High resolution structure of the HDGF PWWP domain: A potential DNA binding domain

STEPHEN M. LUKASIK,[1] TOMASZ CIERPICKI,[2] MATTHEW BORLOZ,[2] JOLANTA GREMBECKA,[2] ALLEN EVERETT,[3] AND JOHN H. BUSHWELLER[1,2]

[1]Department of Chemistry and [2]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22906, USA
[3]Division of Pediatric Cardiology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA

(RECEIVED August 3, 2005; FINAL REVISION November 3, 2005; ACCEPTED November 7, 2005)

Abstract

Hepatoma Derived Growth Factor (HDGF) is an endogenous nuclear-targeted mitogen that is linked with human disease. HDGF is a member of the weakly conserved PWWP domain family. This 70–amino acid motif, originally identified from the *WHSC1* gene, has been found in more than 60 eukaryotic proteins. In addition to the PWWP domain, many proteins in this class contain known chromatin remodeling domains, suggesting a role for HDGF in chromatin remodeling. We have determined the NMR structure of the HDGF PWWP domain to high resolution using a combination of NOEs, J-couplings, and dipolar couplings. Comparison of this structure to a previously determined structure of the HDGF PWWP domain shows a significant difference in the C-terminal region. Comparison to structures of other PWWP domains shows a high degree of similarity to the PWWP domain structures from Dnmt3b and mHRP. The results of selected and amplified binding assay and NMR titrations with DNA suggest that the HDGF PWWP domain may function as a nonspecific DNA-binding domain. Based on the NMR titrations, we propose a model of the interaction of the PWWP domain with DNA.

Keywords: HDGF; PWWP; growth factor; heart development; dipolar couplings; protein structure/folding; protein–nucleic acid interactions; DNA-binding domains; structure; structural proteins; NMR spectroscopy; heteronuclear NMR

The endothelium controls vascular smooth muscle cell (VSMC) proliferation and migration through the release of cytokines, growth factors, and matrix proteins (Casscells 1992; Dodge et al. 1993; Frid et al. 2004; Herve et al. 2005; McMullen et al. 2005). In particular, growth factors liberated at the time of vascular injury have been demonstrated to directly stimulate VSMC proliferation in vivo (Walker et al. 1986; Winkles et al. 1987; McNamara et al. 1993). After surface-receptor binding and activation, most growth-factor complexes are internalized, transported to lysosomes, and degraded (Honegger et al. 1987; Schlessinger 2000; Dikic 2003). However, it has recently been demonstrated that several exogenous growth factors, such as PDGF and bFGF, exert their effect within the nucleus of the cell (Rakowicz-Szulczynska et al. 1986; Baldin et al. 1990; Gabriel et al. 1991).

Hepatoma Derived Growth Factor (HDGF) is an endogenous nuclear-targeted mitogen that is linked with human disease. HDGF is coexpressed with PCNA in proliferating VSMC atherosclerotic plaques (Everett et al. 2000). It has also been demonstrated that HDGF expression is predictive of survival in patients with non-small-cell lung cancer (Ren et al. 2004; Iwasaki et al.

Reprint requests to: John H. Bushweller, Department of Chemistry and Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22906, USA; e-mail: jhb4v@virginia.edu; fax: (434) 982-1616.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.051751706.

2005). Although lacking a secretory leader sequence, HDGF was originally purified from medium conditioned by the human hepatoma-derived cell line (HuH-7) and from the rat metanephrogenic cell line 7.1.1 (Nakamura et al. 1994; Oliver and Al-Awqati 1998). HDGF fits the definition of a true growth factor, as exogenous HDGF is mitogenic for fibroblasts (Nakamura et al. 1994), endothelial cells (Oliver and Al-Awqati 1998; Everett et al. 2004), several hepatoma cell lines (Kishima et al. 2002), and neuronal cells (Zhou et al. 2004).

HDGF is a member of a relatively new family of proteins containing the weakly conserved PWWP domain. The PWWP domain was first characterized from the *WHSC1* gene (Stec et al. 2000). It contains a conserved 70–amino acid sequence and has been found in ~60 eukaryotic proteins. In addition to a PWWP domain, proteins in this family frequently contain known chromatin association domains such as the bromodomain, chromodomain, SET domain, and Cys-rich Zn-binding domains (Stec et al. 2000). This strongly suggests a role in chromatin regulation or modification for these proteins, and potentially, the PWWP domain itself. Initially, the PWWP domain was hypothesized to be a site for protein–protein interaction. However, the PWWP domain of Dnmt3b has been shown to interact with DNA (Qiu et al. 2002; Chen et al. 2004; Ge et al. 2004). Indeed, it has recently been demonstrated that Dnmt3a and Dnmt3b are associated with chromatin throughout the entire cell cycle and that this association is mediated via the PWWP domain (Ge et al. 2004).

We have determined the high-resolution solution structure of the N-terminal PWWP domain of HDGF using NMR spectroscopy. Comparison of this structure to a previously determined HDGF PWWP domain structure shows a significant difference in the C-terminal region of the domain. We have also compared this structure with the PWWP domain structures from Dnmt3b and mHRP. In addition, we have used selected and amplified binding and NMR titrations to map a potential DNA-binding site on the protein.

Results

Defining PWWP domain boundaries in HDGF

We expressed and purified a portion of HDGF corresponding to amino acids 0–142 through affinity and size-exclusion chromatography. The zero numbering represents an N-terminal Gly cloning artifact remaining after rTEV cleavage. To identify the core PWWP domain, this fragment was subjected to limited proteolysis with subtilisin. The cleaved protein band was extracted from an SDS–polyacrylamide gel and analyzed by MALDI-TOF mass spectrometry, resulting in identification of 0–110 as a stable domain amenable to structural analysis. This result was confirmed by measurements of the $\{^1H\}^{15}N$ NOE in 110 and 142 amino acid constructs (Fig. 1B).

