Putney, S. D., Royal, N. J., de Vegvar, H. N., Herlihy, W. C., Biemann, K., & Schimmel, P. (1981a) *Science 213*, 1497-1501.

Putney, S. D., Sauer, R. T., & Schimmel, P. (1981b) *J. Biol. Chem. 256*, 198-204.

Regan, L. (1986) Ph.D. Thesis, Massachusetts Institute of Technology.

Regan, L., Dignam, J. D., & Schimmel, P. (1986) *J. Biol. Chem. 261*, 5241-5244.

Regan, L., Bowie, J., & Schimmel, P. (1987) *Science 235*, 1651-1653.

Riordan, J. F., & Giese, R. W. (1977) *Methods Enzymol. 47*, 31-40.

Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989) *Science 246*, 1135-1142.

Schreier, A. A., & Schimmel, P. (1972) *Biochemistry 11*, 1582-1589.

Starzyk, R. M., Webster, T. A., & Schimmel, P. (1987) *Science 237*, 1614-1618.

Starzyk, R. M., Burbaum, J. J., & Schimmel, P. (1989) *Biochemistry 28*, 8479-8484.

Waye, M. W., Winter, G., Wilkinson, A. J., & Fersht, A. R. (1983) *EMBO J. 2*, 1827-1829.

Zelwer, C., Risler, L., & Brunie, S. (1982) *J. Mol. Biol. 155*, 63-81.

Cadmium-113 NMR Studies of the DNA Binding Domain of the Mammalian Glucocorticoid Receptor†

Tao Pan, Leonard P. Freedman,‡ and Joseph E. Coleman*

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510, and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Received March 30, 1990; Revised Manuscript Received June 26, 1990

ABSTRACT: The DNA binding domain of the mammalian glucocorticoid hormone receptor (GR) contains nine highly conserved cysteine residues, a conservation shared by the superfamily of steroid and thyroid hormone receptors. A fragment [150 amino acids (AA) in length] consisting of GR residues 407-556, containing within it the entire DNA binding domain (residues 440-525), has been overexpressed and purified from *Escherichia coli* previously. This fragment has been shown to contain 2.3 ± 0.2 mol of Zn(II) per mole of protein [Freedman, L. P., Luisi, B. F., Korszun, Z. R., Basavappa, R., Sigler, P. B., & Yamamoto, K. R. (1988) *Nature 334*, 543]. Zn(II) [or Cd(II) substitution] has been shown to be essential for specific DNA binding. ^{113}Cd NMR of a cloned construct containing the minimal DNA binding domain of 86 AA residues [denoted GR(440-525)] with ^{113}Cd(II) substituted for Zn(II) identifies 2 Cd(II) binding sites by the presence of 2 ^{113}Cd NMR signals each of which integrates to 1 ^{113}Cd nucleus. The chemical shifts of these two sites, 704 and 710 ppm, suggest that each ^{113}Cd(II) is coordinated to four isolated $-S^-$ ligands. Shared $-S^-$ ligands connecting the two ^{113}Cd(II) ions do not appear to be present, since their T_1s differ by 10-fold, 0.2 and 2.0 s, respectively. Addition of a third ^{113}Cd(II) or Zn(II) to $^{113}Cd_2$GR(440-525) results in occupancy of a third site, which introduces exchange modulation of the two original ^{113}Cd NMR signals causing them to disappear. Addition of EDTA to the protein restores the original two signals. ^{1}H-^{113}Cd heteronuclear multiple quantum spectroscopy of GR(440-525) shows that the major protons coupled to ^{113}Cd are a group of β-protons assignable to Cys residues. A small variable signal corresponding to a ϵ-CH_3 of Met coupled to ^{113}Cd suggests that the thioether of a Met may be a ligand to the third ^{113}Cd(II) ion. Binding of the third ^{113}Cd(II) causes significant increase of the Cd-S charge transfer absorption bands. We propose that GR(440-525) can form three Zn(II) binding sites involving a combination of the nine Cys and one Met residues as ligands, all of which are highly conserved among the superfamily of steroid and thyroid hormone receptors. Both circular dichroism and ^{1}H NMR show the folding of GR(440-525) to be dependent on the presence of Zn(II) or Cd(II). Removal of the metal ions causes GR(440-525) to completely unfold from its native structure. Both Zn(II) and Cd(II)GR(440-525) have very similar ^{1}H NMR spectra, suggesting almost identical structures for the two metal derivatives of the DNA binding domain of GR.

The mammalian glucocorticoid hormone receptor (GR)[1,2] belongs to the superfamily of steroid and thyroid hormone receptors which have striking amino acid sequence homologies within their DNA binding domains [for a review, see Evans

† This work was supported by NIH Grants DK09070 and GM21919 and by NSF Grant DMB 8716787 (to Keith Yamamoto). The 500-MHz NMR was supported by NIH Grant RR03475, NSF Grant DMB-8610557, and ACS Grant RD259. This work is in partial fulfillment of the Ph.D. requirements for T.P.

* Address correspondence to this author at the Department of Molecular Biophysics and Biochemistry, Yale University.

‡ Present address: Cell Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY 10021. L.P.F. was the recipient of a Bank of America-Giannini Foundation fellowship.

[1] Abbreviations: GR, glucocorticoid hormone receptor; GR(407-556), cloned polypeptide fragment of the mammalian glucocorticoid receptor containing amino acid residues 407-556 which span the DNA binding domain, plus 10 residues on the N-terminus and 14 residues on the C-terminus from the cloning vector; GR(440-525), cloned fragment of GR which contains residues 440-525 plus six residues (MKPARP) on the N-terminus and two residues (RL) on the C-terminus contributed by codons from the cloning vector; TSP, sodium (trimethylsilyl)tetradeuteriopropionate.

0006-2960/90/0429-9218$02.50/0

(1988) and Beato (1989)]. The glucocorticoid receptor enhances or represses transcription of a spectrum of steroid-responsive genes by binding to specific DNA sequences termed glucocorticoid response elements, or GREs, upstream of the promoters for these genes. The GR is organized into three functional domains; a N-terminal modulating domain, a central DNA binding domain, and a C-terminal hormone binding and transactivating domain. There may be some overlap of these functions between domains, although the property of specific DNA sequence recognition appears confined to the central domain (Beato, 1989). Within the central DNA binding domain, there are nine highly conserved cysteine residues with the arrangement -$CysX_2CysX_{13}CysX_2Cys$-$X_{15}CysX_5CysX_9CysX_2CysX_4Cys$-. This polypeptide sequence has been proposed to form two independent "zinc finger" motifs similar to those found in the *Xenopus* transcription factor IIIA (Miller et al., 1985). A protein fragment of 150 residues [denoted GR(407–556)] encompassing the entire DNA binding domain (residues 440–525) has been overexpressed and purified from *Escherichia coli* (Freedman et al., 1988). Zinc analysis of the isolated domain showed it to contain 2.3 ± 0.2 mol of Zn(II) per mole of protein when the purification was carried out in the presence of 75 μM $ZnCl_2$. Zn(II) or Cd(II) binding to the protein was shown to be essential for specific DNA binding (Freedman et al., 1988).

