Solution structure of Urm1 and its implications for the origin of protein modifiers

Junjie Xu[†], Jiahai Zhang[†], Li Wang[‡], Jie Zhou[‡], Hongda Huang[†], Jihui Wu[†], Yang Zhong[‡§¶], and Yunyu Shi[†‖]

[†]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology, Hefei, Anhui 230026, People's Republic of China; [‡]School of Life Sciences, Fudan University, Shanghai 200433, People's Republic of China; and [§]Shanghai Center for Bioinformation Technology, Shanghai 201203, People's Republic of China

Protein modifiers are involved in diverse biological processes and regulate the activity or function of target proteins by covalently conjugating to them. Although ubiquitin and a number of ubiquitin-like protein modifiers (Ubls) in eukaryotes have been identified, no protein modifier has been found in prokaryotes; thus, their evolutionary origin remains a puzzle. To infer the evolutionary relationships between the protein modifiers and sulfur carrier proteins, we solved the solution NMR structure of the Urm1 (ubiquitin-related modifier-1) protein from *Saccharomyces cerevisiae*. Both structural comparison and phylogenetic analysis of the ubiquitin superfamily, with emphasis on the Urm1 family, indicate that Urm1 is the unique "molecular fossil" that has the most conserved structural and sequence features of the common ancestor of the entire superfamily. The similarities of 3D structure and hydrophobic and electrostatic surface features between Urm1 and MoaD (molybdopterin synthase small subunit) suggest that they may interact with partners in a similar manner, and similarities between Urm1–Uba4 and MoaD–MoeB establish an evolutionary link between ATP-dependent protein conjugation in eukaryotes and ATP-dependent cofactor sulfuration.

evolution | NMR structure

Protein modifiers are involved in diverse biological processes and regulate the activity or function of target proteins by covalently conjugating to them. The first identified protein modifier was ubiquitin, which is an abundant and ubiquitous protein that is covalently attached to other proteins, either as a tag for targeted protein degradation in the proteosome or as a regulatory post-translational modification (1, 2). Over the past 20 years, a series of ubiquitin-like protein modifiers (Ubls) have been identified that undergo similar cascade enzymatic pathways in the conjugation process (3). Ubiquitin is universal in all eukaryotes (4), and Ubls are prevalent in eukaryotes, yet no protein modifier has been identified in prokaryotes up to now. Therefore, solving the puzzle of the evolutionary origin of protein modifiers has become an important but challenging issue in both structural biology and comparative genomics.

There are mechanistic parallels between activation by ubiquitin and activation by certain sulfur carrier proteins. In *Escherichia coli*, two sulfur carrier proteins, ThiS (involved in thiamin biosynthesis) and MoaD (molybdopterin synthase small subunit), are activated in an ATP-dependent manner by sulfurtransferase, ThiF, and MoeB, respectively, and then form thioesters at the C-terminal diglycin with a sulfur atom (5–7). The mechanism is similar to the activation of ubiquitin by the ubiquitin-activating enzyme E1. In addition, ThiS and MoaD possess the β-grasp fold (8, 9). On the basis of functional and structural similarities, ThiS and MoaD were proposed as prokaryotic homologs of ubiquitin. Recently, this hypothesis was supported by the identification of Urm1 (ubiquitin-related modifier-1) in *Saccharomyces cerevisiae*, which is a unique protein modifier sharing sequence homology with ThiS and MoaD (10) and has been suggested as an evolutionary link between ubiquitin and Ubls in eukaryotes and ancient sulfur carrier proteins. In addition, Uba4, the E1-like enzyme in the Urm1 conjugation pathway, shows strong sequence similarity to MoeB and ThiF (10). However, Urm1 lacks sequence homology with ubiquitin, and its sequence identities with ThiS and MoaD are 20% and 23%, respectively. The structure of *Mus musculus* AAH26994.1 protein, which shares 44% sequence identity with Urm1, provided important structural evidence for the evolutionary relationship (11). However, the function of this protein is unclear. There is no Uba4 counterpart identified in *M. musculus*. It is unclear whether the AAH26994.1 protein functions as a protein modifier. Structural similarity or functional similarity alone cannot establish evolutionary relationship, and, without structural information on *S. cerevisiae* Urm1, this evolutionary link remains unclear.

On the other hand, structural information on Urm1 may also provide new functional insights. In fact, it was found that Urm1 covalently attaches to antioxidant protein Ahp1 to modulate its activity in oxidant-stress response (12). The gene for all urmylation pathway proteins, Urm1 and Uba4, is essential for *S. cerevisiae* viability during budding in vegetative growth and is shown to play a role in invasive growth into agar in the haploid state and in pseudohyphal growth and cell elongation under starvation conditions in the diploid state. A functional cross-link between the TOR (target of rapamycin) signaling pathway and the urmylation pathway was also detected, in which Urm1 was shown to be involved in nutrient sensing (13). Structural analysis of Urm1 would shed light on the binding face of Urm1, which is essential to better understand the function and mechanism of the modification pathway.

