J Biomol NMR (2011) 51:185–190
DOI 10.1007/s10858-011-9552-y

ARTICLE

Zn-binding AZUL domain of human ubiquitin protein ligase Ube3A

Alexander Lemak · Adelinda Yee · Irina Bezsonova · Sirano Dhe-Paganon · Cheryl H. Arrowsmith

Received: 12 May 2011 / Accepted: 22 June 2011
© Springer Science+Business Media B.V. 2011

Abstract Ube3A (also referred to as E6AP for E6 Associated Protein) is a E3 ubiquitin-protein ligase implicated in the development of Angelman syndrome by controlling degradation of synaptic protein Arc and oncogenic papilloma virus infection by controlling degradation of p53. This article describe the solution NMR structure of the conserved N-terminal domain of human Ube3A (residues 24-87) that contains two residues (Cys44 and Arg62) found to be mutated in patients with Angelman syndrome. The structure of this domain adopts a novel Zn-binding fold we called AZUL (Amino-terminal Zn-finger of Ube3a Ligase). The AZUL domain has a helix-loop-helix architecture with a Zn ion coordinated by four Cys residues arranged in Cys-X_4-Cys-X_4-Cys-X_{28}-Cys motif. Three of the Zn-bound residues are located in a 23-residue long and well structured loop that connects two α-helicies.

Keywords Ube3A · E6-AP · E3A ubiquitin ligase · AZUL · Angelman syndrome · Zn-finger · HPV · NMR

A. Lemak · A. Yee · C. H. Arrowsmith
Ontario Cancer Institute, Campbell Family Cancer Research Institute and Department of Medical Biophysics, University of Toronto, and Northeast Structural Genomics Consortium, 101 College Street, Toronto, ON M5G 1L7, Canada

I. Bezsonova (✉)
Department of Molecular Microbial and Structural Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030, USA
e-mail: bezsonova@uchc.edu

S. Dhe-Paganon (✉) · C. H. Arrowsmith (✉)
Structural Genomics Consortium, University of Toronto, 101 College Street, Toronto, ON M5G 1L7, Canada
e-mail: sirano.dhepaganon@utoronto.ca

C. H. Arrowsmith
e-mail: carrow@uhnresearch.ca

S. Dhe-Paganon
Department of Physiology, University of Toronto, 100 College Street, Toronto, ON M5G 1L5, Canada

Biological context

Ube3A is an E3 ubiquitin ligase crucial for normal brain development. Its loss or mutation causes Angelman syndrome (AS), a neurodevelopmental disorder characterized by motor dysfunction, severe mental retardation, speech impairment, seizures, and a high prevalence of autism (Williams et al. 2010). In search of molecular mechanisms of the Ube3A-dependent AS development several of its ubiquitylation targets have been identified, including a synaptic protein Arc that promotes internalization of glutamate receptors. It has been demonstrated that Ube3A regulates synaptic function in the brain by controlling degradation of Arc via ubiquitin-proteasomal pathway (Greer et al. 2010).

Ube3A has also been extensively studied with regard to the role it plays in oncogenic human papilloma virus (HPV) infection. It was initially discovered as a cellular factor recruited by the E6 protein from oncogenic strains of HPV during infection and is often referred to as E6-AP for E6-Associated Protein. HPV E6 hijacks host Ube3A, resulting in the ubiquitylation and degradation of the p53 tumor suppressor, as well as several other cellular proteins leading to cellular immortalization (Beaudenon and Huibregtse 2008). In fact, HPVs are the etiological agents in nearly all cases (99.7%) of cervical cancer (Walboomers et al. 1999), and the HPV E6 protein is one of two viral oncoproteins that is expressed in virtually all HPV-positive cancers. While the crucial role of Ube3A and E6 in p53 degradation and

cancer development is clear, the structural characterization of Ube3A and E6 and especially their complexes with p53 and each other are not well understood.

