



Published in final edited form as:

J Biol Chem. 2003 May 30; 278(22): 20225–20234. doi:10.1074/jbc.M300459200.

Structure and Ubiquitin Interactions of the Conserved Zinc Finger Domain of Npl4*

Bin Wang^{‡,§}, Steven L. Alam^{‡,§}, Hemmo H. Meyer^{¶,||}, Marielle Payne[‡], Timothy L. Stemmler^{**}, Darrell R. Davis^{‡,††,§§}, and Wesley I. Sundquist^{‡,¶¶}

[‡]Department of Biochemistry, University of Utah, Salt Lake City, Utah 84132

^{‡‡}Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84132

^{**}Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, Michigan 48201

^{¶¶}Department of Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06520-8002

Abstract

Ubiquitylated proteins are directed into a large number of different cellular pathways through interactions with effector proteins that contain conserved ubiquitin binding motifs. Here, we report the solution structure and ubiquitin binding properties of one such motif, the Npl4 zinc finger or RanBP2/Nup358 zinc finger (NZF) domain. Npl4 NZF forms a compact module composed of four antiparallel β -strands linked by three ordered loops. A single zinc ion is coordinated by four conserved cysteines from the first and third loops, which form two rubredoxin knuckles. Npl4 NZF binds specifically, but weakly, to free ubiquitin using a conserved ¹³TF¹⁴ dipeptide to interact with the “Ile-44” surface of ubiquitin. Our studies reveal the structure of this versatile class of protein binding domains and provide a means for identifying the subset of NZF domains likely to bind ubiquitin.

Protein ubiquitylation plays an important role in a large number of biological processes, including intracellular proteolysis, DNA repair, transcription, translation, signal transduction, cell cycle progression, organelle assembly, protein trafficking, and virus budding (1–3). Cells therefore dedicate an extensive array of machinery to the enzymology of ubiquitin transfer. Upon ubiquitylation, proteins must be directed to the correct intracellular locale, and pathways that utilize ubiquitin as a targeting signal therefore have

*The Utah Biomolecular NMR Facility is supported by grants from the National Institutes of Health and the National Science Foundation (NIH RR13030 and NSF DBI-0002806).

^{||}Supported by Grants NIH 1R01 GM60919 and HFSP-RG0183/1999M from the NIH and the Human Frontiers Science Project (to Graham Warren).

^{§§}Supported by NIH Grant NIH R01 GM55508.

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

To whom correspondence may be addressed: Depts. of Medicinal Chemistry and Biochemistry, University of Utah, Salt Lake City, UT 84132. Tel.: 801-581-7006; Fax: 801-581-7087; davis@adenosine.pharm.utah.edu. ^{¶¶}To whom correspondence may be addressed: Dept. of Biochemistry, University of Utah, Salt Lake City, UT 84132. Tel.: 801-585-5402; Fax: 801-581-7959; wes@biochem.utah.edu.

[§]These authors contributed equally to the structure determination.

The atomic coordinates and structure factors (code 1NJ3) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

The NMR assignments for this protein are available in the BioMagnetic Resonance Bank (BMRB) database under BMRB accession number 5053 (www.bmrw.wise.edu).

effector proteins that specifically recognize ubiquitylated proteins. A number of such ubiquitin binding proteins are now known, and their study has led to the identification and characterization of several conserved ubiquitin binding motifs, including the ubiquitin-interacting motif (UIM),¹ the ubiquitin-associated domain, and the ubiquitin E2 variant/UBC-like domain (2, 4–7). Each of these motifs forms an independent folding domain that can bind ubiquitin *in vitro*, although the molecular details of ubiquitin recognition are not yet understood in any case.

Given the generality of ubiquitin (Ub) as a targeting signal, it is likely that additional ubiquitin recognition motifs remain to be identified and characterized. One example is the recently discovered Npl4 zinc finger/RanBP2 zinc finger domain (NZF) (see Fig. 1A), which was first identified in Npl4, a protein that binds ubiquitin *in vitro* and mediates several different ubiquitin-dependent functions of the AAA-ATPase p97/Cdc48p. A hallmark of NZF domains is the presence of four regularly spaced cysteine residues, which suggests that the motif may be a metal binding module (8–11). NZF motifs appear to be quite prevalent, and our protein data base searches revealed 243 unique peptide sequences with four cysteines and an asparagine with the same spacing as those of Npl4 NZF (Fig. 1A). This ensemble of sequences also exhibited significant conservation at 12 of 24 of the remaining residues, which presumably reflects the conserved structure and function(s) of this common domain.

In addition to Npl4 itself, several other NZF-containing proteins are known to be involved in ubiquitin-dependent processes. One example is Vps36p, a protein required for vacuolar protein sorting of ubiquitylated proteins in yeast (11). Isolated NZF domains from both Npl4 and Vps36p bind mono- and poly-Ub chains *in vitro*, consistent with the idea that these NZF domains serve to recognize ubiquitylated protein substrates *in vivo* (10). However, not all NZF-like domains bind ubiquitin, suggesting that those that do may represent a specialized subset of a much larger domain superfamily that shares a common three-dimensional fold (10).

The prevalence of the NZF domain and its apparent role in ubiquitin recognition in several important biological pathways led us to study the molecular basis for NZF/Ub interactions. Toward this end, we have analyzed the sequence conservation across putative NZF domains, characterized the metal and ubiquitin binding properties of the Npl4 and Vps36p NZF domains, determined the solution structure of the NZF domain from Npl4, mapped the interaction surfaces in the Npl4 NZF/Ub complex, and tested the importance of a conserved, dipeptide motif (TF) at NZF positions 13 and 14 for ubiquitin binding.

EXPERIMENTAL PROCEDURES

Data Base Searches for Putative NZF Domains

The SwissProt, PIR, PRF, and KEGG GENE data bases were searched with the pattern $X_6CX_2CX_3NX_6CX_2CX_5$, which yielded 243 unique matches in 166 different proteins (duplicates were removed manually). The apparent correlation between Asn-16 and Trp-7 (72%; see Fig. 1A) was tested by a search with the pattern $X_4WXCX_2CX_{10}CX_2CX(5)$, which produced 161 matches (in 114 proteins), 84% of which had asparagine at the variable position corresponding to Asn-16 in Npl4 NZF. Searches with the more stringent RanBP2-like pattern (12) (and see, on the World Wide Web,

¹The abbreviations used are: UIM, ubiquitin-interacting motif; Ub, ubiquitin; NZF, Npl4 zinc finger; BME, β -mercaptoethanol; GST, glutathione *S*-transferase; EXAFS, extended x-ray absorption fine structure; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; HSQC, heteronuclear single-quantum coherence; E2, ubiquitin conjugating enzyme; E3, ubiquitin-protein isopeptide ligase.

www.expasy.ch/cgi-bin/nicedoc.pl?PDOC50199) $X_4WXCX_2CX_3NX_6CX_2CX_5$, with both the Trp and Asn positions fixed, revealed greater conservation at the variable positions but was judged overly stringent, since it excluded the known NZF domain of Vps36p (10). Searches of the PDBSTR data base yielded no matches, indicating that the motif has not previously been characterized structurally.

Protein Expression and Purification

DNA encoding NZF domains from rat Npl4 (amino acids 580–608) and yeast Vps36p (amino acids 177–205) were cloned into pGEX-4T expression vectors. These constructs encoded GST-NZF fusion proteins with thrombin cleavage sites between the GST and NZF domains (13). Mutations were introduced into the Npl4 expression construct by QuikChange™ mutagenesis (Stratagene).

Protein expression in DH5α *Escherichia coli* carrying the expression plasmids was induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside ($A_{600} = 0.4$). After 4 h at 23 °C, the cells were harvested by centrifugation and stored at –70 °C prior to protein purification. All steps in Npl4 NZF protein purification were performed at 4 °C, except where noted. Pellets from 6 liters of cells were resuspended in 40 ml of buffer A (10 mM sodium phosphate, pH 7.4, 150 mM NaCl, 5 mM β-mercaptoethanol (BME), 10 μM ZnCl₂), sonicated to lyse the cells, and centrifuged for 1 h at 39,000 × *g* to clear insoluble debris. The soluble GST-Npl4 NZF fusion protein was purified by affinity chromatography on an FF 16/10 glutathione-Sepharose column (Amersham Biosciences). The bound protein was loaded and washed with ~10 column volumes of buffer A and then eluted with 20 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0, 5 mM BME, 10 μM ZnCl₂.

Purified fractions (~24 ml) were pooled and dialyzed for 16 h against 2 liters of thrombin cleavage buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM BME, 2.5 mM CaCl₂, 10 μM ZnCl₂), quantified by optical absorption ($\epsilon_{280} = 47,330 \text{ M}^{-1} \text{ cm}^{-1}$), incubated at 23 °C for 16 h with 1 unit of thrombin protease (Novagen)/mg of GST-Npl4 NZF, and then concentrated to ~3 ml by ultrafiltration (Amicon 3). Cleaved Npl4 NZF was separated from GST and uncleaved fusion protein by size exclusion chromatography (Superdex-75, APBiotech) in 100 mM NaCl, 20 mM sodium phosphate, pH 8.0, 5 mM BME, 10 μM ZnCl₂. This procedure typically yielded ~4 mg of pure Npl4 NZF. ¹⁵N- and ¹⁵N/¹³C-labeled NZF peptides were expressed and purified as described above except that *E. coli* was grown in M9 minimal medium with ¹⁵NH₄Cl or ¹⁵NH₄Cl/¹³C₆-glucose as the sole nitrogen and carbon source, respectively. Prior to atomic absorption and extended x-ray absorption fine structure (EXAFS) spectroscopic analyses, unbound zinc was removed by gel filtration of pure Npl4 NZF (Superdex-75) in a degassed buffer lacking zinc (20 mM sodium phosphate pH 5.5, 150 mM NaCl, 5 mM BME).