In this paper, we solve the solution NMR structure of Urm1 protein in *S. cerevisiae* to infer the evolutionary relationships between the protein modifiers and sulfur carrier proteins and explore the function and interaction pattern of the Urm1 conjugation system. We use both structural comparison and phylogenetic analysis of the ubiquitin superfamily to test the hypothesis that Urm1 is a "molecular fossil" in the superfamily and has the most conserved structural features of the family's common ancestor.

Results and Discussion

Solution Structure of Urm1. C-terminal His-tagged Urm1 was expressed in *E. coli*. The isotopically labeled and purified protein sample possesses good solubility and stability in the NMR buffer. Secondary structural elements (SSEs) of Urm1 were identified by a typical NOE pattern; chemical shift index and hydrogen exchange generally confirmed the results (Fig. 6, which is published as supporting information on the PNAS web site). The NMR data used for structure calculations are summarized in Table 1. A final

EVOLUTION

Table 1. NMR and structural statistics

NMR restraints in the structure calculation	
Intraresidue	419
Sequential ($\|i - j\| = 1$)	475
Medium-range ($\|i - j\| < 5$)	302
Long-range ($\|i - j\| \geq 5$)	407
Hydrogen bonds	52
Total distance restraints	1,655
Dihedral angle restraints	79
Lennard–Jones potential energy, kcal·mol^{-1}	-267.58 ± 12.32
rmsd from idealized covalent geometry	
Bonds, Å	0.0008 ± 0.00004
Angles, °	0.2650 ± 0.0023
Impropers, °	0.1020 ± 0.0030
rmsd from experimental restraints	
Distance, Å	0.0036 ± 0.0005
Constrained dihedral, °	0.0155 ± 0.0086
Coordinate rmsd for residues 2–95, Å	
All backbone atoms	0.58
All heavy atoms	1.04
Ramachandran plot, % residues	
Most favored regions	74.5
Additional allowed regions	23.0
Generously allowed regions	2.0
Disallowed regions	0.5

None of the structure exhibits distance violations >0.5 Å or dihedral angle violations >5°.

set of 20 structures with lowest energy (Fig. 1A) was selected for structural statistic calculation. The structure of Urm1 contains one five-strand β-sheet and four α-helices. The β-sheet is arranged in the order 21534, as in ubiquitin; four helices together back the curved β-sheet on the concave side; and the C terminus (residues 95–99) is flexible and protrudes from the globular fold as a tail (Fig. 1). Interaction between the inner face of the α2-helix and the concave face of the β-sheet form a hydrophobic core, which is essential to maintain the compact fold. The short one-turn α4-helix follows the β4-strand tightly. Long-range NOEs were observed from the N terminus (residues 33–35) of α2-helix to the residues (80–83) after the C terminus of α4-helix. The α1-helix together with α3-helix close the C-terminal part of the globular fold. The most similar structure to Urm1 is that of MoaD-related protein from *Thermus thermophilus* (TtMoaD). The SSEs and overall structures of Urm1 and TtMoaD are almost the same. The rms deviation
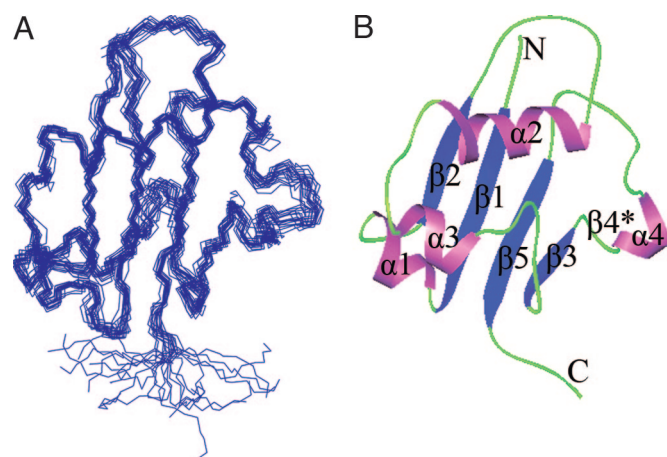

Fig. 1. NMR structure of Urm1. (*A*) Backbone overlay of 20 NMR structures with the lowest energy from the final CNS v.1.1 calculation. (*B*) Ribbon representation of Urm1 (strand β4* was identified by NOE connectivity).

(rmsd) between structure of TtMoaD [Protein Data Bank (PBD) ID code 1V8C], and the average structure of Urm1 is 2.3 Å with a corresponding Z score of 10.6 [calculated by the DALI algorithm (14)].

Partner-Binding Interface Related to Function. As a protein modifier, the function of Urm1 concerns multiple protein–protein interactions. Hydrophobic regions exposed to solvents are most likely to be the partner-binding regions. Multiple-sequence alignment reveals, besides residues involved in hydrophobic core, three hydrophobic residues (Ile-66, Leu-68, and Leu-76) that are type-conserved in Urm1 orthologs (Fig. 2). These residues, together with Leu-9, form a hydrophobic patch on the exposed side of the β-sheet (Fig. 3A). A similar hydrophobic patch on the surface of MoaD (Fig. 3B) is involved in the interaction between MoaD and its partners (9, 15). ThiS and other protein modifiers also possess the exposed hydrophobic region (8, 16). By analogy, the hydrophobic patch also may be essential for interaction between Urm1 and its binding partners, such as Uba4.

