THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 20, pp. 15057–15064, May 18, 2007
Printed in the U.S.A.

Intrinsic Disorder and Autonomous Domain Function in the Multifunctional Nuclear Protein, MeCP2*[S]

Received for publication, January 30, 2007, and in revised form, March 12, 2007 Published, JBC Papers in Press, March 19, 2007, DOI 10.1074/jbc.M700855200

Valerie H. Adams[‡], Steven J. McBryant[‡], Paul A. Wade[§], Christopher L. Woodcock[¶], and Jeffrey C. Hansen[‡1]

From the [‡]Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523, the [¶]Department of Biology, University of Massachusetts, Amherst, Massachusetts 01003, and the [§]Laboratory of Molecular Carcinogenesis, NIEHS, National Institutes of Health, Research Triangle Park, North Carolina 27709

To probe the tertiary structure and domain organization of native methyl CpG-binding protein 2 (MeCP2), the recombinant human e2 isoform was purified to homogeneity and characterized by analytical ultracentrifugation, CD, and protease digestion. The location of intrinsic disorder in the MeCP2 sequence was predicted using the FoldIndex algorithm. MeCP2 was found to be monomeric in low and high salt and over a nearly 1000-fold concentration range. CD indicated that the MeCP2 monomer was nearly 60% unstructured under conditions where it could preferentially recognize CpG dinucleotides and condense chromatin. Protease digestion experiments demonstrate that MeCP2 is composed of at least six structurally distinct domains, two of which correspond to the well characterized methyl DNA binding domain and transcriptional repression domain. These domains collectively are organized into a tertiary structure with coil-like hydrodynamic properties, reflecting the extensive disorder in the MeCP2 sequence. When expressed as individual fragments, the methyl DNA binding domain and transcriptional repression domain both could function as nonspecific DNA binding domains. The unusual structural features of MeCP2 provide a basis for understanding MeCP2 multifunctionality *in vitro* and *in vivo*. These studies also establish an experimental paradigm for characterizing the tertiary structures of other highly disordered proteins.

Methyl CpG-binding protein 2 (MeCP2)[2] is a 53-kDa nuclear protein named for its methylated DNA-binding capacity (1, 2). Accordingly, MeCP2 can preferentially recognize methylated DNA and act as a methylation-dependent transcriptional repressor *in vitro* and *in vivo* (3, 4). MeCP2 is also involved in the maintenance of condensed chromosomal superstructures *in vivo* (5, 6) and *in vitro* (7, 8) and regulates mRNA splicing *in vivo* (9). MeCP2 is able to interact with many different macromolecules and macromolecular complexes, including unmethylated and methylated DNA (10–12), nucleosomes and chromatin (7, 8, 13), transcriptional co-repressors (14), a histone H3 methyltransferase (15), Dnmt1 DNA methyltransferase (16), PU.1 (17), and Y box-binding protein 1 and other splicing factors (18). MeCP2 has two well defined functional domains. Residues 78–162 are required to specifically recognize methylated CpG dinucleotides and have been termed the methyl DNA binding domain (MBD) (19). The minimal sequence needed to repress transfected DNA has been called the transcriptional repression domain (TRD) and consists of residues 207–310 (20). Despite the extensive interest in MeCP2 function, very little is known beyond the tertiary structure of the protein outside of the MBD, and even this domain is not well understood at the biochemical level. For example, fully one-half of the residues required to recognize a single methylated CpG are disordered in the NMR structure of the MBD (21–23). The importance of deciphering the structural basis of MeCP2 function is further underscored by its central role in the neurological disorder, Rett Syndrome, which is caused by a number of different nonsense, missense, and frameshift mutations scattered throughout the *Mecp2* gene (24–26).

Intrinsically disordered proteins have one or more long regions that do not on their own fold into α-helices or β-sheets/turns (27–31). Structural genomics studies indicate that many eukaryotic proteins contain one or more intrinsically disordered regions, including a large number involved in genome regulation (32, 33). Proteins such as the core and linker histones have unstructured terminal regions that serve as combinatorial interaction domains (31). Others such as yeast SIR3p have long internal intrinsically disordered regions (34). The HMGA family of high mobility group proteins are examples of intrinsically disordered proteins that mostly lack secondary and tertiary structure (35). The MeCP2 monomer has been reported to be asymmetric and yields an anomalous molecular mass when characterized by gel filtration and SDS-PAGE (36), each of which can be characteristic of an intrinsically disordered protein (27, 28). To better understand MeCP2 tertiary structure and determine how MeCP2 structure and function may be influenced by intrinsic disorder, we have purified the recombinant human protein and characterized it by analytical ultracentrifugation, circular dichroism, and protease mapping. The

* This work was supported by individual grants from the Rett Syndrome Research Foundation (to P. W., J. C. H., and C. W.), National Institutes of Health Grants GM45916 and GM66834 (to J. H.) and GM70897 (to C. W.), and grants from the Intramural Research Program of NIEHS, National Institutes of Health (to P. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains three supplemental figures.

[1] To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, Colorado State University, 1870 Campus Delivery, Fort Collins, CO 80523. Tel.: 970-491-5586; Fax: 970-491-0494; E-mail: Jeffrey.C.Hansen@colostate.edu.

[2] The abbreviations used are: MeCP2, methyl CpG-binding protein 2; MBD, methyl DNA binding domain; TRD, transcriptional repression domain; HMGA, high mobility group A; MALDI/TOF-MS, matrix-assisted laser desorption/ionization/time of flight mass spectrometry; CTD, C-terminal domain.

location of order and disorder was predicted using the FoldIndex program. We find that the MeCP2 monomer is composed of at least six biochemically distinct domains, which are organized into a tertiary structure that is 60% unstructured and has coil-like hydrodynamic properties. When expressed as individual fragments, the MBD and TRD function as autonomous DNA binding domains *in vitro*. These results demonstrate that the MeCP2 tertiary structure is characteristic of an intrinsically disordered protein and provide a structural basis for understanding the multifunctionality of MeCP2 *in vitro* and *in vivo*.

