SUPPLEMENTARY MATERIAL FOR

Solution NMR Structures of Homeodomains from Human Proteins

ALX4, ZHX1, and CASP8AP2 Contribute to the Structural Coverage of the Human Cancer Protein Interaction Network

Xianzhong Xu • Surya VSRK Pulavarti • Alexander Eletsky • Yuanpen Janet Huang • Thomas B. Acton • Rong Xiao • John K. Everett • Gaetano T. Montelione • Thomas Szyperski

Xianzhong Xu • Surya VSRK Pulavarti • Alexander Eletsky • Thomas Szyperski (✉)
Department of Chemistry, The State University of New York at Buffalo, and Northeast Structural Genomics Consortium, Buffalo, NY 14260, USA
E-mail: szypersk@buffalo.edu.

Yuanpeng Janet Huang • Thomas B. Acton • Rong Xiao • John K. Everett • Gaetano T. Montelione
Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

Gaetano T. Montelione
Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway NJ 08854, USA
Materials and Methods

Residues 462-532 of human zinc fingers and homeoboxes protein 1 [ZHX1(462-532)] containing homeodomain 2 (residues from 464-526), residues 209-280 of human homeobox protein aristaless-like 4 [ALX4(208-280)] containing a homeodomain (residues from 214-273), and the C terminal homeodomain comprising residues 1916-1982 of human caspase 8-associated protein 2 [CASP8AP2(1916-1982)] were cloned into a pET15 vector, yielding the plasmids HR7907F-462-532AV6HT.2, HR4490C-209-280-NHT.2, and HR8150A-1916-1982-Av6HT.3, respectively. The resulting constructs contained non-native residues at the N-terminus to facilitate protein purification [MSGLNDIFEAQKIEWHEHHHHHENLYFQSHM for both ZHX1(462-532) and CASP8AP2(1916-1982) and MDVEAWLDERVPLVETHHHHHHENLYFQSHM for ALX4(209-280)]. Cells of the rare codon enhanced *Escherichia coli* BL21 (DE3) pMGK strain, were transformed with HR7907F-462-532AV6HT.2, HR4490C-209-280-NHT.2, or HR8150A-1916-1982-Av6HT.3 plasmids, and cultured in MJ9 minimal medium containing (15NH4)2SO4 and U-13C-glucose as sole nitrogen and carbon sources. [U-13C, 15N]-ZHX1(462-532), [U-13C, 15N]-ALX4(209-280), and [U-13C, 15N]-CASP8AP2(1916-1982) were purified using an AKTAxpress system (GE Healthcare) with three-step protocol consisting of 1st IMAC (HisTrap HP), 2nd IMAC(HisTrap HP) after His tag cleavage with TEV protease, and gel filtration (HiLoad 26/60 Superdex 75) chromatography. The final yield of purified [U-13C, 15N]-ZHX1(462-532) (> 98% monodisperse by light scattering with molecular weight (MW) of 9.19 kDa by MALDI-TOF mass spectrometry) was about 10.1 mg/L. Each protein has nonnative SHM residues in the N terminus after TEV cleavage of the His tag. The final yield of purified [U-13C, 15N]-ALX4(209-280) (> 98% monodisperse by light scattering with MW of 9.88 kDa by MALDI-TOF mass spectrometry) was about 1.8 mg/L. The
final yield of purified $[U^{-13}\text{C};^{15}\text{N}]-\text{CASP8AP2(1916-1982)}$ (> 98% monodisperse by light scattering with MW of 8.72 kDa by MALDI-TOF mass spectrometry) was about 11.4 mg/L. In addition, uniformly $^{15}\text{N}$- and 5% biosynthetically directed fractionally $^{13}\text{C}$-labeled samples {[$5\%^{13}\text{C}; U^{-15}\text{N}]$-ZHX1(462-532), ($5\%^{13}\text{C}; U^{-15}\text{N}]$-ALX4(209-280), and [$5\%^{13}\text{C}; U^{-15}\text{N}]$-CASP8AP2(1916-1982)) were generated using a mixture of 95% natural abundance and 5%-$^{13}\text{C}$-glucose as a carbon source in the culture medium. The final samples of $[U^{-13}\text{C};^{15}\text{N}]$-ZHX1(462-532), $[U^{-13}\text{C};^{15}\text{N}]$-ALX4(209-280), $[U^{-13}\text{C};^{15}\text{N}]$-CASP8AP2(1916-1982), [$5\%^{13}\text{C}; U^{-15}\text{N}]$-ZHX1(462-532), [$5\%^{13}\text{C}; U^{-15}\text{N}]$-ALX4(209-280) and [$5\%^{13}\text{C}; U^{-15}\text{N}]$-CASP8AP2(1916-1982) were prepared at respective concentrations of approximately ~0.4, ~0.9, ~0.8, ~0.3, ~0.6, and ~0.6 mM, respectively in 90% H$_2$O/10% D$_2$O at pH 6.5 and containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM CaCl$_2$, 50 µM DSS, 0.02% NaN$_3$ for ZHX1(462-532) and CASP8AP2(1916-1982), and in 90% H$_2$O/10% D$_2$O at pH 7.5 and containing 10 mM Tris-HCl, 100 mM NaCl, 10 mM DTT, 50 µM DSS, 0.02% NaN$_3$ for ALX4(209-280). The oligomerization state of ZHX1(462-532), ALX4(209-280), and CASP8AP2(1916-1982) were analyzed by static light scattering after an analytical gel-filtration column (Agilent Technologies) in the respective buffer (Figs. S1, S2, S3).

