Molecular Evolution of Keap1: Two Keap1 Molecules with Distinctive IVR Structures are Conserved among Fish

Li Li[‡,§], Makoto Kobayashi[‡,§,1], Hiroshi Kaneko[§], Yaeko Nakajima-Takagi[§], Yuko Nakayama[‡] and Masayuki Yamamoto[‡,§,¶]

From the [‡]Environmental Response Project, Japan Science and Technology Agency and the [§]Institute of Basic Medical Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Japan, and the [¶]Department of Medical Biochemistry, Tohoku University Graduate School of Medicine, 2-1 Seiryo-cho, Aoba-ku, Sendai 980-8575, Japan

Running Head: Molecular evolution of Keap1

Address correspondence to: Makoto Kobayashi, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8575, Japan. Tel: +81-29-853-8457. Fax: +81-29-853-5977. E-mail: makobayash@md.tsukuba.ac.jp

Keap1 is a BTB-Kelch type substrate adaptor protein of the Cul3-dependent ubiquitin ligase complex. Keap1 facilitates the degradation of Nrf2, a transcription factor regulating the inducible expression of many cytoprotective genes. Through comparative genome analyses, we found that amino acid residues comprising the pocket of Keap1 that interacts with Nrf2 are highly conserved among Keap1 orthologs and related proteins in all vertebrates and in certain invertebrates, including flies and mosquitoes. The interaction between Nrf2 and Keap1 appears to be widely preserved in vertebrates. Similarly, cysteine residues corresponding to Cys273 and Cys288 in the intervening region of mouse Keap1, which are essential for the repression of Nrf2 activity in cultured cells, are conserved among Keap1 orthologs in vertebrates and invertebrates, except fish. We found that fish have two types of Keap1, Keap1a and Keap1b. To our surprise, Keap1a and Keap1b contain the cysteine residue corresponding to Cys288 and Cys273, respectively. In our analysis of zebrafish Keap1a and Keap1b activities, both Keap1a and Keap1b were able to facilitate the degradation of Nrf2 protein and repress Nrf2-mediated target gene activation. Individual mutation of either residual cysteine residue in Keap1a and Keap1b disrupted the ability of Keap1 to repress Nrf2, indicating that the presence of either Cys273 or Cys288 is sufficient for fish Keap1 molecules to fully function. These results provide an important insight into the means by which Keap1 cysteines act as sensors of electrophiles and oxidants.

The transcription factor Nrf2 induces the expression of phase 2 detoxifying and antioxidant proteins in response to electrophilic insults (1). These induced proteins contribute to the prevention of oxidative damage and chemically-induced cancer in animals. The importance of Nrf2 in this induction and resulting chemoprevention has been demonstrated by a number of experiments using Nrf2-deficient mice (2). The electrophile response is regulated through a *cis*-acting element called the antioxidant responsive element (ARE) or electrophile responsive element (EpRE) within the regulatory region of each gene (3). Nrf2 binds to the ARE/EpRE sequence as a heterodimeric complex with small Maf proteins through a basic region-leucine zipper domain (4). Under normal homeostatic conditions, Nrf2 protein is targeted for proteasomal degradation and has a short half-life. This degradation is positively controlled by Keap1, a member of the BTB-Kelch protein family (5,6). Keap1 binds to Nrf2 and promotes its degradation as a substrate-specific adaptor protein for the Cul3 ubiquitin ligase complex (7). When oxidative/electrophilic stress signals disrupt the Nrf2-Keap1-Cul3 complex, ubiquitination of Nrf2 is blocked and Nrf2 becomes stable (8). Consequently, the expression of a battery of cytoprotective genes is induced as Nrf2 accumulates in the nucleus.

Keap1 is composed of three major domains: a Broad-Complex,□Tramtrack and Bric-a-brac (BTB) domain, a double glycine□repeat (DGR) domain and an intervening region (IVR) domain (1). The BTB domain functions to dimerize Keap1 (9), while the DGR domain serves as a binding site for Nrf2 (5) and actin (10). Our group and that of Hannink have determined the crystal structure of the Keap1 DGR domain and identified its interface with Nrf2 (11,12). Involvement of the Keap1 IVR domain in the

ubiquitination of Nrf2 has been demonstrated (8,13). In cultured cells, mutation of Cys273 or Cys288 in the IVR domain to alanine or serine reduced Keap1-dependent ubiquitination and increased Nrf2 stability, suggesting that these residues are crucial for the Nrf2-repressing activity of Keap1 (13-15).

We previously isolated homolog genes of Nrf2 and Keap1 in zebrafish and established that the Nrf2-dependent induction of cytoprotective genes is conserved among vertebrates (16,17). We thus speculated that the Nrf2-Keap1 system of cytoprotection is also conserved in vertebrates. To our surprise, zebrafish Keap1 protein does not contain a cysteine residue corresponding to Cys273 in mouse Keap1, yet still repressed the activity of Nrf2 in zebrafish embryos (16). In this paper, we compared the amino acid sequences of the Keap1-related proteins of various vertebrates and invertebrates by comparative genome analysis. Critical amino acids in the Nrf2-interacting surface of the DGR domain are highly conserved among these proteins, but are completely different in other mouse BTB-Kelch proteins. This indicates that Keap1 is the only BTB-Kelch protein that regulates Nrf2 activity and also implies the presence of the Nrf2-Keap1 system in invertebrates. Interestingly, fish have two Keap1 genes, which we refer to as Keap1a and Keap1b. Keap1a has a cysteine residue corresponding to Cys273, but not to Cys288, in mouse Keap1, while the case is vice versa for Keap1b. We analyzed the activities of zebrafish Keap1a and Keap1b using zebrafish embryos and demonstrated that either protein can promote Nrf2 degradation; both Cys273 and Cys288 are important for Keap1 activity, but either one is enough in fish.

