ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/11984497

Solution Structure of the DNA Binding Domain of the Human Forkhead Transcription Factor AFX (FOXO4) †

ARTICLE *in* BIOCHEMISTRY · JUNE 2001

Impact Factor: 3.02 · DOI: 10.1021/bi001663w · Source: PubMed

CITATIONS
57

READS
22

4 AUTHORS, INCLUDING:

Mats Wikström
Amgen
36 PUBLICATIONS **1,217** CITATIONS

SEE PROFILE

Available from: Mats Wikström
Retrieved on: 29 January 2016

Articles

Solution Structure of the DNA Binding Domain of the Human Forkhead Transcription Factor AFX (FOXO4)†

Johan Weigelt,* Isabel Climent, Karin Dahlman-Wright,‡ and Mats Wikström

Department of Structural Chemistry, Biovitrum Division, Pharmacia, S-112 87 Stockholm, Sweden

Received July 17, 2000; Revised Manuscript Received March 6, 2001

ABSTRACT: AFX is a human forkhead transcription factor. Based on results from studies of the orthologous transcription factor DAF-16 in *Caenorhabditis elegans*, it was suggested that some of the metabolic defects in both type I and type II diabetes may be due to unregulated activity of AFX. In the present study, we report the high-resolution NMR solution structure of the DNA binding domain of AFX. It is the first structure of the DNA binding domain from a small subfamily of forkhead transcription factors (i.e., AFX, FKHR, FKHRL1, FKHRL1P1, and FKHRP1). Despite rather low sequence identity for a protein within the forkhead family, the structure is remarkably similar to those of the DNA binding domains of HNF3-γ and FREAC-11, and to a lesser extent the DNA binding domain of Genesis which displays a slightly altered orientation of the DNA recognition helix. The high degree of structural similarity between the DNA binding domains of different forkhead transcription factors implies that the repositioning of helix 3, observed for Genesis, cannot be a general feature for modulation of the DNA binding specificity. Other mechanisms that could influence the DNA binding specificity are discussed.

Forkhead transcription factors serve as regulatory keys in embryogenesis, tumorigenesis and maintenance of differentiated cell states (*1*). The conserved forkhead DNA binding domain (which belongs to the winged helix superfamily of proteins) encompasses about 100 amino acid residues. Three major α-helices are packed against each other, resting on a small three-stranded antiparallel β-sheet from which two loops ("wings") protrude. The highest degree of sequence conservation among forkhead domains is found in three helices. Helix 3 has been identified as being responsible for most of the direct base contacts with DNA, although other contacts with DNA have also been observed (*2, 3*).

AFX (*human*) belongs to a small subfamily (AFX, FKHR, FKHRL1, FKHRL1P1, and FKHRP1) of forkhead transcription factors (*4*). Its designation according to the recently introduced nomenclature for winged helix/forkhead transcription factors is FOXO4 (*5*). Studies performed in *Caenorhabditis elegans* revealed that the orthologous transcription factor DAF-16 is involved in an insulin-like signaling pathway (*6, 7*). Based on these results, it was suggested that some of the metabolic defects caused by declines in insulin signaling, in both type I and type II diabetes, may be due to unregulated activity of AFX and/or any of the other DAF-16 orthologues (*6*).

AFX has been shown to be associated with oncogenesis. A chromosomal translocation that results in gene fusions of the mixed-lineage leukemia transcription factor (MLL)[1] and AFX is observed in a subtype of acute leukemia (*8*). A number of other forkhead transcription factors have also been implicated in tumorigenesis, and similar translocations have been identified (*1*). Furthermore, recent evidence shows that AFX-like transcription factors are involved in cell cycle regulation (*9*).

The transcriptional activity of AFX is regulated by phosphorylation, and it has been demonstrated that in response to insulin, AFX is phosphorylated through a phosphatidylinisitol-3-OH-kinase/protein kinase B [PI(3)K/PKB] pathway (*10*). An identical insulin-dependent phosphorylation process has been observed for FKHR (*11*). Lack of PKB activity does not fully inhibit phosphorylation of AFX, and a second Ras/Ra1 insulin-dependent pathway was also identified (*10*). Phosphorylation of AFX results in repression of its transcriptional activity. This is achieved, at least in part, by translocation of AFX from the cell nucleus to the cytoplasm (*12*). It has been speculated whether the transcriptional repression of AFX and FKHR occurs at more than one level; i.e., the phosphorylation might also affect the DNA binding properties of the proteins (*13, 14*).

† The coordinates for the 20 NMR conformers of the DNA binding domain of AFX have been deposited at the Protein Data Bank (accession code 1E17). The ^{1}H, ^{13}C, and ^{15}N chemical shifts have been deposited at the BioMagResBank (http://www.bmrb.wisc.edu/) database (accession number 4675).

* Correspondence should be addressed to this author. Phone: +46 (0)8 697 24 94. Fax: +46 (0)8 697 23 19. Email: Johan.Weigelt@biovitrum.com.

‡ Present address: Department of Biosciences, Karolinska Institutet, S-141 57 Huddinge, Sweden.

[1] Abbreviations: MLL, mixed-lineage leukemia transcription factor; PI(3)K, phosphatidylinositol-3-OH-kinase; PKB, protein kinase B; NOE, nuclear Overhauser effect; dDNA, duplex DNA; IRS, insulin response sequence.

10.1021/bi001663w CCC: $20.00 © 2001 American Chemical Society
Published on Web 04/25/2001

A wealth of data have been reported for forkhead transcription factors. For a review, see Kaufmann and Knöchel (*1*). Three-dimensional structures, however, have so far only been determined for the DNA binding domains of HNF3-γ, Genesis, and FREAC-11 (*2*, *3*, *15*, *16*). The present work, describing the NMR solution structure of the DNA binding domain of AFX, is the first report of a high-resolution structure of the DNA binding domain from an AFX-like forkhead transcription factor.

MATERIALS AND METHODS

Plasmid Construct. The DNA binding domain of the human Afx gene comprising amino acid residues G82−A207 (hAfxDBD) was amplified by PCR using AfxDBD5′ (GACGACGGATCCGGGGGCTGTAACAGGTCCTC) and AfxDBD3′ (GACGACGGATCCTCAGGCTTTACTGCGGCCCCG) as primers. The amplification was done using 25 cycles of 95 °C 15 s denaturing, 57 °C 30 s annealing followed by a 72 °C 1 min extension, so that the fragment contained a *Bam*HI site at the 5′- and 3′-ends. The PCR fragment was then digested and gel purified, and finally subcloned into the *Bam*HI site of pET15b (Novagen) in-frame with the His6-tag to produce pHis-Afx-DBD. Restriction analysis and DNA sequencing were used to confirm the sequence.

