A strong 13C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins.

Pierre Barraud, Mario Schubert, Frédéric H-T Allain

A strong ^{13}C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins

Pierre Barraud†, Mario Schubert & Frédéric H.-T. Allain†

Institute of Molecular Biology and Biophysics, ETH Zurich, CH-8093 Zürich, Switzerland

† Corresponding authors: P Barraud, Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstrasse 20, CH-8093 Zürich, Switzerland. Tel.: +41 44 633 0716, Fax: +41 44 633 1294. E-mail: pierre.barraud@mol.biol.ethz.ch / FH-T Allain, Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstrasse 20, CH-8093 Zürich, Switzerland. Tel.: +41 44 633 3940, Fax: +41 44 633 1294. E-mail: allain@mol.biol.ethz.ch

Abstract

Zinc is the second most abundant metal ion incorporated in proteins, and is in many cases a crucial component of protein three-dimensional structures. Zinc ions are frequently coordinated by cysteine and histidine residues. Whereas cysteines bind to zinc via their unique S^{γ} atom, histidines can coordinate zinc with two different coordination modes, either $N^{\delta 1}$ or $N^{\varepsilon 2}$ is coordinating the zinc ion. The determination of this coordination mode is crucial for the accurate structure determination of a histidine-containing zinc-binding site by NMR. NMR chemical shifts contain a vast amount of information on local electronic and structural environments and surprisingly their utilization for the determination of the coordination mode of zinc-ligated histidines has been limited so far to ^{15}N nuclei. In the present report, we observed that the ^{13}C chemical shifts of aromatic carbons in zinc-ligated histidines represent a reliable signature of their coordination mode. Using a statistical analysis of ^{13}C chemical shifts, we show that $^{13}C^{\delta 2}$ chemical shift is sensitive to the histidine coordination mode and that the chemical shift difference $\delta\{^{13}C^{\varepsilon 1}\} - \delta\{^{13}C^{\delta 2}\}$ provides a reference-independent marker of this coordination mode. The present approach allows the direct determination of the coordination mode of zinc-ligated histidines even with non-isotopically enriched protein samples and without any prior structural information.

Keywords

NMR spectroscopy; ^{13}C chemical shifts; histidine; Zn-ligated histidine; zinc-binding protein; coordination mode of histidine; histidine tautomers; chemical shift analysis

Introduction

NMR chemical shifts are strongly sensitive to local electronic environment and therefore contain an immeasurable amount of information on protein structures and dynamics (Wishart and Case 2002). This information is largely used in the different steps of the structure determination process by NMR, namely chemical shift resonance assignment (Wuthrich 1986; Oh et al. 1988; Grzesiek and Bax 1993), secondary structure and backbone angle restraints predictions (Wishart and Sykes 1994; Cornilescu et al. 1999; Shen et al. 2009), determination of side-chain rotamer conformation (London et al. 2008; Hansen et al. 2010; Hansen and Kay 2011), detection of cis-peptide bonds (Schubert et al. 2002), prediction of the oxidation state of cysteines (Sharma and Rajarathnam 2000), and even complete determination of three-dimensional structures (Cavalli et al. 2007; Shen et al. 2008; Wishart et al. 2008).

A large number of proteins incorporate metal ions as part of their structure (hence called metalloproteins). According to various estimates, about one-quarter to one-third of all proteins require metal ions to perform their biological function (Rosenzweig 2002; Dupont et al. 2010). Zinc is only the second most abundant metal ion in living organisms after iron, but regarding the large variety of zinc-binding protein domains, zinc is the most widespread incorporated metal ion in protein structures. Indeed, in eukaryotes, about 9 % of all proteins are predicted to include at least one zinc ion in their structure (Dupont et al. 2010; Andreini et al. 2006b). Zinc is commonly coordinated by four ligands provided by protein side chains in structural sites and by only three protein ligands in catalytic sites in which the coordination sphere is completed by a water molecule involved in catalysis (Auld 2001). The most common protein side chains involved in zinc-coordination are by far cysteines and histidines, found with similar usage, followed by aspartate and glutamate (Andreini et al. 2006a).

Contrary to cysteine ligands, for which the atom that directly contacts the zinc ion is unique (Cys S^{γ}), histidine ligands bind to zinc ions with two different coordination modes involving either one or the other of the two endocyclic nitrogens $N^{\delta 1}$ and $N^{\varepsilon 2}$ (Figure 1a). It is important to notice that these coordination modes are intrinsically associated with the tautomeric form of the histidine ring, where a $N^{\delta 1}$ coordination mode is uniquely compatible with a protonation on $N^{\varepsilon 2}$ and vice versa (Figure 1a and b). Since the two endocyclic nitrogens are 2.2 Å apart, the correct determination of the histidine coordination mode is of primary importance for the accurate structure determination by NMR of a histidine-containing

zinc-binding site.

Two different approaches are currently used for this purpose. In the first approach, a preliminary structure is determined without any information on the histidine-coordination mode, and this one is latter deduced from the analysis of the bundle of structures (Lee et al. 1989; Omichinski et al. 1990; Dempsey et al. 2004; Malgieri et al. 2007; Bessière et al. 2008; Cordier et al. 2008; Briknarová et al. 2011). This method is effective when NOE patterns to $H^{\delta 2}$ and $H^{\varepsilon 1}$ aromatic protons are sufficiently different to unambiguously position a single endocyclic nitrogen in the vicinity of the zinc ion. Note that this approach can sometimes be unreliable, since the differences between corresponding distances present in the two alternative structures, formed via $N^{\delta 1}$ or $N^{\varepsilon 2}$ coordination, can be rather small. The second approach aims at determining the tautomeric form of the histidine ring through the measurement of the ^{15}N chemical shifts of the endocyclic nitrogens $N^{\delta 1}$ and $N^{\varepsilon 2}$ (Legge et al. 2004; Ramelot et al. 2004; Kostic et al. 2006; Liew et al. 2007; Estrada et al. 2009; Peroza et al. 2009; Eustermann et al. 2010). This is achieved via the measurement of optimized long-range ^{15}N-^{1}H HSQC or HMQC (Pelton et al. 1993). The identification of the $N^{\delta 1}$ and $N^{\varepsilon 2}$ chemical shifts takes advantage of the fact that $N^{\varepsilon 2}$ is connected to two close proton-neighbours via a $^{2}J_{NH}$ coupling ($H^{\delta 2}$ and $H^{\varepsilon 1}$) whereas $N^{\delta 1}$ couples to only one such neighbour ($H^{\varepsilon 1}$) (Figure 1b). As far as the $N^{\delta 1}$ and $N^{\varepsilon 2}$ chemical shifts can be identified, this approach is unambiguous. Indeed, the chemical shifts of the protonated nitrogen and the non-protonated nitrogen in a neutral histidine ring display an average difference of about 80 ppm, with protonated nitrogen chemical shifts being around 170 ppm and the non-protonated ones, with a lone pair, around 250 ppm (Figure 1b). Therefore, the only possible ambiguities related to this method come from chemical shifts overlaps between different histidine residues which could render confident assignments impossible. This method requires a ^{15}N-labelled protein sample, and this definitely creates limitations in the case of samples obtained via chemical synthesis, a method commonly used for the production of small zinc-binding peptides.

Here we present a simple and reliable method for the identification of the coordination mode of histidines based on ^{13}C aromatic carbon chemical shifts. We made use of the Biological Magnetic Resonance Data Bank (BMRB) to conduct a statistical analysis on ^{13}C aromatic carbon chemical shifts of zinc-coordinated histidines and demonstrate that ^{13}C aromatic chemical shifts, which can be easily obtained even with non-labelled protein samples, constitute an excellent marker for the determination of the coordination mode of histidines.

Methods

Data collection and analysis

A database of histidine ^{13}C aromatic carbons $C^{\delta 2}$ and $C^{\varepsilon 1}$ chemical shifts was constructed from the BMRB database (http://www.bmrb.wisc.edu/). Basic shell scripts were written to identify zinc-coordinated histidines and to extract $C^{\delta 2}$ and $C^{\varepsilon 1}$ chemical shifts. The database was divided in two groups depending on the annotation present in the BMRB file regarding the coordination mode of the corresponding histidine ($N^{\delta 1}$ or $N^{\varepsilon 2}$ coordination). Only histidines for which both $C^{\delta 2}$ and $C^{\varepsilon 1}$ chemical shifts were available were used in subsequent chemical shift analysis. Wrong annotation of the coordination mode in the BMRB entry as compared with the associated PDB coordinates was present in 3 cases (bmrb entries 10108, 6374 and 11061) and corrected in the constructed database. The coordination mode of one histidine was wrongly assigned when compared to closely related protein structures (bmrb entry 6682) and was also corrected for subsequent analysis. This selection resulted in 223 histidines coordinating zinc via $N^{\varepsilon 2}$ and 43 histidines via $N^{\delta 1}$. Statistical analysis of the chemical shifts were carried out with Gnumeric (http://projects.gnome.org/gnumeric/). Chemical shifts were grouped in 0.5 ppm bins for analyzing the statistics of occurrence.

In addition, a database with ^{13}C aromatic carbons $C^{\delta 2}$ and $C^{\varepsilon 1}$ chemical shifts of non-coordinated histidines was constructed from the Reference Corrected Database (RefDB) (http://refdb.wishartlab.com/) (Zhang et al. 2003) with similar basic shell scripts. Histidines having both $C^{\delta 2}$ and $C^{\varepsilon 1}$ chemical shifts assigned were used for analysis.

Results

Initial questions and genesis of the project

In the structure determination process of a protein domain with an unexpected zinc-binding site (Barraud et al. 2011), we noticed the presence of a downfield shifted $C^{\delta 2}$ of a histidine side chain ($C^{\delta 2}$ = 126.1 ppm). This is rather unusual in proteins where most histidine $C^{\delta 2}$ have chemical shifts around 120 ppm. We wondered whether this downfield shifted $C^{\delta 2}$ was a characteristic of zinc-coordinated histidines. On one hand, downfield shifted $C^{\delta 2}$ in histidines are presented as a characteristic of zinc-coordinated histidines in some reports (De Guzman et al. 2000; Ramelot et al. 2004; Kornhaber et al. 2006). But on the other hand, we also noticed numerous examples where zinc-coordination was not associated with a particularly downfield shifted $C^{\delta 2}$ (Zeng et al. 2008; Huang et al. 2009; Eustermann et al. 2011), showing that confusion exists in the literature about this point. Interestingly, the analysis of $C^{\delta 2}$ chemical shifts from few particular enzymes with histidines involved in catalytic triads or from imidazole derivatives indicated that $C^{\delta 2}$ chemical shift would be a sensitive probe of histidines tautomeric forms (Sudmeier et al. 2003; Reynolds et al. 1973). Altogether, this suggest that a downfield shifted $C^{\delta 2}$ in histidines would not be a signature of zinc-coordination but rather a signature of the histidine tautomeric form.