EXPERIMENTAL PROCEDURES

Isolation of cDNA — A partial cDNA fragment encoding zebrafish Keap1b was prepared by PCR using specific primers designed based on genomic DNA information. A λZAP-II 15-19 hour stage cDNA library (18) was screened to isolate a full-length Keap1b cDNA clone using the partial cDNA clone as a probe. The probe was labeled using an AlkPhos Direct DNA labeling kit and positive plaques on the membrane filters were detected with CDP-Star as substrate according to the manufacturer's instructions (GE Healthcare).

Radiation hybrid mapping — Radiation hybrid mapping using panel LN54 was performed as described in Hukriede *et al.* (19) using specific primers for each Keap1 gene. The sequences of each primer were as follows: *keap1a*, 5'-AGGATTTCTCCGCCATTGTG and 5'-CCTTGAAGTTGCTGGTGAAC; *keap1b*, 5'-ATGACGGAGTGTAAGGCGG and 5'-CAGGCCGTTGGTGAACATG.

Plasmid construction — The plasmid pCS2keap1b was constructed by subcloning the open reading frame of zebrafish *keap1b* into the *Bam*HI and *Xba*I sites of the vector pCS2. To construct pSPkeap1aC, cDNA encoding the C-terminus region (amino acids 353-601) containing the 3'-UTR of zebrafish *keap1a* was inserted into the *Not*I and *Sal*I sites of the vector pSPORT1. The plasmid pKSkeap1bN was generated by inserting cDNA encoding the N-terminal region (amino acids 8-188) of *keap1b* into the *Bam*HI and *Xho*I sites of pBluescript II KS. To construct pCS2nrf2NTnGFP, cDNA encoding the N-terminus region (amino acids 1-305) of zebrafish *nrf2* plus two repeats of SV40 nuclear localizing signal (DPKKKRKV) were subcloned into the *Bam*HI site of pCS2eGFP. The cDNA fragments for 3 x FLAG-tag (MDYKDHDGDYKDHDIDYKDDDDK) and 3 x hemagglutinin (HA)-tag (MEYPYDVPDYAAEYPYDVPDYAAEYPYDVPD YAAKLE) were subcloned into the *Bam*HI and *Eco*RI sites of pCS2 to generate pCS2FL and pCS2HA, respectively. The plasmids pCS2FLkeap1a, pCS2FLkeap1b and pCS2FLnrf2 were constructed by inserting the open reading frames of *keap1a*, *keap1b* and *nrf2*, respectively, into the *Hin*dIII and *Xba*I sites of pCS2FL. pCS2HAkeap1a and pCS2HAkeap1b were prepared by inserting the open reading frames of *keap1a* and *keap1b*, respectively, into the *Hin*dIII and *Xba*I sites of pCS2HA. The constructs pCS2FLkeap1aC264S and pCS2FLkeap1bC247S were made by introducing Cys to Ser point-mutations by PCR into pCS2FLkeap1a and pCS2FLkeap1b, respectively. pKSgstp1N was constructed by subcloning the cDNA for the N-terminal region (amino acids 1-135) of *gstp1* into the *Bam*HI and *Sal*I sites of pBluescript II KS. All constructs were verified by DNA sequencing. Plasmids pCS2nrf2, pCS2keap1a (previously named pCS2Keap1) and pCS2eGFP were described previously (17,20).

Expression analysis − Zebrafish embryos and larvae were obtained by natural mating. All experiments were carried out using a wild-type AB strain. The expression of *keap1a*, *keap1b* and *gstp1* genes was

analyzed by reverse transcriptase-PCR (RT-PCR) and whole mount *in situ* hybridization. For RT-PCR analysis, total RNA was prepared from adult tissues or the whole bodies of embryos and larvae using QIAzol (Qiagen). First-strand cDNA was synthesized by incubation at 25˚C for 15 min and at 42˚C for 45 min with murine leukemia virus reverse transcriptase (SuperScript II, Invitrogen) and random hexamer oligonucleotide primers. From the 20 μl first strand reaction, 0.025-0.1 was used for PCR using the following primers: *keap1a*, 5'-ATGATATGTCCAAGAAAGAAG and 5'-TCATGAGGAAATCGCAGCAG; *keap1b*, 5'-ACGGAGTGTAAGGCGGAG and 5'-ACCTGGCTGAAGTTCATG; *gstp1*, 5'-CTAGGAGCAGCTTTGAAACGCAC and 5'-TGGCCAGAACATTTTCAAAGC. *ef1α*, 5'-GCCCCTGCCAATGTA and 5'-GGGCTTGCCAGGGAC. The expression of *ef1α* was used to standardize the amount of cDNA. Real-time RT-PCR was performed to quantitate *gstp1* expression using an ABI Prism 7700 (Applied Biosystems) and probes labeled with a reporter fluorescent dye (TaqMan probe) as previously described (21). TaqMan probes, primers and cDNAs were added to the master mix containing the reagents for PCR (Eurogentec). The sequences of the specific primers and probes were as follows: *gstp1*, 5'-CAACGCCATGCTGAGACATC (sense), 5'-GAAGATCTTCAACGCCGTCG (antisense) and 5'-FAM-AACATGCTGCATATGGCAAAAACGAC AGT-TAMRA (probe); *ef1α*, 5'-CGTGGTAATGTGGCTGGAGA (sense), 5'-CTGAGCGTTGAAGTTGGCAG (antisense) and 5'-FAM-AGCAAGAACGACCCACCCATGGAG-T AMRA (probe). Whole mount *in situ* hybridization was performed as described previously (22) using RNA probes transcribed from pSPkeap1a, pKSkeap1b and pKSgstp1N.