For $[U^{-13}\text{C};^{15}\text{N}]$-ZHX1(462-532), 2D $[^{15}\text{N},^1\text{H}]$ HSQC [1] (0.2 h), aromatic non-constant-time 2D $[^{13}\text{C},^1\text{H}]$-HSQC (0.5 h), aliphatic non-constant-time 2D $[^{13}\text{C},^1\text{H}]$ HSQC (0.9 h), aromatic constant-time 2D $[^{13}\text{C},^1\text{H}]$ HSQC [1] (0.2 h), aliphatic constant-time 2D $[^{13}\text{C},^1\text{H}]$ HSQC [1] (0.2 h), GFT (4,3)D $\underline{\alpha}^\text{\beta}$ $\underline{\alpha}^\text{\alpha}$ (14.7 h) [2], 3D HNCO [3] (1.6 h), GFT (4,3)D HNN$\underline{\alpha}^\text{\beta}$ $\underline{\alpha}^\text{\alpha}$ (14.0 h) [2], GFT (4,3)D $\underline{\alpha}^\text{\beta}$ $\underline{\alpha}^\text{\alpha}$ (15.6 h) [4, 5], aromatic and aliphatic GFT (4,3)D HCCH (5.7 and 8.5 h) [5-7], long-range 2D $[^{15}\text{N},^1\text{H}]$ HSQC for His side chains [8] (15.3 h), and 3D $^{15}\text{N}$/$^{13}\text{C}$-aliphatic/$^{13}\text{C}$-aromatic-resolved $[^1\text{H},^1\text{H}]$-NOESY (28.3 h) [7] were acquired on a Varian
INOVA 600 MHz spectrometer equipped with a cryogenic probe. Aliphatic 3D (H)CCH-TOCSY [9] (14.2 h), aromatic and aliphatic GFT (4,3)D HCCCH (1.6 and 14.0 h) and 3D $^{15}$N/$^{13}$C-aliphatic/$^{13}$C-aromatic-$[^{1}H,^{1}H]$-NOESY ($\tau_{\text{mix}} = 70$ ms) (42.5 h) were also acquired on a Varian INOVA 750 MHz spectrometer equipped with a cryogenic probe. 2D $[^{13}$C, $^{1}$H] HSQC spectra with 28, 42, and 56 ms (6.0, 7.0, 8.0 h, respectively) constant-time delays were recorded with the [5% $^{13}$C; U$^{15}$N]-ZH1(462-532) sample using a Varian INOVA 600 MHz spectrometer equipped with a cryogenic probe in order to obtain stereo-specific assignments for isopropyl groups of Val and Leu residues [10].

