

Structural insights into the similar modes of Nrf2 transcription factor recognition by the cytoplasmic repressor Keap1

Balasundaram Padmanabhan,^a Kit I. Tong,^b Akira Kobayashi,^b Masayuki Yamamoto^{b,c,*} and Shigeyuki Yokoyama^{a,d,*}

^aGenomic Sciences Center, Yokohama Institute, RIKEN, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, ^bGraduate School of Comprehensive Human Sciences, Center for TARA, and JST-ERATO Environmental Response Project, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba 305-8577, Japan, ^cDepartment of Medical Biochemistry, Tohoku University Graduate School of Medicine, 2-1 Seiryō-cho, Aoba-ku, Sendai 980-8575, Japan, and ^dDepartment of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: masi@mail.tains.tohoku.ac.jp, yokoyama@biochem.s.u-tokyo.ac.jp

The cytoplasmic repressor Keap1 regulates the function of transcription factor Nrf2 which plays critical roles in oxidative and xenobiotic stresses. The Neh2 domain of Nrf2 interacts with Keap1 at the bottom region of the Kelch/ β -propeller domain which is formed by double-glycine repeat and C-terminal region domains (Keap1-DC). The structure of Keap1-DC complexed with an Nrf2 peptide containing a conserved DLG motif has been determined at 1.9 Å resolution. The Keap1-bound DLG peptide possesses a hairpin conformation, and it binds to the Keap1 protein at the bottom region of the β -propeller domain. The intermolecular interaction occurs through their complementary electrostatic interactions. Comparison of the present structure with the recently reported Keap1-DC complex structure suggests that the DLG and ETGE motifs of Neh2 in Nrf2 bind to Keap1 in a similar manner but with different binding potencies.

Keywords: oxidative stress; Nrf2 transcription factor; Keap1; β -propeller domain; structure of the complex.

1. Introduction

Higher animals have developed elaborate defense mechanisms to counteract the damage that can be provoked by electrophiles and reactive oxygen species (ROS) [see the review by Kobayashi & Yamamoto (2006), and references therein]. Reactive molecules, such as reactive oxygen species, electrophilic chemicals and heavy metals, damage biological macromolecules and impair normal cellular functions. Oxidative and xenobiotic stresses are known to cause diverged pathological processes, including cancer, cardiovascular disease, diabetes and neurodegeneration. A battery of genes encoding detoxifying and anti-oxidative stress enzymes/proteins is coordinately induced following exposure to electrophiles and ROS.

Nrf2 is the transcription factor for cytoprotective enzymes that counteract oxidative and electrophilic attacks (Itoh *et al.*, 1997). The cellular concentration of Nrf2 remains low in homeostatic/unstressed conditions, modulated by Keap1 (Kelch-like ECH associated protein 1) and proteasomal degradation (Itoh *et al.*, 2004). Nrf2 contains a conserved N-terminal regulatory domain termed Neh2, two trans-activation domains, and a C-terminal bZIP domain.

The Keap1 protein is the direct binding partner of Nrf2 and is the major negative regulator of cytoprotective gene expression (Itoh *et al.*, 1999). Keap1 possesses four characteristic domains: the BTB domain, the intervening region (IVR); the double glycine repeat or

kelch repeat (DGR) and the C-terminal region (CTR). The BTB domain, like its other structural homologs (Stogios *et al.*, 2005), is believed to serve in dimerization of Keap1 in the cytoplasm (Zipper & Mulcahy, 2002).

Keap1 represses Nrf2 transcription functions and targets Nrf2 for ubiquitin-dependent degradation, under unstressed conditions (Wakabayashi *et al.*, 2003; Zhang & Hannink, 2003; Motohashi & Yamamoto, 2004), by associating with the Neh2 domain of Nrf2 (Itoh *et al.*, 1999). Upon exposure to oxidative stress or electrophilic attack, Keap1 loses its ability to repress Nrf2, allowing Nrf2 to escape proteasomal degradation (Kobayashi *et al.*, 2006). Nrf2 then translocates to the nucleus and transcriptionally activates a battery of cytoprotective genes.