Extended X-ray absorption edge spectroscopy of the Zn(II) complex of GR(407–556) revealed a pattern compatible with a tetrahedral coordination geometry consisting of four sulfur ligands for both Zn(II) sites. Freedman et al. (1988) concluded from these data that there are two metal binding sites utilizing eight of the nine highly conserved cysteine residues. Which eight of the nine conserved Cys residues actually constitute the ligands has been a matter of some controversy. It has been reported that site-directed mutagenesis of these conserved Cys residues, one at a time, indicates that mutation of the most C-terminal Cys (Cys^{500} in the GR) to Ser^{500} does not significantly affect the transcriptional activation of the receptor (Severne et al., 1988). On the other hand, studies of missense mutants isolated in vivo came to the opposite conclusion, namely, that Cys^{500} is indispensable for transcriptional activation (Schena et al., 1989).

No direct evidence has yet been obtained for the formation of either of the two alternate "zinc fingers" outlined above. In addition to the nine cysteines, other highly conserved residues, His^{451} and Met^{505}, are located in the vicinity of the putative cysteine ligands and are also potential ligands to Zn(II). Met^{505} is required for specific DNA binding, since mutation of this residue to Thr^{505} or Ile^{505} (Schena et al., 1989) or Gly^{505} (Hollenberg & Evans, 1988) abolishes the DNA binding affinity. The recent discovery that the GAL4 transcription factor from *Saccharomyces cerevisiae* forms a binuclear $Zn(II)_2Cys_6$ cluster suggests that alternate arrangements of ligands in the so-called Cys_2Cys_2 "zinc-fingers" are possible (Pan & Coleman, 1990a,b). We report here the characterization of an 86 amino acid DNA binding subdomain of the GR containing both putative "zinc-fingers", a characterization carried out with ^1H NMR, ^{113}Cd NMR, and ^1H–^{113}Cd heteronuclear multiple quantum spectroscopy. Metal-dependent conformational changes in the DNA binding domain have been detected by circular dichroism.

Materials and Methods

The GR(440–525) construct or "DART" was produced from a plasmid constructed by one of us (L.P.F.) in the laboratory of Keith R. Yamamoto at the University of California, San Francisco, which contains the DNA sequence of the entire DNA binding domain of GR (residues 440–525) plus codons for six residues (MKPARP) at the 5′ end and codons for two residues (RL) at the 3′ end contributed from the cloning vector. Transcription is from a T7 promoter with the plasmid in the host BL21(DE3) which carries a chromosomal copy of the T7 gene 1 under the control of the lac promoter (Studier & Moffatt, 1986).

Purification of the GR(440–525) Construct. Fifteen grams (wet weight) of IPTG-induced cells was suspended in 40 mL of lysis buffer [50 mM Tris-HCl (pH 8.0)/200 mM KCl/1 mM EDTA/1 mM dithiothreitol/0.5 mM PMSF] and sonicated twice for 3 min on ice. Polymin P was then added over a 10-min period to 0.8%, followed by centrifugation at 25000*g* for 20 min at 4 °C. After dialysis of the supernatant against standard column buffer [(STD) 10 mM Tris-HCl (pH 8.0)/1 mM EDTA/1 mM β-mercaptoethanol/10% (v/v) glycerol plus 50 mM NaCl and 50 μM $ZnCl_2$], the dialysate was loaded onto a preequilibrated Trisacryl SP column and washed sequentially with STD buffer plus 50 mM NaCl, STD buffer plus 150 mM NaCl, and STD buffer plus 300 mM NaCl. The 300 mM NaCl fractions were combined and dialyzed against STD buffer plus 50 mM NaCl and 50 μM $ZnCl_2$. The dialysate was then loaded onto a Bluegel–agarose (Bio-Rad) column, washed with STD buffer plus 50 mM NaCl, and eluted with STD buffer plus a 50–1000 mM NaCl gradient. The GR(440–525) polypeptide elutes in the fractions containing 300–800 mM NaCl and is >95% pure.

Preparation of Apo- and Cd(II)GR(440–525). Purified GR(440–525) contains ~2 mol of Zn(II) per mole of protein. The apoprotein can be prepared by dialysis against 50 mM acetate (pH 5.0)/150 mM NaCl/10 mM EDTA/10% (v/v) glycerol/1 mM β-mercaptoethanol, followed by dialysis against metal-free buffer at pH 5.0. The pH could then be returned to neutral upon addition of 2 M metal-free Tris-HCl, pH 8.0. With this procedure, an apo-GR(440–525) containing less than 0.02 mol of Zn(II) per mole of protein could be prepared. ^{113}Cd(II)GR(440–525) was obtained upon addition of 3 mol of ^{113}Cd(II) per mole of Zn(II) bound to the protein, followed by incubation for at least 12 h at room temperature. The excess ^{113}Cd(II) and the free Zn(II) displaced by the Cd(II) can then be removed by dialysis. Gel retardation assays employing the Zn(II) and Cd(II) derivatives of the fragment of 150 amino acid residues, GR(407–556), show both metal derivatives of the longer fragment to form specific complexes with its GRE DNA sequence (Freedman et al., 1988). We have carried out similar gel retardation assays with the shorter 86 amino acid residue construct, GR(440–525), using a 17 bp DNA fragment containing a consensus GRE sequence. Like the longer fragment, both the native Zn(II) and Cd(II)-exchanged GR(440–525) show formation of specific protein–DNA complexes (data not shown).

^{113}Cd NMR was performed on a Bruker AM 500 spectrometer (110.9 MHz for ^{113}Cd) with a 10-mm tunable broad-band probe at 25 °C. The acquisition conditions and buffers used are listed in the figure legends. Samples were 1.8 mL of ~0.35 mM GR(440–525). Spectral width was 15.2 kHz (136 ppm), and a 60° pulse was used throughout.

[2] The amino acid sequences within the DNA binding domain of the glucocorticoid receptors from the rat and human proteins are identical; therefore, the findings in this paper apply to both proteins. The particular cDNA used to construct GR(440–525) came from a rat cDNA library, and the numbering of the residues used in this paper applies to the rat protein rather than the human protein.