For [U-13C,$^{15}$N]-ALX4(209-280), 2D $[^{15}$N, $^{1}$H] HSQC [1] (0.2 h), aromatic constant-time 2D $[^{13}$C,$^{1}$H]-HSQC [1] (0.5 h), aliphatic constant-time 2D $[^{13}$C,$^{1}$H] HSQC [1] (0.8 h), GFT (4,3)D $\text{C}^{\alpha\beta}\text{C}^{\alpha}(\text{CO})\text{NHN}$ (14 h)[2], 3D HNCO [3] (2.0 h), GFT (4,3)D HNN$\text{C}^{\alpha\beta}\text{C}^{\alpha}$ (28.0 h) [2], 3D HBHA(CBCACO)NH [11] (7.3 h), 3D-HN(CA)CO [47] (13.1 h) and 3D $^{15}$N/$^{13}$C-aliphatic/$^{13}$C-aromatic-$[^{1}H,^{1}H]$-NOESY (29 h) [6] were acquired on a Varian INOVA 600 MHz spectrometer equipped with a cryogenic probe. A 2D $[^{13}$C, $^{1}$H] HSQC (1.5 h) spectrum with 28 ms constant-time delay were recorded with the [5% $^{13}$C; U$^{15}$N]-ALX4(209-280) sample using the same spectrometer in order to obtain stereo-specific assignments for isopropyl groups of Val and Leu residues [10]. In addition, for [U-13C,$^{15}$N]-ALX4(209-280), 2D $[^{15}$N, $^{1}$H] HSQC [1] (0.2 h), aromatic constant-time 2D $[^{13}$C,$^{1}$H] HSQC [1] (0.6 h), aliphatic 2D constant-time $[^{13}$C,$^{1}$H]-HSQC [1] (0.5 h), aliphatic [12] and aromatic [13] 3D (H)CCH (8.5 and 2.2 h), aromatic and aliphatic GFT (4,3)D HCCCH (4.3 and 17.3 h) [5,6,7], and aliphatic 3D (H)CCH-TOCSY [9] (13.8 h), and long-range 2D $[^{15}$N,$^{1}$H] HSQC for His side chains [8] (7.0 h), 3D $^{15}$N/$^{13}$C-aliphatic/$^{13}$C-aromatic-$[^{1}H,^{1}H]$-NOESY [7] (58.5 h) were acquired on a Varian INOVA 750 MHz spectrometer equipped with a cryogenic probe.
For $[U^{13}\text{C};^{15}\text{N}]-\text{CASP8AP2}$, 2D $[^{15}\text{N},^{1}\text{H}]\text{HSQC}$ [1] (0.3 h), 3D HNCO [3] (1.7 h), 3D $^{15}\text{N}/^{13}\text{C}$-aliphatic/$^{13}\text{C}$-aromatic-edited $[^{1}\text{H},^{1}\text{H}]$-NOESY (27.4 h) [7], aliphatic non-constant-time 2D $[^{13}\text{C},^{1}\text{H}]-\text{HSQC}$ (1.0 h), aromatic non-constant-time 2D $[^{13}\text{C},^{1}\text{H}]-\text{HSQC}$ (1.2 h), aliphatic constant-time 2D $[^{13}\text{C},^{1}\text{H}]-\text{HSQC}$ [1] (0.9 h), aromatic constant-time 2D $[^{13}\text{C},^{1}\text{H}]\text{HSQC}$ [1] (1.0 h), GFT (4,3)D $^{\alpha\beta}$$^{\alpha}$CO$^{\alpha\beta}$C$^{\alpha}$NHN (14.0 h) [2], GFT (4,3)D HNN$^{\alpha\beta}$$^{\alpha}$C$^{\alpha}$ (16.6 h)[2], 3D HBHA(CBCACO)NH [11] (13.7 h), 3D HN(CA)CO (13.0 h), aliphatic [12] and aromatic [13] 3D (H)CCH (9.4 and 4.7 h), aliphatic and aromatic GFT (4,3)D HCCH [5,6,7] (8.5, 5.4 h), and aliphatic 3D (H)CCH-TOCSY [9] (19.7 h) were acquired on a Varian INOVA 600 MHz spectrometer equipped with a cryogenic probe. 2D $[^{13}\text{C},^{1}\text{H}]$ HSQC spectra with 28, 42, and 56 ms (6, 5, and 4 h, respectively) constant-time delays were recorded with the [5% $^{13}\text{C};U^{15}\text{N}$]-CASP8AP2 sample using a Varian INOVA 500 MHz spectrometer in order to obtain stereo-specific assignments for isopropyl groups of Val and Leu residues [10]. In addition, for $[U^{13}\text{C},^{15}\text{N}]-\text{CASP8AP2}$, 2D $[^{15}\text{N},^{1}\text{H}]\text{HSQC}$ [1] (0.7 h) and 3D $^{15}\text{N}/^{13}\text{C}$-aliphatic/$^{13}\text{C}$-aromatic-resolved $[^{1}\text{H},^{1}\text{H}]$-NOESY [7] (27 h) were acquired on a Varian INOVA 750 MHz spectrometer equipped with a cryogenic probe.

