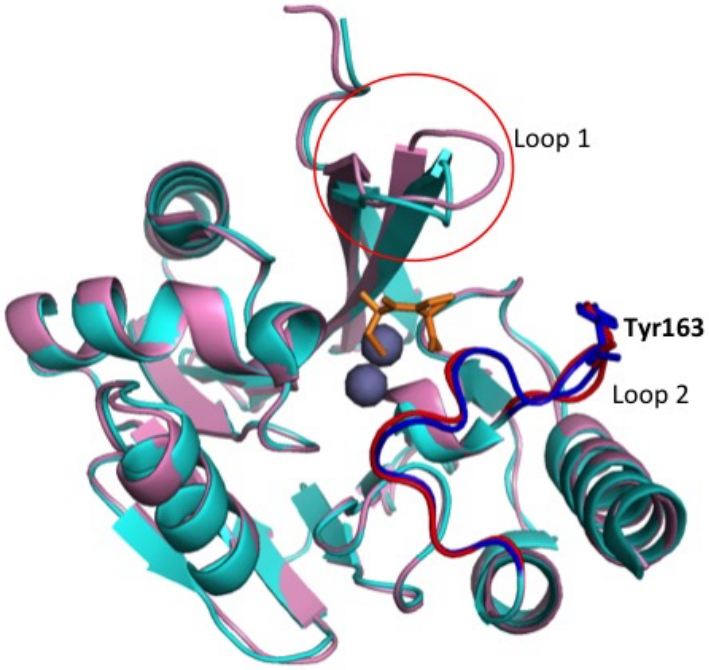


Figure 2.21 Overlay of crystal structures of IMP-1 determined in the presence (cyan with the loop 2 highlighted in blue; PDB ID: 4C1F; Brem et al., 2015) and absence of captopril (pink with the loop 2 highlighted in red; PDB ID: 1DDK; Concha et al., 2000). Zn atoms are shown as spheres and the captopril molecule is shown as orange sticks.

To explore the flexibility of loop 2, the solvent-exposed residue Tyr163 in the loop 2 region was substituted by Tby and the *tert*-butyl signal was monitored by 1D ^{1}H-NMR spectroscopy in the free and inhibitor bound forms of the protein (Figure 2.22).

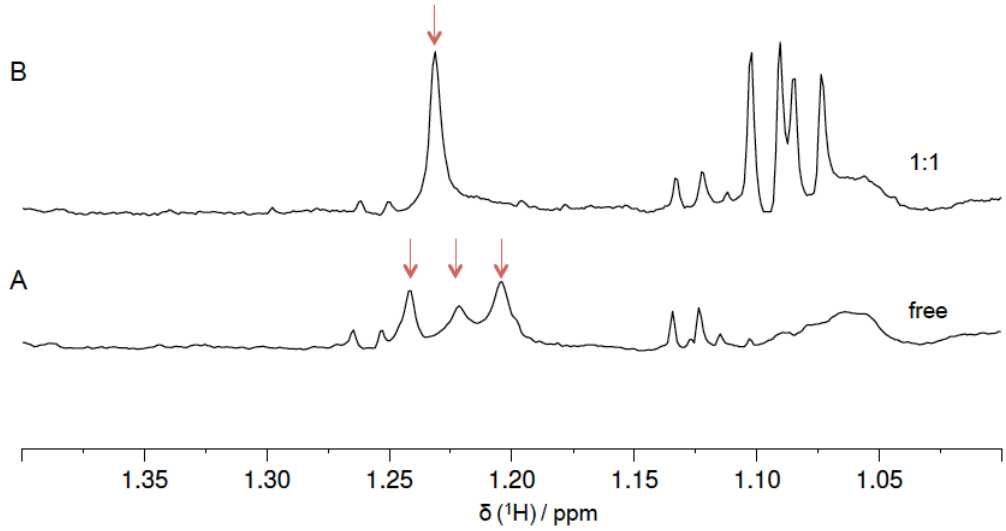


Figure 2.22 Selected spectral region from 1D ^{1}H-NMR spectra recorded of the 290 μM Tyr163Tby mutant of IMP-1 in the presence or absence of captopril. The spectra were recorded at 298 K in 20 mM MES, pH 6.5, 100 mM NaCl. The peaks of the *tert*-

butyl signal are marked by arrows. (A) Free IMP-1 Tyr163Tby. (B) Same as (A) but with captopril added in 1:1 ratio.

Unexpectedly, three *tert*-butyl peaks were observed for the Tby residue at position 163, indicating conformational heterogeneity. In contrast, only a single *tert*-butyl signal was observed in the presence of the inhibitor, indicating that the bound inhibitor rigidifies the loop. Furthermore, the *tert*-butyl peaks were uncharacteristically broad in the free protein, suggesting chemical exchange between the different conformational species. To confirm the presence of chemical exchange between the three peaks, a [^{1}H,^{1}H]-NOESY spectrum was measured (Figure 2.23).

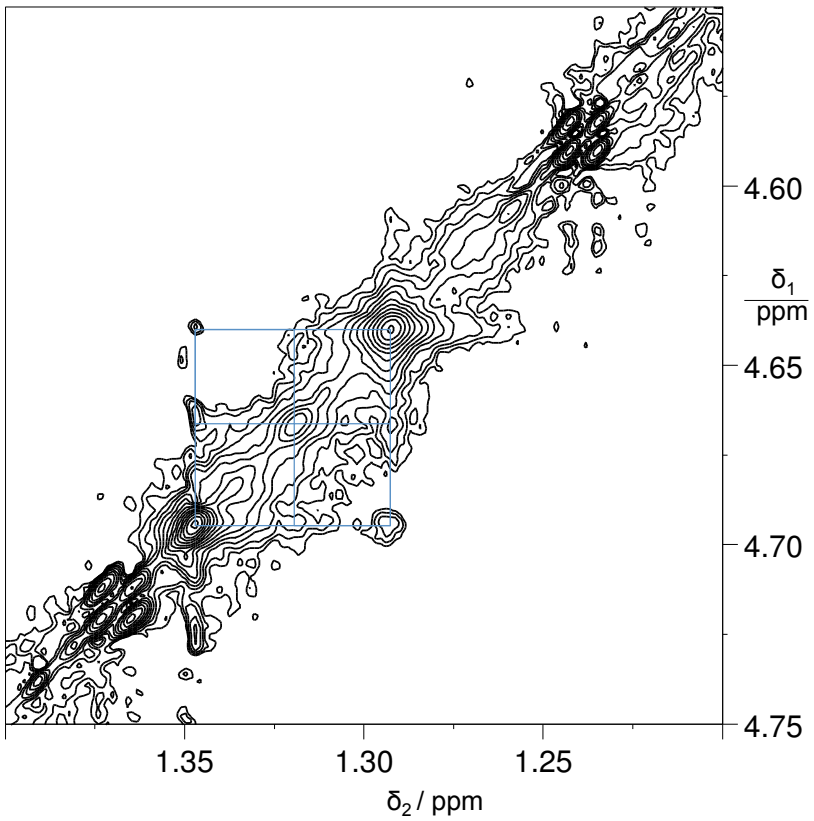


Figure 2.23 Selected spectral region from a [^{1}H,^{1}H]-NOESY spectrum recorded of IMP-1 Tyr163Tby with a mixing time of 400 ms, showing the diagonal and cross-peaks of the three *tert*-butyl peaks. The spectrum was measured at 310 K in 20 mM MES, pH 6.5, 100 mM NaCl. The cross-peaks are connected by lines.

The observation of cross-peaks between the three *tert*-butyl signals confirms the chemical exchange between the three species, marking them as different in conformation rather than chemistry. Therefore, loop 2 of IMP-1 is more flexible than indicated by the crystal structures.

2.4 Conclusion

Studying the structure of the active-site loops 1 and 2 in IMP-1 is difficult because no resonance assignments have been determined for these polypeptide segments in the wild-type protein. IMP-1 is an enzyme, where conformational exchange prevents the observation and assignment of many of the backbone amides. In this situation, the introduction of ^{13}C-Tby as an NMR tag into the active-site loop 1 allowed the measurement of PCS values of the *tert*-butyl group. The loop 1 presents an open conformation in its free form and shows two different conformations, open and closed, in the presence of an inhibitor. While significant differences in PCSs could be observed in the presence or absence of inhibitor, data with a different NMR tag are required to confirm that the PCSs report on the conformation of the loop 1 rather than the tag. Conversely, the loop 2 displays a higher degree of conformational flexibility in the absence of the inhibitor, but adopts a single conformation in the presence of the inhibitor which suggests that the loop becomes more rigid upon inhibitor binding.

Even if structural details of the loop conformations could not be determined in the present work, the NMR measurements report with high sensitivity on the flexibility of the loop regions, which, in a screen for improved inhibitors of IMP-1, can be used to detect not only the occurrance of inhibitor binding but also the conformational consequences in the protein.

2.5 Future work

In this study, Tby was used as an NMR probe to measure PCSs in loop 1 of IMP-1. Close contacts between the side chain of the Tby residue and the side chain of Leu4 are predicted for the protein structure with the loop in the open conformation. It would be informative to carry out additional PCS measurements by incorporating a different unnatural amino acid as NMR probe such as trifluoromethyl phenylalanine (Tmf). Compared to Tby, Tmf has a less bulky side chain which would allow fewer contacts with the side chain of Leu4. PCSs of the trifluoromethyl group could be measured by 1D ^{19}F-NMR spectroscopy. It would be particularly interesting, whether slow chemical exchange between two conformations still prevails for Tmf in the presence of captopril, as this would affirm whether the interaction between the side chains of Tby and Leu4 influences the conformational exchange of loop 1.

2.6 References

Abraham, E.P. and Chain, E. (1940) An enzyme from bacteria able to destroy penicillin. *Nature* **146**, 837.

Ambler, R.P. (1980) The structure of β-lactamases. *Philos. T. R. Soc. Lon. B.* **289**, 321-331.

Arakawa, Y., Murakami, M., Suzuki, K., Ito, H., Wacharotayankun, R., Ohsuka, S., Kato, N. and Ohta, M. (1995) A novel integron-like element carrying the metallo-β-lactamase gene *bla*$_{IMP}$. *Antimicrob. Agents Chemother.* **39**, 1612–1615.

Bebrone, C. (2007) Metallo-β-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem. Pharmacol.* **74**, 1686-1701.

Brem, J., van Berkel, S.S., Zollman, D., Lee, S.Y., Gileadi, O., McHugh, P.J., Walsh, T.R., McDonough, M.A. and Schofield, C.J. (2015) Structural basis of metallo-β-lactamase inhibition by captopril stereoisomers. *Antimicrob. Agents Chemother.* **60**, 142-150.

Bush, K., Jacoby, G.A. and Medeiros A.A. (1995) A functional classification scheme for β-lactamases and its correlation with molecular structure. *Antimicrob. Agents chemother.* **39**, 1211-1233.

Carfi, A., Pares, S., Duee, E., Galleni, M., Duez, C., Frere, J-.M. and Dideberg, O. (1995) The 3-D structure of a zinc metallo-*β*-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J.* **14**, 4914-4921.

Carruthers, T.J., Carr, P.D., Loh C.-T., Jackson, C.J. and Otting, G. (2014) Fe^{3+} located in the dinuclear metallo-β-lactamase IMP-1 by pseudocontact shifts. *Angew. Chem. Int. Ed.* **53**, 14269-14272.

Concha, N.O., Rasmussen, B.A., Bush, K. and Herzberg, O. (1996) Crystal structure of the wide-spectrum binuclear zinc β-lactamase from *Bacteroides fragilis. Structure* **4**, 823-836.

Concha, N.O., Janson, C.A., Rowling, P., Pearson, S., Cheever, C.A., Clarke, B.P., Lewis, C., Galleni, M., Frere, J-.M., Payne, D.J., Bateson, J.H. and Abdel-Meguid, S.S. (2000) Crystal Structure of the IMP-1 metallo-β-lactamase from *Pseudomonas aeruginosa* and its complex with a mercaptocarboxylate inhibitor: binding determinants of a potent, broad-spectrum inhibitor. *Biochemistry* **39**, 4288-4298.

Dal Peraro, M., Vila, A.J. and Carloni, P. (2004) Substrate binding to mononuclear metallo-β-lactamase from *Bacillus cereus*. *Proteins* **54**, 412-423.

Daiyasu, H., Osaka, K., Ishino, Y. and Toh, H. (2001) Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Lett.* **503**, 1-6.

Davies, J. and Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417-433.

de la Cruz, L., Nguyen, T.H.D., Ozawa, K., Shin, J., Graham, B., Huber, T. and Otting, G. (2011) Binding of low-molecular weight inhibitors promotes large conformational changes in the dengue virus NS2B-NS3 protease: fold analysis by pseudocontact shifts. *J. Am. Chem. Soc.* **133**, 19205-19215.

Essack, S.Y. (2001) The development of β-lactam antibiotics in response to the evolution of β-lactamases. *Pharm. Res.* **18**, 1391-1399.

Galleni, M., Lamotte-Brasseur, J., Rossolini, G.M., Spencer, J., Dideberg, O., Frere, J-.M. and the metallo-β-lactamase working group (2001) Standard numbering scheme for class B β-lactamase. *Antimicrob. Agents Chemother*. **45**, 660-663.

Garau, G., Di Guilmi, A.M. and Hall, B.G. (2005a) Structure-based phylogeny of the metallo-β-lactamases. *Antimicrob. Agents Chemother*. **49**, 2778-2784.

Garau, G., Bebrone, C., Anne, C., Galleni, M., Frere, J-.M. and Dideberg, O. (2005b) A metallo-beta-lactamase enzyme in action: crystal structures of the monozinc carbapenemase CphA and its complex with biapenem. *J. Mol. Biol.* **345**, 785-795.

Garcia-Saez, I., Mercuri, P.S., Papamicael, C., Kahn, R., Frere, J-.M., Galleni, M., Rossolini, G.M. and Dideberg, O. (2003) Three-dimensional structure of FEZ-1, a monomeric subclass B3 metallo-β-lactamase from *Fluoribacter gormanii*, in native form and in complex with D-captopril. *J. Mol, Biol.* **325**, 651-660.

Garcia-Saez, I., Docquier, J.-D., Rossolini, G.M. and Dideberg, O. (2008) The three-dimensional structure of VIM-2, a Zn-beta-lactamase from *Pseudomonas aeruginosa* in its reduced and oxidised form. *J.Mol. Biol.* **375**, 604-611.

Hall, B.G., Salipante, S.J. and Barlow, M. (2003) The metallo-β-lactamases fall into two distinct phylogenetic groups. *J. Mol. Evol*. **57**, 249-254.

Haruta, S., Yamaguchi, H., Yamamoto, E.T., Eriguchi, Y., Nukaga, M., O'Hara, K. and Sawai, T. (2000) Functional analysis of the active site of a metallo-β-lactamase proliferating in Japan. *Antimicrob. Agents Chemother.* **44**, 2304–2309.

Hinchliffe, P., González, M.M., Mojica, M.F., González, J.M., Castillo, V., Saiz, C., Kosmopoulou, M., Tooke, C.L., Llarrull, L.I., Mahler, G., Bonomo, R.A., Vila, A.J. and Spencer, J. (2016) Cross-class metallo-β-lactamase inhibition by bisthiazolidines reveals multiple binding modes. *Proc. Natl. Acad. Sci. USA* **113**, 3745-3754.

Horton, L.B., Shanker, S., Mikulski, R., Brown, N.G., Phillips, K., Lykissa, E., Prasad, B.V.V. and Palzkill, T. (2012) Mutagenesis of zinc ligand residue Cys221 reveals plasticity in the IMP-1 metallo-β-lactamase active site. *Antimicrob. Agents Chemother.* **56**, 5667-5677.

Laraki, N., Franceschini, N., Rossolini, G.M., Santucci, P., Meunier, C., de Pauw, E., Amicosante, G., Frere, J-.M. and Galleni, M. (1999) Biochemical characterization of the *Pseudomonas aeruginosa* 101/1477 metallo-β-lactamase IMP-1 produced by *Escherichia coli*. *Antimicrob. Agents Chemother.* **43**, 902–906.

Levy, S.B. and Marshall, B. (2004) Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med*. **10**, 122-129.

Livermore, D.M. (1995) β-Lactamases in laboratory and clinical resistance. *Clin. Microbiol. Rev.* **8**, 557-584.

Majiduddin, F.K., Materon, I.C. and Palzkill, T.G. (2002) Molecular analysis of β-lactamase structure and function. *Int. J. Med. Microbiol.* **292**, 127-137.

Moali, C., Anne, C., Lamotte-Brasseur, J., Groslambert, S., Devreese, B., Beeumen, J.V., Galleni, M. and Frere, J-.M. (2003) Analysis of the importance of the metallo-β-lactamase active site loop in substrate binding and catalysis. *Chem. Biol*. **10**, 319-329.

Queenan, A.M. and Bush, K. (2007) Carbapenemases: the versatile β-lactamases. *Clin. Microbiol. Rev.* **20**, 440-458.

Rasmussen, B.A. and Bush, K. (1997) Carbapenem-hydrolyzing β-lactamases. *Antimicrob. Agents Chemother*. **41**, 223-232.

Richmond, M.H. and Sykes, R.B. (1973) The β-lactamases of Gram-negative bacteria and their possible physiological role. *Adv. Microb. Physiol.* **9**, 31-88.

Schmitz, C., Stanton-Cook, M., Su, X.-C., Otting, G. and Huber, T. (2008) Numbat: an interactive software tool for fitting Δχ-tensors to molecular coordinates using pseudocontact shifts. *J. Biomol. NMR* **41**, 179-189.

Siemann, S., Brewer, D., Clarke, A.J., Dmitrienko, G.I., Lajoie, G. and Viswanatha, T. (2002) IMP-1 metallo-β-lactamase: effect of chelators and assessment of metal requirement by electrospray mass spectrometry. *Biochim. Biophys. Acta* **1571**, 190–200.

Siu, L.K. (2002) Antibiotics: action and resistance in Gram-negative bacteria. *J. Microbiol. Immunol. Infect*. **35**, 1-11.

Sykes, R. (2010) The 2009 Garrod lecture: the evolution of antimicrobial resistance: a Darwinian perspective. *J. Antimicrob. Chemother.* **65**, 1842-1852.

Toney, J.H., Hammond, G.G., Fitzgerald, P.M.D., Sharma, N., Balkovec, J.M., Rouen, G.P., Olson, S.H., Hammond, M.L., Greenlee, M.L. and Gao, Y.D.J. (2001) Succinic acids as potent inhibitors of plasmid-borne IMP-1 metallo-*β*-lactamase. *Biol. Chem.* **276**, 31913-31918.

Ullah, J.H., Walsh, T.R., Taylor, I.A., Emery, D.C., Verma, C.S., Gamblin, S.J. and Spencer, J. (1998) The crystal structure of the L1 metallo-β-lactamase from *Stenotrophomonas maltophilia* at 1.7 A resolution. *J. Mol. Biol.* **284**, 125-136.

Walsh, T.R., Toleman, M.A., Poirel, L. and Nordmann, P. (2005) Metallo-β-lactamases: the quiet before the storm? *Clin. Microbiol. Rev.* **18**, 306-325.

Wang, Z., Fast, W. and Benkovic, S.J. (1999) On the mechanism of the metallo-β-lactamase from *Bacteroides fragilis*. *Biochemistry* **38**, 10013-10023.

Watanabe, M., Iyobe, S., Inoue, M. and Mitsuhashi, S. (1991) Transferable imipenem resistance in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **35**, 147-151.

Young, D.D., Young, T.S., Jahnz, M., Ahmad, I., Spraggon, G. and Schultz, P.G. (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894-1900.

Zhao, W-.H. and Hu, Z-.Q. (2011) IMP-type metallo-β-lactamase in Gram-negative bacilli: distribution, phylogeny, and association with integrons. *Crit. Rev. Microbiol.* **37**, 214-226.

CHAPTER 3

Study of oligonucleotide binding by the *E. coli* single-stranded DNA binding protein using NMR spectroscopy

3.1 Introduction

3.1.1 Single-stranded DNA binding proteins

DNA is the main storage mode of the hereditary information of life. In order for cells to maintain and replicate their genome, double-helical DNA needs to be separated into its two complementary single-stranded DNA (ssDNA) strands. As ssDNA is prone to nucleolytic and chemical attacks, ssDNA-binding proteins (SSB) evolved to bind and protect ssDNA, when it emerges as an intermediate during DNA processing. SSBs not only protect ssDNA from degradation, but also attract SSB interacting proteins (SIP), which are involved in DNA replication, recombination and repair (Chase and Williams, 1986; Lohman and Ferrari, 1994; Meyer and Laine, 1990). Since SSBs must cover all ssDNA as it becomes accessible, they are highly abundant in all kingdoms of life and they bind to ssDNA with high affinity and very little, if any, sequence specificity (Meyer and Laine, 1990).

The T4 gene 32 protein from the T4 bacteriophage was the first SSB to be discovered and characterized (Alberts and Frey, 1970), followed by *E. coli* SSB (Sigal et al., 1972). Since then, many prokaryotic and eukaryotic SSBs have been identified in bacteria (de Vries and Wackernage, 1993), plasmids (Golub and Low, 1985), humans (Fairman and Stillman 1988; Wobbe et al., 1987), yeast (Brill and Stillman, 1989) and many other organisms, illustrating their functional importance.

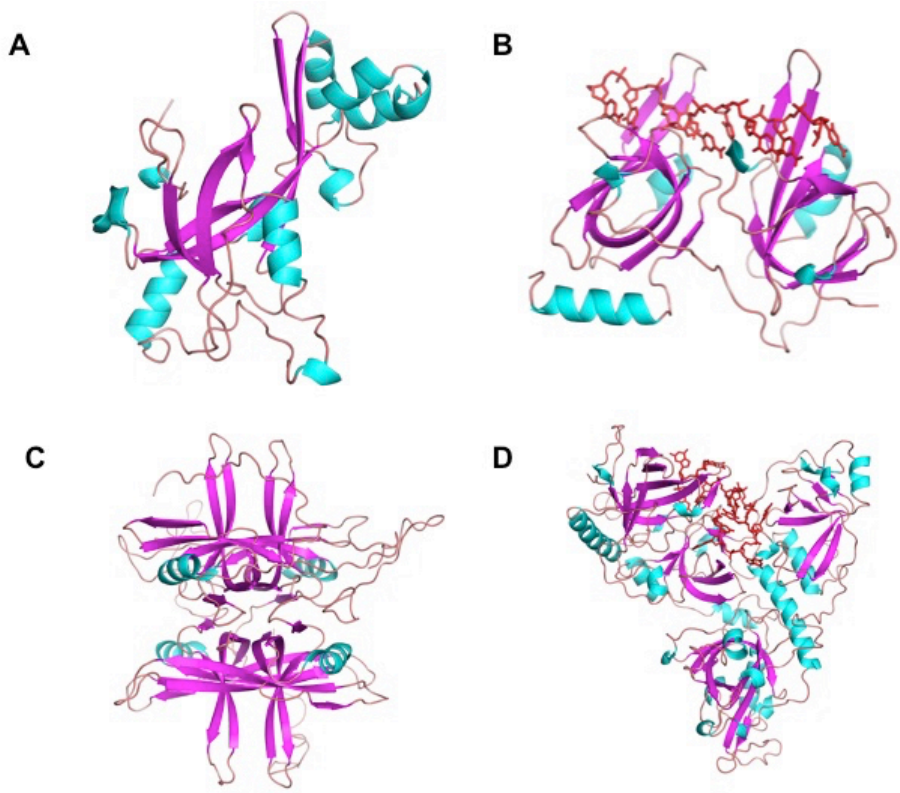


Figure 3.1 Ribbon diagrams of the 3D structures of some ssDNA binding proteins. ssDNA segments, when present, are displayed in stick representations (orange). (A) T4 SSB (PDB ID: 1GPC; Shamoo et al., 1995); (B) human replication protein A (PDB ID: 1L1O; Bochkareva et al., 2002); (C) *E. coli* SSB (PDB ID: 1QVC; Matsumoto et al., 2000); (D) telomere end binding protein (PDB ID: 1OTC; Horvath et al., 1998).

3.1.2 *E. coli* SSB and its binding modes

To date, *E. coli* SSB is the best-characterised prokaryotic SSB protein. It functions as a homotetramer. Each subunit of *E. coli* SSB consists of two domains. The N-terminal domain (residues 1-112) is composed of an oligonucleotide/oligosaccharide binding fold (OB fold), which binds to ssDNA (Lohman and Ferrari, 1994). Chemical modification studies and fluorescence quenching experiments suggested that tryptophan (Trp) residues are important for ssDNA binding (Overman et al., 1988; Curth et al., 1993). Trp-54 and Trp-88 were shown to play a major role in ssDNA binding as nearly complete fluorescence quenching was observed for these residues (Curth et al, 1993). Furthermore, Trp-40 and Trp-54 were shown to form stacking interactions with ssDNA bases (Khamis et al., 1987). In addition, crosslinking experiments (Merrill et al., 1984)

and mutational studies (Casas-Finet, 1987; Bayer, 1989) indicated that Phe-60 is involved in ssDNA binding while also affecting the stability of the tetramer.

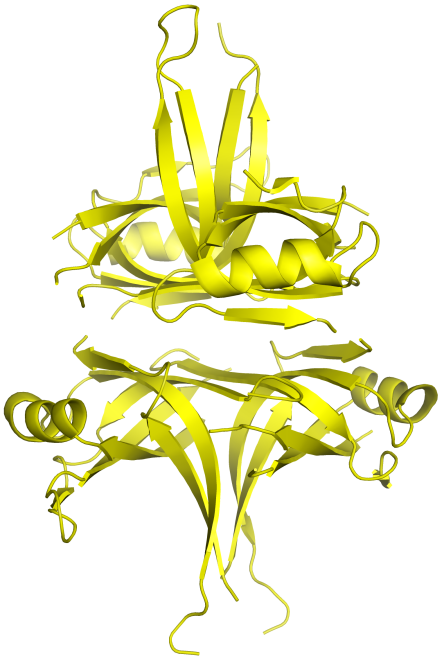


Figure 3.2 X-ray crystal structure of the *E. coli* SSB tetramer (PDB ID: 1SRU; Savvides et al., 2004).

The C-terminal domain of SSB (residues 113-177) consists of a long intrinsically disordered and flexible polypeptide chain with eight predominantly negatively charged residues (DFDDDIPF) at the C-terminal tip (Shereda et al., 2008). These eight residues are highly conserved among bacterial SSBs (Lu and Keck, 2008). They bind to the OB domain (Shishmarev et al., 2014) and, by binding to SIPs, they assist with recruiting SIPs to ssDNA (Shereda et al., 2008). The NMR experiments showed that the C-terminal tip is in fast exchange between unbound and bound states to the OB domain. Furthermore, the C-terminal tip is mostly in the unbound state even in the absence of ssDNA (Su et al., 2014).

