Structural analysis of missense mutations occurring in the DNA-binding domain of HSF4 associated with congenital cataracts

Zaiyu Xiao¹, Ling Guo²,¹, Yang Zhang¹, Liwei Cui², Yujie Dai², Zhu Lan², Qinghua Zhang³,⁴, Sheng Wang³,⁵, Wei Liu²,³

¹ College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
² Institute of Immunology, Army Military Medical University of PLA, Chongqing 400038, China
³ Department of Obstetrics and Gynecology, Daping Hospital, Army Medical University of PLA, Chongqing 400038, China

ARTICLE INFO
Keywords:
Congenital cataract
Heat shock factor 4
DNA-binding domain
Molecular dynamics simulation

ABSTRACT
Congenital cataract (CC) is the major cause of childish blindness, and nearly 50% of CCs are hereditary disorders. HSF4, a member of the heat shock transcription factor family, acts as a key regulator of cell growth and differentiation during the development of sensory organs. Missense mutations in the HSF4-encoding gene have been reported to cause CC formation; in particular, those occurring within the DNA-binding domain (DBD) are usually autosomal dominant mutations. To address how the identified mutations lead to HSF4 malfunction by placing adverse impacts on protein structure and DNA-binding specificity and affinity, we determined two high-resolution structures of the wild-type DBD and the K23N mutant of human HSF4, built DNA-binding models, conducted in silico mutations and molecular dynamics simulations. Our analysis suggests four possible structural mechanisms underlining the missense mutations in HSF4-DBD and cataractogenesis: (i), disruption of HSE recognition; (ii), perturbation of protein-DNA interactions; (iii), alteration of protein folding; (iv), other impacts, e.g. inhibition of protein oligomerization.

1. Introduction
Cataract is an eye disease arising from the loss of ocular lens transparency and can be classified as congenital and age-related according to the disorder onset. Congenital cataract (CCs) is a lens opacity present at birth and form the leading cause of childish blindness (Lund et al., 1992). Approximately 50% of CC cases may have a genetic background (Francis and Moore, 2004). It has been known that mutations in the genes encoding various crystallin proteins, cytoskeletal proteins, gap junction proteins or transcription factors, e.g. FOXE3, HSF4, MAF, and PITX3, may lead to CC formation (Anand et al., 2018; Hejtmancik, 2008; Shiels and Hejtmancik, 2017). According to recent sequencing and phenotyping data, the inheritance of CC may be familial and usually autosomal dominant (Anand et al., 2018; Berry et al., 2018).

HSF4 belongs to the heat shock factor (HSF) family conserved from fungi to humans. The members of this family serve as central regulators in maintaining cellular protein homeostasis or mediating cell differentiation and development (Akerfelt et al., 2010; D. Westerheide et al., 2012; Vihervaara and Sistonen, 2014). Malfunction of HSFs is often linked to severe human diseases such as cancer and neurodegenerative disorders (Dai et al., 2007; Neef et al., 2011; Scherz-Shouval et al., 2014). The human genome encodes six HSF proteins, among which HSF1, 2 and 4 have been extensively studied during past decades, but physiological functions of the others remain undetermined (Gomez-Pastor et al., 2018). HSF1 is the master regulator of the heat shock response against proteotoxic stress conditions, while HSF2 is more involved in specific developmental processes such as corticogenesis and spermatogenesis (Akerfelt et al., 2010; Vihervaara and Sistonen, 2014; Widlak and Vydra, 2017). HSF4 is not implicated in heat shock response but required for cell growth and differentiation during the development of sensory organs, e.g. eye lens, in cooperation with HSF1 (Fujimoto et al., 2008, 2004). There are two HSF4 isoforms, HSF4a and HSF4b, resulting from alternative splicing. HSF4a is the canonical isoform that contains an additional 30 amino acids and can activate transcription, while HSF4a inhibits the expression of other HSFs (Bjork and Sistonen, 2010; Pirkkala et al., 2001).

