FOR THE RECORD

Structural analysis of UBL5, a novel ubiquitin-like modifier

TERESA MCNALLY,[1] QIULONG HUANG,[1] RICHARD S. JANIS, ZHIHONG LIU, EDWARD T. OLEJNICZAK, AND REGINA M. REILLY

Global Pharmaceutical Research and Discovery, Abbott Laboratories, Abbott Park, Illinois 60064-6100, USA

(RECEIVED March 13, 2003; FINAL REVISION April 11, 2003; ACCEPTED April 11, 2003)

Abstract

UBL5 is a widely expressed human protein that is strongly conserved across phylogeny. Orthologs of UBL5 occur in every eukaryotic genome characterized to date. The yeast ortholog of UBL5, HUB1, was reported to be a ubiquitin-like protein modifier important for modulation of protein function. However, unlike ubiquitin and all other ubiquitin-like modifiers, UBL5 and its yeast ortholog HUB1 both contain a C-terminal di-tyrosine motif followed by a single variable residue instead of the characteristic di-glycine found in all other ubiquitin-like modifiers. Here we describe the three-dimensional structure of UBL5 determined by NMR. The overall structure of the protein was found to be very similar to ubiquitin despite the low ~25% residue similarity. The signature C-terminal di-tyrosine residues in UBL5 are involved in the final β sheet of the protein. This is very different to the di-glycine motif found in ubiquitin, which extends beyond the final β sheet. In addition, we have confirmed an earlier report of an interaction between UBL5 and the cyclin-like kinase, CLK4, which we have determined is specific and does not extend to other cyclin-like kinase family members.

Keywords: UBL5; beacon; ubiquitin; NMR spectroscopy; structure determination; cyclin-like kinase

Supplemental material: See www.proteinscience.org.

UBL5 was initially identified in a screen for highly expressed genes in human iris (Friedmann et al. 2001). The gene encodes a protein of 73 amino acids with a molecular weight of 8.5 kD and a pI of 8.6. The amino acid sequence of UBL5 is identical to that of Beacon (Collier et al. 2000), a protein reported to be involved in feeding behavior and development of obesity and type 2 diabetes in the Israeli sand rat *Psammomys obeseus*. Quantitative RT-PCR of Beacon expression in the rat hypothalamus demonstrated a correlation between beacon mRNA levels, body fat percentage, and plasma insulin levels. Recent reports indicate that UBL5 may also interact with the cyclin-like kinase CLK4 (Collier et al. 2000). Cyclin-like kinases have been reported to be involved in the regulation of the activity of PTP1B (Moeslein et al. 1999) a protein that is involved in the development of both diabetes and obesity.

Based on sequence homology and structure prediction algorithms it is predicted that the protein UBL5 has a structure similar to ubiquitin (Friedmann et al. 2001). The sequence similarity is, however, weak, and the protein does not contain the characteristic di-glycine residues at its C terminus that are required for ubiquitin-like modifiers to conjugate to their target proteins (Hochstrasser 1996; Jentsch and Pyrowolakis 2000; Glickman and Ciechanover 2002). The yeast ortholog of UBL5, HUB1, has been re-

Reprint requests to: Edward T. Olejniczak, Global Pharmaceutical Research and Discovery, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064-6100, USA; e-mail: Edward.olejniczak@abbott.com; fax: (847) 938-2478.

[1]These authors contributed equally to this work.

Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RMSD, root-mean-squared deviation; CLK, cyclin-like kinase; ppm, parts per million.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.0382803.

ported to have ubiquitin-like modifier activity (Jentsch and Pyrowolakis 2000; Dittmar et al. 2002). UBL5 has been shown to be interchangeable with HUB1 in a yeast-based assay that measured its ability to rescue HUB1 mutants from a mating projection defect (Dittmar et al. 2002). These data from yeast suggest that UBL5 like HUB1, covalently attaches to other proteins and based on its sequence does so in a manner distinct from the di-glycine containing ubiquitin-like proteins.