Due to the presence of four OB domains in the tetramer, SSB can bind long ssDNA strands in multiple binding modes, which are characterised by the length of occluded ssDNA segments. This has been observed by electron microscopy, where Griffith et al. (1984) observed different ssDNA binding modes when varying the protein to DNA ratio, as well as in solution, where Lohman and Overman (1985) reported

different ssDNA binding modes at different salt conditions, which resulted in different quenching behaviour of the intrinsic SSB fluorescence.

Among the different binding modes reported, the $(SSB)_{35}$ and $(SSB)_{65}$ modes are the two major binding modes observed *in vitro* (Lohman and Overman, 1985). The relative abundance of these two binding modes depends on the solution conditions. The valance of any salt and its concentration, as well as the ratio of protein to DNA are the two main factors, which determine the relative abundance and stabilities of these two binding modes. In addition, there are effects from the pH and temperature (Lohman and Overman, 1985; Bujalowski and Lohman, 1986; Bujalowski et al., 1988; Wei et al., 1992).

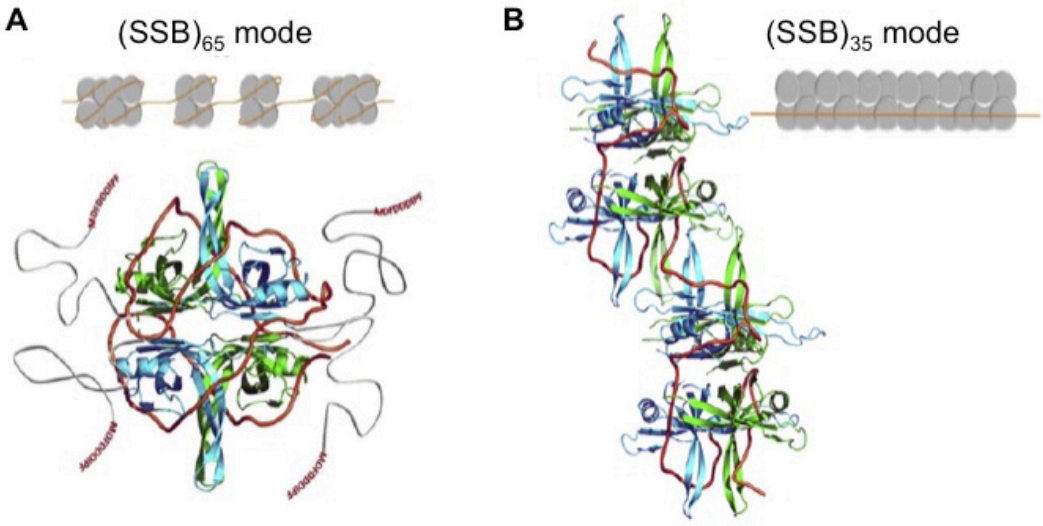


Figure 3.3 *E. coli* SSB binding modes. (A) Schematic representation of the SSB-ssDNA interaction in the $(SSB)_{65}$ binding mode, with 65 nucleotides of DNA (orange ribbon) wrapped around each SSB tetramer. (B) Schematic drawing of a hypothetical model of the SSB-ssDNA complex in the $(SSB)_{35}$ binding mode, where pairs of SSB tetramers interact with ssDNA segments of about 70 nucleotides (orange line) using an average of only two subunits of each tetramer. Reproduced from Kozlov et al. (2017).

The $(SSB)_{35}$ mode is favoured at low monovalent salt concentrations ([NaCl] < 10 mM) and high ratios of protein to DNA. In this mode ssDNA wraps around, on average, only two of the four subunits in the SSB tetramer. For long segments of ssDNA, SSB can bind with very high cooperativity, forming long protein clusters (Figure 3.3B; Griffith et al., 1984; Lohman et al., 1986; Ferrari et al., 1994). On the

other hand, the $(SSB)_{65}$ mode is favoured at moderate to high monovalent salt concentrations ([NaCl] > 200 mM). In this mode, ssDNA can occupy all four possible binding sites on SSB (Figure 3.3A). Unlike the $(SSB)_{35}$ mode, the $(SSB)_{65}$ mode displays limited cooperativity between tetramers, forming at most dimers of tetramers (octamers; Chrysogelos and Griffith, 1982; Griffith et al., 1984; Bujalowski and Lohman, 1987; Overman et al., 1988).

Since ssDNA binds to the four subunits of SSB with salt-dependent negative cooperativity, the relative stability of the two main binding modes can be modulated by the monovalent salt concentration in solution. The negative cooperativity effect is strong at low salt and decreases with increasing salt concentration. At low salt concentration, ssDNA can bind tightly to the first two subunits but binding to the other two subunits is much weaker (Lohman and Bujalowski, 1988; Bujalowski and Lohman, 1986). Salt dependence of the negative cooperativity effect is the main reason for the transition between the two major binding modes with changing salt concentrations.

A crystal structure of a chymotryptic fragment of the SSB tetramer bound to two molecules of dC_{35} has been published, where the chymotryptic digestion served to remove the unstructured C-terminal polypeptide segment of SSB (Figure 3.4; Raghunathan et al., 2000). Based on this structure, models have been proposed for the topologies of ssDNA wrapping around SSB in the $(SSB)_{65}$ and $(SSB)_{35}$ modes. In the model proposed for the $(SSB)_{65}$ mode, the ssDNA wraps around all four subunits of SSB with a topology similar to a baseball seam. In this model, the 5' and 3' ends of the ssDNA enter and exit the tetramer in close proximity (Raghunathan et al., 2000). The model proposed for the $(SSB)_{35}$ mode shows partial interactions with three monomer units, which on average equal interactions with two subunits (Figure 3.3B). In this model, the ssDNA enters and exits the tetramer on opposite sides (Raghunathan et al., 2000). Notably, the crystal structure did not detect electron density for all nucleotides, making these models to a large extent speculative.

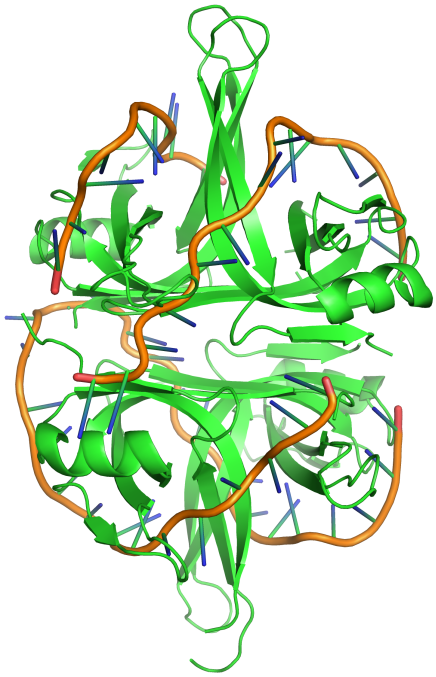


Figure 3.4 X-ray crystal structure of a chymotryptic fragment of *E. coli* SSB bound to two 35-mer single-stranded DNA molecules (PDB ID: 1EYG; Raghunathan et al., 2000).

3.1.3 ssDNA binding polarity on SSB

In addition to *E. coli* SSB, a number of SSB structures from other organisms have been determined in the presence of ssDNA (Chan et al., 2009; Yadev at al., 2012; Antony et al., 2012 George et al., 2012, Gamsjaeger et al., 2015). Many of them show a high degree of structural similarity to *E. coli* SSB and similar topology of ssDNA binding. One important difference between *E. coli* SSB and SSBs from other organisms, however, is the polarity with which ssDNA binds to the OB domains of the SSBs.

The nucleic acid binding polarity on OB folds may serve as an arbiter in solving the question whether OB folds all derive from a common ancestral origin or whether they arose independently during evolution. In a review, Theobald et al. (2003) proposed a “standard polarity” of DNA for wrapping around OB folds of various proteins, where the standard polarity is defined by ssDNA binding with the 5’-end directed towards strands β4 and β5 and the 3’-end directed in the approximate direction of the strand β2.

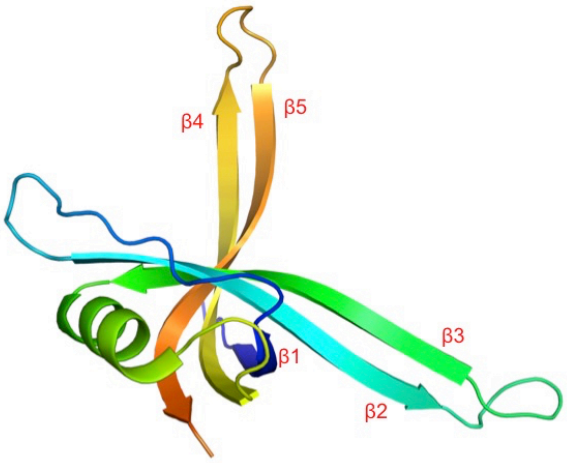


Figure 3.5 Example of a standard OB domain. This particular OB fold is from the *Helicobacter pylori* SSB (PDB ID: 2VW9; Chan et al., 2009).

Interestingly, the polarity of ssDNA binding reported for *E. coli* SSB is the opposite from that observed for the SSBs of all other organisms, as far as structures with bound ssDNA have been published (Theobald et al., 2003). This raises the question, whether the resolution of the crystal structure of *E. coli* SSB in complex with dC_{35} is sufficient to determine the polarity of the ssDNA. Unfortunately, the original diffraction data are unavailable, precluding a new refinement of the structure with different ssDNA models.

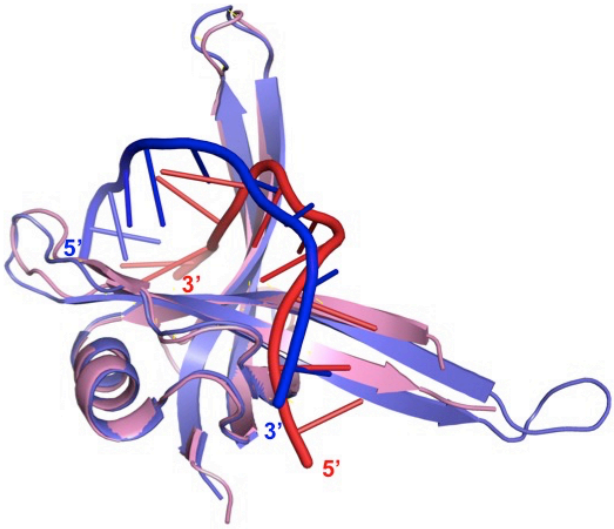


Figure 3.6 Overlay of the *H. pylori* SSB monomer (light blue with oligonucleotide strand in dark blue; PDB ID: 2VW9; Chan et al., 2009) and the *E. coli* SSB monomer (light pink with oligonucleotide strand in red; PDB ID: 1EYG; Raghunathan et al., 2000), highlighting the opposing polarities with which ssDNA binds to the OB domain. The 5'- and 3'-ends of the ssDNA strands are labelled.

3.14 Research objectives

In the present work, *E. coli* SSB-ssDNA interactions were studied in solution by NMR spectroscopy. In particular, we used the unnatural amino acid ^{13}C-*tert*-butyltyrosine (^{13}C-Tby) as an NMR probe. Changes in the *tert*-butyl signal observed upon addition of different oligonucleotides under varying solution conditions were used to investigate the SSB binding modes and their salt dependence, as well as the ssDNA binding polarity on SSB.

3.2 Materials and Methods

3.2.1 Expression and purification of *p*-cyanophenylalanyl-tRNA synthetase (*p*CNF-RS)

The pETMCSI vector (Neylon et al., 2000) was used to express *p*CNF-RS with an N-terminal His_6-tag followed by a tobacco etch virus (TEV) protease recognition site *in vivo*. The plasmid was transformed into *E. coli* BL21(DE3) cells, which were grown at 37 °C in 1 L LB medium containing 100 µg/mL ampicillin. At OD_{600} = 0.8 the cells were induced with 1 mM IPTG. After induction, the cells were incubated at 25 °C and harvested by centrifugation after overnight expression (~16 h). The cell pellet was resuspended in buffer A (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5% glycerol) and lysed by two passes through a French press at 12,000 psi and centrifuged at 30,000 *g* for 1 h at 4 °C.

The clarified lysate was loaded onto a 5 mL Co-NTA column (GE Healthcare, USA) equilibrated with buffer A and the protein was eluted with a linear gradient of 0-500 mM imidazole in buffer A. The fractions were analysed by 12% SDS-PAGE. The fractions containing the protein were pooled and the N-terminal His_6 tag was removed by incubation with His_6-tagged TEV protease at 4 °C for 16 h in a buffer containing 50 mM Tris-HCl, pH 8, 500 mM NaCl and 2 mM 2-mercaptoethanol. The mixture was passed again over the Co-NTA column to separate the cleaved His_6-tag and TEV protease from *p*CNF-RS. The purified *p*CNF-RS was concentrated and stored in buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol) at -70 °C.

3.2.2 Preparation of total tRNA with optimized suppressor tRNA

The pEvol plasmid encoding optimized suppressor-tRNA, $tRNA_{CUA}$ (Young et al., 2010), was transformed into BL21(DE3) cells. A mixture of total tRNA was produced and purified as described (Ozawa et al., 2012).

3.2.3 Cell-free protein expression

The gene for SSB Y97U, where U stands for an unnatural amino acid incorporated in response to a TAG codon, was cloned into the pETMCSI vector (Neylon et al., 2000) with a C-terminal His_6-tag for protein purification. Due to a cloning artefact, residues 132–141 from the flexible C-terminal domain were missing, but this modification is unlikely to perturb the DNA-binding mode. SSB Y97^{13}C-Tby, where ^{13}C-Tby stands for Tby with a uniformly ^{13}C-labelled *tert*-butyl group, was produced by cell-free protein synthesis. The reaction was conducted at 30 °C for 16 h in a dialysis system following a published protocol (Apponyi et al., 2008). To incorporate the unnatural amino acid ^{13}C-Tby, purified *p*CNF-RS and total tRNA mixture containing $tRNA_{CUA}$ were added to the inner cell-free reaction mixture and 1 mM of ^{13}C-Tby was added into the outer buffer mix.

3.2.4 Purification of SSB Y97^{13}C-Tby protein

The inner cell-free reaction mixture was centrifuged at 20,000 *g* in a bench-top centrifuge for 1 h at 4 °C. The supernatant was half diluted with buffer A (50 mM Tris-HCl, pH 7.5, 300 mM NaCl) and passed through a 1 mL Co-NTA gravity column (GE Healthcare, USA). Following a wash step with two column volumes of buffer A, the column was washed with two column volumes of 1 M $MgCl_2$ to remove bound ssDNA from the protein. Subsequently, the column was washed with buffer A containing 15 mM imidazole and the protein was eluted with buffer A with 300 mM imidazole. The purified protein was dialysed against NMR buffer (20 mM phosphate buffer, pH 7.2, 300 mM NaCl) and concentrated up to 40 μM (tetramer).

3.2.5 Oligonucleotides

All deoxyoligonucleotides used for NMR (Table 3.1) were purchased from Integrated DNA Technologies (IDT) in HPLC-purified form.

Table 3.1 Sequences of oligonucleotides used.

Name	Oligonucleotide sequence
$(dT)_{67}$	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
$(dT)_{35}$	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
ssDNA-3'	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGTGTT
ssDNA-5'	5'-TTGTGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
ssDNA-3'x4	5'-TTTTTTTTGGGTTTTTTTTTTTTTTGGGTTTTTTTTTTTTTTGGGTTTTTTTTTTTTTGGGTTTT
ssDNA-5'x4	5'-TTTTGGGTTTTTTTTTTTTTGGGTTTTTTTTTTTTTGGGTTTTTTTTTTTTTGGGTTTTTTTT
ssDNA-3'x4	5'-TTTTTTTTGGGTTTTTTTTTTTTTTGGGTTTTTTTTTTTTTTGGGTTTTTTTTTTTTTGGGTTTT
$(dT)_{35}$*	5'-TT*TTTTTTTTTTTTTTTTTTTTTTTTTTT*TTTTTT

- In $(dT)_{35}$*, stars denote phosphorothioate groups

3.2.6 NMR sample preparation

Different amounts of oligonucleotide stock solution were added to SSB Y97^{13}C-Tby in NMR buffer and 10% D_2O (v/v) was added to reach a final protein concentration of 10 μM (tetramer concentration) in a volume of 180 μL.

3.2.7 NMR spectroscopy

All NMR experiments were recorded on a Bruker 800 MHz NMR spectrometer equipped with a 5 mm TCI cryoprobe, using 3 mm NMR tubes at 25 °C. Parameters for the ^{13}C-HSQC spectra: t_{1max} = 21.2 ms, t_{2max} = 150.7 ms, total recording time per spectrum 20 min. $^1J_{CC}$ couplings were refocused by a selective 180°(^{13}C) pulse on the resonance of the quaternary carbon of the *tert*-butyl group in the middle of the t_1-evolution period.

3.2.8 EPR sample preparation

The oligonucleotide $(dT)_{35}$* containing a phosphorothioate group at sites 3 and 30 (Table 3.1) was purchased from Integrated DNA Technologies (USA) in HPLC-purified form. Five-fold molar excess of C10-Gd^{3+} tag (Wu et al., 2017) was added to a 1 mM solution of the modified oligonucleotide and the solution was diluted 3-fold by adding a 3:2 (v/v) mixture of TEAA (pH 7.5) and DMF buffer. The reaction mixture was incubated at room temperature overnight with shaking. Subsequently, the samples were washed thoroughly by exchanging to EPR buffer (20 mM phosphate buffer, pD 7.2, 300 mM NaCl in D_2O) using an Amicon centrifugal filter unit.

Two EPR samples were prepared by adding $(dT)_{35}$ oligonucleotide ligated with C10-Gd^{3+} tag to a wild-type SSB protein sample in 1:1 ratio (Sample 1) and 2:1 ratio (Sample 2). Perdeuterated glycerol was added to a final concentration of 20% (v/v) to reach a final protein concentration of 100 μM.

3.3 Results and discussion

3.3.1 SSB expression and purification

^{13}C-Tby was successfully incorporated into SSB by cell-free protein synthesis to produce SSB Y97^{13}C-Tby with C-terminal His_6-tag. The purification protocol developed by Dr Yao Wang (section 3.2.4) was used to obtain pure protein. The His_6-tag greatly simplified the purification, as the purification protocol for SSB without His_6-tag requires multiple columns to obtain SSB without bound ssDNA. Minimizing the loss of protein during purification was important as *in vitro* expression yields only relatively small quantities of protein and ^{13}C-Tby is expensive. In the following, 'SSB' refers to SSB Y97^{13}C-Tby with C-terminal His_6-tag.

3.3.2 Modelling

The crystal structure published by Raghunathan et al. (2000) is of a chymotryptic fragment of *E. coli* SSB bound to two 35-mer single-stranded poly-dC DNAs (PDB ID: 1EYG). This structure was used to model the positions of the Tby residues in the tetramer. The model was generated using the software PyMOL (Figure 3.7). Assuming that the presence of the *tert*-butyl groups does not change the backbone or side-chain conformations of SSB, the modelling indicates that the *tert*-butyl groups are within contact distance of ssDNA. Short distances between the *tert*-butyl groups and ssDNA were confirmed by intermolecular NOEs (see Chapter 4). Therefore, the NMR resonances of the *tert*-butyl groups are expected to be sensitive reporters of the binding of ssDNA.

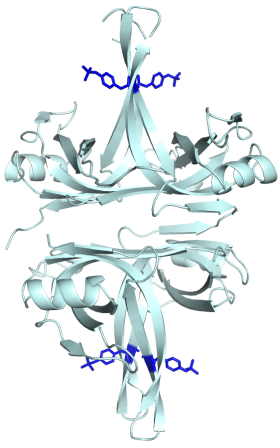


Figure 3.7 Overview of the SSB tetramer (cyan) in a ribbon representation with Tby (blue) modelled at position 97 using the program PyMOL (PDB ID: 1EYG).

3.3.3 NMR spectroscopy

All NMR experiments were performed with samples containing 10 μM SSB tetramer (40 μM monomer). Each ^{13}C-HSQC spectrum took only 20 minutes to record, illustrating the exceptional sensitivity that can be obtained by ^{13}C-Tby as an NMR probe.

As references, ^{13}C-HSQC spectra were measured of free SSB tetramer at high (300 mM) and low (10 mM) NaCl concentrations. A single cross-peak was observed for the *tert*-butyl group, showing that all four subunits of the tetramer experience on average the same chemical environment (Figure 3.8). In the low-salt conditions, the ^{1}H-chemical shift changed slightly up-field (Figure 3.8B).

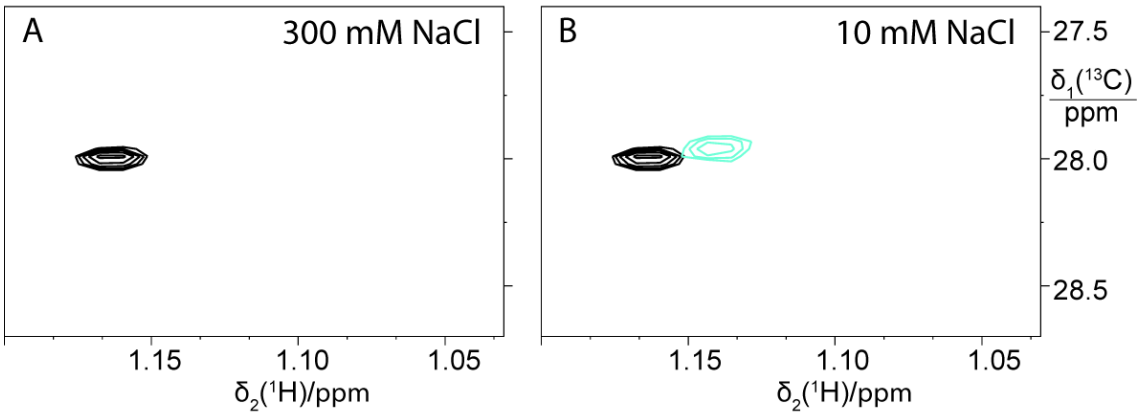


Figure 3.8 ^{13}C-HSQC spectra of free SSB Y97^{13}C-Tby. The spectral region shown displays the cross-peaks observed for the *tert*-butyl group. (A) Free SSB at high salt. (B) Superimposition of two spectra with free SSB at high salt (black) and low salt (cyan).

3.3.4 Salt dependence of binding of a single molecule of $(dT)_{67}$ to SSB

Initially, an attempt was made to detect the two main ssDNA binding modes reported in the literature (Lohman and Overman, 1985; Bujalowski and Lohman, 1986), $(SSB)_{65}$ (favoured at high salt) and $(SSB)_{35}$ (favoured at low salt), using the oligonucleotide $(dT)_{67}$.

At high NaCl concentration, addition of an equimolar amount of $(dT)_{67}$ shifted the peak of the *tert*-butyl signal from its position in the free SSB tetramer (Figure 3.9A). As in free SSB, only a single peak was observed in the presence of $(dT)_{67}$, which is in agreement with the $(SSB)_{65}$ mode, where the ssDNA interacts with all four possible binding sites on the SSB tetramer (Raghunathan et al., 2000). In addition, each *tert*-butyl group in the SSB tetramer interacts with ssDNA in a similar manner.

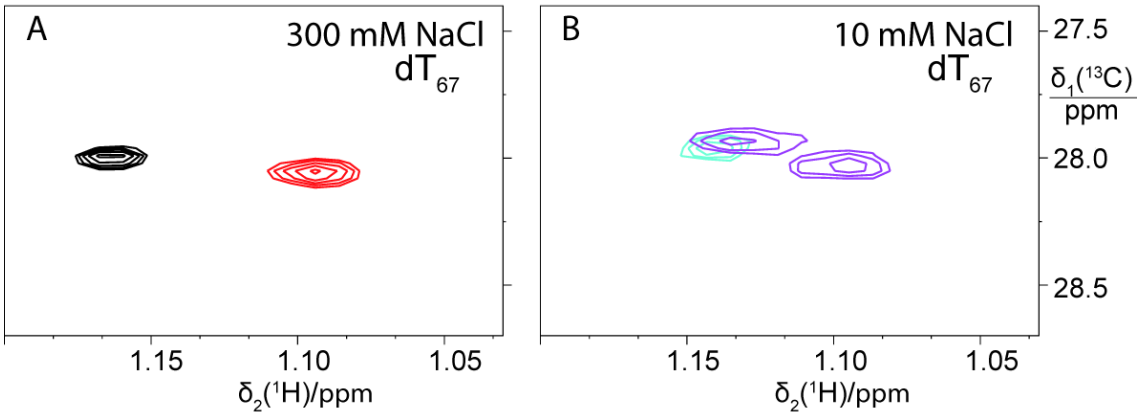


Figure 3.9 Superimposition of two ^{13}C-HSQC spectra of SSB Y97^{13}C-Tby. The spectral region shown displays the cross-peaks observed for the *tert*-butyl group. (A) SSB with 10 µM $(dT)_{67}$ (red) and free SSB (black) at high salt. (B) SSB with 10 µM $(dT)_{67}$ (purple) and free SSB (cyan) at low salt.

When the SSB-$(dT)_{67}$ complex was measured at low NaCl concentration, two peaks were observed for the *tert*-butyl group with similar intensities, one of them at the same position as for the high salt condition and the other near the position observed for free SSB (Figure 3.9B). This result is in full agreement with the other main binding mode of SSB, $(SSB)_{35}$, where only two out of four possible binding sites of the SSB tetramer are occupied with ssDNA. This binding mode is well-established for low-salt conditions (Lohman and Overman, 1985; Bujalowski and Lohman, 1986; Lohman et al., 1986).

Both at high and low salt concentrations, formation of the SSB-$(dT)_{67}$ complex was accompanied by an increase in line widths of the *tert*-butyl signal, as expected for the increased molecular weight of the complex versus the free protein. Interestingly, the line width of the *tert*-butyl peak was larger at 10 mM than 300 mM NaCl concentration, regardless of the presence or absence of ssDNA. A possible explanation for this

observation could be that the low salt conditions favour inter-tetramer interactions between SSB molecules (Griffith et al, 1984; Lohman et al. 1986; Ferrari et al., 1994; Kozlov et al., 2015), in agreement with the much greater solubility of SSB in the presence of 300 mM NaCl.

3.3.5 Binding of $(dT)_{35}$ to SSB at high salt concentration

Experiments were also conducted with the shorter oligonucleotide $(dT)_{35}$. Previous studies indicated that a single molecule of $(dT)_{35}$ binds to two subunits of the SSB tetramer (Bujalowski and Lohman, 1986). Furthermore, SSB has been reported to bind ssDNA in the $(SSB)_{35}$ mode at high protein-to-DNA ratio (Chrysogelos and Griffith, 1982; Griffith et al., 1984; Lohman et al., 1986). Therefore, an equimolar ratio of $(dT)_{35}$ is expected to bind in the $(SSB)_{35}$ mode regardless of salt concentration.