Data mining and primary chemical shift analysis

In order to understand the exact correlation between zinc-coordination, tautomeric form and downfield shifted $C^{\delta 2}$ in histidines, we decided to perform a systematic survey of the chemical shifts of aromatic carbons in zinc-coordinated histidines. For this, we identified zinc-coordinated histidines in all BMRB entries and extracted their ^{13}C aromatic carbon chemical shifts, namely $C^{\delta 2}$ and $C^{\varepsilon 1}$, and retained pairs of chemical shifts for each histidines for which both $C^{\delta 2}$ and $C^{\varepsilon 1}$ were present. This resulted in a database of zinc-coordinated histidines containing a total of 266 entries (See Supplementary Table 1). We then divided the entries in two groups according to the histidine coordination mode reported in the corresponding BMRB entry. The large majority of the histidines (85%) displayed a $N^{\varepsilon 2}$ coordination mode and the rest (15%) a $N^{\delta 1}$ coordination mode. This preference for zinc coordination via $N^{\varepsilon 2}$ is in accordance with previous analysis of crystal structures of metalloproteins (Chakrabarti 1990). A two-dimensional landscape of $C^{\varepsilon 1}$ versus $C^{\delta 2}$ chemical shifts is shown

on Figure 2a with different symbols for the $N^{\delta 1}$ and $N^{\varepsilon 2}$ coordination modes. This shows a clear separation between the two coordination modes, the two groups differing predominantly in their $C^{\delta 2}$ chemical shift. However, few outliers seem to contradict the clear distinction between the $N^{\delta 1}$ and $N^{\varepsilon 2}$ coordination modes. For these five points (numbered 1-5 on Figure 2a), we manually checked the consistence between the annotated coordination mode in the BMRB entry and the effective coordination mode reported in the corresponding structure deposited in the Protein Data Bank (PDB). Out of these five points, three were indeed wrongly annotated in the BMRB entry (point 1, 2 and 3 on Figure 2a), and were then corrected in our database for the subsequent analysis. In addition, the structure corresponding to point 4 (PDB code 1ZR9; BMRB entry 6682) was solved assuming a $N^{\varepsilon 2}$ coordination mode without any direct experimental evidence (Hayes et al. 2008) even if closely related proteins structures had shown unambiguously that zinc-coordination via this histidine was in fact occurring with its $N^{\delta 1}$ (Möller et al. 2005; Andreeva and Murzin 2008). We also corrected this entry in our database. Chemical shifts of the corrected database are presented on Figure 2b. Finally, only a single point out of 266 (point 5 on Figure 2a) was found further away from the average of the histidines with an $N^{\delta 1}$ coordination mode. It should be pointed out that the coordination mode of this histidine has been determined by measuring an optimized long-range ^{15}N-^{1}H HSQC (Kwon et al. 2003). This particular shift might therefore reflect either an incorrect chemical shift assignment or a very special electronic environment (two tryptophan residues are indeed present in the vicinity of this zinc-binding site; PDB code 1NKU). This point remained in the database that was used for statistical analysis.

Chemical shift of $C^{\delta 2}$ aromatic carbons is a sensitive probe of histidine coordination mode

Our statistical analysis revealed a clear separation between the histidine $C^{\delta 2}$ chemical shifts of the two coordination modes (Figure 2b). Statistical values for the aromatic carbons chemical shifts $C^{\delta 2}$ and $C^{\varepsilon 1}$ of our entire zinc-coordinated histidine database were calculated and are given in Table 1. Whereas the $^{13}C^{\varepsilon 1}$ chemical shift is rather insensitive to the coordination mode, there is in average a significant chemical shift difference of about 8.3 ppm for the $^{13}C^{\delta 2}$ chemical shift between the $N^{\delta 1}$ coordination mode (average value $\overline{\delta}$ = 127.42 ppm) and the $N^{\varepsilon 2}$ coordination mode (average value $\overline{\delta}$ = 119.09 ppm). This chemical shift difference is not due to the zinc coordination *per se,* but reflects the nature of the tautomeric form of the histidine. To be precise, the $N^{\varepsilon 2}$ coordination mode is uniquely compatible with the $N^{\delta 1}$-H tautomeric form (Figure 1) and as a consequence, the $^{13}C^{\delta 2}$ chemical shift has a characteristic

value of about 127 ppm. Vice versa, the $N^{\delta 1}$ coordination mode is uniquely compatible with the $N^{\varepsilon 2}$-H tautomeric form and the $^{13}C^{\delta 2}$ chemical shift has then a typical value of about 119 ppm. This shows that $^{13}C^{\delta 2}$ chemical shift in zinc-coordinated histidines constitute an extremely potent probe for the determination of the coordination mode of histidines. However, the chemical shift is a relative measure that depends strongly on correct calibration to a standard. Even if the standard procedure for calibrating chemical shifts of biomolecules is well documented (Wishart et al. 1995), correct calibration is not always properly conducted (Zhang et al. 2003). Incorrect chemical shift referencing could then potentially lead to a wrongly predicted coordination mode of histidines if the $^{13}C^{\delta 2}$ chemical shift were used exclusively. To circumvent this problem, we sought to establish a simple reference-independent parameter which will constitute a more robust probe for the determination of the coordination mode of histidines.

The chemical shift difference $C^{\varepsilon 1}$-$C^{\delta 2}$ is a reference independent marker of the coordination mode of histidines

As the $^{13}C^{\varepsilon 1}$ chemical shift is rather insensitive to the coordination mode of histidines (Figure 2 and Table 1), the chemical shift difference $C^{\varepsilon 1}$-$C^{\delta 2}$ retains all the information content of the $^{13}C^{\delta 2}$ chemical shift regarding the coordination mode of histidines, but in addition constitutes a reference independent marker of this coordination mode. The complete histogram showing the distribution of this chemical shift difference $\Delta_{\varepsilon\delta}$ is presented on Figure 3. In addition, statistical values are given in Table 2. Remarkably, average and median values are very close to each other for both coordination modes showing that both distributions are rather symmetrical which is also visible on the histogram plot (Figure 3). In case of an $N^{\varepsilon 2}$ coordination mode, $\Delta_{\varepsilon\delta}$ is distributed around an average value of 12.32 ± 0.82 ppm whereas for an $N^{\delta 1}$ coordination mode, $\Delta_{\varepsilon\delta}$ is found around 19.77 ± 1.49 ppm (Table 2). Altogether, we can formulate the following principles for the determination of the coordination mode of histidines. If the chemical shift difference between the ε1 and the δ2 aromatic carbons is larger than 17 ppm, the corresponding histidine should coordinate zinc through an $N^{\delta 1}$ coordination mode whereas if $\Delta_{\varepsilon\delta}$ is less than 17 ppm it should coordinate zinc via an $N^{\varepsilon 2}$ coordination mode (Figure 2b and 3). Remarkably, the two groups of histidines are clearly separated with almost no overlap between the $N^{\delta 1}$ and the $N^{\varepsilon 2}$ coordination modes. The present method would have allowed the determination of the correct coordination mode in more than 99 % of

the zinc-ligated histidines of the BMRB, and is therefore extremely potent. This shows that aromatic carbon chemical shifts in histidines are only slightly affected by the many different structural environments assumed to be present in our large database of zinc-ligated histidines. Precisely, the variations due to each local electronic environment remain smaller than the intrinsic chemical shift difference of the two different histidine tautomeric forms, which allow to predict the histidine coordination mode with an almost absolute confidence. Nevertheless, we want to mention that it is always possible that in case of a very special electronic environment, extremely shifted chemical shifts could be observed (see point 5 in Figure 2a).

Discussion

We show in the present report that ^{13}C aromatic carbon chemical shifts can be used to distinguish between the $N^{\delta 1}$ and the $N^{\varepsilon 2}$ coordination mode of histidines in zinc-binding proteins. This provide an alternative method to the currently used approaches for the determination of the coordination mode of histidines. This method has several advantages that will be discussed below.

NMR structure calculations with unlabelled protein samples of small zinc-finger proteins, obtained via chemical synthesis is a competitive approach, as compared with recombinant protein production, regarding the size of these small zinc-binding domains (ranging from about 20 to 60 residues). This approach is therefore quite common in the NMR community working with small zinc-knuckles or zinc-fingers (see for example references (Sharpe et al. 2002; Isernia et al. 2003; Dempsey et al. 2004; Sharpe et al. 2005; Matsui et al. 2007; Cordier et al. 2008; Bourbigot et al. 2008; Briknarová et al. 2011; Bazzi et al. 2011; Malgieri et al. 2011)). In these cases, the measurement of long-range ^{15}N-^{1}H correlation spectra is simply not possible regarding the very low natural abundance of ^{15}N nuclei. Therefore, the determination of the coordination mode of histidines in these studies relied on either preliminary structure calculations that should then be very precise or on previously determined close homologous structures. However, structures determined without heteronuclear labelling are in general less precise which could therefore lead to incorrect evaluation of the coordination mode. The method presented in this report, which is based on the aromatic ^{13}C chemical shift, is perfectly applicable to unlabelled protein samples. The aromatic carbon assignment could easily be obtained with a natural abundance ^{13}C-^{1}H HSQC together with $H^{\delta 2}$–$H^{\varepsilon 1}$ correlations in homonuclear 2D COSY or TOCSY spectra (King and Wright 1982; Miura and Ichikawa 1991; Xia et al. 1995). The coordination mode can then be determined with a direct measurement and not

with an indirect method making use of preliminary structure calculations. It also does not rely on previously published structures, therefore avoiding two approaches that could strongly bias the structure determination of new classes of zinc-binding domains.

Moreover, with the present method, the information on the histidine coordination mode is available directly after chemical shift assignment prior to any structure calculation. This could for example allow the early identification of an unusual zinc-binding site with an unconventional histidine coordination mode. Moreover, the discrimination revealed in this study could be used by structural and chemical shift databases as internal check for the determination of histidine coordination modes. If such a chemical shift based verification would be used for BMRB depositions, structures and chemical shifts with wrong coordination modes for zinc-ligated histidines could be prevented (see points 1-4 on Figure 2a). Remarkably, in the database we constructed, there were very few points with a wrong annotation of their coordination mode, most of which were actually correctly determined in terms of structure but wrongly annotated in the database. Overall, this indicates the high quality of the zinc-binding protein structures determined by NMR. This point was definitely crucial to conduct such chemical shift analysis that relied on the correct determination of the histidine coordination mode in the BMRB entries used for analysis.

Histidine side-chains contain an aromatic ring that could in principle have a strong influence on the chemical shift of neighbouring nuclei. Particularly, zinc-sites with one, two or three histidines in their coordination sphere might have significantly different histidine aromatic carbon shifts due to mutual ring-current effects. We analyzed this potential effect on $C^{\delta 2}$ and $C^{\varepsilon 1}$ chemical shifts but no clear correlations between the number of histidines and these carbon chemical shifts appeared in our analysis, indicating that this effect is too small to be significant.

