Mol. Cells *33*, 163-171, February 29, 2012
DOI/10.1007/s10059-012-2220-6

Molecules and Cells

Dissection of the Dimerization Modes in the DJ-1 Superfamily

Hoi Jong Jung[1,4,5], Sangok Kim[2,5], Yun Jae Kim[1], Min-Kyu Kim[1], Sung Gyun Kang[1,3], Jung-Hyun Lee[1,3], Wankyu Kim[2,*], and Sun-Shin Cha[1,3,*]

The DJ-1 superfamily (DJ-1/ThiJ/PfpI superfamily) is distributed across all three kingdoms of life. These proteins are involved in a highly diverse range of cellular functions, including chaperone and protease activity. DJ-1 proteins usually form dimers or hexamers *in vivo* and show at least four different binding orientations *via* distinct interface patches. Abnormal oligomerization of human DJ-1 is related to neurodegenerative disorders including Parkinson's disease, suggesting important functional roles of quaternary structures. However, the quaternary structures of the DJ-1 superfamily have not been extensively studied. Here, we focus on the diverse oligomerization modes among the DJ-1 superfamily proteins and investigate the functional roles of quaternary structures both computationally and experimentally. The oligomerization modes are classified into 4 types (DJ-1, YhbO, Hsp, and YDR types) depending on the distinct interface patches (I-IV) upon dimerization. A unique, rotated interface via patch I is reported, which may potentially be related to higher order oligomerization. In general, the groups based on sequence similarity are consistent with the quaternary structural classes, but their biochemical functions cannot be directly inferred using sequence information alone. The observed phyletic pattern suggests the dynamic nature of quaternary structures in the course of evolution. The amino acid residues at the interfaces tend to show lower mutation rates than those of non-interfacial surfaces.

INTRODUCTION

Human DJ-1, a representative member of the DJ-1/ThiJ/PfpI superfamily (hereafter called the DJ-1 superfamily), is implicated in the pathogenesis of Parkinson's disease (Bonifati et al., 2003). Human DJ-1 is involved in various cellular functions, including protection against environmental stress (Moore et al., 2005), apoptosis (Junn et al., 2005), and regulation of tumor suppressors (Kim et al., 2005). For example, DJ-1 plays a protective role against oxidative stress (Lee et al., 2003; Shendelman et al., 2004; Yang et al., 2005). DJ-1 also acts as a redox-sensitive negative regulator of the PI3K/Akt survival signalling pathway and blocks the Daxx-ASK1 proapoptotic pathway by sequestering Daxx from ASK1 (Junn et al., 2005; Kim et al., 2005; Yang et al., 2005).

Proteins belonging to the DJ-1 superfamily are evolutionarily distributed in Archaea, Bacteria, and Eukaryota (Lucas and Marin, 2007). Structural comparison among DJ-1 family members revealed that the core structure consists of a central six-stranded parallel sheet and eight helices packing on either face of the sheet (Lee et al., 2003). In the core structure, a "nucleophile elbow" motif is present, carrying an absolutely conserved cysteine which is commonly found in α/β hydrolases (Ollis et al., 1992). In spite of their homologous tertiary structures, no conservation in the biochemical activity of the DJ-1 superfamily is apparent (Du et al., 2000; Halio et al., 1997; Horvath and Grishin, 2001; Lee et al., 2003; Mizote et al., 1999; Quigley et al., 2003; Shendelman et al., 2004). Human DJ-1 has chaperone activity (Lee et al., 2003), and *Escherichia coli* Hsp31 has both peptidase and chaperone activities (Lee et al., 2003; Mujacic and Baneyx, 2007; Sastry et al., 2002; Shendelman et al., 2004). However, *E. coli* YhbO and *Deinococcus radiodurans* DR1199 exhibit neither chaperone nor peptidase activity, although they are implicated in stress responses like human DJ-1 and *E. coli* Hsp31 (Abdallah et al., 2006; 2007; Fioravanti et al., 2008). In this regard, it is worth noting that the quaternary structures of YhbO and DR1119 are different from those of DJ-1 and Hsp31.

Alternative oligomerization modes are a source of generating functional diversity among homologs with a similar tertiary structure (Kim et al., 2006) and can be critical for protein function (Li et al., 2009). Generally, DJ-1 superfamily members form dimers or hexamers, which are important for their stability or their biochemical activity (Gorner et al., 2007; Lee et al., 2003; Wei et al., 2007; Wilson et al., 2005). DJ-1, DR1199, YDR533Cp, and Hsp31 are dimers, although their dimeric conformations are distinct, whereas PH1704 forms a hexamer (Abdallah et al., 2006; Du et al., 2000; Fioravanti et al., 2008; Lee et al., 2003; Quigley et al., 2003; Wilson et al., 2004). The

[1]Marine Biotechnology Research Center, Korea Ocean Research and Development Institute, Ansan 426-744, Korea, [2]Ewha Research Center for Systems Biology, Division of Molecular and Life Sciences, Ewha Womans University, Seoul 120-750, Korea, [3]Department of Marine Biotechnology, University of Science and Technology, Daejeon 305-333, Korea, [4]Present address: Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries, Anyang 431-810, Korea, [5]These authors contributed equally to this work.
*Correspondence: wkim@ewha.ac.kr (WK); chajung@kordi.re.kr (SSC)

Received October 11, 2011; revised November 12, 2011; accepted November 14, 2011; published online January 2, 2012

Keywords: DJ-1 superfamily, DJ-1/ThiJ/PfpI superfamily, quaternary structure

hexameric structure of PH1704 can be considered as a trimer of dimers (Du et al., 2000). Recently, we showed that the dimeric form of DJ-1 can stack one by one to form filamentous aggregates (Cha et al., 2008). Consequently, the dimeric form can be regarded as the basic oligomerization unit in the DJ-1 superfamily.

Here, we investigate distinct dimerization/oligomerization modes or orientations for virtually all the DJ-1 superfamily members whose 3D structures are available. The functional roles of quaternary structures are also examined both computationally and experimentally. Since DJ-1 superfamily proteins are widely distributed taxonomically and have highly diverse sequences, quaternary structures, and cellular function, they are ideal candidates to investigate various factors affecting functional and structural diversification in the course of protein evolution. Several previous studies mainly focused on sequence conservation or active sites (Bandyopadhyay and Cookson, 2004; Wei et al., 2007). Nevertheless, the quaternary structures of DJ-1 superfamily members have not been investigated thoroughly. Improved classification and comparison systems for the DJ-1 superfamily will help to elucidate the evolutionary relationship among members in this huge protein family.