EXPERIMENTAL PROCEDURES

Expression and Purification of MeCP2—Recombinant human MeCP2 isoform e2 (486 residues) was purified as described (37), with the exceptions that the expression host BL21RP+ was used and the clarified supernatant was allowed to rock overnight in the presence of the chitin agarose. Purified proteins contained the vector-derived sequence, EFLEGSSC, linked to their C-terminal ends. Full-length MeCP2 and the TRD and MBD fragments were eluted with a 300–1000 mM NaCl step gradient (in 100 mM steps), dialyzed into storage buffer (25 mM Tris, pH 7.5, 5% glycerol, 10 mM NaCl, 2 μM β-mercaptoethanol), and stored at 4 °C. Mass spectroscopy analysis (Bruker Ultraflex MALDI/TOF) was performed at the Colorado State University (CSU) Macromolecular Resources core facility. The full-length MeCP2/pTYB1 plasmid was used as the template for generating the MBD and TRD fragments. Primers were designed to amplify DNA fragments corresponding to amino acids 74–172 (MBD) MBD5′, 5′-GAC ATA TGG TGC CGG AAG CTT CTG CC; MBD3′, 5′-CAG AAT TCT TTC TGC TCT CG CCG GGA; 198–305 (TRD) TRD5′, 5′-GA CAT ATG CCC AAG GCG GCC ACG TCA GAG GGT; and TRD3′, 5′-CAG AAT TCG GGG AGT ACG GTC TCC TGC ACA TACGG ATAG. Polymerase chain reaction was used to generate the fragments. The corresponding products were digested with NdeI and EcoR1, gel-purified, and ligated into similarly digested and purified pTYB1 vector DNA. The pTYB1 clones were sequenced (CSU Core Facility) to confirm fidelity and continuity of the open reading frame, and successful clones were selected.

DNA and Chromatin Binding Assays—The preparation of the DNA and nucleosomal arrays for electrophoretic mobility shift assay have been previously described (7). Methylated 198 DNA was generated by incubating 208-12 DNA in the presence of SssI and verified by digesting the DNA with the methylation-sensitive enzyme AvaI. The methylated DNA was then digested with EcoRI, which released two fragments: a 198-bp fragment that contains 10 methylated CG sites and a 10-bp fragment (all enzymes from New England Biolabs). The methylated 198 DNA fragment was purified using a PCR cleanup column (Qiagen) to remove the 10-bp fragments.

Analytical Ultracentrifugation—Sedimentation velocity experiments were performed with either a Beckman XL-I or a Beckman XL-A analytical ultracentrifuge using the absorbance optical system as described (38). Boundaries were analyzed using the method of Demeler *et al.* (39, 40) using Ultrascan (version 7.3). This analysis yields an integral distribution of sedimentation coefficients, $g(s)$. Sedimentation coefficients (s) were corrected to that in water at 20 °C ($s_{20,w}$). The solvent densities (ρ) were calculated in Ultrascan. The partial specific volume of full-length MeCP2 (0.731 cm^3/g at 20 °C) was calculated from the primary amino acid sequence within Ultrascan. Modeling of hydrodynamic parameters was performed within Ultrascan. Sedimentation equilibrium experiments were performed at 5 °C using charcoal-filled Epon 6-sector centerpieces. Scans were collected with the absorbance optical system at 229 and 280 nm by using an average of 20 scans collected at a radial step resolution of 0.001 cm. Overlays of successive scans taken 4 h apart at each speed confirmed that the samples had reached equilibrium. The equilibrium concentration gradients were edited and globally fitted within Ultrascan. The frictional ratio (f/f_o) was calculated from the known molecular mass and measured sedimentation coefficient using Ultrascan.

Circular Dichroism and Prediction Algorithms—CD spectroscopy was performed on a Jasco-720 spectropolarimeter at 20 °C. Proteins were dialyzed extensively against 10 mM K_2HPO_4, pH 7.9, 20 mM NaCl prior to obtaining measurements. A total of 15 spectra from full-length MeCP2 and the TRD and MBD fragments (all at 5 μM) were collected and averaged to obtain the final spectra. The spectra were baseline-subtracted *versus* the dialysis buffer. The molar ellipticity [Θ] was obtained by normalization of the measured ellipticity (Θ, millidegree), where $[\Theta] = (\Theta \times 100)/(nlc)$, n is the number of residues, c is the total concentration (mM), and l is the cell path length (cm). The percentages of secondary structure types were determined from the spectra using the CONTINLL, SELCON3, and CDSSTR methods within CDPro analytical software (41). The SDP48 basis set was used to deconvolute the CD spectrum. The full-length MeCP2 protein sequence (National Center for Biotechnology Information (NCBI) accession number AAH11612) was analyzed by the following algorithms: DPM, DSC, GOR3, HNNC, MLRC, PHD, Predator, and SOPM) (42). The results are reported as the mean ± standard deviation of the values returned by each algorithm. The FoldIndex program (43) was accessed through the Bioinformatics & Biological Computing site. The window and step parameters were set at 6 and 2, respectively.

Trypsin Digestion and N-terminal Sequencing—MeCP2 was dialyzed into 25 mM Tris-HCl, pH 7.5, 10% glycerol, and 150 mM NaCl (digestion buffer). Trypsin (Roche Applied Science) was dissolved in 1 mM HCl, and serial dilutions were prepared (5.0–0.3125 μg/ml into digestion buffer). 10 μl of each trypsin dilution was incubated with 10 μl of MeCP2 (0.8 mg/ml) for 30 min on ice. The reactions were quenched with 5× SDS sample buffer, and the samples were boiled for 3 min and loaded onto a 12.5% SDS-PAGE gel and electrophoresed. Protein bands were visualized by staining the gel with Coomassie Brilliant Blue. Images were visualized by white light epi-illumination and digitized in a GelLogic 200 (Eastman Kodak Co.). For N-terminal sequencing, 8 μg of MeCP2 was digested with either 5 or 2.5 μg/ml trypsin as described. Following electrophoresis, the proteins were electroeluted onto a polyvinylidene difluoride membrane and visualized with RAPIDstain (Calbiochem). Protein bands 1–7 (see Fig. 3*B*) were excised and submitted to the CSU core facility for

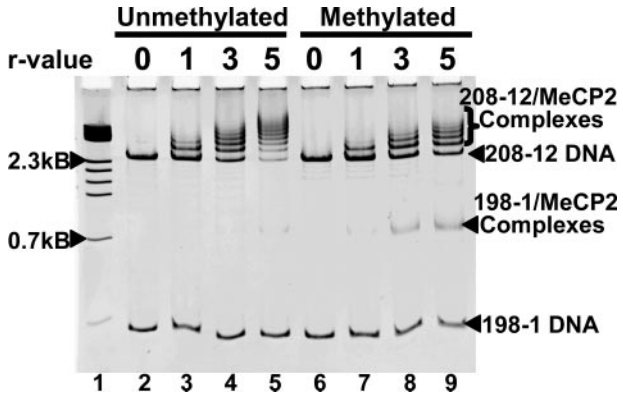

FIGURE 1. **DNA binding properties of purified MeCP2.** A methylated DNA competition assay was performed. MeCP2 at the indicated *r*-values (0.36, 1.09, or 1.82 pmol) was premixed with 1.46 pmol of unmethylated 208-12 competitor DNA and incubated for 25 min at room temperature. 0.36 pmol of unmethylated (*lanes 2–5*) or methylated (*lanes 6–9*) 198-1 DNA was added (final volume was 11 μl (Tris-EDTA, pH 7.5; 100 mM NaCl)), and the mixtures were incubated for 25 min. The samples were electrophoresed on a 5% native polyacrylamide gel as described (53). The *r*-value was defined as the mol of protein added/mol of 208-bp repeat. DNA size standards were loaded in *lane 1*. Following electrophoresis, the gel was visualized using UV transillumination and digitized, and the image was inverted using a GelLogic 200 (Kodak).