Finally, one should emphasize once more that the particular downfield shifted $C^{\delta 2}$ is not due to the zinc coordination *per se,* but reflects the nature of the tautomeric form of the histidine. The observation of a particularly downfield shifted $C^{\delta 2}$ is not a signature of the presence of a zinc-ligated histidine. However, the $N^{\delta 1}$-H tautomeric form of neutral histidine is less common and is present only in particular cases where this tautomeric form is stabilized (Pelton et al. 1993; Day et al. 2003; Sudmeier et al. 2003; Schubert et al. 2007). To illustrate the relative abundance of the $N^{\delta 1}$-H tautomeric form compared to the $N^{\varepsilon 2}$-H tautomeric form, we constructed another chemical shift database by extracting aromatic chemical shifts of non-coordinated histidines from the Reference Corrected Database (RefDB) (Zhang et al. 2003).

This resulted in about 1000 histidines with both $C^{\varepsilon 1}$ and $C^{\delta 2}$ chemical shifts. These data are displayed on Figure 4 and compared with the ^{13}C aromatic chemical shifts of the zinc-coordinated histidines. Most histidines that are not coordinated to zinc have their $C^{\delta 2}$ chemical shift around 120 ppm and only a small proportion (about 5 %) display a downfield shifted $C^{\delta 2}$ chemical shift of more than 124 ppm. Importantly, $C^{\delta 2}$ chemical shifts of protonated histidines are also found around 120 ppm, similar to the $N^{\varepsilon 2}$-H tautomer (Sudmeier et al. 2003). Therefore, whereas a downfield shifted $C^{\delta 2}$ is a true signature of the $N^{\delta 1}$-H tautomeric form, ^{13}C aromatic chemical shifts do not allow to distinguish the $N^{\varepsilon 2}$-H tautomer from the protonated form of the histidine ring (Sudmeier et al. 2003).

Moreover, ^{13}C aromatic chemical shifts cannot distinguish between coordinated and non-coordinated histidines (Figure 4) and it is therefore important to mention that the present method allows to determine the histidine coordination mode, but the knowledge on which histidine within a protein sequence is coordinated to zinc must come from another source of information, *e.g.* preliminary structure calculations. Nonetheless, this information is also not provided by the other approaches aiming at determining the coordination mode of histidines, and must in any case be deduced from other data.

Remarkably, the $N^{\varepsilon 2}$ coordination mode which corresponds to the chemical shift signature of the $N^{\delta 1}$-H tautomeric form is very common in zinc-coordinated histidines (Figure 3). As a consequence, most histidines with a downfield shifted $C^{\delta 2}$ chemical shift are actually zinc-coordinated histidines and this can therefore explain the confusion in the literature about the origin of downfield shifted $C^{\delta 2}$ chemical shift. The present report with a comprehensive survey of the ^{13}C aromatic chemical shifts of histidines should definitely help clarify the relations between zinc-coordination, tautomeric form and downfield shifted $C^{\delta 2}$ in histidines.

Acknowledgment

We are grateful to Thomas Aeschbacher, Fionna Loughlin and Lawrence P. McIntosh for helpful discussions. Research in the Allain laboratory is supported by the Swiss National Science Foundation (Nr 31003E-131031) and the SNF-NCCR structural biology. P.B. was supported by the Postdoctoral ETH Fellowship Program. *Author contributions:* P.B. designed the project and analyzed the data; P.B., M.S. and F.H.-T.A. wrote the manuscript; all authors discussed the results and approved the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

Andreeva A, Murzin AG (2008) A fortuitous insight into a common mode of RNA recognition by the dsRNA-specific zinc fingers. Proc Natl Acad Sci U S A 105 (52):E128-129.

Andreini C, Banci L, Bertini I, Rosato A (2006a) Counting the zinc-proteins encoded in the human genome. J Proteome Res 5 (1):196-201.

Andreini C, Banci L, Bertini I, Rosato A (2006b) Zinc through the three domains of life. J Proteome Res 5 (11):3173-3178.

Auld DS (2001) Zinc coordination sphere in biochemical zinc sites. Biometals 14 (3-4):271-313

Barraud P, Emmerth S, Shimada Y, Hotz H-R, Allain FH-T, Bühler M (2011) An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. EMBO J 30 (20):4223-4235.

Bazzi A, Zargarian L, Chaminade F, Boudier C, De Rocquigny H, Rene B, Mely Y, Fosse P, Mauffret O (2011) Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyriboses in the binding polarity of NC. Nucleic Acids Res 39 (9):3903-3916.

Bessière D, Lacroix C, Campagne S, Ecochard V, Guillet V, Mourey L, Lopez F, Czaplicki J, Demange P, Milon A, Girard J-P, Gervais V (2008) Structure-function analysis of the THAP zinc finger of THAP1, a large C2CH DNA-binding module linked to Rb/E2F pathways. J Biol Chem 283 (7):4352-4363.

Bourbigot S, Ramalanjaona N, Boudier C, Salgado GF, Roques BP, Mely Y, Bouaziz S, Morellet N (2008) How the HIV-1 nucleocapsid protein binds and destabilises the (-) primer binding site during reverse transcription. J Mol Biol 383 (5):1112-1128.

Briknarová K, Thomas CJ, York J, Nunberg JH (2011) Structure of a zinc-binding domain in the Junin virus envelope glycoprotein. J Biol Chem 286 (2):1528-1536.

Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci U S A 104 (23):9615-9620.

Chakrabarti P (1990) Geometry of interaction of metal ions with histidine residues in protein structures. Protein Eng 4 (1):57-63

Cordier F, Vinolo E, Véron M, Delepierre M, Agou F (2008) Solution structure of NEMO zinc finger and impact of an anhidrotic ectodermal dysplasia with immunodeficiency-related point mutation. J Mol Biol 377 (5):1419-1432.

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13 (3):289-302

Day RM, Thalhauser CJ, Sudmeier JL, Vincent MP, Torchilin EV, Sanford DG, Bachovchin CW, Bachovchin WW (2003) Tautomerism, acid-base equilibria, and H-bonding of the six histidines in subtilisin BPN' by NMR. Protein Sci 12 (4):794-810.

De Guzman RN, Liu HY, Martinez-Yamout M, Dyson HJ, Wright PE (2000) Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP. J Mol Biol 303 (2):243-253.

Dempsey BR, Wrona M, Moulin JM, Gloor GB, Jalilehvand F, Lajoie G, Shaw GS, Shilton BH (2004) Solution NMR structure and X-ray absorption analysis of the C-terminal zinc-binding domain of the SecA ATPase. Biochemistry 43 (29):9361-9371.

Dupont CL, Butcher A, Valas RE, Bourne PE, Caetano-Anollés G (2010) History of biological metal utilization inferred through phylogenomic analysis of protein structures. Proc Natl Acad Sci U S A 107 (23):10567-10572.

Estrada DF, Boudreaux DM, Zhong D, St Jeor SC, De Guzman RN (2009) The Hantavirus Glycoprotein G1 Tail Contains Dual CCHC-type Classical Zinc Fingers. J Biol Chem 284 (13):8654-8660.

Eustermann S, Brockmann C, Mehrotra PV, Yang J-C, Loakes D, West SC, Ahel I, Neuhaus D (2010) Solution structures of the two PBZ domains from human APLF and their interaction with poly(ADP-ribose). Nat Struct Mol Biol 17 (2):241-243.

Eustermann S, Videler H, Yang J-C, Cole PT, Gruszka D, Veprintsev D, Neuhaus D (2011) The DNA-binding domain of human PARP-1 interacts with DNA single-strand breaks as a monomer through its second zinc finger. J Mol Biol 407 (1):149-170.

Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. J Biomol NMR 3 (2):185-204

Hansen DF, Kay LE (2011) Determining valine side-chain rotamer conformations in proteins from methyl 13C chemical shifts: application to the 360 kDa half-proteasome. J Am Chem Soc 133 (21):8272-8281.

Hansen DF, Neudecker P, Vallurupalli P, Mulder FAA, Kay LE (2010) Determination of Leu side-chain conformations in excited protein states by NMR relaxation dispersion. J Am Chem Soc 132 (1):42-43.

Hayes PL, Lytle BL, Volkman BF, Peterson FC (2008) The solution structure of ZNF593 from Homo sapiens reveals a zinc finger in a predominantly unstructured protein. Protein Sci 17 (3):571-576.

Huang A, de Jong RN, Wienk H, Winkler GS, Timmers HTM, Boelens R (2009) E2-c-Cbl recognition is necessary but not sufficient for ubiquitination activity. J Mol Biol 385 (2):507-519.

Isernia C, Bucci E, Leone M, Zaccaro L, Di Lello P, Digilio G, Esposito S, Saviano M, Di Blasio B, Pedone C, Pedone PV, Fattorusso R (2003) NMR structure of the single QALGGH zinc finger domain from the Arabidopsis thaliana SUPERMAN protein. Chembiochem 4 (2-3):171-180.

King G, Wright PE (1982) A two-dimensional NMR method for assignment of imidazole ring proton resonances of histidine residues in proteins. Biochem Biophys Res Commun 106 (2):559-565.

Kornhaber GJ, Snyder D, Moseley HNB, Montelione GT (2006) Identification of zinc-ligated cysteine residues based on 13Calpha and 13Cbeta chemical shift data. J Biomol NMR 34 (4):259-269.

Kostic M, Matt T, Martinez-Yamout MA, Dyson HJ, Wright PE (2006) Solution structure of the Hdm2 C2H2C4 RING, a domain critical for ubiquitination of p53. J Mol Biol 363 (2):433-450.

Kwon K, Cao C, Stivers JT (2003) A novel zinc snap motif conveys structural stability to 3-methyladenine DNA glycosylase I. J Biol Chem 278 (21):19442-19446.

Lee MS, Gippert GP, Soman KV, Case DA, Wright PE (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science 245 (4918):635-637

Legge GB, Martinez-Yamout MA, Hambly DM, Trinh T, Lee BM, Dyson HJ, Wright PE (2004) ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. J Mol Biol 343 (4):1081-1093.

Liew CK, Crossley M, Mackay JP, Nicholas HR (2007) Solution structure of the THAP domain from Caenorhabditis elegans C-terminal binding protein (CtBP). J Mol Biol 366 (2):382-390.

London RE, Wingad BD, Mueller GA (2008) Dependence of amino acid side chain 13C shifts on dihedral angle: application to conformational analysis. J Am Chem Soc 130 (33):11097-11105.

Malgieri G, Russo L, Esposito S, Baglivo I, Zaccaro L, Pedone EM, Di Blasio B, Isernia C, Pedone PV, Fattorusso R (2007) The prokaryotic Cys2His2 zinc-finger adopts a novel fold as revealed by the NMR structure of Agrobacterium tumefaciens Ros DNA-binding domain. Proc Natl Acad Sci U S A 104 (44):17341-17346.

Malgieri G, Zaccaro L, Leone M, Bucci E, Esposito S, Baglivo I, Del Gatto A, Russo L, Scandurra R, Pedone PV, Fattorusso R, Isernia C (2011) Zinc to cadmium replacement in the A. thaliana SUPERMAN Cys His zinc finger induces structural rearrangements of typical DNA base determinant positions. Biopolymers 95 (11):801-810.