Although several predictions have been made for the structure of UBL5, no experimental data are available. Here, we describe the three-dimensional structure of UBL5 and compare it to that of ubiquitin. We also report on our studies of the interaction of UBL5 with CLK4 and other cyclin-like kinase family members using yeast two hybrid studies.

Results

Yeast two-hybrid interactions of UBL5

UBL5 has been reported to interact with the cyclin-like kinase CLK4 (Collier et al. 2000). CLK4 is a member of the LAMMER family of protein kinases, which includes three other closely related proteins (Schultz et al. 2001). Yeast two-hybrid analysis was used to determine whether UBL5 could interact with CLK1 and CLK2, the most closely related family members with the highest sequence similarity to CLK4. Constructs comprising the full length of CLK1 and CLK2 genes in a two-hybrid bait vector were transformed into yeast strain AH109 containing the UBL5 gene fused to the GAL4 activating domain. CLK1 and 2 are known to undergo alternative splicing that, due to the presence of a stop codon in each intervening exon, results in a truncated protein product (Menegay et al. 2000). This differential splicing has not, as yet, been reported for CLK4, but such splice variants were identified in our attempts to isolate the CLK4 transcript from cDNA libraries by RT-PCR.

The only construct that grew on nonpermissive minimal medium was the CLK4 full-length construct, and not the C-terminally truncated splice variant. None of the aforementioned constructs grew on permissive minimal medium when transformed with the GAL4-activating domain alone, suggesting that the CLK4 interaction was specific to UBL5. Because the splice variant of CLK4 did not interact with UBL5, it can be assumed that the interaction between UBL5 and CLK4 is mediated by the C terminus of the protein, where the kinase domain resides. Interestingly, this is where CLK4 and, in particular CLK1, are the most similar in sequence. Because the interaction was detected using yeast two-hybrid screening it is still not known if the interaction is covalent or not.

Protein structure determination

An N-terminal His-tagged version of UBL5 was expressed in *Escherichia coli*, isotopically labeled, and purified. After cleavage of the His tag the protein was found to aggregate at concentrations above 0.1 mM over a wide pH range and was thus unsuitable for NMR structural studies. The uncleaved protein was stable and was used in the structural studies. The ^{15}N HSQC spectrum of the tagged protein was essentially identical to that of the wild-type protein, indicating that the fold of the protein is preserved. The backbone and side-chain resonances of the protein were assigned from standard heteronuclear three-dimensional NMR experiments (see Materials and Methods) recorded on a protein sample that was uniformly labeled with ^{15}N and ^{13}C.

The structure of the UBL5 was determined from a total of 952 unambiguous NMR-derived distance and torsion angle restraints along with 307 ambiguous distance restraints. Figure 1A depicts a backbone (N,C$^{\alpha}$,C′) superposition of 10 low-energy structures that were derived from the NMR data using the program CNX (MSI). Excluding the N-terminal tag the atomic root-mean-squared-deviation (RMSD) about the mean position is 0.51 ± 0.06 Å for the backbone atoms and 1.02 ± 0.09 Å for all heavy atoms. There are no dihedral-angle violations greater than 5° and no NOE violations greater than 0.4 Å. Only covalent geometry, NOE, torsion, and repulsive terms were included in the structure refinement. Even so, the Lennard-Jones energy is large and negative (−380 ± 13 kcal mole^{-1}), indicating that the structures have favorable nonbonded contacts. A summary of the structural statistics is given in Table 1 of the supplementary material. Analysis of the average-minimized structures (1–73) with the program PROCHECK showed that 70% of the residues for UBL5 lie in the most favored region of the Ramachandran plot, while an additional 28.7% lie in allowed regions (Laskowski et al. 1993).