1D $^{15}\text{N}$ T$_1$ and T$_2$ (CPMG) gradient-enhanced relaxation experiments [14] were recorded for $[U^{13}\text{C};^{15}\text{N}]-\text{ZHIX1(462-532)}$, ALX4(209-280) and CASP8AP2(1916-1982) samples with a Varian INOVA 750 MHz spectrometer, using the following delays, T$_1$ (s): 0.1, 0.2, 0.3, 0.4, 0.7, 1.0, 1.5, 2.0; T$_2$ (ms): 10, 30, 50, 70, 90, 110, 130, 150, 170. Spectra were processed with VNMRJ and integrated of the range 9.5-8.2 ppm for ZHIX1(462-532), 9.3-8.2 ppm for ALX4(209-280), and 9.0-8.4 ppm for CASP8AP2(1916-1982) using VNMRJ 2.1B (Varian). The following T$_1$ and T$_2$ relaxation times were extracted by fitting exponential decay curves. ZHIX1(462-532): $T_1 = 460 \pm 14$ ms, $T_2 = 101 \pm 4$ ms; ALX4(209-280): $T_1 = 567 \pm 6$ ms, $T_2 = 89 \pm 6$ ms;
CASP8AP2(1916-1982): $T_1 = 591\pm46$ ms, $T_2 = 109\pm19$ ms. Isotropic rotational correlation times ($\tau_c$) were estimated in the 'slow molecular tumbling limit' approximation according to [15]:

$$\tau_c = \frac{1}{4\pi^2 v_N} \sqrt{\frac{T_1}{6T_2}}$$

The resulting $\tau_c$ values of about 6±0.3, 7±0.3, and 5±0.7 ns for ZHX1(462-532), ALX4(209-280), and CASP8AP2(1916-1982), respectively, indicate that all three proteins are monomeric in solution, consistent with the static light-scattering results after an analytical gel-filtration column (Agilent Technologies). The slightly longer $\tau_c$ value obtained for ALX4(209-280) is very likely due to the long flexible tails at both termini. Such tails are known to slow down the overall rotational tumbling.