Solution structure of the HDGF PWWP domain

Independent of other studies, we have determined the structure of the HDGF PWWP domain by NMR spectroscopy (Fig. 1). Assignments of ^1H, ^{13}C, and ^{15}N chemical shifts were based on standard heteronuclear 3D NMR experiments (Sattler et al. 1999). A nearly complete set of chemical shifts has been obtained for the fragment containing residues 8–96 (see Materials and Methods). For structure determination, we collected 3D ^{15}N-edited NOESY and 3D ^{13}C-edited NOESY spectra. The assignment of NOESY cross-peaks and structure calculations were achieved in an automated manner using CANDID/CYANA (Herrmann et al. 2002). In addition to NOEs, we utilized TALOS-derived ϕ and ψ dihedral restraints (Cornilescu et al. 1999) and a set of 23 interstrand hydrogen bonds identified for regular β-strands based on NOE patterns (Table 1). In order to further refine the structure of the PWWP domain of HDGF, we collected a large set of residual dipolar couplings (RDCs) for protein weakly aligned in charged polyacrylamide gels (Cierpicki and Bushweller 2004). Interestingly, we could obtain useful samples only in gels with very strong positive charge and in the presence of 200 mM NaCl in order to reduce electrostatic interactions. Attempts to achieve different alignment by using negatively charged and zwitterionic gels failed, most likely due to the very high pI of the protein (pI = 9.2) and strong interaction with the gels. We collected a large set of RDCs for protein weakly aligned in a 75 + M polyacrylamide gel (Cierpicki and Bushweller 2004). Four sets of RDCs ($^1D_{HN}$, $^1D_{NC'}$, $^2D_{HNC'}$, and $^1D_{C'C\alpha}$) have been measured using an HNCO based set of experiments (Yang et al. 1999). The refinement of the structure has been carried out in CNS (Brunger et al. 1998) with distance restraints obtained from automatic assignment in CANDID/CYANA, dihedral angle restraints, hydrogen bonds, $^3J_{HNH\alpha}$ couplings, and RDCs (Table 1). The final set of 26 lowest-energy structures is shown in Figure 1A.

A well-defined core structure is observed for residues 9–33 and 44–96. Indeed, the backbone RMS deviation for this portion of the protein is 0.31 Å, representing the most accurate structure of this domain to date. Significant disorder is observed for the N and C termini, as well as the poorly conserved loop containing residues 34–43. Measurement of the $\{^1H\}^{15}N$ NOE shows depressed values for these regions, confirming the increased dynamics in these regions (Fig. 1B).

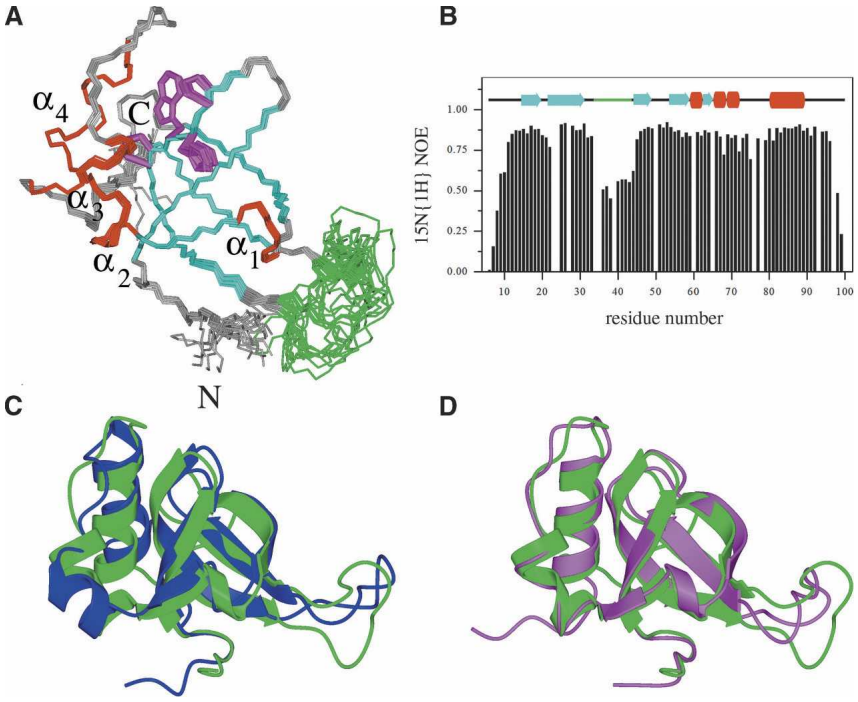

Figure 1. Structure of the HDGF PWWP domain. (*A*) Final set of 26 lowest-energy conformers with secondary structural elements color-coded as follows: α1–4, red; β1–5, cyan; loop L2 (residues 34–43), green; and side chains of the PWWP motif (P26-F27-W28-P29), magenta. (*B*) Heteronuclear NOE results with secondary structure elements color-coded as in *A*. (*C*) Comparison of this structure (green) with the previously solved HDGF PWWP domain structure (PDB code 1ri0) (blue). (*D*) Comparison of this structure with HRP-3 PWWP domain structure (PDB code 1n27) (magenta).

The conserved HDGF β-barrel consists of five antiparallel strands (β_1, 15–18; β_2, 26–32; β_3, 45–49; β_4, 53–58; β_5, 63–65). Strands β_1 and β_2 are linked by a β–β arch (MKGYP), which allows the side chains of Met20 and Tyr23 to point into the core of the barrel. Strands β_2 and β_3 are connected by the flexible loop L_2. This region shows little sequence similarity between PWWP homologs. Strands β_3 and β_4 are connected by a type II′ β turn. A 3_{10}-helix (α_1) is present between strands β_4 and β_5. The characteristic i to i + 3 hydrogen bonding pattern is observed for the carbonyls of Gly59 and Pro60 and the amides of Asp62 and Leu63, respectively.

The C-terminal portion of the PWWP domain consists of two adjacent helical turns α_2 and α_3 (residues 66–69 and 69–73) and the 10-residue long α_4 helix (residues 82–91). The α_2 is stabilized through hydrogen bonding between the backbone carbonyl of Pro65 and the amide of Ser69. In addition, hydrogen bonding is observed between the carbonyl of Ser69 and amide of Phe73 in α_3. The C-terminal α_4 contains a regular pattern of i to i + 4 hydrogen bonds for residues 82–91.

The conserved PWWP motif is located at the beginning of β-strand 2 (PHWP). Pro24 forms a β-bulge, which, along with the bend created by Pro27, is necessary for the formation of the β-barrel. In addition, the indole ring of Trp26 packs against α_1, while the side chain of His25 packs against the side chain of the highly conserved Lys19 residue. Indeed, mutation of Lys19 to Ser resulted in disruption of the fold, as assessed by ^{15}N-^{1}H HSQC spectra (data not shown), likely as a result of the intimate packing of this residue with His25. Indeed, we observe several NOEs between the side-chain protons of these residues.