HSFs share similar module organizations in their sequences, which

Abbreviations: CC, congenital cataract; HSF, heat shock factor; DBD, DNA-binding domain; HSE, heat shock element; MD, molecular dynamics
* Corresponding authors.
E-mail addresses: zhangql1123@163.com (Q. Zhang), shengwang@hust.edu.cn (S. Wang), wei.liu.2005@gmail.com (W. Liu).
These authors contributed equally to this work.
https://doi.org/10.1016/j.jsbxb.2019.100015
Received 11 August 2019; Received in revised form 12 November 2019; Accepted 13 November 2019
Available online 15 November 2019
2590-1524/ © 2019 Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).
usually comprise an N-terminal DNA binding domain (DBD), an adjacent oligomerization domain also referred to as the hydrophobic repeat region (HR-A and -B), an intrinsically disordered regulatory domain and a C-terminal activation domain. DBD is the most conserved peptide region (HR-A and -B), an intrinsically disordered regulatory domain. Usually comprise an N-terminal DNA binding domain (DBD), an activation domain and a C-terminal activation domain. DBD is characterized by a winged helix-turn-helix motif in its structural core (Feng et al., 2016b; Harrison et al., 1994; Vuister et al., 1994). When binding to a target gene, the second helix in this motif is inserted into the major groove of DNA and specifically contacts the GAA triplet present in an HSE (Jaeger et al., 2016; Littlefield and Nelson, 1999; Neudegger et al., 2016). Upon activation, HSF1 and HSF2 are converted into a DNA-binding component through homotrimerization or heterotrimerization. HSF4, however, is constitutively trimerized and thus has constant DNA-binding activity (Akerfelt et al., 2010; Gomez-Pastor et al., 2018). Each DBD in an HSF trimer binds to a single nGAAn repeat, and therefore three repeats are required for optimal binding (Gonsalves et al., 2011; Periscic et al., 1989; Xiao et al., 1991).

Despite being relatively less studied, HSF4 has arisen researchers’ extensive interest since a link of severe CC occurrence in Chinese and Danish families with 4 missense mutations in HSF4-DBD was reported (Bu et al., 2002). The correlation of this transcription factor with CC formation was further demonstrated in HSF4-deficient mice, which displayed inclusion-like structures in lens fiber cells soon after birth (Fujimoto et al., 2004; Min et al., 2004). After that, a number of mutations were identified in different ethnicities for both autosomal dominant and recessive congenital cataract (Table 1) (Anand et al., 2018; Berry et al., 2018; Enoki et al., 2010; Gillespie et al., 2014; Hejtmancik, 2008; Jing et al., 2014; Shi et al., 2008). Interestingly, all known autosomal dominant mutations lie within DBD, while recessive mutations are located outside (Anand et al., 2018; Berry et al., 2018). Despite the increasing number of identified CC-associated mutations, mechanistic understanding of how these mutations induce HSF4 malfunction is limited due to the insufficient biochemical/structural studies revealing clear pathways from genotypes to phenotypes.

In this study, we attempted to explore conformational changes in HSF4-DBD and protein-DNA interactions induced by the CC-associated mutations. To this end, we determined the crystal structures of the wild type and a mutant DBD at atomic resolutions, built a DNA-binding model, conducted in silico mutations and molecular dynamics (MD) simulations. Our results suggest diverse mechanisms underlying the identified mutations in HSF4.

2. Materials and methods

2.1. Protein expression

The nucleotide sequence encoding the DNA-binding domain of human HSF4 (residues 17–122) were amplified by PCR and inserted into a pET-22b (+) plasmid (Novagen), which was subsequently transformed into E. coli host strain BL21(DE3). Bacteria were grown in LB medium containing 100 μg ml⁻¹ ampicillin at 310 K. Expression of the recombinant protein was induced with 0.3 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 4 h at 303 K.

2.2. Protein purification

Protein purification was conducted following a previously established protocol (Feng et al., 2016a). In short, the harvested bacteria were lysed using a high-pressure crusher at 277 K in the lysis buffer containing 50 mM NaH₂PO₄/Na₂HPO₄, pH 8.0, 500 mM NaCl and 25 mM imidazole. After removal of insoluble debris, the supernatant was loaded onto a HisTrap HP 5 ml column (GE Healthcare), which was eluted with 300 mM imidazole. The pooled fractions were then desalted and applied onto a HiTrap SP 5 ml column (GE Healthcare) pre-equilibrated with 20 mM HEPES pH 7.5, 100 mM NaCl. A linear-gradient elution with increasing NaCl concentration from 100 mM to 1.0 M was developed for protein elution. A step of size-exclusion chromatography was performed afterward to further improve the protein purification using a HiLoad 16/600 Superdex 75 column (GE Healthcare) and an elution buffer consisting of 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT, 0.2 mM EDTA and 5% glycerol.

The expression plasmid of the K23N mutant was built following the instructions of the Muta-direct™ Kit (SBS Genetech) using the pET22b- HSF4-DBD plasmid as the template. Subsequent expression and purification processes were same as those for the wild-type protein. The purified proteins were concentrated to 60 mg ml⁻¹ and stored at 193 K until being used for crystallization.