Ube3A is a 100 kDa (875 amino acid residues) protein. Its catalytic ubiquitin ligase domain is located at the C-terminus of the protein (residues 545–875) and is the founding member of the HECT class of homologous E3 ubiquitin ligase domains (Homologous to E6-AP C-Terminus domain). HECT domains bind to an E2 forming a thioester intermediate with the C-terminus of activated ubiquitin via a catalytic cysteine residue. The structure of the HECT domain has been determined and shown to consist of two lobes with a catalytic site at their junction and an E2-binding site on the N-lobe of the domain. The extensive amino-terminal region of the Ube3A is potentially responsible for its substrate specificity and is not yet structurally characterized. Most of the Angelman-associated point mutations lie within the carboxyl-terminal HECT domain of the protein and disrupt its catalytic activity, however, there have been reported at least three mutations (Cooper et al. 2004; Bai et al. 2011) in Angelman patients that lie in the non-catalytic amino-terminal portion (residues 1–544) and may affect substrate binding, subcellular localization, or protein stability.

In this work we present the NMR solution structure of a highly conserved amino-terminal region of Ube3A (residues 24–87, pdbID:2KR1). The region contains 5 conserved cysteine residues one of which (C44) was found to be mutated in patients with Angelman syndrome. We show that the domain forms a previously unknown Zn-finger, which we call AZUL (Amino-terminal Zn-finger of Ube3a Ligase) that may potentially play a role in Ube3A substrate recognition.

Methods and results

Cloning, expression and purification

DNA fragments encoding the N-terminal zinc binding domain were amplified by PCR from BC002582 cDNA template and sub-cloned into the pET28-MHL vector (http://www.sgc.utoronto.ca/SGC-WebPages/toronto-vectors.php) with an N-terminal polyhistidine tag. It was expressed in *Escherichia coli* strain BL21(DE3) growing in M9-minimal medium supplemented with ^{15}N ammonium chloride (1 g/L) and ^{13}C glucose (2 g/L). The protein was purified to homogeneity using Ni–NTA affinity chromatography as described previously (Yee et al. 2002). The purified protein was exchanged into the NMR buffer containing 10 mM MOPS, 450 mM NaCl, 10 mM DTT, 10 μM zinc sulphate, 1 mM benzamidine, 0.01% sodium azide, 1× protease inhibitor cocktail (Roche), pH 6.5.

Oligomeric state of the sample was characterized as monomeric by gel filtration on Superdex 200. The number of zinc atoms was deduced from LC-mass spectrometry and (Agilent Technologies) and corresponded to 1:1 stoicheometry.

Structure determination

NMR spectra were recorded at 25°C on Bruker Avance 600 or 800 MHz spectrometers equipped with cryo probes. The assignments of ^{1}H, ^{15}N and ^{13}C resonances were obtained by an ABACUS (Lemak et al. 2011) approach using the following experiments: HNCO, CBCA(CO)NH, HBHA(CO)NH, HNCA, (H)CCH-TOCSY and H(C)CH-TOCSY (Kay 1995). Peak picking was performed manually using Sparky (Goddard and Kneller 2003). The restraints for backbone φ and ψ torsion angles were derived from chemical shifts of backbone atoms using TALOS (Delaglio et al. 1995). Automated NOE assignment and structure calculations were performed using CYANA (version 2.1) (Güntert 2004). A total of 96% of NOESY peaks were assigned after seven iterative cycles of automated structure calculation and NOE assignment. The final 20 lowest-energy structures were refined within the CNS package (Brünger et al. 1998) by performing a short constrained molecular dynamics simulation in explicit solvent (Linge et al. 2003). Resulting structures were analyzed using MOLMOL (Koradi et al. 1996) and PSVS validation software (Bhattacharya et al. 2007).

The Zn-coordinated residues were identified by analysis of both the Cys chemical shifts and the initial structures that were calculated using only dihedral angle and NOE distance restrains. The $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts of the Cys44, Cys49, Cys54 and Cys83 residues are typical for zinc-coordinated cysteines with the probabilities (Kornhaber et al. 2006) of 0.95, 0.96, 0.80 and 0.99, respectively. Additional distance restraints between zinc ion and its ligands were introduced in the calculations of the final structure. These non-experimental distance restraints were developed based on the analysis of the geometrical properties of zinc binding sites observed in a dataset of high quality protein crystal structures (Alberts et al. 1998). These include: restraints of 2.25–2.40 Å between Zn and Sγ of Cys residues and restraints of 3.28–3.38 Å between Zn and Cβ of Cys residues. The distances between Sγ atoms of Cys residues that coordinate the zinc atom were also restrained within bounds of 3.70–4.00 Å.