Application of RDCs has a significant impact on the accuracy of the structure. Although the structure of the PWWP domain of HDGF obtained from a CANDID/CYANA calculation is based on a large number of experimental restraints, including 391 long-range NOEs, its accuracy judged from analysis of RDCs is moderate. The quality factors representing the agreement between structure and experimental RDCs (Cornilescu et al. 1998) calculated for all $^{1}D_{HN}$ couplings and the lowest-energy conformers from CYANA are ~40%. Application of RDCs leads to a substantial decrease of the quality factors to 12% with no NOE distance violations larger than 0.2 Å and no dihedral angle violations exceeding 2°. The difference between the unrefined and refined set of HDGF structures is 1.00 ± 0.11 Å and 1.72 ± 0.12 Å for backbone and heavy atoms, respectively. Furthermore, application of RDCs results in an

Table 1. *Statistics for assignment of NOESY spectra and calculation of HDGF structure*

Assignment using CANDID/CYANA[a]		
^{13}C NOESY (aliphatic)/assigned	2040/1920 (94%)	
^{13}C NOESY (aromatic)/assigned	292/265 (91%)	
^{15}N NOESY/assigned	953/802 (84%)	
Structure refinement using CNS		
No. NOE distances	1415	
Intraresidual	519	
Sequential ($\|i-j\| = 1$)	340	
Medium-range ($1 < \|i-j\| < 5$)	165	
Long range ($\|i-j\| > 4$)	391	
No. hydrogen bond restraints[b]	46	
No. dihedral angle restraints ($\psi + \phi$)[c]	38 + 41	
No. $J_{HNH\alpha}$ couplings	54	
No. RDCs	258	
$^{1}J_{HN}$ (Trp side chains)	64 (2)	
$^{1}J_{NC'}$	62	
$^{2}J_{HNC'}$	66	
$^{1}J_{C'C\alpha}$	64	
RMS deviations from distance restraints (Å)	0.0071 ± 0.0007[d]	
RMS deviations for hydrogen bond restraints (Å)	0.0197 ± 0.0024	
RMS deviations for dihedral angles (deg)	0.345 ± 0.016[d]	
RMS deviations for J-couplings (Hz)	0.48 ± 0.03	
RMS deviations for RDC (Hz)		
$^{1}J_{HN}$	1.1 ± 0.05	
$^{1}J_{NC'}$	0.23 ± 0.007	
$^{2}J_{HNC'}$	0.57 ± 0.015	
$^{1}J_{C'C\alpha}$	0.52 ± 0.013	
RMS deviations for covalent geometry		
bond lengths (Å)	0.0015 ± 0.00006	
bond angles (deg)	0.341 ± 0.006	
impropers (deg)	0.254 ± 0.011	
Ramachandran plot statistics (%)[e]	Residues 9–96	Residues 9–33, 44–96
residues in most favored regions	81.5	90.7
residues in additional allowed regions	13.5	7.3
residues in generously allowed regions	2.7	–
residues in disallowed regions	2.3	2.0
RMS deviations for NMR ensemble (Å)		
backbone	1.09 ± 0.35	0.31 ± 0.07
heavy atoms	1.65 ± 0.30	1.15 ± 0.16

[a] Number of collected and assigned peaks from 3D NOESY spectra.
[b] Two distance restraints for each hydrogen bond.
[c] Backbone dihedral angles from TALOS analysis (Cornilescu 1999).
[d] Largest NOE distance violation is < 0.2 Å and dihedral angle violations do not exceed 2°.
[e] Ramachandran plot statistics according to PROCHECK analysis (Laskowski 1996).

increase of structure precision and significant improvement of the Ramachandran plot statistics. Indeed, the percentage of residues in the most favored regions increases from 77.4% to 90.7%.

Selected and amplified binding assay (SAAB)

In order to try to identify a specific DNA-binding site for the PWWP domain of HDGF, we used a SAAB assay. This approach has previously been utilized to determine specific DNA-binding sites for Ciz1 (Warder and Keherly 2003), the MT domain of MLL (Ayton et al. 2004), and RAP1 (Graham and Chambers 1994), for example. A 72-bp oligonucleotide was synthesized, comprising a randomized 30-bp central region flanked by known sequences. Primers targeted at the known sequence were utilized for subsequent library amplification and for the production of the initial double-stranded DNA library. Six rounds of selection were performed utilizing the amplified PCR product as a binding library after the first round. His-tagged $HDGF_{1–142}$ bound to Ni-NTA beads was incubated with the DNA library and washed six times. The sample was boiled to release the new DNA library, which was

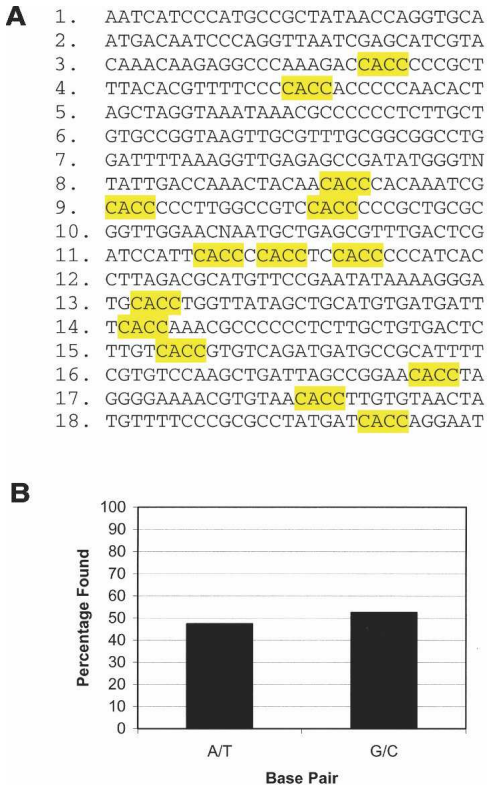

Figure 2. HDGF PWWP domain SAAB assay results. (*A*) A total of 20 clones were sequenced, yielding 18 unique inserts. The CACC element identified in 10 clones (61%) is outlined in yellow. (*B*) Percentage of G/C and A/T found in the clones sequenced.

further amplified by PCR. Reactions were monitored using TAE-agarose gels and visualized with ethidium bromide. The final library was cloned into PCR blunt (Invitrogen) and sequenced with the M13 reverse primer.