We have recently reported that the 'ETGE' motif in the Neh2 domain of Nrf2 binds to Keap1 at the bottom region of the β -propeller domain (also called the Kelch domain) (Padmanabhan *et al.*, 2006). As the other conserved 'DLG' motif in the Neh2 domain is also functionally important for association with Keap1, we have attempted to determine the C-terminal region of mouse Keap1, containing the β -propeller domain (hereafter Keap1-DC; amino acid residues 309–624), complexed with a 15-mer peptide containing the DLG motif. We report here the preliminary crystal structure analysis of the Keap1-DC complex with the DLG motif containing peptide at 1.9 Å resolution.

Table 1

Summary of data collection and refinement statistics for the Keap1-DC-Neh2 peptide complex.

Data collection	
Source	RAXIS IV ⁺⁺
Wavelength (Å)	1.5418
Space group	<i>P6₁</i>
Unit cell (Å)	<i>a</i> = <i>b</i> = 103.13, <i>c</i> = 56.14, γ = 120°
Resolution (Å)	20.0–1.9
Completeness (%)†	98.2 (88.0)
Redundancy	8.6 (3.6)
<i>R</i> _{merge} (%)‡	8.6 (39.8)
Refinement statistics	
No. of complex molecules in a.u.	1
Resolution limit (Å)	20.0–1.9
σ cutoff (<i>F</i>)	0
No. of reflections	25143
<i>R</i> _{work} / <i>R</i> _{free} (%)§¶	17.2/21.1
No. of protein residues	295
No. of peptide residues	6
No. of SO ₄ ions	7
No. of water molecules	307
Average <i>B</i> -factor (Å ²)	
Protein	32.5
Water	45.5
r.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.55

† Numbers in parentheses are values in the highest resolution shell (1.95–1.90 Å). ‡ $R_{\text{merge}} = \sum_i \sum_h |I(h)_i - \langle I(h) \rangle| / \sum_i \sum_h I(h)_i$, where $I(h)$ is the observed intensity of reflection h , $\langle I(h) \rangle$ is the mean intensity of reflection h over all measurements of $I(h)$, \sum_h is the sum over all reflections and \sum_i is the sum over i measurements of reflection h . § $R_{\text{work}} = \sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$. ¶ R_{free} was calculated with 5% of data omitted from the refinement.

2. Materials and methods

2.1. Crystallization and data collection

The mouse Keap1-DC domain was expressed in *Escherichia coli* and purified as previously described (Padmanabhan *et al.*, 2005). The crystals for the Keap1-DC complex were obtained by the co-crystallization method. The DLG motif containing peptide (²²ILWRQ-DIDLGVSRV³⁶, mouse Nrf2), purchased from Promega, was mixed with Keap1-DC (4 mg ml⁻¹) (10:1 molar ratio) and incubated at 277 K for about one day before setting up the crystallization experiment. The co-crystals were obtained in a drop containing 0.8 M lithium sulfate, 0.5 M ammonium sulfate and 0.1 M sodium citrate, pH 5.2, at 293 K. Complete diffraction data were collected under cryogenic conditions, using a Rigaku RA-Micro7 Cu $K\alpha$ rotating-anode X-ray generator, operated at 40 kV and 20 mA and equipped with a Rigaku RAXIS IV⁺⁺ imaging-plate area detector and an X-stream low-temperature system. The diffraction data were indexed and processed using *HKL2000* (Otwinowski & Minor, 1997). The results of the data reduction statistics are given in Table 1.

