NMR studies of metalloproteins

Hongyan Li*, Hongzhe Sun

Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong  E-mail: hylichem@hku.hk

Abstract: Metalloproteins represent a large share of the proteomes, with the intrinsic metal ions providing catalytic, regulatory and structural role critical to protein functions. Structural characterization of metalloproteins and identification of metal coordination features including numbers and types of ligands and metal-ligand geometry, and mapping the structural and dynamic changes upon metal binding are of significance towards understanding biological functions of metalloproteins. NMR spectroscopy has long been used as an invaluable tool for structure and dynamic studies of macromolecules. Here we focus on the application of NMR spectroscopy in characterization of metalloproteins, including structural studies, identification of metal coordination sphere by hetero-/homo-nuclear metal NMR spectroscopy. Paramagnetic NMR as well as ^{13}C directly detected proton-less NMR spectroscopy will also be addressed for paramagnetic metalloproteins. Moreover, these techniques offer great potentials in studies of other non-metal binding macromolecules.

Keywords: ^{13}C direct detection; metalloprotein; metal coordination; NMR spectroscopy; paramagnetic metalloprotein.

Contents

1 Introduction
2 The Contribution of NMR to structural metalloproteins
   2.1 Conventional method for structure determination of metalloproteins
   2.2 Utilization of chemical shifts to generate structures
3 Identification of metal coordination
   3.1 Homonuclear and heteronuclear metal NMR spectroscopy
   3.2 Chemical shift perturbation
4 NMR in studies of paramagnetic metalloproteins

4.1 Paramagnetism-based structural restraints

4.2 $^{13}$C-detected protonless NMR

5 Perspectives

Reference

1 Introduction

Metal ions play important roles in life science and the molecular mechanism of metal-dependent life process and the entirety of metal within a cell or tissue/organ are the topics of emerging field of metallomics and metalloproteomics [1-3]. Metal ions usually must associate with proteins (and other biomolecules), i.e. metalloproteins, to prevent the toxic effects of metal excess. Metalloproteins are one of the most diverse classes of proteins with the intrinsic metal ions providing catalytic, regulatory and structural role critical to protein function, and are found in plants, animals and many microorganisms. It has been estimated that metalloproteins account for approximately one-quarter to one-third of all the proteins in the human body [4]. A recent study revealed that much of microbial metalloproteomes still remains uncharacterized [5]. Systematic bioinformatics survey on 1371 metalloenzymes with known structures showed that about 40% of enzyme-catalyzed reactions involve metal ions e.g. magnesium, zinc and iron [6]. Enormous efforts have been devoted toward understanding the structure and function of the metalloproteins and such knowledge was also used to design a new functional metalloprotein [7] and to rationalize and to search for new metalloproteins by bioinformatics approach [6, 8]. However, it is impossible currently from gene sequence to predict the numbers and types of metal an organism assimilates from its environment or uses in its metalloproteins because the geometry and composition of metal binding site are diverse and poorly-recognized [9, 10]. Therefore, understanding of the function of metalloproteins comes from individually characterization of the structures of the proteins and chemical states of the metal centers by various spectroscopic techniques including NMR spectroscopy, circular dichroism (CD), elec-
tronic absorption spectroscopy (UV), small angle X-ray absorption as well as extended X-ray absorption fine structure (EXAFS).

Over the last three decades, NMR spectroscopy has been developed into a very important and versatile analytical technique both in the chemical and biological sciences. It has been used within the framework of Structural Genomic (SG) projects worldwide for determination of structures of proteins at atomic level under physiologically relevant conditions [11-13]. Moreover, NMR spectroscopy is applicable to study the interactions of proteins with other molecules including proteins, nucleic acids even small molecules which are mainly based on the sensitivity of the chemical shifts towards changes in the chemical environments [14, 15]. Application of this technique to structural studies is limited to small proteins (30-35 KDa) even with the aid of isotopic labeling ($^{13}$C, $^{15}$N and $^2$H) although backbone assignments of a 723-residue enzyme with a molecular weight of 81.4 KDa were achieved [16]. Recently, a few breakthroughs have been made in this field. This includes the first NMR structure of a seven-helix transmembrane protein determined in membrane-mimetic environments [17] and the first structure determined in living cells by in-cell NMR [18]. Apart from its application in structural studies, NMR spectroscopy is also able to monitor the internal motion of biomolecules ranging from subnanoseconds to beyond seconds. Characterization of dynamics of biomolecules, such as folding transition will be great help for our understanding the biological function of the biomolecules.

Application of NMR spectroscopy in studies of metalloproteins in principle is the same as other proteins if proteins contain diamagnetic metal ions. In the case of paramagnetic metalloproteins, things are getting more challenging since paramagnetic metals affect longitudinal and transverse relaxation rates of observed nuclei. However, recent advances in the hardware and methodology have enable structures of such metalloproteins to be determined including $^{13}$C-detected experiments, solid-state NMR and discovery of paramagnetic relaxation enhancement (PRE) [19-21]. In this review, we will address mainly the application of NMR spectroscopy in studies of metalloproteins including contribution of NMR in structural characterization of metalloproteins, in particularly, special attention will be
devoted to utilization of NMR in characterization/identification of the metal binding sites and its coordination environment as well as in probing conformational changes of metalloproteins upon metal binding and release. The techniques used for paramagnetic metalloproteins will be also included briefly since a number of reviews in this field can be found [19-22]. Systematic review on application of metal NMR spectroscopy will not be made and interested readers are directed elsewhere [23].

2 The Contribution of NMR to structural metalloproteins

2.1 Conventional method for structure determination of metalloproteins

NMR spectroscopy can be applied to structurally characterize diamagnetic metal containing metalloproteins similarly as other proteins. Amongst the structures deposited to the Brookhaven Protein Data Bank (PDB), ca. 15% of the structures were resolved by NMR. Here, we will give a brief summary since the detailed methodology can be found in numerous references [11, 24-26].

In structural studies of proteins/metalloproteins, concentrations of about 1 mM are typically required and proteins must be soluble and stable over a period of time (weeks). For small proteins with several tens of amino acids e.g. metallothioneine [27, 28], it is sufficient using $^{15}$N-labeled samples to determine structures of the proteins. However, if proteins can be overexpressed in a bacterial system (e.g. E. coli), it is desirable to overexpress the protein with uniform enrichment of $^{15}$N and $^{13}$C even for small proteins. This will make full use of multidimensional heteronuclear NMR experiments to increase the spectral resolution. Unlike other proteins, expression of metalloproteins in a bacterial system usually requires specific metal ions to be supplemented in the medium to induce overexpression of the targeted proteins [14, 29]. Alternatively, metal ions have to be incorporated into the proteins after purification; in particular if the metal ions play structural roles, otherwise the proteins may not be stable for structural characterization. For example, HypA from Helicobacter pylori precipitates easily in the absence of zinc which serves a structural role [15]. However, caution has to be taken during metal incorporation since excess metal ions may also cause protein aggregation.
A major bottleneck in solving protein structures by NMR is the highly peak-picking and assignment of chemical shifts and NOEs. The strategy of assignment process and structure calculation can be found in an excellent review [30]. In general, for a $^{15}$N/$^{13}$C-labelled protein, a series of double/triple resonance experiments are recorded for resonance assignments. Backbone assignments are derived from HNCA, HNCOCA, CBCACONH, HNCO and HNCCAO whereas side-chain protons and carbon atoms are assigned from HCCH-TOCSY, HCCH-COSY, HBHACONH, C(CO)NH and H(CCO)NH for [24, 25]. The chemical shifts of backbone and side-chain are then used to assign NOEs ($^{15}$N-$^{13}$C-HSQC-NOESY) to derive inter-protein distance restraints. Usually structure determination process goes though several iterations of compiling a NOESY peak list, assignment of NOE cross-peaks to sequence-specific interactions, structure generations and assessment, refinement of NOESY peak lists and reassignment of the cross-peaks, which can be carried out automatically [31, 32]. In addition to distance restraints, dihedral angle restraints usually are obtained from several experiments e.g. HNHA [33], HNHB [34] or predicated from TALOS, a program that empirically predicts backbone angles ($\phi, \Psi$) based on the chemical shifts of H$^\alpha$, C$^\alpha$, C$^\beta$, C$'$ and N [35], as well as the H-bond restraints derived from H-D exchange experiments. For an elongated macromolecules, RDC as an additional restraints is necessary for structure determination. Information about RDC can be found in Chapter 4 of this book. In order to get a relatively good quality of the structures, numbers of the NMR restraints used for structural determination are usually of the order of 10-20 independent interatomic distances per amino acid plus some dihedral restraints and as well as atom-atom vector directions. The quality of calculated structures has to be evaluated using programs PROCHECK, WHATIF etc. and detailed description can be found in a recent review [13].