We were unable to determine a specific DNA sequence present in all of the sequenced clones (Fig. 2). The most common sequence was the C-rich 5′-CACC-3′ that appeared in 61% of the final library, usually as one copy per clone (Fig. 2A). The CACC element was found repeated in two of the 18 unique clones sequenced. No other consensus sequence was found with repeatability >5%, as assessed with the program MEME (Bailey and Elkan 1994). In previously reported SAAB assays, each analyzed sequence in the final library contained multiple copies of the specific high-affinity DNA element. For example, SAAB analysis of clones from the MT domain of MLL showed three to seven copies of the CpG dinucleotide sequence (Ayton et al. 2004). This lack of selection may be attributed either to a very weak interaction of the PWWP domain with the CACC DNA element or nonspecific binding, similar to what has been described for the PWWP domain of Dnmt3b (Chen et al. 2004). We further assessed whether the PWWP domain of HDGF selectively recognized C/G or A/T-rich DNA elements (Fig. 2B). Of the clones sequenced, 53% were G/C, while 47% were A/T (Fig. 2B). Therefore, the PWWP domain of HDGF does not discriminate between AT and GC base pairs, suggestive of a nonspecific DNA interaction.

Mapping of interactions between HDGF and DNA

In order to map the DNA-binding site on the HDGF PWWP domain, we carried out NMR titrations of 0.2 mM ^{15}N-labeled PWWP domain with various concentrations of the DNA element 5′-TACAA**CACC**CACA AA-3′. Numerous protein resonances were shifted by addition of 0.5, 1, and 2 equivalents of DNA (Fig. 3A,C). Very similar chemical-shift perturbations were observed for a mutant of this DNA (CACC → CATT), consistent with a nonspecific interaction (data not shown). The observation of only a single set of peaks and the slow saturation of protein chemical shifts upon titration with DNA is consistent with fast exchange and relatively weak binding. Indeed, titration measurements by NMR are consistent with a K_d of $\sim 10^{-5}$ M; however, this may be an apparent K_d that does not reflect a single unique binding site.

Most of the chemical-shift changes in the HDGF PWWP domain are localized to one face of the protein, consisting of residues found in the β1/β2 β–β arch region (residues Lys19–Tyr23), the 3_{10} helix (residues Gly59–Leu63), the two α_2 and α_3 helical turns (Tyr66–Phe73), and the N terminus (residues Lys8–Tyr10). Electrostatic calculations show that the location of the putative DNA-binding site overlaps with a patch of very high positive electrostatic potential on the protein (Fig. 3D,E). This patch includes several Lys residues (8, 19, 21, 61, 72, 75, and 80) and Arg79, all of which are solvent-exposed. This site has also been identified as a heparin binding site, consistent with heparin's polyanionic composition (Sue et al. 2004).

We also analyzed chemical shifts of the DNA fragment based on homonuclear 2D TOCSY and NOESY spectra. Titration of DNA with HDGF results in a continuous stretch of relatively weak perturbations along the entire DNA sequence (Fig. 3B). This is consistent with a nonspecific interaction and with multiple binding sites on DNA. The strongest chemical-shift perturbations are observed for 5 bp located in the middle of the 15-mer oligonucleotide. Interestingly, the size of this DNA fragment agrees well with that of the mapped surface of HDGF involved in DNA binding. Based on the chemical shift titration, we created a model of the HDGF DNA complex (Fig. 3C). Although this model has limited accuracy, it clearly identifies a surface of the protein that is

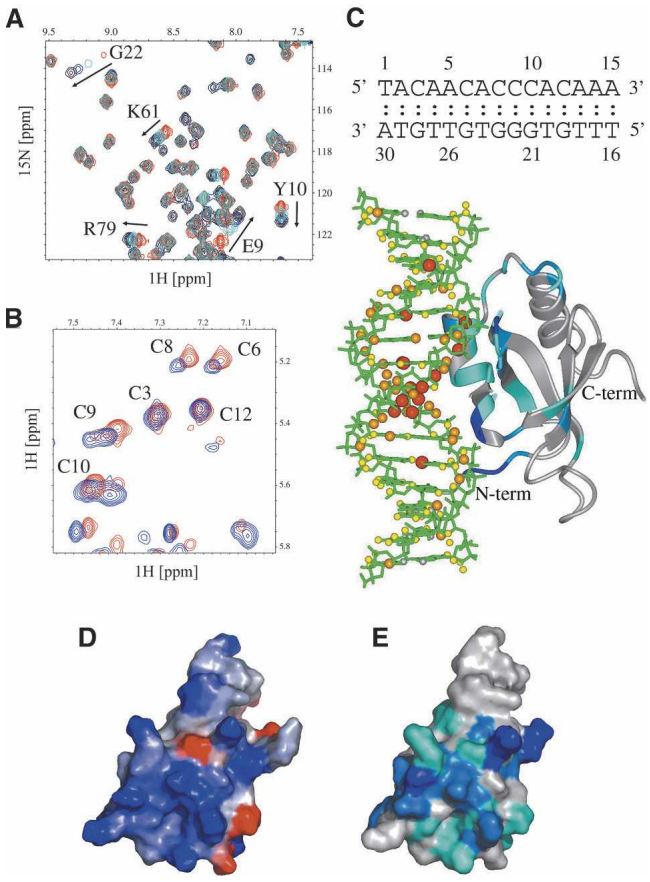

Figure 3. Mapping of HDGF–DNA interactions by NMR. (*A*) ^{15}N-^{1}H HSQC of $HDGF_{0–110}$ PWWP domain (red) titrated with 0.5 (cyan), 1 (blue), and 2 (black) molar equivalents of DNA. (*B*) Fragment of 2D NOESY of DNA (blue) and DNA–$HDGF_{0–110}$ (red) in a 1.5:1 molar ratio. (*C*) Model of the complex between DNA and HDGF based on chemical-shift perturbations observed in the PWWP domain and DNA. Chemical-shift differences for $HDGF_{0–110}$ and DNA were calculated according to the formulas $\Delta\delta_{HDGF} = \sqrt{\Delta\delta_{HN}^2 + 0.11*\Delta\delta_N^2}$ and $\Delta\delta_{DNA} = |\Delta\delta_H|$, respectively. Chemical-shift changes for HDGF are color-coded according to the following scheme: blue, $\Delta\delta_{HDGF} > 0.1$; light blue, $0.05 \geq \Delta\delta_{HDGF} > 0.1$; and cyan, $0.01 \geq \Delta\delta_{HDGF} > 0.05$. Chemical-shift perturbations of DNA protons are proportional to the sphere radii and colored as follows: red, $\Delta\delta_{DNA} > 0.025$; orange, $0.025 \geq \Delta\delta_{DNA} > 0.015$; and yellow, $\Delta\delta_{DNA} \leq 0.015$. (*D*) Electrostatic map of the $HDGF_{0–110}$ PWWP domain (blue, positive electrostatic potential; red, negative electrostatic potential). This map was generated using the GRASS server (Nayal et al. 1999); (*E*) Contour plot of the $HDGF_{0–110}$ PWWP domain in the same orientation as in *C* with residues colored according to the magnitude of the observed chemical-shift change upon addition of DNA.

complementary to double-stranded DNA and consistent with the pattern of observed chemical-shift changes.