A ribbon representation (Carson 1987) of the averaged minimized NMR structure of UBL5 is shown in Figure 1B. The first β-sheet of the protein starts at its native N terminus. This sheet is connected by a tight turn to another strand of antiparallel β-sheet. An α-helix follows, and this is connected to a third short sheet. The final sheet of the protein runs parallel with the second β-sheet and antiparallel to the short third sheet. The final β-sheet of the protein ends at the next to last residue of the protein, tyrosine 72. The C-terminal di-tyrosine residues have the characteristic chemical shifts of residues in a β-sheet, and have the expected cross-strand NOEs to residues on strands 2 and 3. These data unambiguously indicate that these residues are involved in the final β-sheet of the protein and are not exposed residues on the surface of the protein.

Discussion

UBL5 protein is extremely well conserved across phylogeny; all mammalian sequences identified are identical at the amino acid level. This strong conservation, from yeast to mammals, suggests an important function for UBL5. In a

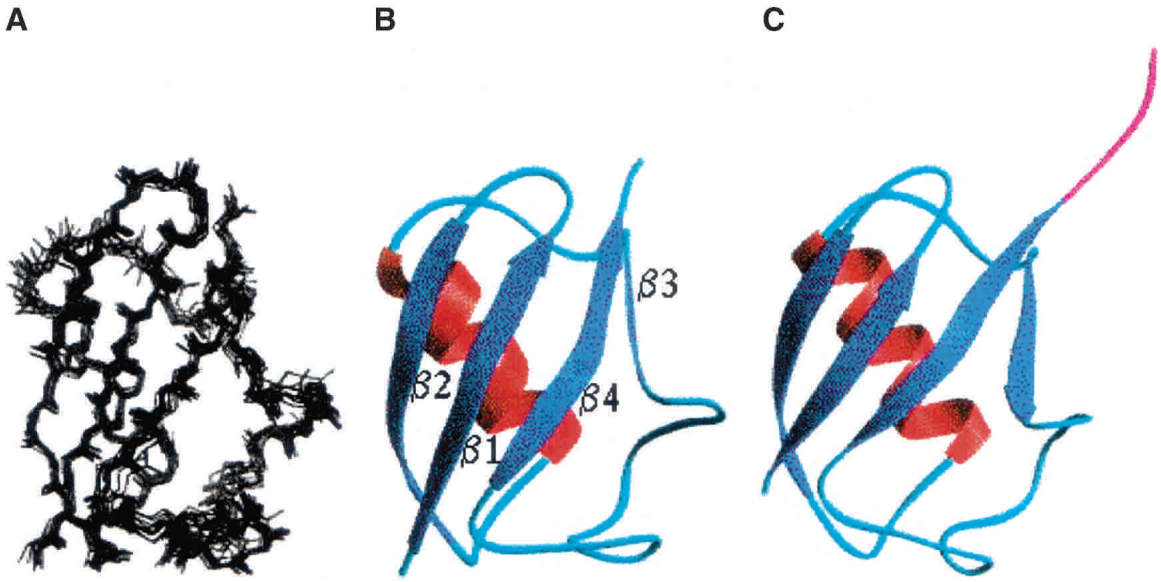

Figure 1. (*A*) Backbone (N,C$^{\alpha}$,C) superposition of 10 low-energy NMR-derived structures for UBL5. The root-mean-square deviation (RMSD) to the mean coordinate positions is 0.51 ± 0.06 Å for the backbone atoms (N,Cα,C′) and 1.02 ± 0.09 Å for all heavy atoms. (*B*) Ribbons depiction of the average-minimized structure for UBL5. The sheets are numbered in the same manner as ubiquitin. (*C*) Ribbons depiction of ubiquitin.

recent publication (Dittmar et al. 2002) the *Saccharomyces cerevisiae* HUB1 was demonstrated to play a role in polarized cell morphogenesis. Under different growth conditions, HUB1 becomes conjugated onto several proteins according to growth conditions including SPH1 and YDL223c. The strong sequence conservation between HUB1 and UBL5 as well as its ability to substitute for HUB1 in yeast are all evidence that UBL5 is also a ubiquitin-like protein modifier (Dittmar et al. 2002).