Urm1 Arg-20 sits in a position similar to Arg-11 in MoaD relative to the C-terminal tail and hydrophobic patch, which contributes to form a hydrogen bond or ionic bond with MoeB. Furthermore, the Urm1 Asp-13 is similar to MoaD Glu-12 in structural location (Fig. 3 *A* and *B*). The Asp-13 and Arg-20 of Urm1 also are type-conserved among all of the Urm1 orthologs (Fig. 2). Therefore, the two electrostatic residues also may be essential in Urm1–Uba4 interactions, similar to the MoaD–MoeB complex. In addition, the overall electrostatic surface of Urm1 is dominated by negative potential as in MoaD (Fig. 3 *C* and *D*). The dispersion of positive potential of the two proteins also is comparable. In contrast, the surfaces of other modifiers are usually divided roughly into "acidic" face and "basic" face as in ubiquitin (17–19). The positive potential region formed by Arg-6, Arg-54, Arg-42, and Arg-72 in ubiquitin is important for its activating and binding. In contrast, the corresponding regions on Urm1 and MoaD exhibit a combination of negative and hydrophobic characters. In Urm1, Arg-62 sits in a position similar to Lys-50 in MoaD relative to the C-terminal tail and may be involved in binding interactions. Therefore, the Urm1 conjugation pathway may employ a different recognition mechanism from ubiquitination, although it is more similar to MoaD–MoeB recognition.

In general, the solved Urm1 structure provides some insights into the function and interaction of the Urm1 protein. Similar 3D structure and hydrophobic and electrostatic surface features of Urm1 and MoaD suggest that they may interact with partners in a similar manner. Similarity between Urm1–Uba4 and MoaD–MoeB demonstrates an evolutionary relationship between protein modifiers and certain sulfur carrier proteins and a link between ATP-dependent protein conjugation in eukaryotes and ATP-dependent cofactor sulfuration.