Finally, we carried out a titration experiment using $HDGF_{0–110}$ and a single-stranded poly-GC oligonucleotide. No chemical-shifts changes were observed, indicating that the protein does not interact with single-stranded DNA.

Discussion

Comparison of the HDGF PWWP domain and mHRP-3 structures

The solution structure of the HDGF PWWP domain was previously reported by Sue et al. (2004). The two structures were aligned using residues 9–33 and 44–96 (Fig. 1C). This excludes the unstructured loop L2, which varies between homologous PWWP domains. The backbone global r.m.d. deviation for these residues is 3.06 ± 0.11 Å. A backbone RMS deviation of 1.78 ± 0.10 Å was obtained for the β-barrel region (residues 9–33, 44–65) and 4.32 ± 0.20 Å for the C-terminal region (66–96). Sue et al. (2004) reported that the C-terminal portion of HDGF PWWP domain contains two α helices. In contrast, our data clearly shows that the C-terminal domain of $HDGF_{0–110}$ consists of two short turns (residues 66–73) and an extended α helix (82–91). We have no evidence of a continuous i to i + 4 hydrogen bonding pattern in this region, observing hydrogen bonding only between Pro65 and Ser69 and between Ser69 and Phe73. It is worthwhile to emphasize that significant improvement in the accuracy of this region was obtained upon introduction of RDCs in the refinement procedure. In order to assess the previous structure of HDGF using RDCs, we calculated quality factors using $^{1}D_{HN}$ couplings. We obtained a very high Q value of $75.8\% \pm 2.2\%$ that reveals poor agreement between the structure and the experimental dipolar couplings. This observation further demonstrates that our structure has significantly improved accuracy.

The PWWP domain structure of mHRP-3 has recently been determined (Nameki et al. 2005). The construct utilized has 75% identity to the PWWP domain of HDGF. Comparison with our structure indicates a backbone RMS deviation between the proteins of 1.35 ± 0.06 Å for residues 10–33 and 44–90 (Fig. 1D). This includes an RMS deviation of 1.40 ± 0.08 Å (residues 10–33, 44–65) for the conserved β-barrel and 1.24 ± 0.14 Å for the C-terminal domain (residues 66–90). Some small differences between the two structures can be traced to the amino acid sequence, such as the absence of Val95 in HPR-3, resulting in the lack of a hydrophobic stabilization interaction with the α_4 helix. The sequence and structural similarity between these two proteins is strongly suggestive of a very similar function.

Comparison to the Dnmt3b PWWP domain

Although the function of the PWWP domain is presently unknown, it is found in many chromatin-associated proteins (Stec et al. 2000). DNA-binding activity of the PWWP domain from the DNA methyltransferase

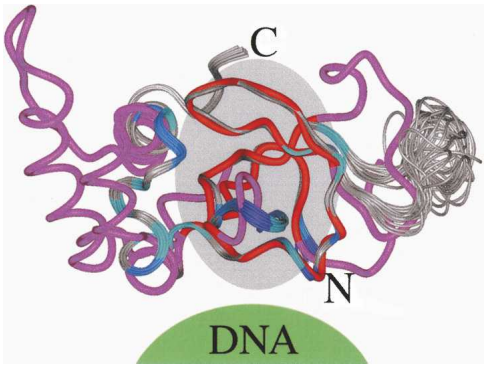

Figure 4. Comparison of the $HDGF_{1–110}$ PWWP domain solution structure with the X-ray structure of the Dnmt3b PWWP domain (PDB code 1khc). A set of 26 conformers of the $HDGF_{1–110}$ PWWP domain is shown in gray, with additional coloring of residues mapped in the DNA-binding experiment (cyan, light blue, and blue; see Fig. 3C for explanation). The structure of Dnmt3b is shown in magenta. DNA-binding site elucidated from NMR titration experiment is represented as a green sphere.

Dnmt3b was demonstrated by in vitro experiments (Qiu et al. 2002). Recently, it was shown that this domain is essential for chromatin targeting of Dnmt3a and Dnmt3b, being localized to heterochromatin in interphase and at specific loci during metaphase (Ge et al. 2004). While the PWWP domain alone binds to metaphase chromosomes, additional regions are required for heterochromatin association.

A comparison of the PWWP domains from HDGF and Dnmt3b is shown in Figure 4 (Qiu et al. 2002). Despite low sequence identity for the β-barrel cores of the two proteins (29%), the structures are very similar. The backbone RMS deviation between β-barrel residues of $HDGF_{0–110}$ and Dnmt3b PWWP domains is 0.70 ± 0.02 Å. As shown previously, this structural motif is conserved throughout all PWWP domains (Qiu et al. 2002; Slater et al. 2003; Sue et al. 2004; Nameki et al. 2005). However, significant differences between $HDGF_{0–110}$ and Dnmt3b are present in the C-terminal α-helical domain, which is substantially longer in Dnmt3b (67 residues in Dnmt3b vs. 25 residues in $HDGF_{0–110}$). Overall, the highest structural similarity between the two domains exists for the interface, which in the case of $HDGF_{0–110}$ is involved in DNA binding (Fig. 4), suggestive of a similar function for this domain in the two proteins.

PWWP domain function

Our results are inconclusive as to a specific DNA-binding site for the PWWP domain of HDGF. SAAB results indicate some preference for a 5′-CACC-3′ sequence, found in 61% of the unique sequenced clones. CACC-binding elements have been previously identified in regulatory regions of muscle-specific genes (Biesiada et al. 1999; Esser et al. 1999). In particular, zinc finger domains such as Sp1 and RIN ZF have been shown to regulate gene expression using this element. Recently, Dellow et al. (2001) have identified a new class of CACC transcription factors, HCB1 and HCB2 (heart CACC-box binding factors). Binding of both proteins to DNA is not affected by chelation agents. This suggests that HCB proteins are not zinc finger DNA-binding domains. The protein sequences of these transcription factors have yet to be determined. It has been demonstrated that Dnmt3b associates with DNA. Recently, Chen et al. (2004) showed that the Dnmt3b PWWP domain binds DNA nonspecifically. Competition experiments utilizing salmon sperm DNA inhibited binding to specific targeted probes. This is in agreement with our results of a weak or nonspecific interaction of the PWWP domain of HDGF with DNA.