MATERIALS AND METHODS

Data collection and phylogenetic analysis

DJ-1 superfamily protein sequences were collected using both a keyword (DJ-1) and homology-based search from the protein database at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) based on annotation as DJ-1. Additionally, sequences of known 3D structures were collected from SCOP (http://scop.mrc-lmb.cam.ac.uk). The quaternary structures and patch type were classified by visual inspection after structural superposition and further verified by referring to SCOPPI (http://www.scoppi.org) (Winter et al., 2006).

DJ-1 protein sequences were aligned using ClustalW 2.0 at default parameters (Larkin et al., 2007). Putatively multi-domain proteins of unusually long sequences were excluded by manual inspection of aligned sequences. After the filtering step, our final data consisted of 934 sequences of the DJ-1 superfamily. Phylogenetic trees were constructed by the Neighbor-Joining (NJ) method with 2,000 bootstrap replicates using MEGA 4.0 (Saitou and Nei, 1987), where Poisson correction was used for NJ. For a detailed analysis within each group, we used both NJ and UPGMA in MEGA 4.0 (Tamura et al., 2007). For the inspection of conserved residues, Tcoffee (Notredame et al., 2000) was used for multiple alignments as well as ClustalW 2.0.

Protein expression and purification

The *TON1285* gene from *Thermococcus onnurineus* NA1 was amplified by polymerase chain reaction, digested with *Nde*I and *Xho*I, and inserted downstream of the T7 promoter of pET-28a or pET-24a (Novagen). The resulting construct in pET-28a contained residues 1-168 with an additional 20 residues including six histidines at its N-terminus (MGSSHHHHHHSSGLV PRGSH), and the constructs in pET-24a contained residues 1-168 with an additional 8 residues including six histidines at its C-terminus (LEHHHHHH) or no histidine tag.

After verifying the DNA sequence, plasmid DNA was transformed into *E. coli* strain BL21 (DE3). The transformed cells were grown in Luria-Bertani (LB) medium (Merck) containing 50 μg ml^{-1} kanamycin to an OD_{600} of 0.6 at 210 K, and expression of TON1285 was induced with 1 mM isopropyl-D-thiogalactoside (IPTG; Duchefa). After 4 h induction at 310 K, cells were harvested by centrifugation. The harvested cells were washed, resuspended with 20 mM Tris-HCl pH 7.5 and 500 mM NaCl and disrupted by sonication. The cell debris was discarded by centrifugation at 20,000 $\times$ *g* for 30 min at 277 K. The resulting supernatant was loaded onto a nickel-nitrilotriacetic acid column (Ni-NTA; Qiagen). The column was washed with buffer containing 20 mM Tris-HCl pH 7.5, 500 mM NaCl, and 30 mM imidazole. TON1285 was eluted with the same buffer containing 500 mM imidazole. The partially-purified protein fraction by Ni-NTA was dialysed into 20 mM Tris-HCl pH 7.5 and 1 mM dithiothreitol solution and loaded onto a column packed with 30 ml HiTrap Q HP resin (GE Healthcare). The bound proteins were then eluted with a gradient to 20 mM Tris-HCl pH 7.5, 1 M NaCl, and 1 mM dithiothreitol. The eluted fraction containing TON1285 was concentrated and subsequently loaded onto a Superdex 75 HR 16/60 column (GE Healthcare) pre-equilibrated with buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 10 mM dithiothreitol. The purified TON1285 was concentrated to approximately 15 mg ml^{-1} for crystallization.

The *ydeA* gene (GenBank Accession No. BAA19347) from *Bacillus subtilis* was amplified by polymerase chain reaction, digested with *Nde*I and *Hin*dIII, and inserted downstream of the T7 promoter of pET-21a (Novagen). The resulting construct contained residues 1-197 with no histidine tag. After verifying the DNA sequence, plasmid DNA was introduced into *E. coli* strain Rosetta (DE3). The transformed cells were grown in LB medium (Merck) containing 0.1 mg ml^{-1} ampicillin to an OD_{600} of 0.6 at 310 K, and expression of YdeA was induced with 1 mM IPTG (Duchefa). After 4h induction at 310 K, cells were harvested, resuspended in 20 mM Tris-HCl pH 7.5, disrupted by sonication, and finally centrifuged at 20,000 $\times$ *g* for 30 min at 277 K. The resulting supernatants were loaded onto an SP-Sepha-rose Fast Flow column (GE Healthcare). Then the unbound fraction was consecutively loaded on a Q-Sepharose Fast Flow column (Amersham Biosciences) that was pre-equilibrated with 20 mM Tris-HCl pH 7.5. The bound proteins were then eluted with a gradient to 20 mM Tris-HCl pH 7.5 and 500 mM NaCl. The eluted fraction containing YdeA was concentrated and subsequently loaded onto a Superdex 75 16/60 column (GE Healthcare). DJ-1 was purified as described (Lee et al., 2003).

Protein crystallization and structure determination

Crystallization screening was performed with commercially available screening kits from Hampton Research, Emerald BioStructures, and Axygen Biosciences using the microbatch crystallization method. Small drops composed of 1 μl protein solution and an equal volume of crystallization reagent were pipetted under a layer of a 1:1 mixture of silicon oil and paraffin oil in 72-well HLA plates (Nunc). Crystals of TON1285 were grown at 295 K in a precipitant containing 20% PEG 8000, 0.1 M HEPES pH 7.5, and 0.2 M ammonium sulfate. The crystals were frozen at 100 K using a Cryostream cooler (Oxford Cryosystem) after brief immersion in a cryoprotectant solution containing 15% glycerol in the same precipitant solution. A 1.78 Å resolution data set was collected using an ADSC Quantum 315 CCD on beamline 4A of the Pohang Light Source, Republic of Korea. Diffraction data were processed using *DENZO* and scaled using *SCALEPACK* from the *HKL-2000* program suite (Otwinowski and Minor, 1997). The crystals belonged to the hexagonal space group $P6_3$, with unit-cell parameters a = b = 87.11, c = 80.12 Å. The calculated crystal volume per unit of molecular mass (V_M) was 2.3 Å^3 Da^{-1} with a solvent content of 45.8% by volume when the asymmetric unit was assumed to contain two TON1285 monomers. Phasing information was obtained by molecular replacement using the program *MOLREP*