Results

The resulting ensemble of 20 structures is shown in (Fig. 1). The backbone trace of the molecule is shown in

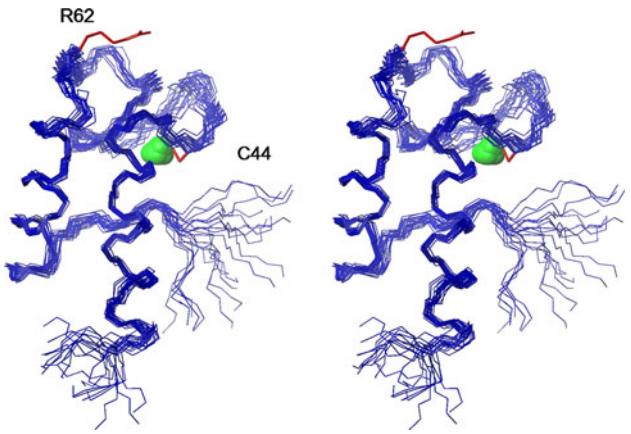

Fig. 1 Solution structure of the AZUL domain from Ube3A ubiquitin ligase. The ensemble of 20 structures is shown in a stereo view. The backbone of the domain is shown in *blue* and the side chains of Cys44 and Arg62 are *labeled* and *colored red*. MOLMOL software was used to create this figure (Koradi et al. 1996)

blue and the coordinated Zn atom is shown as a green sphere. The structure is well defined with a backbone r.m.s.d. of 0.72 ± 0.13 Å and the heavy atoms r.m.s.d. of 1.36 ± 0.18 Å (Table 1). Four residues at the amino-terminus and three residues at the C-terminus of the molecule are disordered in the NMR ensemble. The domain has a helix–loop–helix architecture with two helices oriented at a 53.6° angle with respect to each other. The N- and C-terminal helices are 15 and 13 residues long, respectively. Both helices are amphipathic with predominantly charged residues located on one side of the helix and hydrophobic residues located on the opposite side. The hydrophobic face of the C-terminal helix is composed of Ala68, Ala69, Ala72, Leu75 and Tyr76, and residues Ala29, Ile33, Tyr36 and Leu40 make up the hydrophobic side of the N-terminal helix. The two hydrophobic surfaces are facing each other forming the hydrophobic core of the domain.

The helices are connected with a 23 residue long loop (Fig. 2). The loop contains four Cys residues (44, 49, 54 and 57) and the well defined structure of such a long loop is held together by an atom of Zinc coordinated in tetrahedral geometry by three out of four cysteine residues present in it, namely Cys44, Cys49 and Cys54. Unlike Cys57, these three Cys residues are highly conserved (Fig. 3). The fourth Zn-coordinating group is provided by another conserved cysteine, Cys83, located in the extended C-terminus of the domain (Fig. 2, insert). Based on its $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts values Cys57 is unlikely to be Zn-coordinated, however, in our structure it still remains in close proximity (within 4 Å) to the coordinated metal.

Interestingly, we noticed that a subset of resonances in this domain was doubled. The doubled resonances were not of equal intensity; the minor peaks were about half the intensity of the major peaks, indicating potentially two

Table 1 NMR data and refinement statistics

NMR distance and dihedral constraints		
Distance restraints		
Total NOE		1248
sequential		293
Non-sequential		955
Intra-residual		340
Inter-residual		908
Medium-range		312
Long-range		303
Hydrogen bonds		23
Zinc-ligand distance restraints		12
Dihedral angles		
φ		45
ψ		45
Structure statistics		
Violations		
Distance constraints (mean ± SD) (Å)		0.014 ± 0.002
Dihedral angle constraints (mean ± SD) (°)		0.526 ± 0.119
Max. distance constraints violation (Å)		0.31
Max. dihedral angle violation (°)		4.8
Deviations from idealized geometry		
Bond lengths (Å)		0.013 ± 0.001
Bond angles (°)		0.87 ± 0.02
Impropers (°)		1.75 ± 0.13
Ramachandran plot[a]		
Core (%)		83.5
Allowed (%)		16.5
Generous (%)		0.0
Disallowed (%)		0.0
Average pairwise r.m.s.d.[b] (Å)		
Heavy		1.36 ± 0.18
Backbone		0.72 ± 0.13
	Raw	Z-score
Global quality scores[c]		
Procheck (phi-psi)[a]	−0.22	−0.55
Procheck (all)[a]	−0.23	−1.36
MolProbity clash	8.22	0.11
RPF scores[d]		
Recall	0.957	
Precision	0.944	
Dp-score	0.813	