The ^{13}C-HSQC spectrum of the SSB tetramer with a 1:1 ratio of $(dT)_{35}$ measured at high salt revealed two cross-peaks (3.11A), one corresponding to free SSB and the other near the cross-peak observed for the *tert*-butyl signal in the $(SSB)_{65}$ mode. Both peaks have similar intensities as expected for the $(SSB)_{35}$ mode. Notably, however, the chemical shift values are slightly different from those observed in the reference recorded with free SSB and in the 1:1 complex of SSB with $(dT)_{67}$. This may indicate some difference in the way, in which ssDNA contacts SSB in the $(SSB)_{35}$ and $(SSB)_{65}$ modes, as suggested in the literature. As depicted in Figure 3.10, the binding path of ssDNA in the $(SSB)_{35}$ mode has been proposed to connect two opposing subunits, i.e. a monomer in one half and another monomer in the other half of the tetramer, whereas in the $(SSB)_{65}$ mode, the first half of the ssDNA interacts with adjacent monomers in the same half of the tetramer (Raghunathan et al., 2000). Most likely, however, binding of ssDNA to one of the monomers causes slight changes in the environment of the *tert*-butyl group in the adjacent unoccupied monomer and, similarly, an unoccupied binding site influences an occupied binding site in the adjacent monomer, considering that the two Tby residues are close in space (see Figure 3.7). This makes it difficult to distinguish NMR signals of occupied monomers from those that merely neighbour an occupied monomer.

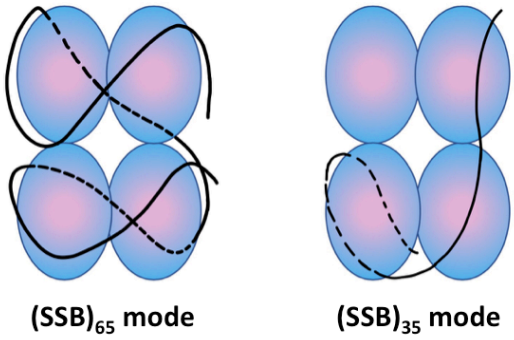


Figure 3.10 Models depicting the proposed ssDNA binding paths of the $(SSB)_{65}$ and $(SSB)_{35}$ binding modes.

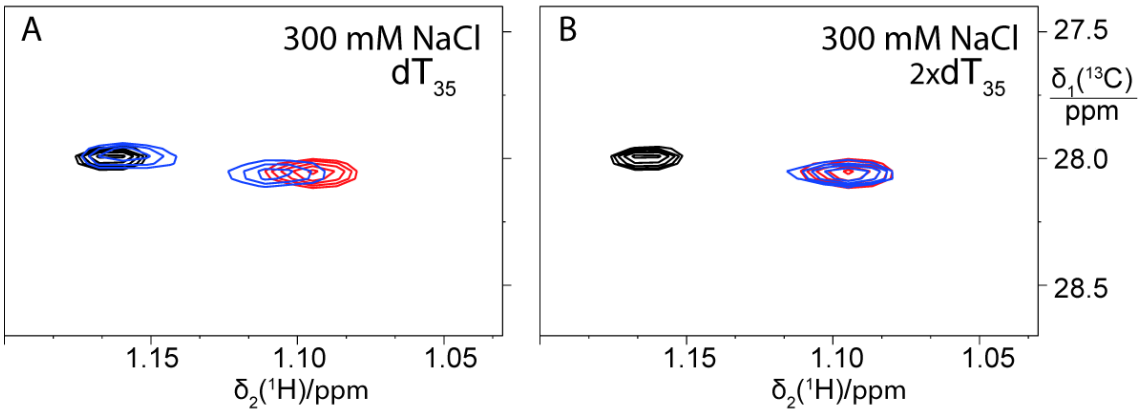


Figure 3.11 Superimposition of three ^{13}C-HSQC spectra of SSB Y97^{13}C-Tby at high salt. The spectral region shown displays the cross-peaks observed for the *tert*-butyl group in free SSB (black) and SSB with 10 µM $(dT)_{67}$ (red). The third spectrum (blue) is of SSB with (A) 10 µM $(dT)_{35}$ or (B) 20 µM $(dT)_{35}$.

The observation of two resolved cross-peaks in the spectrum of SSB with $(dT)_{35}$ in 1:1 ratio (Figure 3.11A) indicates that the rate of ssDNA exchanging between different monomers of SSB is slow on the NMR time scale, as fast exchange would lead to an average peak for bound and free SSB. With a peak separation of about 40 Hz, the residence time of $(dT)_{35}$ in its binding sites must be at least 50 ms.

In the presence of two molecules of $(dT)_{35}$ all four binding sites on the SSB tetramer are expected to be saturated with ssDNA. SSB-ssDNA complexes formed between SSB with two molecules of $(dT)_{35}$ or a single molecule of $(dT)_{70}$ have previously been reported to show similar ssDNA binding paths at high salt concentrations (Kozlov and Lohman, 1998; 2006), i.e. the length of ssDNA did not

change the binding mode. Indeed, a single *tert*-butyl cross-peak was observed in the presence of two molecules of $(dT)_{35}$ per SSB tetramer, and this cross-peak appeared at the same chemical shifts as for SSB with one molecule of $(dT)_{67}$ (Figure 3.11B). These results indicate that at high salt concentration a single $(dT)_{67}$ molecule and two molecules of $(dT)_{35}$ wrap around an SSB tetramer in the same way.

Viewing the tetramer in the orientation of Figure 3.7 there are, in principle, two different modes in which $(dT)_{35}$ could bind to two monomers in the SSB tetramer, by binding to (i) two monomers in the top half of the tetramer or (ii) one monomer in the top half and one in the bottom half. The deviation in chemical shifts of the cross-peaks obtained with a 1:1 ratio of protein to $(dT)_{35}$ from those of free SSB and the SSB-$(dT)_{67}$ complex (Figure 3.11A) suggests model ii, but interpretation i cannot be excluded. Therefore, EPR spectroscopy was used to study the conformation of a doubly-tagged $(dT)_{35}$ molecule bound to SSB at high salt. To tag $(dT)_{35}$ with two paramagnetic EPR tags, a modified $(dT)_{35}^{*}$ oligonucleotide was ordered, which contained phosphorothioate groups at positions 3 and 30 (Table 3.1) as sites for ligation with EPR tags. The C10-lanthanide tag has previously been used successfully to tag DNA containing a phosphorothioate group in order to generate PCSs in NMR spectra (Wu et al., 2017). The present work used the C10 tag loaded with a Gd^{3+} ion attached to each of the phosphorothioate groups in the modified $(dT)_{35}$ oligonucleotide in order to measure the distances between the two Gd^{3+} ions by a DEER experiment. A shorter distance was expected for $(dT)_{35}$ binding in mode i than in mode ii.

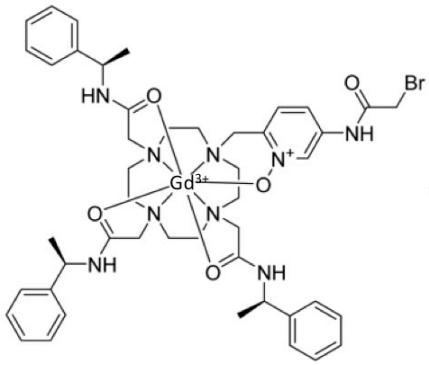


Figure 3.12 Chemical structure of the C10 tag.

Figure 3.13 shows the Gd^{3+}-Gd^{3+} distance obtained for protein to $(dT)_{35}$ ratios of 1:1 and 1:2. The distance of 5.6 nm obtained with the 1:1 ratio suggests that $(dT)_{35}$ binds to two SSB monomers in mode ii, whereas a distance of about 3 nm would have been expected in mode i. Indeed, a shorter distance of about 3.5 nm emerged in addition at the protein to $(dT)_{35}$ ratio of 1:2 (Figure 3.13).

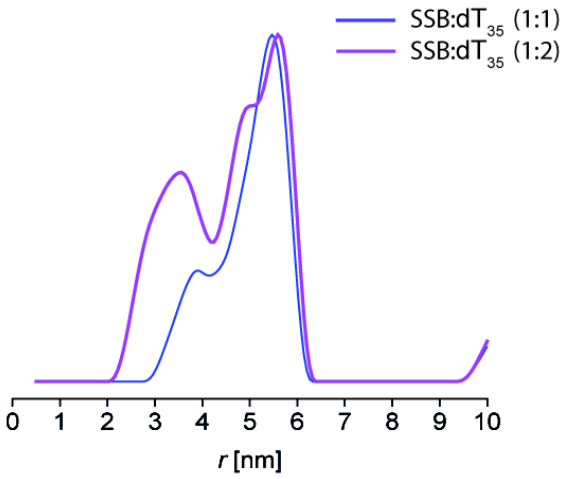


Figure 3.13 DEER distance distributions obtained for SSB-$(dT)_{35}$ complexes, where the $(dT)_{35}$ oligomer was labelled at two different sites with the C10-Gd^{3+} tag. Data recorded by the group of Prof. Daniella Goldfarb at the Weizmann Institute, using a W-band pulse EPR spectrometer.

3.3.6 Binding of $(dT)_{35}$ to SSB at low salt concentration

In the cytosol of *E. coli*, the concentration of NaCl never rises to levels higher than 10 mM (Milo and Phillips, 2016). Therefore, the binding of $(dT)_{35}$ to SSB was also investigated at low salt concentration. Negative cooperativity of ssDNA binding has been reported to increase with a decrease in monovalent salt concentration (Lohman and Bujalowski, 1988). To study this effect by NMR, SSB containing ^{13}C-Tby was used to monitor the binding of $(dT)_{35}$ in 1:1 and 1:2 ratios (Figure 3.14). For either ratio, two *tert*-butyl cross-peaks were observed as expected for the $(SSB)_{35}$ binding mode, where one cross-peak corresponded to the bound form and the other to the free form of SSB. However, the peak intensities were different than predicted.

Kozlov and Lohman (2011) had shown that, at low salt concentrations, the first $(dT)_{35}$ molecule binds to SSB with high affinity, whereas the second molecule of $(dT)_{35}$ binds more weakly due to negative binding cooperativity. Therefore, a 1:1 ratio of SSB

to $(dT)_{35}$ should display two *tert*-butyl cross-peaks of equal intensities, as was observed at high salt concentration. Instead, the spectrum obtained showed two peaks, where the cross-peak pertaining to free SSB was about 3-fold more intense than the cross-peak belonging to an OB domain with bound ssDNA. It thus seems that only a single binding site is completely saturated in the presence of an equimolar amount of $(dT)_{35}$ at low salt (Figure 3.14A).

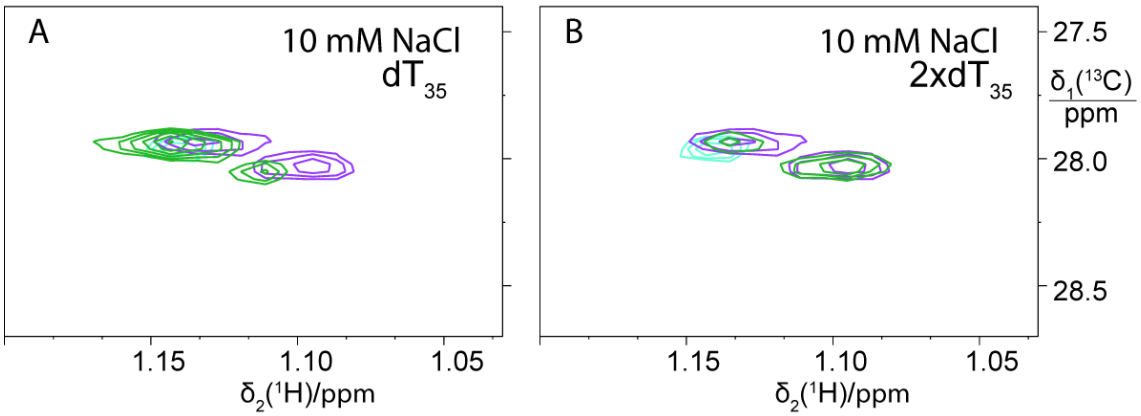


Figure 3.14 Superimposition of three ^{13}C-HSQC spectra of SSB Y97^{13}C-Tby at low salt. The spectral region shown displays the cross-peaks observed for the *tert*-butyl group in free SSB (cyan), SSB with 10 µM $(dT)_{67}$ (purple) and either (A) 10 µM $(dT)_{35}$ or (B) 20 µM $(dT)_{35}$ (green).

In the presence of two molecules of $(dT)_{35}$ and based on the results obtained at high salt (Figure 3.11B), a similar outcome as obtained with a single molecule of $(dT)_{67}$ would be expected also at low salt. Indeed, two cross-peaks were present in the spectrum, but the peak intensities were different than predicted. Unexpectedly, the spectrum shows a more intense cross-peak for the bound than the free state, indicating that more than two Tby residues interact with $(dT)_{35}$ provided in 2:1 ratio (Figure 3.14B). Therefore, the second $(dT)_{35}$ molecule seems to bind more easily than the first one, which is the hallmark of cooperative rather than non-cooperative binding.

Clearly, the difference in the length of the oligonucleotides matters at low salt. It has been reported that longer ssDNA segments such as $(dT)_{70}$ can bind to two SSB tetramers at low salt concentration ([NaCl] < 10 mM) with high inter-tetramer cooperativity which, for longer segments of ssDNA, allows the formation of long protein clusters (Griffith et al, 1984; Lohman et al. 1986; Ferrari et al., 1994; Kozlov et al., 2015). In the model structure of the $(SSB)_{35}$ mode by Raghunathan et al. (2000), a

70 nucleotides long ssDNA strand interacts with two SSB tetramers, binding to two OB domains of each tetramer. The NMR results could thus be explained if, at low salt, some weak affinity between SSB tetramers assisted in ssDNA binding. For short oligonucleotides such as $(dT)_{35}$, dimerization of SSB tetramers into octamers would be more difficult to achieve, but once an octamer has formed, it may bind additional $(dT)_{35}$ molecules more easily.

3.3.7 ssDNA binding polarity of SSB

According to the crystal structure 1EYG (Raghunathan et al., 2000), *E. coli* SSB binds ssDNA with the opposite polarity compared to all other available structures of SSB-ssDNA complexes from different organisms. This suggests that SSBs are capable of binding ssDNA in either polarity and there is no biophysical explanation to why OB domains prefer one polarity over the other (Theobald et al., 2003). Notably, however, *E. coli* SSB is the first SSB co-crystal structure with ssDNA that has been determined and its resolution is moderate at 2.8 Å (Raghunathan et al., 2000). Therefore, independent verification of the polarity of ssDNA binding would be valuable and an experiment was designed in order to determine the ssDNA binding polarity on SSB using solution NMR.

Noting that the chemical shift of the *tert*-butyl group in a Tby residue is sensitive to its chemical environment, proximity to guanosine nucleotides would be expected to result in different chemical shifts than proximity to thymidine nucleotides. Based on this assumption, different oligonucleotides were designed in an attempt to determine whether the ssDNA binding polarity of *E. coli* SSB is the same in solution as in the published crystal structure.

First, a single short stretch of guanosine nucleotides was placed in $(dT)_{67}$ oligomers, (i) with the dG nucleotides closer to the 3'-end (referred to as ssDNA-3') and (ii) with the dG nucleotides closer to the 5'-end (referred to as ssDNA-5'; see Table 3.1). Assuming that the 3' and 5' ends of the oligonucleotides are located as shown in the high-salt binding mode of Figure 3.15, a single *tert*-butyl group is expected to interact with guanosine nucleotides in ssDNA-3' whereas none of the *tert*-butyl groups will interact with guanosine nucleotides in ssDNA-5', if the ssDNA wraps around SSB

in the same polarity as in the published crystal structure (Figure 3.15). The opposite result is expected, if ssDNA wraps around SSB with the opposite polarity.

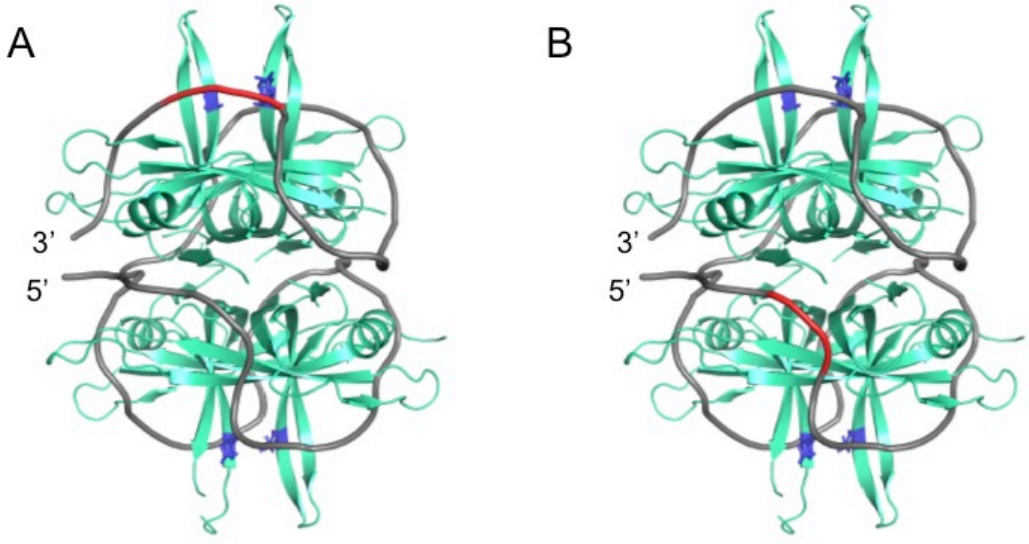


Figure 3.15 Structural model of the SSB Y97Tby tetramer with ssDNA bound in the $(SSB)_{65}$ mode, based on the crystal structure 1EYG (Raghunathan et al., 2000). The expected locations of segments of guanosine nucleotides are highlighted in red for (A) ssDNA-3' and (B) ssDNA-5'.

^{13}C-HSQC spectra of SSB at high salt concentration were measured with an equimolar ratio of ssDNA-3' or ssDNA-5'. The spectrum with ssDNA-3' showed an additional cross-peak shifted downfield by about 0.05 ppm from the position observed for Tby residues near thymidine nucleotides (Figure 3.16A). In contrast, only the cross-peak characteristic of Tby near thymidine nucleotides was observed with ssDNA-5' (Figure 3.16B). Assuming that the 5' and 3' ends of the ssDNA are positioned as indicated in Figure 3.15, this indicates that ssDNA indeed wraps around *E. coli* SSB with a polarity opposite to the "standard polarity" shown by the other published OB fold structures (Theobald et al., 2003).

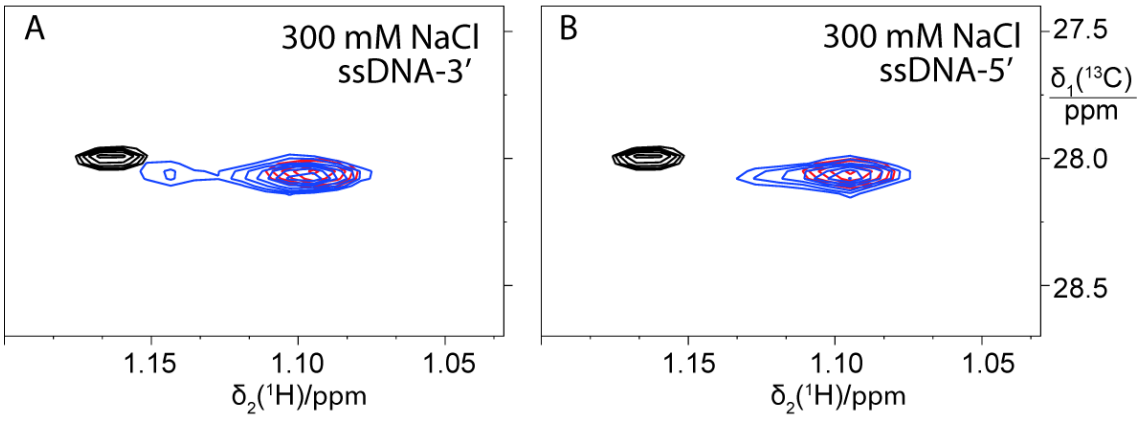


Figure 3.16 Superimposition of three ^{13}C-HSQC spectra of SSB Y97^{13}C-Tby at high salt. The spectral region shown displays the cross-peaks observed for the *tert*-butyl group of free SSB (black), SSB with 10 µM $(dT)_{67}$ (red) and (A) SSB with 10 µM ssDNA-3' or (B) SSB with 10 µM ssDNA-5' (blue).

For further confirmation, two additional oligonucleotides were designed, where four stretches of guanosine nucleotides were inserted with equal spacing, ssDNA-3'x4 and ssDNA-5'x4 (Table 3.1). According to the model of Figure 3.17, all four *tert*-butyl groups are expected to interact with guanosine nucleotides in ssDNA-3'x4 while none would interact with guanosine nucleotides in ssDNA-5'x4.

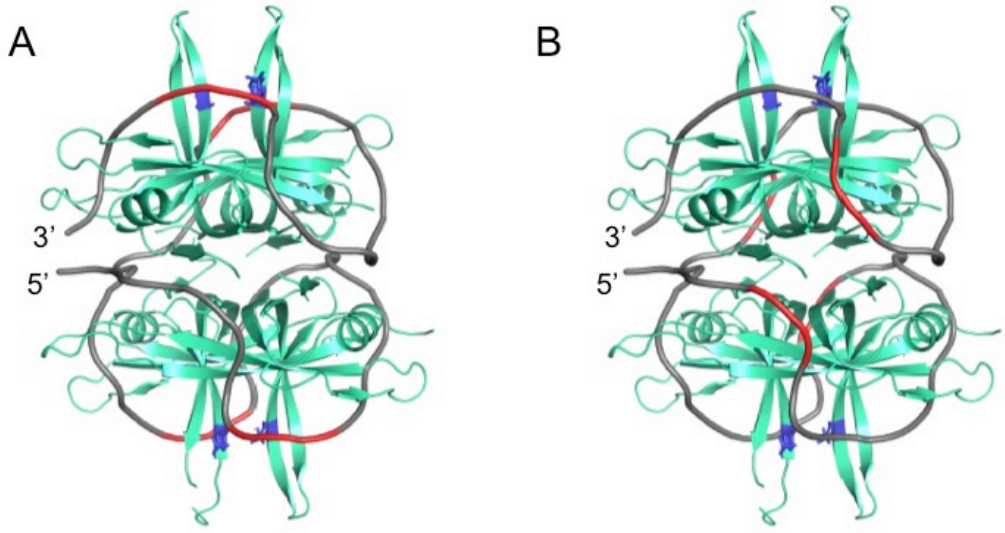


Figure 3.17 Structural model of the SSB Y97Tby tetramer with ssDNA bound in the $(SSB)_{65}$ mode based on the crystal structure 1EYG (Raghunathan et al., 2000). The expected locations of segments of guanosine nucleotides are highlighted in red for (A) ssDNA-3'x4 and (B) ssDNA-5'x4.

While single cross-peaks were expected for the complexes with either oligonucleotide sequence, albeit at different chemical shifts, this result was obtained only with ssDNA-5'x4 (Figure 3.18B). In contrast, the complex with ssDNA-3'x4 produced a series of weak cross-peaks (Figure 3.18A), which shows that the four *tert*-butyl groups experience different chemical environments. Therefore, the presence of guanosines clearly perturbs the environment of the *tert*-butyl groups most strongly in the case, when the model of Figure 3.17A predicts them to be next to guanosines. The peaks are also shifted downfield by up to 0.05 ppm, as expected based on the results obtained with ssDNA-3' (Figure 3.16A), while the cross-peak observed with ssDNA-5'x4 is only slightly shifted relative to the position expected for the thymidine-rich segments (Figure 3.18B).

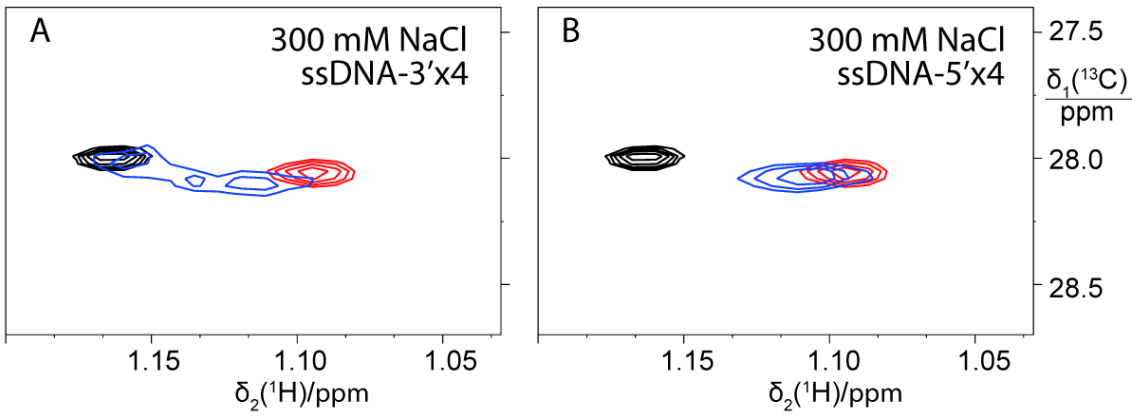


Figure 3.18 Superimposition of three ^{13}C-HSQC spectra of SSB Y97^{13}C-Tby at high salt. The spectral region shown displays the cross-peaks observed for the *tert*-butyl group in free SSB (black), SSB with 10 µM $(dT)_{67}$ (red) and (A) SSB with 10 µM ssDNA-3'x4 or (B) SSB with 10 µM ssDNA-5'x4 (blue).

While the experimental data suggest that the ssDNA binding polarity on *E. coli* SSB in solution is the same as in the crystal structure, this result is not entirely conclusive. First, it should be noted that the experiments were performed at high salt conditions, which facilitate diffusional migration of SSB along ssDNA (Zhou et al., 2011; Roy et al., 2009). Furthermore, the electron density of the ssDNA strand in the crystal structure has some discontinuities (Raghunathan et al., 2000), indicating that the ssDNA is disordered in some segments. This makes the actual model of the $(SSB)_{65}$ binding mode ambiguous. Due to these reasons, it is possible that the guanosine

nucleotides in ssDNA-3'x4 do not interact with the *tert*-butyl groups of SSB Y97^{13}C-Tby as expected based on the model shown in Figure 3.17.

3.4 Conclusion

The introduction of a ^{13}C-labelled Tby residue into SSB proved to generate a very sensitive probe for studying SSB-ssDNA complexes without perdeuteration as is usually required for solution NMR studies of systems larger than about 35 kDa. The main SSB binding modes, $(SSB)_{65}$ and $(SSB)_{35}$, were confirmed under high-salt and low-salt conditions, respectively. The effort made to determine the ssDNA binding polarity on SSB yielded somewhat ambiguous results. In general, the results agree with the polarity reported by the published crystal structure by Raghunathan et al. (2000) but considerable uncertainties arise from ambiguities in the $(SSB)_{65}$ model and possible diffusion of ssDNA on SSB.

