Published in final edited form as:

Structure. 2016 August 2; 24(8): 1372–1379. doi:10.1016/j.str.2016.05.013.

Dynamic Local Polymorphisms in the Gbx1 Homeodomain Induced by DNA Binding

Andrew Proudfoot[1], **Michael Geralt**[1], **Marc-Andre Elsliger**[1], **Ian A. Wilson**[1], **Kurt Wüthrich**[1,2,*], and **Pedro Serrano**[1,*]

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA, and Joint Center for Structural Genomics (http://www.jcsg.org), La Jolla, CA 92037, USA

[2]Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

SUMMARY

The *Gastrulation Brain Homeobox 1 (Gbx1)* gene encodes the Gbx1 homeodomain that targets TAATTA motifs in dsDNA. Residues Glu17 and Arg52 in Gbx1 form a salt bridge, which is preserved in crystal structures and MD simulations of homologous homeodomain–DNA complexes. In contrast, our NMR studies show that DNA-binding to Gbx1 induces dynamic local polymorphisms, which include breaking of the Glu17–Arg52 salt bridge. To study this interaction, we produced a variant with Glu17Arg and Arg52Glu mutations, which exhibited the same fold as the wild-type protein, but a two-fold reduction in affinity for dsDNA. Analysis of the NMR structures of the Gbx1 homeodomain in the free form, the Gbx1[E17R,R52E] variant, and a Gbx1 homeodomain–DNA complex showed that stabilizing interactions of the Arg52 side chain with the DNA backbone are facilitated by transient breakage of the Glu17–Arg52 salt bridge in the DNA-bound Gbx1.

Keywords

Binding affinity; NMR structure determination; J-UNIO; protein; DNA complex

INTRODUCTION

The *Gastrulation Brain Homeobox 1* (*Gbx1*) gene is involved in the development of neurons in the dorsal and ventral spinal cord during embryogenesis (Buckley et al., 2013). Two different types of dorsal neuron populations are present in the developing embryo, i.e., six early-born populations, called dl1-dl6, and two late-born populations, called dlLA and dlLB (Meziane et al., 2013). Gbx1 is expressed during the second phase of neurogenesis and is

*Correspondence to: Pedro Serrano (serrano@scripps.edu), Kurt Wüthrich wuthrich@scripps.edu. Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

Author contributions

AP and MG produced, expressed and purified the proteins. AP and PS determined and validated the structures. AP performed CD and ITC experiments. AP, MAE, IAW, KW and PS designed the research strategy. AP, KW and PS wrote the manuscript.

associated with the dlLA sub-population of neurons, which undergo GABAergic differentiation (John et al., 2005). The expression of *Gbx1* in transgenic mice is therefore first detected at embryonic day 7.5 (E7.5) (Rhinn et al., 2004). At E8.25, expression is detected in the prospective hindbrain, and it is identified in regions of the developing forebrain by E10.5. By E11.5, Gbx1 is expressed in the spinal cord, and by E14 expression is restricted only to a narrow layer in the dorsal horn within the spinal cord (Meziane et al., 2013; Rhinn et al., 2004; Water et al., 2003). Mice heterozygous for a *Gbx1* knockout mutation presented with no behavioral or phenotypic abnormalities, while mice homozygous for the null mutation displayed developmental defects, which impact central nervous system organisation and function. By ten weeks of age, $Gbx1^{-/-}$ mice exhibited a locomotive defect, which specifically affected hind-limb gait, resulted in unevenness in walking, and is not related to muscle strength or motor co-ordination (Buckley et al., 2013; Meziane et al., 2013).

Gbx1 is one of two *Gbx* genes, *Gbx1* and *Gbx2*, in the *Gbx* gene cluster. Both genes encode DNA-binding transcription factors which are members of the *Antennapedia* (ANTP) super class of homeodomains and are related to the *Drosophila unplugged* gene (Rhinn et al., 2004; Waters et al., 2003). Gbx1 is highly conserved in higher eukaryotes, with 91% overall sequence identity between the mouse and human proteins (Sievers et al., 2011).

Homeodomains form structures with three α-helices folded into a compact globular domain containing a helix-turn-helix motif (Qian et al., 1989; Billeter et al., 1990), where the third helix serves as the recognition helix (Billeter et al., 1993; Otting et al., 1990)Otting et al., 1990; Billeter et al., 1993). Key amino acids involved with DNA binding are highly conserved in homeodomain sequences (Banerjee-Basu and Baxevanis, 2001). Furthermore, several amino acid pairs have been identified that contribute to protein stability through the formation of salt bridges between the different α-helices. The most highly conserved salt bridge in the homeodomain family links Glu17 and Arg52 in helices 1 and 3, and salt bridges between residues Glu19 and Arg30, and between Arg31 and Glu42 are also quite common (Torrado et al., 2009). Comparison of the different individual molecules in the crystallographic asymmetric unit (Longo et al., 2007) and molecular dynamics studies (Babin et al., 2013; Flader et al., 2003) identified small rearrangements of the Glu17 and Arg52 salt bridge and the recognition helix upon binding to DNA, whereas NMR solution structures showed larger conformational changes, which may include disruption of the Glu17 interaction with Arg52 (Billeter et al., 1993; Torrado et al., 2009; Baird-Titus et al., 2006; Chaney et al., 2005).

In this study, we investigate changes in the Gbx1 homeodomain induced by DNA binding. To gain further insights into the role of the Glu17/Arg52 interaction, we designed the Arg17/Glu52 variant and used circular dichroism (CD) spectroscopy and isothermal titration calorimetry (ITC) to assess the effects on protein stability and binding affinity to DNA. NMR structure determination was then used to provide a structural basis for the observed changes in binding affinity and protein stability.