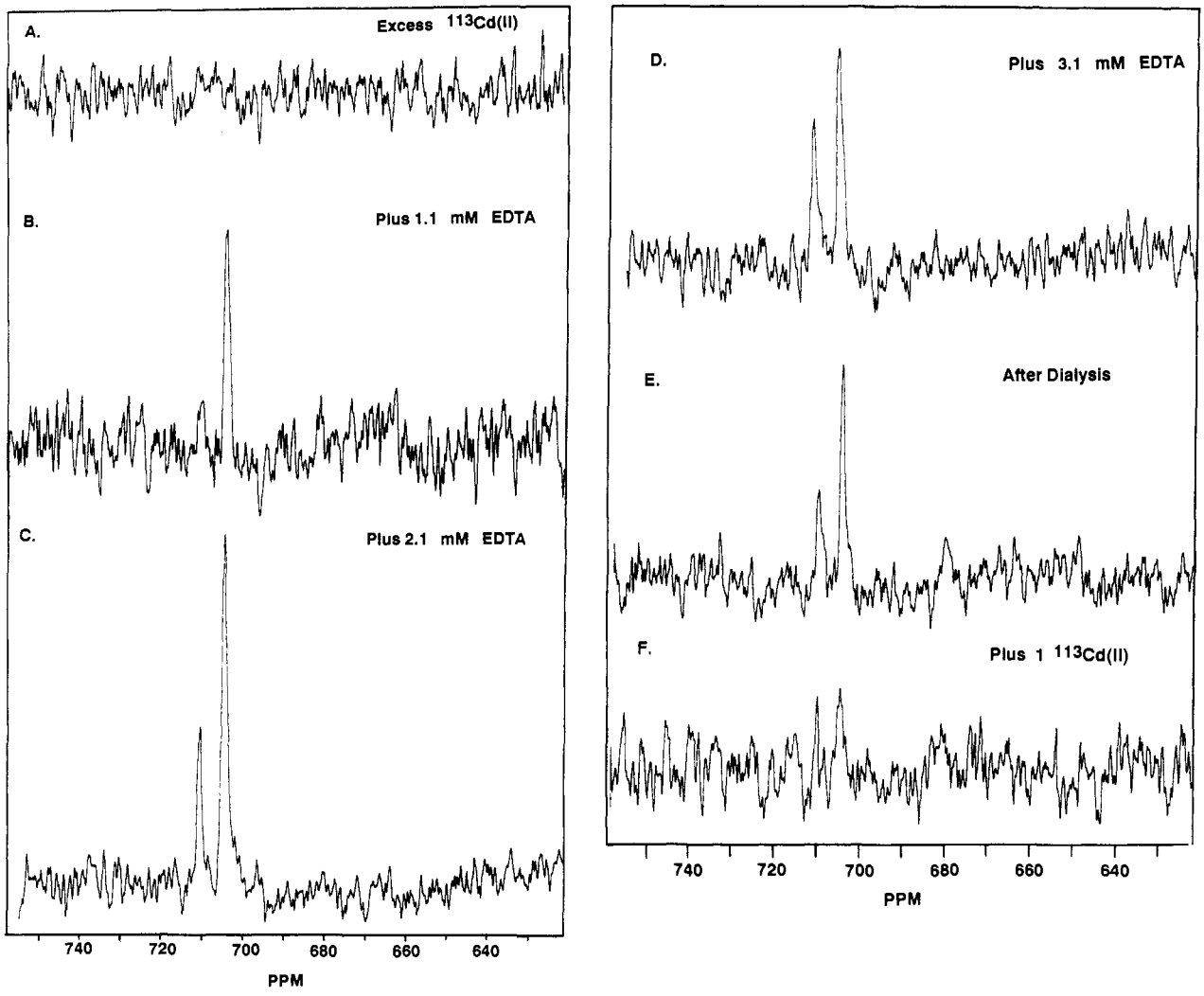

FIGURE 1: ^{113}Cd NMR spectra of GR(440–525) in 50 mM phosphate, pH 7.8. The number of transients is given in parentheses, and the relaxation delay was 1.5 s. A 50-Hz line broadening was applied for spectral enhancement. (A) In the presence of excess ^{113}Cd(II) (11 000). (B) Plus 1.1 mM EDTA (22 000). (C) Plus 2.1 mM EDTA (25 500). (D) Plus 3.1 mM EDTA (11 600). (E) Following dialysis of sample D against 50 mM phosphate, pH 8.0 (9000). (F) Plus 1 equiv of ^{113}Cd(II) per protein molecule (15 500).

Zinc and cadmium analyses were performed by atomic absorption spectroscopy using an IL157 spectrometer (Instrumentation Laboratories, Lexington, MA).

Results

^{113}Cd NMR of GR(440–525). ^{113}Cd NMR can be used as a probe to determine the number of sulfur donors involved in Cd(II) ligation [Pan & Coleman, 1989, 1990a; Giedroc et al., 1989; for a review, see Armitage and Otvos (1982)]. The ^{113}Cd NMR spectrum of GR(440–525) in the presence of excess ^{113}Cd(II) shows no ^{113}Cd NMR signals (Figure 1A). Upon the addition of EDTA (1.1 mM), one signal at 704 ppm, relative to that for 0.1 M $^{113}Cd(ClO_4)_2$, is observed (Figure 1B). Upon the addition of more EDTA (2.1 and 3.1 mM), a second signal at 710 ppm can be detected (Figure 1B–D). No other peaks are found when ^{113}Cd spectra are collected from 490 to 760 ppm. The EDTA-bound ^{113}Cd(II) can be removed by dialysis, and the original two ^{113}Cd resonances at 704 and 710 ppm remain (Figure 1E). Following dialysis, this sample contained ~1.8 mol of ^{113}Cd(II) per mole of protein.

All samples of the ^{113}Cd(II)-substituted GR(440–525) which show the fully developed ^{113}Cd NMR signals contained ~2 mol of cadmium per mole of protein and no zinc. If dialysis of the ^{113}Cd(II)-exchanged protein is terminated too soon, the protein passed through a gel filtration column only, or dialyzed against metal-free water rather than 40 mM Tris-HCl, the metal content of the protein remains ~3 mol per mole of protein, and no ^{113}Cd signals can be observed. When 1 additional equiv of ^{113}Cd(II) was added to the sample containing 1.8 mol of ^{113}Cd(II) per mole of protein, both signals disappeared again (Figure 1F). The above phenomena are similar to findings with ^{113}Cd NMR studies of bacterial alkaline phosphatase (Gettins & Coleman, 1983), in which the presence of ^{113}Cd(II) bound at a second site (B) diminishes the signal arising from the ^{113}Cd(II) bound at a first site (A) through conformational modulation of site A by metal ion exchange at site B (see Discussion).

The signal of $^{113}Cd_2$GR(440–525) at 710 ppm has approximately half the intensity of the ^{113}Cd signal at 704 ppm when the relaxation delay is set at 1.5 s (Figure 1C). The difference in amplitude between the 710 and 704 ppm signals is entirely accounted for by differences in T_1 of the two ^{113}Cd nuclei (Figure 2). When the spectrum of the same sample is collected by setting relaxation delays to 0.75, 1.5, and 3.6 s, respectively, the increase in intensity of the 710 ppm peak relative to that of the 704 ppm peak is evident (Figure 2). Estimates of the T_1 values based on spectra collected at three different time delays suggest T_1 for the 704 ppm peak to be 0.2–0.4 s while T_1 for the 710 ppm peak is 2–3 s.