In addition to general strategies described above, metal-based NMR parameters are also offering great help in evaluation of structures of metalloproteins especially for those metalloproteins that their folding is highly metal-dependent. To incorporate metal cluster constrains into structural calculation, residues that coordinate to metal ions (e.g. Zn$^{2+}$) must be identified first either by mutagenesis studies or
by physical characterization such as UV absorption spectroscopy, EXAFS and NMR spectroscopy [36-38]. Providing that metal coordination residues and geometries are unveiled, metal cluster restraints can be obtained based on relevant crystal structures of either macromolecules or small molecules. Usually, metal cluster restraints are not used in initial structure generation to avoid bias especially during the automated assignment procedure in CYANA [39], instead, metal ligand restraints are incorporated in the last step of structural calculation as well as in the refinement stage[40, 41].

The application of NMR spectroscopy in protein structure determination actually started with a small metalloprotein, metallothoinine (MT) [42]. Metallothioines are a class of low molecular weight (typically 6-7 kDa) cysteine-rich proteins. The proteins are lack of well-defined secondary structure and their fold is dictated mostly by a clustered network of the cysteine residues and the metal ions usually represented by Zn^{2+}, Cu^{+} and Cd^{2+} [43, 44]. Since the first solution structure of rabbit liver Cd_{7}MT2 [42], numerous three dimensional structures of metallothiones from different isoforms (MT1/MT2/MT3) or different species such as blue crab and mammalian (rabbit, rat and human) have been resolved by NMR spectroscopy [45] with only one structure (rat liver Cd_{5}Zn_{2}MT2) determined by X-ray crystallography [46]. The protein consists two dynamic metal-thiolate clusters and folded into two domains (α, β) and structural mobility of the protein has made it difficult to be crystallized. The metal cluster restraints e.g. Cd-S bond lengths, as well as Cd-S-Cd, S-Cd-S and CysC⁶-S-Cd bond angles from the X-ray crystal structures of model cadmium complexes and rat liver Cd_{5}Zn_{2}MT2 were often incorporated with other distance and angle restraints in structure calculation. Recently, a new member of metallothionine MT3 with the conserved CPCP motif in the N-termini is involved in the growth inhibitory activity and is down-regulated in the brain of Alzheimer’s patients [47]. The solution structures of both human [27], and mouse MT3 [48] resolved by NMR spectroscopy for the C-terminal α-domain, Fig. 8.1a, revealed a similar Cd₄Cys₁₁ cluster as well as a very similar tertiary folds to MT1/2. However, a loop in the acidic hexapeptide insertion is found and is slightly longer in human MT3 than in mouse MT3. The first
solution structure of Cd₇MT-nc of the Antarctic fish *Notothenia coriiceps* was also determined [28]. The position of the 9th cysteine of Cd₇MT-nc is different from mammalian MT which results in a structural change of the domain, in particular in the orientation of the loop (Lys50-Thr53), Fig. 8.1b, and in turn to a different charge distribution with respect to mammalian MT [28]. Interestingly, an intrigue class of histidine-containing metallothionines have also been identified in fungi and bacterial [49]. The histidine residue has been thought to be able to modulate zinc affinity and reactivity. Solution structure of one of this class of MTs, Zn₄SmtA from cyanobacterium *Synechococcus* PCC 7942 was determined [41], Fig. 8.1c, revealing a Zn₄Cys₄His₂ cluster with a topology similar to that of the Zn₄Cys₁₁ cluster of the α-domain of mammalian MT. However, the two ZnCysHis sites and one of the ZnCys₄ site readily exchange Zn²⁺ for exogenous Cd²⁺. Moreover, SmtA contains a short α-helix and two small antiparallel β-sheets surrounded the inert zinc site, which resemble zinc finger portions of GATA and LIM proteins. Such a structure of SmtA probably renders its function of specific protein and/or DNA recognition [41].

NMR spectroscopy has enormous contribution to structural biology of metalloproteins, particularly in zinc-binding proteins. Zinc, the second most abundant metal found in eukaryotic organisms, plays important catalytic and structural roles in variety biological processes. Binding of zinc is able to stabilize the folded conformations of domains, which renders a proper function of the protein [50]. Zinc finger proteins are the most abundant class of zinc-binding proteins that contain conserved cysteines and histidines coordinated to zinc. Intensive structural and functional studies have established the invariance of the ββα framework of the Cys₄His₂ zinc finger module and provided sound basis for understanding the nature of DNA recognition [40, 51-55]. Diverse structures of zinc fingers also account for their diverse functions such as DNA recognition, RNA packaging, transcriptional activation, regulation of apoptosis, protein folding and assembly as well as lipid binding [52, 56, 57]. Comprehensive reviews in this area can be found [56-58]. Here, we will only highlight some of the recent studies. Solution structure of Gfi-1 zinc finger 3-5 complex with a 16-mer consensus DNA
Fig. 8.1 (a) Overlaid view of structures of human MT3 (PDB: 2f5h) in orange and mouse MT3 (PDB: 1J19) in pale green with the cadmium ions shown as spheres (left), and the four metal-thiolate cluster of human MT3 in the α-domain with the metal ions and sulfur atoms from cysteines shown in cyan and yellow respectively (right); (b) NMR structures of the *Notothenia coriiceps* Cd:MT-nc with the α-domain (PDB: 1m0g) shown in the left and the β-domain (PDB: 1m0j) shown in the right. The metal thiolate clusters are also shown with the cadmium
ions shown as spheres in skyblue and sulfurs as sticks in yellow; (c) Solution structure of bacterial SmtA (PDB: 1jjd) with the zinc ions shown in pale cyan as spheres (left) and the metal cluster $\text{Zn}_4\text{Cys}_9\text{His}_2$ (right) with the zinc ions shown in pale cyan, sulfurs in yellow and the nitrogens of histidines in blue.