RESULTS

Others have shown for nine different homeodomains that replacement of either Glu17 or Arg52 (residue numbering according to Torrado et al., 2009; in the Gbx1 constructs used in our work, these positions are shifted to Glu23 and Arg58, but to facilitate correlations with other homeodomains we also use the numbering in Torrado et al., 2009; see Fig. 1a) by a neutral amino acid caused decreased protein stability, resulting in protein aggregation during purification (Chi et al., 2005). In a different approach, data mining showed that homeodomains with a salt bridge between residues 17 and 52 invariably contain the positively charged residue at position 52. Here, a Gbx1[E17R,R52E] double mutant was generated, i.e., the two residues involved in the salt bridge were swapped (Fig. 1a), and the effects on protein stability and the interaction with DNA were studied. Gbx1[E17R,R52E] purified with similar yield as the wild type Gbx1 homeodomain, indicating that the mutations preserved protein stability. This was then confirmed with thermal denaturation curves monitored by CD spectroscopy (Fig. 1b), which are sigmoidal in shape and indicate that all regular secondary structures are melted at 75 °C. The mid-points of unfolding are 52.5 °C for the wild-type Gbx1 and 47.3 °C for the Gbx1[E17R,R52E] homeodomains. This data provided an initial indication that the homeodomain fold is preserved in the variant protein, with slightly reduced thermal stability.

Binding Studies with the Palindromic DNA Duplex CGACTAATTAGTCG (14Dd)

The DNA binding affinities of the wild type Gbx1 and Gbx1[E17R,R52E] homeodomains were measured using isothermal titration calorimetry (ITC). We selected the palindromic DNA duplex CGACTAATTAGTCG (14Dd), because it contains the TAATTA sequence element recognized by Gbx1 and enabled the recording of high-quality NMR spectra. Binding experiments were carried out by titrating 100 μM protein solutions into a 7.5 μM solution of the DNA duplex 14Dd (Fig. 1, c and d). Gbx1 and Gbx1[E17R,R52E] complexation with DNA is exothermic. The titration curves could be fitted to a single-site binding model, and the stoichiometry parameters (N) indicated that one protein molecule binds to one DNA duplex. The binding affinity of 14Dd to Gbx1 is 35 nM, and to Gbx1[E17R,R52E] it is 68 nM. The higher affinity for the wild type protein can be related to the only difference between the two proteins in this region, i.e., the replacement of Arg52 by Glu, since an electrostatic interaction between Arg52 and a phosphate group of the DNA has been observed in NMR structures of other homeodomain–DNA complexes (Billeter et al., 1993; Baird-Titus et al., 2006; Chaney et al., 2005), which would be absent after the Arg52Glu replacement, since Arg17 is separated from DNA contacts by 14 Å.

Formation of a 1:1 Gbx1–DNA complex was directly documented by NMR studies of the binding of the uniformly ^{15}N-labeled Gbx1 homeodomain to unlabelled 14Dd. In the absence of DNA, 69 of the 70 expected backbone amide resonances were observed in 2D [^{15}N,^{1}H]-HSQC spectra. Upon formation of a 1:1 complex, all 70 expected protein resonances were observed, with large chemical shift changes relative to the free homeodomain manifesting the complex formation. At sub-stoichiometric concentrations of 14Dd, the intensity and number of peaks present in the spectrum decreased, indicating intermediate rate exchange on the chemical shift time-scale between free and DNA-bound

Gbx1. These experiments also showed that the Gbx1 homeodomain was amenable for NMR structure determination in both the free and the DNA-bound forms.

NMR Structures of the Gbx1 and Gbx1[E17R,R52E] Homeodomains

NMR structures were determined using the automated J-UNIO protocol (Serrano et al., 2012) in combination with the torsion angle dynamics algorithm CYANA 3.0 (Güntert et al., 1997). Statistical data on the structure determinations are listed in Table 1. The structure of the Gbx1 homeodomain (Fig. 2, a and c) is well defined, with RMSDs of 0.59 and 1.02 Å, respectively, for the backbone and all heavy atoms. The helices α1 and α2 with residues 9–23 and 27–38, are arranged anti-parallel to each other, and helices α2 and α3, with residues 41–60, form a helix-turn-helix motif. Residues 1–6 at the N-terminal are flexibly disordered. 14 residues exhibit solvent accessibility below 15% and form a hydrophobic core of the protein. These include Phe8, Leu16, Phe20, Leu40, Val45, Trp48 and Phe49, which are either invariant or highly conserved in all homeodomains (Banerjee-Basu and Baxevanis, 2001; Chi, 2005; Clarke, 1995).

The structure of the Gbx1[E17R,R52E] homeodomain (Fig. 3) contains the same regular secondary structure elements as Gbx1. Superposition of Gbx1 and Gbx1[E17R,R52E] yields a backbone RMSD value of 0.98 Å and shows differences in the orientation of helix α3 from residue 52 onwards (Fig. 3). Peak doubling was observed in the 3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY spectrum for residues 14 to 23 and 52 to 58 (Figs. S1–S3), indicating that the solution structure of Gbx1[E17R,R52E] includes a dynamic polymorphism, with slow exchange between locally different structures. In the bundle of 20 NMR conformers (Fig. 3a) the Glu52 sidechain adopts two different orientations, which are hence both compatible with the ensemble of observed noe distance constraints; in the two states it interacts, respectably, with the the guanidinium group of Arg17 or the ammonium group of Lys55 (Fig. 3b).