2.3. Protein crystallization

The protein samples of wild-type DBD and the K23N mutant were diluted to 30 mg ml⁻¹ before crystallization trials. Screening for initial crystallization conditions was carried out by means of sitting-drop vapor-diffusion using five commercial kits from Hampton Research (California, USA) and a Gryphon-LCP robot (Art Robbins Instruments, USA). Drops were set up by mixing 0.4 µl protein solution and 0.4 µl reservoir solution equilibrating with 40 µl reservoir solution in the well. The subsequent optimizations were manually conducted using the hanging-drop vapor diffusion method with drops consisting of 1 µl protein and 1 µl reservoir solution. Crystals of both the wild-type and mutant proteins were grown at 0.1 M Tris, pH 8.0 and 3.4 M sodium nitrate.

2.4. Diffraction data collection and processing

The crystals were directly mounted in nylon cryoloops (Hampton Research) and flash-cooled in a stream of liquid nitrogen. X-ray Table 1

| Nucleotide Position | AA Position | Secondary Structure | Conservation | Cataract Class | Possible Structural Impact |
|---------------------|-------------|---------------------|--------------|----------------|---------------------------|
| c. 57C > A          | A19D        | Helix α1            | High         | Lamellar       | NO                        |
| c. 69G > T          | K23N        | Helix α1            | Very high    | Gerullean      | DNA-binding perturbation  |
| c. 103C > T         | H35Y        | Loop                | Very low     | Congenital     | NO                        |
| c. 179C > A         | P60H        | Helix α2            | High         | Nuclear        | Protein misfolding        |
| c. 182A > G         | Q61R        | Helix α2            | Very low     | Senile         | Wing movement             |
| c. 190A > G         | K64E        | Helix α2            | Very high    | Lamellar       | DNA-binding perturbation  |
| c. 218G > A         | R73H        | Helix α3            | Very high    | Congenital     | HSE-recognition disruption|
| c. 256A > G         | I86V        | Wing                | Very low     | Senile         | NO                        |
| c. 341 T > C        | L114P       | Helix α2            | High         | Lamellar       | Protein misfolding        |
| c. 355G > A         | R116H       | Helix α2            | Very low     | Lamellar       | C-terminus movement       |
| c. 362C > T         | R119C       | C-terminal loop     | Very high    | Lamellar       | HSE-recognition disruption|

Notes: AA = amino acid; NO = not obvious.
2.5. Structure determination and refinement

The structure of wild-type HSF4-DBD was determined by means of molecular replacement using the recently solved structure of human HSF1-DBD (PDB entry 5HDG) (Feng et al., 2016b) as a search model. After the automatic model building using Phenix.autobuild (Terwilliger et al., 2008), the structure was refined using Phenix.refine (Afonine et al., 2012) with several rounds of manual remodeling between refinement cycles using the modeling toolkit Coot (Emsley et al., 2010). Structure of the K23N mutant was solved by molecular replacement as well using the wild-type structure as a search model, and refined in the same way. Statistics of data collection and structure refinement are summarized in Table 2. All structural representations were generated using the molecular visualization program PyMOL (Schrodinger LLC, 2015).

2.6. Modeling of HSF4-DBD bound to DNA and in silico mutation

To generate a model of HSF4-DBD bound to DNA similar to the recently reported structures of HSF1/2-DBDs in complex with DNA containing two binding sites (Jaeger et al., 2014; Neudegger et al., 2016), PDB entry 5DHG was used as the template. The two HSF2-DBDs existing in that structure were replaced with the refined structure of HSF4-DBD through superimposition, and thus a model of two DBD copies bound to a palindromic 12-bp DNA duplex containing TCTGAAA at the center was generated. After auto-building of the missed wing using homology modeling, energy minimization, and short-time MD simulation were performed to exclude steric conflicts. The DNA molecule in this model was then replaced by a 24-bp B-form ds-DNA comprising the above palindromic sequence at the middle region generated using Nucgen from the Amber14 package (Case et al., 2014). One DBD copy in the resultant model was subjected to in silico mutations including A19D, K23N, H35Y, L114P, Q61R, K64E, I86V, and R116H, while the other remained unchanged. The model remaining wild type in both DBDs served as the reference system for MD simulations.

2.7. MD simulations

Simulations were performed using the Amber14 package (Case et al., 2014) in parallel on an 8-GTX1080Ti-GPU workstation following a five-step protocol including ensemble construction, minimization, heating, equilibration, and production. The force field parameters retrieved from ff14SB and ff99bsc0 for protein and DNA respectively were applied. Each ensemble was explicitly solvated in a box using the TIP3 water model. An appropriate amount of Na+ and Cl- ions were added to each ensemble to mimic the physiological saline milieu and ensure zero net charges. After minimization, heating, and equilibration, one 100 million steps of NPT simulations with 1-fs step size at 300 K were calculated in the process of production. The embedded tools in Amber were used to analyze the resultant trajectories from which a final stable conformation for each DBD-DNA complex was extracted. For structural comparison, all mutant models were aligned to the output coordinates from the reference system.