Thrombin cleavage left two nonnative residues at the N terminus of Npl4 NZF (NH₂-Gly-Ser₂). These residues are included in our numbering scheme but are not shown in any of the figures. Purified Npl4 NZF was characterized by SDS-PAGE, N-terminal sequencing (NH₂-Gly-Ser-Thr-Ser-Ala), and matrix-assisted laser desorption ionization mass spectrometry (calculated mass = 3374.7 Da; observed mass = 3374.7 Da). The stoichiometry of bound zinc was analyzed by determining the protein concentration using optical spectroscopy ($\epsilon_{280} = 5840 \text{ M}^{-1} \text{ cm}^{-1}$), and the zinc concentration was analyzed by flame atomic absorption spectrometry. Unlabeled and ¹⁵N-labeled ubiquitin were expressed and purified as described (14).

X-ray Absorption Spectroscopy

Pure Npl4 NZF was dialyzed against 2 liters of XAS buffer (either 20 mM sodium phosphate, pH 5.5, 50 mM NaCl, 5 mM BME or 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM BME), concentrated by ultrafiltration (Amicon 3), adjusted to a final concentration of 30% (w/v) glycerol, and snap-frozen in a Lucite cell covered in Kapton tape at a final protein concentration of 1 mM.

Data were collected at beamline 7-3 at the Stanford Synchrotron Radiation Laboratory using an Si(220) double crystal monochromator detuned 50% for harmonic rejection. Fluorescence excitation spectra were recorded using a 30-element germanium solid-state array detector. Samples were maintained at 10 K using a continuous flow liquid helium cryostat (Oxford). XAS spectra were measured at 5-eV steps in the pre-edge, 0.35-eV steps in the edge (9640–9690 eV), and 0.05-Å⁻¹ increments in the EXAFS region, integrating from 1–30 s in a k^3 weighted manner over 45 min. EXAFS data represent the average of nine scans. X-ray energies were calibrated with simultaneous measurement of a zinc foil absorption spectrum, assigning the first inflection point to 9659 eV. Similar spectra were obtained at both pH 5.5 and 8.0.

XAS data were analyzed using the Macintosh OS X version of the EXAFSPAK program, which integrates Feff version 7.0 to generate theoretical models (see, on the World Wide Web, www-ssrl.slac.stanford.edu/exafspak.html). Data reduction employed second-order polynomial base-line flattening in the pre-edge and three-region cubic spline flattening throughout the EXAFS region. Data were converted to k space using $E_0 = 9680$ eV. Meaningful EXAFS data were limited to 10,325 eV due to monochromator imperfections. EXAFS data of crystallographically characterized model complexes were fit using amplitude and phase functions calculated in Feff version 7.0 for a zinc-sulfur interaction. This resulted in a scale factor of 1 and an E_0 of -15.25 , which were used for data fitting of the NZF protein. Coordination numbers were fixed at half-integer values, whereas R and σ_2 were allowed to float during the final fits.

NMR Spectroscopy

Sample Preparation

Samples for structure determinations were ~1.0 mM Npl4 NZF in NMR buffer (20 mM sodium phosphate, pH 5.5, 50 mM NaCl, 5 mM BME, 10 μ M ZnCl₂, in 90% H₂O, 10% D₂O). Samples were degassed and flame-sealed under argon in NMR sample tubes to reduce cysteine oxidation.

Data Collection and Assignments

NMR spectra were recorded at 18 °C on a Varian Inova 600-MHz spectrometer equipped with a tripleresonance ¹H/¹³C/¹⁵N probe and z axis pulsed field gradients. Backbone and side chain assignments were made using a standard suite of triple resonance experiments (15, 16), except that two-dimensional versions of the HNCACB and HN(CA)CO experiments were collected and analyzed (Fig. 2). Side chain assignments were completed using H(CCO)NHTOCSY, (H)C(CCO)NH-TOCSY, ¹⁵N-edited TOCSY, and ¹³C-edited NOESY experiments, with aromatic resonances assigned using a combination of ¹H/¹³C HSQC and ¹³C-edited NOESY experiments centered on the aromatic carbon resonances (125 ppm). Stereospecific assignments for 28 of 36 β -methylene protons were obtained using a combination of HNHB, HN(CO)HB (17), ¹⁵N-edited TOCSY, and NOESY data. Resonance assignments were complete except for amide proton and nitrogens for the first three amino acids, methionine methyls, and His H ϵ 1 aromatic protons.

Three-dimensional ^{15}N -edited NOESY-HSQC (18, 19) and three-dimensional ^{13}C -edited NOESY-HSQC (20, 21) (140-ms mixing times) were used to generate distance restraints. Three-bond coupling constants ($^3J_{\text{HN-HA}}$) were obtained from a three-dimensional HNHA experiment (22). Hydrogen-bonded amides were identified using a long range HNCO experiment with the N-CO coupling period tuned to 133.3 ms ($\sim 1/4 \ ^1J_{\text{NCO}}$ for the one bond coupling (23)). Amide temperature coefficients ($\Delta\delta/\Delta T$) (24–26) were calculated from the observed shifts in $^1\text{H}/^{15}\text{N}$ -HSQC spectra recorded at 5, 10, 15, 20, 25, and 30 °C. Surface-exposed amides for the free and ubiquitin-bound forms of NZF were characterized in SEAhsqc spectra (27) tuned to a water-amide exchange time of 100 ms. All spectra were processed using FELIX 97 (MSI).

Structure Determination

Npl4 NZF structures were refined using torsion angle dynamics in DYANA (28) and then regularized in CNS (29) through a gentle simulated annealing with 362 NOE interproton restraints, four hydrogen-bonding restraints, and 13 J_{HNHA} scalar coupling restraints (30). Initial rounds of refinement used only NOE data, which defined the general fold of the domain and revealed the stereochemistry about the coordinated zinc. Final refinements added EXAFS-derived restraints for zinc-S γ (2.33–2.37 Å), (zinc-C β 3.25–3.61 Å), and S γ -S γ distances (3.78–4.15 Å) to ensure approximate tetrahedral zinc coordination as well as hydrogen-bonding restraints and $^3J_{\text{HNHA}}$ coupling constants.

NOE assignments and structure calculations were independently validated using the automated assignment/structure calculation package (CANDID within CYANA) (31), which produced essentially the same structure as was determined manually. Structures were analyzed using PROCHECK-NMR (32), MOLMOL (33), and INSIGHT II (MSI) (Table I). Structure figures were created with PYMOL (DeLano Scientific).

Chemical Shift Perturbation Mapping Experiments

Chemical shift perturbation experiments were performed at 18 °C in NMR buffer. To identify the ubiquitin binding sites on the Npl4 NZF domain, unlabeled ubiquitin was titrated into 1.5 mM ^{15}N -labeled Npl4 NZF at final concentrations of 0–1.5 mM. To identify the Npl4 NZF binding site on ubiquitin, unlabeled Npl4 NZF was titrated into 0.15 mM ^{15}N -labeled ubiquitin at final concentrations of 0–0.15 mM. $^1\text{H}/^{15}\text{N}$ HSQC spectra (18) were collected at each titration point, and normalized chemical shift changes (δ) were calculated using the equation $\delta = 25((\delta_{\text{HN}})^2 + (\delta_{\text{N}}/5)^2)^{0.5}$ (34, 35). Amide chemical shift assignments for human ubiquitin were obtained from the VLI Research, Inc. Web site at www.vli-research.com/ubshifts.htm.

Ubiquitin Binding

Ubiquitin binding affinities for wild type and mutant GST-Npl4 NZF proteins and wild type GST-Vps36p NZF captured on anti-GST antibody biosensor surfaces were quantified as described previously (36). All measurements were performed at 20 °C in 25 mM Tris-HCl, pH 7.0, 5 mM BME, 2 μM ZnCl₂.

RESULTS

The NZF Is a Zinc Binding Module

The Npl4 and Vps36p NZF constructs used in our studies span the conserved NZF regions and the sequence and numbering scheme for the rat Npl4 NZF polypeptide (Npl4 NZF) are given in Fig. 1A. Npl4 NZF was expressed in *E. coli*, and the soluble protein was purified to homogeneity under nondenaturing conditions (see “Experimental Procedures”). The presence of a single, covalently bound zinc ion was demonstrated using atomic absorption

spectroscopy (0.93 mol of zinc/mol of Npl4 NZF) and EXAFS spectroscopy (Fig. 1, *B* and *C*). The best fit model to the transformed EXAFS spectrum revealed that the zinc coordination sphere contained four sulfur ligands ($n = 4.56$) and an average Zn–S bond length of 2.33 Å. The Debye-Waller factor (0.00454 Å²) was consistent with a slightly distorted tetrahedral geometry about the zinc center. We therefore conclude that Npl4 NZF coordinates a single, tetrahedral Zn²⁺ ion using the four conserved cysteines.

NMR Studies of Npl4 NZF

Nearly complete resonance and stereospecific assignments for Npl4 NZF were obtained using standard ¹H/¹³C/¹⁵N heteronuclear NMR experiments (see Fig. 2 and “Experimental Procedures”). As expected for a Zn(Cys)₄ metal center, the four cysteine Cβ resonances exhibited chemical shifts (31.5–32.5 ppm) that were intermediate between free cysteine (28.3 ppm) and oxidized cysteine (41.2 ppm). The structure of Npl4 NZF was initially calculated using 362 NOE restraints, which were sufficient to define both the protein structure and the zinc coordination geometry. Subsequent refinements included restraints for observed hydrogen bonds (eight restraints), *J*_{H_NH_A} scalar-coupling (13 restraints), and tetrahedral zinc coordination (4 experimental, 10 idealized restraints). The final ensemble of 20 low energy structures was of high quality (Fig. 3 and Table I), with root mean square deviations from the mean structure of 0.11 ± 0.05 (backbone) and 0.72 ± 0.15 Å (all heavy atoms).

Structure of Npl4 NZF

Npl4 NZF contains four short β-strands (S1–S4) that form two orthogonal hairpins (S1/S2 and S3/S4). The S1/S2 hairpin exhibits a canonical β-hairpin hydrogen bonding pattern, and the three expected hydrogen bonds were observed directly in HNC(O)-LR spectra (23). In contrast, the S3/S4 “hairpin” consists of two twisted, antiparallel strands that lack canonical cross-strand hydrogen bonding. The four strands are connected by a compact central loop (L2) that contains overlapping type-1 (Gln-17–Thr-20) and reverse-γ (Gly-19–Gly-21) turns and two short metal binding loops (L1 and L3) that form “rubredoxin knuckles” (Figs. 3 and 4) (37–39).