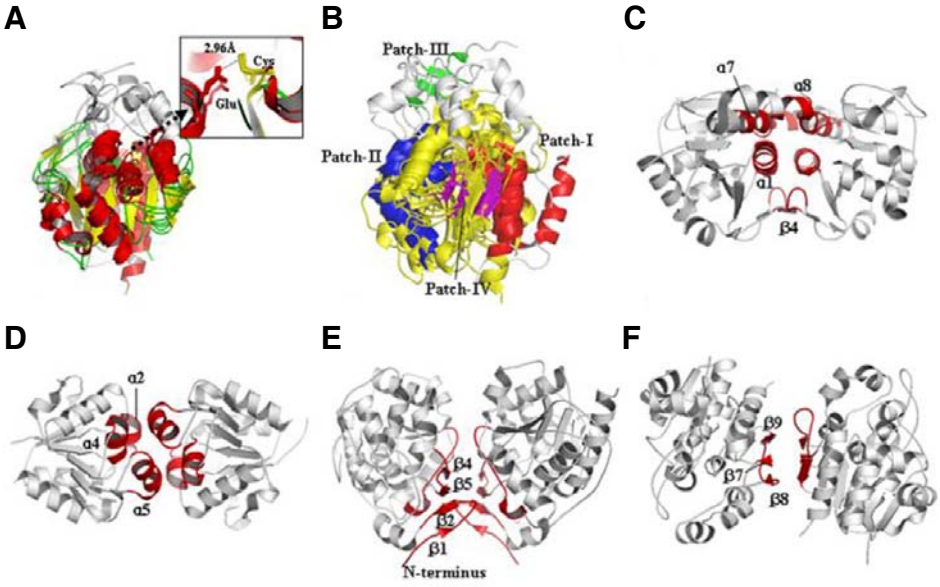

Fig. 1. Tertiary and quaternary structures of the DJ-1 superfamily. (A) Superposed 3D structures of canonical DJ-1 superfamily members (YhbO, DJ-1, Hsp31, and YDR533Cp) reveal common features of the DJ-1 superfamily (α-helices in red, β-strands in yellow, and extended regions in gray). DJ-1, Hsp31, and YDR533Cp yielded root mean squared deviations of 3.09 Å, 5.59 Å, and 3.44 Å to YhbO, respectively. (B) Four distinct patches (I-IV) are shown in different colors on the superposed 3D structures (patch I in red, patch II in blue, patch III in green, and patch IV in magenta). Little spatial overlap exists between different patches. (C) The DJ-1 type dimer by patch I. (D) The YhbO type dimer by patch II. (E) The Hsp type type dimer by patch III. (F) The YDR type dimer by patch IV. The patch regions of dimeric interfaces are shown in red in (C-F).

in the CCP4 program suite (Vagin and Teplyakov, 2010). The structure of PH1704 [PDB code 1G2I, (Du et al., 2000)] was used as a search model. Model building was performed using *COOT* (Emsley et al., 2010), and refinement was performed with a maximum-likelihood algorithm implemented in *CNS* (Brunger et al., 1998). The subsequent refinement and manual refitting of the initial model reduced R and R_{free} values to 15.4% and 21.3%, respectively. The ideality of the stereochemistry of the final model containing residues 1-166 was verified by *PROCHECK* (Laskowski et al., 1996). The Ramachandran plot indicates 97.3% of the non-glycine residues are in the most favored regions, and the remaining 2.7% of the residues are in the allowed region.

Chaperone activity assay

Chaperone activity was measured by monitoring the effects of DJ-1 and YdeA on the thermal aggregation of citrate synthase (Sigma) at 319 K. DJ-1 and YdeA in 20 mM Tris pH 7.5 buffer and 1mM dithiothreitol with 0.5 mM H_2O_2 was preincubated with citrate synthase for 5 min. After preincubation, the absorbance was measured on a UV-2401PC spectrophotometer (SHIMAZU) equipped with a constant temperature cell holder.

Proteolytic activity assay

Protease activities were assayed by monitoring the production of MCA from a fluorogenic peptide, AAF-MCA. TON1285 (0.15 mM) was incubated for 20 min with 0.3 mM substrate in 20 mM Tris pH 7.5, 1 mM dithiothreitol, and 100 mM NaCl from 303 K to 363 K. MCA fluorescence was measured at 360 ± 40 nm excitation and 460 ± 40 nm emission on an F-2000 fluorescence spectrophotometer (HITACHI).

RESULTS AND DISCUSSION

Dimerization modes of the DJ-1 superfamily

The DJ-1 superfamily members share a similar tertiary structure except for extended or terminal regions (Fig. 1A) (Gorner et al., 2007; Lee et al., 2003; Wei et al., 2007; Wilson et al., 2005). In the shared tertiary structure, two functional residues are highly conserved; a cysteine residue at the nucleophilic elbow and a nearby glutamate (or aspartate) residue (Fig. 1A). The cysteine residue (Cys106 in DJ-1) serves as a nucleophile in peptidases/proteases in the DJ-1 superfamily (Lee et al., 2003). Furthermore, its oxidation is believed to be critical to the cytoprotective activity of DJ-1 (Canet-Aviles et al., 2004; Neumann et al., 2004). The glutamate residue (Glu18 in DJ-1) is responsible for the decreased pK_a value of Cys106, facilitating its oxidation in DJ-1 (Witt et al., 2008).

The majority of the DJ-1 superfamily members are dimers. However, in spite of similar tertiary structures, close examination of their dimeric conformations reveal the highly diverged nature of their quaternary structures. At least four interface patches (patches I-IV) mediate four distinct dimerization modes in the DJ-1 superfamily. Interestingly, dimerization is achieved through the same patch of the two subunits in a symmetric conformation in all known cases. For instance, patch I of one subunit makes contact with the same patch I in another subunit, not with patches II-IV. Patch I is responsible for the dimerization of DJ-1, Atu0886, SCRP-27A, YajL, YdeA, YP001094981.1, and ThiJ/PfpI-like proteins. In DJ-1, patch I consists of α1, β4, the N-terminal part of α7, the C-terminal part of α8, and two loop regions, which are between β3 and β4 and between β11 and α7 (Fig. 1C). Hereafter, the dimeric form generated by patch I is called the DJ-1 type. It is notable that the dimeric contact of SCRP-27A is different from those observed in other DJ-1 type dimers, where one patch I is rotated by ~90° with respect to the other patch I. Therefore, we denote this rotated contact observed in SCRP-27A as patch I'. This rotated contact seems to be related to higher order oligomerization, which is discussed in the next section.