Overexpression and Purification of the DNA Binding Domain of AFX. Escherichia coli cells BL21(DE3) (Novagen) harboring pHis-Afx-DBD was grown in minimal medium with $(^{15}NH_4)_2SO_4$ and 100% ^{13}C-labeled and/or unlabeled glucose as the sole nitrogen and carbon sources. Three different isotopically enriched (U-^{15}N labeled, U-^{13}C/^{15}N double labeled, and 10%-^{13}C/U-^{15}N partially double labeled) samples of AFX-DBD were prepared. The resulting protein construct contained 24 additional residues (*his-tag*, GSSHHHHHHSSGLVPRGSHMLEDP) at the N-terminus included for purification purposes and not removed.

Cells were harvested by centrifugation and resuspended (10 mL/g of cells) in buffer A (20 mM sodium phosphate, 0.5 M NaCl, pH 7.4) supplemented with 1 mg/mL lysozyme (Sigma) and Complete (Boehringer) protease inhibitor. The suspension was stirred at 4 °C for 90 min, followed by further lysis using sonication. The lysate was clarified by centrifugation, and the supernatant was loaded onto a Ni^{2+}-charged 5 mL HiTrap chelating column (Amersham Pharmacia Biotech) equilibrated in buffer A. Following an initial column wash with buffer A containing 50 mM imidazole, proteins were eluted using a linear gradient of imidazole (50−600 mM in 10 column volumes). Protein fractions of >95% purity (as determined by SDS−PAGE) were pooled and dialyzed against NMR sample buffer [20 mM sodium phosphate, 100 mM NaCl, 0.02% (w/v) NaN_3, 1 mM EDTA, pH 6.3]. Further concentration of the protein was carried out using an Ultrafree centrifugal filter unit (Millipore).

Preparation of dDNA Samples. The DNA duplex used in NMR interaction studies was as follows: IRS_A d-5′-CAAGCAAAACAAACCA-3′ and IRS_B d-5′-TGTTCAAAATAATCTG-3′. Single-stranded oligonucleotides (Midland Certified Reagent Co.) were dissolved in NMR sample buffer and annealed following standard protocols.

NMR Spectroscopy. All NMR measurements were performed in either H_2O or D_2O at 31 °C on Varian UNITY Inova 600 and 800 MHz NMR spectrometers. The protein concentration was approximately 1.5 mM, and sample buffers were supplemented with 1 mM Pefabloc protease inhibitor (Boehringer). H_2O samples contained 10% (v/v) D_2O. Data were processed using nmrPipe/nmrDraw (*17*) and analyzed in ANSIG (*18*). Azara (provided by Wayne Boucher and the Department of Biochemistry, University of Cambridge) was used to prepare spectral plots.

The following spectra were recorded to obtain resonance assignments: 2D ^1H-^{15}N HSQC (*19*), 3D NOESY-^{15}N-HSQC (*20*), 3D HN(CO)CA (*22*), 3D HN(CA)CB (*21*), 3D HNCO (*21*), 3D HNHA (*23*), 3D HNHB (*24*), 3D HCCH-TOCSY (D_2O) (*25*), 3D NOESY-^{13}C-HSQC (D_2O) (*26*), 3D NOESY-^{13}C-HSQC (aromatics) (*26*), 2D TOCSY-relayed-CT-^{13}C-HMQC (D_2O, aromatics) (*27*), 2D CT-^{13}C-HSQC (D_2O, aromatics), and 2D CT-^{13}C-HSQC (methyls) (*28*). Stereospecific assignments for β-methylene protons were obtained from qualitative $^3J_{N,H\beta}$ couplings and short-range NOE patterns. Stereospecific assignments for isopropyl groups were obtained from the CT-^{13}C-HSQC spectrum recorded with the 10%-^{13}C/U-^{15}N labeled sample (*29*).

NMR distance restraints were collected from three different NOESY spectra: 3D NOESY-^{15}N-HSQC (τ_m = 40 ms), 3D NOESY-^{13}C-HSQC (τ_m = 40 ms), and 3D NOESY-^{13}C-HSQC (optimized for aromatic residues) (τ_m = 60 ms). $^3J_{HN\alpha}$ couplings were measured in the 3D HNHA spectrum.

^{15}N-{^1H} steady-state NOE measurements were performed using the pulse sequence described by Farrow et al. (*30*).

2D ^1H-^{15}N-HSQC (*19*) spectra were recorded at 20 °C on 1:1 mixtures of AFX-DBD with dDNA at 0.1 mM to verify complex formation of AFX-DBD with IRS_A and IRS_B.

NMR Structure Calculation. The NMR solution structure of AFX-DBD was calculated using the program X-PLOR v. 3.84 (*31*) modified to include terms for $^3J_{HN\alpha}$ couplings (*32*) and a conformational database refinement (*33*). The resonance lines of approximately 40 residues at the N-terminus (his-tag + ~15) and 30 residues at the C-terminus were very narrow and displayed no long-range NOE cross-peaks. Therefore, only residues Ser92−Gly181 were included in the structure calculations. Distance restraints were grouped into four distance ranges: strong, 1.8−2.8 Å; medium, 1.8−3.6 Å; weak, 1.8−5.0 Å; and very weak, 1.8−6.0 Å. Center averaging (*31*) was used, and consequently pseudo-atom corrections had to be applied for distances involving methyl protons, aromatic ring protons, and nonstereospecifically assigned methylene protons (*34*). Hydrogen bond restraints were employed in areas of regular secondary structure displaying characteristic NOE cross-peaks. Two distance restraints were used for each hydrogen bond (1.6 Å < d_{HN-O} < 2.5 and 2.6 Å < d_{N-O} < 3.5 Å). For $^3J_{HN\alpha}$ values smaller than 5 Hz, restraints for the ϕ angle were added as −60° ± 20°. $^3J_{HN\alpha}$ values greater than 8 Hz yielded ϕ angle restraints of −120° ± 40°. In addition, the $^3J_{HN\alpha}$ values were also included directly in the refinement protocol. Further restraints for ϕ and ψ were added based on the consensus chemical shift index: helical residues, −60° ± 40° (ϕ) and −50° ± 50° (ψ), β-strand residues, −120° ± 40° (ϕ) and 130° ± 50° (ψ). The χ_1 angle was restrained, when possible, to one or more of the intervals −60° ± 60°, 60° ± 60°, or 180° ± 60° based on qualitative $^3J_{NH\beta}$ couplings and short-range NOE patterns [H^{β}_i−H^{α}_i, H^{β}_i−H^{N}_i, and H^{β}_i−$H^{N}_{(i+1)}$].