The rubredoxin knuckle was initially recognized in the iron-binding centers of rubredoxins and has subsequently been observed at the zinc coordination sites in a number of proteins (39–42). Rubredoxin knuckles are six-residue loops that connect two β-strands and display two metal binding Cys ligands (from loop positions 1 and 4). The canonical knuckle configuration orients the backbone amide nitrogens at positions 3 and 4 to hydrogen-bond with the first cysteine Sγ atom and orients the backbone amide at position 6 to hydrogen-bond with the second cysteine Sγ. This unusually tight metal binding loop is favored by a glycine residue at position 5, which adopts a positive backbone φ torsion angle. Variations in the details of this hydrogen bonding scheme are seen in proteins within the rubredoxin family, for example in rubrerythrin (Protein Data Bank entry 1DVB) (43).

The knuckles in Npl4 NZF are slightly different from canonical rubredoxin knuckles, because both NZF knuckles lack glycines at position 5 (Figs. 1*A*, 3*C*, and 4). This opens up the loops slightly and alters the predicted hydrogen bonding pattern (see Fig. 3*C* and legend). Moreover, in the second knuckle, the side chain amide nitrogen of Asn-16 donates a hydrogen bond to the first cysteine (Cys-23, loop position 1). Nevertheless, the overall trajectories in the two Npl4 knuckles are similar to one another and to other rubredoxin knuckles (Fig. 4).

As summarized in Table II, nearly all of the residues that are conserved across NZF domains perform identifiable structural roles. The NZF tertiary structure is stabilized by a small

hydrophobic “core,” which lies just below the zinc and consists primarily of the conserved Trp-7 residue (Figs. 1A and 3). The Trp-7 He1 proton is hydrogen-bonded to the Thr-20 carbonyl oxygen of strand 3, and the indole ring makes a series of hydrophobic contacts with residues from both strands S1 and S4. One face of the Trp-7 ring is shielded from solvent by the guanidinium group of Arg-30, and the other is shielded by the Asn-16 side chain. The conserved Asn-16 side chain also serves to bridge strands 2 and 3, since its amide nitrogen and oxygen hydrogen-bond across to the Thr-20 hydroxyl and Cys-23 sulfur, respectively (Fig. 3C).

Interaction Surfaces in the NZF-Ub Complex

The Npl4 NZF motif can bind both mono- and poly-Ub, as analyzed by affinity copurification (10). Backbone amide NMR chemical shift perturbation experiments were used to map the interaction surfaces of both proteins within the 1:1 Npl4 NZF-Ub complex (Fig. 5). The Npl4 NZF-Ub complex was in fast exchange, as is typical for complexes with dissociation constants in the high micromolar range (see below). Chemical shift changes observed in titration experiments therefore reflected the population-weighted average of shifts for the free and bound species.

A series of $^1\text{H}/^{15}\text{N}$ HSQC spectra of ^{15}N -labeled ubiquitin were collected during the stepwise titration of 0–1 eq of unlabeled Npl4 NZF (Fig. 5). Comparison of the spectra revealed that 15 of 70 observable ubiquitin backbone amide resonances shifted significantly upon complex formation ($\delta > 2$; see Fig. 5A). The shifted ubiquitin residues generally clustered about the three C-terminal strands of the β -sheet (with the exception of Thr-14), which presumably corresponds to the NZF binding surface. A prominent feature of the binding site is an exposed, hydrophobic surface formed by Ub residues Leu-8, Ile-44, Val-70, Leu-71, and Leu-73 (highlighted in Fig. 5B). This corresponds to the Ile-44 surface of ubiquitin, which has been shown by genetic analyses to function in endocytosis, proteasomal degradation, and human immunodeficiency virus budding (44, 45). Ubiquitin residue Lys-48 also shifted significantly upon NZF binding, which is notable because polyubiquitin chains linked via the Lys-48 side chain target proteins for proteasome degradation. Our studies do not reveal, however, whether NZF binding is favored or disfavored by conjugation at this position.

In the reciprocal experiment, a series of $^1\text{H}/^{15}\text{N}$ HSQC spectra of ^{15}N -labeled NZF were collected during the stepwise titration of 0–1 eq of unlabeled ubiquitin (Fig. 5C). 10 of 25 observable NZF backbone amide resonances shifted significantly upon ubiquitin binding ($\delta > 2$). All 10 shifted residues were clustered about the rubredoxin knuckles above the zinc coordination site and include the zinc ligands Cys-12, Cys-23, and Cys-26 as well as Thr-13, Phe-14, and Met-25. Like its complementary counterpart on ubiquitin, the binding surface on NZF is primarily hydrophobic but also includes residue Thr-13. Thus, it appears that the interface between NZF and Ub is likely to resemble many other protein-protein interfaces that exhibit buried hydrophobic cores as well as key complementary hydrophilic interactions that provide orientation and specificity.

Ubiquitin Binding by the NZF Domain

To quantify the interaction between Npl4 NZF and free ubiquitin, we performed biosensor binding experiments in which pure recombinant ubiquitin was allowed to bind to immobilized GST-NZF (Fig. 6A). Ubiquitin bound to the Npl4 NZF surface with rapid, reversible kinetics, and the interaction was specific as ubiquitin did not bind to a control GST surface (Fig. 6A, *inset*). A fit of the Npl4 NZF/Ub binding data to a simple 1:1 model yielded a dissociation constant (K_D) of $122 \pm 2 \mu\text{M}$. We speculate that this relatively weak binding affinity may reflect the fact that Npl4 NZF normally recognizes ubiquitylated

proteins (rather than free ubiquitin) and may therefore gain additional binding energy through contacts with the conjugated C-terminal region of ubiquitin, additional contacts elsewhere in the Npl4 protein complex, and/or avidity effects.

We note that the affinity of the Npl4 NZF/Ub interaction is similar to those observed for other well characterized ubiquitin recognition motifs. One example is the UIM, a widespread motif with well established ubiquitin binding functions. Measurements from several groups show that UIMs bind to ubiquitin with dissociation constants in the high micromolar range. Specifically, Shekhtman and Cowburn (46) used NMR chemical shift titrations to estimate that the UIM motif from Hr bound ubiquitin with a dissociation constant of $230 \pm 50 \mu\text{M}$, and Raiborg *et al.* (47) performed biosensor experiments to show that ubiquitin bound to a large, immobilized Hrs fragment spanning the UIM (residues 1–289), with a dissociation constant of $300 \mu\text{M}$.

Biosensor binding experiments were also performed to quantify the interaction of ubiquitin with the immobilized NZF domain of yeast Vps36p (Fig. 6B). Both the affinity ($K_D = 199 \pm 17 \mu\text{M}$) and specificity (not shown) of the Vps36p NZF/Ub interaction were similar to those of the Npl4 NZF/Ub interaction. The observation that Vps36p NZF binds ubiquitin is not surprising from a biological perspective, since Vps36p is required for the sorting of ubiquitylated proteins into the yeast vacuole (11). However, the observation is of interest from a structural perspective, because Vps36p NZF has an asparagine residue in place of the conserved Trp-7 residue that forms the hydrophobic core of Npl4 NZF. Our experiments demonstrate that the Vps36p NZF can nevertheless form a ubiquitin binding module, despite the absence of this hydrophobic core residue.

Our chemical shift mapping experiments indicated that ubiquitin probably binds to Npl4 NZF by contacting the exposed $^{13}\text{TF}^{14}$ Npl4 dipeptide. As discussed below, NZF motifs appear to fall into several different classes that are characterized by distinct dipeptide sequences at positions 13 and 14. We suggest that NZF motifs that bind ubiquitin contain $^{13}\text{TF}^{14}$, whereas those that bind other ligands have other pairs of amino acids at positions 13 and 14. To test this idea, the Npl4 NZF $^{13}\text{TF}^{14}$ residues were mutated singly and in pairs to LV and NE, which are the residue pairs found in NZF motifs from RanBP2 and Mdm2, respectively. As shown in Fig. 6C, all four single mutations increased the K_D for ubiquitin binding at least 25-fold, and the two double mutants did not bind detectably to ubiquitin ($K_D > 10 \text{ mM}$). We therefore conclude that the $^{13}\text{TF}^{14}$ dipeptide is an essential component of the ubiquitin binding epitope of Npl4 NZF.

DISCUSSION

Our studies demonstrate that the Npl4 NZF domain forms a tightly folded zinc binding module and provide a structural rationale for nearly every conserved residue in this domain family (Table II). NZF now joins the ubiquitin-associated domain (48, 49), ubiquitin E2 variant/UBC-like domain (16, 50, 51), and UIM2 as a structurally characterized ubiquitin binding domain. Interestingly, at least three of these structurally distinct motifs (the NZF, ubiquitin E2 variant/UBC-like domain, and UIM) contact ubiquitin on the hydrophobic patch that surrounds the exposed surface of the C-terminal three strands of the β -sheet (the Ile-44 surface) (16, 46, 49, 53–55),² highlighting the importance of this surface for ubiquitin function.

²R. Fisher, B. Wang, S. Alam, W. I. Sundquist, and C. P. Hill, manuscript in preparation.

Similarities between Npl4 NZF and Other Metalloproteins

The Npl4 NZF bears an unexpected resemblance to the “ β ribbon” zinc fingers found in a series of transcription-related proteins from archaea and eucarya, including TFII_s and RPB9 (56–59). In all of these proteins, a single Zn^{2+} ion is coordinated by four Cys residues presented by the knuckles that link two β -hairpins. As shown in Fig. 4A, zinc binding loops from Npl4 NZF and TFII_s are similar and overlay with heavy atom r.m.s. deviation of less than 1.2 Å. The overall topologies of the zinc binding domains of TFII_s (Protein Data Bank entry 1TFI), RPB9 (1QYP), and Npl4 are also similar, although Npl4 NZF is more compact than the other two and seems to represent a “stripped down” version of the fold. More generally, the zinc center of Npl4 NZF is also structurally related to the iron binding sites of the rubredoxin protein family, particularly to ruberythrin (Fig. 4B). Despite these structural similarities, however, there are no obvious functional connections between Npl4 NZF and the other proteins, beyond the possibility that all may mediate protein-protein interactions. Instead, the similarities appear to reflect the fact that the NZF fold provides a particularly favorable geometry for metal binding.