(Fig. 8.2a). Solution structure of Gfi-1 zinc finger 3-5 complex with a 16-mer consensus DNA (Fig. 8.2a) demonstrated zinc fingers 3-5 bind into the major groove of the target DNA reminiscent of canonical Cys$_2$His$_2$ zinc-finger domains, which provide valuable insight into the structure determinants for DNA binding specificity as well as molecular rationales for a naturally occurring mutation that causes acute myeloid leukemia [59]. Poly(ADP-ribose) (PAR) is an important post-translational modification in higher eukaryotes. Solution structures of two PBZ modules (PAR-binding zinc finger) of PNK-like factor (APLF) and the PDB domain of Drosophila melanogaster CG1218-PA reveal a novel type of Cys$_2$His$_2$ zinc finger and provide structural basis for PBZ-PAR recognition. Intriguingly, Cys$_3$His$_1$ zinc coordination of the PBZ modules is structurally and functionally dissimilar from canonical double stranded DNA-binding TFIIIA-type zinc finger, rather they resemble single-stranded RNA-binding Cys$_3$His$_1$ tandem zinc fingers (TZFs). Both of them are lack of secondary structures but have rigid backbone conformations as a result of zinc binding [60, 61]. Zinc finger proteins are also able to bind to RNA. The NMR structure of tandem zinc finger (TZF) domain of the protein TIS11d bound to the RNA sequence 5′-UUAUUUAUU-3′ (Fig. 8.2b) reveals a pair of novel Cys$_3$His$_1$ fingers which independently recognizes the four nucleotide sequence UAUU and the sequence specificity in RNA recognition is achieved by a network of intermolecular hydrogen bonds [62]. This structure provides insights into RNA-binding function of this family of Cys$_3$His$_1$ zinc finger proteins [62]. The Cys$_3$His$_1$ zinc finger motif is also found in the structure of SAP30 polypeptide of the Sin3 corepressor complex which adopts a novel fold comprising two $\beta$-strands and two $\alpha$-helices with the zinc organized center. Such a structure may also function as a double-stranded DNA-binding motif [63]. The zinc finger CW (zf-CW) domain with a motif about 60 residues is frequently
found in proteins involved in epigenetic regulation. Interestingly, NMR structure of human zf-CW domain and PWWP domain containing proteins reveal a new fold in which a zinc is coordinated tetrahedrally by four conserved Cys residues [64]. Such a structure partially resembles the plant homeo domain (PHD) finger bound to the histone tail, implicating a similar function of zf-CW domain [64]. This kind of Cys$_4$ motif is widely found in other metalloproteins such as [NiFe] hydrogenases accessory protein HypA. Solution structure of HypA from Helicobacter pylori (Fig. 8.2c) showed that zinc coordinated to four cysteines donated from loops and no apparent secondary structure found in the zinc-domain [15]. The X-ray structure of HypA from Thermococcus kodakaraensis KOD1 further confirmed such a zinc coordination sphere [65].

Fig. 8.2 (a) Ribbon diagram of the Gfizf35-DNA complex (PDB: 2kmk) with the zinc ions shown as gray spheres and the side-chains of two coordinated histidines and two cysteines shown in sticks and DNA shown in yellow sticks; (b) Solution structure of the RNA complex of TIS11d (PDB: 1rgo) with the zinc ions gray spheres coordinated to three cysteines and one histidine and RNA shown as yellow sticks; (c) Solution structure of HypA from Helicobacter pylori (PDB: 2kdx) with the zinc ions in gray sphere coordinated to four cysteine sulfurs.
### 2.2 Utilization of chemical shifts to generate structures

Protein NMR chemical shifts are highly sensitive to local structure and reflect a wide array of structure factors including backbone and side-chain conformation, secondary structure, hydrogen bonds, and the orientation/position of aromatic rings. Chemical shift data can be used in conjunction with protein sequence information and reasonable force field to generate 3D structure models using the method of CHEMSHIRE or CS-ROSETTA [66-68]. The Chemical-Shift-ROSETTA (CS-ROSETTA) is a robust protocol available for de novo protein structure generation. The method uses experimental chemical shifts of $^{13}$C$^{\alpha}$, $^{13}$C$^\beta$, $^{13}$C', $^{15}$N, $^1$H$^{\alpha}$, and $^1$H$^N$ as an input to select polypeptide fragments in existing protein structures (e.g. PDB data bank) in conjunction with the standard ROSETTA Monte Carlo fragment assembly and energy minimization protocol [67, 68]. The CS-ROSETTA has been further combined with CYANA using unassigned NOESY data to direct Rosetta trajectories toward the native structure and produces a more accurate models than CS-ROSEAAR alone [69]. Moreover, chemical shifts has been further extended in determination of protein-protein complex structures via CamDock method [70]. The method that utilizes chemical shifts to generate structures may provide potentially a new direction for high-throughput NMR structure determination of proteins including metalloproteins although such a method has not yet been applied in metalloproteins so far.

### 3 Identification of metal coordination

#### 3.1 Homonuclear and heteronuclear metal NMR spectroscopy

Metalloprotein functionality depends on subtle interaction between properties of the metal ion, dictated by its coordination chemistry. Our present knowledge in terms of structure-function of metalloproteins in particular the role of metal ions involved varies considerably from proteins to proteins. Therefore, identification of metal coordination parameters including numbers and types of ligands and metal-ligand geometry and mapping the structural and dynamic changes upon metal binding are of significance towards understanding biological functions of metalloproteins. Today, NMR spectroscopy is one of the leading techniques for this pur-
pose. Applicability of homonuclear metal NMR and heteronuclear $^1$H-metal HMQC to directly monitor protein-metal interactions highly rely on the properties of the nuclei. Some of the metal (e.g. $^{113}$Cd) NMR has been used extensively to identify coordination sphere of the metal ions. Moreover, the coupling constants between the NMR active metals and nuclei of the protein provide insight into the identity and geometry of the metal ligands [71]. Many metal (e.g. $^{43}$Ca, and $^{67}$Zn) NMR are less powerful and hardly used owing to the fact that these nuclei have the spin quantum number $I$ greater than $\frac{1}{2}$ which led to lower sensitivity and poor resolution and broadening due to large quadrupolar moment although ultrahigh field improve it. Several reviews have systematically summarized the application of heteronuclear NMR spectroscopy in biological and medicinal chemistry as well as in study of metalloproteins [23, 71-73]. Here, we will highlight some of the recent progresses as a snapshot of using metal NMR to identify metal coordination.

Cadmium is one of the most widely used metal nuclei for probing metal-protein interactions, despite its toxic properties. It has two NMR active nuclei $^{113}$Cd and $^{111}$Cd (spins of $\frac{1}{2}$) with the former being slightly more sensitive and therefore usually used as a preferred nucleus. At natural abundance, the sensitivity of $^{113}$Cd is very low (ca 7.6 folds of $^{13}$C), therefore isotopic enrichment (ca 96%) of $^{113}$Cd is usually needed to ensure reasonable quality of spectra to be acquired in a relatively short period of time (few hours for ca 0.5 mM samples). $^{113}$Cd or $^{113}$Cd-1H NMR spectroscopy has been utilized in the study of a variety of metalloproteins where the native Zn$^{2+}$, Ca$^{2+}$, Mg$^{2+}$, Mn$^{2+}$, Fe$^{2+}$ and Cu$^{2+}$ can be substituted by $^{113}$Cd given that the adaptable ligand coordination number and geometry of Cd$^{2+}$ is similar to Zn$^{2+}$ and ionic radius of Cd$^{2+}$ (0.97 Å) is similar to that of Ca$^{2+}$ (0.9 Å) [27, 74-77]. Moreover, the substitution of the native zinc from metalloenzymes and DNA-binding proteins by cadmium caused almost no changes in their structures and functions [78, 79].
Fig. 8.3 Chemical shifts of $^{113}$Cd for structurally characterized $^{113}$Cd-substituted metalloproteins relative to external 0.1 M [Cd(ClO$_4$)]. The chemical shift positions are represented by gray bars and coordinating atoms are highlighted in blue with typical proteins listed besides. Here S represents sulfur from cysteine, S* represents sulfur from methionine, O represents oxygen from carboxylate or water and N represents nitrogen from histidine [71].