Microinjection of zebrafish embryos — Synthetic capped RNA was made with an SP6 mMESSAGE mMACHINE *in vitro* transcription kit (Ambion) using linearized DNA of the pCS2 derivatives described above. For expression in whole bodies, RNA was injected into yolk at the one-cell stage using an IM300 microinjector (Narishige). GFP expression was examined under the GFP Plus (480 nm excitation, 505 nm emission) filter of a MZFLIII microscope (Leica) equipped with a 600CL-CU digital camera (Pixera).

In vitro translation and co-immunoprecipitation — HA- and FLAG-tagged Keap1 proteins were *in vitro*-translated separately by TNT Coupled Wheat Germ Extract Systems (Promega) using pCS2 derivatives as DNA templates. *In vitro*-translated Keap1 proteins were mixed in binding buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl and 0.1% NP40) and incubated with an affinity matrix immobilized anti-HA antibody (Roche Diagnostics, 3F10) at 4˚C for 4 hours with gentle mixing on a rotator. The beads were collected by centrifugation at 12,500 x *g* for 5 seconds and washed three times in binding buffer. Precipitated proteins were eluted in SDS-sample buffer and resolved by 12% SDS-PAGE, followed by immunoblotting using anti-HA (Roche Diagnostics, 12CA5) and anti-FLAG antibodies (Sigma, M2, peroxidase conjugate) as described previously (22).

RESULTS

Identification of the second Keap1 in zebrafish — By virtue of recent progress in the zebrafish genome project, we came across a novel Keap1-related gene that shows a higher similarity to mammalian Keap1 than previously reported zebrafish Keap1 (16). A partial cDNA was isolated by RT-PCR using specific primers whose design was based on genomic DNA information. We screened a zebrafish cDNA λ-phage library using this partial cDNA as a probe and isolated a full-length cDNA clone.

We referred to this gene as *keap1b* and the previous *keap1* was renamed *keap1a*. The deduced amino acid sequence of the Keap1b cDNA product showed 81% and 78% identity to the BTB and DGR domains, respectively, of mouse Keap1 protein (Fig. 1*A*). These values are quite high compared with Keap1a whose identities to the BTB and DGR domains are only 49% and 55%, respectively. We mapped both Keap1 genes using an LN54 hybrid panel (19) and found that *keap1a* and *keap1b* are localized on zebrafish chromosomes 2 and 6, respectively. The latest information from the zebrafish genome project supported these mapped sites and further demonstrated that synteny was found between *keap1b* and the human *KEAP1* locus on chromosome 19p13.2 (Supplemental Fig. 1).

Neh2 is the domain in Nrf2 that interacts with the DGR domain in Keap1 (5). Within the Neh2 domain, we found that the motifs ETGE and DLG are critical for the interaction with Keap1 (16,23). Recently, we identified the region of the Keap1 DGR domain responsible for binding to the ETGE and DLG motifs

by structural analysis of the mouse Keap1 protein (11,24). The amino acid residues important for binding to the ETGE motif have been recognized as Ser363, Arg380, Asn382, Arg415, Arg483, Ser508, Tyr525, Gln530, Ser555 and Ser602. Those important for binding to the DLG motif are Asn382, Arg415, Arg483, Ser508, Ser555, Tyr572, Phe577, Ser602 and Gly603 (Fig. 1*B*, *white characters highlighted in black*). Mutation analyses of mouse and human Keap1 proteins have demonstrated that Tyr334, Gly364, Gly430, His436 and Phe478 (Fig. 1*B*, *white characters highlighted in grey*), in addition to Arg380, Asn382, Arg415, Arg483, Tyr525 and Tyr572, are critical for inhibiting Nrf2 activity (11,12). Interestingly, all these residues, except Asn382 and Tyr572, are conserved in both zebrafish Keap1a and Keap1b, suggesting that both proteins can interact with Nrf2. Indeed, zebrafish Keap1a has been shown to interact with Nrf2 and to inhibit its activity (16). Although Mayven is a protein with the highest homology to Keap1 in the DGR domain among mouse BTB-Kelch proteins (25), it possesses only two of the 13 critical Nrf2-interacting residues in mouse Keap1 (Fig. 1*B*, *mM*). This case is similar to KLHL20 and KLHL5, two other Keap1-related proteins (Supplemental Table 1). These results suggest that the activity of Nrf2 is regulated by two Keap1 proteins, Keap1a and Keap1b, in zebrafish and by a single Keap1 protein in mouse, which may be the only BTB-Kelch protein that can facilitate Nrf2 degradation. Here, we propose to define Keap1 as a BTB-Kelch protein carrying the evolutionarily conserved Nrf2-interacting surface.

Unlike Keap1, we could not find a second Nrf2 gene in the zebrafish genome database. Nrf2 is a member of the CNC protein family whose members are NF-E2 p45, Nrf1, Nrf2, Nrf3, Bach1 and Bach2 (1). Among them, genetic loci of mammalian NF-E2 p45, Nrf1, Nrf2 and Nrf3 genes have been mapped close to those of HoxC, HoxB, HoxD and HoxA, respectively (34). Interestingly, the zebrafish genome has two copies of HoxA, HoxB and HoxC clusters but only one HoxD cluster (35). We assume that the second Nrf2 gene in zebrafish had been lost together with the second HoxD cluster during evolution.