3. Results

3.1. The DBD in HSF4 displays a conserved overall structure with other HSFs

HSF4-DBD (residues 17–122) was crystallized in space group P222, same as the DNA-free structure of HSF1-DBD determined recently in our lab (PDB entry 5HDG) (Feng et al., 2016b). The diffraction resolution at 1.2 Å, which was the highest among all available HSF structures, allowed us to refine a model with excellent quality (Table 2). Under such a resolution, the 2Fo - Fc density was so sharp to define the precise position of each non-hydrogen atom (Fig. 1A), although residues within the wing (residues 88–96) were completely invisible and had to be omitted from the model (Fig. 1A), which is consistent with most reported DBD structures (Feng et al., 2016b; Harrison et al., 1994; Littlefield and Nelson, 1999; Neudegger et al., 2016).

HSF4-DBD shares high degree of sequence homology and identical structural topology with other HSF-DBDs, in particular with human HSF1 (sequence identity = 77%, RMSD = 0.697 Å) and HSF2 (sequence identity = 70%, RMSD = 0.617 Å) (Fig. 1C and 1D). Only structural topology with other HSF-DBDs, in particular with human HSF1 (sequence identity = 77%, RMSD = 0.697 Å) and HSF2 (sequence identity = 70%, RMSD = 0.617 Å) (Fig. 1C and 1D). Only slight structural differences were observed, e.g. discontinued helix α2 in HSF4 vs integrated one in other HSFs. For further analysis of the CC-associated mutations, a DNA-binding model was built by replacing HSF2-DBD in the PDB entry of 5DHG (Jaeger et al., 2016) with the refined model of HSF4-DBD. In this model, two copies of HSF4-DBD bind to a palindromic DNA molecule containing two binding sites arranged in the tail-to-tail orientation (Fig. 1E). Starting from this model, we conducted in silico mutations in only one DBD copy followed by MD simulations for 100 ns, with all ensembles displaying reasonable RMSD fluctuations (Supplementary Figs. 3–5).

3.2. CC-associated mutations occur in either highly or poorly conserved positions

To analyze relative conservation of the mutation site associated with congenital cataract, the sequence and structure of HSF4-DBD was input to the ConSurf server (Landau et al., 2018), where a conservation profile of individual amino acids was generated by psi-blasting with
150 homologous sequences retrieved from the REF90 database (Suzek et al., 2007). The reported positions can be classified into three groups (Table 1 and Supplementary Fig. 1): I, invariant (K23, K64, R73, and R119); II, highly conserved (A19, P60 and L114); III, non-conserved (H35, Q61, I86 and R116), meaning that no CC-associated mutations occur at moderately conserved positions. As expected, most conserved amino acids (group I and II) are exposed on the surface assumingly contacting DNA, while all poorly conserved residues are present on the surface.
DNA-distal surface (Fig. 2A). The only exception is L114, a partially buried hydrophobic residue.

3.3. Mutations R73H and R119C very likely destroy HSE recognition

R73 and R119 are highly conserved Arg residues presumably contacting bases within the major groove (Jaeger et al., 2016; Neudegger et al., 2016). Their counterparts in human HSF2, R63, and R109 for example, respectively form bidentate hydrogen bonds with the guanine in GAA (Supplementary Fig. 2A) and a solvent-mediated hydrogen bond with the purine preceding the TTA triplet in the complementary strand (Supplementary Fig. 2B). The guanidine side chains of these two arginines are necessary for the conserved hydrogen bonds essential for HSE recognition. We hence reason that any replacement of either R73 (Fig. 2B) or R119 (Fig. 2C) would very likely abolish the activity to recognize HSEs by destroying the base-specific interactions.

3.4. Mutations P60H and L114P probably affect protein folding

As an imide acid, proline is rarely found at internal positions in helices, which would otherwise mediate the formation of kinks because its cyclic pyrrolidine side-chain restricts the dihedral angles of the preceding amino acid (Barlow and Thornton, 1988). A strictly conserved Pro residue, however, is present at the center of a bulged kink in helix α2 in HSFs (Feng et al., 2016b; Harrison et al., 1994; Vuister et al., 1994), e.g. P60 in HSF4 (Fig. 2D), and was found to be irreplaceable in maintaining proper folding and protein solubility (Hardy and Nelson, 2000). Based on that study, it seems likely that the mutation of P60H may lead to insoluble HSF4 expression due to aberrant folding kinetics. L114 is a buried hydrophobic amino acid located in a 310 helix (η2) close to the C-terminus of HSF4-DBD. The occurrence of the L114P mutation means that a Pro residue is introduced at the internal of a 310 helix, which would generally destruct the helix integrity because of the
restricted torsion angles. Consistently, MD simulation showed the deformation of \( \eta_2 \) from a two-turned to a single-turned helix upon this mutation (Fig. 2E). Considering the role of P60 in protein folding, we suppose that the replacement of L114 with a Pro residue may also alter the solubility of newly synthesized HSF4.