3.5 Future work

It may be possible to gain more decisive information on the ssDNA binding polarity on SSB by biochemical rather than NMR spectroscopic techniques. For example, it has been shown that ssDNA can be cleaved site-specifically by radicals formed in Cu^{2+}-catalysed oxidation of ascorbate (Sergey et al., 1988). This could be exploited by introducing a Cu^{2+} ion at two individual sites (Trp88 and Tyr97) on the OB domain by incorporating a metal chelating unnatural amino acid or a tag. In the presence of ascorbate, the Cu^{2+} ion would result in ssDNA cleavage at those sites. Cleavage of an ssDNA oligomer with a fluorophore attached at the 5'-end could then be analysed following separation of the different fragments by urea polyacrylamide gel electrophoresis (PAGE) using a fluorescence scanner. The different length fragments produced by copper at either position 88 or 97 would give information about the ssDNA binding polarity on SSB. A potential difficulty associated with this experiment, however, would be the possibility of digesting already digested ssDNA fragments, which may be difficult to distinguish from ssDNA with single cuts.

It may also be worthwhile to attempt stabilization of ssDNA binding on SSB by the addition of a counter-ion such as Mg^{2+}. In view of the very low solubility of SSB in the presence of intermediate concentrations of NaCl, however, it may be difficult to prepare the complex in sufficiently high concentration to perform NMR spectroscopy. In *E. coli*, Mg^{2+} ions are primarily bound to ATP (at concentrations of 30 – 100 mM)

while free Mg^{2+} ions are present at concentrations of 0.01 – 1 mM concentration (Milo and Phillips, 2016), which offers a wide range of conditions qualifying as physiological.

3.6 References

Alberts, B.M. and Frey, L. (1970) T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. *Nature* **227**, 1313–1318.

Antony, E., Weiland, E.A., Korolev, S. and Lohman, T.M. (2012) *Plasmodium falciparum* SSB tetramer wraps single-stranded DNA with similar topology but opposite polarity to *E. coli* SSB. *J. Mol. Biol.* **420**, 269-283.

Apponyi, M., Ozawa, K., Dixon, N.E. and Otting, G. (2008) Cell-free protein synthesis for analysis by NMR spectroscopy. *Methods in Molecular Biology 426, Structural proteomics: high-throughput methods B. Kobe, M. Guss, T. Huber Eds., Humana Press, Totowa, USA* pp. 257-268

Bayer, I., Fliess, A., Greipel, J., Urbanke, C. and Maass, G. (1989) Modulation of the affinity of the single-stranded DNA-binding protein of *Escherichia coli* (*E. coli* SSB) to poly(dT) by site-directed mutagenesis. *Eur. J. Biochem.* **179**, 399-404.

Bochkareva, E., Korolev, S., Lees-Miller, S.P., Bochkarev, A. (2002) Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* **21**, 1855-1863.

Brill, S.J. and Stillman, B. (1989) Yeast replication factor-A functions in the unwinding of the SV40 origin of DNA replication. *Nature* **342**, 92–95.

Bujalowski, W. and Lohman T.M. (1986) *Escherichia coli* single-strand binding protein forms multiple, distinct complexes with single stranded DNA. *Biochemistry* **25**, 7799–7802.

Bujalowski, W. and Lohman, T.M. (1987) Limited co-operativity in protein-nucleic acid interactions: A thermodynamic model for the interactions of *Escherichia coli* single strand binding protein with single-stranded nucleic acids in the "beaded", $(SSB)_{65}$ mode. *J. Mol. Biol.* **195**, 897-907.

Bujalowski, W., Overman, L.B. and Lohman, T.M. (1988) Binding mode transitions of *Escherichia coli* single strand binding protein-single-stranded DNA complexes. Cation, anion, pH, and binding density effects. *J. Biol. Chem.* **263**, 4629-4640.

Casas-Finet, J.R., Khamis, M.I., Maki, A.H. and Chase, J.W. (1987) Tryptophan 54 and phenylalanine 60 are involved synergistically in the binding of *E. coli* SSB protein to single-stranded polynucleotides. *FEBS Lett.* **220**, 347–352.

Chan, K.-W., Lee, Y.-J., Wang, C.-H., Huang, H. and Sun, Y.-J. (2009) Single-stranded DNA-binding protein complex from *Helicobacter pylori* suggests an ssDNA-binding surface. *J. Mol. Biol.* **388**, 508-519.

Chase J.W. and Williams K.R. (1986) Single-standed DNA binding proteins required for DNA replication. *Annu. Rev. Biochem.* **55**, 103–136.

Chrysogelos, S. and Griffith, J. (1982) *Escherichia coli* single-strand binding protein organizes single-stranded DNA in nucleosome-like units. *J. Proc. Natl. Acad. Sci. USA* **79**, 5803–5807.

Curth, U., Greipel, J., Urbanke, C. and Maass, G. (1993) Multiple binding modes of the single-stranded DNA binding protein from *Escherichia coli* as detected by tryptophan fluorescence and site-directed mutagenesis. *Biochemistry* **32**, 2585–2591.

de Vries, J. and Wackernagel, W. (1993) Cloning and sequencing of the Serratia marcescens gene encoding a single-stranded DNA-binding protein (SSB) and its promoter region. *Gene* **127**, 39–45.

Fairman, M.P. and Stillman, B. (1988) Cellular factors required for multiple stages of SV40 DNA replication in vitro. *EMBO J.* **7**, 1211–1218.

Ferrari, M.E., Bujalowski, W. and Lohman, T.M. (1994) Co-operative binding of *Escherichia coli* SSB tetramers to single-stranded DNA in the $(SSB)_{35}$ binding mode, *J. Mol. Biol.* **236**, 106–123.

Gamsjaeger, R., Kariawasam, R., Gimenez, A.X., Touma, C., McIlwain, E., Bernardo, R.E., Shepherd, N.E., Ataide, S.F., Dong, Q., Richard, D.J., White, M.F. and Cubeddu, L. (2015) The structural basis of DNA binding by the single-stranded DNA-binding protein from *Sulfolobus solfataricus*. *Biochem. J.* **465**, 337-346.

George, N.P., Ngo, K.V., Chitteni-Pattu, S., Norais, C.A., Battista, J.R., Cox, M.M. and Keck, J.L. (2012) Structure and cellular dynamics of *Deinococcus radiodurans* single-stranded DNA (ssDNA)-binding protein (SSB)-DNA complexes. *J. Biol. Chem.* **287**, 22123-22132.

Griffith, J.D., Harris, L.D. and Register, J., III (1984) Visualization of SSB-ssDNA complexes active in the assembly of stable RecA DNA filaments. *Cold Spring Harbor Symp. Quant. Biol.* **49**, 553–559.

Golub, E.I. and Low, K.B. (1985) Conjugative plasmids of enteric bacteria from many different incompatibility groups have similar genes for single-stranded DNA-binding proteins. *J. Bacteriol.* **162**, 235–241.

Horvath, M.P., Schweiker, V.L., Bevilacqua, J.M., Ruggles, J.A. and Schultz, S.C. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* **95**, 963-974.

Khamis, M.I., Casas-Finet, J.R., Maki, A.H., Murphy, J.B. and Chase, J.W. (1987) Investigation of the role of individual tryptophan residues in the binding of *Escherichia coli* single-stranded DNA binding protein to single-stranded polynucleotides. A study by optical detection of magnetic resonance and site-selected mutagenesis. *J. Biol. Chem.* **262**, 10938–10945.

Kozlov, A.G. and Lohman, T.M. (1998) Calorimetric studies of *E. coli* SSB protein-single-stranded DNA interactions. Effects of monovalent salts on binding enthalpy. *J. Mol. Biol.* **278**, 999-1014.

Kozlov, A.G. and Lohman, T.M. (2006) Effects of monovalent anions on a temperature-

dependent heat capacity change of *Escherichia coli* SSB tetramer binding to single-stranded DNA. *Biochemistry* **45**, 5190-5205.

Kozlov, A.G. and Lohman, T.M. (2011) *E. coli* SSB tetramer binds the first and second molecules of $(dT)_{35}$ with heat capacities of opposite sign. *Biophys. Chem.* **159***,* 48-57.

Kozlov, A.G., Weiland, E., Mittal, A., Waldman, V., Antony, E., Fazio, N., Pappu, R.V. and Lohman, T.M. (2015) Intrinsically disordered c-terminal tails of *E. coli* single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA. *J. Mol. Biol.* **427**, 763-774.

Kozlov, A.G., Shinn, M.K., Weiland, E.A. and Lohman, T.M. (2017) Glutamate promotes SSB protein-protein interactions via intrinsically disordered regions. *J. Mol. Biol.* **429**, 2790-2801.

Lohman, T.M. and Overman, L.B. (1985) Two binding modes in *Escherichia coli* single strand binding protein-single-stranded DNA complexes. Modulation by NaCl concentration. *J. Biol. Chem.* **260**, 3594–3603.

Lohman, T.M., Overman, L.B. and Datta, S. (1986) Salt-dependent changes in the DNA binding co-operativity of *Escherichia coli* single strand binding protein. *J. Mol. Biol.* **187**, 603–615.

Lohman, T.M. and Bujalowski, W. (1988) Negative cooperativity within individual tetramers of *Escherichia coli* single strand binding protein is responsible for the transition between the $(SSB)_{35}$ and $(SSB)_{56}$ binding modes. *Biochemistry* **27**, 2260-2265.

Lohman, T.M. and Ferrari, M.E. (1994) *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.* **63**, 527–570.

Lu, D. and Keck, J.L. (2008) Structural basis of *Escherichia coli* single-stranded DNA-binding protein stimulation of exonuclease I. *Proc. Natl Acad. Sci. USA* **105**, 9169–9174.

Matsumoto, T., Morimoto, Y., Shibata, N., Kinebuchi, T., Shimamoto, N., Tsukihara, T. and Yasuoka, N. (2000) Roles of functional loops and the C-terminal segment of a single-stranded DNA binding protein elucidated by X-Ray structure analysis. *J. Biochem.* **127**, 329-335.

Merrill, B.M., Williams, K.R., Chase, J.W. and Konigsberg, W.H.J. (1984) Photochemical cross-linking of the *Escherichia coli* single-stranded DNA-binding protein to oligodeoxynucleotides. Identification of phenylalanine 60 as the site of cross-linking. *Biol. Chem.* **259**, 10850–10856.

Meyer, R.R. and Laine, P.S. (1990) The single-stranded DNA binding protein of *Escherichia coli*. *Microbiol. Rev.* **54**, 342–380.

Milo, R. and Phillips, R. *Cell biology by the numbers.* New York: Garland science, Taylor and Francis group, 2016.

Neylon, C., Brown, S.E., Kralicek, A.V., Miles, C.S., Love, C.A. and Dixon, N.E. (2000) Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* **39**, 11989–11999.

Overman, L.B., Bujalowski, W. and Lohman, T.M. (1988) Equilibrium binding of *Escherichia coli* single-strand binding protein to single-stranded nucleic acids in the $(SSB)_{65}$ binding mode. Cation and anion effects and polynucleotide specificity. *Biochemistry* **27**, 456-471.

Ozawa, K., Loscha, K.V., Kuppan, K.V., Loh, C.-T., Dixon, N.E. and Otting, G. (2012) High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochem. Biophys. Res. Commun.* **418**, 652–656.

Raghunathan, S., Kozlov, A.G., Lohman, T.M., and Waksman, G. (2000) Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat. Struct. Biol*. **7**, 648–652.

Savvides, S.N., Raghunathan, S., Fütterer, K., Kozlov, A.G., Lohman, T.M. and Waksman, G. (2004) The C-terminal domain of full length *E. coli* SSB is disordered even when bound to DNA. *Protein Sci.* **13**, 1942-1947.

Sergey, A.K., Tatiana, G.A., Sergey, V.M. and Valentin, V.V. (1988) Site-specific cleavage of single-stranded DNAs at unique sites by copper-dependent redox reaction. *Nature* **335**, 186-188.

Shamoo, Y., Friedman, A.M., Parsons, M.R., Konigsberg, W.H. and Steitz, T.A. (1995) Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature* **376**, 362-366.

Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M., and Keck, J.L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* **43**, 289-318.

Shishmarev, D., Wang, Y., Mason, C.E., Su, X.-C., Oakley, A.J., Graham, B., Huber, T. Dixon, N.E. and Otting, G. (2014) Intramolecular binding mode of the C-terminus of *E. coli* single-stranded DNA binding protein (SSB) determined by nuclear magnetic resonance spectroscopy. *Nucleic Acids Res.* **42**, 2750-2757.

Sigal, N., Delius, H., Kornberg, T., Gefter, M.L. and Alberts, B. (1972) A DNA unwinding protein isolated from *Escherichia coli*: its interaction with DNA and with DNA polymerases. *Proc. Natl. Acad. Sci. USA* **69**, 3537-3541.

Su, X.-C., Wang, Y., Yagi, H., Shishmarev, D., Mason, C.E., Smith, P.J., Vandevenne, M., Dixon, N.E. and Otting, G. (2014) Bound or free: interaction of the C-terminal domain of *Escherichia coli* single-stranded DNA-binding protein (SSB) with the tetrameric core of SSB. *Biochemistry* **53**, 1925-1934.

Theobald, D.L., Mitton-Fry, R.M., and Wuttke, D.S. (2003) Nucleic acid recognition by OB fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115-133.

Wei, T.-F., Bujalowski, W. and Lohman, T.M. (1992) Cooperative binding of polyamines induces the *Escherichia coli* single-strand binding protein-DNA binding mode transition. *Biochemistry* **31**, 6166-6174.

Wobbe, C.R., Weissbach, L., Borowiec, J.A., Dean, F.B., Murakami, Y., Bullock, P. and Hurwitz, J. (1987) Replication of simian virus 40 origin-containing DNA in vitro with purified proteins. *Proc. Natl. Acad. Sci. USA* **84**, 1834-1838.

Wu, Z., Lee, M.D., Carruthers, T.J., Szabo, M., Dennis, M.L., Swarbrick, J.D., Graham, B. and Otting, G. (2017) New lanthanide tag for the generation of pseudocontact shifts in DNA by site-specific ligation to a phosphorothioate group. *Bioconjugate Chem.* **28**, 1741-1748.

Yadav, T., Carrasco, B., Myers, A.R., George, N.P., Keck, J.L. and Alonso, J.C. (2012) Genetic recombination in *Bacillus subtilis*: a division of labor between two single-strand DNA-binding proteins. *Nucleic Acid Res.* **40**, 27640-27652.

Young, D.D., Young, T.S., Jahnz, M., Ahmad, I., Spraggon, G. and Schultz, P.G. (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894-1900.

CHAPTER 4

Chemical tagging with *tert*-butyl and trimethylsilyl groups for measuring intermolecular nuclear Overhauser effects in a large protein-ligand complex

Authors' contributions

In this study, I carried out all the protein sample preparations including protein expression, purification and ligation with the TMS tag and the NMR measurements. The chemical synthesis of the tag molecules was performed by the group of A/Prof. Bim Graham at Monash University, including Drs Luke Adams and Luigi Aurelio. Advice on how to handle single-stranded DNA binding protein (SSB) and initial expression constructs were provided by Dr Yao Wang.

4.1 Introduction

The overall aim of structural biology is to study the relationship between protein structure and biological function. In the past few decades, solution NMR spectroscopy has emerged as one of the best techniques to study these systems in their native states under near-physiological conditions. Very detailed analyses of macromolecular structure, dynamics and interactions of smaller macromolecules (< 30 kDa) were made possible with the emergence of multidimensional NMR combined with improved strategies for labelling proteins with NMR-active isotopes such as ^{15}N and ^{13}C (Bax, 1994; Clore and Gronenborn, 1991).

Many biologically important processes, however, are associated with larger macromolecular complexes, which are bigger than 30 kDa. In this case, solution NMR spectroscopy becomes challenging. One difficulty arises, because large molecules tumble slowly in solution, which results in faster relaxation rates of transverse magnetisation (Foster et al., 2007; Chen et al., 2015). Another problem is simply associated with the presence of an increased number of signals in the NMR spectrum, so that peak overlap makes the spectra hard to interpret.

The effect of dipole-dipole interactions on nuclear relaxation can be reduced by using deuterated samples. In particular, uniformly perdeuterated samples result in much slower T_2 relaxation rates of the remaining ^{1}H spins, because ^{2}H spins have a much smaller magnetic moment than ^{1}H spins and therefore cause much less relaxation (Sattler and Fesik, 1996; Gardner and Kay, 1998). Nevertheless, perdeuteration can also be disadvantageous as the deuterated sites of the macromolecule become NMR-silent. To minimise peak overlap in NMR spectra of high-molecular weight systems, proteins can be prepared where selected amino-acid types are isotope-labelled. This approach is not only expensive, but also does not resolve the difficulty to obtain site-specific information on the proteins, as the NMR signals of the selectively labelled amino acids still need to be assigned to specific amino-acid residues in the protein.

Due to their short-range nature, ^{1}H-^{1}H NOEs provide exceptionally detailed information to define 3D structures at atomic resolution (Foster et al., 2007). For large macromolecular systems, however, it is challenging to assign NOESY cross-peaks in the spectral region of amino-acid side-chain resonances, as peak overlap and fast relaxation make it difficult to resolve individual peaks.

In this situation, O-*tert*-butyltyrosine (Tby) provides an exceptional solution to address many of the issues associated with structural studies of high-molecular weight systems by NMR spectroscopy (Chen et al., 2015). A Tby residue can be site-specifically incorporated into proteins in response to an amber stop codon. The *tert*-butyl group appears as a narrow and intense singlet in the 1D ^{1}H-NMR spectrum. The narrow width of the signal results from rapid rotations of the methyl groups as well as rotations of the *tert*-butyl group around the carbon-oxygen bond and its intensity comes from nine chemically equivalent ^{1}H spins without any resolved two-bond or three-bond couplings.

The typical chemical shift of the *tert*-butyl group in Tby is around 1.3 ppm with minor deviations subject to the local chemical environment. In many cases, the ^{1}H-NMR signal of the *tert*-butyl signal can readily be picked in the 1D ^{1}H-NMR spectrum as a narrow peak that is more intense than other signals from the protein, but occasionally it has also been found to be masked by protein signals making its assignment ambiguous. In this situation, an outstanding alternative for a *tert*-butyl group is to use a trimethylsilyl (TMS) group, as the TMS group produces a singlet at

about 0 ppm, where only very few protein resonances are found. Therefore, the unambiguous assignment of the TMS signal is much easier.

4.2 Research objectives

This study explored the possibility to observe intermolecular NOEs in a high-molecular weight protein-DNA complex (*E. coli* SSB in complex with a 67-mer single-stranded DNA molecule) without isotope labelling. The molecular weight of the system is 95 kDa and most ^{1}H-NMR signals of the protein are too broad to allow resolving individual NOEs. A site-specifically introduced *tert*-butyl group (introduced via the unnatural amino acid Tby) or a TMS group (introduced by chemical modification of a single cysteine residue), however, deliver narrow lines in the ^{1}H-NMR spectrum, greatly increasing the chances to resolve cross-peaks in 2D NOESY spectra. Indeed, NOESY cross-peaks could be observed, which occurred in the presence but not in the absence of DNA, as expected for intermolecular NOEs. The work also introduced ^{13}C-labelled Tby as a site-selective NMR probe for monitoring changes in chemical environments with exceptional sensitivity.

4.3 References

Bax, A. (1994) Multidimensional nuclear-magnetic-resonance methods for protein studies. *Curr. Opin. Struct. Biol.* **4**, 738-744.

Chen, W.-N., Kuppan, K.V., Lee, M., Jaudzems, K., Huber, T. and Otting, G. (2015) O-*tert*-Butyltyrosine, an NMR tag for high-molecular weight systems and measurements of submicromolar ligand binding affinities *J. Am. Chem. Soc.* **137**, 4581-4586.

Clore, G.M. and Gronenborn, A.M. (1991) Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390-1399.

Foster, M.P., McElroy, C.A. and Amero, C.D. (2007) Solution NMR of large molecules and assemblies. *Biochemistry* **46**, 331-340.

Gardner, K.H. and Kay, L.E. (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 357-406.

Sattler, M. and Fesik, S. W. (1996) Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* **4**, 1245-1249.

4.4 Reproduction of the published article

Jabar, S., Adams, L., Wang, Y., Aurelio, L., Graham, B. and Otting, G. (2017) Chemical tagging with *tert*-butyl and trimethylsilyl groups for measuring intermolecular nuclear Overhauser effects in a large protein-ligand complex. *Chem. Eur. J.* **23**, 13033-13036.

Chemical tagging with *t*-butyl and trimethylsilyl groups for measuring intermolecular NOEs in a large protein-ligand complex

Shereen Jabar,[†][a] Luke A. Adams,[†][b] Yao Wang,[a] Luigi Aurelio,[b] Bim Graham,*[b] and Gottfried Otting*[a]

Abstract: Intermolecular ^{1}H-^{1}H nuclear Overhauser effects (NOE) present a powerful tool to assess contacts between proteins and binding partners, but are difficult to identify for complexes of high molecular weight. We show that intermolecular NOEs can readily be observed following chemical labeling with *t*-butyl or trimethylsilyl groups. Proteins can be furnished with *t*-butyl or TMS groups site-specifically using genetically encoded unnatural amino acids or by chemical modification of single cysteine residues. No isotope labeling is required. The approach is demonstrated with the 95 kDa complex between tetrameric *E. coli* single-stranded DNA binding protein (SSB) and single-stranded DNA.

The short-range nature of the nuclear Overhauser effect (NOE) makes it an excellent tool for obtaining intermolecular distance restraints in structural studies of protein-ligand complexes in solution. Increasing molecular mass of a complex, however, not only increases the complexity of its ^{1}H-NMR spectrum, but also the ^{1}H-NMR line widths due to increased ^{1}H-relaxation rates, making the assignment of specific intermolecular ^{1}H-^{1}H NOEs challenging. The conventional approach is to improve spectral resolution by labeling with stable isotopes (^{13}C and ^{15}N) and using more involved NMR pulse sequences such as half-filters[1,2] and three-dimensional NMR spectra.[3] Unavoidably, however, long pulse sequences entail significant signal loss due to relaxation. To combat the enhanced relaxation rates associated with high molecular weight systems, perdeuterated samples are frequently prepared to limit the dipolar relaxation pathways of the remaining ^{1}H-NMR spins,[4] but perdeuteration can change the properties of the protein[5] and the dilution of ^{1}H-NMR spins decreases the chances of observing intermolecular ^{1}H–^{1}H contacts in the protein-ligand complex.

Here we show that intermolecular NOEs can be observed for proteins of high molecular weight at natural isotopic abundance by the use of *t*-butyl or trimethylsilyl (TMS) groups as chemical tags. These groups can be site-specifically installed in proteins either by using genetically encoded unnatural amino acids or by chemical modification of a single cysteine residue. We have previously shown that O-*t*-butyl-tyrosine (Tby) can easily be incorporated into proteins in response to an amber stop codon and that, even in the context of proteins with >100 kDa molecular weight, solvent-exposed *t*-butyl groups produce a narrow and intense ^{1}H-NMR signal that can readily be observed.[6]

The TMS group equally features a singlet of nine equivalent ^{1}H spins, but no unnatural amino acid with a TMS group has been reported that can be incorporated into proteins in response to a stop codon. TMS groups are attractive because of their chemical shift of about 0 ppm, where signal overlap with other resonances in the ^{1}H-NMR spectrum is minimal. To introduce a TMS group, we labelled a cysteine residue with a TMS tag (Figure 1).

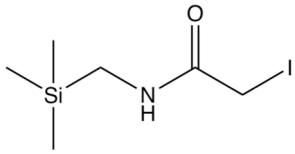


Figure 1. Chemical structure of the trimethylsilyl (TMS) tag used in the present work. The tag reacts with proteins by alkylation of cysteine residues.

We applied these labeling strategies to the complex between single-stranded DNA binding protein (SSB) from *Escherichia coli* and single-stranded DNA (ssDNA) presented by a 67-thymidine oligomer, dT_{67}. *E. coli* SSB is a homotetramer and the complex with dT_{67} has a molecular mass of 95 kDa. A co-crystal structure of *E. coli* SSB in complex with two dC_{35} oligonucleotides has been published (PDB ID 1EYG).[7] The crystals were obtained following removal of the flexible C-terminal domain by chymotryptic digestion and show the ssDNA wrapping around the OB domains of the SSB tetramer. It is known from biochemical experiments that, in the presence of high salt (>200 mM NaCl), a SSB tetramer occludes a stretch of 65 nucleotides of ssDNA, indicating that each OB domain in the tetramer binds to ssDNA.[8–10] Although the crystal structure is missing electron density for parts of the dC_{35} oligomers, it suggests how ssDNA may wrap around SSB in this binding mode (referred to as $(SSB)_{65}$). Unexpectedly, however, the ssDNA binding sites of different SSB monomers accommodate the ssDNA in different ways, with different orientations of the nucleotide bases relative to the protein (Figure 2).[7]

[a] Ms. S. Jabar, Dr. Y. Wang, Prof. G. Otting
Research School of Chemistry
Australian National University
Canberra, ACT 2601 (Australia)
E-mail: gottfried.otting@anu.edu.au

[b] Dr. L. A. Adams, Dr. L. Aurelio, Prof. B. Graham
Monash Institute of Pharmaceutical Sciences
Monash University
Parkville, VIC 3052 (Australia)
E-mail: bim.graham@monash.edu

† The first two authors contributed equally. Supporting information for this article is given via a link at the end of the document.

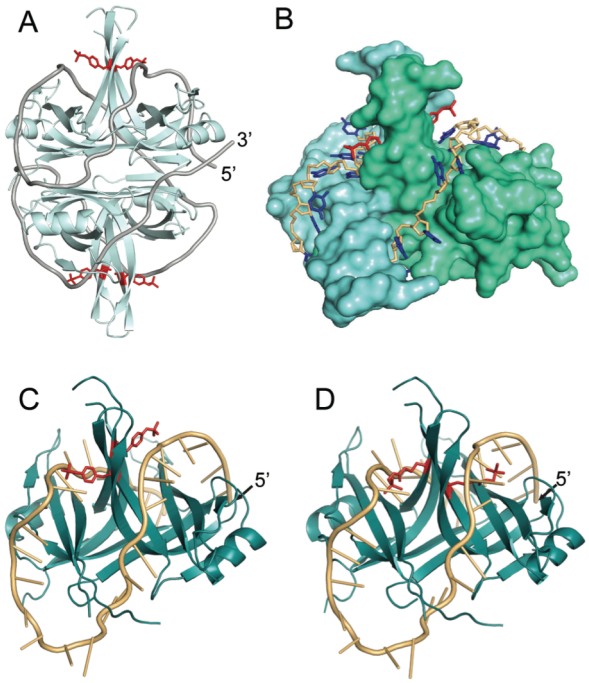


Figure 2. ssDNA binding to the OB domains of *E. coli* SSB as reported by the co-crystal structure with dC_{35}.[7] The side chains of mutated residues were modeled on the conformations of the wild-type residues and are shown in a stick representation (red). (A) Overview of the SSB tetramer with Tby modeled at position 97, showing the OB domains in a ribbon representation (cyan) and the phosphodiester backbone of ssDNA as a tube (grey). The ssDNA was modeled as a 65-mer, extrapolating from ssDNA segments observed in the crystal structure. (B) OB domain dimer corresponding to the top half of the tetramer shown in (A), slightly rotated for an improved view of the ssDNA binding grooves. The protein is shown in a surface representation (cyan and green), with the side chains of the Tby residues shown as bonds (red). Bonds of the ssDNA of the crystal structure are shown in gold, with the nucleotide bases in blue. (C) OB domain dimer and ssDNA in a ribbon representation, with sticks indicating the directions of the nucleotide bases. Note the different orientations of the nucleotide bases near the two modified amino acid residues. Tby residues are shown at position 97. (D) Same as (C), except with cysteine residues at position 98 modeled with the TMS tag of Figure 1 (red).