Keap1 is present in vertebrates and in some invertebrates — To identify the range of species in which Keap1 is present, we searched the Ensemble (http://www.ensembl.org/index.html) and DDBJ/EMBL/GenBank (http://blast.ddbj.nig.ac.jp/top-e.html) databases for Keap1-related proteins. As well as in mammals, Keap1 genes were found in chicken, frogs (*Xenopus laevis*, *Xenopus tropicalis*), fugu, *Tetraodon*, medaka fish, stickleback, ascidians (*Ciona intestinalis*, *Ciona savignyi*), mosquitoes (*Aedes aegypti*, *Anopheles gambiae*) and *Drosophila.* A phylogenetic tree based on the amino acid sequences of their DGR domains classified the Keap1 proteins into 5 subgroups: (1) vertebrate Keap1; (2) fish Keap1a; (3) fish Keap1b; (4) ascidian Keap1; (5) invertebrate Keap1 (Fig. 1*C*). No Keap1-related genes were found in nematode or yeast. We noted that all these Keap1 proteins carry 13 critical Nrf2-interacting residues, with the exceptions of Asn382 and Tyr572 for fish Keap1a and Tyr525 for invertebrate Keap1 (Supplemental Table 1). The results suggest that Keap1 regulates Nrf2 or related proteins in these organisms in a manner similar to that in mammals.

Keap1a and Keap1b are conserved among fish, but not in other vertebrates, signifying that both proteins are essential to the fish Nrf2-Keap1 system. Keap1b rather than Keap1a may represent the ortholog of vertebrate Keap1, since conserved synteny was observed between human *KEAP1* and fish Keap1b loci (Supplemental Fig. 1). No synteny was found between human Keap1 and fish Keap1a genes or with ascidian or invertebrate Keap1. This implies that Keap1b may be the proper homolog of vertebrate Keap1.

Keap1a and Keap1b repress Nrf2 activity despite their lack of a cysteine residue corresponding to mouse Keap1 Cys273 and Cys288, respectively — All fish Keap1a and Keap1b lack a cysteine residue corresponding to Cys273 and Cys288, respectively, while both these cysteines are conserved even in ascidian and invertebrate Keap1 proteins (Fig. 2). This finding was surprising since both Cys273 and Cys288 in the IVR were demonstrated to be crucial for the Nrf2-repressing activity of mouse Keap1 (13-15). To elucidate whether zebrafish Keap1a and Keap1b can repress the inducible function of Nrf2, we tested the extent of their repression on the Nrf2-mediated inducible expression of the endogenous *gstp1* gene in zebrafish embryos. The *gstp1* gene encodes a Pi-class glutathione *S*-transferase and was strongly induced in both electrophile-treated larvae and Nrf2-overexpressing embryos (16,26). Its promoter contains an evolutionarily conserved ARE/EpRE sequence that is critical for both Nrf2 binding and promoter activity (26). *In vitro* synthesized zebrafish Keap1a or Keap1b

mRNA (200 pg) was co-injected with Nrf2 mRNA (100 pg) into zebrafish embryos at the one-cell stage (Fig. 3*A*). At midgastrula, *gstp1* expression was analyzed by whole mount *in situ* hybridization analysis. Nrf2-induced expression of *gstp1* was reduced by co-overexpression of either Keap1a or Keap1b (Fig. 3*B*), indicating that both Keap1a and Keap1b possess the ability to repress Nrf2 activity. To confirm this, we used FLAG-tagged Keap1 proteins to standardize the protein expression level of each Keap1 by immunoblotting (Supplemental Fig. 2). Seventy-five pg of Keap1a mRNA and 200 pg of Keap1b mRNA expressed similar amounts of Keap1 proteins in zebrafish embryos. Only full-length proteins were overexpressed in embryos. The FLAG-tagged constructs were used to compare the Nrf2-repression activity of Keap1a and Keap1b by real-time RT-PCR analyses (Fig. 3*C*). Sixty pg of Nrf2 mRNA were co-injected with various amounts of Keap1a or Keap1b mRNAs (Fig. 3*C*). The dose effects of Keap1 mRNA on Nrf2 repression were similar between Keap1a and Keap1b, suggesting that the activities of Keap1a and Keap1b to repress Nrf2 activity are comparable, at least in zebrafish embryos.

Both Keap1 proteins promote Nrf2 degradation — Mouse Keap1 has been shown to promote the degradation of Nrf2 as a substrate-specific adaptor protein for the Cul3 ubiquitin ligase complex (7). To elucidate whether zebrafish Keap1 proteins also promote Nrf2 degradation, we examined the effects of Keap1 co-overexpression on the level of Nrf2 protein. FLAG-tagged Nrf2 protein overexpressed in zebrafish embryos by mRNA injection was detectable by whole-mount immunostaining using anti-FLAG antibody (Fig. 4*A*). This antibody staining disappeared when we co-overexpressed either Keap1a or Keap1b, advocating the promotion of Nrf2-degradation as the means by which these Keap1 proteins repress Nrf2. To confirm this, we overexpressed an Nrf2 GFP-fusion protein (Nrf2NTnGFP) in zebrafish embryos and tested its stability in the presence or absence of Keap1a or Keap1b by observing GFP expression (Fig. 4*B*). Note that the N-terminal domain of zebrafish Nrf2 was used to construct the GFP-fusion protein, since this region corresponds to the mouse Nrf2 protein that was shown to be sufficient for Keap1 dependent degradation in both cultured cells and mouse intestine (27). GFP expression was observed in Nrf2NTnGFP-overexpressing embryos, while GFP expression was dramatically lower when either Keap1a or Keap1b was co-overexpressed (Fig. 4*B*) (Nrf2NTnGFP, 53.7%, n=93; Nrf2NTnGFP +Keap1a, 0%, n=132; Nrf2NTnGFP +Keap1b, 0%, n=73; no injection, 0% n=100). These results demonstrated that both Keap1a and Keap1b repress Nrf2 activity by facilitating its degradation, as is the case for mouse Keap1.