3.5. Crystal structure of the K23N mutant reveals significant conformational changes

K23 is a highly conserved amino acid located in helix \( \alpha_1 \), but spatially close to the C-terminus of helix \( \alpha_2 \). We crystalized this mutant in space group \( P2_12_1_2 \) (Table 2). The asymmetric unit accommodates two protein monomers with an extended hybrid \( \beta \)-sheet formed at the dimeric interface (Fig. 3A). The overall scaffold of the mutant showed high similarity with the wild-type protein (RMSD = 0.186 Å) (Fig. 3B). A striking difference, however, was observed in stand \( \beta_3 \), which became much longer than that in the wild-type DBD (Fig. 3C). The change occurring in this \( \beta \)-strand gives more constraints to the downstream wing loop, which was, though, invisible in density.

We further built a DNA-binding model using the refined structure and did an MD simulation. Compared with the wild-type protein, helix \( \alpha_2 \) moves along with the helical axis to its C-terminus, and the side chain of K64 is inserted into the DNA minor groove, deviating from the horizontal orientation over the minor groove in the wild-type DBD (Fig. 3D). This comparison means that upon the K23N mutation, conserved interactions of K64 with the DNA backbone are replaced by contacts with bases, which is probably adverse to HSE recognition. In another regard, considering that the wing is involved in protein-protein interactions in an HSF trimer (Jaeger et al., 2016; Littlefield and Nelson, 1999; Neudegger et al., 2016), we speculate that this mutant with a more constrained wing may exhibit weaker DNA-binding affinity due to deterioration of the synergetic binding to multiple repeats within an HSE.

3.6. Mutations Q61R, K64E, and R116H may perturb DNA-binding by inducing conformational changes in local regions

Q61 is a poorly conserved amino acid in helix \( \alpha_2 \) (Fig. 1C), the first helix in the helix-turn-helix motif. When Q61 is replaced by arginine, cascaded conformational changes happened during a simulation. A new salt bridge between R61 and E57 replaced an old one between two preceding residues, K56 and E57, and the side chain of K56 was pushed away (Fig. 4A). Subsequently, E95* in the neighboring DBD was pulled closer to reoriented K56, which further led to an apparent wing movement in that DBD, and as a result, the side chain of K82 was inserted into the DNA major groove (Fig. 4B). The net consequences of these changes were varnished contacts between K82* and the DNA backbone, perturbation of base-specific contacts in the major groove, and probably devastated wing-mediated protein-protein interactions.

K64 is a highly conserved amino acid in the turn connecting helices \( \alpha_2 \) and \( \alpha_3 \), which is involved in electrostatic contacts with the DNA phosphates (Jaeger et al., 2016; Neudegger et al., 2016). K64E is a pronounced replacement from a positive charge to a negative charge. According to our simulation for this mutant, a salt bridge was formed between E64 and K23 in the vicinity, which led to a slight movement of helix \( \alpha_2 \) (Fig. 4C). More strikingly, the HSF4-DNA affinity greatly drops due to the repulsive force between the negative charges in E64 and the DNA backbone (Fig. 4D).
The C-terminal peptide of HSF4-DBD (residues 114–122) presumably wraps the DNA duplex and directs the orientation of the coiled-coil formed in an HSF4 trimer (Jaeger et al., 2016; Neudegger et al., 2016). This peptide comprises an arginine/lysine cluster that contributes to either a base-specific contact in the major groove (e.g. R119) or electrostatic contacts with the DNA phosphates (e.g. R118 and K120). Its conformation is stabilized by charge-charge interactions among E115, R116, and R118, which is however destabilized by the mutation of R116H. As revealed in the MD simulation, rotation of the E115 side chain by almost 180° induced local conformational change propagated to the very end of DBD, which let the side chain of K120 deviated from the DNA backbone (Fig. 4E and 4F). As a result, the crucial contacts between the C-terminal peptide and DNA were significantly impaired upon this mutation.

3.7. Mutations A19D, H35Y, and I86V induces unnoticeable conformational changes and insignificant impacts on DNA-binding

As a conserved amino acid at the N-terminus of helix α1, A19 is not involved in protein-DNA interactions (Jaeger et al., 2016; Littlefield and Nelson, 1999; Neudegger et al., 2016). Mutation to an aspartic acid introduces a negative charge at this position, which however
unexpectedly induces minute changes in either local or overall conformation, according to the simulation performed for this mutant (Fig. 5A).