N-terminal sequencing (Applied Biosystems Procise Sequencer) and analyzed using Smooth Degree 3 software.

RESULTS

Recombinant human MeCP2 isoform e2 was purified to greater than 95% homogeneity on the large scale (see Fig. 3*B*, *lane 7*) using the Intein system and heparin chromatography (37). MeCP2e2 subsequently will be referred to as MeCP2. MALDI/TOF-MS yielded a MeCP2 molecular mass of 53,541 Da, within 1% of the mass calculated from the amino acid sequence. To assay for methylated DNA binding, purified MeCP2 was bound to unmethylated 208-12 DNA, the complexes were incubated with identical amounts of methylated or unmethylated 198-1 DNA, and the products were analyzed on a native PAGE gel (8). The results indicated that purified MeCP2 preferentially bound to the methylated 198-1 DNA (Fig. 1, compare *lanes 3–5* and *7–9*). The pronounced shifts of the 208-12 competitor further demonstrated that MeCP2 also bound well to unmethylated DNA under the same conditions. The addition of increasing amounts of purified MeCP2 led to a progressive decrease in the mobility of nucleosomal arrays, as observed previously (supplemental Fig. 1) (7, 8). Together, these data show that the purified recombinant human MeCP2 characterized below was able to preferentially recognize methylated DNA, bind to unmethylated DNA, and condense unmethylated nucleosomal arrays into higher order secondary and tertiary chromatin structures.

Hydrodynamic Properties and Oligomeric State—A wide range of sizes has been reported for MeCP2. For example, in the same study, an apparent molecular mass of ~500 kDa was obtained from gel filtration and an apparent molecular mass of ~57 kDa was estimated from sucrose density gradient centrifugation (36). To rigorously assess the oligomeric state of MeCP2 over a large range of protein concentrations and in the presence and absence of salt, the purified protein was characterized by analytical ultracentrifugation. We initially performed sedimentation velocity experiments in low salt buffer covering a 10-fold protein concentration range. *g*(*s*) plots of the diffusion-corrected sedimentation coefficient distributions are shown in Fig. 2*A*. The vertical distributions indicate that MeCP2 sedimented as a single homogeneous 2.2 *s* species at all concentrations examined. The same result was obtained when the sedimentation velocity experiments were performed at 2.9 μM MeCP2, and the NaCl concentration was varied from 2 to 500 mM (Fig. 2*B*). The 2.2 *s* sedimentation coefficient observed in our studies is the same as that obtained using sucrose gradient centrifugation at a single MeCP2 and salt concentration (36).

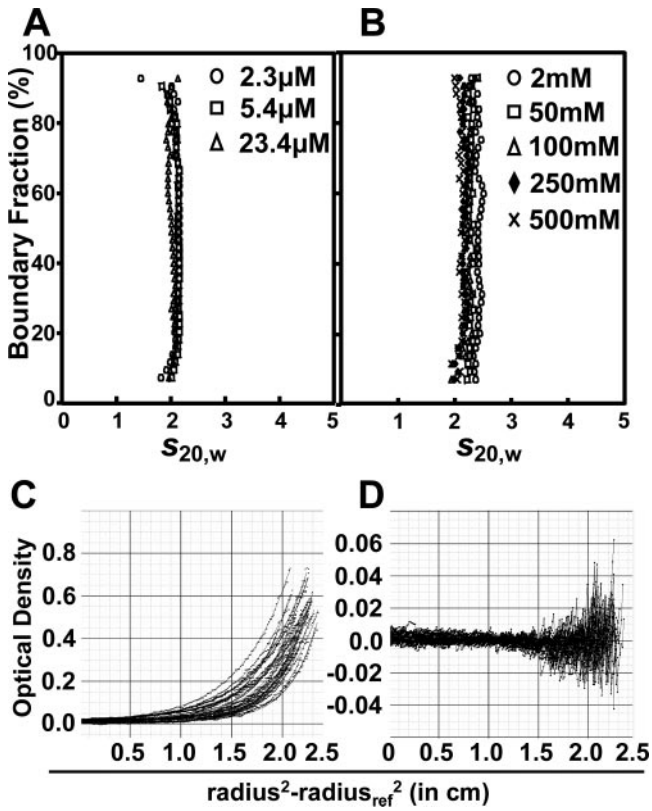

FIGURE 2. **Analytical ultracentrifugation of purified MeCP2.** *A*, sedimentation velocity at different protein concentrations. 2.3 μM (○), 5.4 μM (□), and 23.4 μM (△) MeCP2 in 10 mM Tris HCl, pH 7.5, 2 mM NaCl was sedimented at 50,000 rpm as described under "Experimental Procedures." Scans were collected at 229 (2.3 and 5.4 μM) or 276 nm (23.4 μM). Shown are *g*(*s*) plots obtained after analysis of the boundaries by the method of Demeler and van Holde (40). *B*, sedimentation velocity at different NaCl concentrations. MeCP2 (2.9 μM) in 10 mM Tris HCl, pH 7.5, and 2 (○), 50 (□), 100 (△), 250 (◇), or 500 mM NaCl (×) was sedimented at 48,000 rpm as described under "Experimental Procedures." Scans were obtained at 276 nm. Shown are *g*(*s*) plots. *C*, sedimentation equilibrium. Purified MeCP2 was dialyzed extensively against 20 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 1 μM β-mercaptoethanol prior to the analysis. Samples spanning an initial 30-fold protein concentration range (1.6, 2.68, 3.75, 19.4, 31.2, and 46.9 μM) were sedimented to equilibrium at 24,000, 28,000, and 32,000 rpm. Two equilibrium scans were collected at each speed. The resulting 36 data sets were globally fit to a single ideal species model in Ultrascan. The fits to the data are shown as *solid lines*. *D*, goodness of fit. Shown are the residuals to the fits shown in *panel C*. The variance for the fit to a single ideal component model was 3.46×10^{-5}. The analysis returned a mass of 54,970 Da.