Matsui T, Kodera Y, Endoh H, Miyauchi E, Komatsu H, Sato K, Tanaka T, Kohno T, Maeda T (2007) RNA recognition mechanism of the minimal active domain of the human immunodeficiency virus type-2 nucleocapsid protein. J Biochem 141 (2):269-277.

Miura S, Ichikawa Y (1991) Proton nuclear magnetic resonance investigation of adrenodoxin. Assignment of aromatic resonances and evidence for a conformational similarity with ferredoxin from Spirulina platensis. Eur J Biochem 197 (3):747-757

Möller HM, Martinez-Yamout MA, Dyson HJ, Wright PE (2005) Solution structure of the N-terminal zinc fingers of the Xenopus laevis double-stranded RNA-binding protein ZFa. J Mol Biol 351 (4):718-730.

Oh BH, Westler WM, Darba P, Markley JL (1988) Protein carbon-13 spin systems by a single two-dimensional nuclear magnetic resonance experiment. Science 240 (4854):908-911

Omichinski JG, Clore GM, Appella E, Sakaguchi K, Gronenborn AM (1990) High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution. Biochemistry 29 (40):9324-9334

Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. Protein Sci 2 (4):543-558.

Peroza EA, Schmucki R, Güntert P, Freisinger E, Zerbe O (2009) The beta(E)-domain of wheat E(c)-1 metallothionein: a metal-binding domain with a distinctive structure. J Mol Biol 387 (1):207-218.

Ramelot TA, Cort JR, Goldsmith-Fischman S, Kornhaber GJ, Xiao R, Shastry R, Acton TB, Honig B, Montelione GT, Kennedy MA (2004) Solution NMR structure of the iron-sulfur cluster assembly protein U (IscU) with zinc bound at the active site. J Mol Biol 344 (2):567-583.

Reynolds WF, Peat IR, Freedman MH, Lyerla J, J R (1973) Determination of the tautomeric form of the imidazole ring of L-histidine in basic solution by carbon-13 magnetic resonance spectroscopy. J Am Chem Soc 95 (2):328-331

Rosenzweig AC (2002) Metallochaperones: bind and deliver. Chem Biol 9 (6):673-677

Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. J Biomol NMR 24 (2):149-154

Schubert M, Poon DK, Wicki J, Tarling CA, Kwan EM, Nielsen JE, Withers SG, McIntosh LP (2007) Probing electrostatic interactions along the reaction pathway of a glycoside hydrolase: histidine characterization by NMR spectroscopy. Biochemistry 46 (25):7383-7395.

Sharma D, Rajarathnam K (2000) 13C NMR chemical shifts can predict disulfide bond formation. J Biomol NMR 18 (2):165-171

Sharpe BK, Liew CK, Kwan AH, Wilce JA, Crossley M, Matthews JM, Mackay JP (2005) Assessment of the robustness of a serendipitous zinc binding fold: mutagenesis and protein grafting. Structure 13 (2):257-266.

Sharpe BK, Matthews JM, Kwan AH, Newton A, Gell DA, Crossley M, Mackay JP (2002) A new zinc binding fold underlines the versatility of zinc binding modules in protein evolution. Structure 10 (5):639-648.

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44 (4):213-223.

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci U S A 105 (12):4685-4690.

Sudmeier JL, Bradshaw EM, Haddad KEC, Day RM, Thalhauser CJ, Bullock PA, Bachovchin WW (2003) Identification of histidine tautomers in proteins by 2D 1H/13C(delta2) one-bond correlated NMR. J Am Chem Soc 125 (28):8430-8431.

Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36 (Web Server issue):W496-502.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 6 (2):135-140

Wishart DS, Case DA (2002) Use of chemical shifts in macromolecular structure determination. Methods Enzymol 338:3-34

Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4 (2):171-180

Wuthrich K (1986) NMR of proteins and nucleic acids. John Wiley & Sons, New York

Xia B, Cheng H, Skjeldal L, Coghlan VM, Vickery LE, Markley JL (1995) Multinuclear magnetic resonance and mutagenesis studies of the histidine residues of human mitochondrial ferredoxin. Biochemistry 34 (1):180-187

Zeng L, Yap KL, Ivanov AV, Wang X, Mujtaba S, Plotnikova O, Rauscher r, Frank J, Zhou M-M (2008) Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. Nat Struct Mol Biol 15 (6):626-633.

Zhang H, Neal S, Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 25 (3):173-195

FIGURES LEGENDS

Figure 1

Chemical structure and atom nomenclature of an histidine side chain.

(**a**) Chemical structure and atom nomenclature for the two coordination modes of zinc-ligated histidines, namely the $N^{\delta 1}$ and the $N^{\varepsilon 2}$ coordination modes. (**b**) Chemical structure and characteristic ^{15}N chemical shifts for the two tautomeric form of neutral histidines, namely the $N^{\varepsilon 2}$-H and the $N^{\delta 1}$-H tautomeric form.

Figure 2

Two dimensional plots of $^{13}C^{\delta 2}$ and $^{13}C^{\varepsilon 1}$ chemical shifts of zinc-coordinated histidines as a function of the coordination mode.

(**a**) Initial database prior to any manual corrections. $N^{\delta 1}$ coordinated histidines are shown as black diamonds and $N^{\varepsilon 2}$ coordinated histidines as open circles. Numbered data points were manually checked for the consistency between the deposited structure and the annotated coordination mode in the chemical shifts database. In case of inconsistencies, these data points were corrected to give the final database represented on panel B. Five outliers corresponding to BMRB entries 6374, 11061, 10108, 6682 and 5668 are marked with numbers 1 to 5, respectively. (**b**) Final corrected database showing a clear distinction between the two coordination modes of histidines. The solid line with its corresponding equation ($C^{\varepsilon 1}$ - $C^{\delta 2}$ = 17 ppm) delineate the two clusters of the $N^{\delta 1}$ and the $N^{\varepsilon 2}$ coordination modes. The corresponding statistical values are listed in Table 1.

Figure 3

Number of occurrence of each histidine–zinc coordination mode as a function of the chemical shift difference $\Delta_{\varepsilon\delta}$ ($\Delta_{\varepsilon\delta} = \delta\{^{13}C^{\varepsilon 1}\} - \delta\{^{13}C^{\delta 2}\}$), plotted in 0.5 ppm intervals. The statistical values are listed in Table 2.

Figure 4

Two dimensional plots of $^{13}C^{\delta 2}$ and $^{13}C^{\varepsilon 1}$ chemical shifts for non-coordinated histidines (small open circles), for $N^{\delta 1}$ coordinated histidines (black diamonds) and for $N^{\varepsilon 2}$ coordinated histidines (large open circles).

TABLES

Table 1

Summary of the statistical values of the $^{13}C^{\delta 2}$ and $^{13}C^{\varepsilon 1}$ chemical shifts for 266 histidines coordinating zinc via $N^{\delta 1}$ or $N^{\varepsilon 2}$

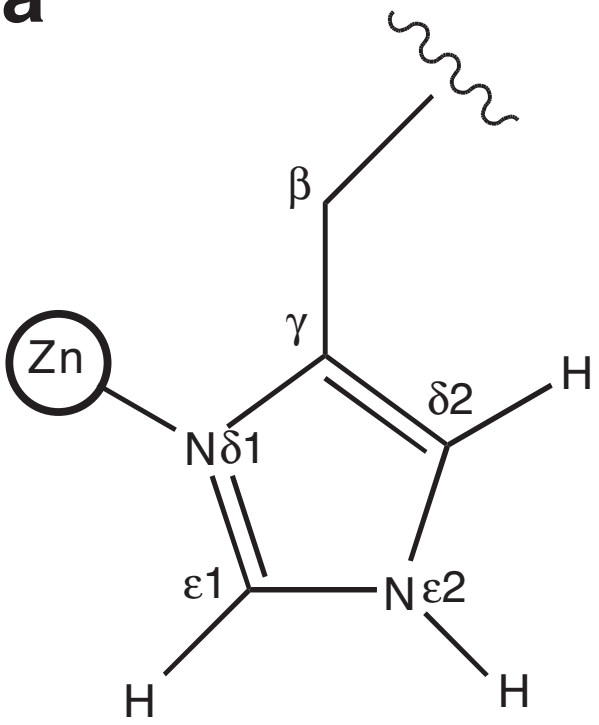
	$N^{\delta 1}$ coordination		$N^{\varepsilon 2}$ coordination	
	$^{13}C^{\delta 2}$ His	$^{13}C^{\varepsilon 1}$ His	$^{13}C^{\delta 2}$ His	$^{13}C^{\varepsilon 1}$ His
Number of shifts	43	43	223	223
Average value $\bar{\delta}$ (ppm)	119.09	138.86	127.42	139.74
Min value δ_{min} (ppm)	116.04	135.37	122.79	137.18
Max value δ_{max} (ppm)	123.85	142.25	132.37	143.1
Standard deviation σ_{δ}(ppm)	1.54	1.30	0.75	0.59

Table 2

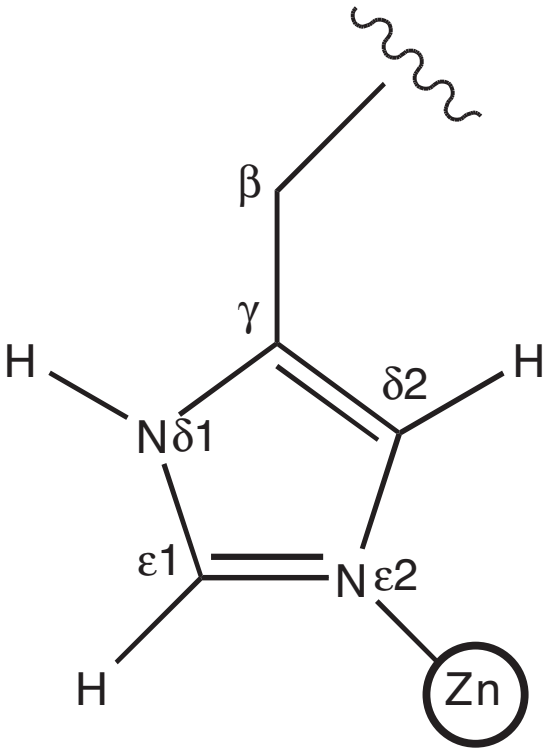
Statistical value of the chemical shift difference $\Delta_{\varepsilon\delta}$ ($\Delta_{\varepsilon\delta} = \delta\{^{13}C^{\varepsilon 1}\} - \delta\{^{13}C^{\delta 2}\}$) as a function of the coordination mode of histidines.