Multidimensional NMR spectra were processed with PROSA 6.4 [16] for ZHX1(462-532), ALX4(209-280), and CASP8AP2(1916-1982). Visualization and analysis of all spectra were performed with XEASY [17], and CARA [18]. Proton chemical shifts were referenced to internal DSS, while $^{13}$C and $^{15}$N chemical shifts were referenced indirectly via gyromagnetic ratios. Sequence-specific backbone (HN, N, C$^\alpha$) and C$'$ resonance assignments were obtained in a semi-automated fashion by analyzing GFT (4,3)D $C^\alpha\beta C^\alpha(CO)NHN$ [2], (4,3)D HNN$C^\alpha\beta C^\alpha$ and 3D HNCO peak lists using the automated backbone resonance assignment program Autoassign [19, 20]. Resonance assignment of side-chains was performed manually using HBHA(CBCACO)NH [except for ZHX1(462-532), for which (4,3)D $H\alpha\beta C^\alpha(CO)NHN$ was used], aromatic and aliphatic (4,3)D HCCCH, aliphatic (H)CCCH TOCSY, and $^{15}$N/$^{13}$C-aliphatic/$^{13}$C-aromatic -resolved NOESY ($\tau_m = 70$ ms). Stereo-specific assignments were obtained for all valine and leucine methyl groups with non-degenerate chemical shifts based on 2D constant-time $[^{13}$C,$^1$H]-HSQC spectra recorded for all three [5% $^{13}$C;U-$^{15}$N]- samples [10].
Assignment completeness was obtained by using the AVS (Assignment Validation Software) suite [21]. Chemical shifts, NOESY peak lists, and raw NOESY time-domain data for ZHX1(462-532), ALX4(209-280), and CASP8AP2(1916-1982) were deposited in the BioMagResBank with BMRB accession numbers 18714, 18805, and 18352, respectively.

$^1$H-$^1$H upper distance constraints for structure calculations were obtained from 3D $^{15}$N/$^{13}$Caliphatic/$^{13}$Caromatic-resolved $[^1$H,$^1$H]-NOESY (Table I). Backbone dihedral angle constraints were derived from chemical shifts using the program TALOS+ [22] (Table I). Automated NOE assignment was performed initially with the programs AutoStructure (Version 2.2.1 [23]) and CYANA 2.1 [24, 25], the resulting consensus peak assignments were verified and corrected by interactive spectral analysis. Subsequently the structure calculation was performed iteratively with CYANA to verify and complete resonance assignments, refine NOESY peak lists and optimize the distance calibration constants. The final 20 conformers out of 100 were further refined by restrained molecular dynamics in explicit water [26] using the program CNS (Version 1.2) [27]. Structural statistics and global quality factors, such as Verify3D [28], ProsaII [29], PROCHECK [30] and MolProbity [31] raw and statistical Z-scores were computed using PSVS 1.4 [32]. The goodness-of-fit between the final ensemble of conformers and the NOESY peak lists was calculated with the AutoStructure/RPF 2.2.1 package [33]. The resulting coordinates for ZHX1(462-532), ALX4(209-280), and CASP8AP2(1916-1982) were deposited in the Protein Data Bank with PDB IDs 2LY9, 2M0C, and 2LR8.