Patch II is involved in the dimerization of YhbO, PH1704, and DR1199. Three helices (α2, α4, and α5) form patch II in YhbO (Fig. 1D). Hereafter, the dimeric contact mediated by patch II is called the YhbO type. Hsp31 is extraordinary in that it has an additional N-terminal extension of 45 residues and an inserted domain near patches I and II (Lee et al., 2003; Quigley et al., 2003; Sastry et al., 2002). Dimeric contact in Hsp31 is mediated by a unique surface named patch III, consisting of the N-

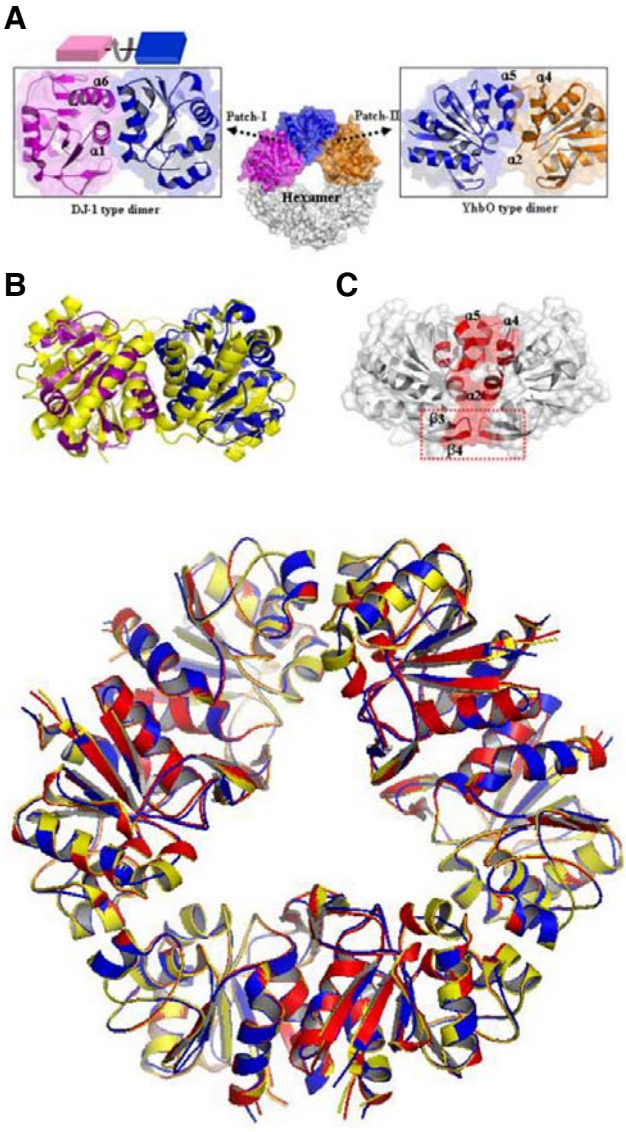

Fig. 2. Quaternary structures of PH1704 and DR1199. (A) PH1704 forms a hexamer, which requires both the DJ-1 type (patch I) and YhbO type (patch II) dimer interfaces. The subunit of the DJ-1 type dimer is represented by two rotated bricks in magenta and blue. (B) Superposition of the PH1704 dimer (patch I type, purple and blue) with SCRP-27A (yellow) shows similar binding orientations. (C) The diagram of the YhbO type dimer through patch II in DR1199. The interface region including the extended loops (dotted red box) is shown in red.

Fig. 3. Comparison of TON1285 proteins with native, N-terminal, and C-terminal tags. Superposition of the N-terminal histidine tagged (yellow), C-terminal histidine tagged (blue), and native TON1285 (red) yielded a root mean squared deviation of 0.5 Å and 0.24 Å to native TON1285, respectively.

terminal 21 residues, the region from β4 to β5, and the loops between β6 and α3 (Fig. 1E). YDR533Cp is similar to Hsp31 but does not have the long N-terminal extension found in Hsp31. The dimeric interface of YDR533Cp is also completely distinct and named patch IV (β7 and the region from β8 to β9; Fig. 1F) (Graille et al., 2004; Wilson et al., 2004). Hereafter, dimeric contact via patches III and IV are called the Hsp type and the YDR type, respectively.

Dimer-based higher order oligomers in the DJ-1 superfamily

Except for patch III which includes a long N-terminal extension, patches I, II, and IV all consist of only the core structure of the DJ-1 superfamily, suggesting the possibility of forming higher order oligomers using a combination of several patch types. It has been reported that for proteins with the YhbO type, dimeric contact can exist as a hexamer in solution (Wei et al., 2007). For example, PH1704 forms a ring-like hexamer by trimerization of the YhbO type dimer (Du et al., 2000). The trimeric interface between the YhbO type dimers is formed by patch I (Fig. 2A). Interestingly, the contact between two PH1704 patch I's is highly similar to the rotated contact mode observed in SCRP-27A described in the previous section (Fig. 2B). This rotated contact allows the completion of the hexameric ring-like architecture of PH1704. Because only the hexameric form exhibits proteolytic activity, the rotated contact between two patch I's can be assumed to be a strategic tool for attaining functional higher order oligomerization. In contrast to PH1704, DR1199 forms only a dimer. Normally, the loop between β3 and β4 (the red box in Fig. 2C) is a part of patch I in the YhbO type dimeric interface. However, the extended version of this loop in DR1199 protrudes towards patch II instead of forming patch I. The disturbed patch I in DR1199 may not allow the rotated conformation required for hexameric ring closure.

YhbO was reported to form a hexamer in solution (Abdallah et al., 2006). However, recombinant YhbO with an additional N-terminal histidine tag shows only the dimeric form of the YhbO type (PDB code:1OI4; Fig. 1D). The histidine tag in the N-terminal extension is ordered in the patch I region, significantly altering native YhbO patch I. Thus, the native YhbO protein

Table 1. Structure-based classification of the DJ-1 superfamily

Sequence group[1]	Type[2]	Interface[3]	Quaternary structure	Catalytic residue	3D structures (PDB ID)	Function	Taxonomic distribution
I	DJ-1	patch I	Dimer	Diad	DJ-1 (1J42, 1P5F, 1PDV, 1PDW, 1PS4, 1Q2U, 1SOA, 1UCF, 2OR3, 2R1U)	Chaperone activity; regulator of the PI3K/Akt survival signaling pathway and the Daxx-ASK1 proapoptotic pathway	Eukaryotes
					YajL (2AB0)	Unknown	Bacteria
					ThiJ/PfpI family protein (3EFE), Atu0886 (2FEX), YdeA (3F5D)	Unknown	Bacteria
					ThiJ/PfpI family protein (3EWN), YP001094981.1 (3BHN)	Unknown	Bacteria Archaea
I′		patch I	Dimer	Diad	SCRP-27A (1VHQ)[5], ElbB (1OY1)[5]	Unknown	Bacteria
II	YhbO	patch II, I	Hexamer	Triad	PH1704 (1G2I), TON1285 (3L18)	Proteolytic activity	Bacteria Archaea
		patch II[4]	Dimer[4] (Hexamer)		YhbO (1OI4)	Involved in hyperosmotic, oxidative, thermal, UV, and pH stresses	Bacteria Archaea
		patch II	Dimer		DR1199 (2VRN)	Involved in hyperosmotic, oxidative, thermal, UV, and pH stresses	Bacteria
III	Hsp	patch III	Dimer	Triad	Hsp31 (1IZY, 1IZZ, 1N57, 1ONS, 1PV2)	Chaperone activity; proteolytic activity	Eukaryotes Bacteria
IV	YDR	patch IV	Dimer	Triad	YDR533Cp[6] (1QVV, 1RW7)	Protection against oxidative stress	Eukaryotes Bacteria

[1]Group by whole sequence homology
[2]Interface type by binding orientation
[3]Dimer interface region
[4]Only the dimer form is available in PDB
[5]Identical proteins with different names
[6]Dimer interfaces are different between 1QVW and 1RW7, all of which are distinct from I-III. 1QVW is taken as a representative.