The C-terminal portion of the PWWP domain differs between proteins (Fig. 4). While the β-barrel region is highly conserved, the C-terminal helical region differs strikingly. Therefore, it is possible that the C-terminal region of the PWWP domain is involved in a protein–protein interaction that targets this domain to an appropriate region of DNA. In this case, a weak interaction with DNA would contribute to specificity.

Materials and methods

Expression and purification of HDGF PWWP domains

cDNA coding for residues 0–142 and 0–110 of HDGF were cloned into the NCOI/XhoI sites of the pHIS Parallel II vector using standard methods (Sheffield et al. 1999). ^{13}C, ^{15}N-labeled protein was generated from minimal medium with ^{15}N ammonium sulfate and ^{13}C glucose (Cambridge Isotope Labs). Unlabeled protein was generated from 2×YT medium (Q-Biogene). In both cases, cells were grown to an OD_{600} of 0.500, at which point they were induced at 30°C for 6 h with 0.8 mM IPTG. His_6-tagged proteins were purified with Ni^{2+} NTA agarose chromatography. The affinity tag was cleaved with rTEV protease. Proteins were further purified by SP-Sepharose (Amersham-Pharmacia) ion exchange chromatography or Sephacryl S-100 size exclusion chromatography. The protein was then dialyzed into NMR buffer, 20 mM KPi (pH 6.5), 50 mM KCl, 1 mM DTT. Protein concentration was determined by absorbance measurements at 280 nm, with an extinction coefficient of 18,500 (cm/M) (Pace et al. 1995).

Limited proteolysis

Limited proteolysis was utilized to determine the PWWP core fragment in HDGF. $HDGF_{0–142}$ was incubated with 0.001 U of subtilisin for 3 h. The resulting fragment was gel extracted and analyzed by MALDI-TOF mass spectrometry at the University of Virginia Keck Core Facility.

NMR spectroscopy

NMR spectra were measured on Varian Inova 500 and 600 MHz NMR spectrometers at 28°C. Protein samples prepared

for structure determination contained 0.8 mM ^{15}N- or $^{13}C/^{15}N$-labeled $HDGF_{0–110}$ in the previously mentioned buffer supplemented with 10% D_2O. Backbone assignments were obtained from 3D HNCACB, CBCA(CO)NH, HN(CA)CO, and HNCO spectra. Side-chain chemical shifts were assigned from CC(CO)NH-TOCSY, HC(CO)NH-TOCSY, and HCCH-TOCSY recorded with 8- and 24-msec mixing times. For the structure determination, we measured 3D ^{15}N-edited NOESY-HSQC (60-msec mixing time) and two ^{13}C-edited NOESY-HSQC spectra (80-msec mixing times) for aliphatic and aromatic regions. An HNHA experiment was recorded to measure $^3J_{HNH\alpha}$ coupling constants (Kuboniwa et al. 1994). All NMR spectra were processed and analyzed using NMRPipe (Delaglio et al. 1995) and Sparky (T.D. Goddard and J.M. Kneller, University of California, San Francisco).

Measurement of dipolar couplings

The sample used for the measurement of dipolar couplings contained 1 mM protein in 20 mM potassium phosphate buffer (pH 6.5) and 200 mM NaCl. The anisotropic coupling constants have been measured in the presence of a positively charged gel compressed to 4% concentration and consisting of 75% (3-acrylamidopropyl)-trimethylammonium chloride and 25% acrylamide (75 + M) (Cierpicki and Bushweller 2004). Four types of couplings—$^1D_{HN}$, $^2D_{HNC'}$, $^1D_{NC'}$, and $^1D_{C'C\alpha}$—were measured using 3D HNCO-based experiments (Yang et al. 1999), and the range for observed RDCs was –20.4 to 15.4, –4.4 to 5.5, –1.6 to 2.3, and –3.2 to 3.9 Hz, respectively. Analysis of RDCs has been carried out in the program PALES (Zweckstetter and Bax 2000) and validation of the structures was based on the quality factor Q, derived from the formula: $Q = \mathrm{rms}(D^{calc} - D^{exp})/\mathrm{rms}(D^{exp})$, where D^{exp} and D^{calc} are experimental and calculated couplings, respectively (Cornilescu et al. 1998).

Structure determination

The assignment of NOESY cross-peaks and structure calculations have been carried out in an automated manner using CYANA 1.0.5 (Herrmann et al. 2002). For this purpose, we used manually assigned chemical shifts and cross-peaks derived from a 3D ^{15}N-edited NOESY (60-msec mixing time) and two 3D ^{13}C-edited NOESY spectra (80-msec mixing times) recorded for the aliphatic and aromatic ^{13}C regions. Further structural restraints included TALOS-derived ϕ and ψ dihedral angles (Cornilescu et al. 1999) and 23 interstrand hydrogen bonds identified based on the pattern of long-range NOEs in neighboring β-strands.

The final structure of the HDGF PWWP domain was calculated using the simulated annealing protocol in CNS (Brunger et al. 1998). The refinement was based on a large set of distances obtained from CYANA/CANDID supplemented with backbone dihedral angles, hydrogen bond restraints, $^3J_{HNH\alpha}$ coupling constants from an HNHA spectrum, and residual dipolar couplings. Parameters of the alignment tensor have been obtained from fitting of RDCs to the preliminary structure of HDGF obtained from CYANA/CANDID using PALES (Zweckstetter and Bax 2000). The final optimized values of the magnitude (D_a) and rhombicity (R) of the alignment tensor are: $D_a = 9.0$ Hz, $R = 0.4$. Force constants for dipolar couplings have been adjusted to reflect RMS deviations for the calculated structure within the order of measurement accuracy. Initial values of the force constants of 0.01 kcal/mol Hz^2 were ramped to 0.75, 0.3, 0.2, and 0.2 kcal/mol Hz^2 for $^1D_{HN}$, $^2D_{HNC'}$, $^1D_{NC'}$, and $^1D_{C'C\alpha}$, respectively. A total of 400 structures have been calculated, and an ensemble of 26 conformers have been selected based on energy criteria for further analysis.