Ubiquitin Binding and Conservation of the NZF Domain

We have identified 243 sequences from 166 different proteins that exactly match the identity and spacing of the four cysteines and single asparagine of Npl4 NZF (see “Experimental Procedures” and Fig. 1A). Others had previously noted the presence of a highly related “zinc finger” sequence motif found in the nuclear pore protein, Ran BP2 (and related proteins) (60–63). Searches based upon the RanBP2/Nup358 zinc finger motif have required the presence of residues equivalent to the Npl4 NZF cysteines as well as Trp-7 and Asn-16 but allowed the spacing between the first two cysteines to vary between 2 and 4 residues (12). These search criteria identify a comparable number of proteins with ours because they include family members that do not match our stringent spacing criteria but exclude NZF proteins like Vps36p (that lack the Trp-7 residue). Thus, it appears that more than 200 known proteins contain NZF motifs, with the precise estimate depending upon the search criteria. Indeed, it will probably be difficult to come up with criteria that will unambiguously define the NZF motif given that there may be a continuum of protein structures that link the NZF and β ribbon zinc motifs.

The prevalence of the NZF motif makes it important to understand the features that dictate different NZF functions. As shown above, the Npl4 and Vps36p NZF motifs both bind specifically to ubiquitin, as do the NZF motifs from two other proteins, TAB2 and RBCK2 (10). However, the NZF domains from two other proteins, Mdm2 and RanBP2/Nup358, do not detectably bind ubiquitin *in vitro* (10). Interestingly, all four known NZF motifs with ubiquitin binding activities display the $^{13}TF^{14}$ dipeptide, whereas the two known NZF motifs that do not bind ubiquitin have very different residues at these positions (LV in RanBP2 and NE in Mdm2). Note that on the basis of structural considerations alone, glycine should be heavily overrepresented at residue 13, since this position in the rubredoxin knuckle adopts a positive ϕ backbone torsion angle (39). Thus, the presence of nonglycine residues at position 13 suggests a functional role in ligand binding.

Consistent with this idea, our chemical shift mapping studies demonstrate that the $^{13}TF^{14}$ dipeptide forms the primary binding site for Ub on Npl4 NZF (with the T13 amide exhibiting the greatest chemical shift change of any NZF residue upon Ub binding). The side chains of Thr-13, Phe-14, and Met-25 form an exposed surface on the Npl4 NZF motif, and our experiments indicate that this is the site of ubiquitin binding. We further suggest that Thr-13 and Phe-14 (and possibly also Met-25) may be generally used by NZF domains whose biological function is to recognize ubiquitin. Consistent with these ideas, our mutagenesis experiments demonstrate that both Thr-13 and Phe-14 are essential for

ubiquitin binding, and preliminary structural studies of the Npl4 NZF·Ub complex confirm that the TF side chains contact the ubiquitin Ile-44 surface.³

To explore this idea further, we tested for correlations between the identities of residues at positions 13 and 14 of the 243 NZF domains identified by our initial search criteria. Strikingly, we found that when a threonine is present at position 13, a large hydrophobic residue (Φ) almost always follows at position 14 (97% correlation, 63/65; Phe = 32, Tyr = 16, Leu = 15). In contrast, when any residue except threonine is present at position 13, a large hydrophobic residue follows only 16% of the time (29/178 total). This analysis again suggests that the NZF motif forms a scaffold that presents the surface-exposed ¹³TF¹⁴ dipeptide (and surrounding residues) to recognize the hydrophobic Ile-44 surface of ubiquitin.

Our sequence analyses also revealed that the LV dipeptide co-varies at NZF positions 13 and 14 (LV; 67% co-variation). We speculate that this dipeptide may also form a recognition epitope for another NZF domain. Interestingly, several proteins, including RanBP2/Nup358 and Nup153, contain multicopy tracts of NZF domains that bind the GDP form of the Ran GTPase (9, 64). These tracts are composed primarily of multiple copies of ¹³LV¹⁴-containing NZF domains, suggesting that RanGDP may be the preferred protein binding partner for this class of NZF domain. The RanBP2 zinc fingers also bind exportin-1 (65), which is apparently yet another protein that can be recognized by the NZF domain.

Biological Implications

The NZF motif was first recognized in the mammalian protein, Npl4, which is a subunit of the heterodimeric UN complex (together with Ufd1) (10). UN is one of at least two alternative adapters that target the AAA-ATPase p97/Cdc48p to specific protein substrates. The UN complex is required for at least three p97/Cdc48p-mediated reactions, including retrotranslocation of ubiquitylated proteins from the ER into the cytosol (ERAD) and the topologically related ubiquitin-dependent processing of transcription factors Spt23 and Mga2 in the ER membrane and subsequent mobilization by Cdc48p (RUP) (66, 67). p97-UN is also required for an as yet undefined reaction that leads to the formation of a closed nuclear envelope after mitosis (68).

Although the exact mechanistic role of UN in ERAD and RUP remains to be determined, UN can bind ubiquitin conjugates by means of the Npl4 NZF domain and can simultaneously recruit the p97 ATPase (10). One possibility is therefore that UN functions by recognizing ubiquitylated protein substrates in the ER membrane and recruiting p97, which then uses the energy of ATP hydrolysis either to pull its substrates out of the membrane or to separate them from other proteins. Interestingly, the other known p97 adapter, p47, which functions in the reassembly of the Golgi apparatus (69) and expansion of the nuclear envelope after mitosis (68), probably also links the ATPase to ubiquitin conjugates (10). In that case, however, ubiquitin recognition is mediated by a ubiquitin-associated domain in p47, and the ubiquitylated substrate remains to be identified.

Vps36p provides another example in which an NZF motif could help to recruit ubiquitylated proteins into a biological pathway. We have shown that the Vps36p NZF can bind ubiquitin (10 and this work), and others have shown that Vps36p is a member of the ESCRT-II complex, which is required for sorting of ubiquitylated proteins into multivesicular bodies (11). It is therefore plausible to suppose that Vps36p uses its NZF motif to recognize ubiquitylated protein substrates, since they are sorted through the multivesicular body

³S. L. Alam, H. H. Meyer, M. Payne, D. R. Davis, and W. I. Sundquist, unpublished results.

pathway. In further analogy to Npl4, there is even the potential that Vps36p could also act as an adaptor for Vps4p, a AAA ATPase that functions late in this pathway (70–72).

Our studies also suggest that many other less well characterized proteins that contain NZF motifs of the “TF” class will also exhibit ubiquitin binding activities. Strengthening this suggestion is the fact that a number of these proteins contain additional domains that are linked to various aspects of ubiquitin biochemistry. Examples include proteins with E3 RING finger motifs that bind ubiquitin-conjugating (E2) enzymes (*e.g.* Ubc7ip3; NCBI entry Q9BYM8) (10, 73), proteins that bind E3 ubiquitin transferases (*e.g.* RYBP, NP_036366) (52), proteins with their own ubiquitin-like domains (*e.g.* Sharpin, NP_112415), and proteins that contain other known ubiquitin binding domains (*e.g.* MGC45404; NP_690000). There are also a number of proteins with NZF motifs of the TF class without previously known links to ubiquitin, and we suggest that investigations in this direction are likely to be fruitful.

In summary, the NZF motif is a zinc binding module that can be incorporated into multifunctional proteins and used to bind ubiquitin, RanGDP, exportin-1, and probably other proteins. Our studies reveal the NZF structure, define the NZF/Ub interface, and provide criteria for distinguishing functional variants of this versatile scaffold.

Acknowledgments

We thank Rebecca Rich and David Myszka (University of Utah Protein Interactions Core Facility) for BIAcore biosensor analyses; Bob Schackmann (University of Utah DNA and Peptide Core Facility) for N-terminal protein sequencing; Dennis Winge (Utah) for atomic absorption spectroscopic analyses; the University of Utah Mass Spectrometry Core Facility for protein mass spectrometry; the SLAC staff; Eric Ross for computer support; Cecile Pickart for the ubiquitin expression clone and protocol; and Chris Hill, Dennis Winge, Mike Summers, and Katie Ullman for critical reading of the manuscript.

REFERENCES

1. Hershko A, Ciechanover A. *Annu. Rev. Biochem.* 1998; 67:425–479. [PubMed: 9759494]
2. Pickart CM. *Annu. Rev. Biochem.* 2001; 70:503–533. [PubMed: 11395416]
3. Hicke L. *Nat. Rev. Mol. Cell. Biol.* 2001; 2:195–201. [PubMed: 11265249]
4. Buchberger A. *Trends Cell Biol.* 2002; 12:216–221. [PubMed: 12062168]
5. Madura K. *Cell Cycle.* 2002; 1:235–244. [PubMed: 12429939]
6. Hofmann K, Falquet L. *Trends Biochem. Sci.* 2001; 26:347–350. [PubMed: 11406394]
7. Hofmann K, Bucher P. *Trends Biochem. Sci.* 1996; 21:172–173. [PubMed: 8871400]
8. Delaney SJ, Hayward DC, Barleben F, Fischbach KF, Miklos GL. *Proc. Natl. Acad. Sci. U. S. A.* 1991; 88:7214–7218. [PubMed: 1714593]
9. Nakielnny S, Shaikh S, Burke B, Dreyfuss G. *EMBO J.* 1999; 18:1982–1995. [PubMed: 10202161]
10. Meyer HH, Wang Y, Warren G. *EMBO J.* 2002; 21:5645–5652. [PubMed: 12411482]
11. Babst M, Katzmann D, Snyder W, Wendland B, Emr S. *Dev. Cell.* 2002; 3:283–289. [PubMed: 12194858]
12. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K, Bairoch A. *Nucleic Acids Res.* 2002; 30:235–238. [PubMed: 11752303]
13. Haj FG, Verveer PJ, Squire A, Neel BG, Bastiaens PI. *Science.* 2002; 295:1708–1711. [PubMed: 11872838]
14. Beal R, Deveraux Q, Xia G, Rechsteiner M, Pickart C. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:861–866. [PubMed: 8570649]
15. Clore GM, Gronenborn AM. *Methods Enzymol.* 1994; 239:349–363. [PubMed: 7830590]
16. Pornillos O, Alam S, Rich RL, Myszka DG, Davis DR, Sundquist WI. *EMBO J.* 2002; 21:2397–2406. [PubMed: 12006492]