When 1 equiv of Zn(II) is added to GR(440–525) containing approximately two ^{113}Cd(II) ions, the same exchange broadening of the two original ^{113}Cd signals is observed (Figure

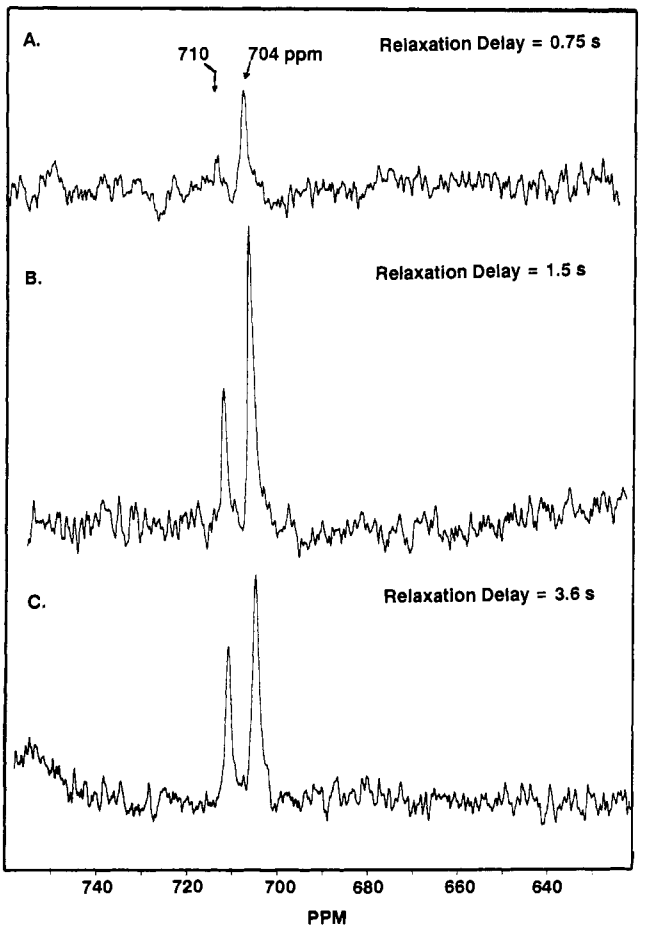

FIGURE 2: ^{113}Cd NMR spectra of the same sample as in Figure 1C except that the relaxation delay is 0.75 s (A) (12000), 1.5 s (B) (25000), or 3.6 s (C) (12800).

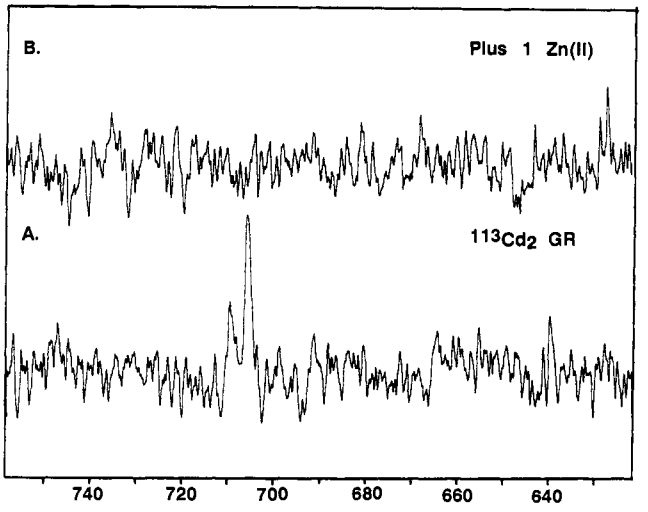

FIGURE 3: ^{113}Cd NMR spectra of (A) $^{113}Cd(II)_2GR(440–525)$ in 40 mM Tris, pH 8.0 (9600). (B) Plus 1 equiv of Zn(II) per protein molecule (10200).

3). Thus, the same third metal binding site can be occupied by either Cd(II) or Zn(II) (see Discussion). The binding of the third Cd(II) or Zn(II) ion does not cause aggregation of GR(440–525) as shown by the protein samples containing 2 mol of $^{113}Cd(II)$ plus 1 mol of Zn(II). Despite exchange modulation of the ^{113}Cd signals, this sample gives rise to a highly resolved ^1H NMR spectrum identical with that of the two $^{113}Cd(II)$ species (see Figures 4 and 5). High-resolution ^1H NMR spectra of GR(440–525) containing two $^{113}Cd(II)$ and one Zn(II) can be obtained because of the different frequencies for exchange modulation for ^{113}Cd NMR (10^3–10^4 s^{-1}) and for ^1H NMR (10^1–10^3 s^{-1}).

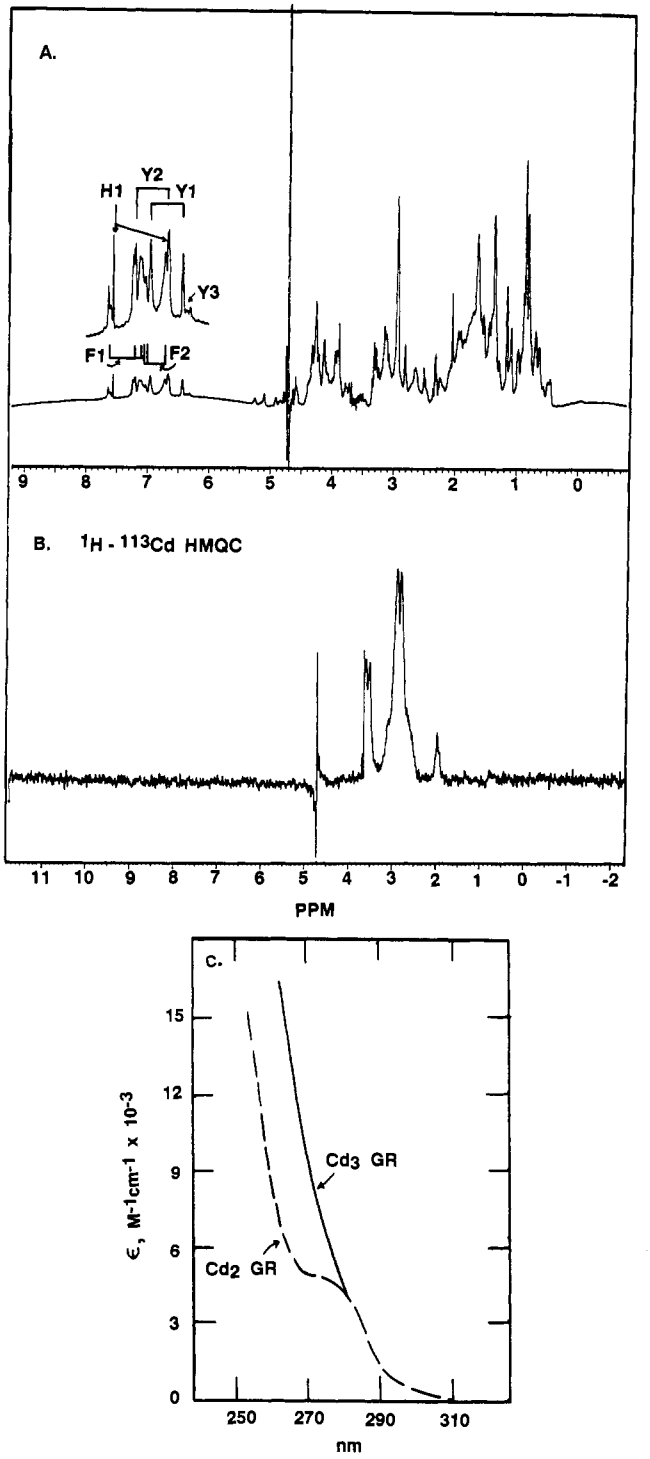

FIGURE 4: (A) ^1H NMR spectrum of GR(440–525) (3 mM) containing 2.1 $^{113}Cd(II)$ and 0.9 Zn(II) in D_2O. The assignment of the aromatic resonances is based on a COSY spectrum of the same sample. (B) ^1H–^{113}Cd HMQC spectrum; pulse sequence of $90_x(^1H)$–τ–$90_\phi(^{113}Cd)$–$t_{1/2}$–$180(^1H)$–$t_{1/2}$–$90_x(^{113}Cd)$–τ–FID using the same sample as in part A; τ = 8 ms. The number of transients was 4096, and the relaxation delay was 1.5 s. (C) UV absorption spectra (260–310 nm) of GR(440–525) containing 2.1 $^{113}Cd(II)$ + 0.9 Zn(II) (Cd_2GR)(---) and 2.9 $^{113}Cd(II)$ + 0.1 Zn(II) (Cd_3GR)(—).