Structural Comparison Within the Ubiquitin Superfamily. To illustrate the structural features of Urm1 protein and compare it with the structures of other proteins in the ubiquitin superfamily, we compared all 26 structures in the superfamily available to date by using a cluster analysis of their structural similarity. In the dendrogram shown in Fig. 4, it is obvious that there is a highly significant structural similarity throughout the entire ubiquitin superfamily, and all of the family members share a similar β-grasp fold. On the other hand, protein modifiers and sulfur carriers each have specific features of folding. The dendrogram shows that the structures in the ubiquitin superfamily can be classified into three clusters: a ubiquitin-related fold, a ThiS-related fold, and a MoaD-related fold (Fig. 4). In the ubiquitin-related fold, the 3_{10} helix (Fig. 4, cyan circle) and the core helix interact through their N-terminal residues, which are perpendicular to each other. In the ThiS-related fold, the short helix (Fig. 4, purple circle) follows the β4-strand immediately and interacts with the core helix N terminus residues by its
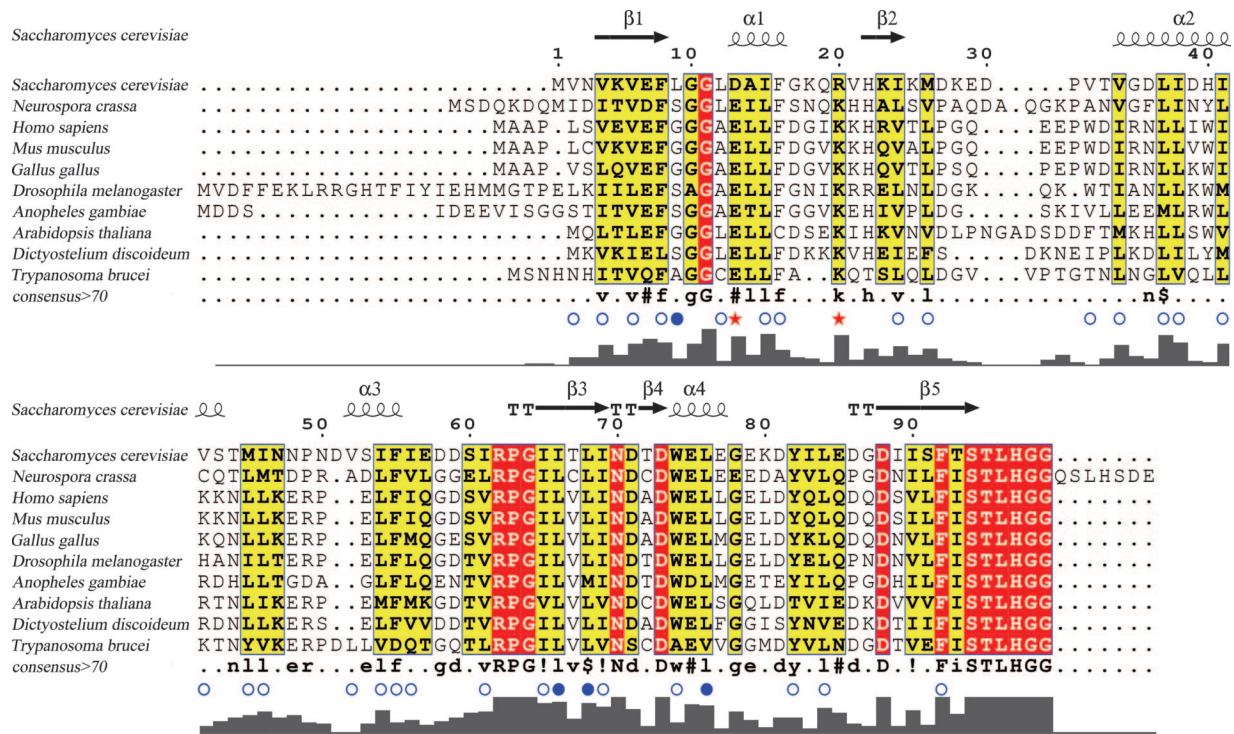
Fig. 2. Multiple sequence alignment of Urm1 from *S. cerevisiae* with its putative orthologs from various other species. Sequences were aligned by using ClustalX. The 70% consensus sequence was generated by ESPript: capital letters indicate identity, and lowercase letters indicate a consensus level of >0.5. !, any one of IV; $, any one of LM; %, any one of FY; #, any one of NDQEBZ. The hydrophobic core residues in ScUrm1 are indicated by open circles on the bottom of the alignment, and the hydrophobic residues exposed outward are indicated by filled circles. Two stars mark the charged residues that may be involved in Urm1-binding interactions.
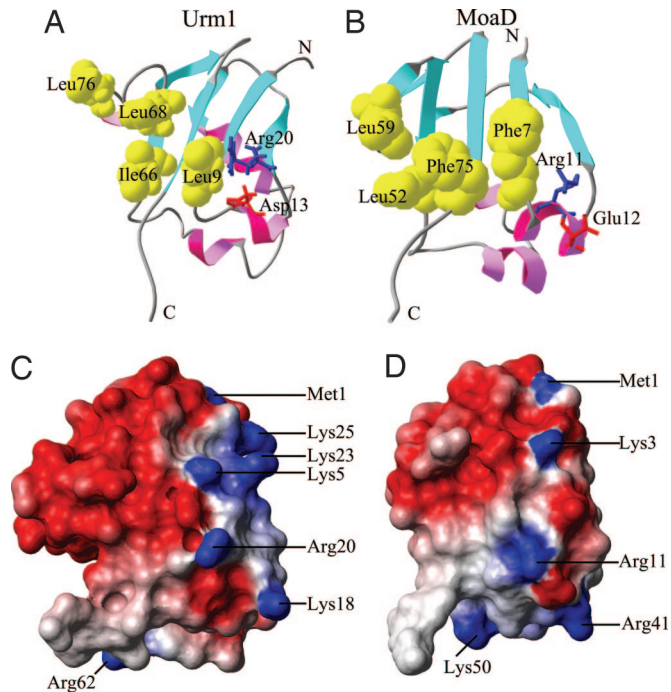

Fig. 3. Structure–function relationship between Urm1 and MoaD (PDB code 1FMA chainD). (*A* and *B*) Solvent-exposed residues (yellow balls) and nearby electrostatic residues (sticks, Arg in blue and Asp in red) were superimposed on the ribbon representation of Urm1 (*A*) and MoaD (*B*). (*C* and *D*) Electrostatic surface diagrams of Urm1 (*C*) and MoaD (*D*). The surface color reflects the magnitude of the electrostatic potential: red, negative; blue, positive; white, neutral. All of the surfaces were observed from the same orientation.

C-terminal region residues. The MoaD-related fold is unique in having two additional α-helixes (Fig. 4, yellow circles) packed together near the C-terminal tail.

It is notable that, in contrast to other protein modifiers, which are classified in the ubiquitin-related fold family, the structures of Urm1 and *M. musculus* AAH26994.1 protein (PDB code 1XO3) are classified in the MoaD-related fold family because Urm1 has two additional α-helixes like MoaD does. Furthermore, Urm1 also contains the SSE characteristic of ThiS: The short helix (Fig. 4, purple circle) follows the β4-strand directly. The characteristic is found only in TtMoaD but none of the other MoaD orthologs, indicating that Urm1 is structurally very similar to the sulfur carrier proteins. High structural similarity combined with a similar ATP-dependent activating mechanism strongly confirms the homology between the two sulfur carrier proteins and Urm1. Furthermore, on the basis of parsimony, the best explanation for the existence of the two significant SSE features in the Urm1 protein structure (one exists in the ThiS fold and the other in the MoaD fold) is that Urm1 has conserved both structural features of the common ancestor, the ThiS proteins have lost one SSE, whereas the MoaD proteins have lost another. The phenomena also imply that the structural features of Urm1 may be the most conserved of the common ancestor of this superfamily.