17. Bax A, Vuister GW, Grzesiek S, Delaglio F, Wang AC, Tschudin R, Zhu G. *Methods Enzymol.* 1994; 239:79–105. [PubMed: 7830604]
18. Mori S, Abeygunawardana C, Johnson MO, van Zijl PC. *J. Magn. Reson. B.* 1995; 108:94–98. [PubMed: 7627436]
19. Zhang O, Kay LE, Olivier JP, Forman-Kay JD. *J. Biomol. NMR.* 1994; 4:845–858. [PubMed: 7812156]
20. Muhandiram DR, Xy GY, Kay LE. *J. Biomol. NMR.* 1993; 3:463–470.
21. Pascal SMM, Yamazaki DR, Forman-Kay JD, Kay LE. *J. Magn. Reson.* 1994; 191:197–201.
22. Kuboniwa H, Grzesiek S, Delaglio F, Bax A. *J. Biomol. NMR.* 1994; 4:871–878. [PubMed: 7812158]
23. Meissner A, Sorensen OW. *J. Magn. Reson.* 2000; 143:387–390. [PubMed: 10729264]
24. Hsu ST, Breukink E, de Kruijff B, Kaptein R, Bonvin AM, van Nuland NA. *Biochemistry.* 2002; 41:7670–7676. [PubMed: 12056898]
25. Ballardina A, Fischman AJ, Gibbons WA, Roy J, Schwartz IL, Smith CW, Walter R, Wyssbrod HR. *Biochemistry.* 1978; 17:4443–4454. [PubMed: 718851]
26. Graham WH, Carter ES 2nd, Hicks RP. *Biopolymers.* 1992; 32:1755–1764. [PubMed: 1472657]
27. Pellicchia M, Meininger D, Shen AL, Jack R, Kasper CB, Sem DS. *J. Am. Chem. Soc.* 2001; 123:4633–4634. [PubMed: 11457263]
28. Guntert P, Mumenthaler C, Wuthrich K. *J. Mol. Biol.* 1997; 273:283–298. [PubMed: 9367762]
29. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. *Acta Crystallogr. D Biol. Crystallogr.* 1998; 54:905–921. [PubMed: 9757107]
30. Garrett DS, Kuszewski J, Hancock TJ, Lodi PJ, Vuister GW, Gronenborn AM, Clore GM. *J. Magn. Reson. B.* 1994; 104:99–103. [PubMed: 8025816]
31. Herrmann T, Guntert P, Wuthrich K. *J. Mol. Biol.* 2002; 319:209–227. [PubMed: 12051947]
32. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. *J. Biomol. NMR.* 1996; 8:477–486. [PubMed: 9008363]
33. Koradi R, Billeter M, Wüthrich K. *J. Mol. Graphics.* 1996; 14:51–55.
34. Cheever ML, Sato TK, de Beer T, Kutateladze TG, Emr SD, Overduin M. *Nat. Cell Biol.* 2001; 3:613–618. [PubMed: 11433291]
35. Grzesiek S, Bax A, Clore GM, Gronenborn AM, Hu JS, Kaufman J, Palmer I, Stahl SJ, Wingfield PT. *Nat. Struct. Biol.* 1996; 3:340–345. [PubMed: 8599760]
36. Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE, Wettstein DA, Stray KM, Cote M, Rich RL, Myszka DG, Sundquist WI. *Cell.* 2001; 107:55–65. [PubMed: 11595185]
37. Adman E, Watenpaugh KD, Jensen LH. *Proc. Natl. Acad. Sci. U. S. A.* 1975; 72:4854–4858. [PubMed: 1061073]
38. Sarisky CA, Mayo SL. *J. Mol. Biol.* 2001; 307:1411–1418. [PubMed: 11292351]
39. Blake PR, Summers MF. *Adv. Biophys. Chem.* 1994; 4:1–30.
40. Hammarstrom A, Berndt KD, Sillard R, Adermann K, Otting G. *Biochemistry.* 1996; 35:12723–12732. [PubMed: 8841116]
41. Martinez-Yamout M, Legge GB, Zhang O, Wright PE, Dyson HJ. *J. Mol. Biol.* 2000; 300:805–818. [PubMed: 10891270]
42. Klug A, Schwabe JW. *FASEB J.* 1995; 9:597–604. [PubMed: 7768350]
43. Sieker LC, Holmes M, Le Trong I, Turley S, Liu MY, LeGall J, Stenkamp RE. *J. Biol. Inorg. Chem.* 2000; 5:505–513. [PubMed: 10968622]
44. Sloper-Mould KE, Jemc JC, Pickart CM, Hicke L. *J. Biol. Chem.* 2001; 276:30483–30489. [PubMed: 11399765]
45. Strack B, Calistri A, Gottlinger HG. *J. Virol.* 2002; 76:5472–5479. [PubMed: 11991975]
46. Shekhtman A, Cowburn D. *Biochem. Biophys. Res. Commun.* 2002; 296:1222–1227. [PubMed: 12207904]

47. Raiborg C, Bache KG, Gillooly DJ, Madshus IH, Stang E, Stenmark H. *Nat. Cell Biol.* 2002; 4:394–398. [PubMed: 11988743]
48. Dieckmann T, Withers-Ward ES, Jarosinski MA, Liu CF, Chen IS, Feigon J. *Nat. Struct. Biol.* 1998; 5:1042–1047. [PubMed: 9846873]
49. Mueller TD, Feigon J. *J. Mol. Biol.* 2002; 319:1243–1255. [PubMed: 12079361]
50. Moraes TF, Edwards RA, McKenna S, Pastushok L, Xiao W, Glover JN, Ellison MJ. *Nat. Struct. Biol.* 2001; 8:669–673. [PubMed: 11473255]
51. VanDemark AP, Hofmann RM, Tsui C, Pickart CM, Wolberger C. *Cell.* 2001; 105:711–720. [PubMed: 11440714]
52. Garcia E, Marcos-Gutierrez C, del Mar Lorente M, Moreno JC, Vidal M. *EMBO J.* 1999; 18:3404–3418. [PubMed: 10369680]
53. Bertolaet BL, Clarke DJ, Wolff M, Watson MH, Henze M, Divita G, Reed SI. *Nat. Struct. Biol.* 2001; 8:417–422. [PubMed: 11323716]
54. Walters KJ, Kleijnen MF, Goh AM, Wagner G, Howley PM. *Biochemistry.* 2002; 41:1767–1777. [PubMed: 11827521]
55. VanDemark AP, Hill CP. *Curr. Opin. Struct. Biol.* 2002; 12:822–830. [PubMed: 12504688]
56. Qian X, Jeon C, Yoon H, Agarwal K, Weiss MA. *Nature.* 1993; 365:277–279. [PubMed: 7626141]
57. Qian X, Gozani SN, Yoon H, Jeon CJ, Agarwal K, Weiss MA. *Biochemistry.* 1993; 32:9944–9959. [PubMed: 8399164]
58. Wang B, Jones DN, Kaine BP, Weiss MA. *Structure.* 1998; 6:555–569. [PubMed: 9634694]
59. Chen HT, Legault P, Glushka J, Omichinski JG, Scott RA. *Protein Sci.* 2000; 9:1743–1752. [PubMed: 11045620]
60. Sukegawa J, Blobel G. *Cell.* 1993; 72:29–38. [PubMed: 8422679]
61. Wu J, Matunis MJ, Kraemer D, Blobel G, Coutavas E. *J. Biol. Chem.* 1995; 270:14209–14213. [PubMed: 7775481]
62. Yokoyama N, Hayashi N, Seki T, Pante N, Ohba T, Nishii K, Kuma K, Hayashida T, Miyata T, Aebi U, Fukui M, Nishimoto T. *Nature.* 1995; 376:184–188. [PubMed: 7603572]
63. Fauser S, Aslanukov A, Roepman R, Ferreira PA. *Mamm. Genome.* 2001; 12:406–415. [PubMed: 11353387]
64. Yaseen NR, Blobel G. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:5516–5521. [PubMed: 10318915]
65. Singh BB, Patel HH, Roepman R, Schick D, Ferreira PA. *J. Biol. Chem.* 1999; 274:37370–37378. [PubMed: 10601307]
66. Bays NW, Hampton RY. *Curr. Biol.* 2002; 12:R366–R371. [PubMed: 12015140]
67. Tsai B, Ye Y, Rapoport TA. *Nat. Rev. Mol. Cell. Biol.* 2002; 3:246–255. [PubMed: 11994744]
68. Hetzer M, Meyer HH, Walther TC, Bilbao-Cortes D, Warren G, Mattaj JW. *Nat. Cell Biol.* 2001; 3:1086–1091. [PubMed: 11781570]
69. Kondo H, Rabouille C, Newman R, Levine TP, Pappin D, Freemont P, Warren G. *Nature.* 1997; 388:75–78. [PubMed: 9214505]
70. Babst M, Wendland B, Estepa EJ, Emr SD. *EMBO J.* 1998; 17:2982–2993. [PubMed: 9606181]
71. Scheuring S, Rohricht RA, Schonning-Burkhardt B, Beyer A, Muller S, Abts HF, Kohrer K. *J. Mol. Biol.* 2001; 312:469–480. [PubMed: 11563910]
72. Bishop N, Woodman P. *J. Biol. Chem.* 2001; 276:11735–11742. [PubMed: 11134028]
73. Martinez-Noel G, Niedenthal R, Tamura T, Harbers K. *FEBS Lett.* 1999; 454:257–261. [PubMed: 10431818]