^1H–^{113}Cd Heteronuclear Multiple Quantum Spectroscopy (HMQC) of GR(440–525). A ^1H NMR spectrum of GR(440–525) containing 2.1 $^{113}Cd(II)$ and 0.9 Zn(II) in D_2O, formed by adding 1 Zn(II) per mole of protein, is shown in Figure 4A. The protons of GR(440–525) coupled to the bound $^{113}Cd(II)$ can be identified by ^1H–^{113}Cd HMQC (Frey et al.,

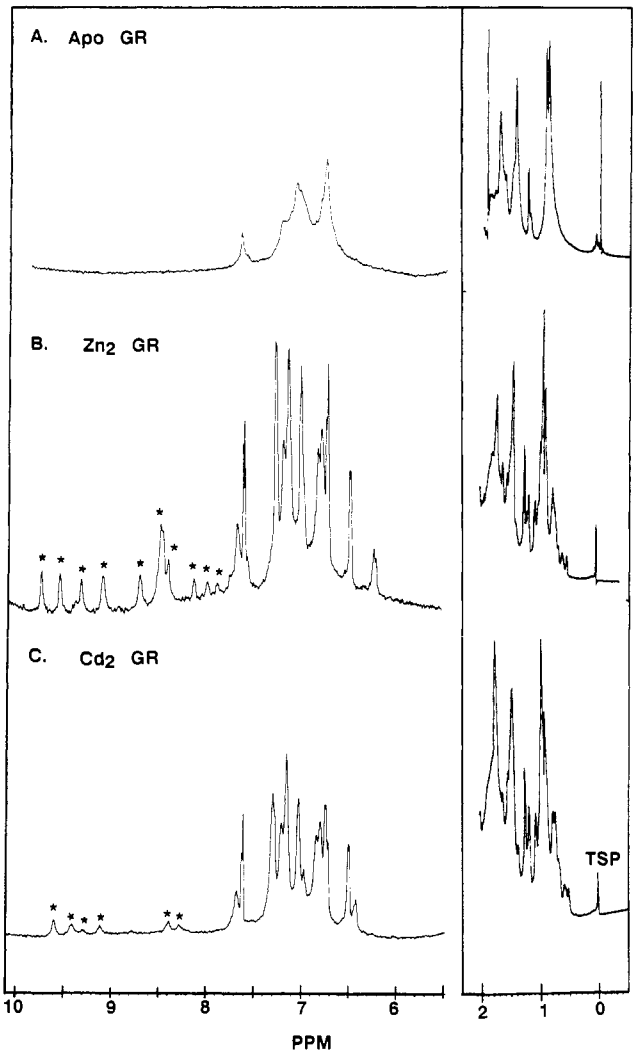

FIGURE 5: ^1H NMR spectra of apo-GR(440–525) (A), Zn(II)GR-(440–525) (B), and ^{113}Cd(II)GR(440–525) (C) in 50 mM phosphate (pH 7.8)/D_2O. Samples were exchanged into D_2O using a G-25 spun column. The spectra were taken after incubation in D_2O for ~20 h. The peaks marked with asterisks are slowly exchanging amide protons.

1985), and such a spectrum for the above sample is shown in Figure 4B. The chemical shifts of the coupled protons are from 3.7 to 2.5 ppm and are consistent with their assignment to the β-protons of cysteine residues. The presence of the small signal at ~2.0 ppm, likely to be the ε-CH_3 of a Met, indicates ligation of some of the ^{113}Cd(II) to a Met residue, probably at the third site. One of the Zn(II) ions of native Zn_2GR(440–525) is more difficult to remove or exchange. Hence, when exchanging ^{113}Cd(II) for Zn(II), it is frequently found that the protein contains ~1 Zn and ~2 ^{113}Cd if the exchange is terminated too soon. Thus, in the 2 ^{113}Cd(II), 1 Zn(II) protein, there is likely to be some Zn(II) → ^{113}Cd(II) mixing between all three sites. We believe the most upfield resonance in the HMQC spectra results from some of the third site being occupied by ^{113}Cd(II), which accounts for the variability and relatively small magnitude of this signal.

Met^{505} is also a residue conserved among the steroid and thyroid superfamily of receptor proteins. Unfortunately, GR(440–525) containing 2.9 ^{113}Cd(II) plus 0.1 Zn(II) gives rise to ^1H NMR spectra with very broad signals (data not shown), suggesting that the exchange modulation by the third Cd(II) is slower than that induced by Zn(II) and thus causes exchange modulation at the ^1H frequency as well. Far-UV absorption spectra of GR(440–525) containing three ^{113}Cd(II)

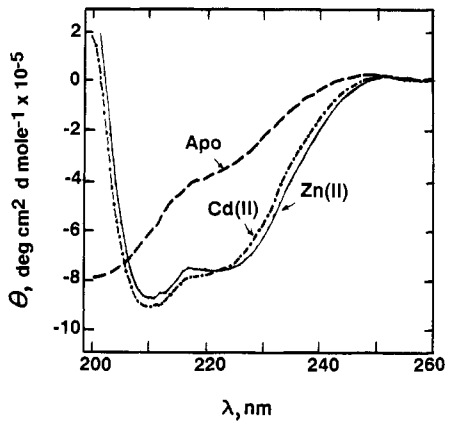

FIGURE 6: Circular dichroism of Zn(II)-, Cd(II)-, and apo-GR-(440–525) (5 μM) in 10 mM Tris/150 mM NaCl/1 mM β-mercaptoethanol/10% glycerol (v/v)/5 μM EDTA, pH 8.0 at 25 °C.

ions show a very substantial increase in the Cd–S charge transfer bands when the third ^{113}Cd(II) is bound (Figure 4C). The third site may thus involve more than one sulfur ligand.

^1H NMR Spectra of Apo-, Zn(II)-, and Cd(II)GR(440–525). The aromatic and upfield aliphatic regions of the ^1H NMR spectra of apo-, Zn(II)-, and Cd(II)GR(440–525) in 50 mM phosphate at pH 7.8 are shown in Figure 5. Removal of Zn(II) causes substantial changes in chemical shifts and line broadening of the ^1H NMR signals (Figure 5A). These changes are comparable to those obtained in the ^1H NMR spectrum of GAL4 protein or T4 gene 32 protein upon Zn(II) removal (Pan et al., 1989; Pan & Coleman, 1990b); i.e., the line broadening is contributed by conformational fluctuation of the apoprotein. Zn(II)- and Cd(II)GR(440–525) have similar conformations as judged by the fact that there are only minor differences in their ^1H NMR spectra (compare Figure 5B and Figure 5C). One finding that does suggest a subtle difference in structure of the two metal ion derivatives is that the exchange rates of a number of amide protons in the Zn(II) protein are substantially slower than in the Cd(II) derivative (Figure 5B,C). The length of the incubation in D_2O of the Zn(II) and Cd(II) samples in Figure 5 was the same.