Phylogenetic Analysis of the Ubiquitin Superfamily. A maximum likelihood (ML) tree showing the phylogenetic relationships within the ubiquitin superfamily was generated based on a sequence data set (Fig. 5). It is believed that the protein modifiers (Urm1 family and ubiquitin-related families) evolved from an ancestral sulfur carrier protein resembling MoaD and ThiS (5). However, the sequence similarity between the two families, MoaD and ThiS, and the Urm1 family is higher than that between these two families and the ubiquitin-related families. To further compare the sequence
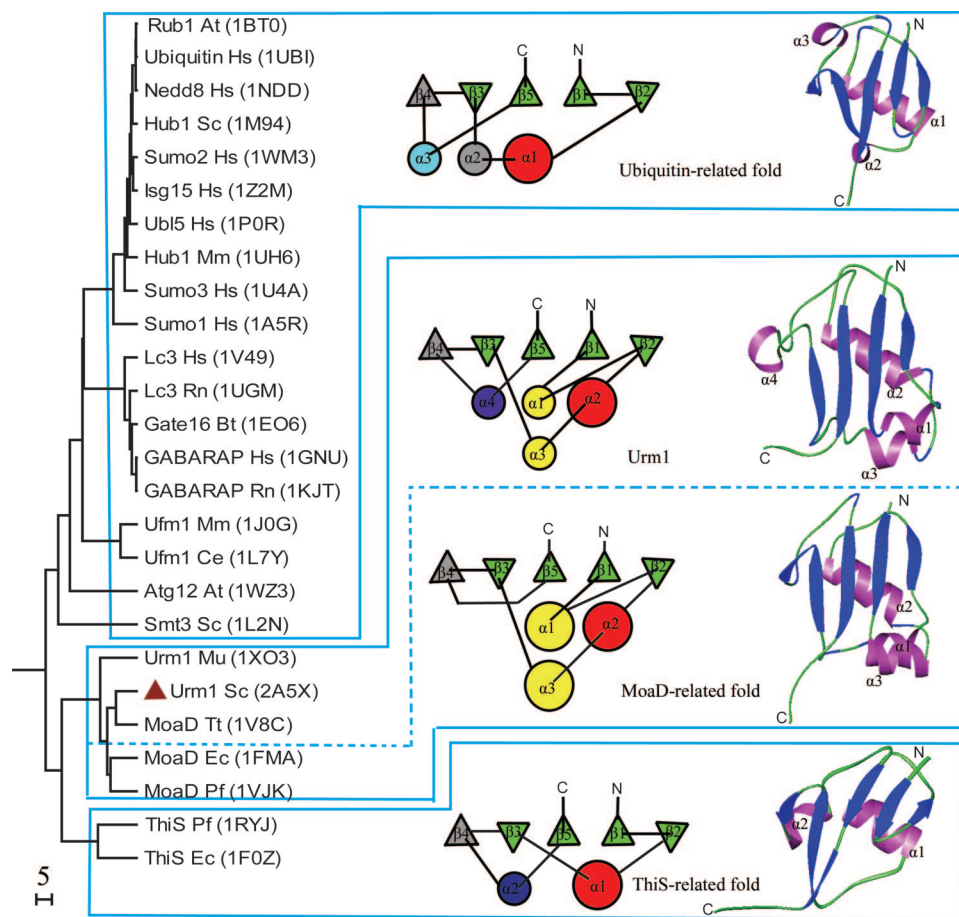
Fig. 4. Structural classification of the ubiquitin superfamily. Ubiquitin homologs were classified into three clusters based on their structural similarities. The diagrams represent the topological structure of each cluster (triangles and circles denote β-strands and helixes, respectively). The corresponding SSEs in spatial structures are colored the same (the β4-strand does not fold into a typical β-strand in some structures, so we use a gray triangle to represent it). Ribbon diagrams of ubiquitin (1UBI), Urm1 (2AX5), MoaD (1FMA chainD), and ThiS (1F0Z) are located on the right.

similarities between four major clades in the ML tree, we partitioned the sequences into segments according to the structural alignment of the superfamily and then mapped the consensus segments onto corresponding clades (Fig. 5). In seven of eight segments, the highest similarity is shown to be between Urm1 and MoaD or ThiS (for the distance matrix, see *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site). Considering the wide distribution of Urm1 and ubiquitin in eukaryotes, which means that Urm1 and the sulfur carrier proteins must have diverged from ubiquitin at a very early time, this result indicates that Urm1 may share a high sequence similarity to its sulfur carrier ancestor, whereas the ubiquitin-related families may have undergone a significant evolutionary change in function. However, it is generally accepted that ubiquitin is well conserved in eukaryotes (5), so this change may have occurred at an early stage in the evolutionary history of the superfamily.