In the present work, we used a construct with C-terminal His-tag (referred to as *Ec*SSB) and introduced a *t*-butyl group by substituting Tyr97 by O-*t*-butyl-tyrosine-(phenyl-3,5-d2) (d2-Tby) or O-*t*-butyl-tyrosine-(*t*-butyl-$^{13}C_4$) (^{13}C-Tby). The Tby residue was incorporated by expressing *Ec*SSB by cell-free synthesis,[11] using a pETMCSI vecto[12] with an amber stop codon at position 97 and adding purified *p*-cyanophenylalanyl-tRNA synthetase, cognate $tRNA_{CUA}$,[13] and Tby. To tag *Ec*SSB with the TMS tag of Figure 1, we expressed the mutant Thr98Cys *in vivo* and incubated the purified protein with a five-fold excess of tag overnight (see the Supporting Information for details).

The ^{13}C-HSQC spectrum of *Ec*SSB Tyr97Tby showed a single cross-peak at a ^{1}H chemical shift of 1.16 ppm with a full line width at half height of about 11 Hz (Figure 3). This chemical shift is within 0.2 ppm of the chemical shift of the free Tby amino acid. In the NOESY spectrum, the singlet of the *t*-butyl group produced intense cross-peaks with resonances at about 6.8 and 7.0 ppm, which can be attributed to the aromatic ring protons of Tby (Figure 4A). Other cross-peaks are probably with neighbouring residues. No attempt was made to assign them.

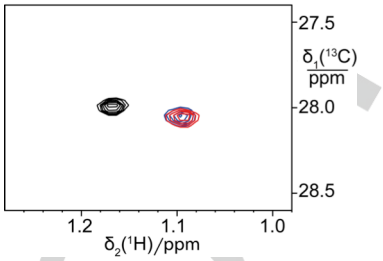


Figure 3. ^{13}C-HSQC spectra of *Ec*SSB with ^{13}C-Tby in position 97. The spectral region shown displays the cross-peak observed for the *t*-butyl group. The spectra were recorded at 25 °C of 40 µM solutions of *Ec*SSB tetramer in 20 mM phosphate buffer, pH 7.2, with 300 mM NaCl. Three spectra are superimposed. Black: free *Ec*SSB. Red: *Ec*SSB with dT_{67}. Blue: *Ec*SSB with dTG_3TGT_{61}.

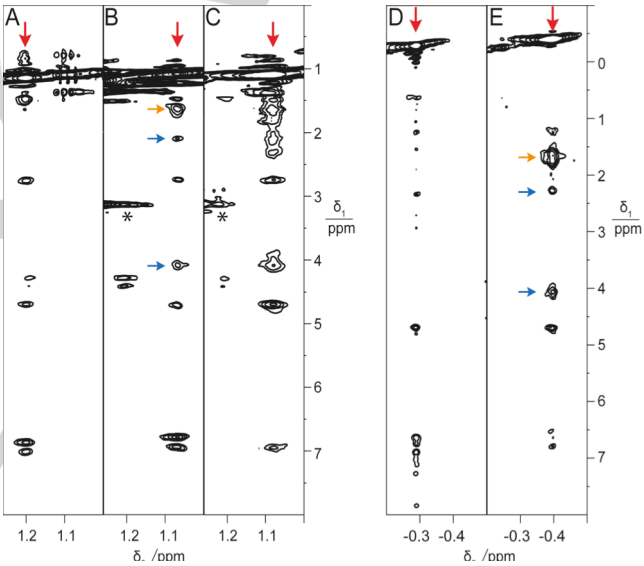


Figure 4. NOESY spectra of *Ec*SSB labelled with *t*-butyl or TMS groups in the presence and absence of ssDNA. The spectral region shown captures the NOE cross-peaks of the ^{1}H-NMR resonances of the *t*-butyl (TMS) group. The spectra were recorded with the same samples and conditions as in Figure 3. Red arrows identify the ^{1}H-NMR resonance of the *t*-butyl (TMS) group. Blue arrows identify NOE cross-peaks between the *t*-butyl (TMS) group and ssDNA. Orange arrows indicate cross-peaks at the frequency of the methyl groups of dT_{67}. (A) NOESY spectrum of free *Ec*SSB Tyr97Tby. (B) NOESY spectrum of *Ec*SSB with a 67-mer ssDNA, where all nucleotides were thymidine except for nucleotides 2, 3, 4 and 6, which were guanosine. The star identifies baseline distortions arising from a low molecular weight impurity. (C) NOESY spectrum of *Ec*SSB with dT_{67}. The ε-positions of the aromatic ring in Tby were deuterated. The star identifies the same baseline artefact as in (B). (D) NOESY spectrum of free *Ec*SSB Thr98Cys-TMS. (E) NOESY spectrum of *Ec*SSB Thr98Cys-TMS with dT_{67}.

To record intermolecular NOEs with ssDNA, we added 67-mer ssDNA in equimolar ratio. Initially, we used two different 67-mers, dT_{67} and dT_{67} with nucleotides 2, 3, 4 and 6 substituted by guanosine (dTG_3TGT_{61}). The ^{13}C-HSQC spectra of either complex showed a single peak at a ^{1}H chemical shift of 1.09 ppm, suggesting that the guanines near the 5'-end of the ssDNA do not approach the *t*-butyl group in any of the OB domains

(Figure 3). Furthermore, the single cross-peak indicates that each OB domain binds ssDNA in the same way or that exchange between different conformations is fast. Complex formation was accompanied by an increase in line width of the ^{1}H-NMR signal to 22 Hz. In the NOESY spectra, cross-peaks with the *t*-butyl group were readily resolved because most NOEs between protein protons were broadened beyond detection (Figure 4). In the presence of ssDNA, the NOESY spectrum showed additional cross-peaks with the *t*-butyl group, which must be intermolecular NOEs with dT_{67} (at 2.1 and 4.1 ppm, compare Figure 4A and B). Based on the chemical shifts, these NOEs are with sugar protons of the ssDNA. A peak at 1.6 ppm is more ambiguous, as it could be an NOE cross-peak with methyl groups of dT_{67} or an artefact associated with the intense diagonal peak. If it were with one or more thymidine methyl groups, a cross-peak with the nearby 6H proton of the base would be expected too. We therefore investigated the possibility that intermolecular NOEs with base protons of dT_{67} overlap with the intramolecular NOEs with the aromatic ring protons of the Tby residue. This was done by preparing a second sample, where the ε-protons, which are *ortho* to the O-*t*-butyl group, were replaced by deuterons. The corresponding NOESY spectrum showed the absence of the intra-residue cross-peak at 6.8 ppm as expected, and the cross-peak at 7.0 ppm was significantly weakened, indicating that it is with the δ-protons of the Tby residue (Figure 4C). These results suggest that dT_{67} projects its phosphodiester backbone towards the residue at position 97, implying that contacts with the bottom of the ssDNA binding groove are formed by nucleotide bases rather than the backbone. The crystal structure 1EYG[7] is ambiguous in this regard (Figure 2).

Due to baseline distortions near the diagonal peak of the *t*-butyl group, it was difficult to probe for intermolecular NOEs with methyl groups of dT_{67}, which would be a sensitive indicator of contacts with the nucleotide bases. To check for such NOEs and probe the orientation of the ssDNA nucleotides, we used the *Ec*SSB Thr98Cys mutant tagged with the TMS tag of Figure 1. The ^{1}H-NMR signal of the TMS group was at about -0.28 ppm (Figure 4D). Its line width was 17 Hz. Following addition of dT_{67}, the resonance shifted to -0.40 ppm (Figure 4E) and its width increased to 27 Hz. The free protein showed NOESY cross-peaks at similar chemical shifts as *Ec*SSB Tyr97Tby, except that two new cross-peaks appeared at 1.6 and 1.8 ppm, which were more intense than all other intermolecular NOEs, indicating that contacts with methyl groups of dT_{67} are readily observable. Residue 98 is located more deeply in the binding pocket of the ssDNA binding groove of the OB domains than residue 97. These NOEs thus support the view that the ssDNA projects one or more pyrimidine rings into the binding groove at this site.

The *t*-butyl group of a solvent-exposed O-*t*-butyl-tyrosine residue features rapid rotations of methyl groups and rotations around carbon-oxygen bonds, which permits the observation of a ^{1}H-NMR signal even in a large, slowly tumbling protein.[6] The present work shows that a quite narrow ^{1}H-NMR signal can also be obtained for a TMS group, which has a resonance frequency close to 0 ppm and thus little overlap with ^{1}H resonances of the protein. The signals of *t*-butyl and TMS groups can readily be identified, making further resonance assignments of the protein unnecessary. Either group can be deployed to observe site-specific intermolecular NOEs. Sensitivity of the NOESY experiment can be maximized by taking advantage of the narrow line shape of the singlet signal and using relatively long acquisition times during the detection period, while short evolution times can be used in the indirect dimension to detect NOEs with the broad resonances of the interaction partners. The example of the 95 kDa *Ec*SSB-dT_{67} complex demonstrates that such intermolecular NOEs can be detected in systems of high molecular weight without isotope labeling and with excellent sensitivity, although we used only dilute solutions containing high salt concentrations and 3 mm NMR tubes in a 5 mm probehead.

It is important to note that it is difficult to draw atomic level structural conclusions from the intermolecular NOEs in the presence of spin-diffusion, which is profound for high molecular weight systems at the mixing time used (200 ms). Furthermore, the chemical modifications may purturb native interactions. Finally, excessively broad ^{1}H-NMR line widths may prevent the observation of NOEs. For example, signals of the H6 protons of dT_{67} may have been too broad to be observable. In general, however, broadening of intramolecular NOE cross-peaks in two dimensions actually facilitates the detection of intermolecular NOEs with *t*-butyl and TMS groups, which are broad only in a single dimension. As even the signal of the TMS group may overlap with protein resonances, observation of a conserved set of NOE cross-peaks with two different chemical probes provides a good way to check for consistent results.

In conclusion, intermolecular NOEs with *t*-butyl and TMS groups can be detected in systems of high molecular weight with remarkable ease without isotope labeling. This opens the door to gain site-specific information on macromolecular interactions, at least in a qualitative way, in situations where isotope labeling and resonance assignments may be costly or impossible, as is often the case, *e.g.*, for proteins that can be produced only with eukaryotic expression systems.

Acknowledgements

We thank Professor Nicholas E. Dixon for discussions and an expression plasmid of wild-type SSB. Financial support by the Australian Research Council is gratefully acknowledged.

Keywords: nuclear magnetic resonance • nuclear Overhauser effect • single-stranded DNA binding protein • *t*-butyl group • trimethylsilyl tag

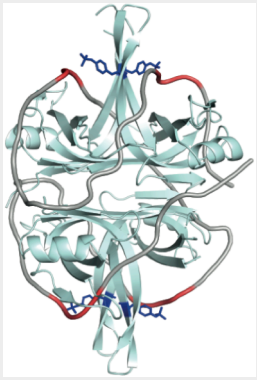
[1] G. Otting, K. Wüthrich, *Quart. Rev. Biophys*. **1990**, *23*, 39–96.
[2] A. L. Breeze, *Prog. NMR Spectr*. **2000**, *36*, 323–372.
[3] L. E. Kay, *Prog. Biophys. Mol. Biol*. **1995**, *63*, 277–299.
[4] C. H. Arrowsmith, Y. S. Wu, *Prog. NMR Spectr*. **1998**, *32*, 277–286.
[5] D. Wade, *Chemico-Biol. Interact*. **1999**, *117*, 191–217.
[6] W.-N. Chen, K. V. Kuppan, M. D. Lee, K. Jaudzems, T. Huber, G. Otting, *J. Am. Chem. Soc.* **2015**, *137*, 4581–4586.
[7] S. Raghunathan, A. G. Kozlov, T. M. Lohman, G. Waksman, *Nat. Struct. Biol*. **2000**, *7*, 648–652.
[8] T. M. Lohman, L. B. Overman, S. Datta, *J. Mol. Biol.* **1986**, *187*, 603–615.
[9] W. Bujalowski, T. M. Lohman, *Biochemistry* **1986**, *25*, 7799–7802.

[10] L. B. Overman, W. Bujalowski, T. M. Lohman, *Biochemistry* **1988**, *27*, 456–471.
[11] K. V. Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa, G. Otting, *Angew. Chem. Int. Ed.* **2012**, *51*, 2243–2246.
[12] C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love, N. E. Dixon, *Biochemistry* **2000**, *39*, 11989–11999.
[13] A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao, P. G. Schultz, *Biochemistry* **2013**, *52*, 1828–1837.

Entry for the Table of Contents

COMMUNICATION

Making contact: Intermolecular nuclear Overhauser effects between large proteins and binding partners can be observed selectively and with high sensitivity following site-specific chemical tagging of the proteins with *t*-butyl or trimethylsilyl groups.



Shereen Jabar, Luke A. Adams, Yao Wang, Luigi Aurelio, Bim Graham, Gottfried Otting**

Page No. – Page No.

Chemical tagging with *t*-butyl and trimethylsilyl groups for measuring intermolecular NOEs in a large protein-ligand complex

Supporting information

Chemical tagging with *t*-butyl and trimethylsilyl groups for measuring intermolecular NOEs in a large protein-ligand complex

Shereen Jabar, Luke A. Adams, Yao Wang, Luigi Aurelio, Bim Graham, and Gottfried Otting

Cell-free protein synthesis of *Ec*SSB Tyr97Tby

The *E. coli* SSB mutant Tyr97Tby was expressed by cell-free synthesis from a pETMCSI vector.[1] The construct included a C-terminal His_6 tag for purification of full-length *Ec*SSB using a Co-NTA column. Due to a cloning artefact, residues 132–141 from the flexible C-terminal domain were missing, but this modification is unlikely to perturb the DNA-binding mode because, although the C-terminus of SSB is known to bind to the ssDNA binding site,[2] the interaction is weak and insignificant in the presence of high concentrations of salt.[3] All NMR experiments were conducted in the presence of 300 mM NaCl to establish the $(SSB)_{65}$ mode, where the C-terminal domain is fully dissociated from the OB domain.[3]

The cell-free synthesis protocol used an *E. coli* S30 extract from which the release factor 1 had been removed,[4] and purified *p*-cyanophenylalanyl-tRNA synthetase (*p*CNF-RS) and cognate $tRNA_{CUA}$[5] were added to the inner buffer and 1 mM Tby to the outer buffer. The protein was purified using a 1 mL Co-NTA gravity column equilibrated with 50 mM Tris-HCl and 300 mM NaCl at pH 7.5. Following a wash step with buffer containing 15 mM imidazole, the column was washed with 1 M $MgCl_2$ to remove bound ssDNA from the protein before eluting the protein with buffer containing 300 mM imidazole. The final expression yield was 0.8 mg of *Ec*SSB Tyr97Tby per mL of reaction mixture. For NMR measurements, the buffer was exchanged for NMR buffer (20 mM phosphate buffer, pH 7.2, with 300 mM NaCl) and the protein concentrated to a final concentration of 40 μM in *Ec*SSB tetramer.

In vivo protein sample preparation

The *E. coli* SSB Thr98Cys mutant was expressed *in vivo* from a pETMCSI vector in *E. coli* BL21 (DE3) cells. The construct contained the same C-terminal modifications as the

construct for *Ec*SSB Tyr97Tby, allowing purification by a Co-NTA column. The cells were incubated at 37 ^{o}C until OD_{600} reached 0.8, induced with 1 mM IPTG and incubated at room temperature for 16 hours. The cells were harvested by centrifugation and were lysed by passing through a French Press. The protein was purified on an FPLC with a 5 mL Co-NTA column, using a similar protocol as for the purification of *Ec*SSB Tyr97Tby. For the tagging reaction, the buffer was exchanged for NMR buffer and glycerol was added to a final concentration of 10% for improved solubility of the TMS tag. The tagging reaction used a five-fold excess of the tag with overnight shaking of the sample in the dark and at room temperature. After removal of excess tag the attachment of the tag was confirmed by mass spectrometry.

Oligonucleotides

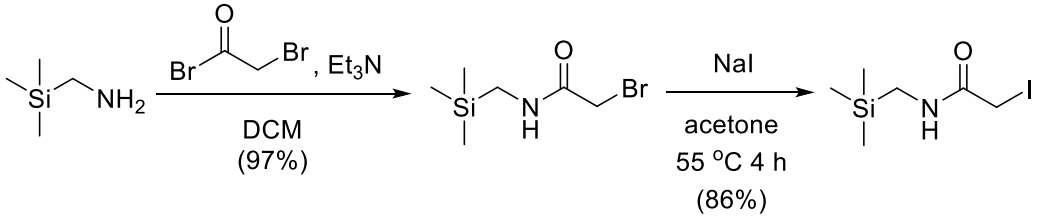
All deoxyoligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT) in HPLC-purified form.

NMR measurements

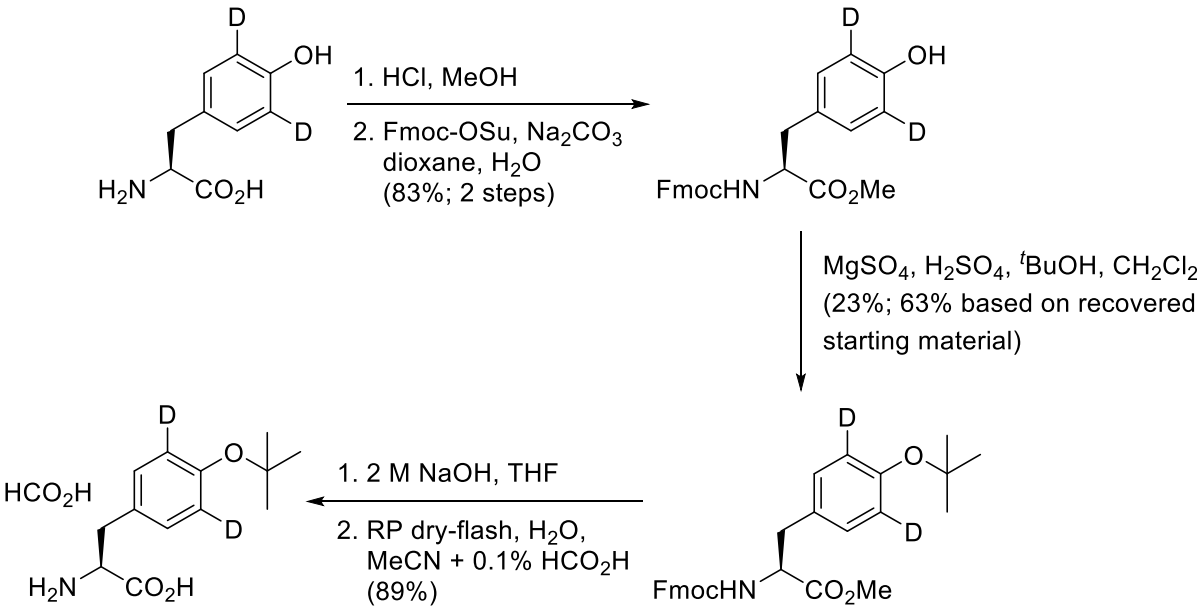
All NMR experiments were recorded on a Bruker 800 MHz NMR spectrometer equipped with a 5 mm TCI cryoprobe, using 3 mm NMR tubes. Parameters for the ^{13}C-HSQC spectra: t_{1max} = 21.2 ms, t_{2max} = 150.7 ms, total recording time per spectrum 20 min. $^{1}J_{CC}$ couplings were refocused by a selective 180^{o}(^{13}C) pulse on the resonance of the quaternary carbon during t_1. Parameters for the NOESY spectra: t_{1max} = 14.3 ms, t_{2max} = 97.5 ms, mixing time 200 ms, total recording time per spectrum 20–40 h. For samples prepared with ^{13}C-Tby, ^{13}C-decoupling was used in both dimensions.

Chemical synthesis

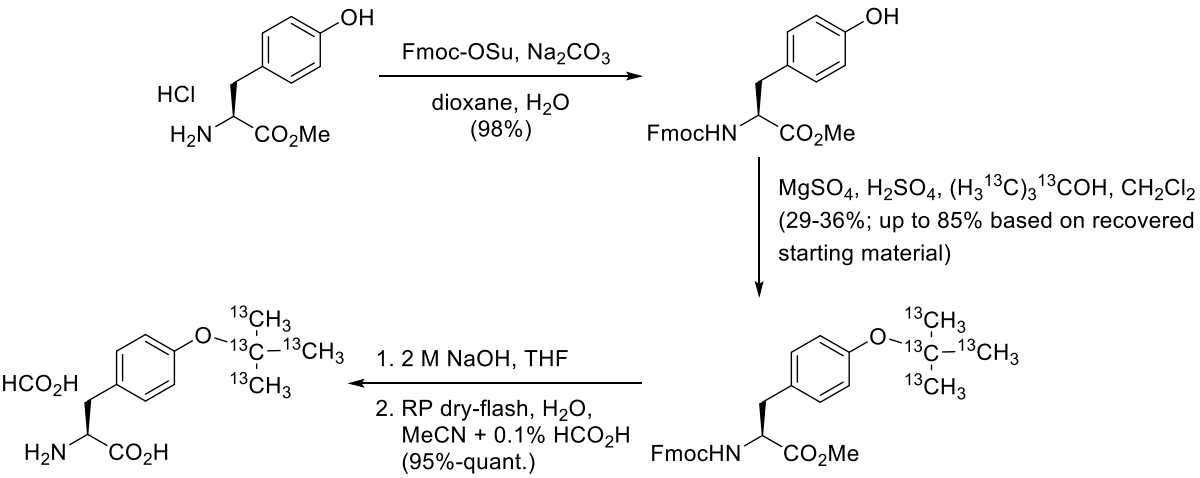
The synthesis of the TMS tag, O-*tert*-butyl-tyrosine-(phenyl-3,5-d2) (d2-Tby) and O-(*tert*-butyl-$^{13}C_4$)-tyrosine (^{13}C-Tby) were carried out as per Schemes S1, S2 and S3, respectively.



Scheme S1. Synthesis of the TMS tag.



Scheme S2. Synthesis of d2-Tby.



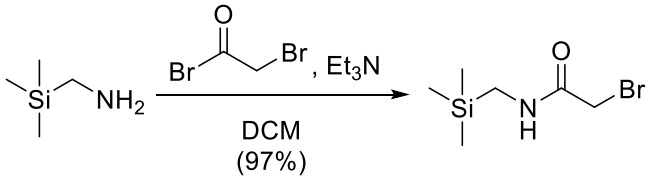
Scheme S3. Synthesis of ^{13}C-Tby.

General procedures and methods

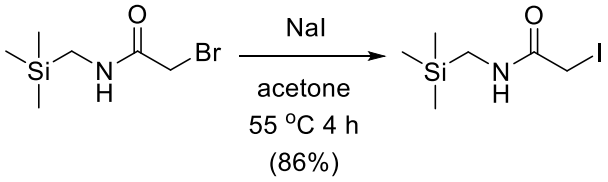
Chemicals and solvents were purchased from standard suppliers and used without further purification. Analytical thin-layer chromatography analysis (TLC) was performed on pre-coated silica gel aluminium-backed plates (Merck Kieselgel 60 F_{254}). Visualisation was by examination under UV light at 254 nm, with subsequent staining by $KMnO_4$. Dry-flash column chromatography[6] was run using Davisil® P60 silica gel (40–63μm).

Optical rotation was measured in a JASCO P-2000 polarimeter at the sodium D line in a cell with 100 mm path length. Nuclear magnetic resonance (NMR) spectra were recorded on either a Bruker Ascend or a Bruker Avance Nanobay III 400 MHz Ultrashield Plus spectrometer at 400 MHz (^{1}H) and 101 MHz (^{13}C). Chemical shifts (δ) are recorded in parts per million (ppm) downfield in relation to tetramethylsilane using the residual solvent as internal standard ($CDCl_3$, δ_H = 7.26 and δ_C = 77.16; D_2O, δ_H = 4.79).[7] Coupling constants (*J*) are reported in Hz with the following splitting abbreviations: singlet (s), doublet (d), triplet (t), quartet (q), broad (br), multiplet (m). High-resolution mass spectra (HRMS) were obtained from an Agilent 6224 TOF LC/MS Mass Spectrometer coupled to an Agilent 1290 Infinity. All data were acquired and reference mass corrected *via* a dual-spray electrospray ionisation (ESI) source.

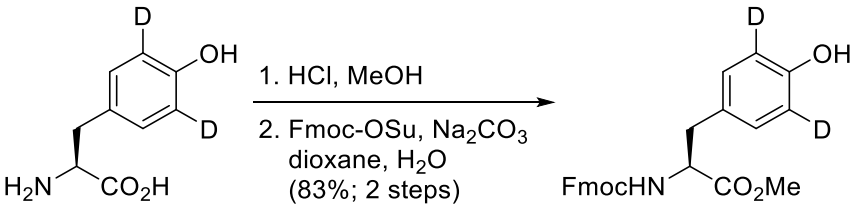
Liquid chromatography mass spectrometry (LCMS) was performed with an Agilent 6120 Series Single Quad coupled to an Agilent 1260 Series HPLC. The following buffers were used: buffer A: 0.1% formic acid in H_2O; buffer B: 0.1% formic acid in MeCN. The following gradient was used with a Poroshell 120 EC-C18 50 x 3.0 mm 2.7 micron column, and a flow rate of 0.5 mL/min and total run time of 5 min: 0–1 min 95% buffer A and 5% buffer B, from 1–2.5 min up to 0% buffer A and 100% buffer B, held at this composition until 3.8 min, 3.8–4 min 95% buffer A and 5% buffer B, held until 5 min at this composition. Mass spectra were acquired in positive and negative ion mode with a scan range of 100-1000 *m/z*. UV detection was carried out at 214 and 254 nm. All retention times (R_t) are quoted in min.