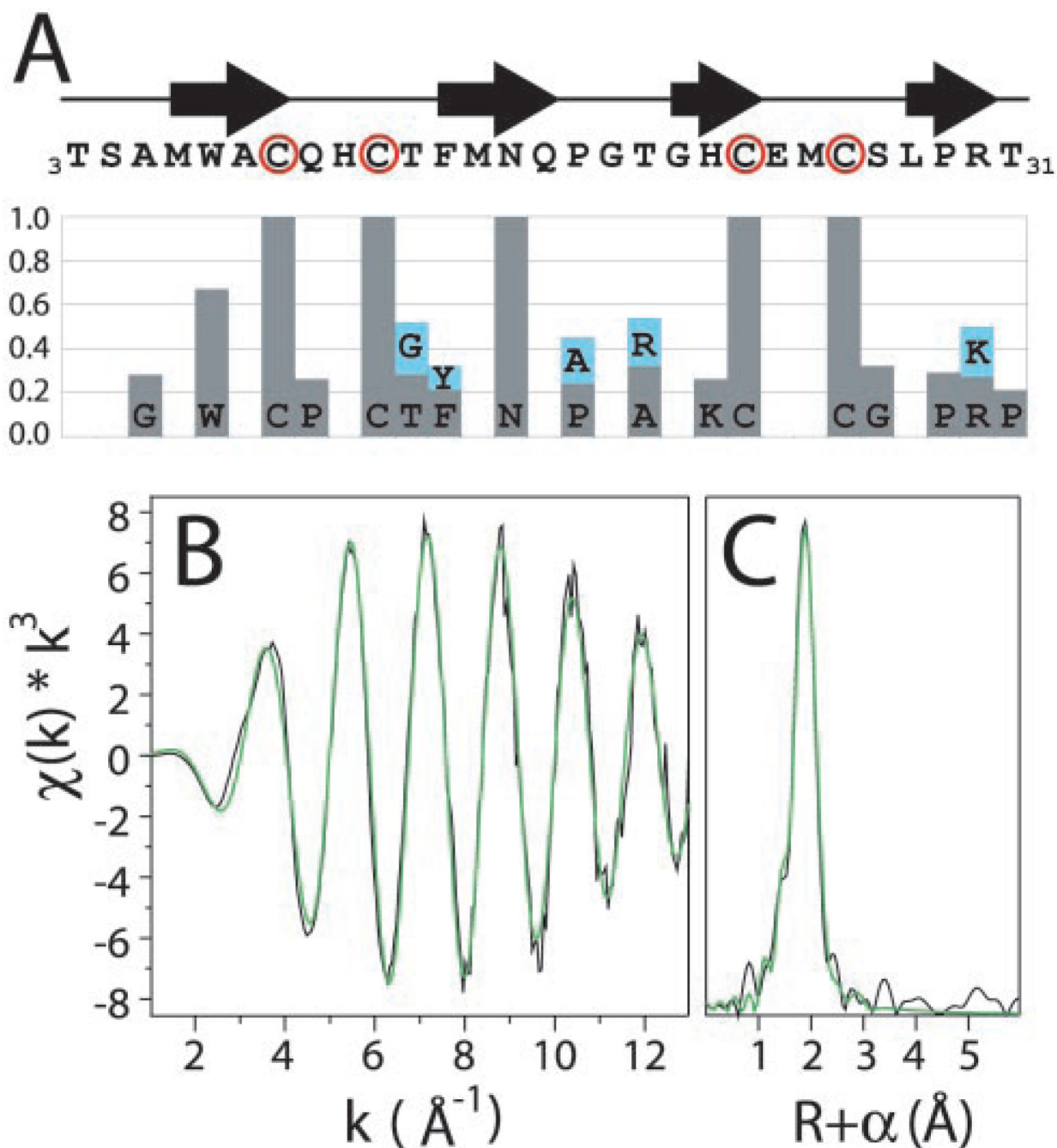


Fig. 1. Primary sequence, conservation, and zinc binding by the Npl4 NZF domain

A, the primary sequence, numbering scheme, and secondary structure of the Npl4 NZF domain are shown above with the four cysteines that coordinate zinc *circled in red*. Shown *below* is the sequence conservation in putative NZF/Ran BP2-like domains from 243 sequences with four cysteines and one asparagine in the same spacing as Npl4 NZF. Residues present in more than 18% of NZF domains are shown explicitly, with *bar heights* representing their relative frequencies (see “Experimental Procedures”). *B*, untransformed EXAFS spectra (*black*) and fit to a $\text{Zn}(\text{Cys})_4$ model (*green*) of Npl4 NZF. *C*, Fourier-transformed EXAFS spectra (*black*) and fit (*green*) of Npl4 NZF.

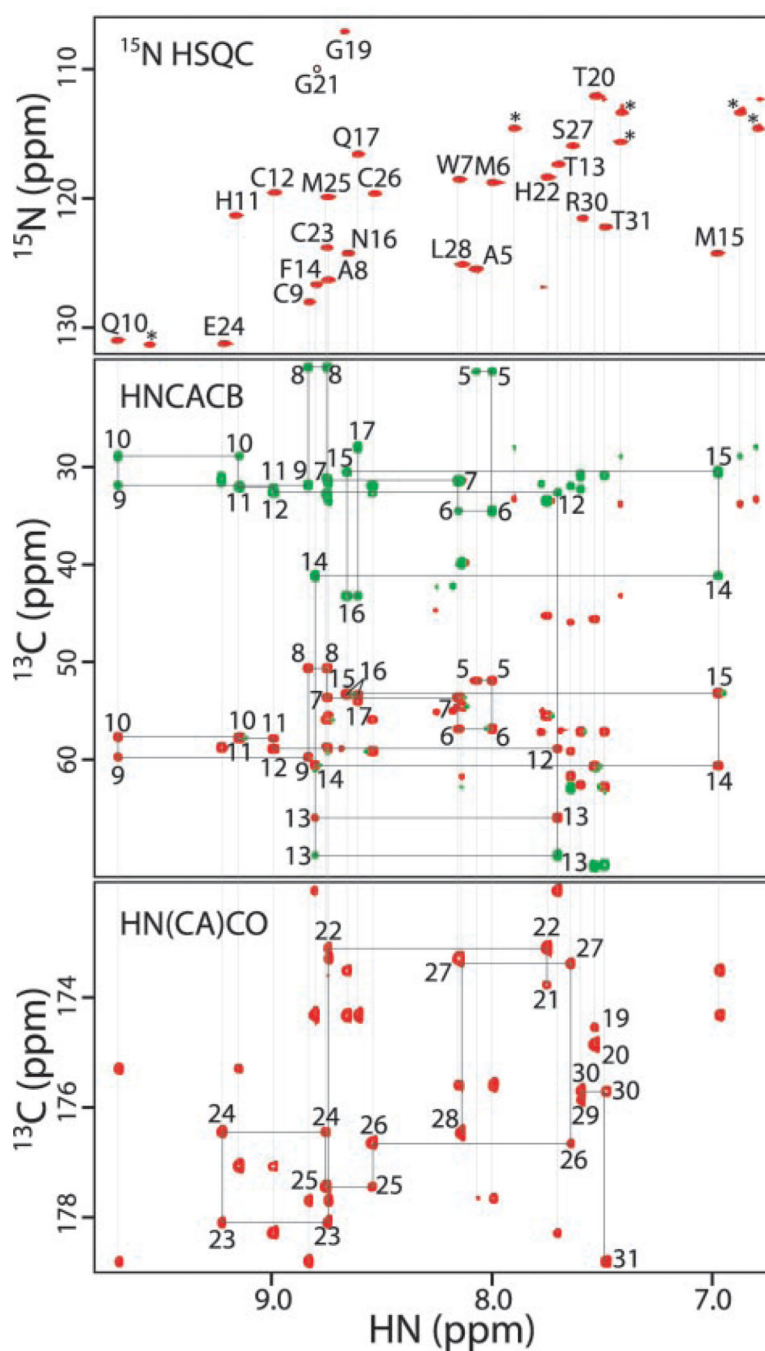


Fig. 2. NMR assignments for the Npl4 NZF domain

Upper panel, $^1\text{H}/^{15}\text{N}$ HSQC spectrum and chemical shift assignments for Npl4 NZF. The *asterisks* denote resonances from side chain amides. *Center panel*, two-dimensional HNCACB spectrum showing intraresidue and sequential ($i-1$) connectivities. *Red peaks* are positive ($\text{C}\alpha$); *green peaks* are negative ($\text{C}\beta$). *Lower panel*, two-dimensional HN(CA)CO spectrum showing intraresidue and sequential ($i-1$) connectivities. *Vertical lines* in the *lower two panels* provide correlations to the appropriate amide resonance in the upper panel. For clarity, residues 1–18 are labeled in the *central panel*, and residues 19–31 are labeled in the *lower panel*. Note that correlations for residues 1–4 and glycine residues 19 and 21 (*open circle* in *panel 1*) were too weak to observe at these contour levels.