Circular Dichroism of Apo-, Zn(II)-, and Cd(II)GR(440–525). The Zn_2GR(440–525) has moderate molar ellipticity in the wavelength region of the peptide bond chromophores, -8.3×10^5 deg cm^2 dmol^{-1} at 208 nm and -7.6×10^5 deg cm^2 dmol^{-1} at 222 nm (Figure 6). A reasonable graphical fit is obtained by a combination of 25% α-helix, 10% β-sheet, and 65% random coil (Greenfield & Fasman, 1969). Circular dichroism of Cd_2GR(440–525) is virtually identical with that of the Zn(II) species (Figure 6). Removal of Zn(II) converts the resultant apoprotein to a largely random-coil conformation (Figure 6). Reconstitution of the apoprotein at pH 8.0 in the presence of β-mercaptoethanol and excess Zn(II) is difficult. Only a small fraction of the apoprotein has refolded after 24 h as revealed by circular dichroic spectra taken 24 h after readdition of a molar excess of Zn(II) (data not shown).

Discussion

It is widely postulated that the DNA binding domain of GR as well as the DNA binding domains of other members of the steroid receptor superfamily form two "Zn fingers" utilizing eight of the nine highly conserved Cys residues as ligands (Evans, 1988; Beato, 1989). Our ^{113}Cd NMR studies show that two metal binding sites with four Cys coordinated to each ^{113}Cd(II) do exist in the DNA binding domain of GR as suggested by ^{113}Cd NMR chemical shifts for $^{113}Cd_2$GR(440–

525) (Figure 1). ^{113}Cd chemical shifts alone cannot be used to differentiate isolated mononuclear metal–sulfur sites from multinuclear metal–sulfur sites in which certain Cys ligands are shared between two metal ions as found in metallothionein [for a review, see Kagi and Kojima (1987)] or more recently for the GAL4 transcription factor (Pan & Coleman, 1990a,b). The fact that the T_1 values for the two ^{113}Cd nuclei bound to GR(440–525) are ~10-fold different (Figure 2) suggests but does not prove that the two sites are somewhat isolated from each other.

While it is not impossible in ^{113}Cd clusters to visualize an arrangement where relaxation of one ^{113}Cd could be different than another, a two-^{113}Cd cluster with shared cysteine ligands requires the ^{113}Cd nuclei to be ~3.5 Å apart in a relatively symmetrical arrangement. Dipolar influences at the two nuclei are therefore likely to be similar. In fact, for the one example of a dinuclear ^{113}Cd(II) cluster found in a protein, GAL4, the two ^{113}Cd resonances show identical T_1s, both 0.1–0.2 s (Pan and Coleman, unpublished data). In rabbit metallothionein, containing both a $Cd(II)_3$ and a $Cd(II)_4$ cluster, one of the ^{113}Cd resonances from the three-metal cluster is unusually broad, suggesting different relaxation properties from the others. On the other hand, metallothionein of crab contains two three-metal clusters, and the T_1 values of the ^{113}Cd signals measured at 110 MHz range from 0.1 to 0.8 s (Dalgarno et al., 1985), all significantly shorter than the slowly relaxing ^{113}Cd in GR(440–525).

The detailed information that has been collected on ^{113}Cd relaxation in Cd(II)–protein complexes containing isolated ^{113}Cd(II) sites suggests that dipolar relaxation rather than chemical shift anisotropy (CSA) often remains the chief relaxation mechanism, even at frequencies as high as 110 MHz (Giedroc et al., 1989). In view of the prominent dipolar contribution to the relaxation of some of the ^{113}Cd–sulfur sites in proteins, it has seemed probable that there are amide backbone–sulfur hydrogen bonds to the liganded $-S^-$ groups as have been observed for a number of Zn(II)– or Fe(II)–sulfur sites in crystal structures (Adman et al., 1975; Berg, 1988) or structures determined by 2D NMR methods (Summers et al., 1990). Such hydrogen bonds would place relaxing protons much closer to the ^{113}Cd(II) ion than the β-protons of the cysteine ligands. The latter are too distant to account completely for the large dipolar contributions observed (Giedroc et al., 1989). The absence of such interactions at the site represented by the 710 ppm signal in GR(400–525) could explain its unusually long T_1 (Figure 2).

A second possible explanation of the 10-fold difference in T_1s for the two ^{113}Cd nuclei in GR(440–525) is that CSA is a major contributor to relaxation. The geometrical features of the two sites would have to be sufficiently different that there is a significantly smaller CSA for the site giving the signal at 710 ppm vs that giving the signal at 704 ppm. While such a circumstance has not been found for the ^{113}Cd-substituted proteins examined thus far, if one were to place 8 β-protons at 3.5 Å from the ^{113}Cd ion and assign CSAs of 100 and 500 ppm, respectively, the T_1s (110 MHz) would be ~2.0 and 0.2 s, respectively, for a molecule with a $\tau_r = 3 \times 10^{-9}$ s, appropriate for GR(440–525) consisting of 94 residues. While it is difficult to speculate on the origin of such a "geometrical" difference in the two sites, a protein structure could significantly distort one complex relative to the other, perhaps a more likely occurrence for separate mononuclear sites. On the basis of the small amount of data available in the literature on the relaxation of ^{113}Cd in clusters, the T_1s are all rather short, <1 s (110 MHz), and CSA appears to make a significant contribution to relaxation (Dalgarno et al., 1985).

The other unexpected property of the ^{113}Cd derivative of GR(440–525) is the presence of a third metal binding site containing sulfur ligands (Figure 4C). The flexibility of metal exchange properties of this site are such that the presence of the metal ion at the third site exchange-modulates the chemical shifts of the first two sites with a frequency corresponding to intermediate exchange, 10^3–10^4 s^{-1}, depending on the $\Delta\delta$ involved. The third metal ion is obviously less tightly bound, since it can be easily stripped off by EDTA (Figure 1).

Disappearance of ^{113}Cd resonances of ^{113}Cd(II)–proteins induced by chemical exchange modulation at frequencies corresponding to the intermediate chemical exchange for the ^{113}Cd NMR resonances is not an unusual observation. Several examples for ^{113}Cd(II)-substituted proteins have been well documented (Schoot Uiterkamp et al., 1980; Coleman et al., 1979). Under some circumstances, the binding of additional ligands to the protein can damp these conformational modulations such that ^{113}Cd signals reappear, e.g., binding of I^- to ^{113}Cd carbonic anhydrase (Schoot Uiterkamp et al., 1980) or inorganic phosphate to alkaline phosphatase (Gettins & Coleman, 1983). Unfortunately, we have not been able to acquire ^{113}Cd NMR spectra of a GR(440–525)–DNA complex with three ^{113}Cd(II) bound, since the complex formed between GR(440–525) and our synthetic 17 bp oligonucleotide containing a 15 bp near-consensus GRE is insoluble.