Table 1: Structural Statistics for the DNA Binding Domain of AFX[a]

assigned NOE cross-peaks	2680
nonredundant distance restraints	1296
dihedral angle restraints	163
hydrogen bonds	23
residual NOE restraint violations	
number of violations >0.1 Å	15 ± 3
sum (Å)	4.4 ± 0.4
maximum (Å)	0.27 ± 0.07
RMS deviations from idealized covalent geometry	
bonds (Å)	0.0024 ± 0.0001
angles (deg)	0.40 ± 0.02
impropers (deg)	0.40 ± 0.02
Lennard−Jones potential energy (kcal mol^{-1})	−328 ± 15
coordinate precision (RMS deviation to the mean)[b]	
backbone atoms N, C$^{\alpha}$, C′ (Å)	0.50 ± 0.12
all heavy atoms (Å)	0.88 ± 0.11
Ramachandran plot appearance[c,d]	
most favored regions (%)	90.3
additionally allowed regions (%)	8.7
generously allowed regions (%)	0.8
disallowed regions (%)	0.1

[a] Only residues 92−181 were included in the structure calculation. [b] Residues 102−136, 145−163, and 173−176. [c] Residues 100−176. [d] From PROCHECK−NMR (*36*).

The structures were calculated using a simulated annealing protocol for torsion angle dynamics (*35*) modified to include conformational database refinement. The conformational database potential was included already in the first steps of torsion angle dynamics, and its weight in the final Powell minimization steps was $w_{ramadb} = 0.5$ ($w_{NOE} = 30$). Twenty conformers were selected from a total of 85 calculated based on the overall X-PLOR energy.

PROCHECK (*36*) was used for structural analysis. Graphic representations of the three-dimensional structures were prepared using the program MOLMOL v. 2.6 (*37*).

RESULTS AND DISCUSSION

Structure of the DNA Binding Domain of AFX. The structure of the DNA binding domain of AFX (AFX-DBD) was determined in aqueous solution at pH 6.3, 31 °C using protein concentrations of about 1.5 mM. The protein construct encompassed residues 82−207 of the native protein preceded by 24 residues encompassing a his6-tag at the N-terminus. The residues outside the structured DNA binding domain (residues 100−180) all showed random coil chemical shifts and very narrow resonance lines resulting in severe resonance overlap in some spectral regions. Nevertheless, complete backbone resonance assignments were obtained for nearly all residues, and side chain resonance assignments were more than 96% complete for residues 95−181 (*38*).

Table 1 shows the structural statistics for the NMR ensemble. The NMR structure of AFX-DBD adopts the expected forkhead winged-helix fold (*2*) (Figure 1). The three major α-helices (H1−H3) are interrupted by a fourth short helix denoted helix 4 (H4), in accordance with the nomenclature from previous studies of forkhead domains, between H2 and H3. The boundaries of this helix are not completely conserved between the different members of the NMR ensemble. In some of the calculated structural models, the N-terminal residues of H4 adopt a 3_{10}-type helical conformation. The short β-strand (S1) between helix 1 and 2 interacts with β-strands 2 and 3 (S2 and S3), formed in the C-terminal part of the domain, to form a twisted antiparallel β-sheet. The disordered loop between β-strands 2 and 3 is referred to as wing 1 (W1). The second wing (W2) consists of the residues C-terminal to β-strand 3. In the present structure, however, these were found to be completely disordered and highly mobile on the nanosecond time scale; i.e., very narrow resonance lines at random coil chemical shifts and no long-range NOEs were observed. The hydrophobic core of the protein is made up of residues from all parts of the sequence and is centered around the interface of helices 1−3. All charged residues are found on the protein surface.

The structure of AFX-DBD is well-defined (RMS deviation for backbone atoms less than 1.0 Å) for residues 100−137, 145−165, and 171−177. Steady-state NOE measurements (^{15}N-{^{1}H}) suggest that the lack of coordinate precision for residues 138−144, 166−170, and the N- and C-terminal parts of the construct truly reflects increased dynamics in these parts of the protein. Figure 2 shows the measured values for the ^{15}N-{^{1}H} steady-state NOEs. A sausage representation of the NMR structure, color-coded according to the measured steady-state NOEs, is displayed in Figure 1d. There is clearly a strong correlation between the measured steady-state NOEs and the backbone coordinate precision. Interestingly, residues 138−144 and 166−170 correspond to the loop preceding helix 3 and the wing 1 loop. These sequence stretches have been identified as being important for the DNA recognition/interaction in related forkhead transcription factors (*39, 40*).

Comparison with Other Forkhead Domains. To date, high-resolution three-dimensional structures of the DNA binding domain have been determined for three different members of the forkhead family of transcription factors: HNF3-γ (X-ray) (*2*), Genesis (NMR) (*3, 15*), and FREAC-11 (NMR) (*16*). The structures have been determined both in complex with DNA (HNF3-γ and Genesis) and in the free form (Genesis and FREAC-11). Very minor structural differences were observed between the DNA-complexed and the free forms of Genesis. This seems to be the case also for FREAC-11; i.e., the secondary C$^{\alpha}$ chemical shifts of FREAC-11 in complex with DNA are highly similar to those observed in the free state (Maria J. P. van Dongen, personal communication).

Figure 3 shows the sequence alignment, similar to the one presented by Guo et al. (*11*), of the DNA binding domains of AFX and other forkhead transcription factors. The sequence conservation is high within the elements of regular secondary structure and in particular among the structurally important hydrophobic residues. AFX is about 85% and 75% identical to FKHR and DAF-16, respectively, but shares only about 40% sequence identity with each of HNF3-γ, Genesis, and FREAC-11. The amino acid sequences of the latter three are 65−70% pairwise identical.