	$N^{\delta 1}$ coordination	$N^{\varepsilon 2}$ coordination
Number of shifts	43	223
Average value $\bar{\Delta}$ (ppm)	19.77	12.32
Median value $\tilde{\Delta}$ (ppm)	19.66	12.38
Standard deviation σ_{Δ} (ppm)	1.49	0.82

a

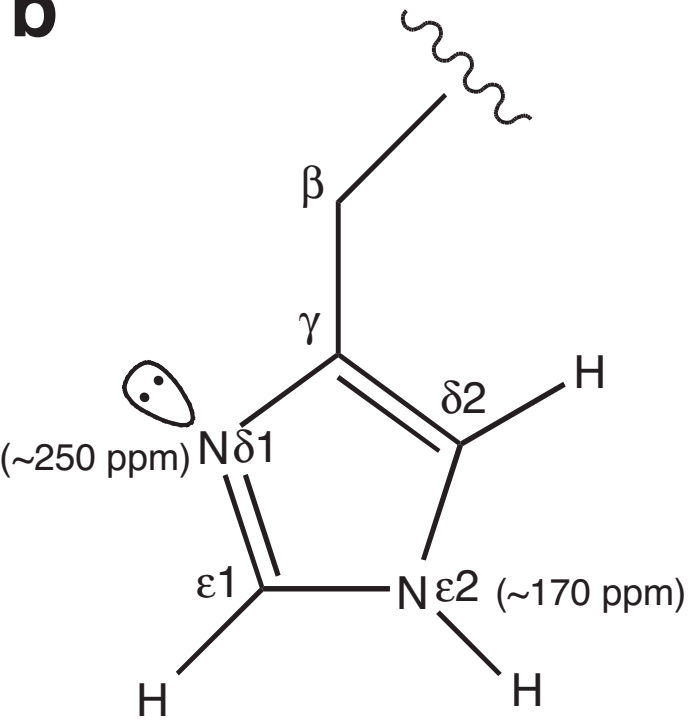


Nδ1 coordination

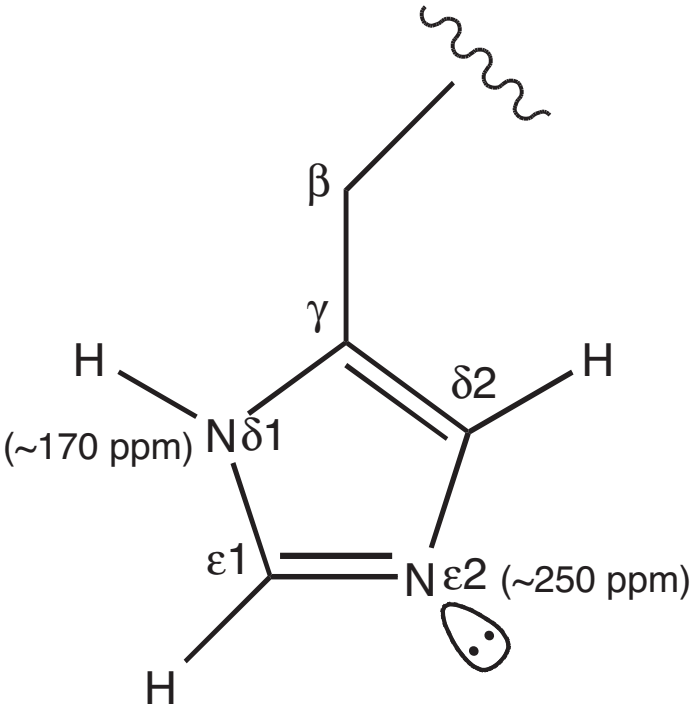


Nε2 coordination

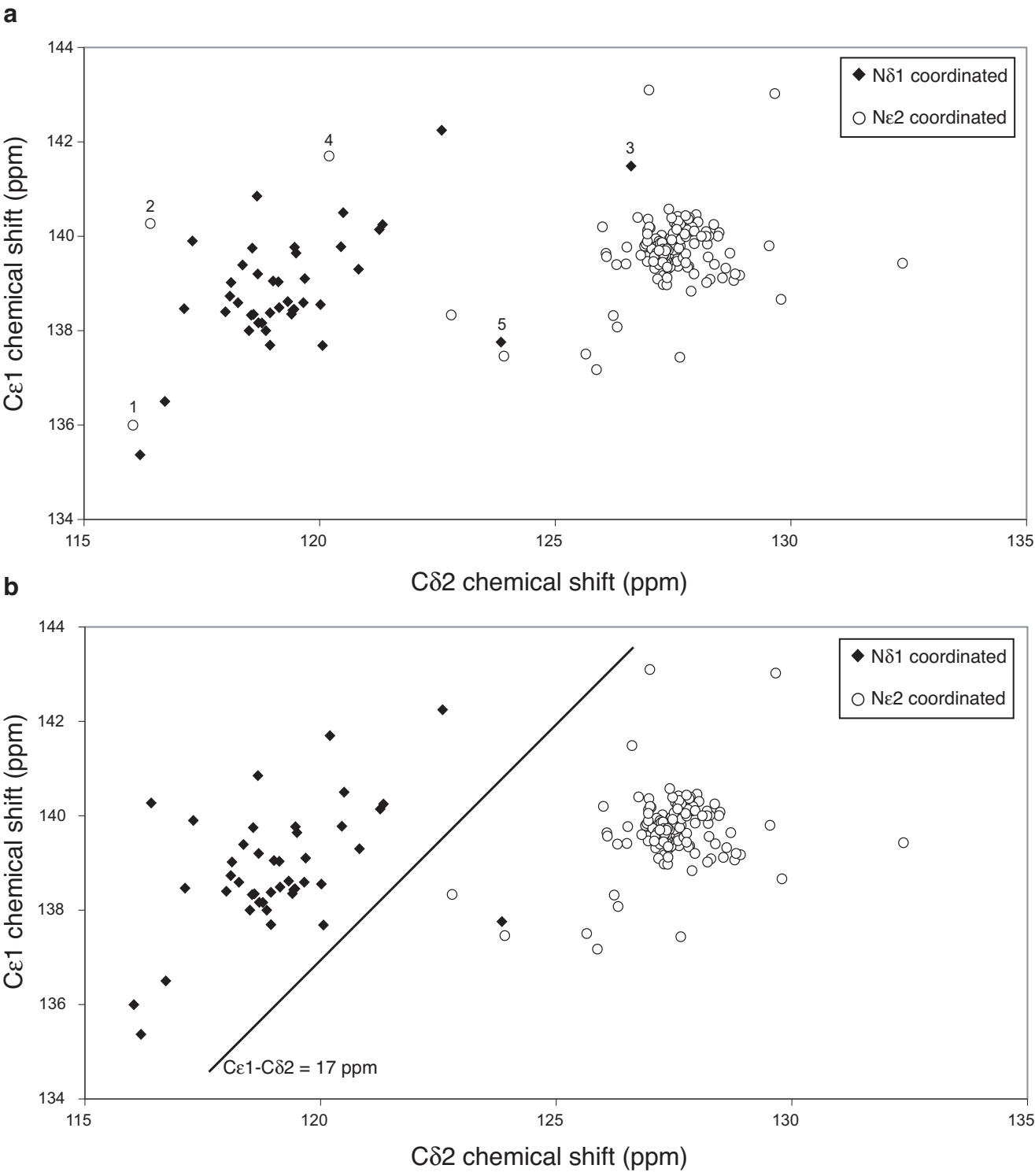
b

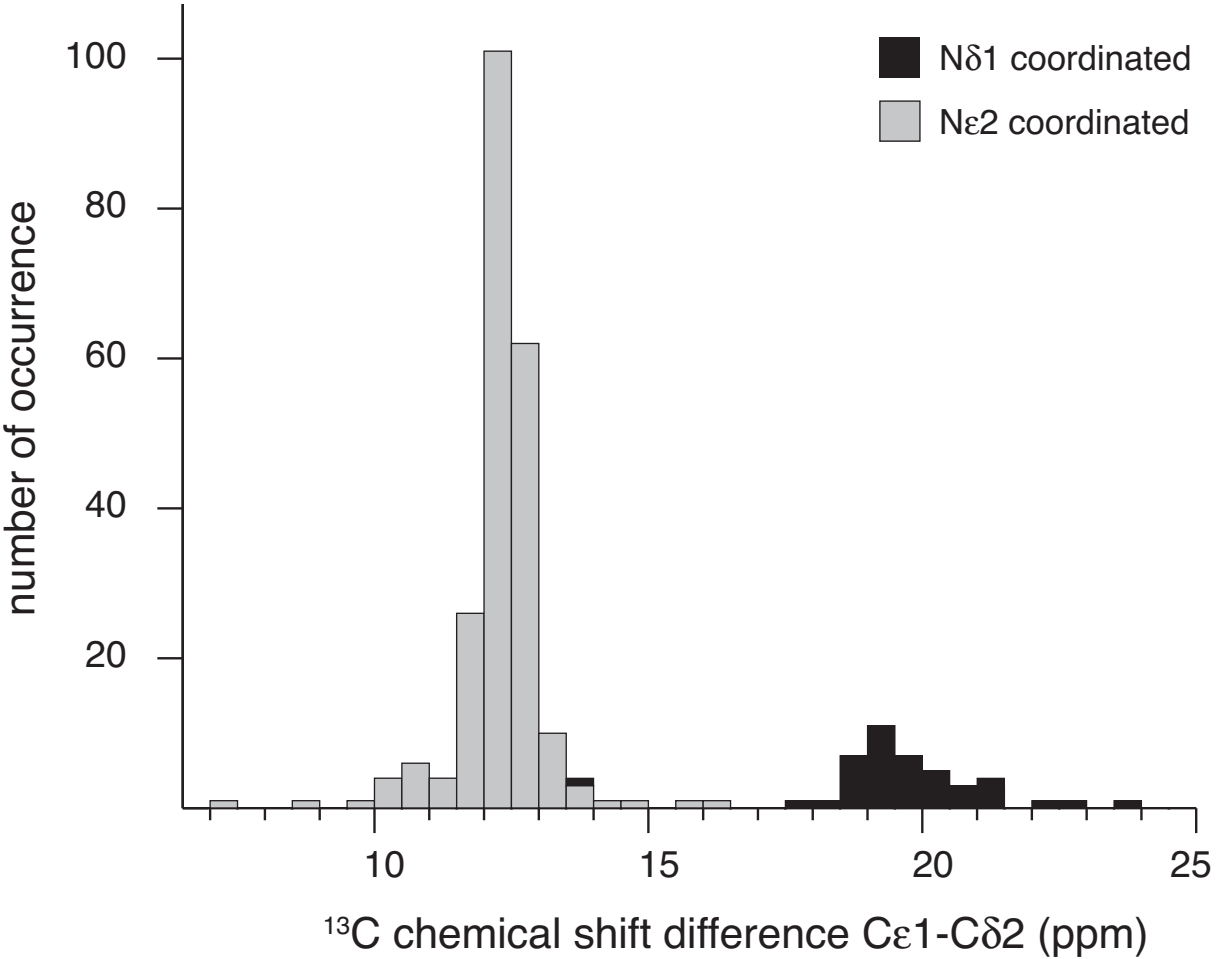


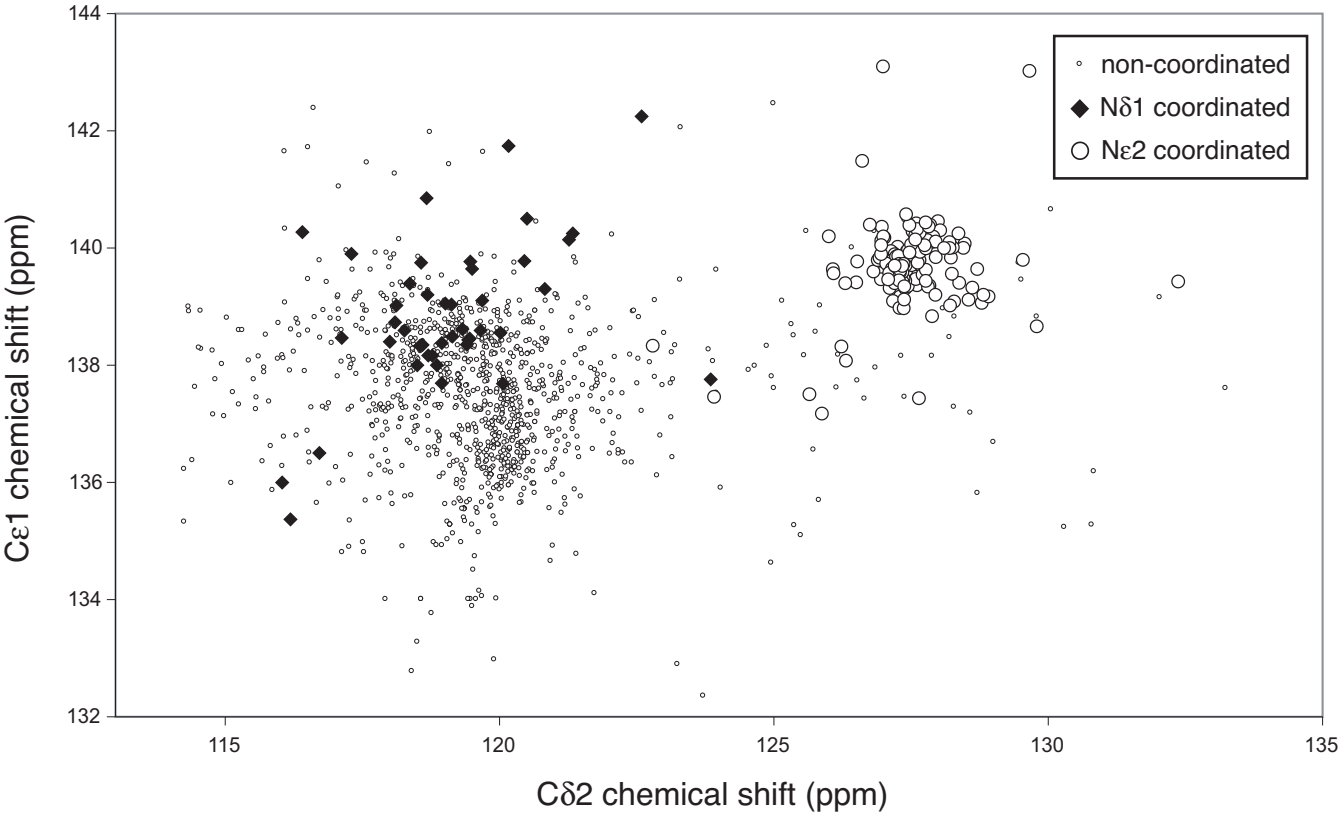
Nε2-H tautomer



Nδ1-H tautomer


a
Nδ1 coordinated
Nε2 coordinated
1
2
3
4
5
Cε1 chemical shift (ppm)
Cδ2 chemical shift (ppm)
144
142
140
138
136
134
115
120
125
130
135
b
Nδ1 coordinated
Nε2 coordinated
Cε1-Cδ2 = 17 ppm
Cε1 chemical shift (ppm)
Cδ2 chemical shift (ppm)


Nδ1 coordinated
Nε2 coordinated
number of occurrence
100
80
60
40
20
10
15
20
25
13C chemical shift difference Cε1-Cδ2 (ppm)


non-coordinated
Nδ1 coordinated
Nε2 coordinated
Cε1 chemical shift (ppm)
Cδ2 chemical shift (ppm)
144
142
140
138
136
134
132
115
120
125
130
135

Supplementary Material

A strong ^{13}C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins

Pierre Barraud, Mario Schubert & Frédéric H.-T. Allain

Supplementary Table 1
Complete chemical shift database used in this study ordered by BMRB entry codes and by histidine coordination modes.

[a] Histidine residue number as reported in the BMRB entry. See also the note regarding histidine numbering in C_2H_2 zinc-fingers at the end of the table.
[b] Topology of the zinc-site in which this histidine is involved. If several histidines are coordinating the same zinc ion, the corresponding histidine is underlined in the motif.

BMRB entry	Protein name	Residue range	Histidine residue number [a]	Zinc site [b]	Cδ2 (ppm)	Cε1 (ppm)	Cε1 - Cδ2 (ppm)	coordination mode
4928	WSTF-PHD	1-51	26	CCHC	120.50	140.50	20.00	Nδ1
5668	m^3A DNA glycosylase I	1-187	175	CHHC	123.85	137.76	13.91	Nδ1
6238	CBP-ZZ	1-52	42	CCHH	118.00	138.40	20.40	Nδ1
6374	AIRE-PHD	1-66	31	CCHC	116.04	136.00	19.96	Nδ1
6658	FLIX3	1-182	144	CCCH	120.02	138.55	18.54	Nδ1