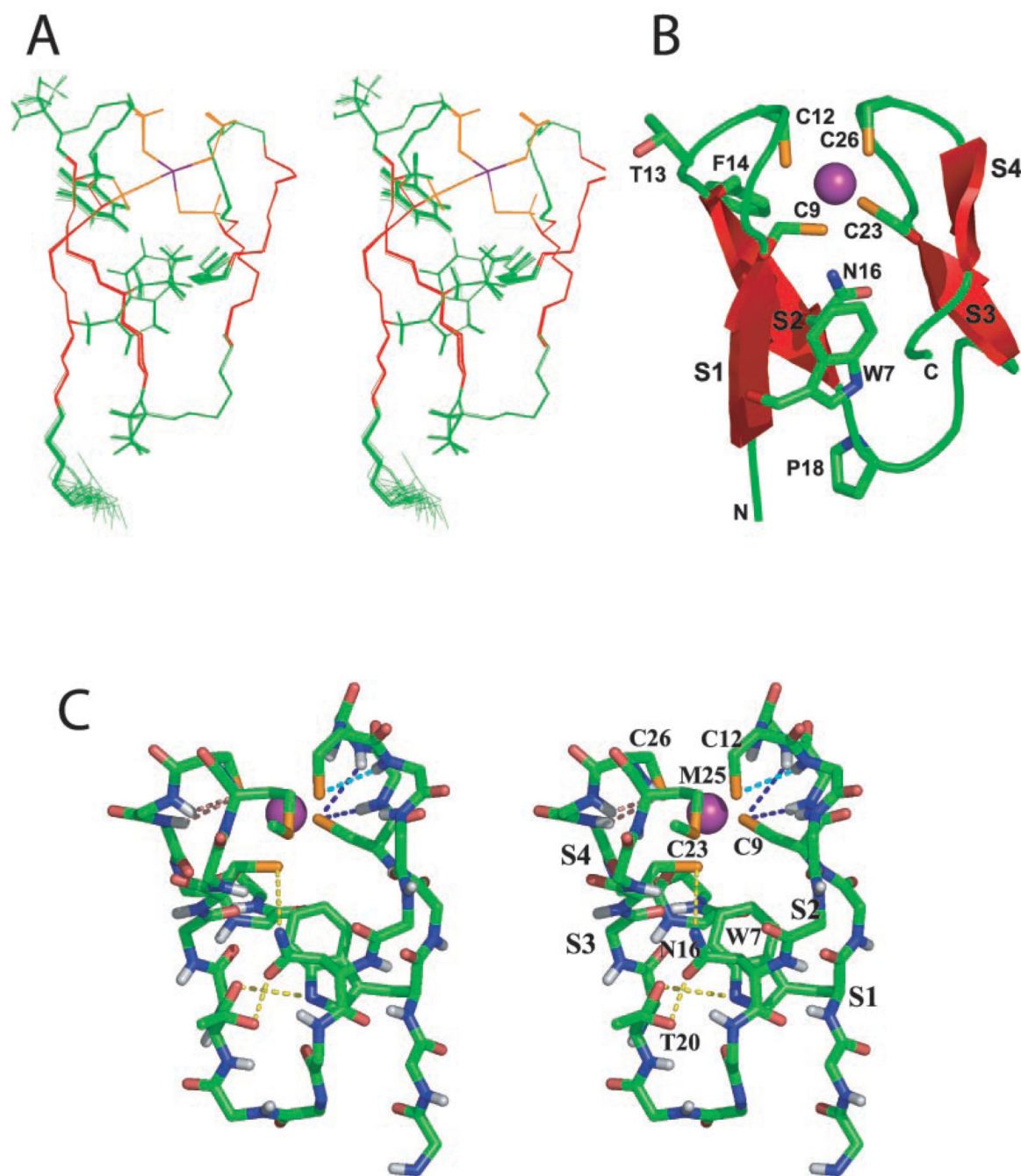


Fig. 3. Structure of the Npl4 NZF domain

A, stereoview superposition of the 20 final structures of Npl4 NZF (backbone trace). Loops are shown in *green*, strands in *red*, Zn²⁺ in *purple*, and the four Sγ ligands in *gold*. A subset of the conserved, well ordered side chains are also shown (in *green*). *B*, *ribbon diagram* of the Npl4 NZF structure. The *orientation* and *color coding* are the same as in *A*. *C*, view of the extensive long range hydrogen bonding network in Npl4 NZF (*dashed lines*). Relative to *A* and *B*, this orientation is given by a 180° rotation about a vertical axis, and the coloring scheme is as follows: *white*, amide protons; *blue*, nitrogens; *red*, carbonyl oxygens; *green*, carbons; *gold*, sulfurs. A subset of the conserved, well ordered side chains are also shown. In addition to the hydrogen bonds normally found in β-sheets and canonical turns, the

following hydrogen bonds were predicted in Insight II (in 20/20 final structures, except where noted): 7Ne1-T20C' (11/20), 9Se-10N, 9S γ -11N, 12S γ -13N (15/20), 16O δ -20O γ , 16N δ 2-23S γ , 26S γ -27N, 26S γ -28N.

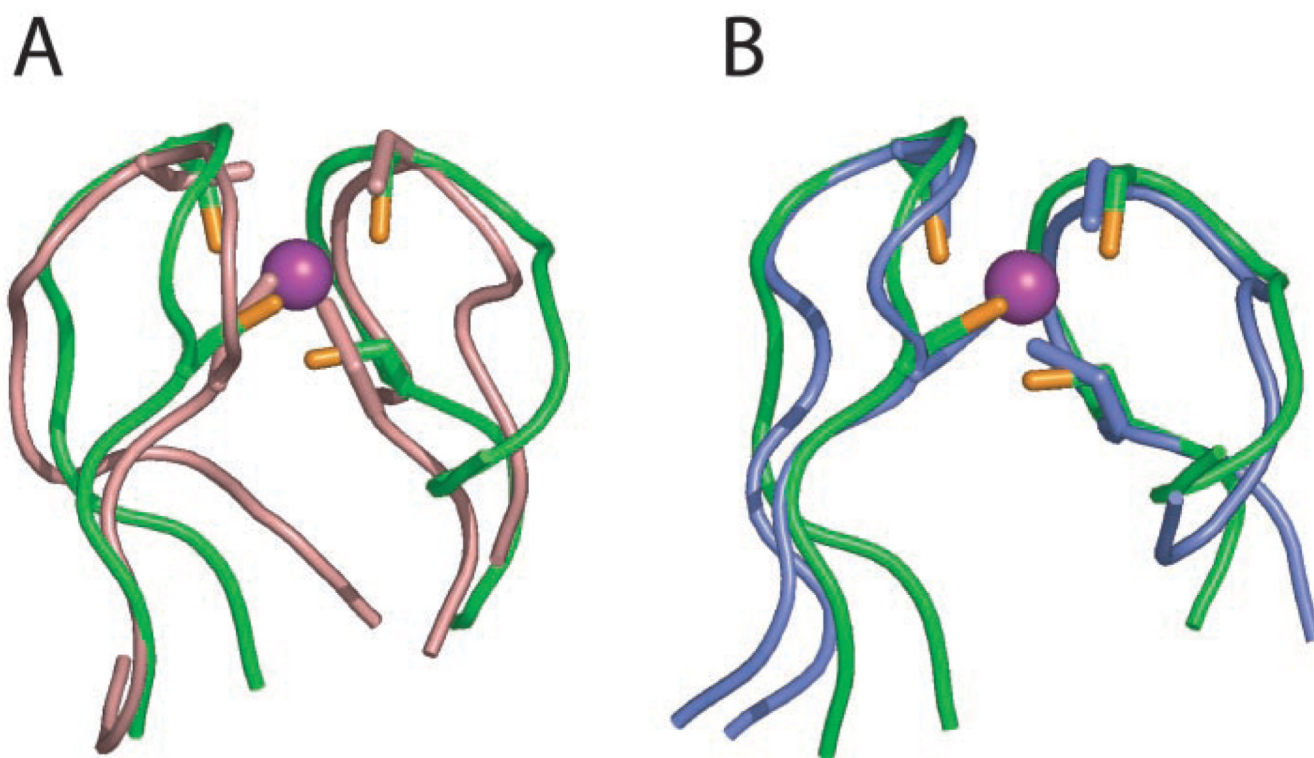


Fig. 4. Structural similarity between the Npl4 NZF zinc site and the rubredoxin knuckles of other metalloproteins

A, superposition of Npl4 NZF (*green*) residues 6–15 (*left side*) and 22–29 (*right side*) onto the zinc binding site of the human transcriptional elongation factor TFIIs (residues 9–18 and 39–46, *salmon*, 1TFI). This superposition gives an r.m.s. deviation of 1.2 Å (backbone). *B*, superposition of Npl4 NZF (*green*) residues 7–15 and 22–30 onto the iron binding site from the rubredoxin of *Clostridium pasteurianum* (residues 4–12 and 38–46, *purple*, 1FHH). The superposition gives an r.m.s. deviation of 0.83 Å (backbone).

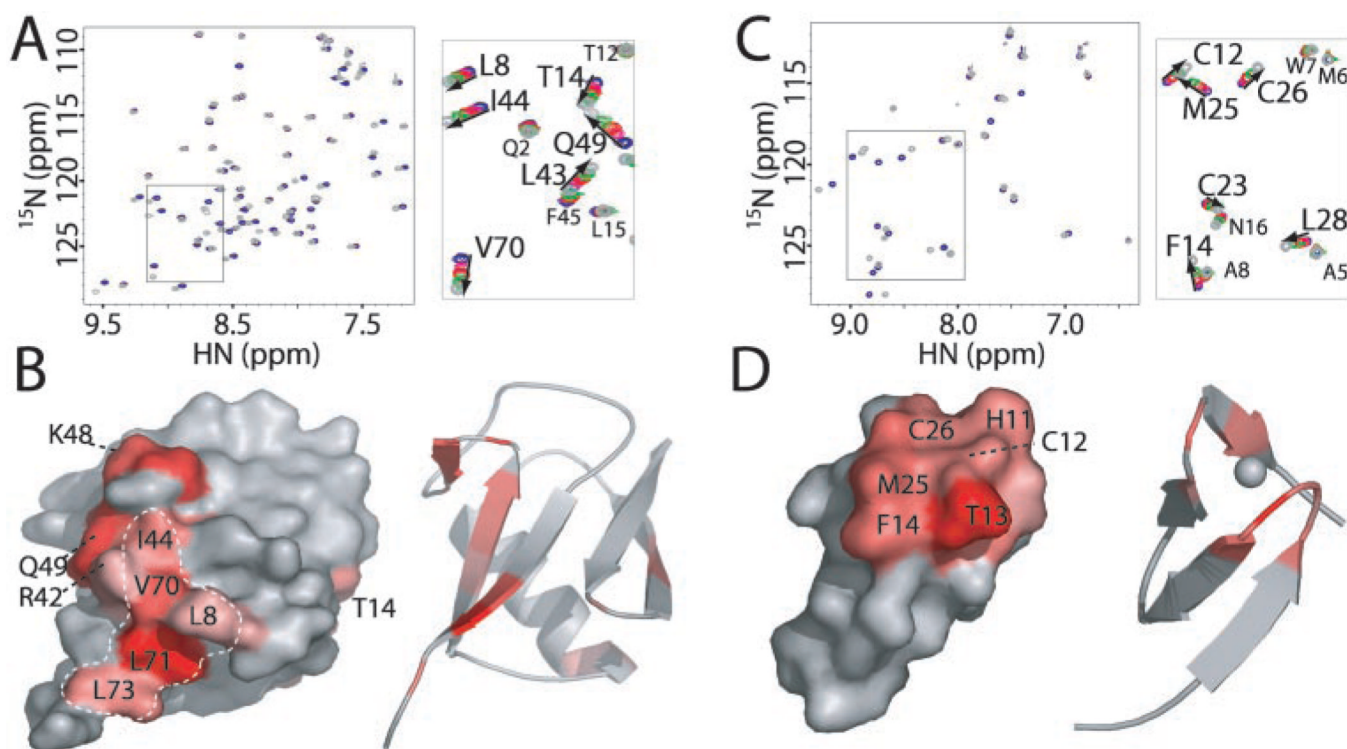


Fig. 5. NMR chemical shift mapping of the interaction surfaces of ubiquitin (A and B) and Npl4 NZF (C and D)

A, overlaid $^1\text{H}/^{15}\text{N}$ HSQC spectra of ubiquitin (0.15 mM) in the presence of 0 (purple), and 1.0 (gray) molar eq of Npl4 NZF. The boxed inset in the left spectrum is expanded at the right and also includes intermediate titration points at 0.125 (magenta), 0.25 (red), and 0.5 (green) eq of Npl4 NZF. B, residues with the greatest chemical shift changes upon the addition of 1 eq of Npl4 NZF are shown mapped onto surface and ribbon representations of ubiquitin. Residues shifted by $\delta \geq 2$ are labeled at the left and are colored using a gradient scheme from red ($\delta = 6$, most shifted) to pink ($\delta = 2$). The hydrophobic patch on the Ile-44 surface of ubiquitin is outlined in white (see text). C, overlaid $^1\text{H}/^{15}\text{N}$ HSQC spectra of Npl4 NZF (1.5 mM) in the presence of 0–1 molar eq of ubiquitin (shown with the same color coding used in A). D, residues with the greatest chemical shift changes upon the addition of 1 eq of ubiquitin are shown mapped onto surface and ribbon representations of Npl4 NZF. Residues with shifts of $\delta \geq 2$ are labeled at the left and are colored using a gradient scheme from red ($\delta = 8$) to pink ($\delta = 2$).