$^{113}$Cd chemical shifts are very sensitive to the nature, number, and geometric arrangement of the coordinated ligands [71] as shown in Fig. 8.3. Such wide chemical shift dispersion not only provides information about the types and numbers of ligand at a particular metal site, but also discriminates multiple metal sites with identical ligand coordination environments. $^{113}$Cd NMR and $^1$H-$^{113}$Cd HMQC has been employed exclusively in identification of metal-thiolate clusters in a family of small proteins e.g. metallothionines [27, 48, 71, 80]. Both homonuclear 1D $^{113}$Cd decoupling studies (Fig. 8.4a) and 2D $^{113}$Cd-$^{113}$Cd COSY (Fig. 8.4b) of $^{113}$Cd-MTs established the existence of two metal thiolate clusters in this protein, while $^1$H-$^{113}$Cd HMQC (Fig. 8.4c) was used to identify sequence-specific cyste-
ine-cadmium coordination bonds. The chemical shift patterns for the two clusters Cd₃Cys₉ and Cd₄Cys₁₂ of human MT₃ as shown in Fig. 8.4a showed seven resonances at analogous position compared with MT₁/₂ with chemical shift ranging from 600 to 690 ppm [27]. However, the resonances from the α-domain (I, V, VI and VII) are slightly shifted in particular for resonance VII, probably attributed to the hexa-peptide insertion in this domain. The homonuclear 2D ¹¹³Cd-¹¹³Cd, Fig. 8.4b, clearly shows the correlation of cadmium signals which confirms the existence of two Cd-thiolate clusters [81]. The 2D ¹H-¹¹³Cd HMQC as shown in Fig. 8.4c is normally used to obtain detailed metal-thiolate connectivity within each of these clusters [27]. Recently, metallothionines from different species such as sea mussel *Mytilus galloprovincialis* (Cd₇MT₁₀) [81], and blue crab *Callinectes sapidus* (MTc) [82] have been studied by combined use of ¹¹³Cd NMR, ¹¹³Cd-¹¹³Cd COSY and ¹H-¹¹³Cd HMQC. Both proteins have 21 cysteine residues with position of cysteines distinct from mammalian MTs. The unique structure and dynamic features of the metal-thiolate cluster in these proteins are obviously seen from their distinct NMR parameters of Cd₇MTC [81, 82]. ¹¹¹Cd NMR has also been applied in studies of His-containing metallothionine, e.g. SmtA. The metal cluster of CdS₄ and CdN₂S₂ were identified [83], and ¹H-¹¹¹Cd HMQC of Cd₇SmtA unequivocally demonstrated couplings of two Cd²⁺ to both Hε₁ or Hδ₂ protons of two histidine residues [41]. Apart from metallothionines, cadmium NMR has also been employed to identify metal coordination environments in various metalloproteins, such as zinc finger [40] and [NiFe] hydrogenease accessory protein HypA [15], where in the latter case zinc ions (substituted by ¹¹³Cd) are coordinated to four cysteine side-chins tetrahedrally. Moreover, it was also used to investigate major zinc binding site on human albumin [74, 84]. The chemical shifts of ¹¹¹Cd of human albumin (ca. 130 and 30 ppm) in combination with other techniques (EXAFS and mutagenesis studies) clearly demonstrated that the major zinc is five-coordinate site with residues of histidine and aspartate.
Fig. 8.4 (a) The 500 MHz NMR spectrum of human $^{113}\text{Cd}-\text{MT3}$ in 15 mM phosphate buffer, pH 7.3 and 310 K. Resonances I, V, VI and VII are assigned to the α-domain and II, III and IV to the β-domain [27]. (b) 2D $^{113}\text{Cd}-^{113}\text{Cd}$ COSY NMR of Cd$_{10}$-MT10 in 17 mM Tris-d$_{11}$ buffer, pH 7.0, 298 K with the one dimensional proton-decoupled $^{113}\text{Cd}$ NMR spectrum at the top. The cadmium connectivity in the α- and β-domains is also shown [81]. (c) The 500 MHz two-dimensional $^1\text{H}-^{113}\text{Cd}$ HMQC spectrum of human $^{113}\text{Cd}-\text{MT3}$ in 15 mM phosphate buffer with $^3J(\text{H},^{113}\text{Cd})$ of 30 Hz. The assignments were also labeled with one-letter amino acid code [27].

Similarly, heteronuclear magnetic resonance spectroscopy with nuclei of spins $\frac{1}{2}$ (e.g. $^{109}\text{Ag}$/$^{107}\text{Ag}$, $^{207}\text{Pb}$ and $^{199}\text{Hg}$) has also been extensively employed in studies of metal coordination environments or the active site structures of metalloproteins. $^{109}\text{Ag}$/$^{107}\text{Ag}$ NMR has so far found limited application in biological systems due to low sensitivity and extremely long spin-lattice relaxation times. The only known metalloprotein studied by 2D $^1\text{H}-^{109}\text{Ag}$ HMQC spectroscopy to date is the silver-substituted yeast metallothionine [85]. The toxic lead ($^{207}\text{Pb}$), although it is not directly biological relevant, is an excellent substitute for Ca$^{2+}$ in calcium-binding allowing them to retain similar structures and function. $^{207}\text{Pb}$ has a moderate resonance frequency, a vast chemical shift range (over thousands) and potentially large spin-spin coupling to neighboring nuclei which limits its application in study of metalloproteins. Although $^{207}\text{Pb}$ has been used to study active sites of model compounds or peptides of metalloenzymes [86, 87], there appears only one application using $^{207}\text{Pb}$ as a probe to study Pb$^{2+}$ binding to the Ca$^{2+}$ site of calcium-binding proteins including calmodulin (CaM) [88]. Binding of $^{207}\text{Pb}$ to both carp and pike parvalbumins gave rise to two $^{207}\text{Pb}$ signals from 750 to 1260 ppm downfield relative to aqueous [Pb(NO$_3$)$_2$]. Similarly, four $^{207}\text{Pb}$ signals, which fall in the same chemical shift window, could be observed for CaM. Both $^{207}\text{Pb}$ and $^1\text{H}$ have demonstrated that Pb$^{2+}$ binds to all four sites simultaneously, in contrast to the behavior of the protein in the presence of Ca$^{2+}$ [88]. The large chemical shift dispersion and remarkable sensitivity to chemical environment of $^{207}\text{Pb}$ signals...
promote extensive studies using model complexes or peptides on mining the relationship of chemical shifts of $^{207}$Pb with the numbers and types of coordinating ligands [86, 87, 89]. A recent study again demonstrated that chemical shifts of $^{207}$Pb can be used to discriminate of PbS$_3$ (from Cys of thiolate-rich peptides) with other ligands such as PbS$_3$O and other O, S, and N donor ligands [89]. Therefore, based on $^{207}$Pb chemical shift maps, Fig. 8.5a, both homonuclear and heteronuclear $^{207}$Pb spectroscopy should provide a useful tool for investigation of Pb$^{2+}$ coordination in more complex biological systems.