Interestingly, some segments of the Urm1 sequence are closer to ThiS, and others are closer to MoaD (Fig. 5): Segments β1 and β2 are more similar between Urm1 and MoaD than either is to ThiS, whereas segments β3–β4 and β5 are more similar between Urm1 and ThiS than either is to MoaD. This observation is consistent with the results in the structural comparison (Fig. 4). It is unlikely that MoaD and ThiS independently evolved a different part of their sequences that still share both sequence and structural similarity with Urm1 protein; therefore, the sequence similarity analysis provides further evidence that Urm1 may have the characters of the ancestral proteins that are most conserved during the evolution of

the ubiquitin superfamily. The sequence similarity analysis also implies that a different selection force must have acted on the sulfur carrier protein families.

The functional divergence analysis provided further confirmation of our previous results. The functional branch length of the Urm1 cluster is 0.121, and the sulfur carrier cluster has a functional branch length of 0.391, compared with the large branch length of 1.509 of the ubiquitin-like proteins. The branch length difference indicates that the ubiquitin-like proteins might have a large fraction of altered functional constraints in their sequence, whereas the Urm1 proteins and sulfur carriers are relatively conservative. The Urm1 family is still the most conservative cluster in the superfamily, which is consistent with our previous analysis.

To understand the structure of the ancestral protein of the ThiS and MoaD family, we reconstructed the sequence and simulated the structure of this putative ancestral protein (Fig. 7, which is published as supporting information on the PNAS web site). There is a highly significant similarity between the reconstructed ancestral structure and the solution NMR structure of Urm1, giving a clue to the evolution of the ubiquitin superfamily. The results obtained from phylogenetic analysis of the ubiquitin superfamily add a critical piece of evidence required to support previous results in the structural comparison; i.e., Urm1 has most conserved the structural features of their common ancestor and can be considered a molecular fossil in the superfamily. An alternative explanation is of an entirely different evolutionary pattern in which two sulfur carrier proteins recombined to

Fig. 5. Phylogenetic relationship of the ubiquitin superfamily and sequence similarity analysis. The unrooted ML tree shows that there are four major clades in the ubiquitin superfamily. Sequences were accordingly clustered and mapped onto the tree. The segments represent the corresponding sequences of SSEs in the structures. Absence of SSEs was shown in broken-line boxes. Colors of the SSEs correspond to those of Urm1 used in Fig. 4. The depth of the color indicates the average similarity of the sequences with the Urm1 family. Ag, *Anopheles gambiae*; Ago, *Ashbya gossypii*; An, *Aspergillus nidulans*; At, *Arabidopsis thaliana*; Bt, *Bos taurus*; Ce, *Caenorhabditis elegans*; Cg, *Cricetulus griseus*; Dd, *Dictyostelium discoideum*; Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Ec, *E. coli*; Gg, *Gallus gallus*; Hs, *Homo sapiens*; Il, *Idiomarina loihiensis*; Mm, *M. musculus*; Nc, *Neurospora crassa*; Pfu, *Pyrococcus furiosus*; Rn, *Rattus norvegicus*; Sc, *S. cerevisiae*; Sp, *Schizosaccharomyces pombe*; Tb, *Trypanosoma brucei*; Tt, *T. thermophilus*.

generate a new protein that functioned as a protein modifier. Obviously, this hypothesis needs further investigation.

Conclusion

The present study reports the solution structure of Urm1 in *S. cerevisiae* and indicates its implications for better understanding the evolutionary history of the ubiquitin superfamily. Combined with the previous evidence of functional similarity, here we show the importance of Urm1 protein in this evolutionary scenario by means of structural comparison and sequence analysis. Among all known members of the ubiquitin superfamily, Urm1 seems very likely to be the unique molecular fossil that has most conserved the structural and sequence features of the common ancestor of the entire superfamily. The similarities of 3D structure and hydrophobic and electrostatic surface features between Urm1 and MoaD suggest that they may interact with partners in a similar manner, yet similarities between Urm1–Uba4 and MoaD–MoeB also establish an evolutionary link between ATP-dependent protein conjugation in eukaryotes and ATP-dependent cofactor sulfuration. For these reasons, we believe that the solved Urm1 protein structure, which is considered to be a critical piece of evidence for inferring the evolutionary origin of the ubiquitin superfamily, would also be extremely informative for further investigation of the complex function and mechanism of the modification pathway during the evolution of protein modifiers.

Materials and Methods

Sample Preparation. The gene encoding wild-type *S. cerevisiae* Urm1 was cloned into the NdeI/XhoI-cleaved plasmid PET22b(+) (Novagen), providing the C-terminal His-tagged (LEHHHHHH) protein. The sequence was confirmed by DNA sequencing

(Takara). The recombinant Urm1 was expressed by using *E. coli* BL21 (DE3). The culture was fermented at 37°C to A_{600} of 1.0 and then induced with 1.0 mM isopropyl β-D-thiogalactoside for 4 h. Cells were harvested and suspended in 50 mM Tris·HCl, pH 7.8/500 mM NaCl. After sonication centrifugation at $100,000 \times g$, the supernatant of lysed cells was collected and purified with a Ni-chelating column (Qiagen). The yield of Urm1 was typically 12–15 mg per liter of culture. Uniformly ^{15}N,^{13}C-labeled Urm1 was prepared with medium containing 0.5 g/liter 99% ammonium chloride and 2.5 g/liter 99% ^{13}C-glucose as the sole nitrogen and carbon source, respectively. NMR samples contained 0.8 mM Urm1, 50 mM phosphate buffer (pH 6.5), 50 mM sodium chloride, 1 mM NaN$_3$ in either 90% H$_2$O/10% ^{2}H$_2$O or 100% ^{2}H$_2$O.