6682	hZNF593	1-116	75	CCHH	120.20	141.70	21.50	Nδ1
7174	Hdm2-RING	429-491	457	CCHH	118.67	140.85	22.18	Nδ1
7210	ING4-PHD	1-63	37	CCHC	118.95	137.69	18.75	Nδ1
7297	RPA1320	1-102	63	CCHH	117.30	139.90	22.60	Nδ1
7297	RPA1320	1-102	7	CCHD	118.70	138.17	19.47	Nδ1
10013	Kiaa1045-RING	1-89	40	CCHC	119.01	139.05	20.04	Nδ1
10098	hC1-PKC-delta	1-84	28	HCCC	120.45	139.78	19.33	Nδ1
10098	hC1-PKC-delta	1-84	67	CCHC	119.46	138.45	19.00	Nδ1
10108	At1g60420	1-89	72	CCHC	119.68	139.10	19.42	Nδ1
10108	At1g60420	1-89	38	HHCC	120.06	137.69	17.62	Nδ1
10247	NEDD9-LIM	1-82	39	CCHC	119.43	138.43	19.00	Nδ1
10247	NEDD9-LIM	1-82	71	CCCH	118.36	139.39	21.03	Nδ1
11036	KAP1-PHD	624-812	648	CCHC	118.27	138.59	20.33	Nδ1
11061	APOBEC2	1-192	60	HCC	116.41	140.27	23.87	Nδ1
11093	ITK-BTK	1-64	16	HCCC	118.95	138.38	19.43	Nδ1
11151	C1-PKC-delta	1-83	18	HCCC	121.34	140.25	18.91	Nδ1
11151	C1-PKC-delta	1-83	56	CCHC	119.32	138.62	19.29	Nδ1
11157	BMX-BTK	1-50	16	HCCC	122.59	142.25	19.66	Nδ1
11161	Tmp30-RING	1-85	39	CHCC	118.78	138.16	19.39	Nδ1
11162	ZNF126-RING	1-78	38	CHCC	118.60	138.35	19.75	Nδ1
11162	ZNF126-RING	1-78	41	CCHC	118.10	138.73	20.64	Nδ1
11163	Tmp5-RING	1-85	39	CHCC	118.55	138.33	19.78	Nδ1
11164	TNF3-RING	1-66	35	CHCC	119.14	138.49	19.35	Nδ1
15059	MLP-LIM	1-194	31	CCCH	119.40	138.35	18.95	Nδ1
15059	MLP-LIM	1-194	141	CCCH	118.12	139.02	20.90	Nδ1
15112	C2-SOD	20-173	92	HHHD	116.72	136.50	19.79	Nδ1
15112	C2-SOD	20-173	101	HHHD	119.47	139.77	20.30	Nδ1
15112	C2-SOD	20-173	109	HHHD	117.12	138.47	21.34	Nδ1
15208	Churchill	2-107	71	HHCC	118.57	139.75	21.18	Nδ1
15796	E2-c-Cbl	355-437	398	CHCC	116.19	135.37	19.18	Nδ1
15948	E3-Arkadia-RING	1-69	40	CCHC	118.50	138.00	19.50	Nδ1
15948	E3-Arkadia-RING	1-69	37	CHCC	118.86	138.00	19.14	Nδ1
17112	C1B-PKC-delta	214-295	231	HCCC	120.82	139.30	18.48	Nδ1
17112	C1B-PKC-delta	214-295	269	CCHC	119.50	139.64	20.14	Nδ1
17113	C1A-PKC-delta	149-228	159	HCCC	121.27	140.14	18.88	Nδ1
17113	C1A-PKC-delta	149-228	197	CCHC	119.12	139.03	19.91	Nδ1
17157	PARP-1-ZNF1	1-108	53	CCHC	118.69	139.20	20.52	Nδ1
17158	PARP-1-ZNF2	103-214	159	CCHC	119.65	138.59	18.94	Nδ1