Keap1a and Keap1b can form homodimers and heterodimers — We previously found that coexpression of Cys273Ala and Cys288Ala mutant proteins of mouse Keap1 substantially restored repressor activity, whereas each Keap1 mutant alone lacked repressor activity (15). This observation implies that Cys273Ala and Cys288Ala form a heterodimer and the simultaneous presence of Cys273 on one monomer and Cys288 on the other is sufficient for the repressor activity. Similarly, it is possible that overexpressed Keap1a or Keap1b in zebrafish embryos form a heterodimer with endogenous Keap1 proteins to share cysteine residues in the same complex. To assess this possibility, we carried out pull-down analysis using *in vitro* translated Keap1a and Keap1b proteins with FLAG- and HA-tags (Fig. 5). Tagged Keap1 proteins were mixed and pulled-down with anti-HA-beads. Precipitated proteins were analyzed by immunoblotting using anti-FLAG and anti-HA antibodies. FLAG-tagged Keap1a protein co-precipitated with both HA-tagged Keap1a and Keap1b. Similarly, FLAG-tagged Keap1b protein was pulled-down with HA-tagged Keap1a and Keap1b. These results demonstrate that Keap1a and Keap1b can form both homodimers and heterodimers.

Keap1a and Keap1b genes are coexpressed in many tissues — Keap1a and Keap1b require simultaneous expression to function as heterodimers. To provide insight into the roles of Keap1a and Keap1b *in vivo*, we examined the tissue distribution of Keap1 mRNA in adult fish (Fig. 6*A*). Total RNA fractions were prepared from various tissues of 10 month old zebrafish males and analyzed by RT-PCR. The amount of cDNA was standardized by the expression level of *ef1α*. While both *keap1a* and *keap1b* were expressed ubiquitously, the expression of *keap1b* was relatively abundant in brain and scarce in gut. We also examined the expression levels of the zebrafish Keap1 genes during the embryonic and larval stages (Fig. 6*B*). RT-PCR analyses demonstrated that *keap1b* was expressed at every stage tested and at similar levels, while *keap1a* expression was quite low

during the embryonic stages and started to increase around the time of hatching (2.5 days). Spatial expression profiles of zebrafish Keap1 genes were assessed at the embryonic stages by whole mount *in situ* hybridization (Fig. 6*C*). Both genes were expressed ubiquitously in the whole body, although some specific regions, such as lens (*arrow*), expressed *keap1a* more strongly than others. Overall, these observations suggest that *keap1a* and *keap1b* are coexpressed in many cells.

Cysteine residues corresponding to Cys273 and Cys288 in mouse Keap1 are important for the Nrf2-repressing activity of Keap1a and Keap1b — The critical cysteine residues in Keap1a and Keap1b must be important for repressing Nrf2 if these two proteins function as heterodimers. To verify this, point-mutations were introduced in these cysteines and the ability to repress Nrf2 was analyzed. In this paper, we refer to the cysteine residues in the IVR domain as ICs (IVR cysteines) to ease comparison among the corresponding cysteines of various Keap1 proteins (see Fig. 2). Cysteine residues corresponding to Cys273 and Cys288 in mouse Keap1 are called IC6 and IC7. We introduced Cys to Ser point mutations in IC7 of Keap1a and in IC6 of Keap1b and examined the effects of these mutations on Nrf2-repressing activity (Fig. 7*A*). We used FLAG-tagged Keap1 proteins to standardize the protein expression level of each Keap1 by immunoblotting. Mutations in Keap1a IC7 and Keap1b IC6 strongly abolished the Nrf2-repressing activity (Fig. 7*B*). IC7 in Keap1a and IC6 in Keap1b are thus essential for the repression of Nrf2 activity.

DISCUSSION

This is the first paper referring to the evolutionary aspects of Keap1, as well as to its definition. Stogios and Privé predicted that more than 53 members of BTB-Kelch proteins exist in human (28). Some of them, such as Mayven, KLHL20 and KLHL5, show relatively high similarity with Keap1. For example, Mayven has a DGR with the highest amino acid sequence identity (44%) to that of Keap1 among the mouse members of the BTB-Kelch family. This value is close to that between zebrafish Keap1a and mouse Keap1 (55%). However, mouse Mayven shares only 2 of the 13 critical amino acid residues of mouse Keap1, which were shown to form the interaction surface for Nrf2 (Supplemental Table 1). In contrast, zebrafish Keap1a shares 11 of them. This indicates that Mayven cannot bind to Nrf2 and is inactive in repressing the function of Nrf2. Indeed, Mayven was not able to repress Nrf2 activity in cultured cells, even when its BTB and IVR domains were swapped with those of mouse Keap1 (Kang, A. Kobayashi and Yamamoto, unpublished data). We anticipate that Keap1 is the only BTB-Kelch protein that regulates Nrf2 activity.