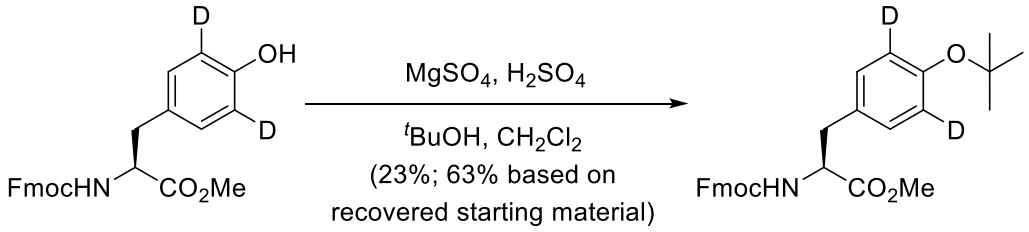
2-Bromo-N-((trimethylsilyl)methyl)acetamide. A solution of (trimethylsilyl)methylamine (500 µL, 3.73 mmol) and Et_3N (520 µL, 3.73 mmol) in anhydrous CH_2Cl_2 (4 mL) was added dropwise to a cooled (0 °C) solution of bromoacetyl bromide (325 µL, 3.73 mmol) in anhydrous CH_2Cl_2 (5 mL), under N_2. The reaction mixture was allowed to warm to room temperature (RT) and stirred for 21 h. The reaction mixture was then diluted with CH_2Cl_2 (50 mL) and washed sequentially with H_2O (25 mL) acidified to pH 1 with 1 M HCl, H_2O (25 mL), brine and dried (Na_2SO_4). Solvent was removed under reduced pressure to give the title compound as a yellow oil (809 mg, 97%), which was used without purification. ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.35 (br s, 1H), 3.90 (s, 2H), 2.79 (d, J = 5.8 Hz, 2H), 0.10 (s, 9H). ^{13}C NMR ($CDCl_3$) δ 165.3 (C), 30.6 (CH_2), 29.7 (CH_2), -2.61 (CH_3). MS (ES) *m/z*: 224 and 226 $[M+H]^+$. LCMS: R_t = 3.24 min; area 100%. HRMS: calcd for $C_6H_{14}BrNOSi + H^+$, 224.0101; found (ES, $[M+H]^+$ obsd), 224.0103.



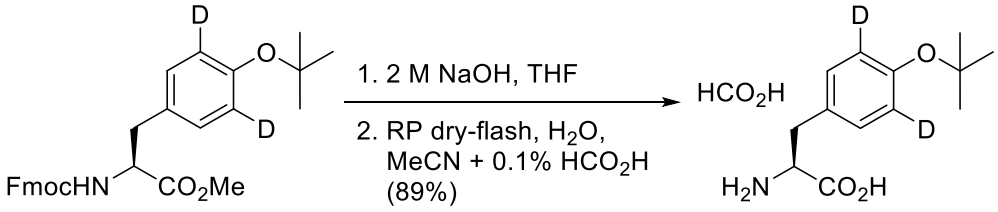
2-Iodo-N-((trimethylsilyl)methyl)acetamide. Sodium iodide (527 mg, 3.51 mmol) was added to a solution of 2-bromo-N-((trimethylsilyl)methyl)acetamide (788 mg, 3.51 mmol) in acetone (6 mL). The suspension was heated, under N_2, at 55 °C for 4 h and then was allowed to cool RT. The reaction mixture was filtered and the filter was washed with acetone. The filtrate was concentrated under reduced pressure and the resultant residue was diluted with Et_2O (50 mL). The organic was washed sequentially with H_2O (25 mL), 10% aqueous sodium thiosulfate (25 mL), brine and dried (Na_2SO_4). Solvent was removed under reduced pressure to give the title compound as a yellow oil that solidified on standing (819 mg, 86%), which was used without purification. ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.02 (br s, 1H), 3.71 (s, 2H), 2.77 (d, J = 5.8 Hz, 2H), 0.09 (s, 9H). ^{13}C NMR ($CDCl_3$) δ 166.6 (C), 31.0 (CH_2), 0.06 (CH_2), -2.52 (CH_3). MS (ES) *m/z*: 272 $[M+H]^+$. LCMS: R_t = 3.11 min; area 100%. HRMS: calcd for $C_6H_{14}INOSi + H^+$, 271.9962; found (ES, $[M+H]^+$ obsd), 271.9963.



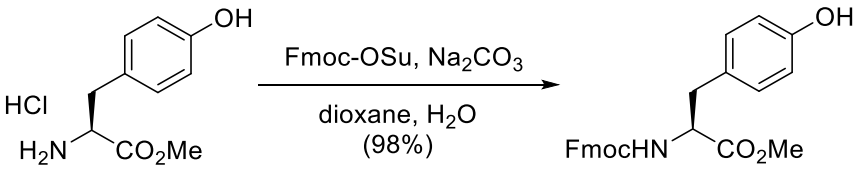
N-(9-Fluorenylmethyloxycarbonyl)-L-tyrosine-(phenyl-3,5-d2) methyl ester. A solution of methanolic HCl was prepared by slow addition of acetyl chloride (1.94 mL, 27.3 mmol) to a cooled (0 °C) suspension of anhydrous $MgSO_4$ (100 mg) in MeOH (11 mL). The mixture was stirred for 15 min prior to the addition of L-tyrosine-(phenyl-3,5-d2) (1.00 g, 5.46 mmol) in one portion. The reaction mixture was then allowed to warm to RT and stirred for 17 h. The reaction solvent was removed under reduced pressure and then toluene was added. Toluene was removed under reduced pressure to give a white solid (L-tyrosine-(phenyl-3,5-d2) methyl ester hydrochloride), which was used without purification. The solid was suspended in a mixture of dioxane (5 mL) and 10% aqueous Na_2CO_3 (11 mL) and cooled to 0 °C. A solution of Fmoc N-hydroxysuccinimide ester (2.03 g, 6.00 mmol) in dioxane (10 mL) was added dropwise to the cooled amino acid mixture. The reaction was then allowed to warm to RT slowly and stirred for 2 d. The reaction mixture was poured onto ice/water (100 mL) and stirred vigorously. Layers were separated and the aqueous was extracted with Et_2O (3 × 25 mL). The combined organics were washed with brine, dried (Na_2SO_4) and solvent was removed under reduced pressure. The crude material was purified by silica dry-flash chromatography, eluting with a gradient of 25–50% EtOAc in petroleum benzine to give the title compound as a pale yellow solid (1.89 g, 83% yield over 2 steps). $[\alpha]^{26}_D$ = +28.4 (*c* = 1.00, $CHCl_3$). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, *J* = 7.4 Hz, 2H), 7.61–7.49 (m, 2H), 7.40 (t, *J* = 7.4 Hz, 2H), 7.31 (td, *J* = 7.4, 1.1 Hz, 2H), 6.94 (s, 2H), 5.27 (br d, *J* = 8.3 Hz, 1H), 4.70–4.58 (m, 1H), 4.44 (dd, *J* = 10.6, 7.0 Hz, 1H), 4.36 (dd, *J* = 10.6, 7.0 Hz, 1H), 4.21 (t, *J* = 7.0 Hz, 1H), 3.73 (s, 3H), 3.07 (dd, *J* = 14.0, 5.6 Hz, 1H), 3.01 (dd, *J* = 14.0, 5.9 Hz, 1H). ^{13}C NMR ($CDCl_3$) δ 172.3 (C), 155.8 (C), 155.0 (C), 144.0 (C), 143.8 (C), 141.5 (C), 130.5 (CH), 127.9 (CH), 127.7 (C), 127.2 (CH), 125.2 (CH), 120.1 (CH), 115.7–115.1 (m, CD), 67.1 (CH_2), 55.1 (CH), 52.5 (CH_3), 47.3 (CH), 37.6 (CH_2). MS (ES) *m/z*: 420 $[M+H]^+$. LCMS: R_t = 3.39 min; area 96%. HRMS: calcd for $C_{25}H_{21}D_2NO_5 + H^+$, 420.1775; found (ES, [M + H]+ obsd), 420.1764.



N-(9-Fluorenylmethyloxycarbonyl)-O-tert-butyl-L-tyrosine-(phenyl-3,5-d2) methyl ester. This procedure was adapted from the literature.[8] The reaction was run in parallel three times on 250 mg scale. Concentrated H_2SO_4 (31.8 μL, 596 μmol) was added to a vigorously stirred suspension of anhydrous $MgSO_4$ (287 mg, 2.38 mmol) in anhydrous CH_2Cl_2 (3.4 mL) in a 10 mL microwave vial at 25 °C. The mixture was stirred for 15 min. A solution of N-(9-fluorenylmethyloxycarbonyl)-L-tyrosine-(phenyl-3,5-d2) methyl ester (250 mg, 596 μmol) in anhydrous CH_2Cl_2 (1.1 mL) and *tert*-butanol (285 μL, 2.98 mmol) was then added. The three reactions were then sealed tightly and stirred vigorously at 25 °C for 4–14 h, with monitoring by LCMS. Reactions were stopped at around 45% conversion, since prolonged reaction times led to decomposition. The three reactions were diluted with CH_2Cl_2 (3 × 5 mL), poured into a cooled (0 °C) solution of saturated aqueous $NaHCO_3/H_2O$ (1:1, 40 mL) and stirred vigorously. Layers were separated and the aqueous layer was extracted with CH_2Cl_2 (3 × 50 mL). The combined organics were washed with brine, dried (Na_2SO_4) and solvent was removed under reduced pressure. The crude material was purified by silica dry-flash chromatography, eluting with a gradient of 5–25% EtOAc in petroleum benzine to give the title compound as a colourless gum (193 mg, 23%). $[\alpha]^{27}_D$ = +31.9 (*c* = 0.74, $CHCl_3$). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, *J* = 7.5 Hz, 2H), 7.58 (dd, *J* = 7.2, 3.9 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.35–7.28 (m, 2H), 6.99 (s, 2H), 5.26 (br d, *J* = 8.2 Hz, 1H), 4.70–4.58 (m, 1H), 4.44 (dd, *J* = 10.6, 7.1 Hz, 1H), 4.35 (dd, *J* = 10.6, 7.1 Hz, 1H), 4.21 (t, *J* = 7.1 Hz, 1H), 3.71 (s, 3H), 3.09 (dd, *J* = 12.8, 4.7 Hz, 1H), 3.05 (dd, *J* = 12.8, 4.9 Hz, 1H), 1.32 (s, 9H). ^{13}C NMR ($CDCl_3$) δ 172.1 (C), 155.7 (C), 154.6 (C), 144.0 (C), 143.9 (C), 141.5 (C), 130.6 (C), 129.8 (CH), 127.9 (CH), 127.2 (CH), 125.2 (CH), 125.2 (CH), 124.0 (t, $J_{C\text{-}D}$ = 23.7 Hz, CD), 120.1 (CH), 78.6 (C), 67.1 (CH_2), 55.0 (CH), 52.4 (CH_3), 47.3 (CH), 37.8 (CH_2), 29.0 (CH_3). MS (ES) *m/z*: 476 $[M+H]^+$. LCMS: R_t = 3.66 min; area 98%. HRMS: calcd for $C_{29}H_{29}D_2NO_5$ + H^+, 476.2401; found (ES, $[M+H]^+$ obsd), 476.2396. N-(9-Fluorenylmethyloxycarbonyl)-L-tyrosine-(phenyl-3,5-d2) methyl ester was recovered as a white solid (302 mg, 40%). Spectroscopic data were in agreement with those reported above.

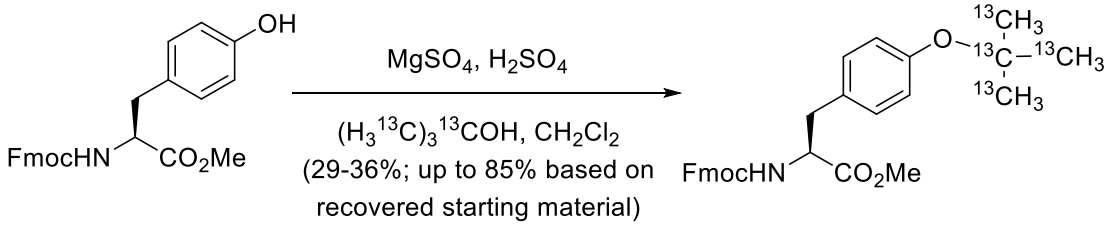


O-tert-Butyl-L-tyrosine-(phenyl-3,5-d2) formate salt. A solution of 2 M NaOH (4 mL) was added to a solution of N-(9-fluorenylmethyloxycarbonyl)-O-*tert*-butyl-L-tyrosine-(phenyl-3,5-d2) methyl ester (181 mg, 381 μmol) in THF (4 mL) and the mixture was stirred at 25 °C for 16 h. THF was removed under reduced pressure and the aqueous was diluted with H_2O (10 mL) and then washed with Et_2O (2 × 10 mL). Residual Et_2O was removed under reduced pressure. The aqueous layer was then subjected to reverse phase silica dry-flash chromatography, eluting with a gradient of 0–30% MeCN in H_2O with 0.1% formic acid to give the title compound as a white solid (150 mg, 89% including 35% w/w excess formic acid as determined by ^{1}H NMR), after lyophilization. $[\alpha]^{27}_D$ = -6.1 (*c* = 1.00, H_2O). ^{1}H NMR (400 MHz, D_2O) δ 8.46 (s, 4.4H, formic acid), 7.28 (s, 2H), 3.98 (dd, *J* = 7.9, 5.2 Hz, 1H), 3.27 (dd, *J* = 14.6, 5.2 Hz, 1H), 3.11 (dd, *J* = 14.6, 7.9 Hz, 1H), 1.36 (s, 9H). ^{13}C NMR (D_2O) δ 173.9 (C), 171.1 (CH), 152.9 (C), 131.2 (C), 130.0 (CH), 81.0 (C), 56.0 (CH), 35.7 (CH_2), 27.8 (CH_3). MS (ES) *m/z*: 240 $[M+H]^+$. LCMS: R_t = 2.83 min; area 100%. HRMS: calcd for $C_{13}H_{17}D_2NO_3 + H^+$, 240.1563; found (ES, $[M+H]^+$ obsd), 240.1555.



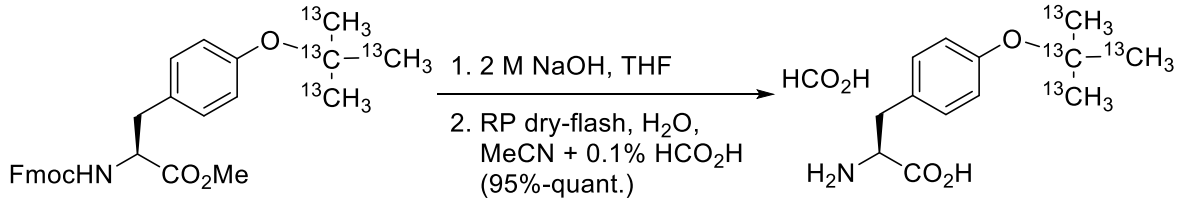
N-(9-Fluorenylmethyloxycarbonyl)-L-tyrosine methyl ester.[9] A solution of Fmoc N-hydroxysuccinimide ester (32.0 g, 95.0 mmol) in dioxane (160 mL) was added dropwise to a cooled (10 °C) suspension of L-tyrosine methyl ester hydrochloride (20 g, 86.3 mmol) in dioxane (80 mL) and 10% aqueous Na_2CO_3 (170 mL). The reaction became very viscous and after 15 min it was allowed to warm to RT, with the addition continuing for a further 1 h. The reaction was then stirred for a further 23 h. The reaction mixture was poured onto ice (~ 1.3 L) and Et_2O (400 mL) was added. Precipitated solids were filtered and then washed with H_2O. The solid was dried under suction and then over KOH under reduced pressure to give the title compound as a white solid (35.3 g, 98%), which was used without purification. $[\alpha]^{27}_D$

= +32.3 (*c* = 1.00, $CHCl_3$), Lit.[10] $[\alpha]^{20}_D$ = +34.7 (*c* = 0.50, $CHCl_3$). ^1H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, *J* = 7.5 Hz, 2H), 7.57 (dd, *J* = 7.0, 4.5 Hz, 2H), 7.40 (t, *J* = 7.4 Hz, 2H), 7.36–7.28 (m, 2H), 6.94 (d, *J* = 8.4 Hz, 2H), 6.73 (d, *J* = 8.4 Hz, 2H), 5.27 (br d, *J* = 8.2 Hz, 1H), 5.16 (br s, 1H), 4.68–4.58 (m, 1H), 4.44 (dd, *J* = 10.6, 7.0 Hz, 1H), 4.36 (dd, *J* = 10.6, 7.0 Hz, 1H), 4.21 (t, *J* = 7.0 Hz, 1H), 3.73 (s, 3H), 3.07 (dd, *J* = 14.1, 5.7 Hz, 1H), 3.01 (dd, *J* = 14.1, 6.0 Hz, 1H). MS (ES) *m/z*: 440 $[M+Na]^+$. LCMS: R_t = 3.43 min; area 94%.



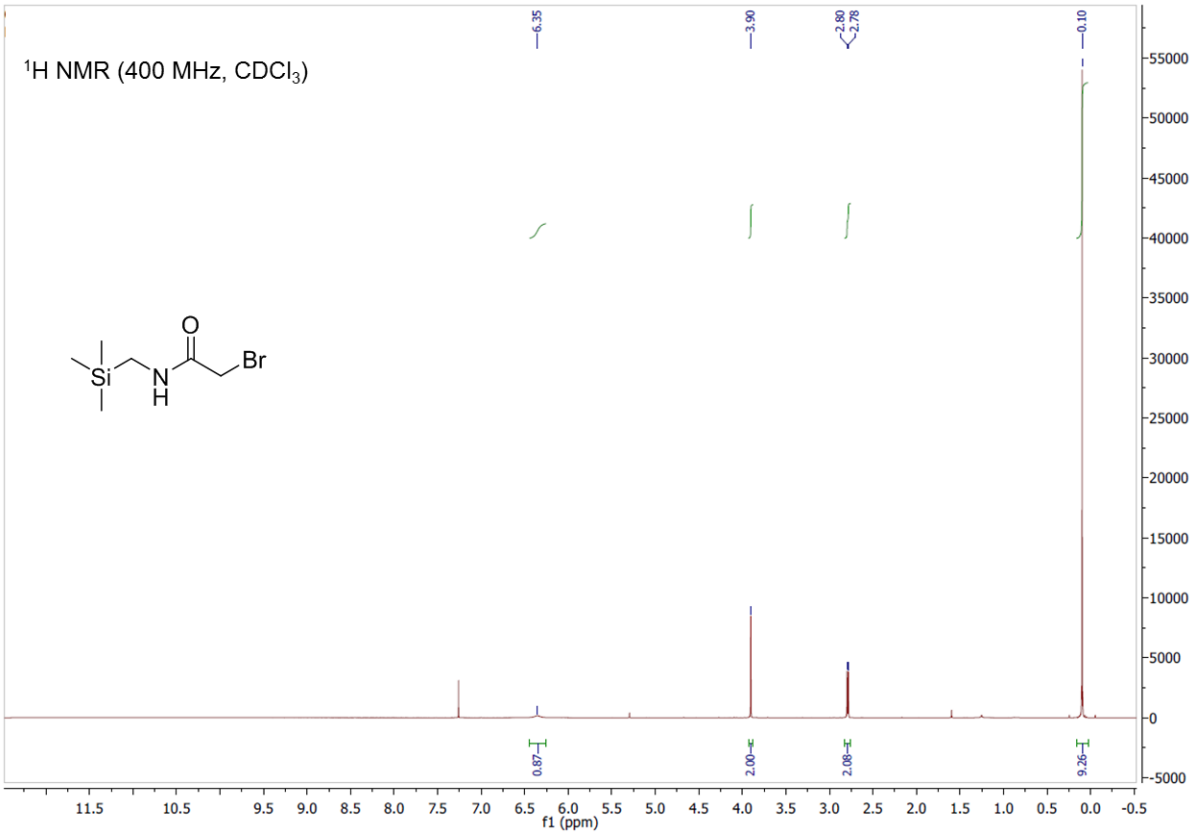
N-(9-Fluorenylmethyloxycarbonyl)-O-(tert-butyl-$^{13}C_4$)-L-tyrosine methyl ester. This procedure was adapted from the literature.[8] The reaction was run in parallel four times on 267 mg scale. Concentrated H_2SO_4 (34.1 μL, 640 μmol) was added to a vigorously stirred suspension of anhydrous $MgSO_4$ (308 mg, 2.56 mmol) in anhydrous CH_2Cl_2 (1.8 mL) in a 10 mL microwave vial at 25 °C. The mixture was stirred for 15 min. A solution of N-(9-fluorenylmethyloxycarbonyl)-L-tyrosine methyl ester (267 mg, 640 μmol) in anhydrous CH_2Cl_2 (600 μL) and *tert*-butanol-$^{13}C_4$ (153 μL, 1.60 mmol) was then added, with additional CH_2Cl_2 (200 μL) used for transfer. The four reactions were then sealed tightly and stirred vigorously at 25 °C for 4–8 h, with monitoring by LCMS. Reactions were stopped around 35% conversion, since prolonged reaction times led to decomposition. The four reactions were diluted with CH_2Cl_2 (3 × 5 mL per vial, 60 mL total), poured into a cooled (0 °C) solution of saturated aqueous $NaHCO_3/H_2O$ (1:1, 40 mL) and stirred vigorously. Layers were separated and the aqueous layer was extracted with CH_2Cl_2 (3 × 50 mL). The combined organics were washed with brine, dried (Na_2SO_4) and solvent was removed under reduced pressure. The crude material was purified by silica dry-flash chromatography, eluting with a gradient of 5–25% EtOAc in petroleum benzine to give the title compound as a colourless gum (364 mg, 30%). $[\alpha]^{27}_D$ = +27.3 (*c* = 1.00, $CHCl_3$). ^1H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, *J* = 7.5 Hz, 2H), 7.58 (dd, *J* = 7.1, 3.9 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.36–7.28 (m, 2H), 6.99 (d, *J* = 8.3 Hz, 2H), 6.91 (d, *J* = 8.3 Hz, 2H), 5.26 (br d, *J* = 8.1 Hz, 1H), 4.70–4.59 (m, 1H), 4.44 (dd, *J* = 10.6, 7.1 Hz, 1H), 4.36 (dd, *J* = 10.6, 7.1 Hz, 1H), 4.21 (t, *J* = 7.1 Hz, 1H), 3.71 (s, 3H), 3.15–2.99 (m, 2H), 1.57–1.11 (m, 9H). ^{13}C NMR ($CDCl_3$) δ 172.1 (C),

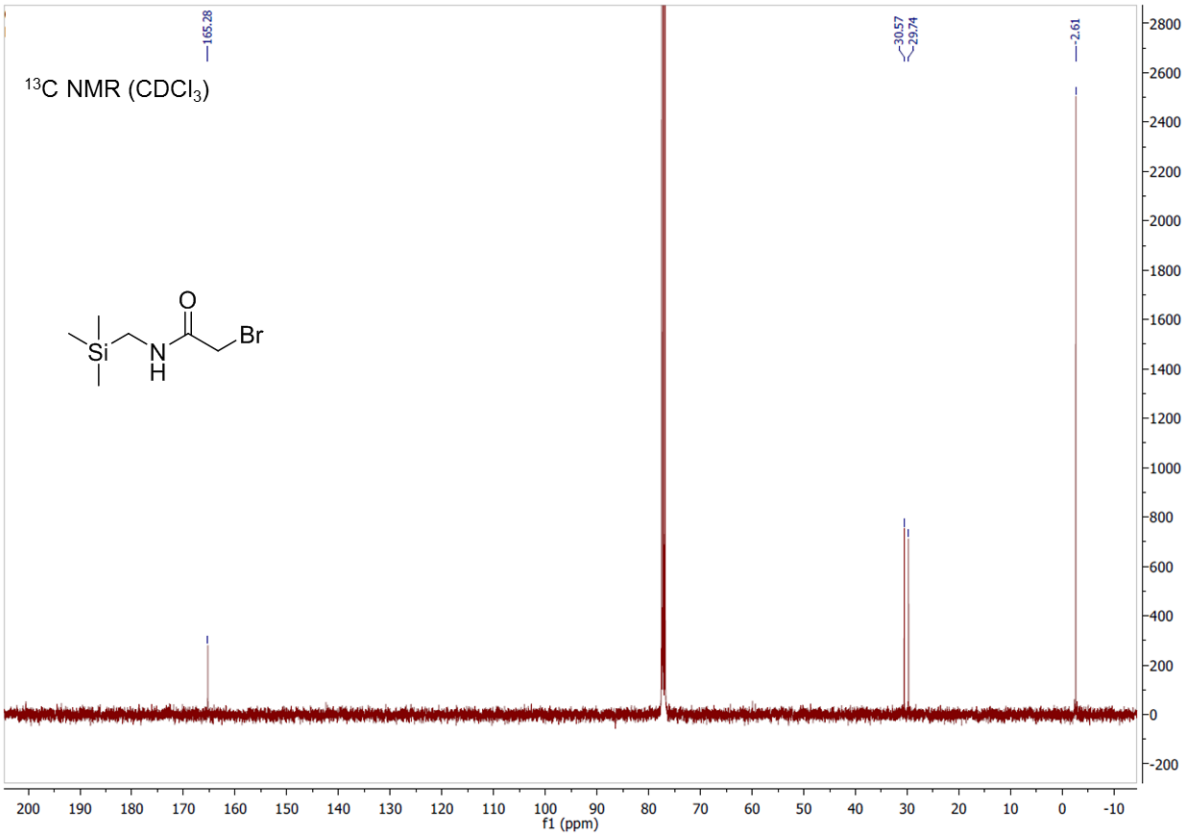
155.7 (C), 154.7 (C), 144.0 (C), 143.9 (C), 141.5 (C), 130.6 (C), 129.9 (CH), 127.9 (CH), 127.2 (CH), 125.2 (CH), 125.2 (CH), 124.3 (CH), 120.1 (CH), 78.6 (q, $J_{13C\text{-}13C}$ = 40.1 Hz, ^{13}C), 67.1 (CH_2), 55.0 (CH), 52.4 (CH_3), 47.3 (CH), 37.8 (CH_2), 29.0 (d, $J_{13C\text{-}13C}$ = 40.1 Hz, $^{13}CH_3$). MS (ES) *m/z*: 478 $[M+H]^+$. LCMS: R_t = 3.66 min; area 97%. HRMS: calcd for $C_{25}{}^{13}C_4H_{31}NO_5$ + H^+, 478.2409; found (ES, $[M+H]^+$ obsd), 478.2426. N-(9-Fluorenylmethyloxycarbonyl)-L-tyrosine methyl ester was recovered as a white solid (518 mg, 49%). Spectroscopic data were in agreement with those reported above.