Structure of the Gbx1 Homeodomain–DNA Complex

The solution structure of Gbx1 was also determined in the complex with DNA (Table 1). Although the global fold seen in the Gbx1 homeodomain (Fig 2, a and c) is maintained, there are functionally important differences between the free and DNA-bound structures, as shown by the backbone RMSD value of 1.37 Å calculated for residues 7 to 60. In the DNA-bound Gbx1 homeodomain the helices α1 and α3 are approximately 3.6 Å further apart than in free Gbx1, which ruptures the salt bridge between Glu17 and Arg52. It is worth noting that this salt bridge has always been present in crystal structures of other homeodomain–DNA complexes (Grant et al., 2000; Jauch et al., 2008; Hovde et al., 2001), while it has been reported to be disrupted in solution structures of such complexes (Billeter et al., 1993; Baird-Titus et al., 2006; Chaney et al., 2005) (Fig. S2). In the Gbx1–14Dd complex, line broadening of the NMR signals for residues Asn51 to Ala54 now indicates intermediate rate exchange between two or multiple different conformations of the residues near the C-terminal of helix α3.

To assign the proton resonances of unlabelled 14Dd bound to the ^{13}C,^{15}N-labelled Gbx1 homeodomain, 2D [^{1}H,^{1}H]-TOCSY and ^{13}C/^{15}N doubly-filtered 2D [^{1}H,^{1}H]-NOESY

experiments with τ_m = 90 ms and 180 ms, respectively, were recorded in H_2O and D_2O solutions. The cytosine H5–H6 and thymidine H6–H7 correlations identified using the 2D [^{1}H,^{1}H]-TOCSY experiment were transferred to the ^{13}C/^{15}N doubly-filtered 2D [^{1}H,^{1}H]-NOESY spectrum acquired in D_2O. Assignment of all other non-labile hydrogen atoms was then conducted using a standard protocol (Wüthrich, 1986), establishing NOE-connectivities of the H6/H8 base protons with sugar protons of the subsequent and preceding nucleotides. The labile protons were assigned using a doubly-filtered 2D [^{1}H,^{1}H]-NOESY spectrum acquired in H_2O solution. NOEs between H2″ of the deoxy-ribose and the protons of cytosine and thymidine bases were identified in the ^{13}C/^{15}N doubly-filtered 2D [^{1}H,^{1}H]-NOESY spectrum. All these data support that the DNA adopted a B-DNA conformation in the complex (Wüthrich, 1986). Using the online server 3D-DART (van Dijk et al., 2009), a standard B-DNA model of 14Dd was therefore generated for the docking experiments described below.

In addition to the structure determination of the 14Dd-bound Gbx1 homeodomain and establishing that 14Dd in the complex with this homeodomain adopts a standard B-DNA conformation, we determined a network of 69 intermolecular NOEs between the uniformly ^{13}C,^{15}N-labelled Gbx1 homeodomain and the unlabelled 14Dd, using a 3D F_2-edited/F_3-filtered ^{13}C-resolved [^{1}H,^{1}H]-NOESY spectrum acquired in D_2O solution with τ_m = 240 ms (Zwahlen et al., 1997). The NOE cross peaks were interactively picked and integrated with XEASY (Bartels et al., 1995) for conversion to upper distance restraints.

To dock the structure of the Gbx1 homeodomain onto the 14Dd B-DNA with the aforementioned distance restraints, we used the HADDOCK program (De Vries et al., 2007; Dominguez et al., 2003; Van Dijk et al., 2006). The resulting Gbx1–DNA complex is shown in Fig. 2e, where the amino acid residues exhibiting intermolecular NOEs with the DNA are identified. The Gbx1 homeodomain interacts with the major groove of the DNA primarily through Val43, Ile47, Gln50, Ala54 and Arg58 of helix α3, and with the minor groove through Ser1, Arg2, Arg3 and Arg4 of the N-terminal arm, respectively, which corresponds to observations in previous structure determinations of homeodomain–DNA complexes (Otting et al., 1990; Billeter et al., 1993; Babin et al., 2013, Baird-Titus et al., 2006; Chaney et al., 2005). In addition, the side chains of Leu28, Arg31, Ser32 and Arg52 are orientated in ways that facilitate interactions with the sugar phosphate backbone of 14Dd. A bend in the Asn51 to Ala54 segment of helix α3 enables tighter contacts of the protein with the DNA and leads to the aforementioned disruption of the Glu17–Arg52 salt bridge. Binding to DNA thus induces conformational changes in the homeodomain which result in improved contacts with the major grove of the DNA.

Due to the line broadening for the NMR signals of Asn51 to Ala54, a search for NOEs between Arg52 and the DNA was not conclusive. Such NOEs have previously been observed in the *bicoid* homeodomain–DNA complex (Baird-Titus et al., 2006). It is interesting that NMR line broadening now indicates that there is intermediate-rate exchange between locally different conformations of the recognition helix in the Gbx1 homeodomain when it is bound to the DNA.

DISCUSSION

The residues Glu17 and Arg52 (residue numbering of Torrado et al., 2009) are highly conserved among homeodomains, and crystal structures (Grant et al., 2000; Jauch et al., 2008; Hovde et al., 2001) and molecular dynamics studies (Flader et al, 2003) of homeodomain–DNA complexes suggested that these two residues form a salt bridge in the free homeodomains which is preserved in their DNA complexes. In apparent contrast, interactions between Arg52 and the DNA backbone had either been directly observed in or inferred by NMR structure determinations of different homeodomain–DNA complexes (Baird-Titus et al., 2006; Billeter et al., 1993; Chaney et al., 2005). The present studies with the Gbx1 homeodomain in solution at ambient temperature now revealed an induced fit recognition mechanism that involves conformational changes in the recognition helix, which includes that breaking of the Glu17–Arg52 salt bridge facilitates Arg52–DNA interactions. The effect of Arg52 side chain interactions with the DNA backbone was directly evidenced by a two-fold decrease in DNA binding affinity when Arg52 was replaced by Glu (Fig. 1, c and d). The solution NMR signals further indicated that interactions of the Gbx1 Arg52 side chain with the DNA may be transient and rather lowly populated, which would explain the relatively small effect on the binding affinity. The dynamic equilibrium between locally different conformations of Gbx1 in the presently studied DNA complex may include species that have been seen in the aforementioned earlier structure determinations of different homeodomain–DNA complexes, and thus also offers a rationale for the observation of the intact 17–52 salt bridge in crystal structures at low temperature. With their short duration, MD simulations (e.g., Babin et al., 2013; Billeter et al., 1996; Flader et al., 2003) explored conformational changes on the pico- to nano-second time scale and may therefore have missed to register Arg52–DNA interactions. Overall, continued studies of the homeodomain–DNA system provide new insight into the structural basis and mechanisms of DNA recognition by proteins, which may be applied to a wide range of physiological processes. While this may be far fetched, it could even include DNA recognition in DNA repair (Cleaver, 2016; Lindahl, 2013; Modrich, 2006; Sancar 1996). The results obtained here also demonstrate the importance of integrative approaches in structural biology which include experimental measurements in solution at ambient temperature.