2.2. Structure determination and refinement

The structure of the Keap1-DC complex was determined by the molecular replacement method, using the apo-form of Keap1-DC structure (Padmanabhan *et al.*, 2006) as a search model and the program *Molrep* from the CCP4 suite (Collaborative Computational Project, Number 4, 1994). It gave a distinct peak with an *R* factor and correlation coefficient of 0.363 and 0.690, respectively, at a resolution between 20 Å and 3.0 Å. The model was refined using *CNS* (Brunger *et al.*, 1998), and several rounds of manual fitting and re-fitting were carried out using the program *O* (Jones *et al.*, 1991), with careful inspection of the $2Fo - Fc$, $Fo - Fc$ and omit electron density maps. During the final stage of refinement, *Refmac5* (Murshudov *et al.*,

1997), which is incorporated into the CCP4 suite, was used for refinement of the structure. The current refined model consists of 301 residues, seven SO₄ ions and 307 water molecules, with final *R*_{work} and *R*_{free} values of 17.2% and 21.1%, respectively, at a resolution of 1.9 Å. In Keap1-DC the electron density for the first 15 residues at the N-terminus and the last 11 residues at the C-terminus, and in the peptide, residues from 22–23 and 30–36 were absent. A Ramachandran analysis using the program *PROCHECK* (Laskowski *et al.*, 1993) showed that 224 residues (90%) were in the most favoured region, 24 residues (9%) were in the additionally allowed region, one residue (0.5%) was in the generously allowed region, and one residue (0.5%) was in the disallowed region. The refinement statistics are summarized in Table 1. The atomic coordinates for the mouse Keap1-DC-Neh2 peptide complex have been deposited in the Protein Data Bank under the accession number 2DYH.

3. Results and discussion

3.1. Overall structure of the Keap1-DC-DLG peptide complex

The structure of this complex was solved by the molecular replacement method using the apo-form of the Keap1-DC structure (Padmanabhan *et al.*, 2006) as a model at 2.5 Å resolution, and refined to 1.9 Å resolution. The Keap1-DC structure possesses a β -propeller domain containing six blades, with sixfold pseudo symmetry (Fig. 1). Each blade is a twisted β -sheet formed by four antiparallel β -strands (β 1– β 4). As found in other β -propeller domains, the β -strand (β 1) in the C-terminal region interacts with the β -strand (β 2) of the N-terminal region (Fig. 1a); thereby the closure of this disc-like molecule is achieved by ‘3 + 1’ division strands derived from N- and C-termini (‘velcro’ closure or ‘molecular clasp’) (Neer & Smith, 1996). The Keap1-DC structure in the present complex is essentially similar to that of the apo-form of Keap1-DC (Padmanabhan *et al.*, 2006) (root mean square deviation of 0.2 Å for main-chain atoms).

The difference Fourier map unambiguously showed the electron density at the peptide bound region. Although we used a peptide that was 15 amino acids long for co-crystallization, from Ile22 to Val36, the electron density was visible only for six amino acids, from Trp24 to Asp29. The Trp24 and Ile28 residues were truncated to Ala as the electron densities of the side-chains were absent. Since the electron density of the side-chain atoms after C ^{δ} was not visible in Arg25 either, these atoms were not included in refinement of the structure. The final *R*_{work} and *R*_{free} values were 17.2% and 21.1%, respectively, for the 301 residues, seven sulfate ions and 307 water molecules, refined to a resolution of 1.9 Å (Table 1).

The DLG peptide binds to Keap1-DC at the bottom side of the six-bladed β -propeller (Fig. 1). The bottom region of the β -propeller domain is highly basic and mainly occupied by arginine residues (Fig. 2). In the interface region the peptide is surrounded by Tyr334, Asn382, Arg483, Tyr525, Tyr572, Phe577, Arg415, Ser508, Gly509, Ser555, Ala556, Ser602 and Gly603 of Keap1. The DLG peptide appears to be positioned close to the fourth, fifth and sixth blades of the Kelch domain; however, the peptide interacts with residues from almost all of the six blades.