H35 and I86 are poorly conserved amino acids exposed on the surface distal to DNA (Fig. 2A). According to the simulations, the replacement of I86 with a valine residue did not change the hydrophobic property at this position, which merely induced trivial changes in the wing (Fig. 5B). Slightly differently, the mutation of H35 to a tyrosine prompted the formation of a local hydrophobic core with the involvement of L36 and F54 (Fig. 5C). The overall conformation and protein–DNA interactions, however, seemed barely affected by this mutation (Fig. 5D).

4. Discussion

A number of mutations in human HSF4-encoding gene have been reported linking with cataractogenesis (Table 1) (Anand et al., 2018; Berry et al., 2018; Bu et al., 2002; Francis and Moore, 2004; Jing et al., 2014; Shi et al., 2008; Shields and Hejtmancik, 2017). Protein trimerization, DNA-binding and transcriptional activities of some HSF4 mutants have been probed using chemical cross-linking, electrophoretic mobilities, DNA-binding and transcriptional activities of some HSF4 mutants after simulation. (C and D) Superimposition of the wild-type DBD and the H35Y mutant after simulation, displaying local (C) and overall (D) conformational differences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The two mutations with proline involvement, however, seem more likely to cause improper folding of HSF4 (Fig. 2D and 2E). The P60H mutation was identified very recently (Li et al., 2016). Although the experimental data for this mutant are unavailable, we believe that this mutation likely results in protein insolubility according to a previous mutational study of yeast HSF (Hardy and Nelson, 1999). Despite being expressed in soluble form, the L114P mutant showed severely inhibited trimerization, DNA-binding and reporter gene transcription (Enoki et al., 2010), and could not be crystallized (data from this work), all inferring significant differences with the wild-type DBD in terms of protein structure and stability, which are usually associated with protein folding.

The MD simulations of the K23N, Q61R, K64E, and R116H mutants suggest that these mutations probably impair DNA-binding specificity and/or affinity by inducing conformational changes in local regions (Figs. 3 and 4). The activity loss of the K23N and K64E mutants has not been reported yet, but their structural impacts are straightforward, both of which abolish the charge-charge interactions between K64 and DNA phosphates (Fig. 3D and 4D). More interestingly, the crystal structure of K23N reveals a marked extension of stand β3, which in principle renders the wing loop less flexible (Fig. 3C). Although this conformation deviation from the wild-type protein might reflect a crystal-packing artifact, there is a possibility that this mutation may also damage the synergic DNA binding among DBDs in an HSF4 trimer, which is usually mediated by the wing (Jaeger et al., 2016; Littlefield and Nelson, 1999; Neudegger et al., 2016). In agreement, relatively weakened DNA-binding activity of this mutant compared with the wild-type protein was detected by EMSAs (Supplement Fig. 6). Mutations Q61R and R116H were originally found in age-related cataracts but later confirmed to cause childhood lamellar cataract in transgenic mice (Jing et al., 2014). These mutations occurring at non-conserved positions...
Z. Xiao, et al.
Journal of Structural Biology: X 4 (2020) 100015
doi.org/10.1016/j.jsbxb.2019.100015.

References

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Uebrumstve, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. Sect. D Biol. Crystallogr. 68, 352–367. https://doi.org/10.1107/S0909444910001308.

Akerfelt, M., Morimoto, R.I., Sistonen, L., 2010. Heat shock factor: integrators of cell stress, development and lifespan. Nat. Rev. Mol. Cell Biol. 11, 545–555. https://doi.org/10.1038/nrm2938.

Amin, J., Ananthan, J., Voellmy, R., 1988. Key features of heat shock regulatory elements. Mol. Cell. Biol. 8, 3761–3769. https://doi.org/10.1128/mcb.8.9.3761.

Anand, D., Agrawal, S.A., Slavotinek, A., Lachke, S.A., 2018. Mutation update of transcription factor genes FOXE3, HSF4, MAF, and PITX3 causing cataracts and other developmental ocular defects. Humat. Mutat. 39, 471–494. https://doi.org/10.1002/humu.23995.

Barlow, D.J., Thornton, J.M., 1988. Helix geometry in proteins. J. Mol. Biol. 201, 601–619. https://doi.org/10.1016/0022-2836(88)90641-9.

Berry, V., Pontioks, N., Moore, A., Ionides, A.C.W., Plagod, V., Cheetham, M.E., Michaelides, M., 2018. A novel missense mutation in HSF4 causes autosomal-dominant congenital lamellar cataract in a British family. Eye 32, 806–812. https://doi.org/10.1038/eye.2017.268.

Bjork, J.K., Sistonen, L., 2010. Regulation of the members of the mammalian heat shock factor family. FEBS J. 277, 4126–4139.