Figure 4 shows an overlay of the different structures of forkhead DNA binding domains. Despite the rather low sequence identity, the structure of AFX-DBD is remarkably similar to the HNF3-γ and FREAC-11 structures. The backbone RMSD for residues in the regular secondary structure elements amounts to 0.6 Å (AFX vs HNF3-γ) and 0.7 Å (AFX vs FREAC-11). The structures of the DNA binding domain of Genesis (free and DNA complexed forms), on the other hand, are less structurally homologous to all other structures; i.e., the backbone pairwise RMSDs lie in the range of 1.9−2.2 Å, the most striking difference

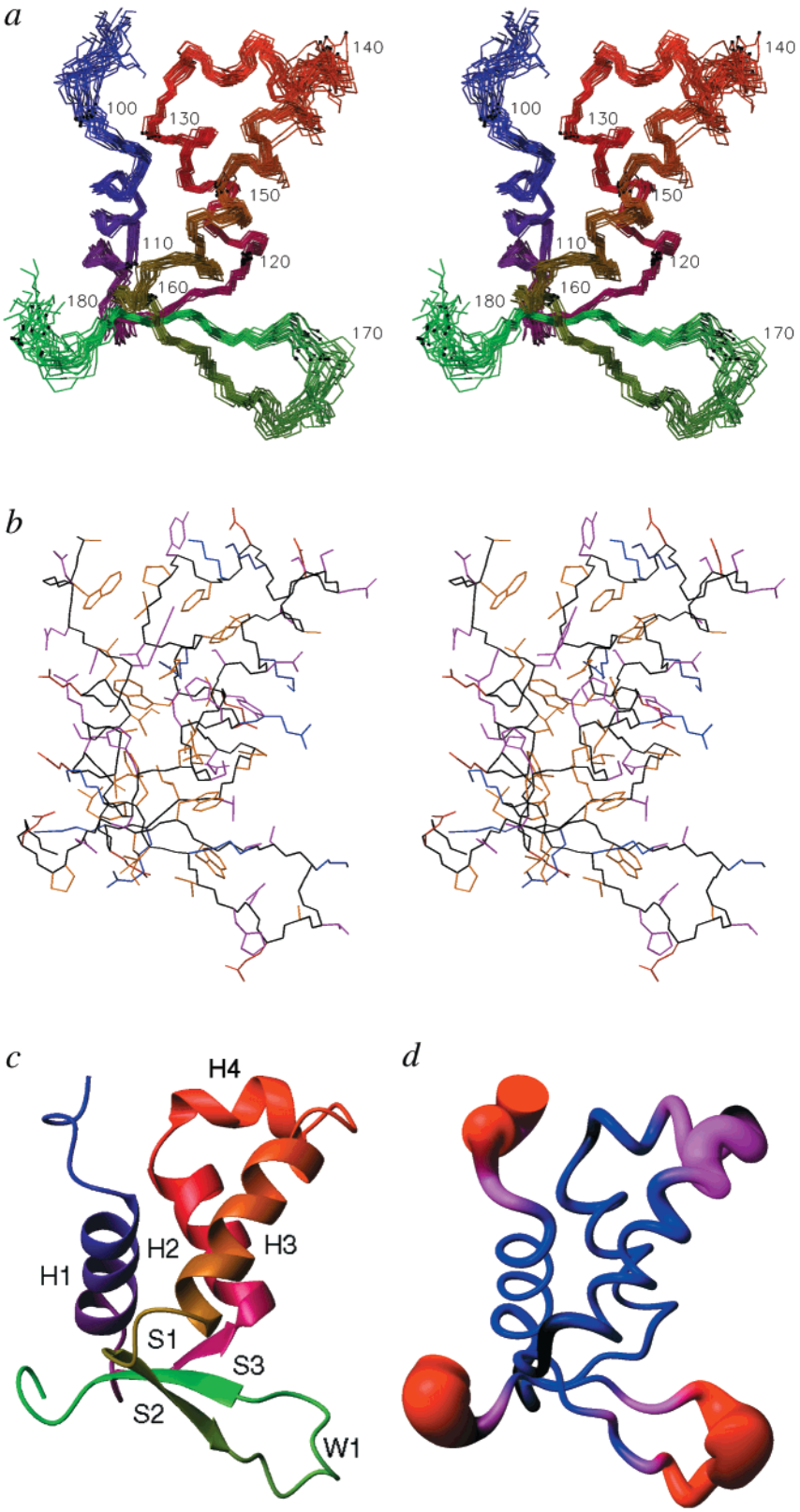

FIGURE 1: Different representations of the NMR solution structure of the DNA binding domain of AFX showing residues 96−181. The DNA interaction helix is facing the viewer. (a) Stereoview showing the ensemble of 20 conformers, using the backbone atoms of residues 102−136, 145−163, and 173−176 for superposition. (b) Stereoview of the conformer closest to the mean structure showing the protein backbone and side chain heavy atom bonds. Side chains are colored yellow (Ala, Phe, Ile, Leu, Met, Pro, Val, and Trp), magenta (Asn, Gln, Ser, Thr, Tyr, and His), blue (Lys and Arg), and red (Asp and Glu). (c) Ribbon diagram of the conformer closest to the mean structure. Helices are labeled according to the standard nomenclature for forkhead domains. (d) Sausage representation of the NMR ensemble. The radius of the "sausage" corresponds to the backbone coordinate precision. The sausage has been color-coded, using the measured ^{15}N-{^{1}H} steady-state NOEs (Figure 2), according to the following scheme: residues with measured values within 2 standard deviations of the mean value found for residues in the 3 major helices, blue; residues with measured values more than 2 (10) standard deviations below the mean value found for residues in the 3 major helices, magenta (red); residues lacking amide NMR assignments, black.

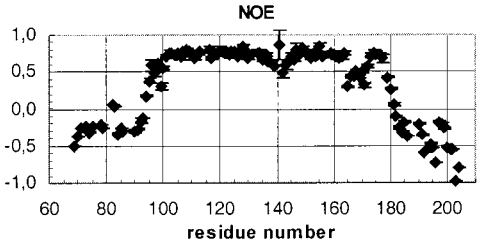

FIGURE 2: Steady-state ^{15}N-{^{1}H} NOEs measured for the backbone amide nitrogen atoms of the DNA binding domain of AFX. Error estimates were calculated from standard error propagation. The atypical value measured for Asn 141 is due to very poor signal-to-noise for the corresponding 2D ^{1}H-^{15}N HSQC cross-peak.

being the different orientation of helix 3 clearly visible in Figure 4. The orientation of helix 3 has been argued to be important for the DNA recognition (*15*), which might explain the different structural feature despite the highly homologous amino acid sequences. It is, however, interesting to note that AFX-DBD is structurally more similar to HNF3-γ and FREAC-11 than is Genesis, which is phylogenetically more closely related.

DNA Complex Formation. A multitude of DNA recognition sites have been reported for forkhead transcription factors (*1*, *10*, *11*, *41*−*43*). The previously reported core consensus recognition motif [5′-(G/A)(T/C)(A/C)AA(C/T)A-3′] (*1*) can be reduced to 5′-(A/C)AA(C/T)A-3′ taking the more recent data into account (*10*, *11*, *41*−*43*). It has been noted that the core sequence is necessary but not sufficient for the protein DNA recognition; i.e., the residues flanking the core motif contribute to the DNA binding specificity (*1*).