We recently proposed "the hinge and latch model" for the interaction between Nrf2 and Keap1 and the induction of cellular defense enzymes (24,29-31). Keap1 dimer recruits its substrate Nrf2 by binding to the evolutionarily conserved DLG and ETGE motifs within the Neh2 domain of Nrf2 (16,23). The structural plasticity of its Neh2 domain allows Nrf2 to link two Keap1 molecules in tandem on either side of the central Neh2 α-helix that exists between the DLG and ETGE motifs, thereby presenting the lysines for E3-catalyzed ubiquitination (29). These lysine residues were shown to be important for Nrf2 degradation (32). In this paper, we have shown that the domain interacting with both the DLG and ETGE motifs is highly conserved among various Keap1 proteins, even in invertebrate Keap1, suggesting that the hinge and latch system may also be conserved. It is plausible that the DLG and ETGE motifs are also conserved among vertebrate and some invertebrate species. Indeed, high conservation of these two motifs was observed by comparative genome analysis (Supplemental Fig. 3). Out of six CNC proteins, Nrf1 and Nrf2 possess the DLG and ETGE motifs. In ascidian, mosquito and fly, only one Nrf1/2-related protein exists that has both DLG and ETGE motifs. The QDxDLG and DxETGE sequences of the DLG and ETGE motifs, respectively, are the only perfectly conserved amino acid sequences in the Neh2 domain of these Nrf1/2-related proteins (Supplemental Fig. 3, *white characters highlighted in black*). Lysine residues also exist between these two motifs in every protein (Supplemental Fig. 3, *red characters*). So, it seems that the DLG and ETGE motifs are quite important for Nrf1/2-related proteins and that Keap1 proteins are important regulators of these proteins, even in invertebrates.

The second topic of this paper covers functional Keap1 proteins lacking either IC6 or IC7. The finding is inconsistent with those we and others observed in cultured cells, that both IC6 and IC7 are indispensable for mouse Keap1 activity (13-15). There are two explanations for this contradiction. Firstly, in Keap1a (lacks IC6) mRNA injected embryos, it is possible that exogenous Keap1a can heterodimerize with endogenous Keap1b. Likewise,

in Keap1b (lacks IC7) mRNA injected embryos, exogenous Keap1b may heterodimerize with endogenous Keap1a. This hypothesis is plausible, since we previously found that coexpression of Cys273Ala and Cys288Ala mutant proteins of mouse Keap1 led to the substantial restoration of repressor activity (15). Moreover, zebrafish Keap1a and Keap1b can form heterodimers and both genes are coexpressed in many cells. However, it was curious to discover that the Nrf2-repressing activities of overexpressed Keap1a and Keap1b were comparable in embryos, in which *keap1b* was dominantly expressed judged on RT-PCR analysis (see Figs. 3 and 6). Similarly, the mRNA expression of *keap1b* was undetectable in adult gut, where *keap1a* was dominantly expressed (see Fig. 6). The second idea is that the ubiquitin ligase machinery may differ in structure between fish and mammals, such that the effectual structure for Keap1 activity may also be distinctive. According to this idea, the tertiary structure of the IVR domain is more important than the presence or absence of each cysteine residue. This is contradictory to the zinc binding model proposed by Dinkova-Kostova et al. (33). They demonstrated that Keap1 is a zinc-containing protein and that alanine substitutions of both Cys273 and Cys288 reduced the binding affinity between Keap1 and zinc to 1/20 and suggested that these two cysteine residues participate in the binding to zinc. At present, it is difficult to adopt a proper hypothesis from these and other theories. In this context, it will be of interest to know whether Keap1a and Keap1b bind zinc in zebrafish embryos. Furthermore, the crystal structures of the IVR domains of various Keap1 proteins should be determined in the future.

REFERENCES

1. Kobayashi, M., and Yamamoto, M. (2006) *Adv. Enzyme Regul.* **46**, 113-140
2. Itoh, K., Chiba, T., Takahashi, S., Ishii, T., Igarashi, K., Katoh, Y., Oyake, T., Hayashi, N., Satoh, K., Hatayama, I., Yamamoto, M., and Nabeshima, Y. (1997) *Biochem. Biophys. Res. Commun.* **236**, 313-322
3. Kobayashi, M., and Yamamoto, M. (2005) *Antioxid. Redox Signal.* **7**, 385-394
4. Katsuoka, F., Motohashi, H., Ishii, T., Aburatani, H., Engel, J. D., and Yamamoto, M. (2005) *Mol. Cell. Biol.* **25**, 8044-8051
5. Itoh, K., Wakabayashi, N., Katoh, Y., Ishii, T., Igarashi, K., Engel, J. D., and Yamamoto, M. (1999) *Genes Dev.* **13**, 76-86
6. Wakabayashi, N., Itoh, K., Wakabayashi, J., Motohashi, H., Noda, S., Takahashi, S., Imakado, S., Kotsuji, T., Otsuka, F., Roop, D. R., Harada, T., Engel, J. D., and Yamamoto, M. (2003) *Nat. Genet.* **35**, 238-245
7. Kobayashi, A., Kang, M. I., Okawa, H., Ohtsuji, M., Zenke, Y., Chiba, T., Igarashi, K., and Yamamoto, M. (2004) *Mol. Cell. Biol.* **24**, 7130-7139
8. Kobayashi, A., Kang, M. I., Watai, Y., Tong, K. I., Shibata, T., Uchida, K., and Yamamoto, M. (2006) *Mol. Cell. Biol.* **26**, 221-229
9. Zipper, L. M., and Mulcahy, R. T. (2002) *J. Biol. Chem.* **277**, 36544-36552
10. Kang, M. I., Kobayashi, A., Wakabayashi, N., Kim, S. G., and Yamamoto, M. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 2046-2051
11. Padmanabhan, B., Tong, K. I., Ohta, T., Nakamura, Y., Scharlock, M., Ohtsuji, M., Kang, M. I., Kobayashi, A., Yokoyama, S., and Yamamoto, M. (2006) *Mol. Cell* **21**, 689-700
12. Lo, S. C., Li, X., Henzl, M. T., Beamer, L. J., and Hannink, M. (2006) *EMBO J.* **25**, 3605-3617
13. Zhang, D. D., and Hannink, M. (2003) *Mol. Cell. Biol.* **23**, 8137-8151
14. Levonen, A. L., Landar, A., Ramachandran, A., Ceaser, E. K., Dickinson, D. A., Zanoni, G., Morrow, J. D., and Darley-Usmar, V. M. (2004) *Biochem. J.* **378**, 373-382
15. Wakabayashi, N., Dinkova-Kostova, A. T., Holtzclaw, W. D., Kang, M. I., Kobayashi, A., Yamamoto, M., Kensler, T. W., and Talalay, P. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 2040-2045
16. Kobayashi, M., Itoh, K., Suzuki, T., Osanai, H., Nishikawa, K., Katoh, Y., Takagi, Y., and Yamamoto, M. (2002) *Genes Cells* **7**, 807-820
17. Takagi, Y., Kobayashi, M., Li, L., Suzuki, T., Nishikawa, K., and Yamamoto, M. (2004) *Biochem. Biophys. Res. Commun.* **320**, 62-69
18. Appel, B., and Eisen, J. S. (1998) *Development* **125**, 371-380