Bu, L., Jin, Y., Shi, Y., Chu, R., Ban, A., Eiberg, H., Andres, L., Jiang, H., Zheng, G., Qian, M., Cui, B., Xia, Y., Liu, J., Bao, L., Zhao, G., Hayden, M.R., Kong, X., 2002. Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. Nat. Genet. 31, 276–278. https://doi.org/10.1038/ng921.

Björkman, H., Ruusuvuori, E., Stenman, M., 2012. Heat shock transcription factor family. FEBS J. 279, 2641–2658. https://doi.org/10.1111/j.1742-4658.2012.08290.x.

Bu, L., Jin, Y., Shi, Y., Chu, R., Ban, A., Eiberg, H., Andres, L., Jiang, H., Zheng, G., Qian, M., Cui, B., Xia, Y., Liu, J., Bao, L., Zhao, G., Hayden, M.R., Kong, X., 2002. Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. Nat. Genet. 31, 276–278. https://doi.org/10.1038/ng921.

Westerheide, D., Raynes, R., Powell, C., Xue, B., Uversky, V., 2012. HSF transcription factor family, heat shock response, and protein intrinsic disorder. Curr. Protein Pept. Sci. 13, 86–103. https://doi.org/10.2174/13892031012799727996.

D.A. Case, T., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D.S., Cheatham, T., D.A. Case, T., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D.S., Cheatham, T., Cieplak, P., Cornell, W., Feller, M., Fox, T., Gohlke, B., Kale, L., Kollman, P.A., LaRue, C., Livernet, S., Merz, K., Michel, C., Nilsson, L., Piana, M., Priolo, S., Roux, B., Smith, A., Strockbine, S., Straatsma, B., Westler, W.M., Woods, D., Yeğen, M., 2005. AMBER 8: an object-oriented software architecture for simulation of biomolecules. J. Comput. Chem. 26, 1668–1684. https://doi.org/10.1002/jcc.20400.

Enoki, Y., Maloda, Y., Furutani, C., Sakurai, H., 2010. DNA-binding and transcriptional activities of human HSF4 containing mutations that associate with congenital and age-related cataracts. Biochim. Biophys. Acta. Mol. Basis Dis. 1802, 749–753. https://doi.org/10.1016/j.bbadis.2010.06.001.

Enoki, Y., Mukoda, Y., Furutani, C., Sakurai, H., 2010. DNA-binding and transcriptional activities of human HSF4 containing mutations that associate with congenital and age-related cataracts. Biochim. Biophys. Acta. Mol. Basis Dis. 1802, 749–753. https://doi.org/10.1016/j.bbadis.2010.06.001.

Feng, H., Liu, W., Wang, D.C., 2016a. Purification, crystallization and X-ray diffraction analysis of the DNA-binding domain of human heat-shock factor 2. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 486–501. https://doi.org/10.1107/S0907444916007492.

Fujimoto, M., Inoue, S., Nakai, A., 2004. HSF4 is required for normal cell growth and differentiation during mouse lens development. EMBO J. 23, 4297–4306. https://doi.org/10.1038/sj.emboj.7600435.

Fujimoto, M., Oshima, K., Shinakawa, T., Imai, Y., Sato, Y., Hori, K., Sasaki, K., Takahashi, J., 2001. Analysis of HSF4 binding regions reveals its necessity for gene regulation during development and heat shock response in mouse lenses. J. Biol. Chem. 276, 29961–29970. https://doi.org/10.1074/jbc.M0062920.

Gillespie, R.L., O'Sullivan, J., Ashworth, J., Blaskar, S., Williams, S., Biswas, S., Kehdi, E., Ramsden, S.C., Clayton-Smith, J., Black, G.C., Lloyd, I.C., Wallis, W., 2010. Personalized diagnosis and management of congenital cataract by next-generation sequencing. Ophthalmology 117, 278–284. https://doi.org/10.1016/j.ijid.2010.06.001.

Francis, P.J., Moore, A.T., 2004. Genetics of childhood cataract. Curr. Opin. Ophthalmol. 15, 10–15.

Fujimoto, M., Inoue, S., Nakai, A., 2004. HSF4 is required for normal cell growth and differentiation during mouse lens development. EMBO J. 23, 4297–4306. https://doi.org/10.1038/sj.emboj.7600435.

Gomez-Pastor, R., Burchiel, E.T., Thiele, D.J., 2018. Regulation of heat shock transcription factors and their roles in physiology and disease. Nat. Rev. Mol. Cell Biol. 19, 4–19. https://doi.org/10.1038/nrm.2017.73.

Gonzales, S.E., Moses, A.M., Razak, Z., Robert, F., Timothy Westwood, J., 2011. Whole-genome analysis reveals that active heat shock factor binding sites are mostly associated with non-heat shock genes in drosophila melanogaster. PLoS One 6, 1–10. https://doi.org/10.1371/journal.pone.0015934.