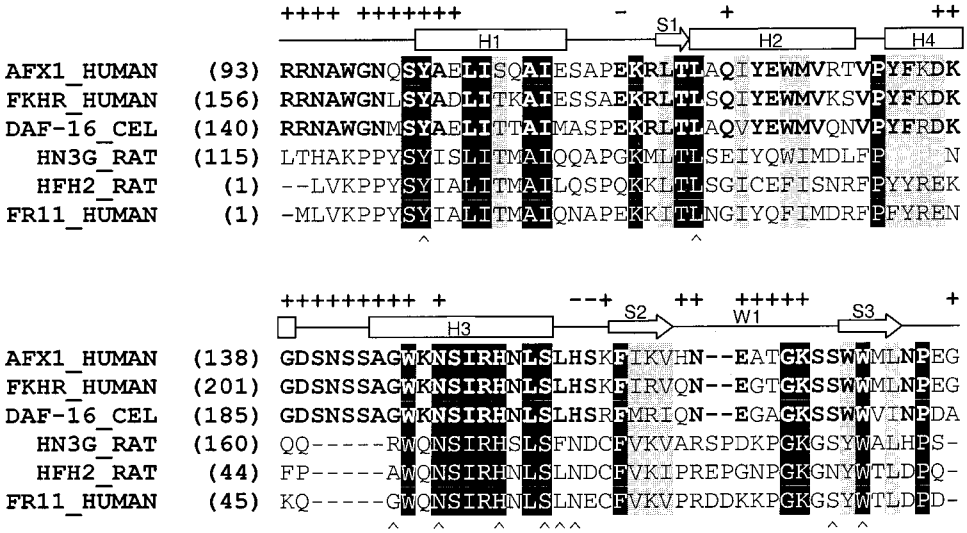

FIGURE 3: Sequence alignment of the DNA binding domain from different forkhead transcription factors: AFX1_HUMAN (AFX current study, *Homo sapiens*, SWISSPROT P98177, trEMBL Y11284); FKHR_HUMAN (FKHR, *Homo sapiens*, SWISSPROT Q12778); DAF-16_CEL (DAF-16, *Caenorhabditis elegans*, SWISSPROT O18676); HN3G_RAT (HNF3-γ, *Rattus norvegicus*, SWISSPROT P32183); HFH2_RAT (HFH2/Genesis, *Rattus norvegicus*, SWISSPROT Q63245); FR11_HUMAN (FREAC-11, *Homo sapiens*, sequence identical to SWISSPROT Q99958 Mesenchyme fork head protein). Black boxes mark fully conserved residues. Shaded boxes mark conserved residues allowing for conservative mutations (K/R, S/T, F/Y/W, and I/L/V/M). Boldface letters indicate residues that are conserved within the AFX-like sequences (AFX, FKHR, and DAF-16). Residues in contact with DNA in the X-ray structure of HNF3-γ are indicated by a caret ($\wedge$). Regular secondary structure elements are indicated at the top. Plus signs (+) denote residues with H^N protons in fast exchange with the solvent, identified from intense exchange cross-peaks in the 3D ^{1}H-^{15}N NOESY-HSQC spectrum. Minus signs (−) mark residues for which the H^N resonance could not be observed.

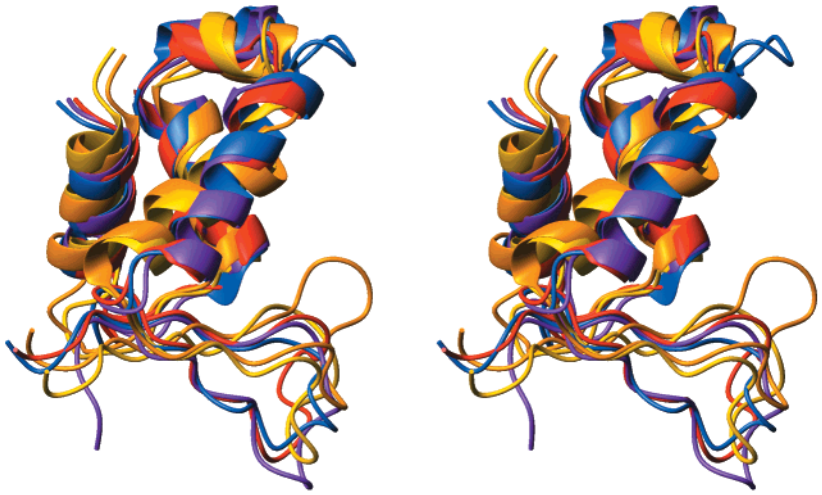

FIGURE 4: Structural comparison (stereoview) of the DNA binding domains of AFX (blue), HNF3-γ (DNA complex) (red), S12 (magenta), Genesis (yellow), and Genesis (DNA complex) (orange). The orientation is identical to that of Figure 1. Backbone heavy atoms of residues corresponding to the secondary structure elements H1, H2, H3, S2, and S3 as indicated in Figure 3 were used for superposition.

The forkhead transcription factor Genesis was shown to interact strongly with a number of different DNA sequences all containing the 5′-AAA(C/T)A-3′ core motif, and a weak interaction was observed with the transthyretin promoter region encompassing the 5′-CAATA-3′ motif, whereas no binding interaction was observed with an HNF3-β-specific sequence despite the presence of the core sequence (*41*). An additional example is the interaction of HNF3-γ with the phosphoenolpyruvate carboxykinase (PEPCK) and insulin-like growth factor-binding protein-1 (IGFBP-1) promoters. Both promoters encompass the core recognition motif, but the affinity of HNF3-γ for the PEPCK promoter is 6 times lower than the affinity for the IGFBP-1 promoter (*44*). The forkhead transcription factor FKHR, which is closely related to AFX, was shown to stimulate IGFBP-1 promoter activity through an insulin response sequence (IRS) (*11*). The IGFBP-1 promoter contains two IRSs: IRSA (5′-CAAAA-CAA-3′) and IRSB (5′-TTATTTTG-3′), which both encompass the 5′-AAA(C/T)A-3′ forkhead recognition motif. The presence of either of these IRSs was shown sufficient to mediate effects of insulin and FKHR on promoter function, whereas single nucleotide mutations abolished the effects. Furthermore, in an electrophoretic band shift assay, AFX was shown to interact with the IGFBP-1 promoter, but not with a mutated sequence (Am2Bm2) (*10*).