In spite of widely-known as protein-modifying agents, the toxic metal Hg$^{2+}$ ion can be used as a probe to substitute the native metal ions for several metalloenzymes in a manner that preserves catalytic activity. Mercury has two NMR-active isotopes, $^{199}$Hg and $^{201}$Hg, with the latter having the spin of $\frac{1}{2}$ and a natural abundance of 16.84%. $^{199}$Hg exhibits several favorable NMR properties for structural and functional study, such as large coupling constants, and wide range of chemical shift dispersion (ca. 5000 ppm) and a relative sensitivity of 5.4 times of $^{13}$C and 8 times of $^{113}$Cd for an equal number of nuclei. Therefore, $^{199}$Hg NMR techniques ($^{199}$Hg and $^1$H-$^{199}$Hg HMQC) have been used as useful tools to probe the metal coordination environment in biological relevant complexes [90, 91] as well as regulatory proteins, copper enzymes, and zinc transcription factor complexes as large as 50 kDa [92-96]. The chemical shift of $^{199}$Hg is very sensitive to the primary coordination spheres including numbers, types of ligands as well as coordination geometry. For examples, $^{199}$Hg bound blue copper proteins gave rise to signals at ca -880 ppm for azurin and -749 ppm for plastocyanin and -706 ppm for rusticyanin [92, 95]. These blue proteins have very similar coordination environments e.g. two histidines and one cysteine and the fourth weakly associated ligand (Met or others) which is slightly different for these proteins. Nevertheless, such a subtle difference can be faithfully reflected by the chemical shifts of $^{199}$Hg. The chemical shift map of $^{199}$Hg can be derived based on various model complexes and proteins, Fig. 8.5b. The large chemical shift dispersion for $^{199}$Hg allows clear differentiation between a variety of M(SR)$_n$ environments. Given that Hg$^{2+}$ is readily exchanged for the native metal ion in many copper, zinc and iron metallo-
proteins [96], $^{199}$Hg NMR methods can play an important role in structural, spectroscopic and chemical studies of metalloproteins and metal-binding domains where the tertiary structure of the folded proteins dictates the geometry of the metal ion.

Fig. 8.5 (a) $^{207}$Pb chemical shifts of various lead model complexes and $^{207}$Pb-substituted proteins relative to external 1 M $[\text{Pb(NO}_3\text{)}_2]$ in 99.9% D$_2$O pH $^*$ 3.3.
The \( \text{pz} \) represents Pyrazolyl ring and \( S\text{-Cys} \) represents a thiolate from cysteine (adapted from ref \[87\], \[89\]). (b) \(^{199}\text{Hg}\) chemical shifts of aliphatic amine/thiol model complexes and \(^{199}\text{Hg}\)-substituted proteins relative to \( \text{Hg}(\text{CH}_3)_2 \) at 298 K. The NRH\(_2\) represents a primary amine and \( SR \) represents a thiolate, and the coordination environments of the Hg proteins include His imidazole nitrogen, Cys thiolate S-Cys and Met thioether S*-Met \[94\].

### 3.2 Chemical shift perturbation

Protein NMR chemical shift is highly sensitive to the exact environment of the atom and can provide valuable insights into structural features including metalligation. For examples, chemical shifts of \(^{13}\text{C}\)\(\beta\) for the zinc bound cysteines (ca. 34 ppm) are significantly downfield shifted relative to those of non-metal bound cysteines (ca. 27 ppm) \[97\]. Such an index has often been used to discriminate zinc bound cysteine residues \[15, 98\]. The chemical shift perturbations upon metal ion binding can be used to estimate the affinity, stoichiometry, and the kinetics of metal binding, and moreover it can be used to identify metal coordination environment. This approach is usually denoted as chemical shift mapping, which has been widely used to study protein-protein interaction \[99, 100\]. \(^1\text{H}\) NMR and \(^1\text{H}\)-\(^1\text{H}\) TOCSY have been used to identify the types of residues binding to metal ions for unlabeled small metalloproteins \[101-103\]. However, two dimensional HSQC especially \(^1\text{H}\)-\(^{15}\text{N}\) HSQC is often employed in studies of metal coordination environment in metalloproteins due to the fact that it is well resolved in comparison with 2D \(^1\text{H}\)-\(^13\text{C}\) HSQC. The identity of each cross-peak in the 2D \(^1\text{H}\)-\(^{15}\text{N}\) HSQC spectra is assigned based on a series of triple resonance experiments. Binding of metal ions would lead to appearance of new peaks or disappearance of original peaks depending on the exchange rates of the apo- and metal-bound forms on the NMR time scales \[99, 104\]. The chemical shifts perturbation (CSP) can be followed in titration experiments, where the concentration of diamagnetic metal ion is increased gradually. When heteronuclear data are available, the binding site is usually predicated by the combined chemical shift perturbation \( \Delta \delta_{\text{comb}} \), which has been showed to be a more reliable approach to
quantitative evaluate titration data [105]. Although several approaches are available to obtain this value as summarized previously [105], in practice, the weighting for $^1$H and $^{15}$N are considered to be the same, the $\Delta \delta_{\text{comb}}$ is usually quantified by the following equation:

$$
\Delta \delta_{\text{comb}} = \sqrt{\frac{\Delta \delta_{\text{HN}}^2 + \frac{1}{25} \Delta \delta_{\text{N}}^2}{2}}
$$

Chemical shift mapping provides information about the location of the metal-binding sites and has been used to identify the metal coordination for numerous metalloproteins [15, 106-108]. For example, upon addition of Cu$^+$ to human Cox17 induced significant chemical shift variations over residues Lys20 and Ala24, and the appearance of the NH signals of Cys22-Cys23, which is thought to serve as the Cu$^+$ binding motif [107]. However, this approach usually has to be used in combination with other physical techniques or biological approaches e.g. mutagenesis to specify the metal binding sites since the chemical shift perturbations mainly stem from either direct binding or conformational changes caused by the metal ions. Binding of Ni$^{2+}$ to *Helicobacter pylori* HypA led to disappearance of signals of Glu3 and Asp40 in the 2D $^1$H-$^{15}$N HSQC spectrum. When combining with mutagenesis study, side-chain 2D $^1$H-$^{15}$N HMQC, UV absorption as well as CD, it was proposed that Ni$^{2+}$ coordinates with His2 (side-chain $N^\delta$), His2 (backbone), Glu3 and Asp40 with a square-planar geometry [15]. Such a binding also induced structural changes which were thought to be important for its downstream receptor’s recognition [15].