Yeast two hybrid has been used to confirm earlier studies that indicated an interaction between UBL5 and CLK4. CLK4 is the most recently identified, and thus most poorly characterized member of the CLK family. These proteins are believed to play a role in phosphorylation of serine and arginine rich non-snRNP splicing proteins (Nayler et al. 1997) and the control of differential splicing, and may also regulate the activity of PTP1B (Moeslein et al. 1999). It remains to be determined whether CLK4 can also phosphorylate PTP1B. The fact that CLK4 appears to be the only member of this protein family that interacts with UBL5 suggests that CLK4 may have a role distinct from that of the other CLKs.

Based on sequence homology it was predicted that UBL5 would be structurally similar to ubiquitin (Friedmann et al. 2001). This is indeed what we have found. In Figure 2, we show the secondary structure of UBL5 and a sequence alignment with ubiquitin. The secondary structures of the two proteins are identical. The additional C-terminal residues important for conjugation in ubiquitin-like protein modifiers are, however, missing in UBL5. The backbone atom RMSD of residues 1–70 of ubiquitin (Cornilescu et al. 1998; PDB entry 1d3z) and 3–72 of UBL5 is 1.53 Å. A ribbons representation of ubiquitin is shown in Figure 1C, which can be compared to UBL5 in Figure 1B. The only

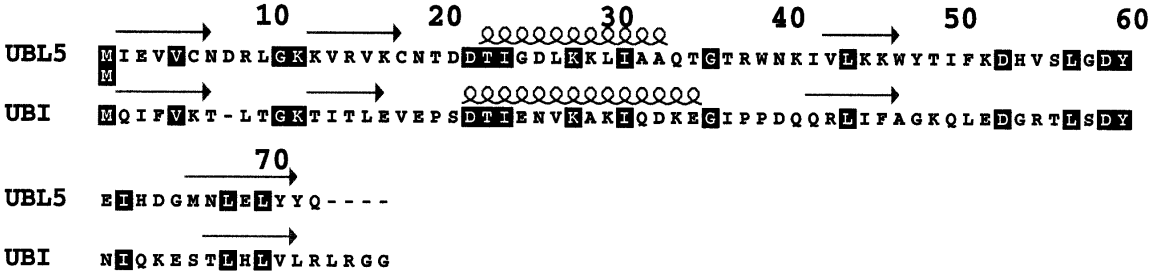

Figure 2. Alignment of UBL5 and Ubiquitin (Ubi) showing the common secondary structure and low sequence identity. The additional C-terminal residues important for conjugation in ubiquitin-like protein modifiers are missing in UBL5.

notable difference between the two structures is the unstructured C terminus found in ubiquitin that extends beyond the end of the highly structured portion of the protein. These residues of ubiquitin are colored magenta in Figure 1C. Based on the strong sequence homology between UBL5 and its orthologs we would expect them to have homologous structures. We would also expect the UBL5 homologs to all have the same structured C terminus, which is very different than that found for ubiquitin and its related modifiers (Bayer et al. 1998; Cornilescu et al. 1998; Rao-Naik et al. 1998; Whitby et al. 1998). This is indeed the case for the yeast ortholog of UBL5, HUB1. The overall fold of HUB1 (PDB entry 1m94) and UBL5 are identical. The backbone RMSD between the two proteins is 1.66 Å.

The residues in the C terminus of UBL5 are likely to be important for conjugation. Studies on HUB1 (Dittmar et al. 2002) suggest that cleavage of the last residue of HUB1 resulted in more efficient HUB1-SPH1 conjugate formation than the wild-type protein. This suggests that the last residue, found to be variable in orthologs of HUB1 and, in fact, different in HUB1 and UBL5, is not important for conjugation. These data may also indicate that C-terminal proteolysis may be an activating step for conjugation.