There is thus compelling evidence in the literature for interaction of AFX with the insulin response elements of IGFBP-1. To verify the suitability of two dDNA 16-mers (IRS_A and IRS_B) for structural studies, 2D ^{1}H-^{15}N-HSQC NMR spectra were recorded on 1:1 mixtures with AFX-DBD. Both spectra were well resolved and displayed large chemical shift changes and broadened resonance lines compared to the spectrum of free AFX-DBD (Figure 5), indicative of complex formation. Moreover, the spectra of the two mixtures (Figure 5b,c) were nearly identical, indicating an identical binding mode of AFX-DBD to the two dDNA oligonucleotides. Virtually all resonances in the HSQC spectrum experienced chemical shift changes upon complex formation, making it difficult to make tentative assignments of the ^{1}H-^{15}N resonances of the DNA-complexed form of AFX-DBD based on the resonance assignments from the native protein alone. The present data confirm binding of AFX-DBD to the individual insulin response elements of IGFBP-1 and clearly identify IRS_A and IRS_B as suitable constructs for further structural studies. To study the DNA complex(es) of AFX-DBD in more detail, a perdeuterated sample of AFX-DBD is most likely needed due to the large molecular mass of the complex(es) (~26 kDa).

Basis for DNA Recognition. Overdier et al. identified 20 amino acid residues preceding the DNA recognition helix as being important for DNA binding specificity (*39*). The binding specificity of HNF3-β was altered to recognize HFH-1-specific DNA binding sites in a chimeric HNF3-β/HFH-1 protein where 20 amino acid residues preceding the DNA recognition helix had been replaced by the corresponding 20 residues from HFH-1 (*39*). No direct base contacts are made by residues in this region (*2, 3*), and the mechanism by which the DNA binding is influenced remains elusive. It has been proposed that the presentation of helix 3 to the DNA may be modulated by the preceding residues (*39*). It can be argued this is the case for the Genesis transcription factor (*3, 15*) where helix 3 adopts an orientation clearly different

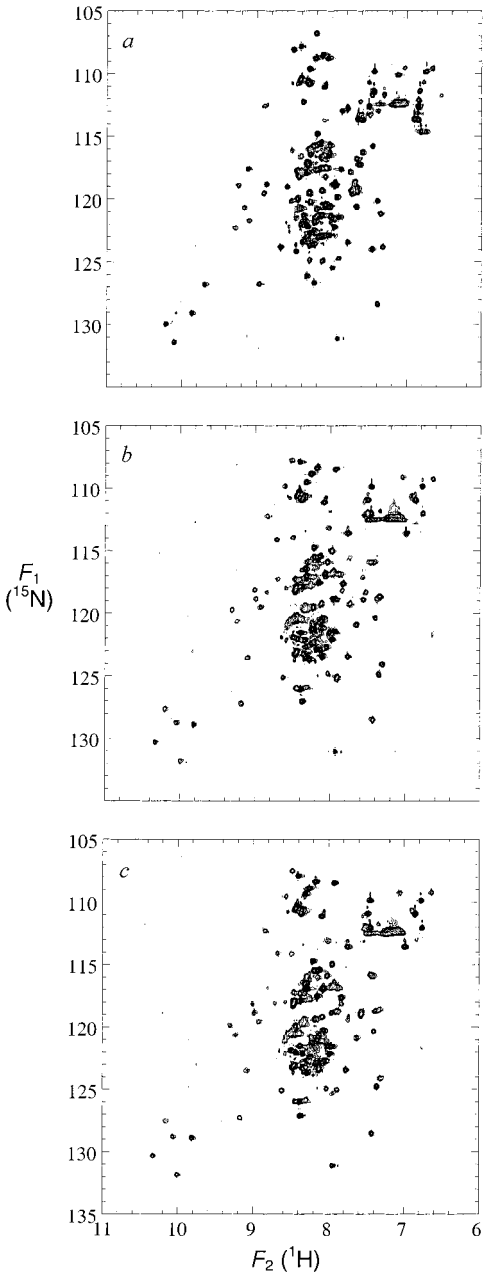

FIGURE 5: 2D ^{1}H-^{15}N HSQC spectra of 1:1 mixtures of AFX-DBD and two double-stranded DNA 16-mers at 0.1 mM concentration. The spectra were recorded at 20 °C and pH 6.3 on a Varian Unity INOVA 600 NMR spectrometer. (a) Reference spectrum. (b) Insulin response sequence A (IRS_A) dCAAGCAAAACAAACCA. (c) IRS_B dTGTTCAAAATAATCTG.

from what is found for HNF3-γ, FREAC-11, and AFX-DBD. It is not obvious, however, that this structural feature is indeed caused by the residues in the region preceding helix 3. The sequence alignment shown in Figure 3 indicates that there is some sequence variability in this sequence stretch. The AFX-like sequences have even a five amino acid insertion in this segment. Yet, this does not significantly alter the positioning of helix 3 in the structure of AFX-DBD compared to the structures of HNF3-γ and FREAC-11. Liao and co-workers argue that the presence of helix 4 in the Genesis structure could account for the repositioning of helix 3 when compared to HNF3-γ (*3, 15*). We find reason to question this. First, as indicated by Figure 3, the amino acid residues of helix 4 itself are highly conserved in all aligned