[a] Values calculated for the ordered regions, as reported by PSVS: residues 4–59

[b] R.m.s.d calculated for residues 4–60

[c] Calculated by PSVS

[d] RPF scores (Huang et al. 2005) reflecting the goodness-of-fit of the structural ensemble to the NMR data

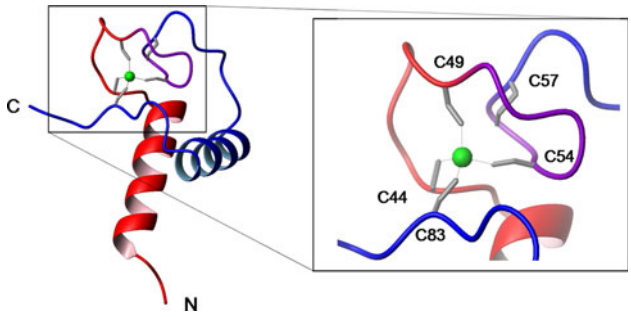

Fig. 2 Coordination of Zn in the AZUL domain. The lowest energy structure from the NMR ensemble is shown as a ribbon with N- and C-terminal helices colored in *red* and *blue*, respectively. The loop connecting two helices is colored gradually from *red* to *blue* for clarity and the Zn atom is shown as a *green sphere*. Side chains of all cysteine residues are shown in gray. The region forming Zn-finger is enlarged in the *insert*. Four cysteines, Cys44, Cys49, Cys54 and Cys83 coordinate atom of Zn and Cys57 remains reduced. MOLMOL software was used to create this figure (Koradi et al. 1996)

distinct conformations of the protein. All nine residues with two sets of NMR resonances (Gly43, Cys44, Gly45, Asn46, Glu47, Cys54, Ala55, Ser56 and Cys57) are located in the loop around the Zn atom and may reflect either a variant arrangement of four out of the potential five Zn-coordinating groups around the metal or exchange between apo- and holo-structures. The observed change in chemical shifts for the N and C atoms is not more than 0.4 ppm (with the exception of 0.7 ppm for backbone nitrogen of Gly43). In particular, both sets of resonances of Cys57 are typical for the reduced state. The change in chemical shifts for majority of H atoms is less than 0.04 ppm and largest change was observed for residue Asn45. CYANA calculations based on two different sets of chemical shifts representing major and minor conformations did not produce two clearly distinct conformations. This result was expected since for each atom with double resonances the strips in NOESY spectra that correspond to two different resonances of the atom have similar pattern of cross peaks. In order to distinguish between the two potential loop conformations in the Zn-binding domain we have used the program CS-Rosetta (Shen et al. 2008) to perform structure calculations based on two different sets of chemical shifts representing major and minor conformations. 15,000 initial models were generated for each set of chemical shifts out of which the best models were selected based on unassigned NOE data. NOESY peak list used in the selection of models corresponding to one of the two sets of chemical shifts does not include the peaks from NOESY spectra strips corresponding to resonances observed only in another set. The best models were then used as starting structures to generate 10,000 more models for each subset by aggressive sampling of the long loop and the C-terminus, and the best clusters were selected. The two resulting ensembles, however, were undistinguishable from each other. One particular feature of the structural ensembles obtained by both CYANA and CS-Rosetta is relatively high flexibility of the Zn-binding loop. For example, in the final ensemble of 20 models with superimposed two helices, the average global displacement of backbone atoms of residues 44–48 are of $\sim$ 1.2 Å which is 5 times as large as the average displacement of atoms in the helices. Corresponding to this loop flexibility, the positions of Zn atom in different models could differ one from other by more than 2 Å. Based on these results we suggest that the observed doubling of the subset of resonances most likely reflect exchange between two close Zn-binding