Purified MeCP2 was next sedimented to equilibrium at six different protein loading concentrations (1.6–47 μM) and at three different rotor speeds. The buffer contained 50 mM NaCl. The subsequent concentration gradients obtained at equilib-

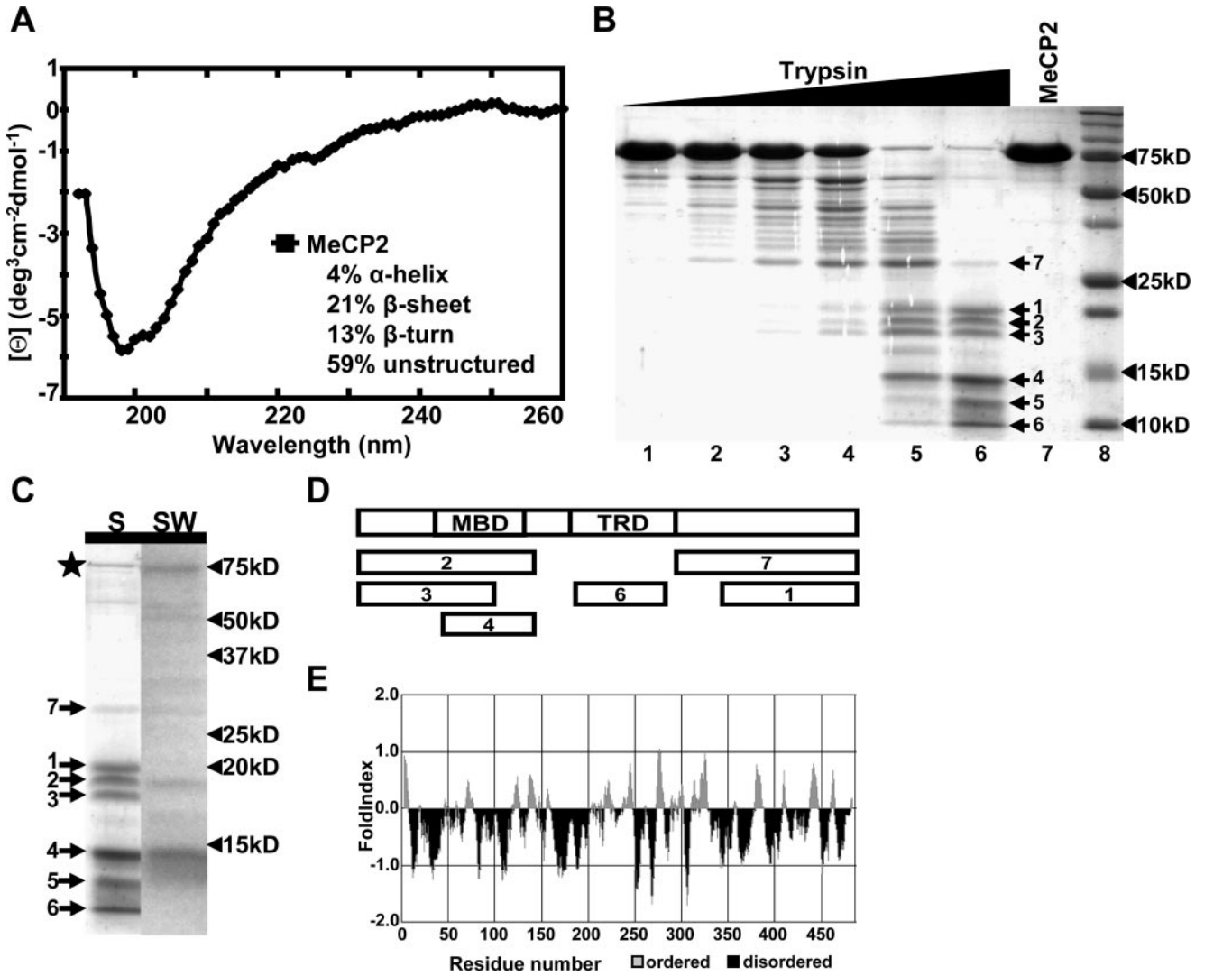

FIGURE 3. **MeCP2 has a disordered tertiary structure.** *A*, circular dichroism. MeCP2 (5 μM) was dialyzed into 10 mM potassium phosphate buffer (pH 7.5), and the far-UV circular dichroism spectrum was recorded in a 0.05-cm path length cell. The spectrum was analyzed with CDPro software, and the percentage of secondary structure (*inset*) was determined as described under "Experimental Procedures." *B*, trypsin digestion. 10 μl of MeCP2 (0.8 μg/μl) was incubated with 0.08, 0.16, 0.32, 0.63, 1.25, 2.5, and 0 μg/ml trypsin (*lanes 1–7*, respectively) for 30 min on ice. Size standards were loaded in *lane 8*. Shown is a 12% SDS-polyacrylamide gel of the digestion products. The seven trypsin-resistant bands excised for N-terminal sequencing are labeled in *lane 7*. *C*, Southwestern (*SW*) analysis. The gel from *panel B* was transferred to an Immobilon-P membrane, and the membrane was incubated with a radiolabeled, methylated oligonucleotide probe as described (8, 13). The *SW lane* shows the resulting 60-min exposure. *Lane 6* from *panel B* (labeled *S*) is shown to the *left* of the *SW lane* for comparison. *D*, schematic representation of the location of the trypsin-resistant bands shown in panel. See "Results" for details. *E*, FoldIndex plot obtained at a window setting of 6 residues.

rium spanned approximately 3 orders of magnitude of protein concentrations, ranging from mid-nanomolar to high micromolar (Fig. 2*C*). 36 different scans were globally fit to a single, ideal species model. The residuals using this model were randomly distributed and showed no systematic deviations (Fig. 2*D*). A molecular mass of 54,970 Da was obtained from the sedimentation equilibrium analysis, within 3% of the mass calculated from the MeCP2 sequence. The frictional coefficient ratio, f/f_o, calculated from the molecular mass and the 2.2 *s* sedimentation coefficient was 2.4. This large frictional coefficient ratio could result from either a rod-like structure with a high axial ratio or a coil-like structure such as that of a denatured protein (34). Together, the data in Fig. 2 demonstrate that MeCP2 is monomeric in the presence of 2–500 mM NaCl and over a nearly 1000-fold range in protein concentration and indicate that MeCP2 is not a well packed, globular protein.