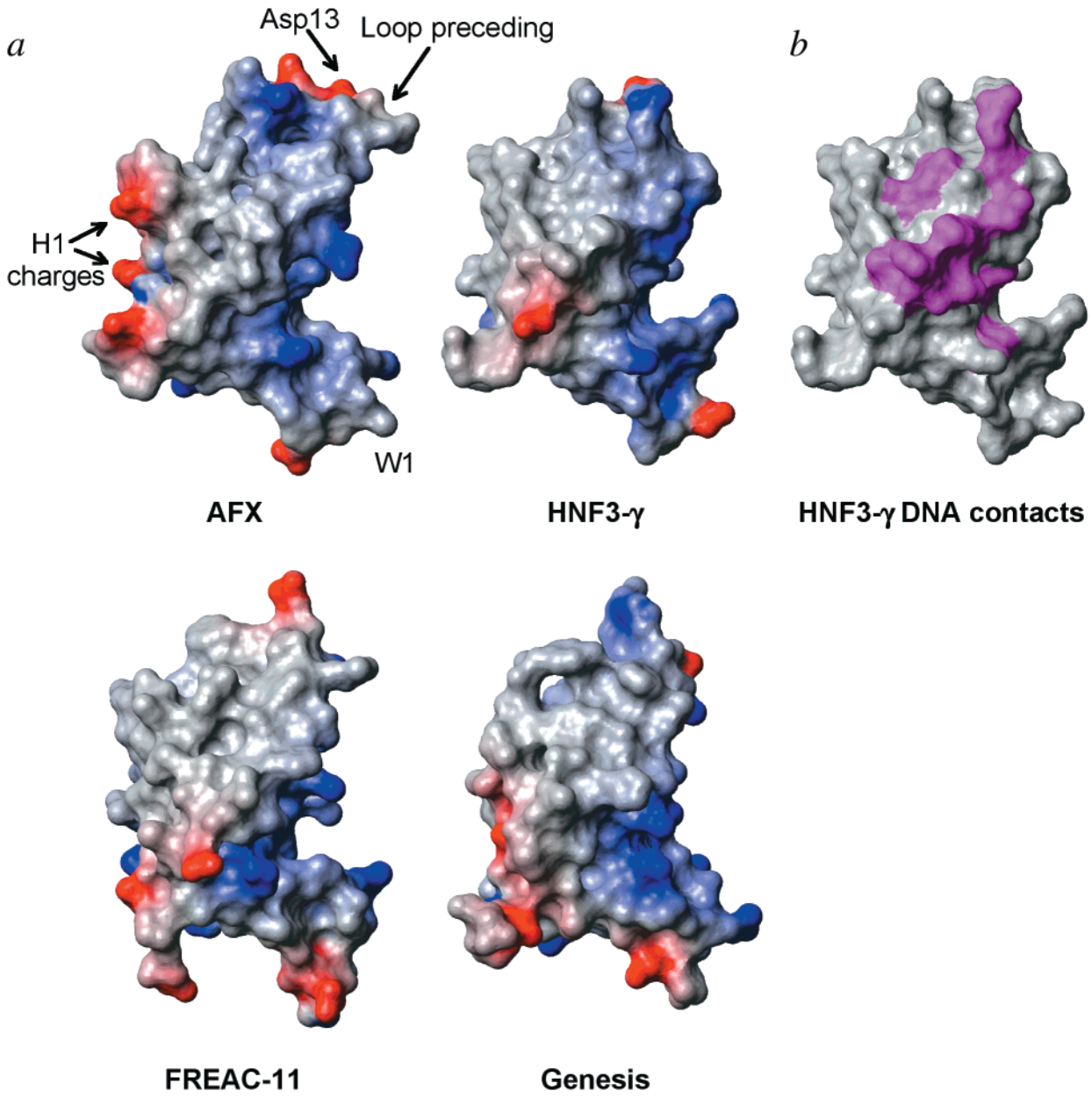

FIGURE 6: (a) Electrostatic potential of the surfaces of the DNA binding domains of AFX, HNF3-γ, FREAC-11, and Genesis. The orientation of the molecules is identical to that of Figures 1 and 4; i.e., the face of the molecule facing DNA is facing the viewer. The electrostatic potential was calculated using the *simplecharge* option in MOLMOL; i.e., charges are attributed only to a few atoms in charged groups. Positive potential is colored blue; negative potential is colored red. (b) Surface representation of HNF3-γ identifying the residues in contact with DNA in the X-ray crystallographic structure. The corresponding surface patches are colored magenta.

sequences. Second, the Liao group state that helix 4 represents a true structural difference and that only a "quasi-helical loop" is present in HNF3-γ. Helix 4 is, however, present also in the FREAC-11 and AFX-DBD structures, and a PROCHECK (*36*) analysis readily identifies helix 4 not only in Genesis, FREAC-11, and AFX-DBD but also in HNF3-γ (a four residue long 3_{10}-type helix). We find there is not conclusive evidence to prove that the residues preceding helix 3 define its positioning. Nevertheless, the importance of this sequence stretch for DNA recognition has been demonstrated (*39*), but other effects than the repositioning of helix 3 must come into play.

Figures 4 and 6 show comparisons of AFX-DBD with other forkhead DNA binding domains. The ribbon diagram of Figure 4 focuses on differences and similarities of the overall backbone conformation of the molecules. Figure 6 displays the electrostatic characteristics of the DNA interaction surface. The DNA interaction surfaces of HNF-3γ and Genesis are very similar (*2*, *3*). Due to the high sequence conservation of the recognition helix of forkhead transcription factors (*1*), it could be assumed that forkhead transcription factors in general, and AFX and FREAC-11 in particular, would display nearly identical binding modes. The altered orientation of helix 3 in Genesis (Figure 4) is clearly unique, and this could perhaps explain the DNA discrimination power of this transcription factor (*3*, *15*). AFX, HNF-3γ, and FREAC-11 all display the same orientation of helix 3, suggesting that the ability for these DNA binding domains to recognize different DNAs must be attributed to other factors. The most obvious differences between the molecules are observed in the electrostatic potentials of the molecular surfaces (Figure 6). In particular, unique surface properties

can be identified for each of the different DNA binding domains even in the regions corresponding to and flanking the DNA interaction residues identified in the HNF3-γ DNA complex. For the particular case of AFX-DBD, the loop preceding helix 3 contains an additional negatively charged residue (Asp139) and a five amino acid residue insertion compared to the other proteins. Wing 1 is shortened by two residues, and it carries only one negatively (Glu166) and one positively charged residue (Lys170), whereas the other sequences contain one negatively and two or more positively charged groups. A further difference is the presence of two negatively charged residues in helix 1 giving rise to the negatively charged surface patches unique for AFX-DBD. It can be argued that altered charge characteristics of the protein surface affect the electrostatic interaction with the phosphate backbone of the DNA and thereby modulate the presentation of the highly conserved recognition helix to the major groove. Small charge changes are known to influence the bending of DNA, as demonstrated for instance in studies of charge-induced DNA bending by mutants of the transcription factor GCN4 (*45*). Moreover, cases where DNA binding is modulated by charged residues contacting DNA bases flanking the core interaction site are also known (*46*). Another striking example of the importance of charge interactions for DNA recognition is the completely altered binding mode of the winged-helix protein hRFX1 to DNA where a loop (wing 1) rather than the recognition helix interacts with the major groove of DNA which is attributed to electrostatic effects (*47*).

The detailed structural basis for DNA recognition of forkhead proteins remains elusive. This is mainly due to the lack of high-resolution structural data on forkhead DNA complexes. To date, only two structures are available: HNF-3γ (*2*) and Genesis (*3*). The HNF-3γ structure was obtained with only a 13-mer DNA which might be sub-optimal; i.e., wing 1 of the protein extended outside the DNA site. The NMR structure of the Genesis DNA complex was obtained using a longer DNA duplex (17-mer) but is relatively poorly resolved and rests on comparatively few intermolecular NOE's. Furthermore the structure calculation was based on a straight *B*-form DNA template, even though evidence obtained for other forkhead transcription factors indicates mild to substantial bending of the DNA upon complex formation (*2*, *48*) (Maria J. P. van Dongen, personal communication). Most likely, several high-resolution structures of different forkhead DNA complexes—or complex structures obtained with several mutants of one forkhead transcription factor—are needed to resolve the issue.