Selected and amplified binding (SAAB)

We carried out a selected and amplified binding (SAAB) selection to determine DNA-binding site preference for the HDGF PWWP domain. We utilized a double-stranded 72 mer containing a 30-bp randomized sequence. We used a similar protocol as described in Lu et al. (1994). Six rounds of selection were performed utilizing the amplified PCR product as a binding library after the first round. His-tagged $HDGF_{0–142}$ bound to Ni-NTA beads was incubated with the DNA library for 2 h and washed four times with TE Buffer. After the final wash, the beads were boiled for 10 min in 30 μL of TE buffer to release the DNA library. Washes and PCR reactions were monitored using TAE-agarose gels and visualized with ethidium bromide. The final library was cloned into PCR blunt (Invitrogen) and sequenced with the M13 reverse primer.

Protein–DNA interactions monitored by NMR

DNA utilized for NMR titrations was annealed by mixing equal aliquots of two oligomers, heating to 85°C, and slow cooling to room temperature. Excess single-stranded oligomers were removed by Q-Sepharose ion exchange chromatography. In order to analyze protein–DNA interactions, we titrated 0.2 mM ^{15}N-labeled HDGF with 0.5, 1, and 2 equivalents of oligonucleotide and used ^1H-^{15}N HSQC spectra to monitor protein chemical-shift changes. The following DNA sequences have been used: 5′-TACAA**CACC**CACAAA-3′ and 5′-TACAA**CATT**CACAAA-3′. In order to analyze chemical-shift changes for the DNA, we carried out proton assignment based on homonuclear 2D TOCSY and 2D NOESY spectra.

Finally, in order to test whether chemical-shift changes and the structure of $HDGF_{0–110}$ are complementary to DNA structure, we used InsightII to build a model of the complex using double-stranded B-DNA matching the sequence used in the NMR experiments. The alignment of DNA with the binding site on HDGF was obtained by visual inspection with particular attention toward matching the pattern of chemical shift changes, shape complementarity, absence of steric clashes (especially with the protein backbone), and favorable electrostatics. Finally, we performed a brief minimization in order to remove steric clashes between the protein side chains and DNA.

Accession number

Coordinates for 26 structures of $HDGF_{0–110}$ PWWP have been deposited in the Protein Data Bank (PDB) with code 2B8A.

Acknowledgments

This work is supported by grants from the United States Public Health Service to J.H.B.

References

Ayton, P.M., Chen, E.H., and Cleary, M.L. 2004. Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol. Cell. Biol.* **24:** 10470–10478.

Bailey, T.L. and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* (eds. R. Altman et al.), pp. 28–36. AAAI Press, Menlo Park, CA.

Baldin, V., Roman, A.M., Bosc-Bierne, I., Amalric, F., and Bouche, G. 1990. Translocation of bFGF to the nucleus is G1 phase cell cycle specific in bovine aortic endothelial cells. *EMBO J.* **9** 1511–1517.

Biesiada, E., Hamamori, Y., Kedes, L., and Sartorelli, V. 1999. Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac α-actin promoter. *Mol. Cell. Biol.* **19:** 2577–2584.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Casscells, W. 1992. Migration of smooth muscle and endothelial cells. Critical events in restenosis. *Circulation* **86:** 723–729.

Chen, T., Tsujimoto, N., and Li, E. 2004. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol. Cell. Biol.* **24:** 9048–9058.

Cierpicki, T. and Bushweller, J.H. 2004. Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins. *J. Am. Chem. Soc.* **126:** 16259–16266.

Cornilescu, G., Marquardt, J.L., Ottiger, M., and Bax, A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120:** 6836–6837.

Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13:** 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277–293.

Dellow, K.A., Bhavsar, P.K., Brand, N.J., and Barton, P J. 2001. Identification of novel, cardiac-restricted transcription factors binding to a CACC-box within the human cardiac troponin I promoter. *Cardiovasc. Res.* **50:** 24–33.

Dikic, I. 2003. Mechanisms controlling EGF receptor endocytosis and degradation. *Biochem. Soc. Trans.* **31:** 1178–1181.

Dodge, A.B., Lu, X., and D'Amore, P.A. 1993. Density-dependent endothelial cell production of an inhibitor of smooth muscle cell growth. *J. Cell. Biochem.* **53:** 21–31.

Esser, K., Nelson, T., Lupa-Kimball, V. and Blough, E. 1999. The CACC box and myocyte enhancer factor-2 sites within the myosin light chain 2 slow promoter cooperate in regulating nerve-specific transcription in skeletal muscle. *J. Biol. Chem.* **274:** 12095–12102.

Everett, A.D., Lobe, D.R., Matsumura, M.E., Nakamura, H., and McNamara, C.A. 2000. Hepatoma-derived growth factor stimulates smooth muscle cell growth and is expressed in vascular development. *J. Clin. Invest.* **105:** 567–575.

Everett, A.D., Narron, J.V., Stoops, T., Nakamura, H., and Tucker, A. 2004. Hepatoma-derived growth factor is a pulmonary endothelial cell-expressed angiogenic factor. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **286:** L1194–L1201.

Frid, M.G., Aldashev, A.A., Crossno, J.T., Jorgensen, J.M., Kale, V.A., and Stenmark, K.R. 2004. Yin and Yang of an endothelial cell: From normal to the extreme in growth, secretion, and transdifferentiation capabilities. *Paediatr. Respir. Rev.* **5 (Suppl. A):** S253–S257.

Gabriel, B., Baldin, V., Roman, A.M., Bosc-Bierne, I., Noaillac-Depeyre, J., Prats, H., Teissie, J., Bouche, G., and Amalric, F. 1991. Localization of peptide growth factors in the nucleus. *Methods Enzymol.* **198:** 480–494.

Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A., and Xu, G.L. 2004. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J. Biol. Chem.* **279:** 25447–25454.

Graham, I.R. and Chambers, A. 1994. Use of a selection technique to identify the diversity of binding sites for the yeast RAP1 transcription factor. *Nucleic Acids Res.* **22:** 124–130.

Herrmann, T., Guntert, P., and Wuthrich, K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319:** 209–227.

Herve, M.A., Buteau-Lozano, H., Mourah, S., Calvo, F., and Perrot-Applanat, M. 2005. VEGF189 stimulates endothelial cells proliferation and migration in vitro and up-regulates the expression of Flk-1/KDR mRNA. *Exp. Cell Res.* **309:** 24–31.