The MeCP2 Tertiary Structure Is Extensively Disordered—To better understand the tertiary structure of MeCP2, the purified protein was characterized by CD spectroscopy and protease digestion. The experimental results were then compared with FoldIndex disorder predictions (43). This same approach has been used to identify extensive disorder in the yeast SIR3p tertiary structure (34). The CD spectrum of purified MeCP2 in low ionic strength buffer is shown in Fig. 3*A*. The most prominent feature of the spectrum was a strong negative peak at 198 nm that arises from the disordered region(s) of a polypeptide chain (44). In contrast, the minimal negative shoulder at ~225 nm indicated that only a small fraction of the polypeptide chain was in an α-helical conformation. The CD spectrum was deconvoluted using CDPro software (41) and a 48-protein reference set that included denatured and/or unfolded proteins (see "Experimental Pro-

cedures"). This analysis yielded values of 4% α-helix, 21% β-sheet, 13% β-turn, and 59% unstructured.

There are 90 consensus trypsin cleavage sites located throughout the 486-residue MeCP2 sequence (the mean and mode distance between consecutive sites is 5.4 and 3 residues, respectively, with a range of 36 residues).[3] Thus, trypsin accessibility is a sensitive probe of global tertiary structure. Purified MeCP2 was digested with increasing amounts of trypsin, and the proteolytic products were visualized by SDS-PAGE (Fig. 3*B*). Of note, the undigested 53-kDa protein yielded an apparent molecular mass of ~75 kDa (Fig. 3*B*, *lane 7*) (36, 37). Digestion of MeCP2 under conditions where most of the protein remained intact produced at least 16 discrete bands, with apparent sizes ranging from ~18–74 kDa (Fig. 3*B*, *lanes 1–4*). At a higher enzyme concentration, most of the high molecular weight bands were depleted, and three new major fragments appeared with apparent sizes of ~10, 12, and 14 kDa (Fig. 3*B*, *lane 5*). At the highest trypsin concentration employed, prominent protease-resistant bands were observed at ~10, 12, 14, 18, 19, and 20 kDa (termed bands 1–6; Fig. 3*B*, *lane 6*). We note that each of these trypsin-resistant bands contained 10–25 trypsin digestion sites that were not utilized. A Southwestern blot of a duplicate gel using a radiolabeled methylated DNA probe was used to identify bands with an intact MBD (8, 13). Of the trypsin-resistant fragments identified in Fig. 3*B*, only band 2 (~19 kDa) and band 4 (~14 kDa) could bind methylated DNA (Fig. 3*C*, compare lanes *S* and *SW*).

The trypsin-resistant bands were excised from the gel and subjected to Edman sequencing. N-terminal sequences were obtained for bands 1–4 and band 6. The C-terminal ends of these bands could only be estimated from sizes returned by SDS-PAGE as we were unable to obtain masses from mass spectroscopy. These estimates took into account the anomalously high masses returned by SDS-PAGE. The results are described below and shown schematically in Fig. 3*D*. Band 1 (~20 kDa apparent; 14 kDa actual) began with residue 355 and was presumed to extend to the C-terminal residue 486. Band 2 (~19 kDa apparent; ~18 kDa actual) began with residue 2 and bound methylated DNA, suggesting that it was composed of residues 2–162. Band 3 (~18 kDa apparent; 14 kDa actual) began with residue 2 and did not bind methylated DNA. This suggests that it spanned residues 2–130/133/135 and was derived from band 2. Band 4 (14 kDa apparent) began with residue 85 and bound methylated DNA, suggesting that its C terminus was at residue 162, 177, or 186 (9, 10, or 11 kDa actual, respectively). Band 2 also may have been derived from band 4. Band 6 (10 kDa apparent; 9 kDa actual) began with residue 212, suggesting that it was composed of residues 212–293. We were unable to sequence band 5 despite repeated attempts but speculate that it may be a slightly larger version of band 6 based on the smearing between the two bands seen only in the SDS gel. During the course of these studies, we also isolated and sequenced the apparent ~27-kDa fragment prominent throughout digestion (band 7) and found that it began with residue 310, suggesting that this band consisted of residues 310–486 (19 kDa actual) and was the direct precursor of band 1 (note that a minor MBD-containing fragment was present at the same size (Fig. 3*C*)). Most of the MeCP2 fragments appeared to exhibit the same anomalous behavior on SDS gels seen with the full-length protein. Even without positive identification of the C-terminal residues, it is apparent that the protease-resistant MeCP2 regions overlapped with the MBD, TRD, and C-terminal residues 355–486. Correspondingly, residues in the region between the MBD and TRD, as well as sites near residues 310 and 355, were relative hot spots for trypsin digestion.

The MeCP2 amino acid sequence was next analyzed using the FoldIndex prediction program. This algorithm uses the hydrophobicity and charge properties of the amino acids to calculate an averaged FoldIndex value for each residue in a protein (43). Residues with positive FoldIndex values are predicted to lie in a classically structured region (*i.e.* helix, sheet, or turn), whereas residues with negative values are predicted to be in a disordered region. Consequently, a plot of FoldIndex value *versus* residue number predicts the location of order and disorder in a protein sequence (43). The MeCP2 FoldIndex plot generated at a window of 6 residues is shown in Fig. 3*E* (and in more detail in supplemental Fig. 2). MeCP2 was predicted to have an unusual structure in which short (5–15 residues) ordered regions punctuated an otherwise highly unstructured sequence. Residues 8–68, 75–120, 160–200, 247–272, and 327–377 were predicted to be almost entirely disordered. Extensive disorder also was predicted in the C-terminal 100 residues. The known and predicted structures of the MBD matched well. The only significant difference was near residues 104–109, which are in a β-strand in the NMR structures (23, 45) and were predicted to be disordered (albeit with small negative FoldIndex values). The FoldIndex predictions closely complement the empirical CD and protease digestion data and further support the conclusion that MeCP2 is an intrinsically disordered protein with an unusual tertiary structure.