Although the Cd(II) or Zn(II) bound to the third site in GR(440–525) can be removed by EDTA or exhaustive dialysis against Tris or phosphate, the reported cellular concentrations of Zn(II) (0.1–1 mM) (National Research Council, 1979) are sufficient to occupy this site in vivo under many conditions. A stoichiometry of 3 mol of metal ions bound per mole of GR(440–525) [2.1 ^{113}Cd(II) and 0.9 Zn(II)] is observed when dialysis is carried out against water alone (Figure 4). Zn(II) analysis of GR(407–556) by Freedman et al. (1988) revealed 2.3 ± 0.2 mol of Zn(II) bound per mole when the sample was dialyzed against 50 mM Tris (pH 7.6)/200 mM NaCl/75 μM $ZnCl_2$. Nonintegral stoichiometries are often caused by differences in sample treatment such as buffer composition or length of the dialysis. GAL4 transcription factor consistently contains 1.1–1.5 mol of Zn(II) per mole when dialyzed against metal-free buffer (Pan & Coleman, 1989); nevertheless, two metal binding sites are present (Pan & Coleman, 1990a,b). The functional significance of the third Zn(II) in GR(440–525) is not easily determined. The GAL4 transcription factor DNA binding domain containing one Zn(II) per molecule will bind DNA specifically in a standard gel retention assay, although the binding affinity may be less compared to that of the two Zn(II) species (Pan & Coleman, 1989; unpublished results). Binding of the second Zn(II) to the GAL4 DNA binding domain does not induce significant additional folding as assayed by circular dichroism (Pan and Coleman, unpublished results). Thus, the second Zn(II) may be required only for fine-tuning of specific DNA binding, an effect that could become critical in vivo. Likewise, the binding of the third Zn(II) to GR may involve only a subtle change in the DNA binding affinity compared to the two-Zn(II) species.

^{1}H–^{113}Cd HMQC indicates possible coordination of the third ^{113}Cd(II) to a Met residue (Figure 4B). Unfortunately, both ^{1}H and ^{113}Cd NMR spectra of $^{113}Cd(II)_3$GR(440–525) give very broad signals. The UV absorption spectrum reveals such a strong increase in Cd–S charge transfer bands upon binding of the third Cd(II) (Figure 4C) that coordination to more than one sulfur atom may be involved. There are three

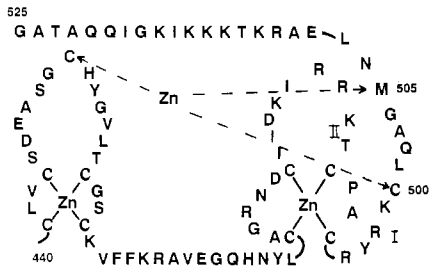

FIGURE 7: Proposed Zn(II) binding model for GR(440–525). The Zn(II)–sulfur bonds at the third site (---) are only meant to indicate possible choices.

potential sulfur ligands (Cys^{450}, Cys^{500}, and Met^{505}) in addition to the eight Cys involved in the basic $^{113}Cd(II)$ complexes giving rise to the NMR signals. Mutagenesis of the DNA binding domain of GR(440–525) indicates that all nine highly conserved Cys residues as well as the conserved Met^{505} are required for specific DNA binding (Severne et al., 1988; Schena et al., 1989; Hollenberg & Evans, 1988). A possible "linear two-zinc-finger" model of the GR DNA binding domain indicating possible interactions with a third Zn(II) is shown in Figure 7. The exchange modulation of the ^{113}Cd NMR signals (Figures 1 and 3) suggests that the metal sites do not exist as isolated subdomains of the protein but that the whole DNA binding domain must form a compact structure which allows close interactions of the first two metal binding subdomains with the third. The compactness of the overall metal-containing domain is supported by the observation that none of the Lys–X or Arg–X bonds in the linker between the two postulated zinc fingers is susceptible to proteolysis by trypsin in the Zn(II) or Cd(II) derivatives (Freedman et al., 1988).

We have relatively little information on the precise nature of the third metal ion binding site except the suggestion based on the HMQC spectrum (Figure 4) that a methionine side chain is involved. It cannot be ruled out that some type of interaction with one of the preexisting $Zn(II)Cys_4$ or $Cd(II)Cys_4$ centers could occur to form a binuclear cluster, an event which could lead to severe conformational modulation of the ^{113}Cd signals from the original two sites. The GR(440–525) contains an extra Cys residue (Cys^{450}) which is present in less than half the members of the superfamily (Evans, 1988; Beato, 1989). Although site-directed mutagenesis reveals that Cys^{450} is not required for transcriptional activation (Severne et al., 1988), Cys^{450} could be a ligand to the third metal ion in GR(440–525). Its absence in other members of the superfamily requires caution in generalizing the observations concerning the third metal binding site to other members of the superfamily. Studies of related hormone receptors such as the estrogen receptor which lack this Cys residue are in progress and will clarify the possible role of this tenth Cys.

It has been shown that the specificity of GRE recognition resides in the vicinity of the "knuckle" part of the first "finger" (Danielsen et al., 1989; Umesono & Evans, 1989). The second "finger" is also required, although it can be swapped with its counterpart from the estrogen receptor without loss of the original specificity (Green et al., 1988). This finding has led to the suggestion that the second finger of the protein is required for dimerization and/or nonspecific contact with the DNA (Beato, 1989). The three positively charged residues within loop I are highly conserved among the steroid superfamily (Figure 7). Loop II has been shown to be responsible for protein–protein interactions (Schena et al., 1989). A parallel example to GR, although in inverse polarity, can be found in the $Zn(II)_2Cys_6$ binuclear cluster of the GAL4 transcription factor where the highly positively charged first loop of six residues within the N-terminal binuclear cluster is responsible for interaction of the protein with the sugar–phosphate backbone, while a more C-terminal loop or turn outside of the cluster appears to confer specific DNA sequence recognition on the GAL4 DNA binding domain (Pan & Coleman, 1990b).

The qualitative kinetics of metal ion binding to the DNA binding domains of GR and GAL4 are, however, considerably different. Cd(II) binds highly cooperatively to the two sites in GAL4 protein, while Cd(II) binds differentially to the three sites in GR(440–525). Binding of Cd(II) to the first two sites is much tighter than to the third site. It is not clear whether these differences are due to the nature of the metal binding sites or to significant differences in the polypeptide folds of the two zinc-containing transcription factors. That there may be significant differences in the folding of the DNA binding domains of GAL4 compared to that of GR is suggested by the dramatic differences in the folding of the two "apo" forms of these proteins as detected by circular dichroism spectra (Figure 6). While removal of the two Zn(II) ions from the DNA binding domain of GAL4, expressed as fragments of either 149 or 63 amino acid residues, causes modest changes in secondary structure, the overall fold of the proteins is maintained (Pan & Coleman, 1989, 1990a). In marked contrast, GR(440–525) of 94 amino acid residues can apparently maintain none of its characteristic secondary and presumably tertiary structure in the absence of the bound metal ions (Figure 6). Thus, the metal ions in GR appear to be involved in maintaining to a large degree the whole protein fold within the DNA binding domain of the hormone receptor.

ACKNOWLEDGMENTS

We thank Keith Yamamoto for permission to use the construct DART in the present study.

Registry No. L-Met, 63-68-3; Zn, 7440-66-6.

REFERENCES

Adman, E., Watenpaugh, K. D., & Jensen, L. H. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4854.

Armitage, I. M., & Otvos, J. D. (1982) *Biol. Magn. Reson. 4*, 79.

Beato, M. (1989) *Cell 56*, 335.

Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99.

Coleman, J. E., Armitage, I. M., Chlebowski, J. F., Otvos, J. D., & Schoot Uiterkamp, A. J. (1979) in *Biological Applications of Magnetic Resonance* (Schulman, R. G., Ed.) p 345, Academic Press, New York.

Dalgarno, D. C., Olafson, R. W., & Armitage, I. M. (1985) *Inorg. Chem. 24*, 3439.

Danielsen, M., Hinck, L., & Ringold, G. M. (1989) *Cell 57*, 1131.

Evans, R. M. (1988) *Science 240*, 889.

Freedman, L. P., Luisi, B. F., Korszun, Z. R., Basavappa, R., Sigler, P. B., & Yamamoto, K. R. (1988) *Nature 334*, 543.

Frey, M. H., Wagner, G., Vasak, M., Sorensen, O. W., Neuhaus, D., Worgotter, E., Kagi, J. H., Ernst, R. R., & Wuthrich, K. (1985) *J. Am. Chem. Soc. 107*, 6847.

Gettins, P., & Coleman, J. E. (1983) *J. Biol. Chem. 258*, 396.

Giedroc, D. P., Johnson, B. A., Armitage, I. M., & Coleman, J. E. (1989) *Biochemistry 28*, 2410.

Green, S., Kumar, V., Theulaz, I., Wahli, W., & Chambon, P. (1988) *EMBO J. 7*, 3037.

Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108.

Hollenberg, S. M., & Evans, R. M. (1988) *Cell 55*, 899.
Kägi, J. H. R., & Kojima, Y., Eds. (1987) *Metallothionein II*, Birkhaeuser, Basel.
Miller, J., Mclachlan, A. P., & Klug, A. (1985) *EMBO J. 4*, 1609.
National Research Council (1979) *Zinc, Report of the Subcommittee on Zinc*, University Park Press, Baltimore, MD.
Pan, T., & Coleman, J. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3145.
Pan, T., & Coleman, J. E. (1990a) *Biochemistry 29*, 3023.
Pan, T., & Coleman, J. E. (1990b) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2077.
Pan, T., Giedroc, D. P., & Coleman, J. E. (1989) *Biochemistry 28*, 8832.
Schena, M., Freedman, L. P., & Yamamoto, K. R. (1989) *Genes Dev. 3*, 1590.
Schoot Uiterkamp, A. J., Armitage, I. M., & Coleman, J. E. (1980) *J. Biol. Chem. 255*, 3911.
Severne, Y., Wieland, S., Schaffner, W., & Rusconi, S. (1988) *EMBO J. 7*, 2503.
Studier, F. W., & Moffat, N. (1986) *J. Mol. Biol. 189*, 113.
Summers, M. F., South, T. L., Kim, B., & Hare, D. R. (1990) *Biochemistry 29*, 329.
Umesono, K., & Evans, R. M. (1989) *Cell 57*, 1139.

Sequence Preferences of DNA Interstrand Cross-Linking Agents: dG-to-dG Cross-Linking at 5′-CG by Structurally Simplified Analogues of Mitomycin C†

Margaret F. Weidner, Snorri Th. Sigurdsson, and Paul B. Hopkins*
Department of Chemistry, University of Washington, Seattle, Washington 98195
Received April 4, 1990; Revised Manuscript Received July 3, 1990

ABSTRACT: The nucleotide sequence preferences of the DNA interstrand cross-linking agents dehydroretronecine diacetate (DHRA), 2,3-bis(acetoxymethyl)-1-methylpyrrole (BAMP), dehydromonocrotaline, and dehydroretrorsine were studied by using synthetic DNA duplex fragments and polyacrylamide gel electrophoresis (PAGE). These agents have structural features in common with the reductively activated aziridinomitosene of mitomycin C (MC). Like MC, they preferentially cross-linked DNA duplexes containing the duplex sequence 5′-CG. For DHRA and BAMP interstrand cross-linked DNA duplexes, PAGE analysis of iron(II)–EDTA fragmentation reactions revealed the interstrand cross-links to be deoxyguanosine to deoxyguanosine (dG-to-dG), again analogous to DNA cross-links caused by MC. Unlike MC, DHRA could be shown to dG-to-dG cross-link a 5′-GC sequence. Furthermore, the impact of flanking sequence on the efficiency of interstrand cross-linking at 5′-CG was reduced for BAMP, with 5′-TCGA and 5′-GCGC being equally efficiently cross-linked. Possible origins of the 5′-CG sequence recognition common to all of the agents are discussed. A model is presented in which the transition state for the conversion of monoadducts to cross-links more closely resembles ground-state DNA at 5′-CG sequences.

The bifunctional alkylating agents are a class of drugs useful in the treatment of human cancer and include the chloroethylnitrosoureas, nitrogen mustards, and mitomycins. These drugs are widely believed to exert their cytotoxic action by reaction with DNA; numerous studies suggest that interstrand cross-linking of duplex DNA is important in this regard (Kohn, 1980). The chemical structures of the conjugates of these drugs and DNA have been extensively studied. Much less work has centered on the identification of sequences at which cross-links are preferentially formed (Mattes, et al., 1988). At the present time, the reactions with DNA of the psoralens are probably the best understood interstrand cross-linking reactions (Hearst, 1989). This paper concerns the mitomycins, cross-linking agents (Iyer & Szybalski, 1963, 1964; Matsumoto & Lark, 1963) about which much less is known. Three laboratories have reported that for reductively activated mitomycins, 5′-CG is the preferred cross-linked sequence (Chawla et al., 1987; Teng et al., 1989; Weidner et al., 1989; Millard et al., 1990b); the combined studies strongly suggest that the aziridinomitosene of mitomycin C (MC, Figure 1)[1] ultimately bridges the exocyclic amino groups (N2) of two deoxyguanosine residues by sequential formation and reaction with DNA of electrophilic sites at carbons 1 and 10 of the mitosene (Tomasz et al., 1987). Very recently, Norman et al. (1990) have reported NMR solution structure studies of an MC cross-linked hexanucleotide duplex and concluded that MC cross-linked DNA does possess this covalent connectivity, with the MC residue deeply nested in the minor groove of a right-handed double helix.

What is the mechanistic origin of 5′-CG recognition by reductively activated mitomycins? Particularly intriguing is the preference of 5′-CG over 5′-GC: Both of these sequences possess a pair of dG N2 groups centrally located in the minor groove of DNA and with similar interatomic spacing (3.6 and 4.0 Å, respectively) in canonical B DNA (Arnott et al., 1976). Furthermore, the sequence that flanks the core 5′-CG sequence

† This work was supported in part by grants from the National Science Foundation (DIR-8220099) and National Institutes of Health (GM 35466 and GM 32681). P.B.H. is an NIH Research Career Development Award recipient (AG 00417) and Sloan Fellow (1988–1992).

[1] Abbreviations: BAMP, 2,3-bis(acetoxymethyl)-1-methylpyrrole; b, broad; d, doublet; DHRA, dehydroretronecine diacetate; DTE, dithioerythritol; DTT, dithiothreitol; m, multiplet; MC, mitomycin C; PAGE, polyacrylamide gel electrophoresis; q, quartet; s, singlet; t, triplet.

0006-2960/90/0429-9225$02.50/0