BMRB entry	Protein name	Residue range	Histidine residue number [a]	Zinc site [b]	Cδ2 (ppm)	Cε1 (ppm)	Cε1 - Cδ2 (ppm)	coordination mode
5668	m^3A DNA glycosylase I	1-187	17	C<u>H</u>HC	126.23	138.32	12.09	Nε2
5987	CBP TAZ1-Zn1	1-100	23	HCCC	128.71	139.64	10.94	Nε2
5987	CBP TAZ1-Zn2	1-100	54	HCCC	128.82	139.20	10.39	Nε2
5987	CBP TAZ1-Zn3	1-100	78	HCCC	127.34	139.69	12.36	Nε2
6005	TIS11d-TZF-Zn1	1-70	28	CCCH	123.91	137.46	13.55	Nε2
6005	TIS11d-TZF-Zn2	1-70	66	CCCH	122.79	138.33	15.54	Nε2
6238	CBP-ZZ	1-52	40	CC<u>H</u>H	126.00	140.20	14.20	Nε2
6275	MTF-1-ZNF1	1-177	28	CCH<u>H</u>	127.20	139.40	12.20	Nε2
6275	MTF-1-ZNF2	1-177	58	CCH<u>H</u>	128.20	140.10	11.90	Nε2
6275	MTF-1-ZNF3	1-177	88	CCH<u>H</u>	128.20	140.00	11.80	Nε2
6275	MTF-1-ZNF4	1-177	117	CCH<u>H</u>	127.20	139.70	12.50	Nε2
6275	MTF-1-ZNF6	1-177	177	CCH<u>H</u>	126.30	139.40	13.10	Nε2
6682	hZNF593	1-116	69	CC<u>H</u>H	128.10	140.00	11.90	Nε2
7174	Hdm2-RING	429-491	452	CC<u>H</u>H	128.21	139.02	10.81	Nε2
7194	TFIIIA-zf46-ZNF1	1-87	22	CC<u>H</u>H	127.77	139.44	11.67	Nε2
7194	TFIIIA-zf46-ZNF1	1-87	26	CCH<u>H</u>	127.28	139.45	12.17	Nε2
7194	TFIIIA-zf46-ZNF2	1-87	52	CC<u>H</u>H	127.58	140.15	12.57	Nε2
7194	TFIIIA-zf46-ZNF2	1-87	56	CCH<u>H</u>	127.38	139.35	11.97	Nε2
7194	TFIIIA-zf46-ZNF3	1-87	80	CC<u>H</u>H	126.09	139.57	13.48	Nε2
7194	TFIIIA-zf46-ZNF3	1-87	85	CCH<u>H</u>	127.08	139.47	12.38	Nε2
10038	hZNF295-Zn1	1-107	35	CCHC	128.79	139.06	10.28	Nε2
10038	hZNF295-Zn2	1-107	61	CCHC	128.38	139.41	11.03	Nε2
10038	hZNF295-Zn3	1-107	88	CC<u>H</u>H	127.13	139.65	12.52	Nε2
10038	hZNF295-Zn3	1-107	93	CCH<u>H</u>	127.30	139.54	12.25	Nε2
10057	IscU	1-130	101	CDCH	129.54	139.80	10.26	Nε2
10108	At1g60420	1-89	35	<u>H</u>HCC	126.61	141.49	14.88	Nε2
10148	Osr2-ZNF1	1-106	36	CC<u>H</u>H	128.24	139.56	11.32	Nε2
10148	Osr2-ZNF1	1-106	40	CCH<u>H</u>	128.48	140.08	11.60	Nε2
10148	Osr2-ZNF2	1-106	64	CC<u>H</u>H	127.02	139.55	12.53	Nε2
10148	Osr2-ZNF2	1-106	68	CCH<u>H</u>	127.99	140.46	12.47	Nε2
10148	Osr2-ZNF3	1-106	92	CC<u>H</u>H	127.30	139.17	11.88	Nε2
10148	Osr2-ZNF3	1-106	96	CCH<u>H</u>	127.26	139.86	12.61	Nε2
10151	hZNF268	887-919	31	CC<u>H</u>H	127.26	139.43	12.17	Nε2
10151	hZNF268	887-919	35	CCH<u>H</u>	127.05	139.70	12.65	Nε2
10152	hZNF268	495-525	31	CC<u>H</u>H	127.39	139.31	11.92	Nε2
10152	hZNF268	495-525	35	CCH<u>H</u>	126.99	139.79	12.79	Nε2
10153	hZNF484	519-551	31	CC<u>H</u>H	127.26	140.02	12.77	Nε2
10153	hZNF484	519-551	35	CCH<u>H</u>	127.95	139.85	11.90	Nε2
10154	hZNF473	557-589	31	CC<u>H</u>H	127.61	139.37	11.76	Nε2
10154	hZNF473	557-589	35	CCH<u>H</u>	127.86	140.20	12.34	Nε2
10155	hZNF484	603-635	31	CC<u>H</u>H	126.98	139.84	12.86	Nε2
10155	hZNF484	603-635	35	CCH<u>H</u>	127.85	140.40	12.55	Nε2
10156	hZNF484	715-747	31	CC<u>H</u>H	127.41	139.62	12.20	Nε2
10156	hZNF484	715-747	35	CCH<u>H</u>	127.88	140.10	12.22	Nε2
10157	B-cell lymphoma 6 protein	626-654	30	CC<u>H</u>H	127.34	139.43	12.09	Nε2
10157	B-cell lymphoma 6 protein	626-654	34	CCH<u>H</u>	128.03	140.30	12.27	Nε2
10158	hZNF473	370-400	31	CC<u>H</u>H	127.44	139.35	11.91	Nε2
10158	hZNF473	370-400	35	CCH<u>H</u>	127.09	139.77	12.68	Nε2
10166	hZNF268	637-667	29	CC<u>H</u>H	127.17	139.92	12.75	Nε2
10166	hZNF268	637-667	33	CCH<u>H</u>	127.50	139.86	12.36	Nε2

BMRB entry	Protein name	Residue range	Histidine residue number [a]	Zinc site [b]	Cδ2 (ppm)	Cε1 (ppm)	Cε1 - Cδ2 (ppm)	coordination mode
10167	hZNF28-homolog	584-616	31	CCHH	127.27	139.60	12.33	Nε2
10167	hZNF28-homolog	584-616	35	CCHH	127.60	140.06	12.46	Nε2
10168	hZNF28-homolog	440-472	31	CCHH	127.22	139.81	12.58	Nε2
10168	hZNF28-homolog	440-472	35	CCHH	127.01	140.18	13.17	Nε2
10169	hZNF224	171-203	31	CCHH	127.09	139.75	12.66	Nε2
10169	hZNF224	171-203	35	CCHH	127.51	139.98	12.47	Nε2
10170	hZNF28-homolog	415-447	31	CCHH	127.12	139.54	12.42	Nε2
10170	hZNF28-homolog	415-447	35	CCHH	127.61	139.72	12.11	Nε2
10171	hZNF224	311-343	31	CCHH	127.22	139.57	12.36	Nε2
10171	hZNF224	311-343	35	CCHH	127.91	140.19	12.29	Nε2
10172	hZNF224	395-427	31	CCHH	128.55	139.12	10.57	Nε2
10172	hZNF224	395-427	35	CCHH	127.42	139.95	12.52	Nε2
10173	hZNF347	592-624	31	CCHH	127.22	139.68	12.45	Nε2
10173	hZNF347	592-624	35	CCHH	127.70	140.02	12.32	Nε2
10174	hZNF347	340-372	31	CCHH	127.32	139.66	12.35	Nε2
10174	hZNF347	340-372	35	CCHH	127.70	140.06	12.35	Nε2
10175	hZNF28-homolog	556-588	31	CCHH	127.03	139.77	12.73	Nε2
10175	hZNF28-homolog	556-588	35	CCHH	127.53	140.29	12.76	Nε2
10176	hZNF28-homolog	596-728	31	CCHH	128.28	139.09	10.80	Nε2
10176	hZNF28-homolog	596-728	35	CCHH	127.78	140.11	12.33	Nε2
10177	hZNF484	687-719	31	CCHH	127.25	139.49	12.24	Nε2
10177	hZNF484	687-719	35	CCHH	127.63	140.12	12.49	Nε2
10178	hZNF473	204-236	31	CCHH	127.17	139.10	11.93	Nε2
10178	hZNF473	204-236	35	CCHH	126.95	139.57	12.62	Nε2
10179	hZNF28-homolog	724-756	31	CCHH	127.59	139.47	11.89	Nε2
10179	hZNF28-homolog	724-756	35	CCHH	127.04	139.87	12.83	Nε2
10180	hZNF224	339-371	31	CCHH	127.34	139.55	12.21	Nε2
10180	hZNF224	339-371	35	CCHH	127.55	139.97	12.42	Nε2
10181	hZNF224	367-399	31	CCHH	127.20	139.82	12.62	Nε2
10181	hZNF224	367-399	35	CCHH	127.51	139.97	12.46	Nε2
10182	hZNF473	725-757	31	CCHH	127.35	139.68	12.33	Nε2
10182	hZNF473	725-757	35	CCHH	128.36	140.25	11.89	Nε2
10183	hZNF484	379-411	31	CCHH	127.26	139.53	12.27	Nε2
10183	hZNF484	379-411	35	CCHH	127.73	140.10	12.37	Nε2
10184	hZNF484	463-495	31	CCHH	127.32	139.57	12.25	Nε2
10184	hZNF484	463-495	35	CCHH	127.64	140.13	12.49	Nε2
10185	hZNF268	607-639	31	CCHH	127.39	139.44	12.05	Nε2
10185	hZNF268	607-639	35	CCHH	126.96	139.87	12.91	Nε2
10186	hZNF95-homolog	369-401	31	CCHH	126.07	139.64	13.57	Nε2
10186	hZNF95-homolog	369-401	35	CCHH	127.51	139.96	12.46	Nε2
10187	hZNF224	479-511	31	CCHH	127.27	139.84	12.57	Nε2
10187	hZNF224	479-511	35	CCHH	127.57	139.97	12.40	Nε2
10188	hZNF347	564-596	31	CCHH	127.30	139.63	12.32	Nε2
10188	hZNF347	564-596	35	CCHH	127.80	140.07	12.27	Nε2
10189	hZNF95-homolog	341-373	31	CCHH	126.52	139.77	13.25	Nε2
10189	hZNF95-homolog	341-373	35	CCHH	128.45	140.00	11.55	Nε2
10190	hZNF95-homolog	425-457	31	CCHH	127.29	138.98	11.69	Nε2
10190	hZNF95-homolog	425-457	35	CCHH	127.06	139.77	12.71	Nε2
10191	hZNF268	719-751	31	CCHH	127.16	139.64	12.48	Nε2
10191	hZNF268	719-751	35	CCHH	127.65	139.99	12.34	Nε2
10192	hZNF224	283-315	31	CCHH	126.95	139.47	12.52	Nε2
10192	hZNF224	283-315	35	CCHH	127.80	140.14	12.34	Nε2