Fig. 3 Multiple sequence alignment of the AZUL domain in various species. Proteins containing regions homologous to the AZUL domain of human Ube3A are aligned. All Cys residues are highlighted in *red* and Arg62 is highlighted in *yellow*. Fully conserved residues are marked with (*) symbol at the bottom. Organism of origin and the GI identification number are shown on the *left-hand side* of each sequence. Secondary structure elements of the human AZUL domain are shown above its sequence for clarity. All aligned proteins are E3 ubiquitin ligases and contain C-terminal HECT domain (not shown) with the exception of protein from *Drosophila persimilis*, which is a Fe(II) oxygenase. Homologous domains are identified using protein blast against non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple sequence alignment is performed using CLUSTALW2 (www.ebi.ac.uk/Tools/clustalw2)

conformations rather than exchange between apo- and holo-structures.

Discussion and conclusions

Ube3A is a critical player in development of the high-risk oncogenic Human Papilloma Virus infection and Angelman syndrome. In the majority of cases, the latter disease is caused by truncating mutations in the Ube3 gene leaving the resulting protein without its catalytic HECT domain. However, several point mutations in the non-truncated full-length protein have been described as causing Angelman syndrome (Cooper et al. 2004; Bai et al. 2011). Most of the mutations are located in the catalytic domain of the Ube3A and disrupt its E3 function, but at least three disease-causing mutations were found in the non-catalytic N-terminal region of the protein, namely C44, S372 and P400. These may be potentially important for substrate recognition. While the amino-terminal part of the Ube3A is long thought to play a regulatory role in the enzyme's function the structural characterization of the region is still lacking.

In this work we present an NMR structure of an N-terminal domain of the Ube3A ligase. It has novel 3D fold and therefore we have named it the AZUL (Amino-terminal Zn-finger of Ube3a Ligase) domain. We show that it has a helix-loop-helix architecture and forms a new type of Zn-binding domain where the Zn-finger is formed by residues in a long loop wrapped around the metal atom (Cys44, Cys49, Cys54) and an extended C-terminus (Cys83) of the domain. The arrangement of the Cys residues in this Zn-finger is Cys-X_4-Cys-X_4-Cys-X_{28}-Cys. The long 23-residue loop connects two helices and remains well structured due to Zn-coordination. Zn-coordination plays an important role in overall domain's structure. It brings C-terminus and loop of the AZUL domain together and thus maintains the positioning of the two helices with respect to each other.

Several kinds of Zn-binding domains have been reported. They are generally small DNA/RNA or protein binding domains that can bind one or more Zn ions through cysteine and histidine side chains arranged in a particular order. The type and arrangement of the Zn-coordinating residues as well as Zn-finger function determine their classes. The three major folds of Zn-fingers include classical C2H2, which coordinates Zn with two cysteine and two histidine residues in an $\beta\beta\alpha$ arrangement, Treble-clef that consists of a β-hairpin with conserved CxxC motif (Zn knuckle) followed by an α-helix that provides additional Zn-binding groups, and Zinc ribbon fold that is made up of two β-hairpins and often forms anti-parallel β-sheet. The AZUL domain coordinates only one atom of Zn and does not contain β-turns/strands or known Zn-binding motifs. The search for proteins with similar 3D folds using DALI server (Holm and Rosenström 2010) did not produce similar Zn-binding domains, thus, making AZUL a novel Zn-finger. All top DALI matches (*Z*-score between 4.0 and 3.0) do not involve in the structural alignment residues from the Zn-ligating loop (residues 44–54) of the AZUL domain. The best match was obtain for elongation factor TS (PDB id code 1tfe, *Z*-score 4.0). The orientation of two helices of AZUL domain and the elongation factor are similar with an r.m.s.d. of 1.7 Å over 36 aligned residues, though sequence identity among the structurally equivalent residues is only 11%. It should be noted that two aligned helices of the elongation factor are connected by very short loop (residues 131–135).