The MBD and TRD Can Function as Autonomous Nonspecific DNA Binding Domains—Given the extensive disorder in the MeCP2 structure, we were curious whether the MBD and TRD could function as isolated fragments. Residues 74–172 and 198–305 (encompassing the MBD and TRD, respectively) were cloned, expressed, and purified to >95% homogeneity as judged by SDS-PAGE (Fig. 4*A*). Analysis of the CD spectrum of the MBD (Fig. 4*B*) returned 10% α-helix, 30% β-strand, 21% β-turn, and 38% unstructured, similar to the secondary structure content of the MBD as determined by NMR (23, 45). The isolated MBD retained a small preference for methylated DNA (data not shown) and also bound nonspecifically to unmethylated DNA (Fig. 4*C*, *lanes 10–13*). The isolated TRD fragment was estimated by CD to be 85% unstructured, 3% α-helix, 7% β-strand, 5% β-turn (Fig. 4*B*). Based on electron microscopy analysis of MeCP2-DNA interactions, we have recently suggested that MeCP2 has a second DNA binding domain that lies within residues 1–295 (8). We therefore characterized the isolated TRD by the same DNA binding assay used to study the wild type protein and isolated MBD. The results indicated that the TRD fragment decreased DNA band mobilities comparably with the MBD (Fig. 4*C*, *lanes 6–9*), and that neither domain alone led to the supershifts seen with full-length MeCP2 (Fig.

[3] ExPASy PeptideCutter site.

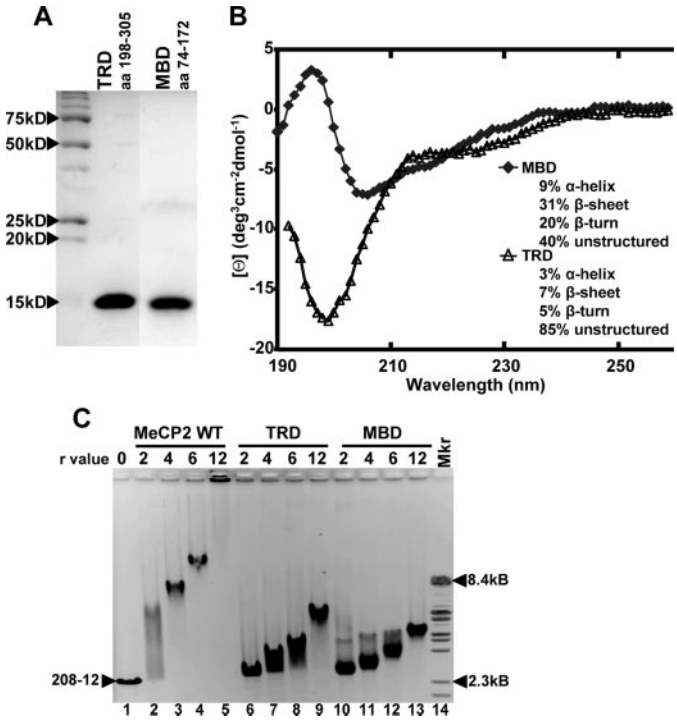

FIGURE 4. **Characterization of isolated MBD and TRD fragments.** *A*, 15% SDS-PAGE gel of the purified MBD and TRD fragments. 5 μg of protein/lane was boiled in 5× SDS loading dye, loaded, and electrophoresed. The protein bands were stained with Coomassie Brilliant Blue and imaged with a GelLogic 200. *B*, circular dichroism. Purified MBD and TRD fragments were characterized as described in the legend for Fig. 3. *C*, DNA binding assay. Full-length MeCP2 (*lanes 2–5*), the TRD (*lanes 6–9*), and the MBD (*lanes 10–13*) at the indicated *r*-values were incubated with 3.6 pmol of 208-12 DNA for 25 min at room temperature (final volume 15 μl in Tris-EDTA, 2.5 mM NaCl) and then electrophoresed on a 0.8% agarose gel (1× Tris acetate EDTA 5V/cm) for 2 h. Size standards were loaded in *lane 1*. The gel was visualized and processed as described in the legend for Fig. 1. Shown is an inverted image of the digitized gel. *WT*, wild type; *Mkr*, marker.

4*C*). When incubated in the presence of 5-fold excess competitor DNA, TRD·DNA and MBD·DNA complexes dissociated to naked DNA (data not shown), indicating that DNA binding was reversible. When tested by the same gel assay, a fragment consisting of residues 2–78 did not shift the DNA over the same protein range (data not shown). The data in Fig. 4 demonstrate that the MBD and TRD have substantially different tertiary structures and that both fragments can function autonomously as nonspecific DNA binding domains *in vitro*.

DISCUSSION

The 53-kDa MeCP2 monomer has a low sedimentation coefficient (2.2 *s*) and high frictional coefficient ratio (f/f_o = 2.4). These hydrodynamic properties are indicative of a rod-like shape if MeCP2 is a rigid, well packed protein. Alternatively, the high f/f_o value could reflect a coil-like conformation, such as that of a denatured protein (see Ref. (34)). CD studies indicate that 60% of full-length MeCP2 is unstructured; likewise, ~40% of the isolated MBD and 85% of the isolated TRD is unstructured. Numerous sites throughout the MeCP2 sequence are trypsin-accessible in the purified protein. The FoldIndex algorithm predicts that the MeCP2 sequence has small regions of secondary structure separated by large regions of disorder. Taken together, the physicochemical data indicate that the tertiary structure of MeCP2 has coil-like hydrodynamic properties and is dominated by many disordered regions. Consistent with this conclusion, individual MeCP2 molecules in electromagnetic images appear as dumbbell-like oblate ellipsoids with no single well defined shape (8). CD experiments suggest that ~35% of the MeCP2 sequence is β-strand/turn. Thus, we speculate that MeCP2 tertiary structure consists of multiple disordered regions that emanate from, and connect, β-sheet secondary structures.

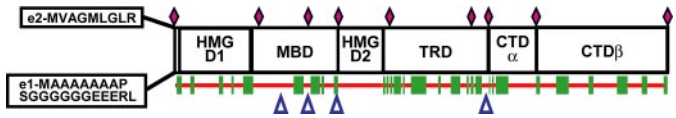

FIGURE 5. **Schematic illustration of MeCP2 domain organization based on biochemical criteria.** The location and sequence of the alternatively spliced region of the e1 and e2 isoforms is shown at the *left*. The boundaries of trypsin-resistant bands are indicated by *purple diamonds*. The four most common missense mutations that combined are responsible for over 30% of Rett cases (R106W, R133C, T158M, and R306C) are indicated by *blue triangles*. The four most common nonsense mutations (R168X, R255X, R270X, and R294X) are not shown. The regions of order predicted by the FoldIndex program at a window of 6 are shown as *green boxes*. See "Discussion" for details.