19. Hukriede, N., Fisher, D., Epstein, J., Joly, L., Tellis, P., Zhou, Y., Barbazuk, B., Cox, K., Fenton-Noriega, L., Hersey, C., Miles, J., Sheng, X., Song, A., Waterman, R., Johnson, S. L., Dawid, I. B., Chevrette, M., Zon, L. I., McPherson, J., and Ekker, M. (2001) *Genome Res.* **11**, 2127-2132
20. Kobayashi, M., Nishikawa, K., and Yamamoto, M. (2001) *Development* **128**, 2341-2350
21. Onodera, K., Shavit, J. A., Motohashi, H., Katsuoka, F., Akasaka, J. E., Engel, J. D., and Yamamoto, M. (1999) *J. Biol. Chem.* **274**, 21162-21169
22. Kobayashi, M., Nishikawa, K., Suzuki, T., and Yamamoto, M. (2001) *Dev. Biol.* **232**, 315-326
23. Katoh, Y., Iida, K., Kang, M. I., Kobayashi, A., Mizukami, M., Tong, K. I., McMahon, M., Hayes, J. D., Itoh, K., and Yamamoto, M. (2005) *Arch. Biochem. Biophys.* **433**, 342-350
24. Tong, K. I., Padmanabhan, B., Kobayashi, A., Shang, C., Hirotsu, Y., Yokoyama, S., and Yamamoto, M. (2007) *Mol. Cell. Biol.* **27**, 7511-7521
25. Prag, S., and Adams, J. C. (2003) *BMC Bioinformatics* **4**, 42
26. Suzuki, T., Takagi, Y., Osanai, H., Li, L., Takeuchi, M., Katoh, Y., Kobayashi, M., and Yamamoto, M. (2005) *Biochem. J.* **388**, 65-73
27. Itoh, K., Wakabayashi, N., Katoh, Y., Ishii, T., O'Connor, T., and Yamamoto, M. (2003) *Genes Cells* **8**, 379-391
28. Stogios, P. J., and Privé, G. G. (2004) *Trends Biochem. Sci.* **29**, 634-637
29. Tong, K. I., Katoh, Y., Kusunoki, H., Itoh, K., Tanaka, T., and Yamamoto, M. (2006) *Mol. Cell. Biol.* **26**, 2887-2900
30. Tong, K. I., Kobayashi, A., Katsuoka, F., and Yamamoto, M. (2006) *Biol. Chem.* **387**, 1311-1320
31. McMahon, M., Thomas, N., Itoh, K., Yamamoto, M., and Hayes, J. D. (2006) *J. Biol. Chem.* **281**, 24756-24768
32. Zhang, D. D., Lo, S. C., Cross, J. V., Templeton, D. J., and Hannink, M. (2004) *Mol. Cell. Biol.* **24**, 10941-10953
33. Dinkova-Kostova, A. T., Holtzclaw, W. D., and Wakabayashi, N. (2005) *Biochemistry* **44**, 6889-6899
34. Kobayashi, A., Ito, E., Toki, T., Kogame, K., Takahashi, S., Igarashi, K., Hayashi, N., and Yamamoto, M. (1999) *J. Biol. Chem.* **274**, 6443-6452
35. Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., and Postlethwait, J. H. (1998) *Science* **282**, 1711-1714

FOOTNOTES

[*]This work was supported by Grants-in-Aid from the Japan Science and Technology Corporation (ERATO) (to M.Y.) and the Ministry of Education, Science, Sports and Culture of Japan (to M.K. and M.Y.).

[1]To whom correspondence should be addressed: Institute of Basic Medical Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Japan. Tel.: +81-29-853-8457; Fax: +81-29-853-5977; E-mail: makobayash@md.tsukuba.ac.jp.

[2]The abbreviations used are: ARE, antioxidant responsive element; BTB, Broad-Complex,□Tramtrack and Bric-a-brac; DGR, double glycine□repeat; eGFP, enhanced GFP; EpRE, electrophile responsive element; HA, hemagglutinin; IC, IVR cysteine; IVR, intervening region; NLS, nuclear localizing signal; pg, picogram; RT-PCR, reverse transcriptase-PCR.

Acknowledgements — We thank M. Eguchi, T. Kinoshita, H. Niu, T. Shimokoube, Y. Terashita and Y. Wada for help in fish maintenance and R. Ide, F. Katsuoka, T. Suzuki, K. Takeuchi and T. Tsujita for help and discussion. We also thank A. Kobayashi, H. Motohashi, T. O'Connor and K. Tong for critical reading of the manuscript.