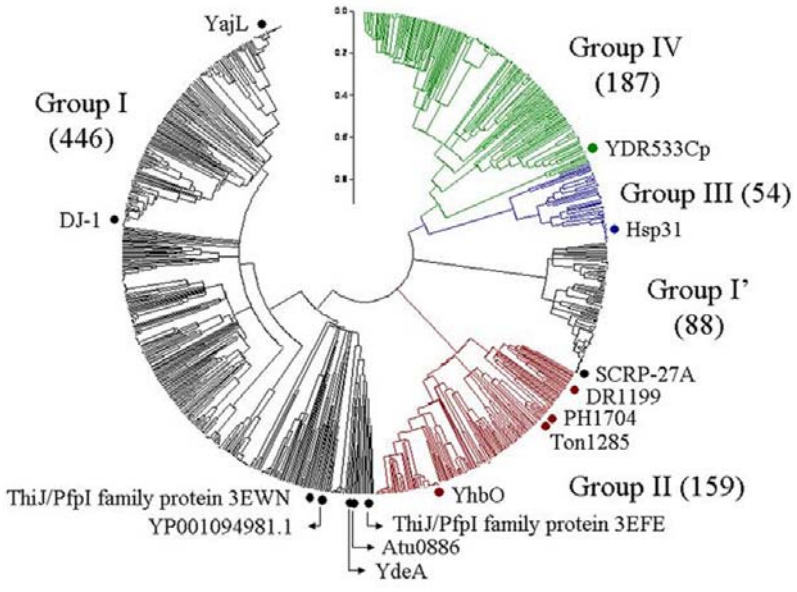

Fig. 4. The phylogenetic tree of the 934 DJ-1/ThiJ/PfpI superfamily members, which were classified into 5 distinct groups (I-IV and I′) based on sequence similarity. The number of proteins in each group is shown in parentheses. The proteins of known 3D structure are indicated using dots in black (patch I), brown (patch II), blue (patch III), and green (patch IV), according to the interface patch type. The sequence similarity groups are shown to agree well with the patch types in the proteins of known 3D structure.

without the artificial N-terminal extension is likely to form a hexamer just like PH1704. Consequently, the dimeric form is most likely the basic building block for the construction of more complex oligomers e.g. hexamer in the DJ-1 superfamily.

To investigate whether an artificial N- or C-terminal extension affects the oligomerization state, we determined the crystal structures of TON1285 from *T. onnurineus* NA1 (Lee et al., 2003). TON1285 is a DJ-1 superfamily member with 89% sequence identity to PH1704. For the test, we prepared three variants of TON1285: wild type protein, an N-terminal extended

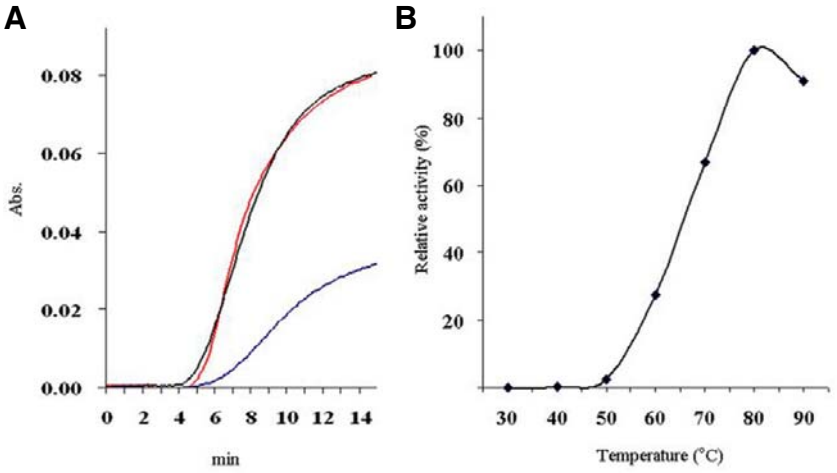

Fig. 5. Chaperone and proteolytic activity of YdeA and TON1285. (A) The measurements of citrate synthase aggregation in the presence of YdeA (red), DJ-1 (blue), and control (no protein, black) to citrate synthase. Although DJ-1 and YdeA both belong to group I, DJ-1 suppressed the aggregation of citrate synthase while YdeA did not. (B) Proteolytic activity of TON1285 from 303 K to 363 K. The activity was detected by cleavage of 7-amido-4-methyl-cou-marin (MCA)-linked peptides Ala-Ala-Phe.

protein, and a C-terminal extended protein (see "Materials and Methods"). In contrast to YhbO, the artificial N-terminal region of TON1285 is not ordered, and the three TON1285 variants show virtually the same hexameric conformation (Fig. 3). It suggests that the N- or C-terminal extension has no influence on the oligomerization state. It should be noted that in general, an N- or C-terminal extension such as the hexa-histidine tag is highly flexible without extensive interactions with the main body of the protein as in TON1285.

The sequence similarity group agrees well with the quaternary structural class but can be heterogeneous in terms of molecular function

We divided the 934 DJ-1 superfamily members into five groups (I-IV and I') based on sequence similarity and conserved active sites (Table 1, Fig. 4). All the members are characterized by a sharp turn between a β-strand and an α-helix, called the "nucleophile elbow," carrying an important cysteine residue (Ollis et al., 1992). The cysteine residue forms either a catalytic triad of Cys-His-Glu/Asp (groups II-IV) or a Cys-Glu/Asp diad (groups I and I'). Although SCRP-27A and its close homologs form a distinct clade in the phylogenetic tree, they share the same diads and interface patch type (patch I) like group I and thus are designated as group I'.

The sequence similarity groups show good parallelism with the quaternary structures or interface patch types (Fig. 4). Group I (446 proteins) included all the known cases of the DJ-1 type (patch I) except for SCRP-27A in group I'. Group I', including SCRP-27A (88 proteins), may have a rotated patch I form, as described in the previous section. All YhbO type dimers belong to group II (159 proteins), including a previously unknown member, TON1285, for which we experimentally validated its quaternary structure (PDB code: 3L18). Two more proteins of distinct quaternary structure, were found, Hsp31 and YDR533Cp. These two proteins also belong to separate groups (III and IV) in accordance with different patch types. Other protein members generally show conserved catalytic triads in both groups III (54 proteins) and IV (187 proteins). At least for groups I and II, multiple proteins are available with reasonable sequence diversity, as well as known 3D structures. Therefore, it is likely that most of the other proteins in groups I and II have the same quaternary structure. That is, sequence and quaternary structure may have diverged largely in parallel in the DJ-1 superfamily.