The sequence of ALX4(209-280) is strictly conserved in the genomes of the following mammals: Pan paniscus (Pygmy Chimpanzee)(Bonobo), Mus musculus (Mouse), Hylobates klossii (Kloss’s gibbon), Pan Troglohytes (Chimpanzee), Ailuropoda melanoleuca (Giant panda), Canis familiaris (Dog)(Canis lupus familiaris), Callithris jacchus (White-tufted-ear
marmoset), Macaca mulatta (Rhesus macaque), Equus caballus (Horse), Myotis lucifugus (Little brown bat), Oryctolagus cuniculus (Rabbit), Cricetulus griseus (Chinese hamster) (Cricetulus barabensis griseus), Loxodonta Africana (African elephant), Heterocephalus glaber (Naked mole rat), Cavia porcellus (Guinea pig), Otolemur garnettii (Small-eared galago) (Garnett’s greater bushbabay), Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii), Spermophilus tridecemlineatus (Thirteen-lined ground squirrel) (Ictidomys tridecemlineatus), and Myotis davidii (David’s myotis).
Fig. S1 Static light scattering results for ZHX1(462-532) before cleavage of the His tag. The NMR sample (10 µl) of [5%13C;U-15N]-ZHX1(462-532) at 20 mM MES pH 6.5, 100 mM NaCl, 10 mM DTT, 1X proteinase inhibitors, 10% D2O was injected onto an analytical gel-filtration column (Superdex200 5/150 GL, GE Healthcare) with the effluent monitored by ultraviolet light (blue trace) and 90° static light-scattering (black trace; miniDAWN TREOS, Wyatt Technology) detectors. The resulting experimental molecular weight is 14.31 kDa (red) (expected: 13.04 kDa).
Fig. S2 Static light scattering results for ALX4(209-280). The NMR sample (10 µl) of \([^{13}C;^{15}N]-ALX4(209-280)\) at 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM DTT, 1X proteinase inhibitors, 10% D₂O was injected onto an analytical gel-filtration column (Superdex200 5/150 GL, GE Healthcare) with the effluent monitored by refractive index (blue trace, Optilab rEX) and 90° static light-scattering (black trace; miniDAWN TREOS, Wyatt Technology) detectors. The resulting experimental molecular weight is 5.69 kDa (red)(expected: 9.87 kDa).
Fig. S3 Static light scattering results for CASP8AP2(1916-1982). The NMR sample (10 µl) of $[^{13}\text{C};^{15}\text{N}]$-CASP8AP2(1916-1982) at 20 mM MES pH 6.5, 100 mM NaCl, 10 mM DTT, 1X proteinase inhibitors, 10% D$_2$O was injected onto an analytical gel-filtration column (Superdex200 5/150 GL, GE Healthcare) with the effluent monitored by ultraviolet light (blue trace) and 90° static light-scattering (black trace; miniDAWN TREOS, Wyatt Technology) detectors. The resulting experimental molecular weight is 11.41 kDa (red) (expected: 8.82 kDa).
**Fig. S4** 2D $[^{15}\text{N},^{1}\text{H}]$ HSQC spectrum of ZHX1(462-532) recorded at 600 MHz $^{1}\text{H}$ resonance frequency. Resonance assignments are indicated using the one-letter amino acid code. Signals arising from side chains (Asn $^{\text{H}}\delta^{2}/^{\text{N}}\delta^{2}$, Arg $^{\text{H}}\varepsilon/N^{\varepsilon}$ and Trp $^{\text{H}}\varepsilon^{1}/^{\text{N}}\varepsilon^{1}$) are labeled with (*) next to the residue number.
Fig. S5 2D $[^{15}\text{N}, ^1\text{H}]$ HSQC spectrum of ALX4(209-280) recorded at 600 MHz $^1\text{H}$ resonance frequency. Resonance assignments are indicated using the one-letter amino acid code. Signals arising from side chains (Asn H$\delta$/N$\delta$, Arg H$\epsilon$/N$\epsilon$ and Trp H$\epsilon$/N$\epsilon$) are labeled with (*) next to the residue number. The two peaks depicted in grey are folded and arise from arginine side chains.
Fig. S6 2D $^{15}\text{N,}^1\text{H}$ HSQC spectrum of CASP8AP2(1916-1982) recorded at 600 MHz $^1\text{H}$ resonance frequency. Resonance assignments are indicated using the one-letter amino acid code. Signals arising from side chains (Asn $H^\delta_2/N^\delta_2$, Arg $H^\varepsilon/N^\varepsilon$ and Trp $H^{\varepsilon_1}/N^{\varepsilon_1}$) are labeled with (*) next to the residue number. Peaks labeled with (#) were not assigned.
Fig. S7 a, b, and c: Stereo view of the 20 conformers representing the solution structures of ZHX1(462-532), ALX4(209-280), and CASP8AP2(1916-1982), obtained after superposition of the Cα atoms of the α-helices for minimal RMSD. Residues 462-472 and 519-532 for ZHX1, residues 209-222 and 270-280 for ALX4, and residues 1916-1930 and 1978-1982 for CASP8AP2 of the disordered N- and C-terminal polypeptide segments were omitted for clarity and the termini are labeled as ‘N’ and ‘C’. d, e, and f: Ribbon diagram of residues 473-518 for ZHX1, 223-269 for ALX4, and 1931-1977 for CASP8AP2 of the lowest-energy conformer of ZHX1(462-532), ALX4(209-280), and CASP8AP2(1916-1982): α-helices are shown in red and yellow and other polypeptide segments are in gray.
References