EXPERIMENTAL PROCEDURES

Protein Production and Purification

An *H. sapiens* Gbx1 construct cloned into a peT23a vector was obtained from Dr. Fumiaki Yumoto and Dr. Robert Fletterick of UCSF within the framework of a PSI:Biology partnership with the JCSG. Mutagenesis of Glu17 to Arg and Arg52 to Glu was performed using QuickChange site-directed mutagenesis (Stratagene) and the following primers. Mutated nucleotides are in bold typeface. Primer 23: 5′ gagcagcttttggaattg**cg**gaaggaatttcattgcaag 3′ and Primer 58: 5′ caagatctggtttcagaat**ga**acgggccaagtggaagc 3′.

Expression of uniformly ^{13}C,^{15}N-labelled Gbx1 and Gbx1[E17R,R52E], in the *E. coli* strain BL21(DE3) (Novagen), was carried out by growing the cells in M9 medium containing [$^{13}C_6$]-D-glucose (4 g/L) and $^{15}NH_4Cl$ (1 g/L) as the sole carbon and nitrogen sources

respectively. Cell cultures were grown at 37 °C with vigorous shaking, to an optical density at 600 nm of 0.6, and the temperature was reduced to 18 °C before expression of the proteins were induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside. The cells were then grown at 18 °C for 20 hours, harvested by centrifugation, resuspended in buffer A (50 mM TRIS-HCL at pH 8.0, 300 mM Sodium Chloride, 20 mM imidazole, 5 mM β-mercaptoethanol, 10 % Glycerol), supplemented with Complete EDTA-free protease inhibitor cocktail tablets (Roche), and lysed by sonication. The cell debris was removed by centrifugation (40,000 × g for 30 minutes) and the supernatant was loaded onto a Ni^{2+} affinity column (HisTrap HP; GE Healthcare) equilibrated with buffer A. The imidazole concentration was increased to 38 mM in order to remove non-specifically bound proteins and then to 300 mM to elute Gbx1. Fractions identified to contain the homeodomain protein, as determined by SDS-PAGE, were digested for 1 hour with 0.08 mg/ml TEV protease at room temperature. Digested proteins were loaded onto a size exclusion column (Superdex^{TM} 75 HiLoad^{TM} 26/60; GE Healthcare) equilibrated with buffer B (20 mM sodium phosphate pH 6.0, 50 mM sodium chloride) and eluted with the same buffer. Fractions containing Gbx1 or Gbx1[E17R,R52E] were pooled and further purified using an AC IEX column (GE Healthcare) equilibrated with buffer B and the proteins were eluted with a 500 mM to 1 M sodium chloride gradient. To buffer exchange the protein back into buffer B, fractions containing Gbx1 or Gbx1[E17R,R52E] were concentrated to 500 μl and diluted 20-fold with buffer B, using 3 kDa cut-off centrifugal filter devices (Millipore). This process was repeated three times. Gbx1 and Gbx1[E17R,R52E] were concentrated to a final protein concentration of 1.2 mM, and for the NMR experiments the protein samples were supplemented with 5% $^{2}H_2O$ (v/v) and 4.5 mM sodium azide.

NMR Spectroscopy

NMR data used for the structure determinations were recorded on Bruker AVANCE 600 MHz and AVANCE 800 MHz spectrometers equipped with 5 mm TXI-HCN z- or xyz-gradient probes. The structures of Gbx1 in the absence and presence of DNA were determined at 298K and 308K respectively. At 600 MHz we recorded a 2D [^{15}N,^{1}H]-HSQC spectrum and the experiments 4D APSY-HACANH, 5D APSY-HACACONH and 5D APSY-CBCACONH (Hiller et al., 2005; Hiller et al., 2008), which were used for the polypeptide backbone assignments, and 2D [^{1}H,^{1}H]-TOCSY and ^{13}C/^{15}N doubly-filtered 2D [^{1}H,^{1}H]-NOESY experiments recorded with τ_m = 90 ms and 180 ms, respectively, which were used for the assignment of DNA protons. 3D ^{15}N-resolved, 3D ^{13}C(aliphatic)-resolved and 3D ^{13}C(aromatic)-resolved [^{1}H,^{1}H]-NOESY experiments were recorded with τ_m = 80 ms and 70 ms for the free and DNA-bound proteins, respectively. A 3D F_2-edited, F_3-filtered ^{13}C-resolved [^{1}H,^{1}H]-NOESY experiment (Zwahlen et al., 1997) was recorded at 800 MHz with τ_m = 240 ms to identify protein–DNA intermolecular NOEs. The NOEs were manually picked and integrated using XEASY (Bartels et al., 1995), and then converted into distance restraints. The 2D [^{15}N,^{1}H]-HSQC spectra used to study binding of Gbx1 to the 14-base pair DNA helix were recorded on a Bruker DRX 700 MHz spectrometer equipped with a 1.7-mm microcoil probehead. Proton chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonic acid sodium salt (DSS). The ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS, using the absolute frequency ratios (Wishart et al., 1995).