NMR Experiments and Structure Calculations. All NMR data were collected at 298 K on a Bruker DMX600 spectrometer. A set of standard triple-resonance spectra was recorded for backbone and side chain assignments. NOE distance restraints were obtained from 3D ^{15}N- and ^{13}C-edited NOESY spectra acquired with a mixing time of 130 ms. After all of the above experiments, the sample was lyophilized and redissolved in 99.96% ^{2}H$_2$O and followed immediately with a series of ^{15}N-HSQC experiments to monitor the disappearance of NH signals to obtain the hydrogen bond information. NMR data were processed with NMRPipe and analyzed with Sparky 3 software. The interproton restraints were classified into four categories: 1.8–3.0 Å, 1.8–4.0 Å, 1.8–5.0 Å, and 1.8–6.0 Å according to NOE intensities. Torsion angle restraints, φ and ψ, for SSEs were mainly generated from the analysis of C$^{\alpha}$, C$^{\beta}$, C′, and H$^{\alpha}$ chemical shifts by using the chemical shift index (20); a few of them were generated from SSEs defined by characteristic NOEs. Hydrogen bond restraints were obtained by assignment of

slow-exchange amide protons located in regular SSEs. Initially, only typical medium- and long-range NOEs defining secondary structures from ^{13}C-edited NOESY spectrum were used together with NOEs obtained from ^{15}N-edited NOESY for structure calculation. Dihedral angle restraints, hydrogen bonds, and all other NOEs were introduced in consecutive steps. The CNS v1.1 program (21) was used to calculate and refine the structures. The statistics on experimental constraints, coordinate precision, and stereochemical quality of the 20 structures with the lowest energy were analyzed with MOLMOL (22) and PROCHECK (23). Analysis of the Ramachandran plot showed that 95.5% of residues were in allowed regions with 74% in the most favored region. The 0.5% of residues in disallowed regions are mainly in the C-terminal tail region.

Structural Comparison. All accessible structural coordinates in the ubiquitin superfamily were extracted from the PDB. Except for Fat10 and UBIM families, all of the families belonging to the ubiquitin superfamily have the representations whose structures had been resolved. A total of 26 structures were used in structural cluster analysis. The structure distance matrix, which indicates the structural similarity (24–26), was calculated with the generated rmsd, and the percentage of structural identity values were calculated by Mammothmult (27). A dendrogram was then constructed based on the structure distance matrix and the unweighted pair-group method using an arithmetic average (Unweighted Pair Group Method with Arithmetic Mean). The diagrams representing topology structures of each cluster (Fig. 4, triangles and circles in the diagrams denote β-strands and helixes, respectively) were produced with Tops (28).

Phylogenetic Analysis. A total of 55 representative sequences from all identified protein modifier families, as well as the ThiS and MoaD families, were obtained from GenBank (accession numbers are listed in Table 2, which is published as supporting information on the PNAS web site). Amino acid sequences within the ubiquitin domain were used for phylogenetic analysis. For structure-resolved proteins, the ubiquitin domains were defined according to their 3D structure. Domains in other sequences were predicted by the DomPred server (29). For the diubiquitin domain proteins, Isg15 and Fat10, the C-terminal ubiquitin domain was used for sequence analysis and structural comparison. The amino acid sequences of the superfamily were aligned by using ClustalX 1.83 (30). Because of the low sequence similarity across the superfamily, the gap separation distance was adjusted to 2.00 for a better alignment (see *Supporting Materials and Methods*), and we manually corrected some sites according to structural information. The phylogenetic tree of the ubiquitin superfamily was constructed with the ML method implemented in PHYML (31), using the model of Jones–Taylor–Thornton plus the four categories of Gamma substitution rates plus Invariable sites. The nonparametric bootstrap test was performed for 100 replicates.

To further understand how the sequence similarity corresponds to the structure in the ubiquitin superfamily, we partitioned the sequences into eight SSEs according to the structural alignment generated by Mammothmult (Fig. 8, which is published as supporting information on the PNAS web site) and then compared the sequence similarity in SSEs between clusters by using MEGA3 (32) under the Jones–Taylor–Thornton model. Considering the short length of the SSEs, uniform rate across sites was assumed in the comparison.

Functional Divergence Analysis. We analyzed the functional divergence of the protein families by using the software DIVERGE (33). We adopted the ML superfamily tree in the previous analysis, defined three clusters [sulfur carriers (ThiS and MoaD), Urm1, and Ubiquitin-like proteins (all other families)], then calculated the functional divergence branch length of each cluster by using the least squares estimation (34, 35).

Reconstruction of Ancestral Protein Sequence. The ancestral protein sequence of the ThiS and MoaD families was reconstructed with the Codeml program in PAML3.14 (36). The user tree provided (Fig. 7*A*) was generated with PHYML by using the same settings and parameters as before, i.e., the Jones–Taylor–Thornton plus gamma substitution rate variation over sites under the hypothesis without molecular clock.