We further asked whether molecular function is also strongly related to either sequence similarity or quaternary structure. The DJ-1 group has at least one member protein, DJ-1, with known chaperone activity (Lee et al., 2003). We examined chaperone activity for *B. subtilis* YdeA in the same group of DJ-1. YdeA showed no chaperone activity, while a more distant member, HSP31, is known to function as a chaperone (Fig. 5A). Protease activity is detected in TON1285, which could be expected due to the high sequence similarity to PH1704 (Fig. 5B). However, DJ-1 was recently reported to have protease activity with the deletion of 15 C-terminal residues in spite of differences in terms of both sequence and patch type (Chen et al., 2010). Therefore, molecular function may have diverged independently from primary and quaternary structure and thus would be difficult to infer simply using sequence information.

Evolution of quaternary structures in the DJ-1 superfamily

As described above, the DJ-1 superfamily has evolved diverse types of quaternary structures with distinct interface patches. It is intriguing to ask how and when these binding modes appeared. Was each interface patch evolved only once at a certain period and then inherited thereafter? Alternatively, each specific binding mode may have been introduced sporadically to multiple taxonomic lineages at different times either through evolution or horizontal gene transfer (HGT). With only a dozen known 3D structures available, it is very challenging to examine these questions unambiguously. However, the binding modes in the DJ-1 superfamily seem to agree well with the groups based on primary sequence (Fig. 4). Therefore, we examined the taxonomic distributions of proteins in each group (I', I-IV), which is likely to reflect binding mode or patch type as well.

In total, the 934 proteins were assigned to 573 species (365 unique species) in the groups I' and I-IV after removing redundancy and filtering ambiguously assigned species. The member proteins in each group are not limited to a single kingdom but are distributed in multiple kingdoms except for group I' (Fig. 6). Group I appears in all three kingdoms, whereas group II appears widely in Archaea and Bacteria but not in Eukaryota. Groups III and IV show similar phyletic patterns; members are found in several phyla of Bacteria and Eukaryota but not in Archaea. Group I' is observed only in a single phylum of proteobacteria, suggesting a highly lineage-specific mode of quaternary structure.

These phyletic patterns are not likely to result purely by inheritance because for example, it contradicts that group II is absent only in a more recent kingdom, eukaryota. We reason

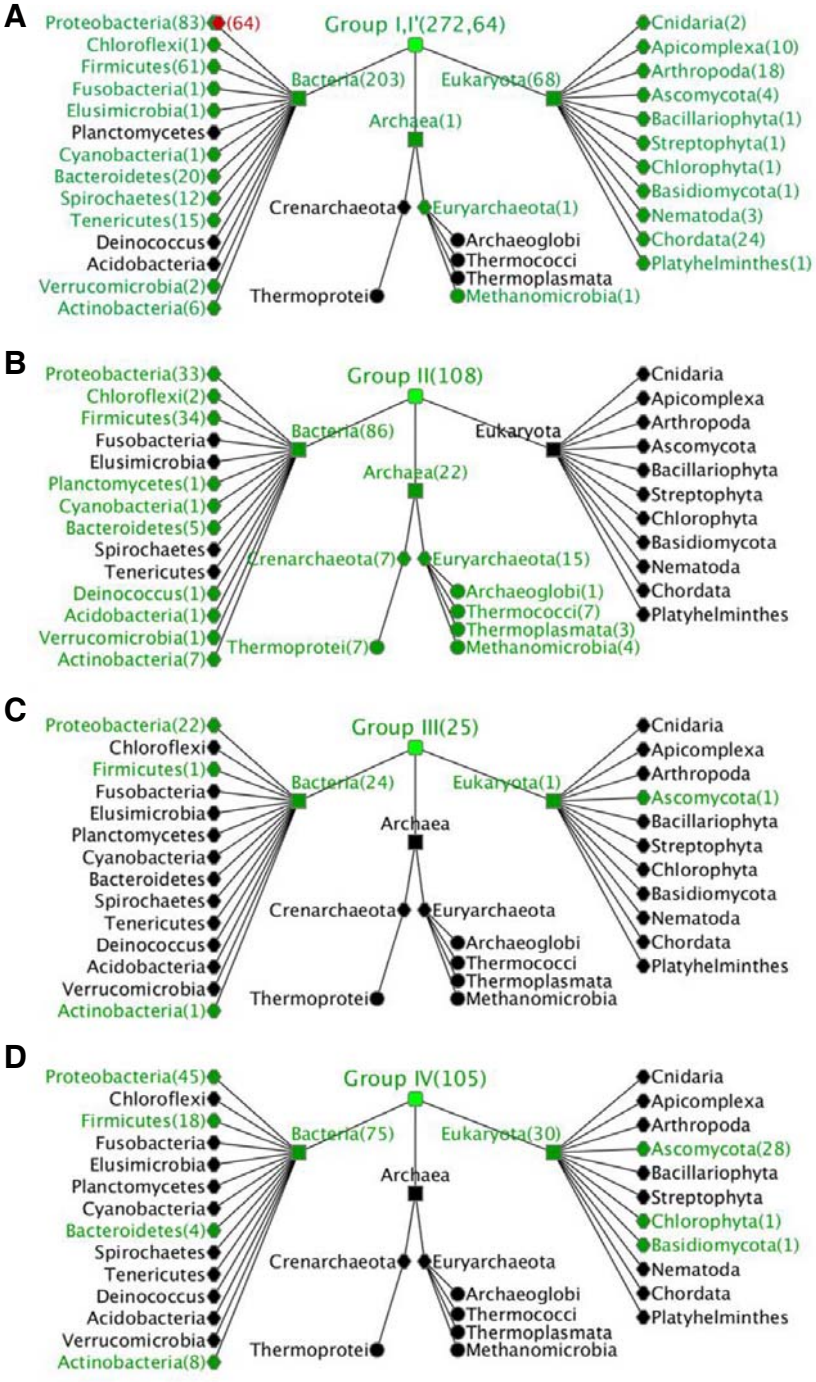

Fig. 6. The taxonomic distribution of groups I' and I-IV in the DJ-1 superfamily. Node color indicates the existence (green) or absence (black) of protein members in each taxon, where the number of unique species is shown in parentheses. Node shape indicates taxonomic ranks such as kingdom (square), phylum (hexagon), and class (circle). (A) Group I (274 species, green) appear in all three kingdoms, but group I' is found only in proteobacteria (64 species, red). (B) Group II (108 species) is distributed in Archaea and Bacteria. (C, D) Groups III (25 species) and IV (105 species) are distributed similarly in Bacteria and Eukaryota.