The DLG peptide possesses a tight four-residue β -hairpin conformation comprising the residues Arg25, Gln26, Asp27, Ile28 and Asp29 (Fig. 2). The structure of this peptide is stabilized by two intramolecular hydrogen bonds: between Arg25 and Ile28, and Trp24 and Asp29. The formation of such a β -hairpin conformation in the peptide allows the residues of the DLG motif to interact substantially with the Keap1-DC amino acid residues.

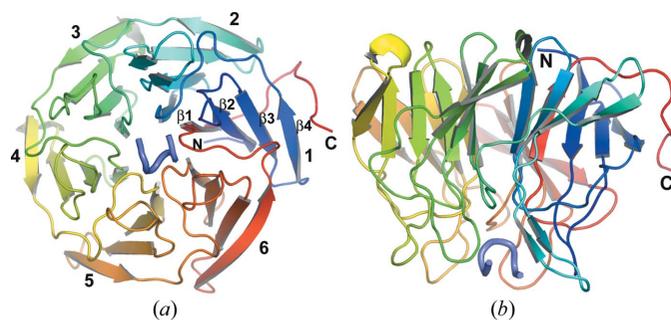


Figure 1
The overall tertiary structure of mouse Keap1-DC complexed with the DLG peptide. Ribbon model of the tertiary structure of the mKeap1-DC β -propeller domain (blue to red) and the DLG peptide (slate). (a) Bottom view. (b) Side view. The figure was generated using *PyMOL* (<http://www.pymol.org/>).

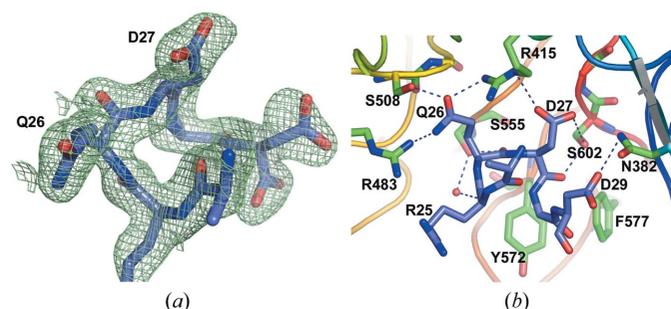


Figure 2
Close-up view of the peptide binding region. (a) Part of the refined DLG peptide showing Gln26 and Asp27. The final electron density $2mF_o - DF_c$ map is contoured at 1.0σ . (b) The interacting residues of the complex are shown by sticks, hydrogen bonds by dashed lines, and a water molecule by a ball.

The Keap1-DC–DLG peptide complex generates eight potential intermolecular electrostatic interactions between the peptide and the β -propeller domain. Gln26 of Neh2 is a conserved residue within the CNC protein family (Katoh *et al.*, 2005) and has significant intermolecular interactions with Ser508, Arg415, Arg483 and Ser555 of Keap1-DC (Fig. 2). For instance, the side-chain of Gln26 is wedged between Arg415 and Arg483. Furthermore, the O ^{ϵ 1} atom of Gln26 contributes a hydrogen bond and a salt bridge to Ser508 and Arg415, respectively, while the N ^{ϵ 2} atom of Gln26 interacts electrostatically with Arg483. In addition to these side-chain interactions, the main-chain carbonyl group of Gln26 is hydrogen bonded to the side-chain of Ser555.

Asp27 of the DLG motif interacts with multiple residues, such as Ser602, Gly603 and Arg415, of Keap1 (Fig. 2). On the other hand, the side-chain of the next residue, Ile28, which adopts a left-handed helical conformation, was absent in the structure. However, this residue is nicely positioned in a hydrophobic pocket produced by Tyr572 and Phe577. Moreover, Asp29, the last residue visible in the electron density map of the DLG peptide, contributes to an electrostatic interaction with Asn382 of Keap1.