The AZUL domain structure explains several earlier findings reported as a result of the mutational analysis of the Angelman syndrome patients. The disease-causing C44Y mutation, for example, was shown to cause rapid degradation of the Ube3A mutant in vivo compared to its wild type counterpart (Cooper et al. 2004). Indeed, as seen in Fig. 2, this mutation affects Cys residue crucial for Zn coordination in the presented structure. It is likely to have a significant effect on the overall domain structure and stability. Another non-HECT mutation found in double-mutant Angelman patients in addition to a severe truncating mutation is R62H (Malzac et al. 1998). Arg62 is not conserved in all organisms (Fig. 3) and the mutation does not cause AS by itself. The AZUL structure shows that Arg62 is located at the surface of the protein with its side chain exposed to the solvent. Thus, substitution of this residue with histidine is not likely to be critical for the domain stability.

Currently, the role of this new Zinc-binding domain is unknown, however, the context in which AZUL and similar domains are found may offer a clue. After performing a search for proteins containing sequences homologous to the AZUL domain in the non-redundant database of NCBI we found that the domain is always accompanied by a C-terminal HECT domain. The only non-E3 ligase exception is a Fe(II) oxygenase protein from *Drosophila persimilis*, however, it shows rather low sequence identity (49%) and lacks the fourth Cys residue responsible for Zn coordination in the Ube3A (Fig. 3). The presence of HECTc domain implies that the function of AZUL domain is closely associated with the HECT domain function. However, the specific role this new Zn-binding domain plays in Ube3A function in the brain and during HPV infection remains to be further investigated.

Acknowledgments Authors are thankful to Peter Loppnau for Ube3A constructs cloning. This work is supported by the US National Institute of Health Protein Structure Initiative (P50-GM62413-01 and GM67965) through the Northeast Structural and the Structural Genomics Consortium a registered charity (number 1097737) that receives funds from the Canadian Institutes of Health Research, the Canadian Foundation for Innovation, Eli Lilly, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Novartis,

the Ontario Minstry of Research and Innovation, Pfizer, and the Wellcome Trust. CHA holds a Canada Research Chair in Structural Proteomics. IB is supported by grant 1P30GM092369 from the NIH.

References

Alberts IL, Nadassy K, Wodak SJ (1998) Analysis of zinc binding sites in protein crystal structures. Protein Sci 7(8):1700–1716

Bai JL, Qu YJ et al (2011) A novel missense mutation of the ubiquitin protein ligase E3A gene in a patient with Angelman syndrome. Chin Med J Engl 124(1):84–88

Beaudenon S, Huibregtse JM (2008) HPV E6, E6AP and cervical cancer. BMC Biochem 9 suppl 1:S4.

Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66(4):778–795

Brünger AT, Adams PD et al (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54(Pt 5):905–921

Cooper EM, Hudson AW et al (2004) Biochemical analysis of Angelman syndrome-associated mutations in the E3 ubiquitin ligase E6-associated protein. J Biol Chem 279(39):41208–41217

Delaglio F, Grzesiek S et al (1995) Nmrpipe—a multidimensional spectral processing system based on unix Pipes. J Biomol NMR 6(3):277–293

Goddard TD, Kneller DG (2003) Sparky—NMR assignment and integration software.

Greer PL, Hanayama R et al (2010) The Angelman syndrome protein Ube3A regulates synapse development by ubiquitinating arc. Cell 140(5):704–716

Güntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378

Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38(Web Server issue): W545–W549

Huang YJ, Powers R, Montelione GT (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 127:1665–1674

Kay LE (1995) Pulsed field gradient multi-dimensional NMR methods for the study of protein structure and dynamics in solution. Prog Biophys Mol Biol 63(3):277–299

Koradi R, Billeter M and Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14(1):51–55, 29–32.

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

Lemak A, Gutmanas A et al (2011) A novel strategy for NMR resonance assignment and protein structure determination. J Biomol NMR 49(1):27–38

Linge JP, Williams MA et al (2003) Refinement of protein structures in explicit solvent. Proteins 50(3):496–506

Malzac P, Webber H et al (1998) Mutation analysis of UBE3A in Angelman syndrome patients. Am J Hum Genet 62(6): 1353–1360

Shen Y, Lange O et al (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105(12):4685–4690

Walboomers JM, Jacobs MV et al (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 189(1):12–19

Williams CA, Driscoll DJ, Dagli AI (2010) Clinical and genetic aspects of Angelman syndrome. Genet Med 12(7):385–395

Yee A, Chang X et al (2002) An NMR approach to structural proteomics. Proc Natl Acad Sci USA 99(4):1825–1830