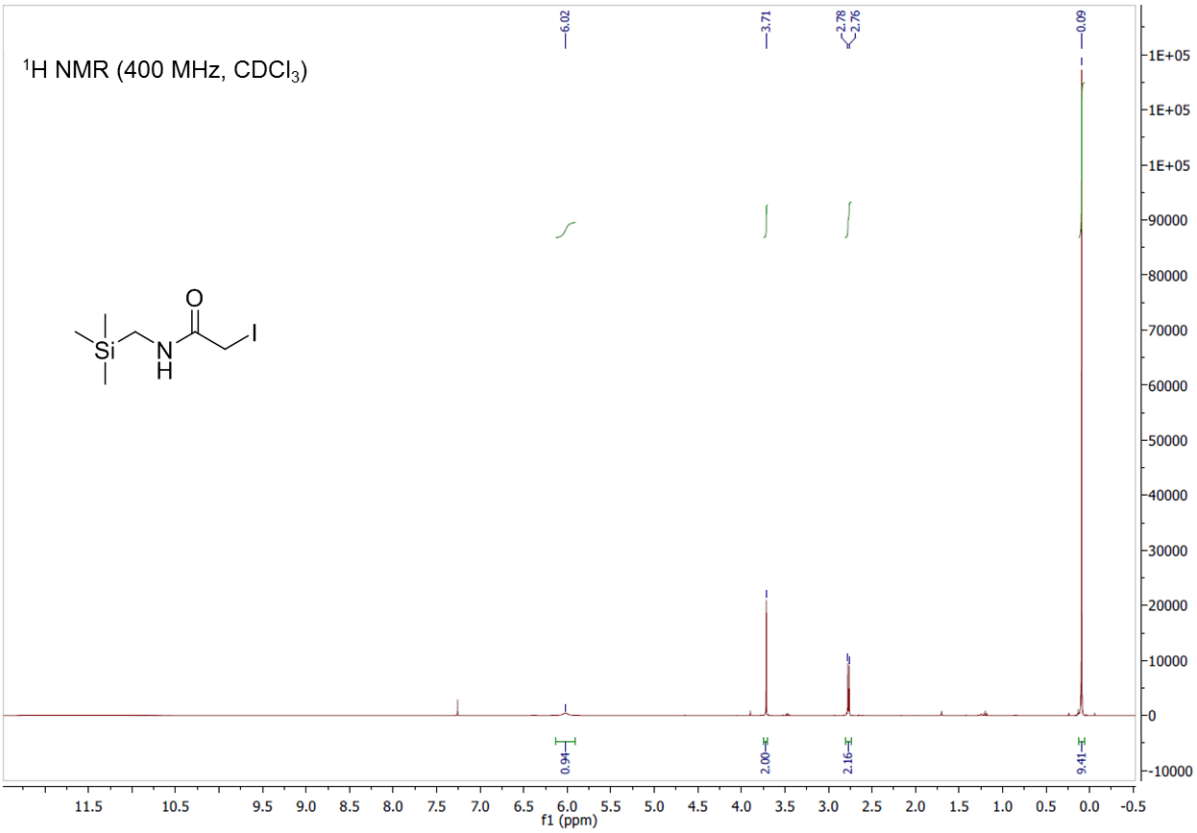


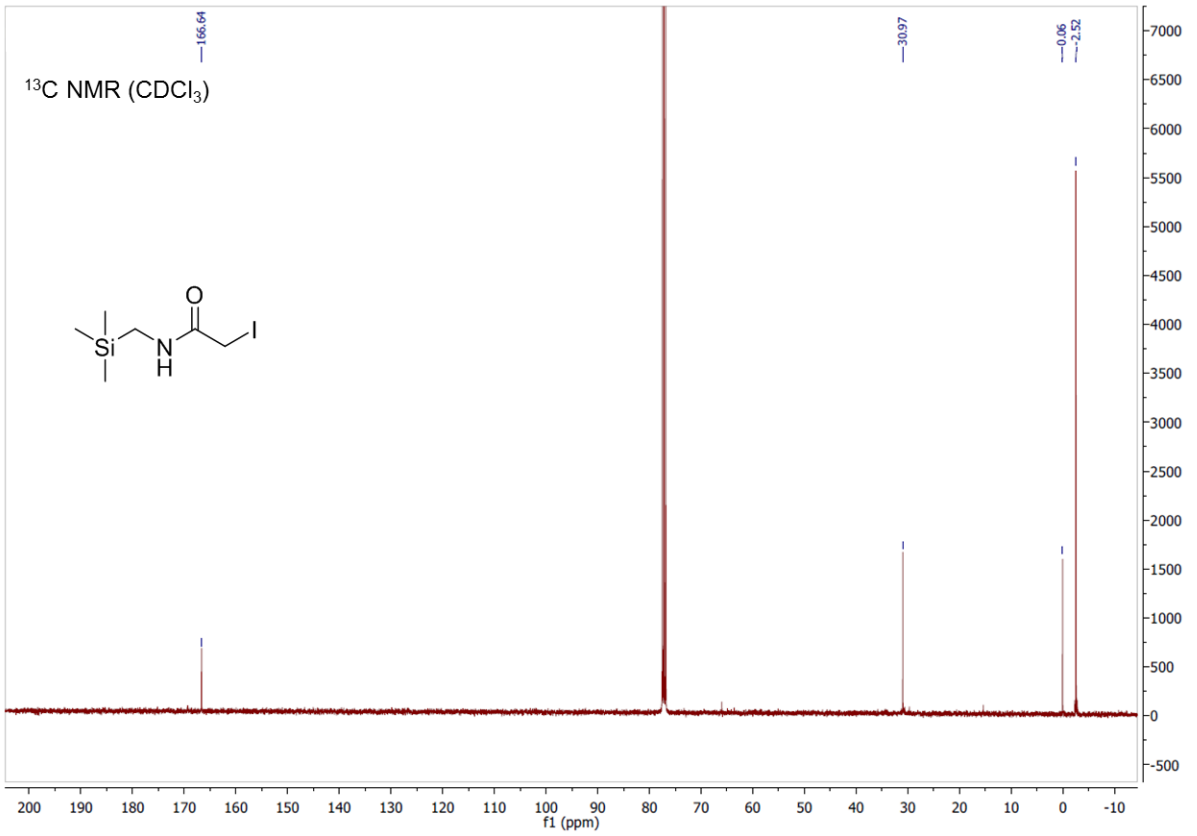
O-(tert-Butyl-$^{13}C_4$)-L-tyrosine formate salt. A solution of 2 M NaOH (3 mL) was added to a solution of N-(9-fluorenylmethyloxycarbonyl)-O-(tert-butyl-$^{13}C_4$)-L-tyrosine methyl ester (198 mg, 415 μmol) in THF (3 mL) and the mixture was stirred at 25 °C for 21 h. THF was removed under reduced pressure and the aqueous mixture was diluted with H_2O (10 mL) and then washed with Et_2O (2 × 10 mL). Residual Et_2O was removed under reduced pressure. The aqueous layer was then subjected to reverse phase silica dry-flash chromatography, eluting with a gradient of 0–30% MeCN in H_2O with 0.1% formic acid to give the title compound as a white solid (129 mg, 98% including 10% w/w excess formic acid as determined by ^{1}H NMR), after lyophilization. $[\alpha]^{27}{}_D$ = -9.1 (*c* = 1.00, H_2O). ^{1}H NMR (400 MHz, D_2O) δ 8.46 (s, 1.7H, formic acid), 7.31–7.25 (m, 2H), 7.13–7.07 (m, 2H), 3.97 (dd, *J* = 7.9, 5.2 Hz, 1H), 3.26 (dd, *J* = 14.6, 5.2 Hz, 1H), 3.10 (dd, *J* = 14.6, 7.9 Hz, 1H), 1.61–1.09 (m, 9H). ^{13}C NMR (D_2O) δ 173.9 (C), 171.0 (CH), 153.0 (C), 131.2 (C), 130.1 (CH), 125.0 (CH), 81.0 (q, $J_{13C\text{-}13C}$ = 40.0 Hz, ^{13}C), 56.0 (CH), 35.7 (CH_2), 27.7 (d, $J_{13C\text{-}13C}$ = 40.0 Hz, $^{13}CH_3$). MS (ES) *m/z*: 242 $[M+H]^+$. LCMS: R_t = 2.84 min; area 98%. HRMS: calcd for $C_9{}^{13}C_4H_{19}NO_3$ + H^+, 242.1572; found (ES, $[M+H]^+$ obsd), 242.1562.

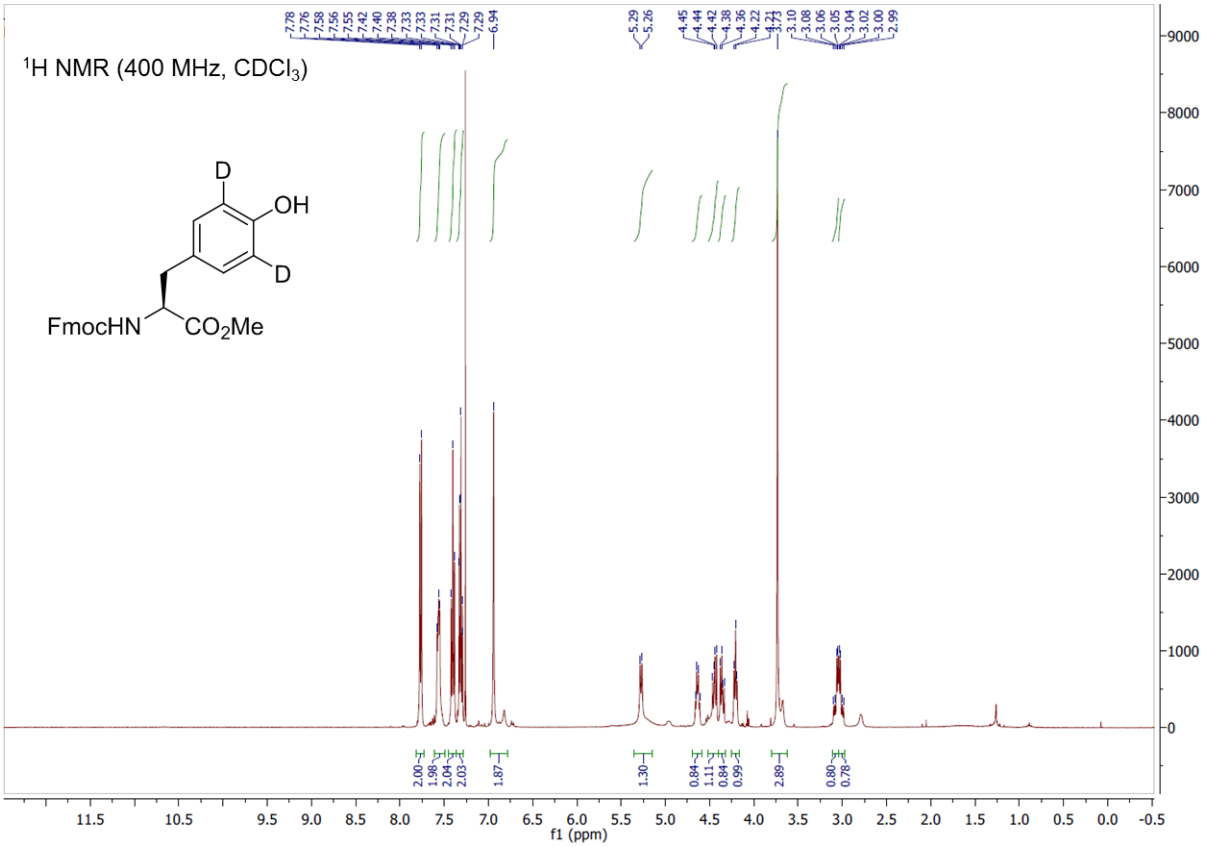
NMR Spectra

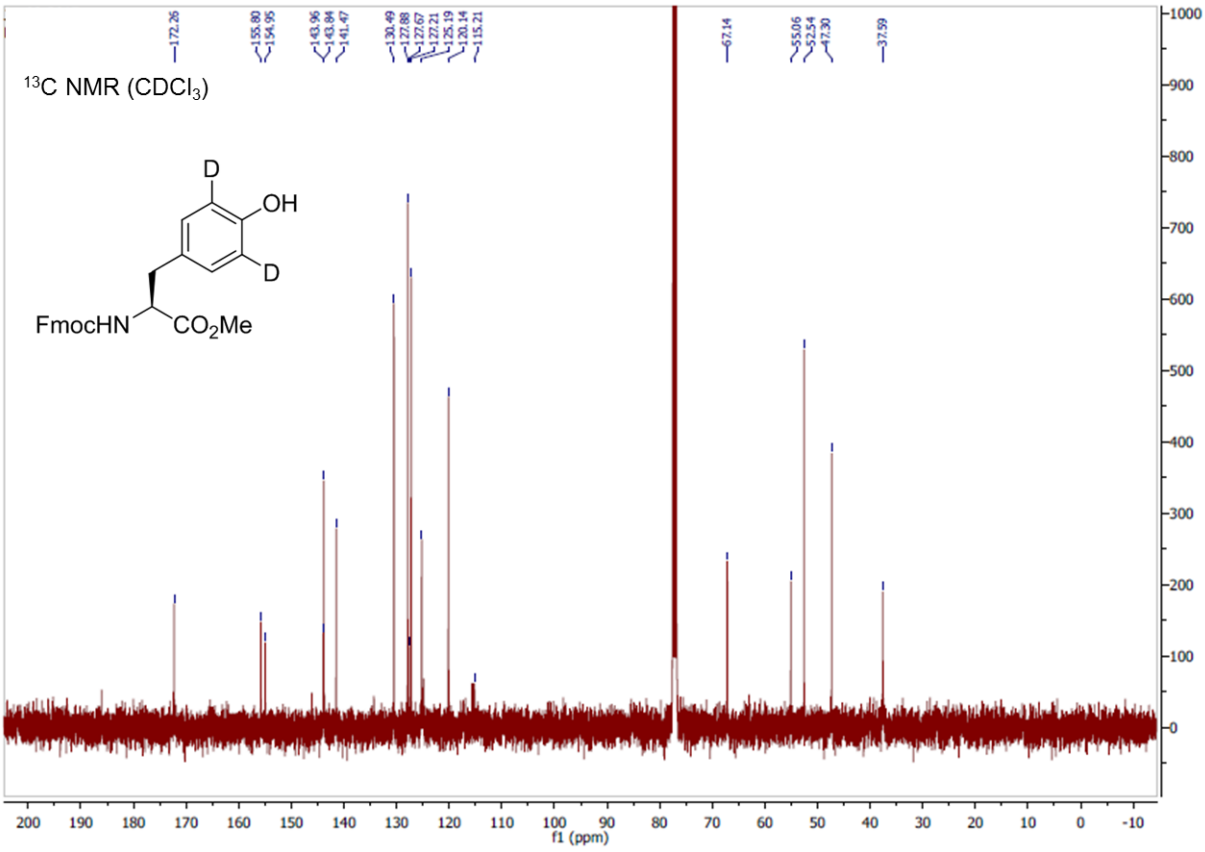

^{1}H NMR (400 MHz, $CDCl_3$)
6.35
3.90
2.80
2.78
0.10
0.87
2.00
2.08
9.26
Si
N
H
O
Br
f1 (ppm)

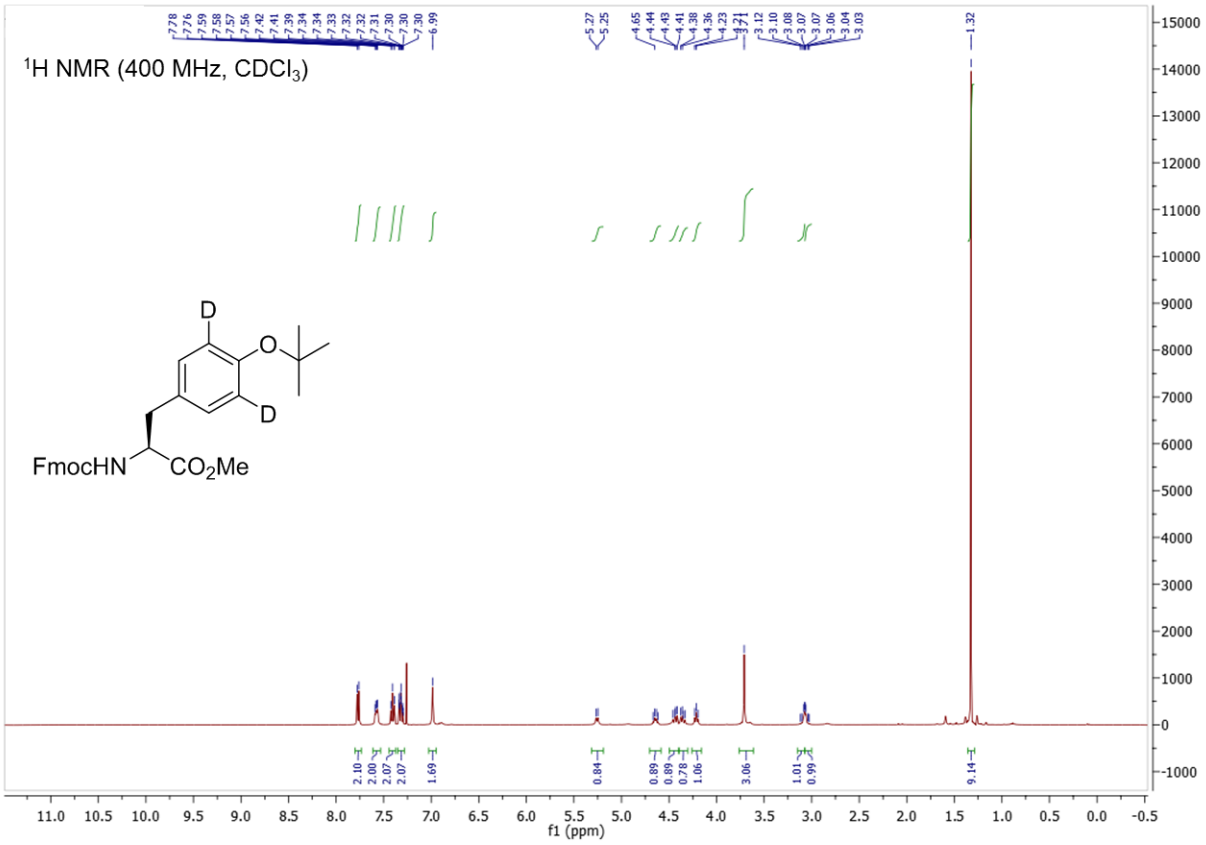

^{13}C NMR ($CDCl_3$)
165.28
30.57
29.74
-2.61
Si
N
H
O
Br
f1 (ppm)

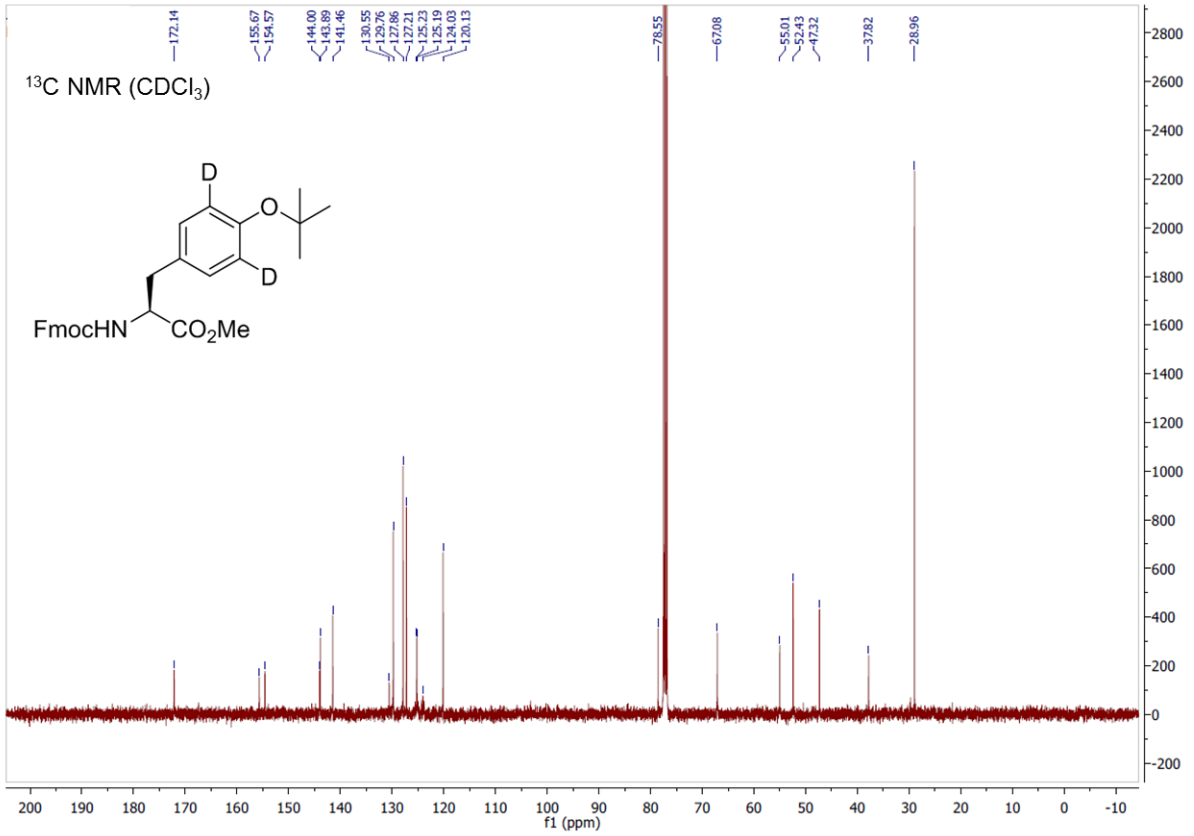

^{1}H NMR (400 MHz, $CDCl_3$)
6.02
3.71
2.78
2.76
0.09
0.94
2.00
2.16
9.41
f1 (ppm)

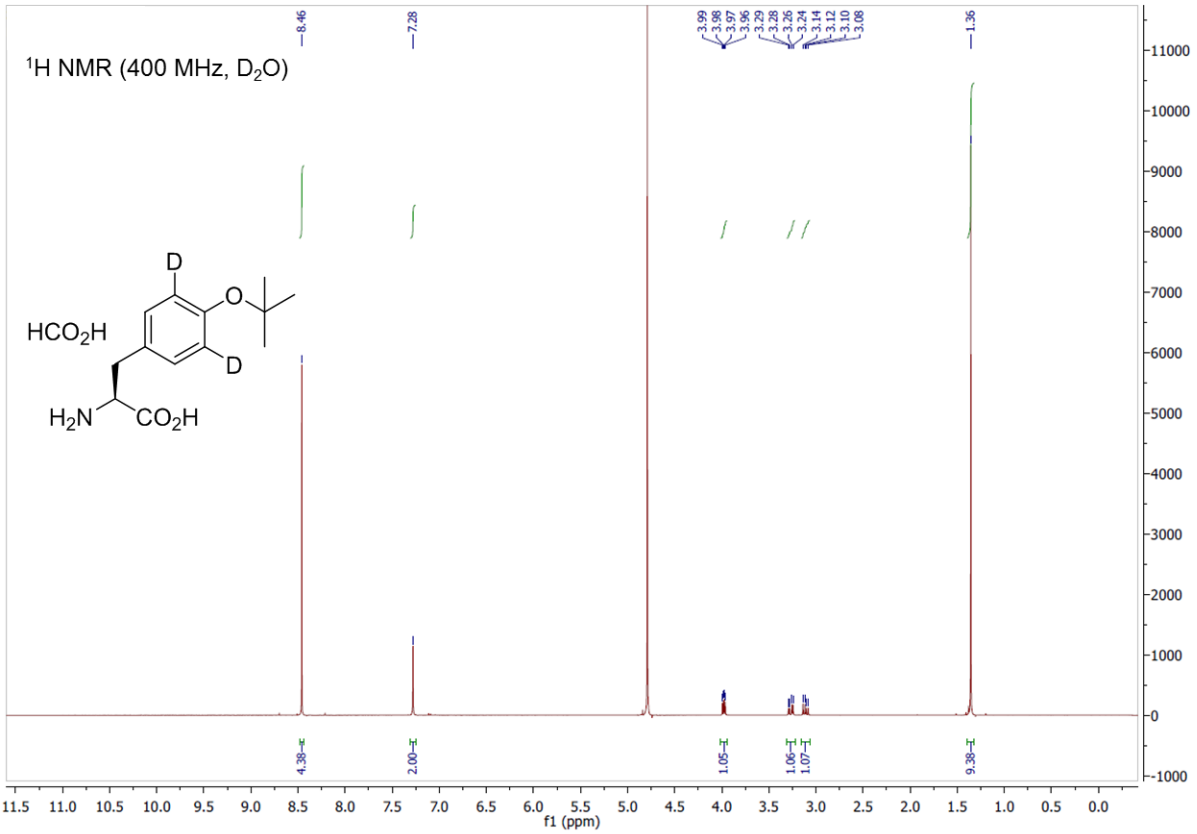

^{13}C NMR ($CDCl_3$)
166.64
30.97
0.06
-2.52
f1 (ppm)

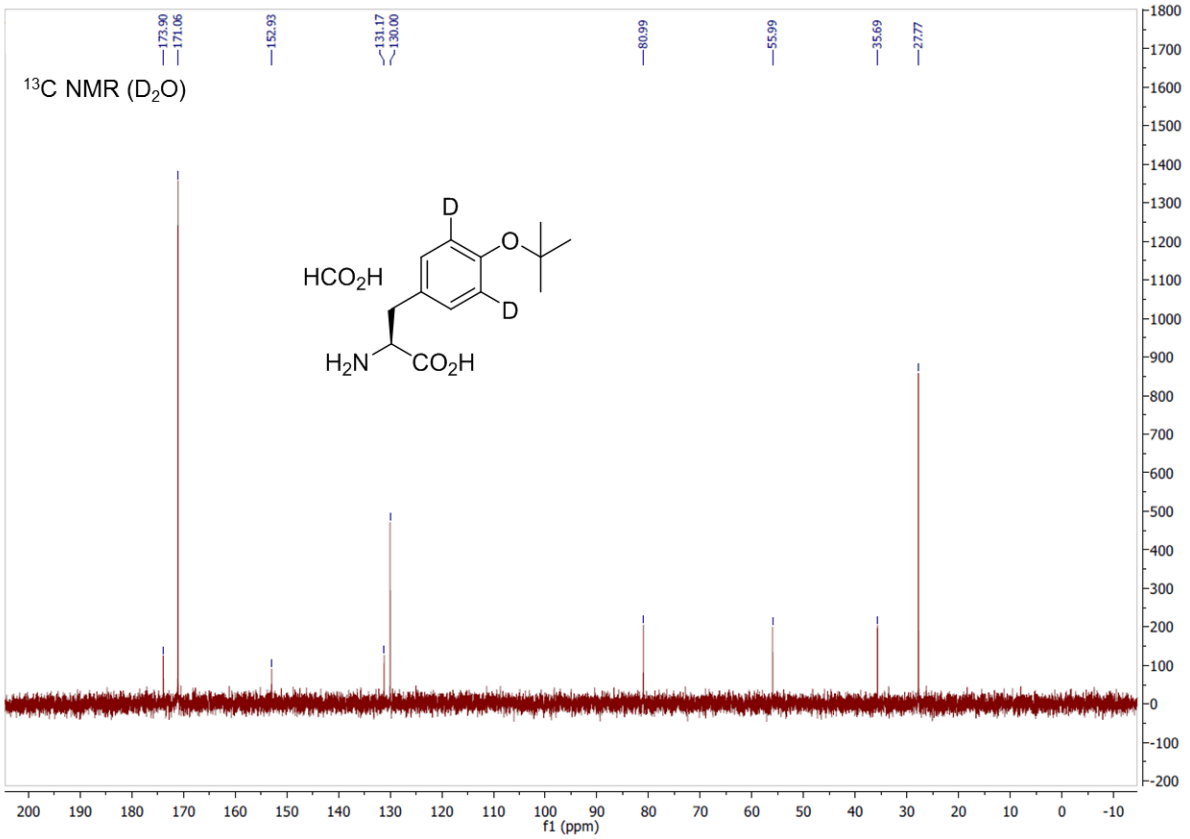

^{1}H NMR (400 MHz, $CDCl_3$)
D
OH
D
FmocHN
CO_2Me
f1 (ppm)