To illustrate the ancestral protein, its reconstructed amino acid sequence of was modeled by submitting the sequence to the 3D-JIGSAW web server (37). In interactive steps, the *T. thermophilus* MoaD (PDB code 1V8C_A) was selected as the modeling template because it shares 29% identity with the reconstructed ancestor sequence over 87 aligned residues.

1. Hershko, A. & Ciechanover, A. (1998) *Annu. Rev. Biochem.* **67,** 425–479.
2. Chen, Z. J., Parent, L. & Maniatis, T. (1996) *Cell* **84,** 853–862.
3. Schwartz, D. C. & Hochstrasser, M. (2003) *Trends Biochem. Sci.* **28,** 321–328.
4. Haas, A. L. & Siepmann, T. J. (1997) *FASEB J.* **11,** 1257–1268.
5. Hochstrasser, M. (2000) *Nat. Cell Biol.* **2,** E153–7.
6. Taylor, S. V., Kelleher, N. L., Kinsland, C., Chiu, H. J., Costello, C. A., Backstrom, A. D., McLafferty, F. W. & Begley, T. P. (1998) *J. Biol. Chem.* **273,** 16555–16560.
7. Rajagopalan, K. V. (1997) *Biochem. Soc. Trans.* **25,** 757–761.
8. Wang, C., Xi, J., Begley, T. P. & Nicholson, L. K. (2001) *Nat. Struct. Biol.* **8,** 47–51.
9. Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. (2001) *Nature* **414,** 325–329.
10. Furukawa, K., Mizushima, N., Noda, T. & Ohsumi, Y. (2000) *J. Biol. Chem.* **275,** 7462–7465.
11. Singh, S., Tonelli, M., Tyler, R. C., Bahrami, A., Lee, M. S. & Markley, J. L. (2005) *Protein Sci.* **14,** 2095–2102.
12. Goehring, A. S., Rivers, D. M. & Sprague, G. F., Jr. (2003) *Eukaryot. Cell* **2,** 930–936.
13. Goehring, A. S., Rivers, D. M. & Sprague, G. F., Jr. (2003) *Mol. Biol. Cell* **14,** 4329–4341.
14. Holm, L. & Sander, C. (1993) *J. Mol. Biol.* **233,** 123–128.
15. Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. (2001) *Nat. Struct. Biol.* **8,** 42–46.
16. Walden, H., Podgorski, M. S., Huang, D. T., Miller, D. W., Howard, R. J., Minor, D. L., Jr., Holton, J. M. & Schulman, B. A. (2003) *Mol. Cell* **12,** 1427–1437.
17. Whitby, F. G., Xia, G., Pickart, C. M. & Hill, C. P. (1998) *J. Biol. Chem.* **273,** 34983–34991.
18. Ding, H., Xu, Y., Chen, Q., Dai, H., Tang, Y., Wu, J. & Shi, Y. (2005) *Biochemistry* **44,** 2790–2799.
19. Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R. & Becker, J. (1998) *J. Mol. Biol.* **280,** 275–286.
20. Wishart, D. S. & Sykes, B. D. (1994) *J. Biomol. NMR* **4,** 171–180.
21. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
22. Koradi, R., Billeter, M. & Wuthrich, K. (1996) *J. Mol. Graphics.* **14,** 29–32, 51–55.
23. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996) *J. Biomol. NMR* **8,** 477–486.
24. Balaji, S. & Srinivasan, N. (2001) *Protein Eng.* **14,** 219–226.
25. Johnson, M. S., Sali, A. & Blundell, T. L. (1990) *Methods Enzymol.* **183,** 670–690.
26. Johnson, M. S., Sutcliffe, M. J. & Blundell, T. L. (1990b) *J. Mol. Evol.* **30,** 43–59.
27. Ortiz, A. R., Strauss, C. E. & Olmea, O. (2002) *Protein Sci.* **11,** 2606–2621.
28. Westhead, D. R., Hatton, D. C. & Thornton, J. M. (1998) *Trends Biochem. Sci.* **23,** 35–36.
29. Marsden, R. L., McGuffin, L. J. & Jones, D. T. (2002) *Protein Sci.* **11,** 2814–2824.
30. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
31. Guindon, S. & Gascuel, O. (2003) *Syst. Biol.* **52,** 696–704.
32. Kumar, S., Tamura, K. & Nei, M. (2004) *Brief Bioinform.* **5,** 150–163.
33. Gu, X. & Vander Velden, K. (2002) *Bioinformatics* **18,** 500–501.
34. Wang, Y. & Gu, X. (2001) *Genetics* **158,** 1311–1320.
35. Gu, X. (1999) *Mol. Biol. Evol.* **16,** 1664–1674.
36. Yang, Z. (1997) *Comput. Appl. Biosci.* **13,** 555–556.
37. Bates, P. A., Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. E. (2001) *Proteins Struct. Funct. Genet. Suppl.* **5,** 39–46.