In summary, we have presented the tertiary structure of UBL5, and have shown that its overall fold is similar to ubiquitin. If UBL5 behaves like a ubiquitin-like modifier (Jentsch and Pyrowolakis 2000; Dittmar et al. 2002) then its mechanism of protein conjugation should be quite distinct from that of other ubiquitin-like modifiers because the structure, described herein, shows that this protein has neither the unstructured C terminus nor the C-terminal di-glycine motif, which is found for other proteins of this class. The widespread expression of UBL5 in all tissue types suggests that this protein and its orthologs have an important functional role. The data presented here may aid in the elucidation of the conjugation mechanism of this class of protein modifiers.

Materials and methods

Plasmid construction for structural studies

Several different UBL5 constructs were prepared and evaluated for their suitability for NMR structural studies. The coding sequence of UBL5 (GenBank accession no. NM_024292, nucleotides 66–287) was amplified by PCR with primers encoding 5′- and 3′-restriction sites. The PCR product was digested and ligated into the Nde I and Xho I sites of the pET28b(+) plasmid (Novagen), providing the N-terminal His-tagged (MGSSHHHHHHSSGLVPRGSH) protein. Constructs were verified by DNA sequencing.

Expression and purification for structural studies

The UBL5 protein used for structural studies was expressed in *E. coli* BL21(DE3) grown on M9 media, and purified using a Ni-NTA affinity chromatography. Uniformly ^{15}N-labeled and uniformly ^{15}N,^{13}C-labeled samples were prepared with media containing either $^{15}NH_4Cl$ or $^{15}NH_4Cl$ plus [U-^{13}C]glucose. NMR samples contained 0.5–1.0 mM protein in either 90% H_2O/10% ^2H_2O or 100% ^2H_2O, 50 mM sodium phosphate (pH 5.7), and 2 mM ^2H-dithiothreitol.

NMR spectroscopy

All NMR data were acquired at 298 K on a Bruker DRX500 or DRX600 NMR spectrometer. Backbone ^1H, ^{13}C, and ^{15}N resonance assignments were achieved with [^{15}N,^{13}C] UBL5 using a suite of triple-resonance experiments (HN(CA)CB, HN(CO-CA)CB, HNCO, and HN(CA)C (Clore and Gronenborn 1998; Kanelis et al. 2001). The side-chain ^1H and ^{13}C NMR signals were assigned from HCCH-TOCSY experiments. NOE distance restraints were obtained from three-dimensional ^{15}N- and ^{13}C-edited NOESY spectra acquired with a mixing time of 80 msec (Ikura et al. 1990).

Structure calculations

UBL5 structures were calculated using a simulated annealing protocol with the program CNX (MSI; Brunger 1992). A square-well potential (F_{NOE} = 50 kcal mole^{-1}) was used to constrain NOE-derived distances. Based on the cross-peak intensities, NOE-derived distance restraints were given upper bounds of 3.0, 4.0, or 5.0 Å. In the refinement stage, additional ambiguous constraints were added, with an upper bound of 6.0 Å, for unassigned cross-peaks that were consistent with the chemical shift table (i.e., error bars of 0.07 ppm for protons, 0.7 ppm for hetero atoms) and the structure. Torsion angle restraints, ϕ and ψ, were generated from analysis of N, C′, C^α, and H^α chemical shifts using the TALOS program (Cornilescu et al. 1999). A force constant of 200 kcal mole^{-1} rad^{-2} was applied to all torsional restraints. Explicit hydrogen bonds were included in α-helices and β-sheets for residues observed to have slowly exchanging amide protons and having appropriate short range NOEs. The program PROCHECK was employed to analyze the geometric quality of the calculated structures in the ensemble (Laskowski et al. 1993). The average minimized coordinates for UBL5 have been deposited with the Protein Data Bank (RCSB018874; PDB number 1POR).

Yeast two-hybrid analysis

Yeast two-hybrid analysis was performed using the Matchmaker 2 two-hybrid kit (Clontech). Interactions were identified following the manufacturer's recommended protocol. The UBL5 gene was cloned in frame with GAL4 in pGBKT7, and expression was verified by Western blot using anti-GAL4 antibody (Clontech). Two hybrid positives were identified by growth on selective LWHA X-Gal-containing media only in the presence of pGBKT7UBL5 and not pGBKT7 alone.