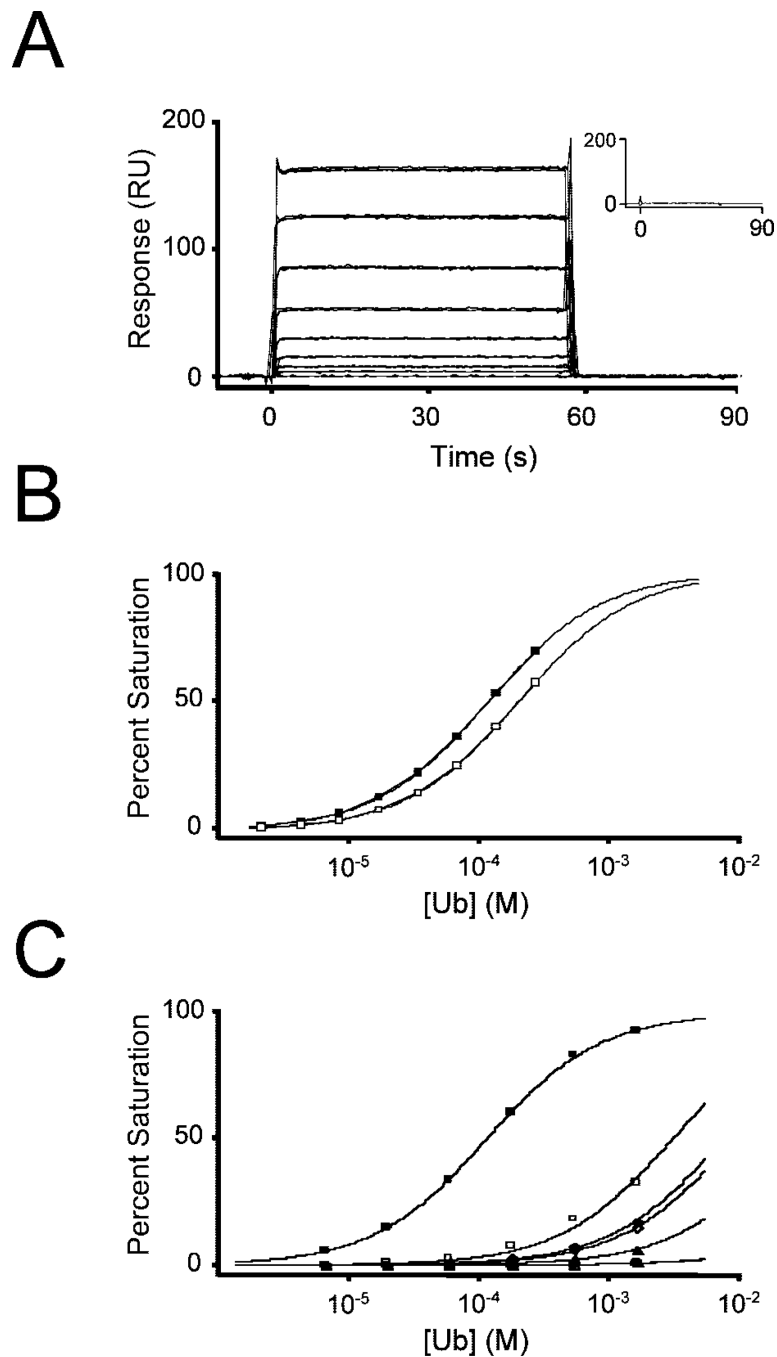


Fig. 6. Quantitation of ubiquitin binding by NZF domains

A, surface plasmon resonance biosensor analysis of the Ub/Npl4 NZF interaction. Ubiquitin was injected in triplicate at concentrations of 0, 2.29, 4.38, 8.75, 17.5, 35.0, 70.0, 140, and 280 μM over GST-Npl4 NZF captured on an anti-GST surface. The *inset* depicts the response obtained for 280 μM ubiquitin injected over recombinant GST captured on an anti-GST surface (negative control). *B*, isotherms for Ub binding to the NZF domains from Npl4 and Vps36p. Fitting the data to simple 1:1 binding models yielded $K_D = 122 \pm 2 \mu\text{M}$ for Npl4 NZF (*filled squares*) and $K_D = 199 \pm 17 \mu\text{M}$ for Vps36p NZF (*open squares*). *C*, isotherms for Ub binding to wild type Npl4 NZF (*filled squares*, $K_D = 108 \pm 13 \mu\text{M}$) and a series of single and double substitution mutants at Npl4 positions 13 and 14. F14V (*open squares*, K_D

= $2830 \pm 30 \mu\text{M}$), T13L (*filled diamonds*, $K_D = 6940 \pm 30 \mu\text{M}$), T13N (*open diamonds*, $K_D = 8450 \pm 30 \mu\text{M}$), F14E (*filled triangles*, $K_D = 10 \text{ mM}$), T13L/F14V (*open triangles*, $K_D = 10 \text{ mM}$), and T13N/F14E (*filled circles*, $K_D = 10 \text{ mM}$).

Table I

Structure statistics for Np14 NZF domain

	$\langle \text{TAD} \rangle^a$	$\langle \text{CNS} \rangle^a$
NOE distance restraints ^b (Å)	362	306
Intraresidue	101	87
Sequential ($ i-j =1$)	111	97
Medium range ($2 < i-j < 5$)	54	44
Long range ($ i-j > 5$)	96	78
Zinc coordination restraints	14	14
Hydrogen bond distance restraints ^c (Å)	0	8
Hydrogen bonds	0	4
Three-bond J_{HNHA} scalar coupling restraints (Hz)	0	13
Stereospecific assignments	28	28
DYANA Target function (Å ⁴)	0.12 ± 0.01	NA ^d
CNS energy	~800 ± 200 ^e	66.5 ± 0.5
Residual distance restraint violations		
Number of violations 0.1 Å	0 ± 0	0 ± 0
Sum of violations (Å or kcal/mol) ^f	1.7 ± 0.1	18.7 ± 0.9
Maximum violation (Å)	0.14	
Residual scalar coupling restraint violations		
Number of violations 1°	NA	0 ± 0
Sum of violations (degrees or kcal/mol) ^f	NA	2.0 ± 0.17
Maximum violation (degrees)	NA	
Van der Waals violations		
Number 0.1 Å	0 ± 0	0 ± 0
Sum of violations (Å or kcal/mol) ^f	0.7 ± 0.1	16.8 ± 0.8
Maximum violation (Å)	0.09	
Ramachandran statistics (residues 5–31) ^g		
Favored	55.8%	60.5%
Allowed	35.7%	26.4%
Generously allowed	8.3%	13.2%
Disallowed	0.3%	0.0%
r.m.s. deviations to the average coordinates ^h (Å)		
Residues 5–31		
Backbone	0.24 ± 0.08	0.11 ± 0.05
Heavy atoms	0.82 ± 0.12	0.72 ± 0.15

^a $\langle \text{TAD} \rangle$ is the ensemble of 20 lowest penalty structures calculated using the program DYANA (28). $\langle \text{CNS} \rangle$ is the same ensemble after 1000 steps (15 ps each) of simulated annealing at 25 K, 1000 slow cooling steps to 0 K, and 10,000 steps of restrained Powell minimization in cartesian space (anneal.inp protocol) (29).

^bOnly meaningful and nonredundant restraints as determined by the DYANA CALIBA function.

^cTwo upper limit distance restraints were used to define each hydrogen bond.

^dNA, not applicable.

^eEnergies for structures input into CNS (from DYANA) were estimated within the generate_easy.inp program after initial regularization without restraints.

^fViolations from DYANA have units of Å, while violation energies from CNS are in kcal/mol.

^gDetermined using PROCHECK-NMR (32).

^hSuperposition and overall rmsds were calculated using the program MOLMOL (33).

Table II

Proposed structural roles for conserved NZF residues

Residue ^a	Structural role(s)
G(A)5	N terminus of strand 1
W7	Hydrophobic core
C9	Zinc coordination
P(Q)10	Rubredoxin knuckle metal binding loop
C12	Zinc coordination
T/G13	Positive ϕ backbone torsion angle at position 5 of the rubredoxin knuckle (G), ubiquitin binding (T)
Y/F/L14	Ubiquitin binding
N16	Interstrand hydrogen bonding
A/P18	$i + 1$ position of type-1 turn
R/A(T)20	Unclear from Np14 NZF structure
K(H)22	Surface-exposed side chain
C23	Zinc coordination
C26	Zinc coordination
G(S)27	Positive ϕ backbone torsion angle at position 5 of the rubredoxin knuckle
P29	C terminus of strand 4
K/R30	Surface-exposed side chain
P(T)31	Domain terminus

^aConserved residues in NZF domain (see Fig. 1). The Np14 NZF residue is given in parentheses when it does not correspond to the consensus sequence.