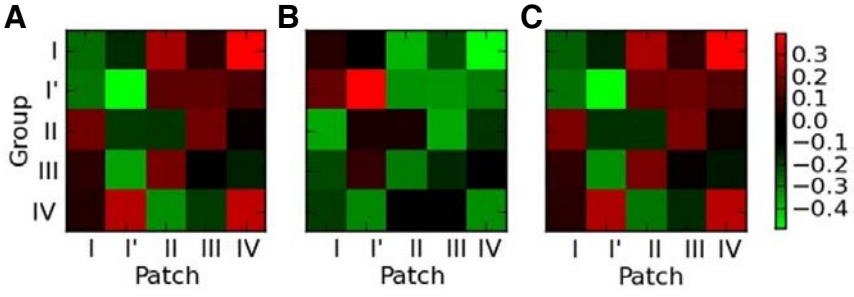

Fig. 7. The mutation rates of interfaces and non-interfacial surfaces. The average mutation rates of (A) interfacial residues (R_I), (B) non-interfacial surfaces (R_S), and (C) the difference in mutation rate between interfaces and non-interfacial surfaces ($R_{I\text{-}S}$). The mutation rates were calculated using rage4site and normalized as a standard Z-value. In (C), the diagonals of groups I, I', and II are negative (green), suggesting relative rate of mutation is lower for the interfaces than the non-interfaces in the DJ-1 family.

that re-invention may have played a more important role than horizontal gene transfer (HGT) based on several observations. HGT is a relatively rare event, particularly in eukaryotes, but many cases exist in which multiple groups are observed in a single phylum (e.g. groups I, III-IV in ascomycota and all five groups in proteobacteria and actinobacteria, etc). If HGT played a major role for the diverse quaternary structures observed in a single taxonomic lineage, it requires frequent HGT events involving eukaryota. Furthermore, certain groups were totally absent in certain taxonomic lineages. For example, group II is absent in Eukaryota, and groups III and IV are not observed in Archaea. It is very unlikely that HGT occurred highly selectively in specific kingdom and, at the same time, very frequently across many taxa. This suggests that the binding modes were dynamically re-invented in most cases and the observed phyletic patterns mainly resulted from combinations of occasional re-invention and inheritance.

The mutation rates are slower at interfaces than non-interfacial surfaces in the DJ-1 family

If dimerization is important for protein function, interfacial residues would be more conserved than other surfaces. We measured the mutation rates of patches I' and I-IV for all the groups tested using rate4site (Pupko et al., 2002). In each group, the patches were defined as the columns in the aligned sequences, where more than half of the residues in each column belong to the corresponding patch among the members of known 3D structures.

For the groups I, I', and II, it is evident that the interfacial patches show lower rates of mutation for the corresponding groups (Fig. 7). For groups III and IV, the member proteins of these groups highly diverged, and it was difficult to align equivalent residues accurately. The aligned interfacial residues were not spatially equivalent in these groups (Supplementary Fig. S1), and accordingly, the same trend was not observed. Nevertheless, the interfacial residues were shown as more conserved at least for the reasonably well-aligned groups. It suggests that invention of a new interface affects the rate of mutation as well as its oligomerization mode.

In general, the groups based on sequence similarity are shown to be consistent with the quaternary structural classes, but their biochemical function could not be directly inferred using sequence information alone. The observed phyletic pattern suggests the dynamic nature of quaternary structures in the course of evolution. The amino acid residues at interfaces tend to show lower rates of mutation than non-interfacial surfaces.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

S.-S. Cha was supported by the Marine and Extreme Genome Research Center program of the Ministry of Land, Transport and Maritime Affairs, the KORDI in-house program (PE98513), and the Development of Biohydrogen Production Technology Using Hyperthermophilic Archaea program of Ministry of Land, Transport and Maritime Affairs and the National Research Foundation of Korea Grant 2009-0084757. W. Kim was supported by the Ewha Global Top 5 Grant 2011 of Ewha Womans University, by the Global Frontier Project grant (NRF-M1AXA-002-2010-0029763), and by the Basic Science Research Program (2011-0014992) of the National Research Foundation funded by the Ministry of Education, Science and Technology of Korea.

REFERENCES

Abdallah, J., Kern, R., Malki, A., Eckey, V., and Richarme, G. (2006). Cloning, expression, and purification of the general stress protein YhbO from *Escherichia coli*. Protein Expr. Purif. *47*, 455-460.

Abdallah, J., Caldas, T., Kthiri, F., Kern, R., and Richarme, G. (2007). YhbO protects cells against multiple stresses. J. Bacteriol. *189*, 9140-9144.

Bandyopadhyay, S., and Cookson, M.R. (2004). Evolutionary and functional relationships within the DJ1 superfamily. BMC Evol. Biol. *4*, 6.

Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C., Squitieri, F., Ibanez, P., Joosse, M., et al. (2003). Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science *299*, 256-259.

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). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905-921.

Canet-Aviles, R.M., Wilson, M.A., Miller, D.W., Ahmad, R., McLendon, C., Bandyopadhyay, S., Baptista, M.J., Ringe, D., Petsko, G.A., and Cookson, M.R. (2004). The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. Proc. Natl. Acad. Sci. USA *101*, 9103-9108.

Cha, S.S., Jung, H.I., Jeon, H., An, Y.J., Kim, I.K., Yun, S., Ahn, H.J., Chung, K.C., Lee, S.H., Suh, P.G., et al. (2008). Crystal structure of filamentous aggregates of human DJ-1 formed in an inorganic phosphate-dependent manner. J. Biol. Chem. *283*, 34069-34075.

Chen, J., Li, L., and Chin, L.S. (2010). Parkinson disease protein DJ-1 converts from a zymogen to a protease by carboxyl-terminal cleavage. Hum. Mol. Genet. *19*, 2395-2408.

Du, X., Choi, I.G., Kim, R., Wang, W., Jancarik, J., Yokota, H., and Kim, S.H. (2000). Crystal structure of an intracellular protease from Pyrococcus horikoshii at 2-A resolution. Proc. Natl. Acad. Sci. USA *97*, 14079-14084.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of coot. Acta Crystallogr. D. Biol. Crystallogr. *66*, 486-501.

Fioravanti, E., Dura, M.A., Lascoux, D., Micossi, E., Franzetti, B., and McSweeney, S. (2008). Structure of the stress response protein DR1199 from *Deinococcus radiodurans*: a member of the DJ-1 superfamily. Biochemistry *47*, 11581-11589.

Gorner, K., Holtorf, E., Waak, J., Pham, T.T., Vogt-Weisenhorn, D.M., Wurst, W., Haass, C., and Kahle, P.J. (2007). Structural determinants of the C-terminal helix-kink-helix motif essential for protein stability and survival promoting activity of DJ-1. J. Biol. Chem. *282*, 13680-13691.