BMRB entry	Protein name	Residue range	Histidine residue number [a]	Zinc site [b]	Cδ2 (ppm)	Cε1 (ppm)	Cε1 - Cδ2 (ppm)	coordination mode
10196	hZNF95-homolog	768-800	31	CCHH	127.27	139.46	12.19	Nε2
10196	hZNF95-homolog	768-800	35	CCHH	127.66	140.00	12.34	Nε2
10197	hZNF224	199-231	31	CCHH	127.13	139.91	12.78	Nε2
10197	hZNF224	199-231	35	CCHH	127.61	140.37	12.76	Nε2
10198	hZNF224	423-455	31	CCHH	126.50	139.41	12.91	Nε2
10198	hZNF224	423-455	35	CCHH	127.64	139.92	12.28	Nε2
10199	hZNF347	312-344	31	CCHH	127.10	139.67	12.57	Nε2
10199	hZNF347	312-344	35	CCHH	127.61	140.00	12.39	Nε2
10200	hZNF473	342-372	31	CCHH	126.96	139.80	12.84	Nε2
10200	hZNF473	342-372	35	CCHH	126.96	140.13	13.17	Nε2
10201	hZNF473	641-673	31	CCHH	127.15	139.57	12.42	Nε2
10201	hZNF473	641-673	35	CCHH	127.62	140.03	12.42	Nε2
10202	hZNF484	491-523	31	CCHH	127.26	139.53	12.27	Nε2
10202	hZNF484	491-523	35	CCHH	127.59	140.42	12.83	Nε2
10203	hZNF484	547-579	31	CCHH	127.22	139.54	12.31	Nε2
10203	hZNF484	547-579	35	CCHH	127.57	140.08	12.52	Nε2
10204	hZNF28-homolog	612-644	31	CCHH	127.19	139.68	12.49	Nε2
10204	hZNF28-homolog	612-644	35	CCHH	127.69	139.99	12.30	Nε2
10205	hZNF28-homolog	668-700	31	CCHH	127.25	139.76	12.51	Nε2
10205	hZNF28-homolog	668-700	35	CCHH	127.44	139.97	12.52	Nε2
10206	hZNF28-homolog	780-812	31	CCHH	127.14	139.72	12.59	Nε2
10206	hZNF28-homolog	780-812	35	CCHH	127.37	138.97	11.61	Nε2
10207	hZNF347	368-400	31	CCHH	127.17	139.61	12.44	Nε2
10207	hZNF347	368-400	35	CCHH	127.10	139.71	12.62	Nε2
10208	hZNF473	315-345	31	CCHH	127.22	139.57	12.35	Nε2
10208	hZNF473	315-345	35	CCHH	127.63	139.93	12.30	Nε2
10209	hZNF473	809-841	31	CCHH	127.16	139.78	12.63	Nε2
10209	hZNF473	809-841	35	CCHH	127.58	140.21	12.64	Nε2
10210	hZNF28-homolog	528-560	31	CCHH	126.81	139.60	12.79	Nε2
10210	hZNF28-homolog	528-560	35	CCHH	127.66	140.26	12.61	Nε2
10216	hZNF268	411-441	31	CCHH	127.37	139.12	11.75	Nε2
10216	hZNF268	411-441	35	CCHH	127.08	139.76	12.67	Nε2
10217	hZNF268	693-723	29	CCHH	127.83	139.36	11.53	Nε2
10217	hZNF268	693-723	33	CCHH	127.26	139.85	12.59	Nε2
10218	hZNF268	355-385	31	CCHH	127.31	139.45	12.14	Nε2
10218	hZNF268	355-385	35	CCHH	126.92	139.82	12.90	Nε2
10219	hZNF268	441-469	29	CCHH	127.22	139.49	12.27	Nε2
10219	hZNF268	441-469	33	CCHH	128.23	139.84	11.61	Nε2
10220	hZNF95-homolog	397-429	31	CCHH	127.38	139.71	12.33	Nε2
10220	hZNF95-homolog	397-429	35	CCHH	127.61	140.32	12.72	Nε2
10221	hZNF224	255-287	31	CCHH	127.15	139.56	12.41	Nε2
10221	hZNF224	255-287	35	CCHH	127.59	139.82	12.23	Nε2
10222	hZNF268	775-807	31	CCHH	127.81	139.34	11.54	Nε2
10222	hZNF268	775-807	35	CCHH	127.14	139.84	12.69	Nε2
10223	hZNF347	760-792	31	CCHH	127.24	139.66	12.42	Nε2
10223	hZNF347	760-792	35	CCHH	127.74	140.02	12.28	Nε2
10224	hZNF268	859-889	31	CCHH	127.25	139.75	12.50	Nε2
10224	hZNF268	859-889	35	CCHH	127.13	139.95	12.82	Nε2
10225	hZNF268	301-331	29	CCHH	127.11	139.32	12.21	Nε2
10225	hZNF268	301-331	33	CCHH	127.58	140.00	12.42	Nε2

BMRB entry	Protein name	Residue range	Histidine residue number [a]	Zinc site [b]	Cδ2 (ppm)	Cε1 (ppm)	Cε1 - Cδ2 (ppm)	coordination mode
10226	hZNF268	273-303	29	CCHH	127.71	139.55	11.84	Nε2
10226	hZNF268	273-303	33	CCHH	127.20	139.89	12.69	Nε2
10227	hZNF268	551-583	31	CCHH	127.16	139.80	12.63	Nε2
10227	hZNF268	551-583	35	CCHH	126.89	139.79	12.90	Nε2
10228	hZNF95-homolog	628-660	31	CCHH	127.36	139.48	12.12	Nε2
10228	hZNF95-homolog	628-660	35	CCHH	127.24	139.79	12.55	Nε2
10229	hZNF268	385-413	29	CCHH	127.35	139.42	12.08	Nε2
10229	hZNF268	385-413	33	CCHH	127.95	140.11	12.16	Nε2
10230	hZNF224	563-595	31	CCHH	127.25	139.87	12.62	Nε2
10230	hZNF224	563-595	35	CCHH	127.57	139.87	12.30	Nε2
10231	B-cell lymphoma 6 protein	598-626	30	CCHH	127.19	139.63	12.43	Nε2
10231	B-cell lymphoma 6 protein	598-626	34	CCHH	127.82	140.37	12.55	Nε2
10232	hZNF95-homolog	796-828	31	CCHH	126.97	140.36	13.40	Nε2
10232	hZNF95-homolog	796-828	35	CCHH	126.93	139.84	12.91	Nε2
10233	hZNF347	284-316	31	CCHH	127.22	139.64	12.42	Nε2
10233	hZNF347	284-316	35	CCHH	127.61	140.06	12.45	Nε2
10234	hZNF347	648-680	31	CCHH	127.23	139.85	12.63	Nε2
10234	hZNF347	648-680	35	CCHH	127.70	139.99	12.29	Nε2
10243	hZNF24	1-72	33	CCHH	127.17	139.61	12.44	Nε2
10243	hZNF24	1-72	37	CCHH	127.53	139.91	12.37	Nε2
10243	hZNF24	1-72	65	CCHH	127.51	139.71	12.20	Nε2
10244	hZNF462	1-88	44	CCHH	127.14	139.64	12.49	Nε2
10244	hZNF462	1-88	48	CCHH	127.47	140.39	12.92	Nε2
10246	CTCF-ZNF1	1-86	34	CCHH	127.88	138.84	10.96	Nε2
10246	CTCF-ZNF1	1-86	39	CCHH	128.91	139.17	10.26	Nε2
10246	CTCF-ZNF2	1-86	66	CCHC	128.62	139.32	10.71	Nε2
10257	BED-protein 2	1-76	54	CCHH	126.75	140.40	13.65	Nε2
10257	BED-protein 2	1-76	59	CCHH	127.47	139.81	12.34	Nε2
10303	hZNF28-homolog	752-784	31	CCHH	127.16	139.63	12.47	Nε2
10303	hZNF28-homolog	752-784	35	CCHH	127.77	140.44	12.67	Nε2
10304	hZNF95-homolog	544-576	31	CCHH	127.17	139.68	12.52	Nε2
10304	hZNF95-homolog	544-576	35	CCHH	127.64	139.99	12.35	Nε2
10305	hZNF347	536-568	31	CCHH	127.05	139.74	12.69	Nε2
10305	hZNF347	536-568	35	CCHH	127.69	140.03	12.34	Nε2
10306	B-cell lymphoma 6 protein	518-541	29	CCHH	127.41	140.58	13.17	Nε2
10306	B-cell lymphoma 6 protein	518-541	34	CCHH	127.77	139.63	11.86	Nε2
10307	hZNF473	426-458	31	CCHH	127.15	139.65	12.50	Nε2
10307	hZNF473	426-458	35	CCHH	127.57	140.07	12.51	Nε2
10308	hZNF473	484-512	29	CCHH	126.95	139.90	12.94	Nε2
10308	hZNF473	484-512	33	CCHH	127.27	139.87	12.60	Nε2
10309	hZNF484	771-803	31	CCHH	127.30	139.66	12.36	Nε2
10309	hZNF484	771-803	35	CCHH	127.66	140.03	12.37	Nε2
10310	hZNF347	396-428	31	CCHH	127.27	139.42	12.16	Nε2
10310	hZNF347	396-428	35	CCHH	127.65	139.79	12.14	Nε2
11130	hZNF512	1-96	81	CCHH	125.87	137.18	11.30	Nε2
11130	hZNF512	1-96	86	CCHH	125.65	137.51	11.86	Nε2
11165	hZNF268	329-359	29	CCHH	127.10	139.58	12.48	Nε2
11165	hZNF268	329-359	33	CCHH	127.58	140.12	12.55	Nε2

BMRB entry	Protein name	Residue range	Histidine residue number [a]	Zinc site [b]	Cδ2 (ppm)	Cε1 (ppm)	Cε1 - Cδ2 (ppm)	coordination mode
11166	hZNF268	581-609	29	CCHH	127.40	139.71	12.31	Nε2
11166	hZNF268	581-609	33	CCHH	127.62	139.75	12.13	Nε2
11167	hZNF347	508-540	31	CCHH	127.66	139.94	12.27	Nε2
11167	hZNF347	508-540	35	CCHH	127.34	139.59	12.24	Nε2
11168	hZNF347	732-764	31	CCHH	127.77	139.99	12.23	Nε2
11168	hZNF347	732-764	35	CCHH	127.30	139.58	12.29	Nε2
11169	hZNF484	435-467	31	CCHH	128.26	140.01	11.75	Nε2
11169	hZNF484	435-467	35	CCHH	127.30	139.62	12.32	Nε2
11170	hZNF484	631-663	31	CCHH	127.75	140.05	12.30	Nε2
11170	hZNF484	631-663	35	CCHH	127.28	139.73	12.45	Nε2
11171	hZNF484	659-691	31	CCHH	127.36	139.73	12.37	Nε2
11171	hZNF484	659-691	35	CCHH	132.37	139.43	7.06	Nε2
15002	CtBP-THAP	1-91	62	CCCH	126.99	140.20	13.21	Nε2
15074	Parkin-IBR	1-80	69	CCCH	127.94	139.20	11.26	Nε2
15160	Polymerase-eta-UBZ domain	1-39	27	CCHH	127.50	140.06	12.56	Nε2
15160	Polymerase-eta-UBZ domain	1-39	31	CCHH	126.99	143.10	16.11	Nε2
15161	ETO-MYND	658-707	695	CCCH	129.79	138.66	8.87	Nε2
15208	Churchill	2-107	66	CCCH	129.66	143.02	13.37	Nε2
15208	Churchill	2-107	59	HHCC	126.96	140.05	13.10	Nε2
16596	hAPLF-PBZ	363-451	392	CCHH	126.31	138.08	11.77	Nε2
16596	hAPLF-PBZ	363-451	398	CCHH	127.47	139.92	12.45	Nε2
16596	hAPLF-PBZ	363-451	434	CCHH	127.64	137.44	9.79	Nε2
16596	hAPLF-PBZ	363-451	440	CCHH	127.66	139.94	12.27	Nε2

Note regarding histidine numbering in C_2H_2 zinc-fingers:
A large number of entries of histidine coordinating zinc via Nε2 come from C_2H_2 zinc fingers solved by the RIKEN Structural Genomics Initiative. Proteins containing the classical C_2H_2 zinc-finger domain are amongst the most abundant in eukaryotic genomes, and RIKEN chose to determine structures of a large number of those C_2H_2 zinc-fingers. The procedure used for residue numbering in these proteins was such that many entries have identical histidine sequence positions (31 and 35). As this could be misinterpreted as duplicated entries in our database, we wanted to make clear here that each of these entries come from truly unique protein domains.