3.2. Comparison with Keap1-DC–ETGE peptide complex structure

We have recently reported the structure of Keap1-DC complexed with Nrf2 peptide containing the ETGE motif (Padmanabhan *et al.*, 2006). Comparison of the Keap1-DC–DLG peptide complex with the Keap1-DC–ETGE peptide complex revealed that both DLG and ETGE fragments interact with Keap1 in a similar manner, by binding through the bottom region of the β -propeller structure (Fig. 3). Superimposition of the structures of these two complexes over the

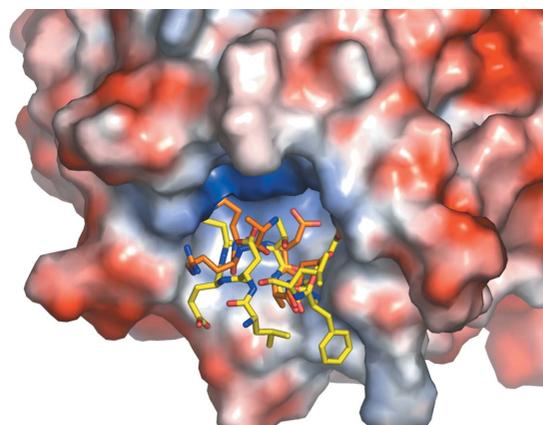


Figure 3
Superimposition of the DLG peptide complex with the ETGE peptide complex. Electrostatic surface potential of mKeap1-DC in the mKeap1-DC peptide complexes. Surface acidic, basic and neutral residues are shown in red, blue and white, respectively. The protein-bound ETGE (yellow) and DLG (slate) peptides are shown by sticks.

main-chain atoms of Keap1-DC revealed that the overall structure of Keap1-DC in these two complexes is almost the same (0.23 Å r.m.s. deviation). However, a small variation was observed in the loop connecting strands β 2 and β 3 of the second blade of the Kelch domain. At the binding region, the side-chains of the Keap1-DC residues possess a very similar conformation within the two complexes, except for Arg415, Arg483 and Asn382 (not shown).

Although both peptide structures are quite flexible, they adopt a similar kind of tight β -turn conformation and also orientate in essentially the same manner with respect to the Keap1-DC structure (Fig. 3). However, the ETGE peptide possesses a comparatively higher number of electrostatic interactions with Keap1-DC than the DLG peptide: approximately 13 in the ETGE peptide complex *versus* eight in the DLG peptide complex. Furthermore, the ETGE peptide is significantly embedded into the Keap1-DC binding cleft compared with the DLG peptide. For instance, the side-chain of Glu79 in the ETGE is well buried in the pocket, whereas the equivalent residue in the DLG, Gln26, is partially buried (Fig. 3). These results therefore suggest that, although the bottom region of Keap1-DC is common for the ETGE and DLG motif interactions, their binding strengths are comparatively different with respect to each other. Our extensive point mutation studies for the ETGE and DLG motifs interactions with Keap1 suggested that the ETGE motif has a higher affinity than the DLG motif (Tong, Kobayashi *et al.*, 2006; Tong *et al.*, 2007).

Data obtained to date suggest that the Keap1 homodimer binds to one Nrf2 by means of two recognition sites, the DLG and ETGE motifs in the Neh2 domain of Nrf2. The Keap1 protein exists as a homodimer in solution (Tong, Kobayashi *et al.*, 2006). The dual recognition sites have physiological relevance, since it has been shown that mutation or deletion of the DLG motif hampers both Keap1-dependent ubiquitination and proteasomal degradation (McMahon *et al.*, 2004; Katoh *et al.*, 2005). Based on the structures of Keap1-DC complexes and functional studies, we have recently proposed a two-site substrate recognition hinge-latch model of the Keap1–Nrf2 system for regulatory mechanism of Nrf2 activation/depression (Tong, Katoh *et al.*, 2006; Tong *et al.*, 2007).