Graille, M., Quevillon-Cheruel, S., Leulliot, N., Zhou, C.Z., de la Sierra Gallay, I.L., Jacquamet, L., Ferrer, J.L., Liger, D., Poupon, A., Janin, J., et al. (2004). Crystal structure of the YDR533c S. cerevisiae protein, a class II member of the Hsp31 family. Structure *12*, 839-847.

Halio, S.B., Bauer, M.W., Mukund, S., Adams, M., and Kelly, R.M. (1997). Purification and characterization of two functional forms of intracellular protease PfpI from the hyperthermophilic archaeon pyrococcus furiosus. Appl. Environ. Microbiol. *63*, 289-295.

Horvath, M.M., and Grishin, N.V. (2001). The C-terminal domain of HPII catalase is a member of the type I glutamine amidotransferase superfamily. Proteins *42*, 230-236.

Junn, E., Taniguchi, H., Jeong, B.S., Zhao, X., Ichijo, H., and Mouradian, M.M. (2005). Interaction of DJ-1 with Daxx inhibits apoptosis signal-regulating kinase 1 activity and cell death. Proc. Natl. Acad. Sci. USA *102*, 9691-9696.

Kim, R.H., Peters, M., Jang, Y., Shi, W., Pintilie, M., Fletcher, G.C., DeLuca, C., Liepa, J., Zhou, L., Snow, B., et al. (2005). DJ-1, a novel regulator of the tumor suppressor PTEN. Cancer Cell *7*, 263-273.

Kim, W.K., Henschel, A., Winter, C., and Schroeder, M. (2006). The many faces of protein-protein interactions: a compendium of interface geometry. PLoS Comput. Biol. *2*, e124.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0.

Bioinformatics *23*, 2947-2948.
Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477-486.
Lee, S.J., Kim, S.J., Kim, I.K., Ko, J., Jeong, C.S., Kim, G.H., Park, C., Kang, S.O., Suh, P.G., Lee, H.S., et al. (2003). Crystal structures of human DJ-1 and *Escherichia coli* Hsp31, which share an evolutionarily conserved domain. J. Biol. Chem. *278*, 44552-44559.
Li, Y.H., Wang, Y.Y., Zhong, S., Rong, Z.L., Ren, Y.M., Li, Z.Y., Zhang, S.P., Chang, Z.J., and Liu, L. (2009). Transmembrane helix of novel oncogene with kinase-domain (NOK) influences its oligomerization and limits the activation of RAS/MAPK signaling. Mol. Cells *27*, 39-45.
Lucas, J.I., and Marin, I. (2007). A new evolutionary paradigm for the Parkinson disease gene DJ-1. Mol. Biol. Evol. *24*, 551-561.
Mizote, T., Tsuda, M., Smith, D.D., Nakayama, H., and Nakazawa, T. (1999). Cloning and characterization of the thiD/J gene of *Escherichia coli* encoding a thiamin-synthesizing bifunctional enzyme, hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase. Microbiology *145*, 495-501.
Moore, D.J., Zhang, L., Troncoso, J., Lee, M.K., Hattori, N., Mizuno, Y., Dawson, T.M., and Dawson, V.L. (2005). Association of DJ-1 and parkin mediated by pathogenic DJ-1 mutations and oxidative stress. Hum. Mol. Genet. *14*, 71-84.
Mujacic, M., and Baneyx, F. (2007). Chaperone Hsp31 contributes to acid resistance in stationary-phase *Escherichia coli*. Appl. Environ. Microbiol. *73*, 1014-1018.
Neumann, M., Muller, V., Gorner, K., Kretzschmar, H.A., Haass, C., and Kahle, P.J. (2004). Pathological properties of the Parkinson's disease-associated protein DJ-1 in alpha-synucleinopathies and tauopathies: relevance for multiple system atrophy and Pick's disease. Acta Neuropathol. *107*, 489-496.
Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205-217.
Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., et al. (1992). The alpha/beta hydrolase fold. Protein Eng. *5*, 197-211.
Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol. *276*, 307-326.
Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., and Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics *18*, S71-77.
Quigley, P.M., Korotkov, K., Baneyx, F., and Hol, W.G. (2003). The 1.6-A crystal structure of the class of chaperones represented by *Escherichia coli* Hsp31 reveals a putative catalytic triad. Proc. Natl. Acad. Sci. USA *100*, 3137-3142.
Sastry, M.S., Korotkov, K., Brodsky, Y., and Baneyx, F. (2002). Hsp31, the *Escherichia coli* yedU gene product, is a molecular chaperone whose activity is inhibited by ATP at high temperatures. J. Biol. Chem. *277*, 46026-46034.
Shendelman, S., Jonason, A., Martinat, C., Leete, T., and Abeliovich, A. (2004). DJ-1 is a redox-dependent molecular chaperone that inhibits alpha-synuclein aggregate formation. PLoS Biol. *2*, e362.
Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. *24*, 1596-1599.
Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. *66*, 22-25.
Wei, Y., Ringe, D., Wilson, M.A., and Ondrechen, M.J. (2007). Identification of functional subclasses in the DJ-1 superfamily proteins. PLoS Comput. Biol. *3*, e10.
Wilson, M.A., St Amour, C.V., Collins, J.L., Ringe, D., and Petsko, G.A. (2004). The 1.8-A resolution crystal structure of YDR533Cp from Saccharomyces cerevisiae: a member of the DJ-1/ThiJ/Pfpl superfamily. Proc. Natl. Acad. Sci. USA *101*, 1531-1536.
Wilson, M.A., Ringe, D., and Petsko, G.A. (2005). The atomic resolution crystal structure of the YajL (ThiJ) protein from *Escherichia coli*: a close prokaryotic homologue of the Parkinsonism-associated protein DJ-1. J. Mol. Biol. *353*, 678-691.
Winter, C., Henschel, A., Kim, W.K., and Schroeder, M. (2006). SCOPPI: a structural classification of protein-protein interfaces. Nucleic Acids Res. *34*, D310-314.
Witt, A.C., Lakshminarasimhan, M., Remington, B.C., Hasim, S., Pozharski, E., and Wilson, M.A. (2008). Cysteine pKa depression by a protonated glutamic acid in human DJ-1. Biochemistry *47*, 7430-7440.
Yang, Y., Gehrke, S., Haque, M.E., Imai, Y., Kosek, J., Yang, L., Beal, M.F., Nishimura, I., Wakamatsu, K., Ito, S., et al. (2005). Inactivation of Drosophila DJ-1 leads to impairments of oxidative stress response and phosphatidylinositol 3-kinase/Akt signaling. Proc. Natl. Acad. Sci. USA *102*, 13670-13675.