Our studies demonstrate that native MeCP2 is composed of at least six biochemically distinct domains (Fig. 5). In two cases (the MBD and TRD), the regions mapped by protease digestion closely overlap with domains defined previously based on function. Residue 85 is a prominent trypsin cleavage site (Fig. 3*B*, band 4), suggesting that N-terminal residues 2–84 define a discrete structural unit within MeCP2. Residues 8–68 are very disordered and have an amino acid composition almost identical to that of the HMGB1 and HMGN1 families of intrinsically disordered proteins (supplemental Fig. 3). Amino acid composition is one of the defining characteristics of intrinsically disordered protein regions (43). We therefore have termed residues 2–77 the HMGD1. We note that the only difference between the MeCP2e2 and MeCP2e1 isoforms lies at the N terminus of the HMGD1. Through alternative splicing, residues 2–9 of the e2 isoform are exchanged for an entirely different 21-residue sequence in the e1 isoform (Fig. 5) (46).

The MBD has been defined as the minimum sequence needed to recognize methylated CpG dinucleotide pairs and encompasses residues 78–162 (19, 23). In our experiments, a fragment with an N terminus at residue 85 recognized methylated DNA (Fig. 3*B*). The structured portion of the MBD consists of a four-stranded β-sheet (residues 93–133) and a three-turn α-helix (residues 135–144). Residues 78–102 are largely disordered (23) and appear to mediate specific interactions with a histone H3 methyltransferase (15). Residues 145–162 are unstructured and pack against one side of the β-sheet in the isolated fragment. The NMR structure of the MBD demonstrates that this protease-resistant domain is more disordered than one would expect for a well packed globular protein, consistent with our CD results. In our studies, an engineered MBD-containing fragment folded independently and bound to both methylated and unmethylated DNA, indicating the potential for autonomous function within the native protein.

The region between the MBD and TRD (residues 163–206) is also predicted to be extensively disordered. The amino acid composition of residues 163–206 closely matches that of the

HMGA1 protein (supplemental Fig. 3). Moreover, residues 188–194 in MeCP2 are identical to those found within the AT hook DNA binding domains of HMGA1 (47). Thus, we refer to this region as the high mobility group-like domain 2, or HMGD2 (Fig. 5). The HMGD2 binds to the *Xenopus* p20 protein (48), and residues 174–190 are the site of a predicted nuclear localization signal.[4] Thus, the HMGD2 appears to be a disordered domain that participates in several different macromolecular interactions.

The minimal TRD sequence needed to repress transiently transfected DNA is residues 207–310 (20) (Fig. 5). The TRD has been shown to interact with histone deacetylase complexes (49–51), the proto-oncogene, c-ski (14), and Dnmt1 DNA methyltransferase (16). Residues 253–269 define a predicted nuclear localization sequence.[4] We have previously speculated that residues 1–294 may have additional DNA binding domains besides the MBD (8). In our experiments, an isolated TRD fragment (residues 198–305) was able to bind nonspecifically to DNA with a relative affinity comparable with the MBD (Fig. 4*C*). The presence of this additional DNA domain may explain how MeCP2 is able to "bridge" or cross-link long unmethylated DNA and nucleosomal arrays into oligomeric suprastructures despite being monomeric (Fig. 1) (7, 8). The isolated TRD fragment was estimated by CD to be 85% unstructured. The FoldIndex plot suggests that the TRD can be subdivided into three distinct regions (residues 200–245, residues 250–275, and residues 276–310).

C-terminal residues 310–486 (Fig. 3*B*, band 7), which we term the CTD, appear very early and are detectable late in trypsin digestion (Fig. 3*B*). This indicates that the TRD-CTD border is hypersensitive to protease digestion and that cleavage at this position releases the CTD as a relatively protease-resistant fragment. Once released from the native protein, the CTD is cleaved at position 354 to generate the protease-resistant 355–486 fragment (Fig. 3*B*, band 2). Based on this biochemical distinction, we refer to residues 310–354 as CTDα and 355–486 as CTDβ. The CTDα may contribute to recognition of methylated DNA in chromatin (8). The CTDβ has several noteworthy features. Residues 366–372 make up a run of 7 disordered histidines. Residues 384–387 define the consensus sequence (PPLP) recognized by group 2 WW motif-containing proteins (18, 52). Truncation at residue 404 significantly reduces the intensity of the MeCP2 footprint on mononucleosomes (13). *In vitro* studies have shown that the CTD is required for chromatin-specific interactions and MeCP2-mediated assembly of secondary and tertiary chromatin structures (8). Thus, the CTD appears to contain histone binding sites, a chromatin-condensing function, and binding sites for other proteins.

Several important ramifications follow from our studies of MeCP2 tertiary structure and domain organization. Despite historical focus on its methylation-dependent actions, MeCP2 should be viewed as a complex, multifunctional protein. In our *in vitro* biochemical experiments, MeCP2 was able to recognize CpG dinucleotides, bind nonspecifically to DNA, and potently condense unmethylated nucleosomal arrays (Fig. 1 and supplemental Fig. 1). *In vivo*, MeCP2 has key roles in transcriptional repression (3, 4), chromatin architecture (5, 6), and RNA splicing (9). As discussed above, all MeCP2 domains have the potential to mediate protein-DNA and/or protein-protein interactions. We speculate that the structural autonomy of these domains is related to MeCP2 multifunctionality and arises from the extensive disorder that permeates the MeCP2 sequence. Each MeCP2 domain not only can potentially function alone, but the disordered tertiary structure may allow different combinations of domains to act together for specific functional purposes. For example, the MBD-HMGD2 combination acts as a matrix attachment region binding fragment (8), whereas the HMGD2-TRD is a general co-repressor-interacting domain (5). Because of the widespread prevalence of intrinsic disorder in the protein world (27–33), the experimental dissection of the MeCP2 structure reported here may serve as a tractable experimental paradigm for characterizing other proteins with highly disordered tertiary structures.

Acknowledgment—We thank Robert Woody for helpful comments and advice.

[4] Identified using PSORTII via the PSORT WWW Server.