FIGURE LEGENDS

FIGURE 1. **Identification of two Keap1 proteins in zebrafish.** *A*, Percentage amino acid sequence identities in

the BTB, IVR and DGR domains between zebrafish and mouse Keap1 proteins. Nucleotide sequence data of zebrafish *keap1b* have been deposited in the DDBJ/EMBL/GenBank databases with the accession number AB271119. *B*, Amino acid sequence alignment of the DGR domains of Keap1 and Mayven proteins. Amino acid residues located in the interaction surface for Nrf2 are highlighted in *black*. GG and W sequences conserved among all BTB-Kelch family proteins are highlighted in *grey*. *White characters highlighted in grey* indicate amino acid residues whose mutations have been shown to reduce Nrf2 repressing activity. The *asterisks* donate stop codons. Abbreviations: *zKa*, zebrafish Keap1a; *zKb*, zebrafish Keap1b; *mK*, mouse Keap1; *mM*, mouse Mayven. *C*, Phylogenetic tree of Keap1 family proteins. Amino acid sequences in the DGR domains were analyzed. The tree was constructed by the NJ method using the CLUSTAL W program (http://clustalw.ddbj.nig.ac.jp/top-j.html). Abbreviations: *aa*, *Aedes aegypti*; *ag*, *Anopheles gambiae*; *c*, chicken; *ci*, *Ciona intestinalis*; *cs*, *Ciona savignyi*; *d*, *Drosophila melanogaster*; *f*, fugu; *h*, human; *m*, mouse; *me*, medaka fish; *te*, *Tetraodon nigroviridis*; *st*, stickleback; *xl*, *Xenopus laevis*; xt, *Xenopus tropicalis*; *z*, zebrafish.

FIGURE 2. **Alignments of 9 cysteine residues and their adjacent amino acid residues in the IVR of various Keap1 proteins.** Note that fish Keap1 proteins lack either IC6 (Keap1a) or IC7 (Keap1b), whereas Keap1 proteins in non-fish species have both of IC6 and IC7 without exception. Cysteine residues are highlighted in *black*. Basic amino acids adjacent to the cysteines are highlighted in *grey*. Abbreviations are as in Fig. 1.

FIGURE 3. **Effects of Keap1a and Keap1b on Nrf2-mediated *gstp1* induction.** *A*, Experimental scheme. Nrf2 and/or Keap1 mRNA were co-injected into embryos at the one-cell stage. After 8 hours, *gstp1* expression in embryos was analyzed by whole-mount *in situ* hybridization (*B*) or by real-time RT-PCR (*C*). *C*, In real-time RT-PCR analysis, FLAG-tagged Keap1 proteins were used instead of non-tagged proteins. The expression levels of *gstp1* were normalized with those of *ef1α* measured in the same cDNA preparations. Values for the Nrf2-overexpressing embryos were arbitrarily set at 100. The amounts of injected mRNA were 0.12, 0.6, 3, 15 and 75 pg for Keap1a and 0.32, 1.6, 8, 40 and 200 pg for Keap1b. *Error bars* indicate standard deviation values from three independent experiments.

FIGURE 4. **Effects of Keap1a and Keap1b on Nrf2 protein stability.** *A*, Immunostaining analysis of overexpressed Nrf2 protein. FLAG-tagged Nrf2 (*FL-Nrf2*) and/or Keap1 mRNA were injected into embryos at the one-cell stage. After 8 hours, the stability of FLAG-tagged Nrf2 protein was analyzed by immunostaining using anti-FLAG antibody. *B*, Expression analysis of Nrf2-GFP fusion protein. mRNA encoding a fusion protein comprising the N-terminal half of Nrf2 protein and enhanced GFP (*eGFP*) protein connected by two copies of SV40 nuclear localizing signal (*NLS*) was injected with or without mRNA encoding Keap1 proteins into one-cell stage embryos and GFP expression was analyzed after 6 hours.

FIGURE 5. **Dimerization of zebrafish Keap1 proteins.** FLAG (*FL*)- and HA-tagged proteins of Keap1a and Keap1b were examined by immunoblotting using anti-FLAG (*upper panel*) or anti-HA (*lower panel*) antibodies. Mixtures of FLAG-tagged and HA-tagged proteins were co-immunoprecipitated with anti-HA conjugated agarose beads. Precipitated proteins were analyzed by immunoblotting.

FIGURE 6. **Expression profiles of *keap1a* and *keap1b*.** *A*, RT-PCR analysis using specific primers for *keap1a*, *keap1b* and *ef1α* and total RNA from tissues of 10 month old male fish. *Br*, brain; *E*, eye; *Gi*, gill; *L*, liver; *Gu*, gut; *S*, spleen; *H*, heart; *K*, kidney; *T*, testis; *Bl*, bladder. *B*, RT-PCR analysis using total RNA from the whole bodies of embryos or larvae at the developmental stages indicated. *C*, *In situ* hybridization analysis. Lateral views of *keap1a* (*upper panels*) and *keap1b* (*lower panels*) expression at the developmental stages indicated. Dominant expression of *keap1a* was observed in lens (*arrow*).

FIGURE 7. **Effects of point mutations of the IVR cysteines.** *A*, Keap1 mutants used in the analysis. *B*, activities of Nrf2 repression of IC-mutated Keap1 proteins. Nrf2 and/or FLAG-tagged Keap1 (FL-Keap1) mRNA were co-injected into embryos at the one-cell stage. The amount of Keap1 mRNA was standardized by the expression level of FLAG-tagged Keap1 protein analyzed by immunoblotting using anti-FLAG antibody (FL-Keap1 protein). After 8 hours, *gstp1* expression in embryos was analyzed by RT-PCR analysis. The

expression of *ef1α* was used to standardize the amount of cDNA.