NMR Structure Determination

The structure determination of free and DNA-bound Gbx1, and of Gbx1[E17R,R52E] followed the J-UNIO protocol (Serrano et al., 2012), with automated assignments performed using the MATCH (Volk et al., 2008) and ATNOS/ASCAN (Fiorito et al., 2008; Herrmann et al., 2002b) algorithms. Structures were determined using the seven-cycle ATNOS/CANDID protocol with the torsion angle dynamics algorithm CYANA-3.0 (Güntert et al., 1997; Herrmann et al., 2002a, b). The 40 conformers with the lowest residual CYANA target function values obtained from the seventh cycle of ATNOS/CANDID/CYANA cycle were subjected to energy minimisation in a water shell with the program OPALp (Luginbuhl et al., 1996), using the AMBER force field (Cornell et al., 1995; Koradi et al., 2000). After structure validation, 20 conformers were selected to represent the NMR structure, and the program MOLMOL (Koradi et al., 1996) was used to analyse these ensembles.

PDB accession numbers

The chemical shifts of Gbx1, Gbx1[E17R,R52E] and DNA-bound Gbx1 have been deposited in the Biological Magnetic Resonance Bank with the accession codes 18944, 25849 and 19511. The atomic coordinates for the bundles of 20 conformers used to represent the solution structures of Gbx1, Gbx1[E17R,R52E] and DNA-bound Gbx1, and for a bundle of 4 conformers used to represent the structure of the Gbx1–DNA complex have been deposited in the Protein Data Bank with the accession codes 2M34, 2N8G, 2ME0 and 2ME6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Joint Center for Structural Genomics (JCSG) through the NIH Protein Structure Initiative (PSI) grant U54 GM094586 from the National Institute of General Medical Sciences (www.nigms.nih.gov). Kurt Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at The Scripps Research Institute. We thank Dr. Robert J. Fletterick and Dr. Fumiaki Yumoto for providing us, under the auspices of a PSI:Biology partnership with the JCSG, with a Gbx1 construct from which the present work was started.

Abbreviations

APSY	automated projection spectroscopy
HSQC	heteronuclear single quantum coherence spectroscopy
JCSG	Joint Center for Structural Genomics
NOESY	nuclear Overhauser effect spectroscopy
PDB	protein data bank
RMSD	root-mean-square deviation
ASCAN	software for automated side-chain resonance assignment

ATNOS	software for automated NMR peak picking
CANDID	software for automated NOE assignment
CYANA	software used for NMR structure calculation
EDTA	ethylenediaminetetraacetic acid
MATCH	software used for backbone chemical shift assignments
NOE	nuclear Overhauser effect
TEV	tobacco etch virus

References

Babin V, Wang D, Rose RB, Sagui C. Binding polymorphism in the DNA bound state of the Pdx1 homeodomain. Plos Comput Biol. 2013; 9:e1003160. [PubMed: 23950697]

Baird-Titus JM, Clark-Baldwin K, Dave V, Caperelli CC, Ma J, Rance M. The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site. J Mol Biol. 2006; 356:1137–1151. [PubMed: 16406070]

Banerjee-Basu S, Baxevanis AD. Molecular evolution of the homeodomain family of transcription factor. Nucleic Acids Res. 2001; 29:3258–3269. [PubMed: 11470884]

Bartels C, Xia TH, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. J Biomol NMR. 1995; 6:1–10. [PubMed: 22911575]

Billeter M, Qian Y, Otting G, Müller M, Gehring W, Wüthrich K. Determination of the three-dimensional structure of the *Antennapedia* homeodomain from Drosophila in solution by ^{1}H nuclear magnetic resonance spectroscopy. J Mol Biol. 1990; 214:183–197. [PubMed: 2164583]

Billeter M, Qian Y, Otting G, Müller M, Gehring W, Wüthrich K. Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain-DNA complex. J Mol Biol. 1993; 234:1084–1094. [PubMed: 7903398]

Billeter M, Güntert P, Luginbühl P, Wüthrich K. Hydration and DNA recognition by homeodomains. Cell. 1996; 85:1057–1065. [PubMed: 8674112]

Buckley DM, Burroughs-Garcia J, Lewandoski M, Waters ST. Characterization of the Gbx1 −/− mouse mutant: A requirement for Gbx1 in normal locomotion and sensorimotor circuit development. Plos One. 2013; 8:e56214. [PubMed: 23418536]

Chaney BA, Clark-Baldwin K, Dave V, Ma J, Rance M. Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome. Biochemistry. 2005; 44:7497–7511. [PubMed: 15895993]

Clarke ND. Covariation of residues in the homeodomain sequence family. Prot Sci. 1995; 4:2269–2278.

Cleaver JE. Profile of Tomas Lindhal, Paul Modrich, and Aziz Sancar, 2015 Nobel laureates in chemistry. Proc Natl Acad Sci USA. 2016; 113:242–245. [PubMed: 26715755]

Chi YI. Homeodomain revisited: a lesson from disease-causing mutations. Hum Genet. 2005; 116:433–444. [PubMed: 15726414]

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A 2nd generation force-field for the simulation of proteins, nucleic-acids and organic-molecules. J Am Chem Soc. 1995; 117:5179–97.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WWB III, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl Acid Res. 2007; 35:W375–W383.

De Vries SJ, van Dijk ADJ, Krzeminski M, van Dijk M, Thureau A, Hsu V, et al. HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. Proteins: Struct Funct Bioinf. 2007; 69:726–733.

Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003; 125:1731–1737. [PubMed: 12580598]

Flader W, Wellenzohn B, Winger RH, Hallbrucker A, Mayer E, Liedl KR. Stepwise induced fit in the pico- to nanosecond time scale governs the complexation of the even-skipped transcriptional repressor homeodomain to DNA. Biopolymers. 2003; 68:139–149. [PubMed: 12548619]

Fiorito F, Herrmann T, Damberger FF, Wüthrich K. Automated amino acid side-chain NMR assignment of proteins using 13C- and 15N-resolved 3D 1H,1H-NOESY. Journal of Biomolecular NMR. 2008; 42:23–33. [PubMed: 18709333]

Grant RA, Rould MA, Klemm JD, Pabo CO. Exploring the role of glutamine 50 in the homeodomain-DNA interface: Crystal structure of engrailed (Gln50 -> Ala) complex at 2.0 angstrom. Biochemistry. 2000; 39:8187–8192. [PubMed: 10889025]

Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol. 1997; 273:283–298. [PubMed: 9367762]

Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR. 2002; 24:171–189. [PubMed: 12522306]

Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol. 2002; 319:209–227. [PubMed: 12051947]

Hiller S, Fiorito F, Wüthrich K, Wider G. Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA. 2005; 102:10876–10881. [PubMed: 16043707]

Hiller S, Wider G, Wüthrich K. APSY-NMR with proteins: practical aspects and backbone assignment. J Biomol NMR. 2008; 42:179–195. [PubMed: 18841481]

Hovde S, Abate-Shen C, Geiger JH. Crystal structure of the Msx-1 homeodomain/DNA complex. Biochemistry. 2001; 40:12013–12021. [PubMed: 11580277]

Jauch R, Ng CKL, Saikatendu KS, Stevens RG, Kolatkar PR. Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog. J Mol Biol. 2008; 376:758–770. [PubMed: 18177668]

John A, Wildner H, Britsch S. The homeodomain transcription factor Gbx1 identifies a subpopulation of late-born GABAergic interneurons in the developing dorsal spinal cord, Dev. Dyn. 2005; 234:767–771.

Koradi R, Billeter M, Wüthrich K. MOLMOL: A program for display and analysis of macromolecular structures. J Mol Graph. 1996; 14:51–55. [PubMed: 8744573]

Koradi R, Billeter M, Guntert P. Point-centered domain decomposition for parallel molecular dynamics simulation. Comput Phys Commun. 2000; 124:139–47.

Lindahl T. My journey to DNA repair. Gen Prot Bioninf. 2013; 11:2–7.

Longo A, Guanga GP, Rose RB. Structural basis for induced fit mechanisms in DNA recognition by the Pdx1 homeodomain. Biochemistry. 2007; 46:2948–2957. [PubMed: 17315980]

Luginbühl P, Güntert P, Billeter M, Wüthrich K. The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR. 1996; 8:136–146. [PubMed: 8914272]

Meziane H, Fraulob V, Riet F, Krezel W, Selloum M, Geffarth M, et al. The homeodomain factor Gbx1 is required for locomotion and cell specification in the dorsal spinal cord. PeerJ. 2013; 1:e142. [PubMed: 24010020]

Modrich P. Mechanisms in eukariotic mismatch repair. J Biol Chem. 2006; 281:30305–30309. [PubMed: 16905530]

Otting G, Qian Y, Billeter M, Müller M, Affolter M, Gehring W, et al. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. 1990; 9:3085–3092. [PubMed: 1976507]

Qian Y, Billeter M, Otting G, Müller M, Gehring W, Wüthirch K. The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell. 1989; 61:573–780.

Rhinn M, Lun K, Werner M, Simeone A, Brand M. Isolation and expression of the homeobox gene Gbx1 during mouse development. Dev Dyn. 2004; 229:334–339. [PubMed: 14745958]

Sancar A. DNA excision repair. Annu Rev Biochem. 1996; 65:43–81. [PubMed: 8811174]

Serrano P, Pedrini B, Mohanty B, Geralt M, Herrmann T, Wüthrich K. The J-UNIO protocol for automated protein structure determination by NMR in solution. J Biomol NMR. 2012; 53:341–354. [PubMed: 22752932]

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7:539. [PubMed: 21988835]

Torrado M, Revuelta J, Gonzalez C, Corzana F, Bastida A, Asensio JL. Role of conserved salt bridges in homeodomain stability and DNA binding. J Biol Chem. 2009; 284:23765–23779. [PubMed: 19561080]

van Dijk M, van Dijk ADJ, Hsu V, Boelens R, Bonvin AMJJ. Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. Nucleic Acids Res. 2006; 34:3317–3325. [PubMed: 16820531]

van Dijk M, Bonvin AMJJ. 3D-DART: a DNA structure modelling server. Nucleic Acids Res. 2009; 37:W235–W239. [PubMed: 19417072]

Volk J, Herrmann T, Wüthrich K. Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. Journal of Biomolecular NMR. 2008; 41:127–38. [PubMed: 18512031]

Waters ST, Wilson CP, Lewandoski M. Cloning and embryonic expression analysis of the mouse Gbx1 gene, Gene Expr. Patterns. 2003; 3:313–317.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, et al. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. Journal of Biomolecular NMR. 1995; 6:135–40. [PubMed: 8589602]

Wüthrich, K. NMR of Proteins and Nucleic Acids. New York: Wiley; 1996.

Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. J Am Chem Soc. 1997; 119:6711–672.