REFERENCES

1. Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992) *Nucleic Acids Res.* **20,** 5085–5092
2. Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992) *Cell* **69,** 905–914
3. Wade, P. A. (2004) *BioEssays* **26,** 217–220
4. Mann, J., Oakley, F., Akiboye, F., Elsharkawy, A., Thorne, A. W., and Mann, D. A. (2006) *Cell Death Differ.* **14,** 275–285
5. Brero, A., Easwaran, H. P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H., and Cardoso, M. C. (2005) *J. Cell Biol.* **169,** 733–743
6. Horike, S., Cai, S., Miyano, M., Cheng, J. F., and Kohwi-Shigematsu, T. (2005) *Nat. Genet.* **37,** 31–40
7. Georgel, P. T., Horowitz-Scherer, R. A., Adkins, N., Woodcock, C. L., Wade, P. A., and Hansen, J. C. (2003) *J. Biol. Chem.* **278,** 32181–32188
8. Nikitina, T., Shi, X., Ghosh, R. P., Horowitz-Scherer, R. A., Hansen, J. C., and Woodcock, C. L. (2006) *Mol. Cell. Biol.* **27,** 864–877
9. Young, J. I., Hong, E. P., Castle, J. C., Crespo-Barreto, J., Bowman, A. B., Rose, M. F., Kang, D., Richman, R., Johnson, J. M., Berget, S., and Zoghbi, H. Y. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 17551–17558
10. Wade, P. A., Jones, P. L., Vermaak, D., Veenstra, G. J., Imhof, A., Sera, T., Tse, C., Ge, H., Shi, Y. B., Hansen, J. C., and Wolffe, A. P. (1998) *Cold Spring Harbor Symp. Quant. Biol.* **63,** 435–445
11. Harikrishnan, K., Pal, S., Yarski, M., Baker, E. K., Chow, M. Z., de Silva, M. G., Okabe, J., Wang, L., Jones, P. L., Sif, S., and El-Osta, A. (2006) *Nat. Genet.* **38,** 964–967
12. Hu, K., Nan, X., Bird, A., and Wang, W. (2006) *Nat. Genet.* **38,** 962–964
13. Chandler, S. P., Guschin, D., Landsberger, N., and Wolffe, A. P. (1999) *Biochemistry* **38,** 7008–7018
14. Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001) *J. Biol. Chem.* **276,** 34115–34121
15. Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., and Kouzarides, T. (2003) *J. Biol. Chem.* **278,** 4035–4040
16. Kimura, H., and Shiota, K. (2003) *J. Biol. Chem.* **278,** 4806–4812
17. Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., and Oikawa, T. (2003) *Oncogene* **22,** 8688–8698
18. Buschdorf, J. P., and Stratling, W. H. (2004) *J. Mol. Med.* **82,** 135–143
19. Nan, X., Meehan, R. R., and Bird, A. (1993) *Nucleic Acids Res.* **21,** 4886–4892
20. Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998) *Nature* **393,** 386–389

21. Free, A., Wakefield, R. I., Smith, B. O., Dryden, D. T., Barlow, P. N., and Bird, A. P. (2001) *J. Biol. Chem.* **276,** 3353–3360
22. Brunner, E., Weitzel, J., Heitmann, B., Maurer, T., Stratling, W. H., and Kalbitzer, H. R. (2000) *J. Biomol. NMR* **17,** 175–176
23. Wakefield, R. I., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A. P., and Barlow, P. N. (1999) *J. Mol. Biol.* **291,** 1055–1065
24. Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999) *Nat. Genet.* **23,** 185–188
25. Bienvenu, T., Carrie, A., de Roux, N., Vinet, M. C., Jonveaux, P., Couvert, P., Villard, L., Arzimanoglou, A., Beldjord, C., Fontes, M., Tardieu, M., and Chelly, J. (2000) *Hum. Mol. Genet.* **9,** 1377–1384
26. Huppke, P., and Gartner, J. (2005) *J. Child Neurol.* **20,** 732–736
27. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) *J. Mol. Graph. Model.* **19,** 26–59
28. Uversky, V. N. (2002) *Eur. J. Biochem.* **269,** 2–12
29. Tompa, P. (2005) *FEBS Lett.* **579,** 3346–3354
30. Dyson, H. J., and Wright, P. E. (2005) *Nat. Rev. Mol. Cell. Biol.* **6,** 197–208
31. Hansen, J. C., Lu, X., Ross, E. D., and Woody, R. W. (2006) *J. Biol. Chem.* **281,** 1853–1856
32. Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., and Iakoucheva, L. M. (2006) *PLoS Comput. Biol.* **2,** e100
33. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) *Biochemistry* **45,** 6873–6888
34. McBryant, S. J., Krause, C., and Hansen, J. C. (2006) *Biochemistry* **45,** 15941–15948
35. Reeves, R. (2001) *Gene* (*Amst.*) **277,** 63–81
36. Klose, R. J., and Bird, A. P. (2004) *J. Biol. Chem.* **279,** 46490–46496
37. Yusufzai, T. M., and Wolffe, A. P. (2000) *Nucleic Acids Res.* **28,** 4172–4179
38. Carruthers, L. M., Schirf, V. R., Demeler, B., and Hansen, J. C. (2000) *Methods Enzymol.* **321,** 66–80
39. Demeler, B., Behlke, J., and Ristau, O. (2000) *Methods Enzymol.* **321,** 38–66
40. Demeler, B., and van Holde, K. E. (2004) *Anal. Biochem.* **335,** 279–288
41. Sreerama, N., and Woody, R. W. (2000) *Anal. Biochem.* **287,** 252–260
42. Combet, C., Blanchet, C., Geourjon, C., and Deléage, G. (2000) *Trends Biochem. Sci.* **25,** 147–150
43. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) *Bioinformatics (Oxf.)* **21,** 3435–3438
44. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999) *Protein Sci.* **8,** 370–380
45. Heitmann, B., Maurer, T., Weitzel, J. M., Stratling, W. H., Kalbitzer, H. R., and Brunner, E. (2003) *Eur. J. Biochem.* **270,** 3263–3270
46. Kriaucionis, S., and Bird, A. (2004) *Nucleic Acids Res.* **32,** 1818–1823
47. Huth, J. R., Bewley, C. A., Nissen, M. S., Evans, J. N., Reeves, R., Gronenborn, A. M., and Clore, G. M. (1997) *Nat. Struct. Biol.* **4,** 657–665
48. Carro, S., Bergo, A., Mengoni, M., Bachi, A., Badaracco, G., Kilstrup-Nielsen, C., and Landsberger, N. (2004) *J. Biol. Chem.* **279,** 25623–25631
49. Nan, X., Campoy, F. J., and Bird, A. (1997) *Cell* **88,** 471–481
50. Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998) *Nat. Genet.* **19,** 187–191
51. Kaludov, N. K., and Wolffe, A. P. (2000) *Nucleic Acids Res.* **28,** 1921–1928
52. Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J. J., Sudol, M., and Tanokura, M. (2004) *J. Biol. Chem.* **279,** 31833–31841
53. Laccone, F., Huppke, P., Hanefeld, F., and Meins, M. (2001) *Hum. Mutat.* **17,** 183–190