Histidine often serves as a metal binding ligand in metalloproteins and can provide both backbone and side nitrogens to coordinate with metal ions such as Zn$^{2+}$ and Ni$^{2+}$. It has been shown that different tautomeric forms of histidine imidazole rings have different, distinguishable signal patterns in a long-range 2D $^1$H-$^{15}$N HMQC spectrum [109] and metal coordination often causes recognizable changes in the NMR spectrum of histidine side-chains and that imidazole nitrogen atoms involved in direct metal coordination have specific chemical shift [15, 40, 98, 110]. This technique has been extensively used to identify Zn$^{2+}$ binding. The
chemical shifts observed for the unprotonated imidazole nitrogen atoms of a zinc finger domain Hdm2(429-491) appeared at ca 215 ppm. The cross-peak pattern in $^1$H-$^{15}$N HMQC spectrum (with $^2$J$_{HN}$) of zinc bound Hdm2 (429-491) showed that His452 and His457 assumed different tautomeric forms with the former being N$^{\delta 2}$-protonated and the latter N$^{\delta 1}$-protonated with chemical shifts around 170 ppm, which demonstrated that Zn$^{2+}$ coordinated to both His452 and His457 via the N$^{\delta 2}$ and N$^{\delta 1}$ respectively [98]. More interestingly, a comparison of the 2D $^1$H-$^{15}$N HMQC spectra of Zn$^{2+}$-bound proteins with $^{113}$Cd$^{2+}$-bound proteins, as shown in Fig. 8.6, allows one to observe the $^{15}$N-$^{113}$Cd coupling, which assists identification of overlapping of histidine side-chains that binds to metal ions [40]. The HMQC spectrum of the zinc-bound domain, shown as black in Fig. 8.6a, clearly shows two of the $^{15}$N resonances His42 and His40 shifted downfield as a result of zinc coordination. Based on the pattern of the cross-peaks, His42 is in the $\varepsilon$ tautomeric state indicative of zinc coordination to the N$^{\delta 1}$ of His42, Fig. 8.6b. Such a method cannot be used to assign the nitrogen atom of His40 due to overlapping of the H$^{\delta 2}$ and H$^{\varepsilon 1}$ resonances of His40. The HMQC spectrum of $^{113}$Cd bound protein, shown as red in Fig. 8.6a, clearly shows coupling between $^{113}$Cd and the N$^{\delta 1}$ of His42 observed on H$^{\varepsilon 1}$, suggesting a covalent bond between them. Importantly, the pattern of the connectivities and the coupling observed to both H$^{\delta 2}$ and H$^{\varepsilon 1}$ from the $^{113}$Cd allows the unambiguous assignment of the $\delta$ tautomer for His40, with the metal coordination via the N$^{\delta 2}$ of the side-chain, Fig. 8.6b [40]. The side-chain 2D $^1$H-$^{15}$N HMQC spectrum of histidine has also been used to identify Ni$^{2+}$ binding atoms in H. pylori HypA and significant downfield shift of His2 upon Ni$^{2+}$ binding indicates that Ni$^{2+}$ binds to His2 through the N$^{\delta 1}$ atom [15].
Fig. 8.6 (a) Comparison of the 500 MHz $^1$H-$^1$5N HMQC spectrum of Zn-bound ZZ domain in black with the $^{113}$Cd-bound form in red. The coupling of $^{113}$Cd to $N_\varepsilon^2$ is observed on $H_\varepsilon^2$ and $H_\varepsilon^2$ for His40; whereas $^{113}$Cd coupling to $N_\delta^1$ is observed on $H_\varepsilon^1$ for His 42. (b) Diagrams showing the connectivities of two zinc-ligand histidine residues H40 and H42 based on the above HMQC spectrum. The coupling of $^{113}$Cd to the attached $^1$5N is useful for identification of metal bound histidine residues and the location of the metal coordination on the histidine rings (adapted from reference [40]).

4 NMR in studies of paramagnetic metalloproteins
Metalloproteins represent a large share of a proteome. A large number of them contains paramagnetic metal ions, which possess unpaired electrons. The presence of a paramagnetic centre causes pronounced effects in NMR spectra and reduces dramatically the intensity of NOEs and the efficiency in the transfer of scalar couplings both in homonuclear and heteronuclear experiments which hampers spectrum assignment and structural determination through standard approaches. However, with the advances in novel experiment design and development of software protocol in recent years, the presence of paramagnetic centre has been used as a precious source of structural information [19, 20]. Extensive reviews regarding this topic can be referred to selected reviews [19-22, 111-113]. Here we will give a very brief description on the paramagnetism-based restraints as well as application of $^{13}$C-detected experiments in structural of paramagnetic metalloproteins.

4.1 Paramagnetism-based structural restraints

In paramagnetic systems, where paramagnetic metal ions are either intrinsic or extrinsic, there are three NMR experimental observables that yield long-range structural information e.g. paramagnetic relaxation enhancement (PRE); pseudo-contact shifts (PCS); residual dipolar couplings (RDC) induced by anisotropic paramagnetic centers. In addition, cross-correlated relaxation (CCR) effects between anisotropic paramagnetic centers and anisotropic parameters of the nuclear spins can also be exploited to generate long-range restraints [19, 111, 112]. Paramagnetic centers with isotropic electron spin distribution (Mn$^{2+}$ and Gd$^{3+}$) produce large PREs due to slow electron relaxation. In contrast, paramagnetic centers with anisotropic electron spin distribution for most paramagnetic metal ions including most of the lanthanides create all four long-range paramagnetic effects, which contains rich structural information [114]. Here, we will focus on the PRE and PCS and their applications. Information about RDC can be found in Chapter 4.

The PRE arises from magnetic dipolar interactions between a nucleus and the unpaired electrons of the paramagnetic center, resulting in an increase in nuclear relaxation rates. In contrast to NOE, where the effects are limited to short
range interaction (< 6Å); the PRE effects are relatively large and can be detected up to 35 Å owing to large magnetic moment of an unpaired electron. There are two mechanisms e.g. the Solomon mechanism and the Curie spin mechanism that contribute the PREs, with the former being predominate for slowly tumbling molecules with long life-times of the electronic spin state (such as Mn$^{2+}$ and Gd$^{3+}$). While the Curie relaxation becomes important when the electronic relaxation is much faster than the rotational tumbling of the molecules, which is the case for the majority of paramagnetic metal ions. Theoretical and experimental aspect of PRE as well as its application in studies of structures of proteins and protein-protein complexes can be found in recent reviews [114-116].

At high magnetic fields (over 500 MHz for $^1$H frequency), the PRE rate, $\Gamma_2$ which arises from the dipole-dipole interaction between a nucleus and unpaired electrons with an isotopic $g$-tensor is conventionally calculated by the Solomon-Bloembergen (SB) equation:

$$\Gamma_2 = \frac{1}{15} \left( \frac{\mu_0}{4\pi} \right)^2 \gamma_i^2 g^2 \mu_B^2 S(S+1)r^{-6} \left\{ 4\tau_c + \frac{3\tau_c}{1 + (\omega / \tau_c)^2} \right\}$$

Where $r$ is the distance between the paramagnetic centre and the observed nucleus; $\mu_0$ is the permeability of vacuum; $\gamma_i$ the nuclear gyromagnetic ratio; $g$, the electron $g$-factor; $\mu_B$, the electron Bohr magneton; $S$, the electron spin quantum number and $\tau_c$, the PRE total correlation time. In practice, $\Gamma_2$ is measured as a difference in transverse relaxation rates between the paramagnetic ($R_{2,\text{para}}$) and diamagnetic ($R_{2,\text{dia}}$) states. A two-time point measurement is recommended as a simple approach of obtaining $\Gamma_2$ rates and their corresponding errors without fitting procedures. In this approach, the $\Gamma_2$ rates are determined from two time points ($T=0$ and $\Delta T$) for transverse relaxation as shown by the following equation [116]:

$$\Gamma_2 = R_{2,\text{para}} - R_{2,\text{dia}} = \frac{1}{T_b - T_a} \ln \frac{I_{\text{dia}}(T_b) I_{\text{para}}(T_a)}{I_{\text{dia}}(T_a) I_{\text{para}}(T_b)}$$

Where $I_{\text{dia}}$, $I_{\text{para}}$ are the peak intensities for the diamagnetic and paramagnetic states, respectively. The choice of time points is important to minimize the error.
For example, if the range of $\Gamma_2$ rates is 0-75 s$^{-1}$, a second time point $T_b$ should be at ca. $1.15/(R_{2\text{dia}+50})$ s, represent a reasonable choice [115, 116].

For isotropic metal ions such as Mn$^{2+}$ and Gd$^{3+}$, the Curie-spin relaxation that could potentially exhibit significant cross-correlation with other relaxation mechanisms, is negligible for medium-size macromolecules [116], the Solomon relaxation is predominant. PRE analysis in such a system is thus simple. The PRE has been used extensively in metalloproteins that possess a rigid intrinsic paramagnetic center [117-119]. Such a strategy has also been extended not only in the NMR structure determination of non-metalloproteins [120-122], in which paramagnetic metal ions (Mn$^{2+}$ or Gd$^{3+}$) or nitroxide radicals were conjugated through appropriate chemical modification [123], but also in the characterization of protein-protein/nuclei acid complexes [124-126]; and membrane-proteins [127], in particular in transient macromolecular interactions [115, 128-131].