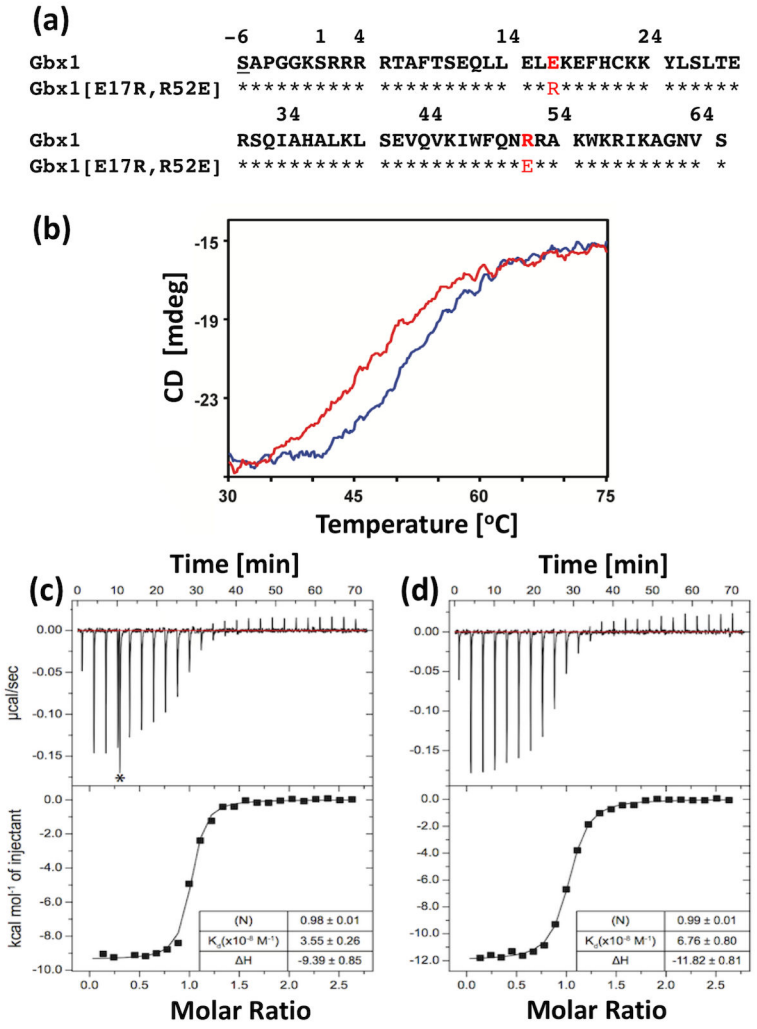

Figure 1.

Amino acid sequences, temperature denaturation and DNA-binding affinities of the Gbx1 and Gbx1[E17R,R52E] homeodomains. (a) Amino acid sequence of the 71-residue Gbx1 construct used. For easy comparison with other homeodomains, the numeration starts with the seventh residue (see the text and Torrado et al., 2009). In free Gbx1, Glu17 and Arg52 (highlighted in red) form a salt bridge. The positions of the α-helices in the NMR structure of Gbx1 are indicated by orange rectangles. (b) Denaturation curves of Gbx1 (blue) and Gbx1[E17R,R52E] (red). The temperature variation of the signal intensity was monitored by the CD intensity at 209 nm. (c) and (d): ITC studies of the binding of the Gbx1 (c) and Gbx1[E17R,R52E] (d) homeodomains to the 14Dd DNA duplex. The top panels show observed heats measured for 24 protein injections at 175-second intervals into a solution containing the DNA. The bottom panels show binding enthalpies calculated with the appropriate molar correction from the individual integrated heat profiles (the data from the fourth injection in (c), indicated with *, was omitted from the analysis because a sharp spike interfered with the integration). The numerical data listed in the bottom panels resulted from fits to a one-site binding model. For additional details, see supporting material.

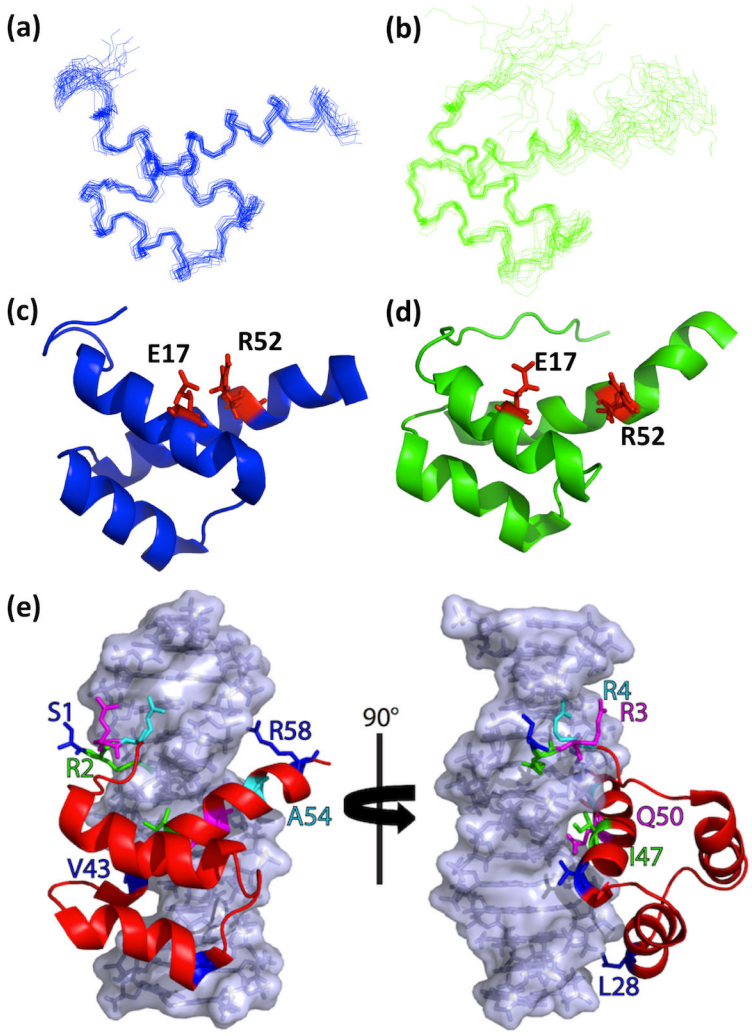

Figure 2.