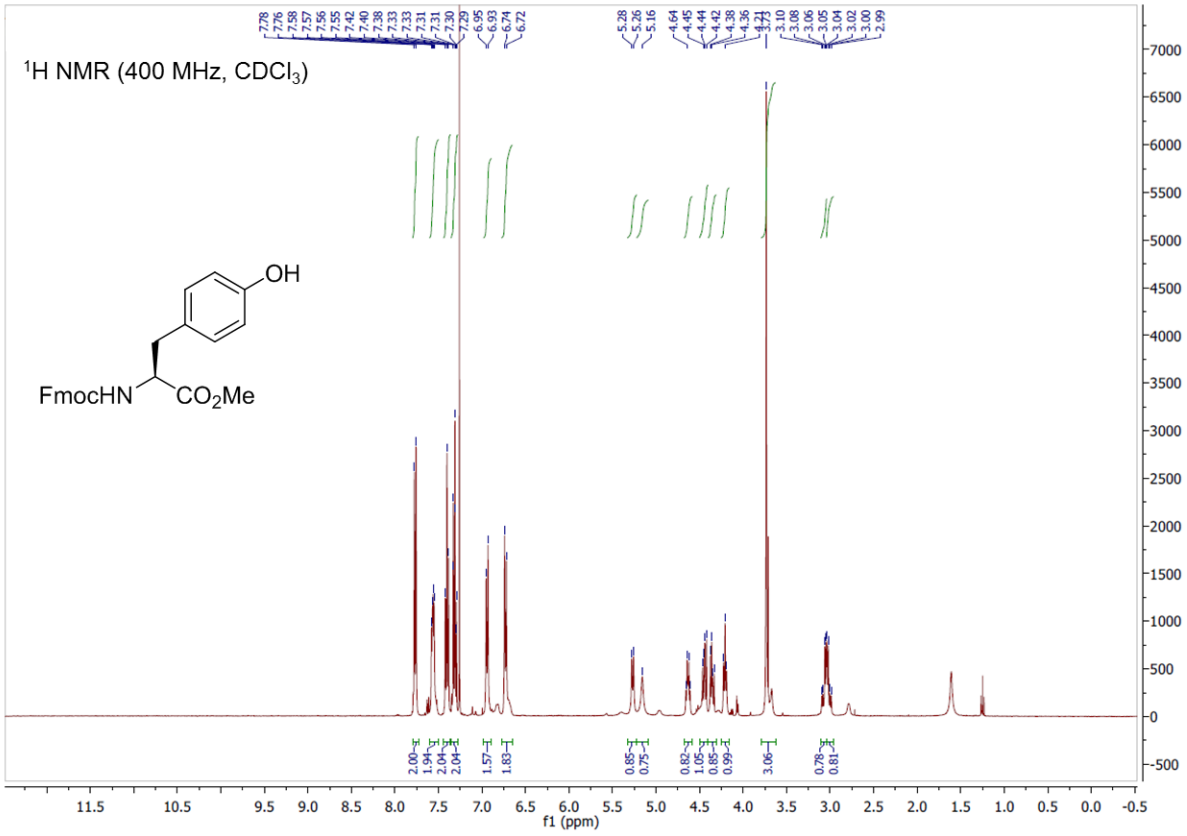

^{13}C NMR ($CDCl_3$)
D
OH
D
FmocHN
CO_2Me
f1 (ppm)

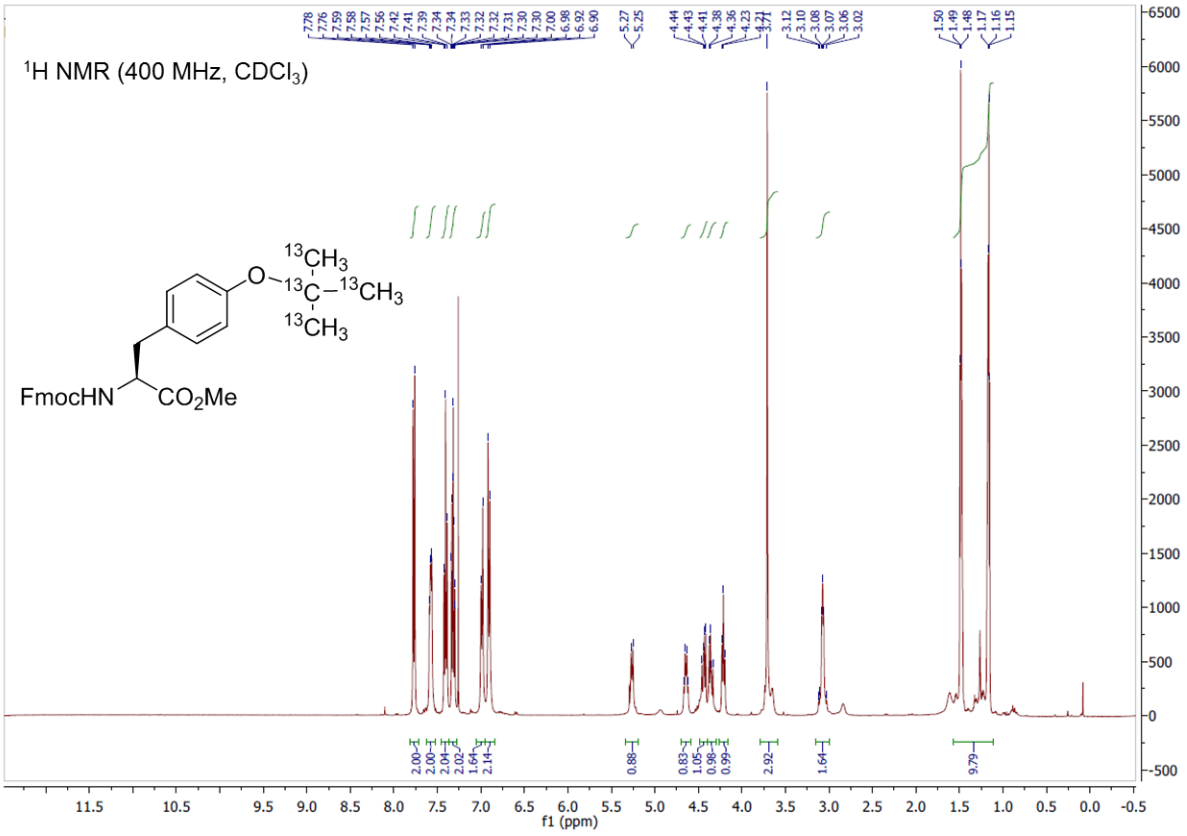

^{1}H NMR (400 MHz, $CDCl_3$)
D
O
D
FmocHN
CO_2Me
f1 (ppm)

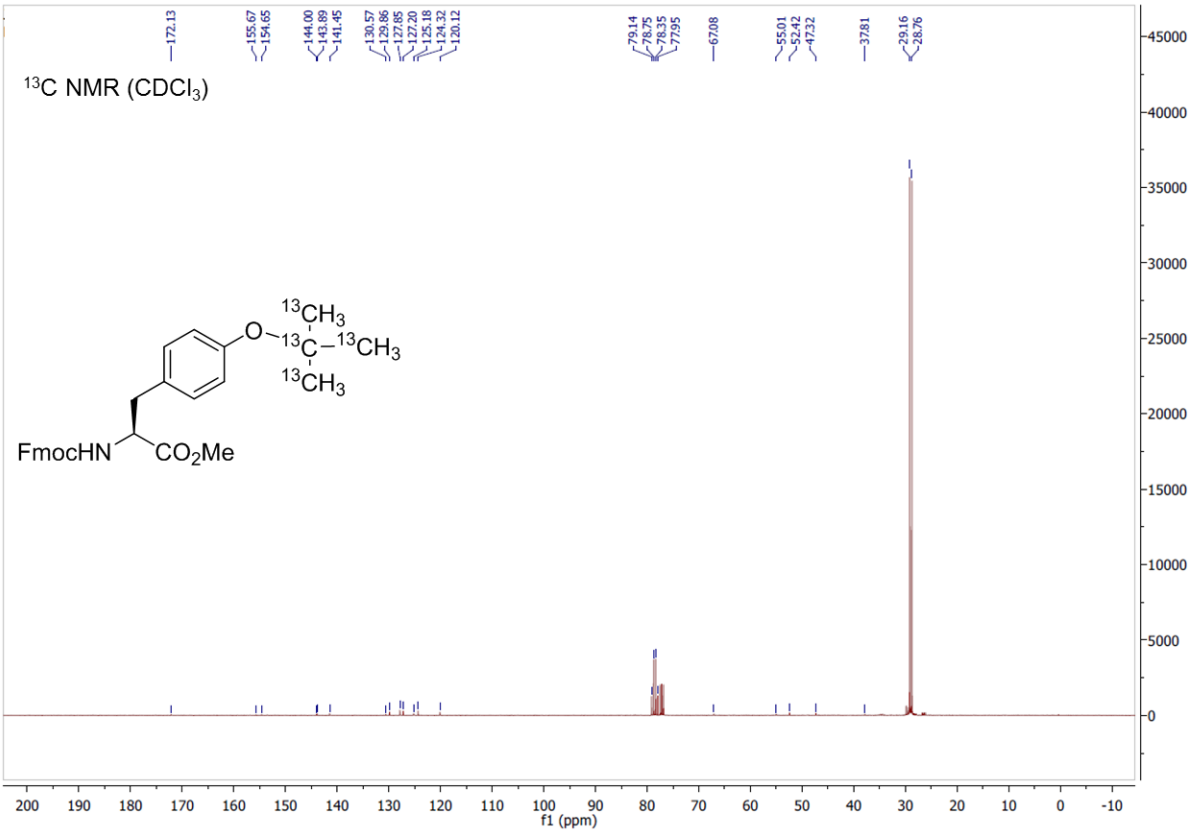

^{13}C NMR ($CDCl_3$)
D
O
D
FmocHN
CO_2Me
f1 (ppm)

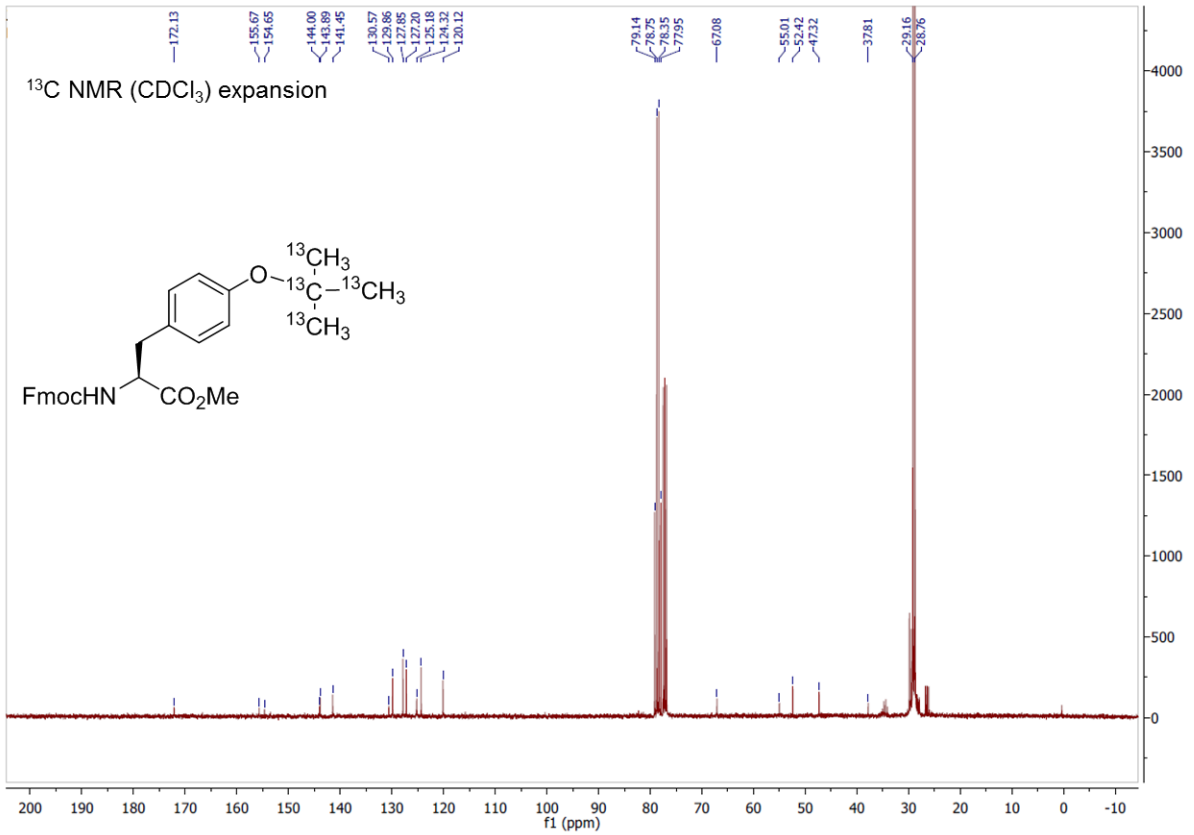

^{1}H NMR (400 MHz, D_2O)
8.46
7.28
3.99
3.98
3.97
3.96
3.29
3.28
3.26
3.24
3.14
3.12
3.10
3.08
1.36
4.38
2.00
1.05
1.06
1.07
9.38
f1 (ppm)

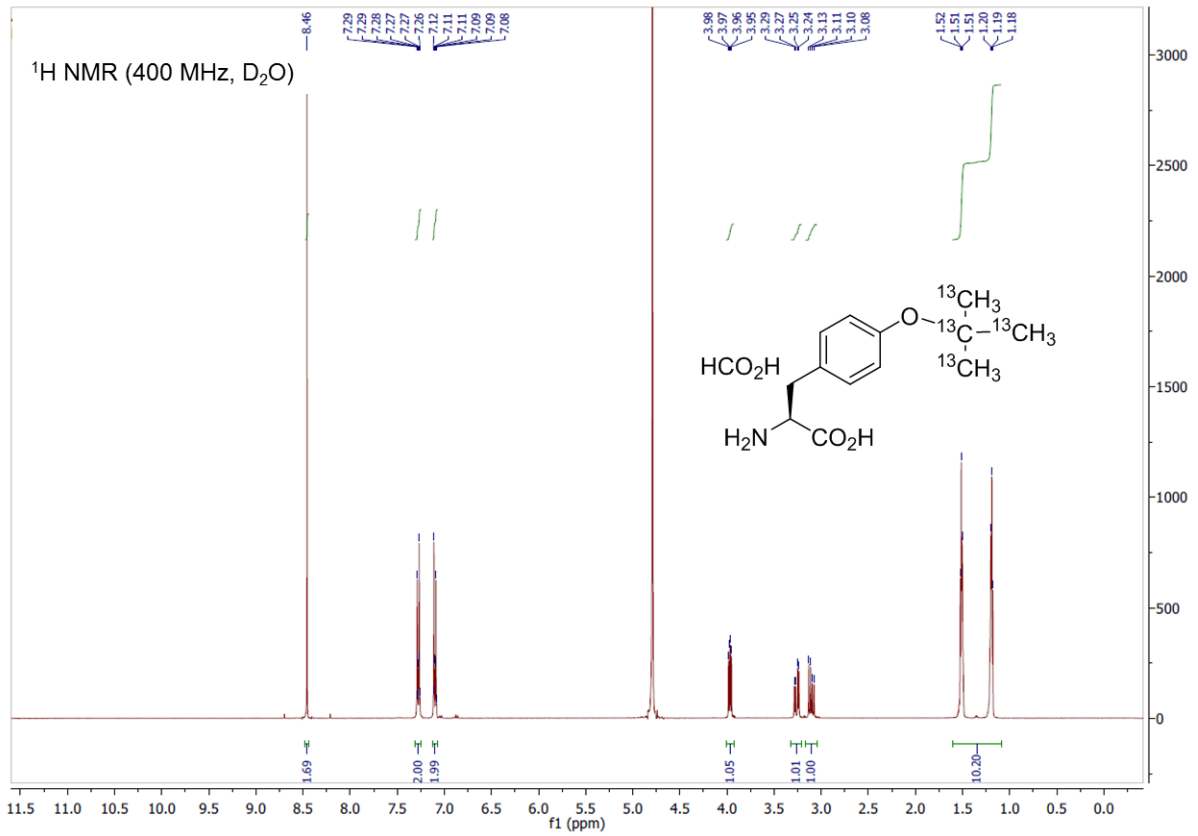

^{13}C NMR (D_2O)
173.90
171.06
152.93
131.17
130.00
80.99
55.99
35.69
27.77
f1 (ppm)

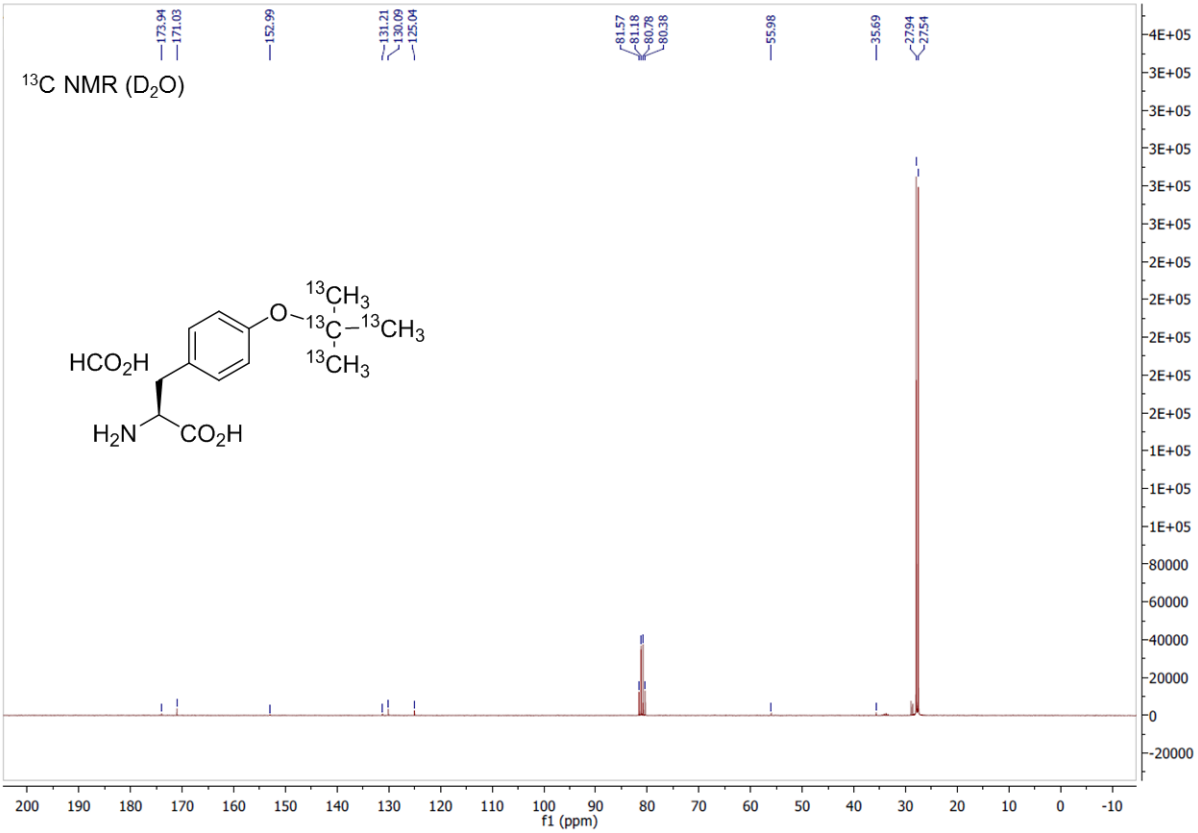

^{1}H NMR (400 MHz, CDCl$_3$)
OH
FmocHN
CO$_2$Me
f1 (ppm)

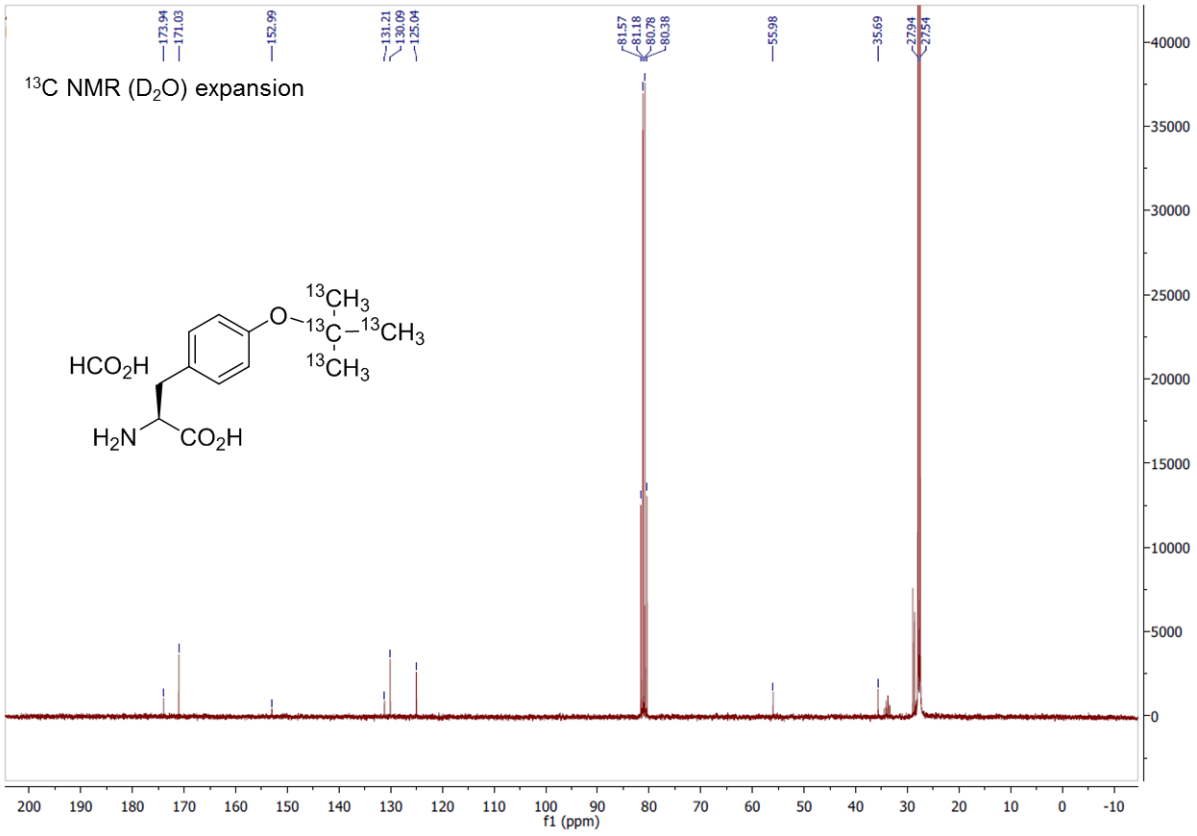

^{1}H NMR (400 MHz, CDCl$_3$)
^{13}CH$_3$
O
^{13}C
^{13}CH$_3$
^{13}CH$_3$
FmocHN
CO$_2$Me
f1 (ppm)


172.13
155.67
154.65
144.00
143.89
141.45
130.57
129.86
127.85
127.20
125.18
124.32
120.12
79.14
78.75
78.35
77.95
67.08
55.01
52.42
47.32
37.81
29.16
28.76
^{13}C NMR ($CDCl_3$)
$^{13}CH_3$
O
^{13}C
$^{13}CH_3$
$^{13}CH_3$
FmocHN
CO_2Me
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


172.13
155.67
154.65
144.00
143.89
141.45
130.57
129.86
127.85
127.20
125.18
124.32
120.12
79.14
78.75
78.35
77.95
67.08
55.01
52.42
47.32
37.81
29.16
28.76
^{13}C NMR ($CDCl_3$) expansion
$^{13}CH_3$
O
^{13}C
$^{13}CH_3$
$^{13}CH_3$
FmocHN
CO_2Me
4000
3500
3000
2500
2000
1500
1000
500
0
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


^{1}H NMR (400 MHz, D_2O)
$^{13}CH_3$
^{13}C
$^{13}CH_3$
$^{13}CH_3$
O
HCO_2H
H_2N
CO_2H
f1 (ppm)


^{13}C NMR (D_2O)
$^{13}CH_3$
^{13}C
$^{13}CH_3$
$^{13}CH_3$
O
HCO_2H
H_2N
CO_2H
f1 (ppm)


173.94
171.03
152.99
131.21
130.09
125.04
81.57
81.18
80.78
80.38
55.98
35.69
27.94
27.54
^{13}C NMR (D_2O) expansion
$^{13}CH_3$
^{13}C
$^{13}CH_3$
$^{13}CH_3$
O
HCO_2H
H_2N
CO_2H
40000
35000
30000
25000
20000
15000
10000
5000
0
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

References

[1] C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love, N. E. Dixon, *Biochemistry* **2000**, *39*, 11989–11999.

[2] D. Shishmarev, Y. Wang, C. E. Mason, X.-C. Su, A. J. Oakley, B. Graham, T. Huber, N. E. Dixon, G. Otting, *Nucleic Acids Res.* **2014**, *42*, 2750–2757.

[3] X.-C. Su, Y. Wang, H. Yagi, D. Shishmarev, C. E. Mason, P. J. Smith, M. Vandevenne, N. E. Dixon, G. Otting, *Biochemistry* **2014**, *53*, 1925–1934.

[4] K. V. Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa, G. Otting, *Angew. Chem. Int. Ed.* **2012**, *51*, 2243–2246.

[5] A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao, P. G. Schultz, *Biochemistry* **2013**, *52*, 1828–1837.

[6] L. M. Harwood, *Aldrichimica Acta* **1985**, *18*, 25.

[7] H. E. Gottlieb, V. Kotlyar, A. Nudelman, *J. Org. Chem.* **1997**, *62*, 7512–7515.

[8] S. W. Wright, D. L. Hageman, A. S. Wright, L. D. McClure, *Tetrahedron Lett.* **1997**, *38*, 7345–7348.

[9] J. G. Adamson, M. A. Blaskovich, H. Groenevelt, G. A. Lajoie, *J. Org. Chem.* **1991**, *56*, 3447–3449.

[10] V. Kumar, N. Yadav, K. P. R. Kartha, *Carbohydr. Res.* **2014**, *397*, 18–26.