In summary, the present Keap1-DC–DLG peptide complex structure revealed that Keap1 shares the common binding site for both the DLG and ETGE motifs of Nrf2 interactions but with different binding strengths. However, additional structure studies on the complete Neh2 of Nrf2 with Keap1-DC are necessary to further

understand the molecular mechanism of the Nrf2 transcription factor which plays an important role in environmental stress.

We thank Drs Y. Nakamura and Chenwei Shang for crystallization and data collection. This work was supported in part by the RIKEN Structural Genomics/Proteomics Initiative, the National Project on Protein Structural and Functional Analyses, the Japanese Ministry of Education, Culture, Sports, Science and Technology and ERATO-JST.

References

- Brünger, 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., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Itoh, K., Chiba, T., Takahashi, S., Ishii, T., Igarashi, K., Katoh, Y., Oyake, T., Hayashi, N., Satoh, K., Hatayama, I., Yamamoto, M. & Nabeshima, Y. (1997). *Biochem. Biophys. Res. Commun.* **236**, 313–322.
- Itoh, K., Tong, K. I. & Yamamoto, M. (2004). *Free Radic. Biol. Med.* **36**, 1208–1213.
- Itoh, K., Wakabayashi, N., Katoh, Y., Ishii, K., Igarashi, K., Engel, J. D. & Yamamoto, M. (1999). *Genes Dev.* **13**, 76–86.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Katoh, Y., Iida, K., Kang, M. I., Kobayashi, A., Mizukami, M., Tong, K. I., McMahon, M., Hayes, J. D., Itoh, K. & Yamamoto, M. (2005). *Arch. Biochem. Biophys.* **433**, 342–350.
- Kobayashi, A., Kang, M. I., Watai, Y., Tong, K. I., Shibata, T., Uchida, K. & Yamamoto, M. (2006). *Mol. Cell. Biol.* **26**, 221–229.
- Kobayashi, M. & Yamamoto, M. (2006). *Advan. Enzyme Regul.* **46**, 113–140.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- McMahon, M., Thomas, N., Itoh, K., Yamamoto, M. & Hayes, J. D. (2004). *J. Biol. Chem.* **279**, 31556–31567.
- Motohashi, H. & Yamamoto, M. (2004). *Trends Mol. Med.* **10**, 549–557.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Neer, E. J. & Smith, T. F. (1996). *Cell*, **84**, 175–178.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Padmanabhan, B., Scharlock, M., Tong, K. I., Nakamura, Y., Kang, M.-I., Kobayashi, A., Matsumoto, T., Tanaka, A., Yamamoto, M. & Yokoyama, S. (2005). *Acta Cryst.* **F61**, 153–155.
- Padmanabhan, B., Tong, K. I., Ohta, T., Nakamura, Y., Scharlock, M., Ohtsuji, M., Kang, M. I., Kobayashi, A., Yokoyama, S. & Yamamoto, M. (2006). *Mol. Cell*, **21**, 689–700.
- Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K. & Privé, G. G. (2005). *Genome Biol.* **6**, R82.
- Tong, K. I., Katoh, Y., Kusunoki, H., Itoh, K., Tanaka, T. & Yamamoto, M. (2006). *Mol. Cell. Biol.* **26**, 2887–2900.
- Tong, K. I., Kobayashi, A., Katsuoka, F. & Yamamoto, M. (2006). *Biol. Chem.* **387**, 1311–1320.
- Tong, K. I., Padmanabhan, B., Kobayashi, A., Shang, C., Hirotsu, Y., Yokoyama, S. & Yamamoto, M. (2007). *Mol. Cell. Biol.* **27**, 7511–7521.
- 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. & Yamamoto, M. (2003). *Nat. Genet.* **35**, 238–245.
- Zhang, D. D. & Hannink, M. (2003). *Mol. Cell Biol.* **23**, 8137–8151.
- Zipper, L. M. & Mulcahy, R. T. (2002). *J. Biol. Chem.* **277**, 36544–36552.