ACKNOWLEDGMENT

We gratefully acknowledge Maria J. P. van Dongen (Pharmacia) and Thomas Lundbäck (Pharmacia) for helpful discussions.

REFERENCES

1. Kaufmann, E., and Knöchel, W. (1996) *Mech. Dev. 57*, 3−20.
2. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature 364*, 412−420.
3. Jin, C., Marsden, I., Chen, X., and Liao, X. (1999) *J. Mol. Biol. 289*, 683−690.
4. Anderson, M. J., Viars, C. S., Czekay, S., Cavenee, W. K., and Arden, K. C. (1998) *Genomics 47*, 187−199.
5. Kaestner, K. H., Knöchel, W., and Martínez, D. E. (2000) *Genes Dev. 14*, 142−146.
6. Ogg, S., Paradis, S., Gottlieb, S., Patterson, G. I., Lee, L., Tissenbaum, H. A., and Ruvkun, G. (1997) *Nature 389*, 994−999.
7. Paradis, S., and Ruvkun, G. (1998) *Genes Dev. 12*, 2488−2498.
8. Borkhardt, A., Repp, R., Haas, O. A., Leis, T., Harbott, J., Kreuder, J., Hammermann, J., Henn, T., and Lampert, F. (1997) *Oncogene 14*, 195−202.
9. Medema, R. H., Kops, G. P. J. L., Bos, J. L., and Burgering, B. M. T. (2000) *Nature 404*, 782−787.
10. Kops, G. J. P. L., de Ruiter, N. D., De Vries-Smith, A. M. M., Powell, D. R., Bos, J. L., and Burgering, B. M. T. (1999) *Nature 298*, 630−634.
11. Guo, S., Rena, G., Cichy, S., He, X., Cohen, P., and Unterman, T. (1999) *J. Biol. Chem. 274*, 17184−17192.
12. Takaishi, H., Konishi, H., Matsuzaki, H., Ono, Y., Shirai, Y., Saito, N., Kitamura, T., Ogawa, W., Kasuga, M., Kikkawa, O., and Nishizuka, Y. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11836−11841.
13. Nakae, J., Park, B.-C., and Accili, D. (1999) *J. Biol. Chem. 274*, 15982−15985.
14. Kops, G. P. J. L., and Burgering, B. M. T. (1999) *J. Mol. Med. 77*, 656−665.
15. Marsden, I., Jin, C., and Liao, X. (1998) *J. Mol. Biol. 278*, 293−299.
16. van Dongen, M. J. P., Cederberg, A., Carlsson, P., Enerbäck, S., and Wikström, M. (2000) *J. Mol. Biol. 296*, 351−359.
17. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
18. Kraulis, P. J., Domaille, P. J., Campbell-Burk, S. L., van Aken, T., and Laue, E. D. (1994) *Biochemistry 33*, 3515−3531.
19. Mori, S., Abeygunawardana, C., Johnson, M. O., and vanZijl, P. C. M. (1995) *J. Magn. Reson. B 108*, 94−98.
20. Talluri, S., and Wagner, G. (1996) *J. Magn. Reson. B 112*, 200−205.
21. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson. B 103*, 203−216.
22. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994) *J. Am. Chem. Soc. 116*, 11655−11666.
23. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
24. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) *J. Magn. Reson. 95*, 636−641.
25. Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. (1993) *J. Magn. Reson. B 101*, 333−337.
26. Zuiderweg, E. R. P., and Fesik, S. W. (1989) *Biochemistry 28*, 2387−2391.
27. Zerbe, O., Szyperski, T., Ottiger, M., and Wüthrich, K. (1996) *J. Biomol. NMR 7*, 99−106.
28. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR spectroscopy: principles and practice*, Academic Press, San Diego.
29. Szyperski, T., Neri, D., Leiting, B., Otting, G., and Wüthrich, K. (1992) *J. Biomol. NMR 2*, 323−334.
30. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
31. Brünger, A. T. (1993) *X-PLOR (Version 3.1): A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
32. Garrett, D. S. (1994) *J. Magn. Reson. B 104*, 99−103.

33. Kuszewski, J., Gronenborn, A. M., and Clore, G. M. (1996) *Protein Sci. 5*, 1067−1080.
34. Wüthrich, K., Billeter, M., and Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.
35. Stein, E. G., Rice, L. M., and Brünger, A. T. (1997) *J. Magn. Reson. 124*, 154−164.
36. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
37. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph. 14*, 51−55.
38. Weigelt, J., Climent, I., Dahlman-Wright, K., and Wikström, M. (2000) *J. Biomol. NMR 17*, 181−182.
39. Overdier, D. G., Porcella, A., and Costa, R. H. (1994) *Mol. Cell. Biol. 14*, 2755−2766.
40. Shiyanova, T., and Liao, X. (1999) *Arch. Biochem. Biophys. 362*, 356−362.
41. Sutton, J., Costa, R., Klug, M., Field, L., Xu, D., Largaespada, D. A., Fletcher, C. F., Jenkins, N. A., Copeland, N. G., Klemsz, M., and Hromas, R. (1996) *J. Biol. Chem. 271*, 23126−23133.
42. Ayala, J. E., Streeper, R. Y., Desgrosellier, J. S., Durham, S. K., Suwanichkul, A., Svitek, C. A., Goldman, J. K., Barr, F. G., Powell, D. R., and O'Brien, R. M. (1999) *Diabetes 48*, 1885−1889.
43. Durham, S. K., Suwnichkul, A., Scheimann, A. O., Yee, D., Jackson, J. G., Barr, F. G., and Powell, D. R. (1999) *Endocrinology 140*, 3140−3146.
44. O'Brien, R. M., Noisin, E. L., Suwanichkul, A., Yamasaki, T., Lucas, P. C., Wang, J.-C., Powell, D. R., and Granner, D. K. (1995) *Mol. Cell. Biol. 15*, 1747−1758.
45. Strauss-Soukop, J. K., and Maher, L. J. (1998) *Biochemistry 37*, 1060−1066.
46. Pan, C. Q., Finkel, S. E., Cramton, S. E., Feng, J.-A., Sigman, D. S., and Johnson, R. C. (1996) *J. Mol. Biol. 264*, 675−695.
47. Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B., and Burley, S. K. (2000) *Nature 403*, 916−920.
48. Pierrou, S., Hellqvist, M., Samuelsson, L., Enerbäck, S., and Carlsson, P. (1994) *EMBO J. 13*, 5002−5012.

BI001663W