Pseudocontact shifts (PCSs) are precious source of structure information and observed only in paramagnetic systems with anisotropic unpaired electrons e.g. Dy$^{3+}$, Tb$^{3+}$ and Fe$^{3+}$. The magnitude of the PCS, $\delta_{\text{pcs}}$, is calculated using the following equation [132]:

$$\delta_{\text{pcs}} = \frac{1}{12\pi r^3} \left\{ \Delta \chi_{ax} (3 \cos^2 \theta - 1) + \frac{3}{2} \Delta \chi_{rh} \sin^2 \theta \cos \phi \right\}$$

$$\Delta \chi_{ax} = \Delta \chi_{zz} - \frac{1}{2} (\chi_{xx} + \chi_{yy}) \quad \text{and} \quad \Delta \chi_{rh} = \chi_{xx} - \chi_{yy}$$

Where $r$ is the distance between the metal ion and the nuclear spin; $\theta$ and $\phi$ are the angles describing the position of the nuclear spin with respect to the principle axes of the magnetic susceptibility tensor $\chi$; and the $\Delta \chi_{ax}$ and $\Delta \chi_{rh}$ are the axial and rhombic components, respectively, of the magnetic susceptibility tensor.

The PCS are manifested by large changes in chemical shifts of the nuclear spins that are exposed to the paramagnetic metal ions and arise from through-space dipolar interactions with rapidly relaxing unpaired electrons. The PCS displays a $r^3$ distance dependence, in contrast to the $r^6$ dependence for the PRE,
which results in a relatively long distance range for the PCS to be detected (ca. 40 Å for Dy³⁺) [133]. In general, the $\delta_{\text{pcs}}$ values can be measured after the complete assignment for the $^1$H-$^{15}$N HSQC spectra of both the diamagnetic and paramagnetic samples is obtained and is calculated as the difference between the chemical-shift values observed for the nuclei in a paramagnetic system and in a diamagnetic analogue. The $\delta_{\text{pcs}}$-derived restraints alone cannot be used to solve the structures. Instead, the PCSs have to be incorporated with NOEs and dihedral-angle restraints to determine structures of proteins or to refine protein structures. The first example of using PCS in a structure refinement was reported on a low-spin Fe³⁺ heme protein [134]. Such a strategy has been extended not only in studies of paramagnetic proteins [135, 136], but also in non-metal binding proteins [137, 138], which were labeled by paramagnetic metal ions such as lanthanides [114, 139], or genetically encoded Co²⁺-binding amino acid [140]. Moreover, PCSs can also be used as restraints in molecular dynamics [141, 142].

4.2 $^{13}$C-detected protonless NMR

Direct-detection of heteronuclei, in particular $^{13}$C offers a valuable alternative to $^1$H detection in the study of biological macromolecules [143-145] as well as paramagnetic proteins [136, 145-147]. The recent development of high magnetic fields as well as the availability of cryogenically cooled probe-heads has improved $^{13}$C sensitivity significantly, which greatly stimulated research of using $^{13}$C-detected experiments on enriched samples to study biological macromolecules [148]. The $^{13}$C detection takes advantage of the slower relaxing $^{13}$C spins and overcomes the drawbacks produced by the fast $^1$H transverse relaxation, which led $^1$H signals broaden beyond detection limits for large proteins. Such an approach is particularly useful in paramagnetic systems since $^{13}$C direct detection is less affected by the paramagnetic center than $^1$H owing to the lower $^{13}$C gyromagnetic ratio, which decreases the paramagnetic dipolar contributions to its relaxation by a factor of around 16 ($\gamma_C/\gamma_H)^2$ [136, 146, 149].

In $^{13}$C direct-detection experiments, several approaches were used to achieve “virtual” decoupling (to remove homonuclear one-bond carbon-carbon couplings)
such as IPAP schemes (in-phase anti-phase) [150-152], in which two FIDs for each increment are recorded and stored separately, one for in-phase and another for anti-phase and two components are combined to remove the splitting; or SE schemes [152, 153] (spin-state selective excitation), in which two different experiments are performed with one being absorptive and another dispersive. One or more of these building blocks (IPAP and SE) can be implemented in any experiments based on $^{13}$C direct-detection. A set of $^{13}$C based experiments, which can be used for the assignment of backbone and side-chains of $^{13}$C/$^{15}$N labeled proteins, are now available and summarized in a recent review [144]. The sequence-specific assignment was achieved by CACO and CANCO, which provide the correlation of each CO to the two neighboring C$^{\alpha}$ nuclei; CACO, CBCACO and $^{13}$C-$^{13}$C TOCSY can provide spin-system assignment [143, 154, 155]. The CON-IPAP experiment is used to correlate backbone nitrogen with CO through the one-bond C$^{\prime}$-N coupling [149]. The $^{13}$C-$^{13}$C NOESY experiments based on dipole-dipole interaction with longitudinal magnetization transfer represent a valuable alternative to COSY experiments based on scalar couplings to detect C-C one bond correlation for large macromolecules and paramagnetic metalloproteins [149, 156, 157]. The $^{13}$C-$^{13}$C NOESY with direct $^{13}$C detection is exploited as a valuable tool to extend the assignment to side-chains in large molecules such as C$^{\prime}$-C$^{\beta}$ if mixing times is long enough (e.g. 800 ms) [156]. The correlations between nuclei not directly bound and not mediated by spin diffusion in $^{13}$C-$^{13}$C NOESY spectra are identified, which would represent a breakthrough of structure determination of large macromolecules in solution by providing distance constraints. However, the sensitivity for $^{13}$C direct detection is required to be improved before long-range correlations to obtain $^{13}$C-$^{13}$C distance constraints can be used in structural characterization of large macromolecules.