Hardy, J.A., Nelson, H.C., 2000. Proline in alpha-helical kink is required for folding kinetics but not for kinked structure, function, or stability of heat shock transcription factor. Protein Sci. 9, 2128–2141. https://doi.org/10.1128/ps.9.9.2128.

Harrison, C.J., Bohn, A.A., Nelson, H.C.M., 1994. Crystal structure of the DNA binding domain of the heat shock transcription factor. Science (80-) 263, 224–227. https://doi.org/10.1126/science.8246072.

Hejmancik, J.F., 2008. Congenital cataracts and their molecular genetics. Semin. Cell Dev. Biol. 19, 134–149. https://doi.org/10.1016/j.semcdb.2007.10.003.

Jaeger, A.M., Pembel, C.W., Sistonen, L., Thiele, D.J., 2016. Structures of HSF2 reveal
mechanisms for differential regulation of human heat-shock factors. Nat. Struct. Mol. Biol. 23, 147–154. https://doi.org/10.1038/nsmb.3150.

Jing, Z., Gangalum, R.K., Bhat, A.M., Nagaoka, Y., Jiang, M., Bhat, S.P., 2014. HSF4 mutation p.Arg116His found in age-related cataracts and in normal populations produces childhood lamellar cataract in transgenic mice. Hum. Mutat. 35, 1068–1071. https://doi.org/10.1002/humu.22610.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., Ben-Tal, N., Gomez-Pastor, R., Burchiel, E.T., Thiele, D.J., 2018. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nat. Rev. Mol. Cell Biol. 33, 299–302. https://doi.org/10.1038/nrm3730.

Li, D., Wang, S., Ye, H., Tang, Y., Qiu, X., Fan, Q., Rong, X., Liu, X., Chen, Y., Yang, J., Lu, Y., 2016. Distribution of gene mutations in sporadic congenital cataract in a Han Chinese population. Mol ViS 22, 539–538.

Littlefield, O., Nelson, H.C.M., 1999. A new use for the “wing” of the “winged” helix-turn-helix motif in the HSF-DNA complex. Nat. Struct. Biol. 6, 464–470. https://doi.org/10.1038/9269.

Lund, A.M., Eiberg, H., Rosenberg, T., Warburg, M., 1992. Autosomal dominant congenital cataract: linkage relations; clinical and genetic heterogeneity. Clin. Genet. 41, 65–69. https://doi.org/10.1111/j.1399-0004.1992.tb03634.x.

Min, J.N., Zhang, Y., Moskophidis, D., Miyachi, N.F., 2004. Unique contribution of heat shock transcription factor 4 in ocular lens development and fiber cell differentiation. Genesis 40, 205–217. https://doi.org/10.1002/gene.20087.

Minor, W., Cymborowski, M., Owunowozi, Z., Chruszcz, M., 2006. HKL-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 855–866. https://doi.org/10.1107/S0907444906019449.

Neef, D.W., Jaeger, A.M., Thiele, D.J., 2011. Heat shock transcription factor 1 as a therapeutic target in neurodegenerative diseases. Nat. Rev. Drug Discov. 10, 140–146. https://doi.org/10.1038/nrd3453.

Neufeld, K., Vihervaara, A., Sistonen, L., 2014. HSF1 at a glance. J. Cell Sci. 127, 261–266. https://doi.org/10.1242/jcs.132605.

Neef, D.W., Jaeger, A.M., Thiele, D.J., 2011. Heat shock transcription factor 1 as a therapeutic target in neurodegenerative diseases. Nat. Rev. Drug Discov. 10, 930–944. https://doi.org/10.1038/nrd3453.

Neufeld, K., Vihervaara, A., Sistonen, L., 2014. HSF1 at a glance. J. Cell Sci. 127, 261–266. https://doi.org/10.1242/jcs.132605.

Neufeld, K., Vihervaara, A., Sistonen, L., 2014. HSF1 at a glance. J. Cell Sci. 127, 261–266. https://doi.org/10.1242/jcs.132605.

Neufeld, K., Vihervaara, A., Sistonen, L., 2014. HSF1 at a glance. J. Cell Sci. 127, 261–266. https://doi.org/10.1242/jcs.132605.

Neufeld, K., Vihervaara, A., Sistonen, L., 2014. HSF1 at a glance. J. Cell Sci. 127, 261–266. https://doi.org/10.1242/jcs.132605.

Neufeld, K., Vihervaara, A., Sistonen, L., 2014. HSF1 at a glance. J. Cell Sci. 127, 261–266. https://doi.org/10.1242/jcs.132605.