Honegger, A.M., Dull, T.J., Felder, S., Van Obberghen, E., Bellot, F., Szapary, D., Schmidt, A., Ullrich, A. and Schlessinger, J. 1987. Point mutation at the ATP binding site of EGF receptor abolishes protein-tyrosine kinase activity and alters cellular routing. *Cell* **51:** 199–209.

Iwasaki, T., Nakagawa, K., Nakamura, H., Takada, Y., Matsui, K., and Kawahara, K. 2005. Hepatoma-derived growth factor as a prognostic marker in completely resected non-small-cell lung cancer. *Oncol. Rep.* **13:** 1075–1080.

Kishima, Y., Yoshida, K., Enomoto, H., Yamamoto, M., Kuroda, T., Okuda, Y., Uyama, H., and Nakamura, H. 2002. Antisense oligonucleotides of hepatoma-derived growth factor (HDGF) suppress the proliferation of hepatoma cells. *Hepatogastroenterology* **49:** 1639–1644.

Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. 1994. Measurement of HN-H α J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR* **4:** 871–878.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8:** 477–486.

Lu, Q., Wright, D.D., and Kamps, M.P. 1994. Fusion with E2A converts the Pbx1 homeodomain protein into a constitutive transcriptional activator in human leukemias carrying the t(1;19) translocation. *Mol. Cell. Biol.* **14:** 3938–3948.

McMullen, M.E., Bryant, P.W., Glembotski, C.C., Vincent, P.A. and Pumiglia, K.M. 2005. Activation of p38 has opposing effects on the proliferation and migration of endothelial cells. *J. Biol. Chem.* **280:** 20995–21003.

McNamara, C.A., Sarembock, I.J., Gimple, L.W., Fenton 2nd, J.W., Coughlin, S.R., and Owens, G.K. 1993. Thrombin stimulates proliferation of cultured rat aortic smooth muscle cells by a proteolytically activated receptor. *J. Clin. Invest.* **91:** 94–98.

Nakamura, H., Izumoto, Y., Kambe, H., Kuroda, T., Mori, T., Kawamura, K., Yamamoto, H., and Kishimoto, T. 1994. Molecular cloning of complementary DNA for a novel human hepatoma-derived growth factor. Its homology with high mobility group-1 protein. *J. Biol. Chem.* **269:** 25143–25149.

Nameki, N., Tochio, N., Koshiba, S., Inoue, M., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Fujikura, Y., Saito, M., et al. 2005. Solution structure of the PWWP domain of the hepatoma-derived growth factor family. *Protein Sci.* **14:** 756–764.

Nayal, M., Hitz, B.C., and Honig, B. 1999. GRASS: A server for the graphical representation and analysis of structures. *Protein Sci.* **8:** 676–679.

Oliver, J.A. and Al-Awqati, Q. 1998. An endothelial growth factor involved in rat renal development. *J. Clin. Invest.* **102:** 1208–1219.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4:** 2411–2423.

Qiu, C., Sawada, K., Zhang, X., and Cheng, X. 2002. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat. Struct. Biol.* **9:** 217–224.

Rakowicz-Szulczynska, E.M., Rodeck, U., Herlyn, M., and Koprowski, H. 1986. Chromatin binding of epidermal growth factor, nerve growth factor, and platelet-derived growth factor in cells bearing the appropriate surface receptors. *Proc. Natl. Acad. Sci.* **83:** 3728–3732.

Ren, H., Tang, X., Lee, J.J., Feng, L., Everett, A.D., Hong, W.K., Khuri, F.R., and Mao, L. 2004. Expression of hepatoma-derived growth factor is a strong prognostic predictor for patients with early-stage non-small-cell lung cancer. *J. Clin. Oncol.* **22:** 3230–3237.

Sattler, M., Schleucher, J. and Griesinger, C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucleic Magn. Reson. Spectrosc.* **34:** 93–158.

Schlessinger, J. 2000. Cell signaling by receptor tyrosine kinases. *Cell* **103:** 211–225.

Sheffield, P., Garrard, S., and Derewenda, Z. 1999. Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. *Protein Expr. Purif.* **15:** 34–39.

Slater, L.M., Allen, M.D., and Bycroft, M. 2003. Structural variation in PWWP domains. *J. Mol. Biol.* **330:** 571–576.

Stec, I., Nagl, S.B., van Ommen, G.J., and den Dunnen, J.T. 2000. The PWWP domain: A potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Lett.* **473:** 1–5.

Sue, S.C., Chen, J.Y., Lee, S.C., Wu, W.G., and Huang, T.H. 2004. Solution structure and heparin interaction of human hepatoma-derived growth factor. *J. Mol. Biol.* **343:** 1365–1377.

Walker, L.N., Bowen-Pope, D.F., Ross, R., and Reidy, M.A. 1986. Production of platelet-derived growth factor-like molecules by cultured arterial smooth muscle cells accompanies proliferation after arterial injury. *Proc. Natl. Acad. Sci.* **83:** 7311–7315.

Warder, D.E. and Keherly, M.J. 2003. Ciz1, Cip1 interacting zinc finger protein 1 binds the consensus DNA sequence ARYSR(0–2)YYAC. *J. Biomed. Sci.* **10:** 406–417.

Winkles, J.A., Friesel, R., Burgess, W.H., Howk, R., Mehlman, T., Weinstein, R., and Maciag, T. 1987. Human vascular smooth muscle cells both express and respond to heparin-binding growth factor I (endothelial cell growth factor). *Proc. Natl. Acad. Sci.* **84:** 7124–7128.

Yang, D., Venters, R.A., Choy, W.Y., and Kay, L.E. 1999. TROSY-based HNCO pulse sequences for the measurement of 1HN-15N, 15N-13CO, 1HN-13CO, 13CO-13Cα and 1HN-13Cα dipolar couplings in 15N, 13C, 2H-labeled proteins. *J. Biomol. NMR* **14:** 333–343.

Zhou, Z., Yamamoto, Y., Sugai, F., Yoshida, K., Kishima, Y., Sumi, H., Nakamura, H., and Sakoda, S. 2004. Hepatoma-derived growth factor is a neurotrophic factor harbored in the nucleus. *J. Biol. Chem.* **27:** 27320–27326.

Zweckstetter, M. and Bax, A. 2000. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122:** 3791–3792.