$^{13}$C direct detection has been successfully applied to paramagnetic proteins, where the contribution to line broadening coming from the paramagnetic centre is so large that $^1$H signals around the metal ion is beyond detectable limits [136, 144, 147, 149, 158]. Such a technique can also be used in generation of paramagnetism-based restraints including PCS, PRE, RDC [148, 159-161]. It has been dem-
onstrated that $^{13}$C directly detected spectra provide an alternative method for the measurement of RDC with precision as good as that from $^1$H detection, but with additional advantage for measuring those broad resonances in $^1$H detection [148]. Direct-detection of $^{13}$C intrinsically offers a way to detect resonances close to the metal ion where $^1$H resonances are too broad to be detected. Indeed, with the aid of $^{13}$C direct detection approach, $^{13}$C resonances as close as 6 Å from the metal ion are detected for CopC, a Cu$^{2+}$ binding protein involved in copper homeostasis, whereas no $^1$H resonance can be detected in a sphere of 11 Å from the metal due to fast relaxation caused by paramagnetic Cu$^{2+}$ [136]. Incorporation of heteronuclear paramagnetism-based restraints e.g. PCSs and longitudinal relaxation rate enhancement allows CopC structures to be resolved with the RMSD of Cu$^{2+}$ determined only by the paramagnetism-based constraints of 1.1 Å [136]. $^{13}$C direct detection technique has also been used for residue-specific assignments of resonances, in particular those near paramagnetic centers (e.g. Ni$^{2+}$, Fe$^{3+}$) such as in a 20 kDa Ni-containing enzyme, acireductone dioxyhenase (ARD) [162] and oxidized human [2Fe-2S] ferredoxin [146], as well as a 19 kDa Fe$^{3+}$ hemophore HasA [147]. In many paramagnetic systems, the longitudinal relaxation rates are influenced to a smaller extent than the transverse relaxation rates. The $^{13}$C-$^{13}$C NOESY experiments are therefore a useful approach to overcome the quench of scalar coupling based transfer in particular for large macromolecules. The use of $^{13}$C direct detected experiments e.g. $^{13}$C-$^{13}$C COSY, $^{13}$C-$^{13}$C NOESY and $^{13}$C-$^{13}$C COCAMQ allows $^{13}$C signals as close as 4 Å to Cu$^{2+}$ to be detected in oxidized monomeric copper, zinc superoxide dismutase (SOD) [149]. The advantage of $^{13}$C-$^{13}$C NOESY experiments for higher molecular weights was seen by comparison of the protein SOD [149, 156]. All of the expected Cα-CO connectivities were detected with higher intensity in the dimeric protein than in the monomeric state. In addition, most of the two bond CO-Cβ cross-peaks were observable for the dimeric SOD when the long mixing times were used [156]. Interestingly, the intrinsic asymmetry of a $^{13}$C-$^{13}$C COSY experiment allows the coordinating residues of a paramagnetic metal ions to be identified easily, provides a unique method to distinguish between monodentate and bidentae coordinating side-chain carbonyls...
Significantly, the $^{13}$C-based strategy in combination with solid-state NMR led to partial sequence-specific (35%) and side-chain assignments for the iron storage protein, ferritin, a very large protein with a molecular mass of 480 kDa and 24-subunits [164, 165]. The solution $^{13}$C-$^{13}$C NOESY spectra for side-chain observation has provided the identification of an iron channel that guides the direction transport of the multimeric Fe$^{3+}$ products from the active site toward the nanocage. The interior of the four-helix bundle is identified as the functional channel based on the observed paramagnetic effects on residues lining the internal face of the four-helix bundle. The NMR data provide a basis for the pathway of iron from the ferrous/dioxygen oxidoreductase site to the central cavity of ferritin [164]. Such studies open new avenues for the application of $^{13}$C direct detection experiments to systems with molecular assemblies larger than 100 kDa.

5 Perspectives

Since the first protein solution structure was determined by high resolution NMR spectroscopy about 25 years ago [166], NMR has been established as the only experimental method that provides both structural and dynamical information at atomic resolution close to physiologically relevant conditions. Protein structure determination in living cells has also been achieved recently by in-cell NMR [18]. However, the limitation of this technique in structural studies lies in low sensitivities and poor resolution when the size of macromolecules increases. Some proteins, in particular metalloproteins might not be stable for a period of time (days or weeks) or have limited solubility. Moreover, new challenges in life science has also promoted development of new NMR methods which will improve sensitivities and reduce acquisition times to fulfill the requirement of characterization of these proteins and their complexes.

Enormous effort has been made to improve NMR instrumentation in terms of experimental sensitivity, which results in availability of high-field magnets, cryogenically cooled probes. In the meantime, tremendous advances in methodology have contributed to an increased interest in the study of molecular systems of increasing size and complexity. The introduction of non-linear sampling scheme (in-
stead of conventional uniform sampling) allows a very fast acquisition of multi-
dimensional NMR [167-169]. Many schemes have been developed to reduce the
spectral dimensionality, thus to speed up the experiments, which enable quick as-
ignment of large proteins [170, 171]. The examples include “GFT” NMR ap-
proach where sub-spectra from joint sampling of indirect dimensions are linearly
recombined and analyzed [172]. In projection reconstruction (“PR”) method, the
corresponding full-dimensional spectrum is reconstructed [171, 173, 174]. Moreo-
ver, various ultrafast NMR techniques including SOFAST/BEST NMR [175, 176]
and Hadmard NMR [177] are also available for studies events of biomolecules
even in real time. All these new schemes deliver appreciable improvement in the
speed of data acquisition and show promise for speed up multidimensional NMR
of normal size proteins [170, 178, 179] and very large proteins [180, 181] as well
as sequence assignment for intrinsically unstructured proteins [178].

$^{13}$C NMR spectroscopy is emerging as a powerful tool to complement $^1$H NMR
spectroscopy in the investigation of biomolecules, in particular for large molecules
and paramagnetic metalloproteins and also for the study of short-lived molecules
[182]. However, the data acquisition time is rather long even for sample with high
concentrations (ca. mM). Implementation of fast NMR methods such as non-
uniform sampling in the indirect dimension significantly reduced experimental
times [183]. Such a strategy will open new avenues to applications of $^{13}$C NMR to
take advantage of the favorable heteronuclear chemical shift dispersion in biologi-
cal systems especially for systems of increasing size.

Chemical shifts of selective nuclei ($^{113/111}$Cd, $^{209}$Pb, $^{195}$Pt) are sensitive towards
types of ligands (N, O and S), numbers and geometries, and will continue to play a
role in characterization of metal-protein local coordination. The ultra-field NMR
facilities direct observation of biologically important metal ions with half integer,
quadrupolar nuclei (e.g. $^{67}$Zn and $^{25}$Mg) [184, 185].

In spite of availability of all these new techniques, their applications in metal-
loproteins are currently sparse. There is an urgent need to promote these advanced
techniques in the scientific community through introducing integrated software
packages for experimental set-up, data processing and analysis. This will enable
protein chemists and bioinorganic chemists, who are not NMR experts, to employ the new techniques in their research. The combination of fast NMR techniques, $^{13}$C directly detected NMR with paramagnetic NMR will offer great potentials in tackle new challenges in life science and open new avenues for NMR spectroscopy to be utilized not only in the characterization of single biomolecules e.g. structural and dynamical studies of proteins/metalloproteins and paramagnetic proteins, short-lived macromolecules and intrinsically unstructured proteins, but also in investigation of more complex systems to give an integrated view of interacting molecular networks. In particular, the ultra-fast NMR opens new perspectives for the NMR structural investigation of unstable protein/metalloprotein samples and real-time site-resolved studies of protein kinetics or monitoring folding/unfolding processes of proteins/metalloproteins caused by ligand or metal binding/release.

Acknowledgements
This work is supported by the Research Grants Council of Hong Kong (HKU7043/06P, HKU2/06C, HKU7042/07P, HKU1/07C, HKU7038/08P, HKU7049/09P and N-HKU752/09), Croucher Foundation.

Reference

64. He F, Umehara T, Saito K et al (2010) Structural insight into the zinc finger CW domain as a histone modification reader. Structure 18: 1127-1139


75. Li H, Otvos JD (1996) 111Cd NMR studies of the domain specificity of Ag+ and Cu+ binding to metallothionein. Biochemistry 35: 13929-13936


79. Pan T, Coleman JE (1990) GAL4 transcription factor is not a "zinc finger" but forms a Zn(II)2Cys6 binuclear cluster. Proc Natl Acad Sci USA 87: 2077-2081
91. Iranzo O, Thulstrup PW, Ryu SB et al (2007) The application of $^{199}$Hg NMR and (199m)Hg perturbed angular correlation (PAC) spectroscopy to define the biological chemistry of Hg(II): a case study with designed two- and three-stranded coiled coils. Chemistry 13: 9178-9190
93. Steele RA, Opella SJ (1997) Structures of the reduced and mercury-bound forms of MerP, the periplasmic protein from the bacterial mercury detoxification system. Biochemistry 36: 6885-6895


enhancement titration measurements. Proc Natl Acad Sci USA 107: 1379-1384