Electronic supplemental material

Table of structural statistics for UBL5.

Acknowledgments

We are grateful to Dr. Terry Opgenorth, Dr. Christine Collins, and Dr. James Trevillyan for critical evaluation of the manuscript.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R., and Becker, J. 1998. Structure determination of the small ubiquitin-related modifier sumo-1. *J. Mol. Biol.* **280:** 275–286.

Brunger, A.T. 1992. *X-PLOR version 3.1*. Yale University Press, New Haven, CT.

Carson, M. 1987. Ribbons. *J. Mol. Graph.* **5:** 103–106.

Clore, G.M. and Gronenborn, A.M. 1998. NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* **2:** 564–570.

Collier, G.R., McMillan, J.S., Windmill, K., Walder, K., Tenne-Brown, J., de Silva, A., Trevaskis, J., Jones, S., Morton, G.J., Lee, S., et al. 2000. Beacon—A novel gene involved in the regulation of energy balance. *Diabetes* **49:** 1766–1771.

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

Cornilescu, G., Marquardt, J.L., Ottiger, M., and Bax, A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120:** 6836–6837.

Dittmar, G.A.G., Wilkinson, C.R.M., Jedrzejewski, P.T., and Finley, D. 2002. Role of a ubiquitin-like modification in polarized morphogenesis. *Science* **295:** 2442–2446.

Friedmann, J.S., Koop, B.F., Raymond, V., and Walter, M.A. 2001. Isolation of a ubiquitin-like (UBL5) gene from a screen identifying highly expressed and conserved iris genes. *Genomics* **71:** 252–255.

Glickman, M.H. and Ciechanover, A. 2002. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiol. Rev.* **82:** 373–428.

Hochstrasser, M. 1996. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30:** 405–439.

Ikura, M., Kay, L.E., Tschudin, R., and Bax, A. 1990. Three-dimensional NOESY-HMQC spectroscopy of a ^{13}C-labeled protein. *J. Magn. Reson.* **86:** 204–209.

Jentsch, S. and Pyrowolakis, G. 2000. Ubiquitin and its kin: How close are the family ties? *Trends Cell Biol.* **10:** 335–342.

Kanelis, V., Forman-Kay, J.D., and Kay, L.E. 2001. Multidimensional NMR methods for protein structure determination. *IUBMB Life* **52:** 291–302.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. Procheck. *J. Appl. Crystallogr.* **26:** 283–291.

Menegay, H.J., Myers, M.P., Moeslein, F.M., and Landreth, G.E. 2000. Biochemical characterization and localization of the dual specificity kinase CLK1. *J. Cell Sci.* **113:** 3241–3253.

Moeslein, F.M., Myers, M.P., and Landreth, G.E. 1999. The CLK family kinases, CLK1 and CLK2, phosphorylate and activate the tyrosine phosphatase, PTP-1B. *J. Biol. Chem.* **274:** 26697–26704.

Nayler, O., Stamm, S., and Ullrich, A. 1997. Characterization and comparison of four serine- and arginine-rich (sr) protein kinases. *Biochem. J.* **326:** 693–700.

Rao-Naik, C., delaCruz, W., Laplaza, J.M., Tan, S., Callis, J., and Fisher, A.J. 1998. The rub family of ubiquitin-like proteins—Crystal structure of arabidopsis Rub1 and expression of multiple rubs in arabidopsis. *J. Biol. Chem.* **273:** 34976–34982.

Schultz, J., Jones, T., Bork, P., Sheer, D., Blencke, S., Steyrer, S., Wellbrock, U., Bevec, D., Ullrich, A., and Wallasch, C. 2001. Molecular characterization of a cDNA encoding functional human CLK4 kinase and localization to chromosome 4q35. *Genomics* **71:** 368–370.

Whitby, F.G., Xia, G., Pickart, C.M., and Hill, C.P. 1998. Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. *J. Biol. Chem.* **273:** 34983–34991.