Solution structures of the Gbx1 homeodomain in the free state and bound to DNA, and of the Gbx1–DNA complex. (a) Backbone of the free Gbx1 homeodomain. A bundle of 20 energy-minimized CYANA conformers, showing the complete protein with residues 1 to 60 (Fig. 1a) superimposed for best fit of residues 6–60 is shown. (b) DNA-bound Gbx1. Same presentation as in (a). (c) and (d) Ribbon presentations of the conformers closest to the mean co-ordinates in (a) and (b), with the side chains of Glu17 and Arg52 shown as red sticks: (c) free Gbx1. (d) DNA-bound Gbx1. (e) Gbx1–DNA complex. The homedomain is shown as a ribbon representation of the backbone and sticks for the side chains of the residues exhibiting intermolecular NOEs, i.e., S1, R2, R3, R4, L28, V43, I47, Q50, A54 and R58. The DNA is in an all-atom space-filling presentation. The drawings have been generated with MOLMOL (Koradi et al., 1996).

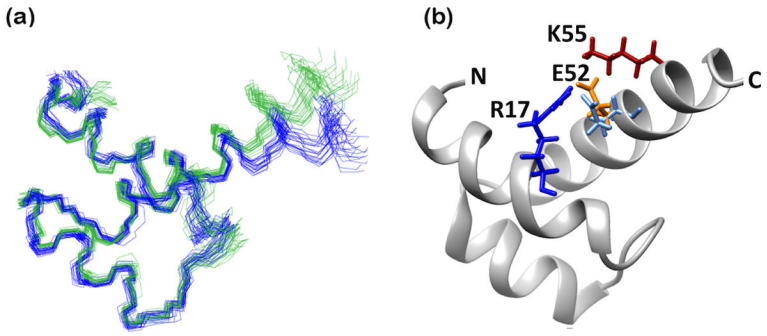

Figure 3.

NMR structure of the Gbx1[E17R,R52E] homeodomain and comparison with Gbx1. (a) Superposition of two bundles of 20 NMR conformers, showing the residues 7–60 of the Gbx1 (blue) and Gbx1[E17R,R52E] (green) homeodomains, which had independently been superimposed for best fit. (b) Ribbon presentation of the conformer closest to the mean coordinates of the Gbx1[E17R,R52E] bundle in (a). The Arg17 (blue) and Lys55 (red) residues are shown in stick representations. Two different orientations of the Glu52 side chain are shown in light blue and orange. These are both compatible with the experimental constraints and enable salt bridge formation with Arg17 or Lys55, respectively. The chain ends and the residues discussed in the text are identified.

Table 1

Input for the structure calculation and characterisation of the ensemble of 20 energy-minimised CYANA conformers used to represent the NMR structures of the Gbx1 homeodomain in the free state and bound to DNA, and the variant Gbx1[E17R,R52E] homeodomain.

Quantity[a]	Gbx1 (298 K)	Gbx1[E17R,R52E] (298K)	DNA-bound Gbx1 (308K)
NOE upper distance limits	1222	1126	897
Intraresidual	354	346	310
Short range	299	283	251
Medium range	374	342	233
Long range	195	155	103
Dihedral angle constraints	294	289	293
Residual target function value (Å ^{2})	0.66 ± 0.07	0.76± 0.09	0.87 ± 0.15
Residual NOE violations			
Number ≥ 0.1 A	5 ± 2	2 ± 1	0 ± 1
Maximum (Å)	0.12	0.13	0.17
Residual dihedral angle violations			
Number ≥ 2.5°	0 ± 0	0 ± 0	0 ± 0
Maximum (°)	1.34	1.41	0.87
Amber energies (kcal/mol)			
Total	−1878 ± 95	−1824 ± 125	−1580 ± 140
Van der Waals	−200 ± 8	−170 ± 15	−146 ± 12
Electrostatic	−2221 ± 97	−2234 ± 139	−1988 ± 145
RMSD from ideal geometry			
Bond lengths (Å)	0.0075 ± 0.0002	0.0076 ± 0.0002	0.0078 ± 0.0001
Bond angles (°)	1.81 ± 0.06	1.92 ± 0.06	1.94 ± 0.05
RMSD to the mean co-ordinates (Å) [b]			
bb Gbx1 (7–60); Gbx1[E17R,R52] (7–60); DNA-bound Gbx1 (7–60)	0.53 ± 0.08	0.55 ± 0.11	1.22 ± 0.44
ha Gbx1 (7–60); Gbx1[E17R,R52] (7–60); DNA-bound Gbx1 (7–60)	0.99 ± 0.10	1.09 ± 0.18	1.77 ± 0.54
Ramachandran plot statistics (%) [c]			
Favoured regions	87.0	86.9	83.1
Allowed regions	98.6[d]	97.8[e]	98[f]

[a] Except for the top six entries, which describe the input generated in the final cycle of the CYANA structure calculation (Serrano et al., 2012; Güntert et al., 1997; Herrmann et al., 2002a,b), the entries refer to the 20 best CYANA conformers after energy minimization with OPALp [34] (see text). Where applicable, the average value for the bundle of 20 conformers and the standard deviation are given. The residue numbering of Torrado et al. (2009) is used (see also Fig. 1a).

[b] bb indicates the backbone atoms N, C^{α}, and C′; ha stands for all heavy atoms. Numbers in parentheses indicate the residues for which the RMSD was calculated

[c] As determined by Molprobity (Davis et al., 2007)

[d] Residues which are outside of the allowed regions in at least one of the 20 conformers include -2, -1, 4, 5, 59, 60, 62, 64

[e] Residues which are outside of the allowed regions in at least one of the 20 conformers include 1,2,3, 5, 7, 61

[f] Residues which are outside of the allowed regions in at least one of the 20 conformers include 1, 2, 28, 39, 62