Polyproline Type II Conformation in the C-Terminal Domain of the Nuclear Pore Complex Protein gp210†

Yair Pilpel,‡ Oren Bogin,‡ Vlad Brumfeld,§ and Ziv Reich*,‡

Departments of Biological Chemistry and Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

Received August 12, 2002; Revised Manuscript Received February 12, 2003

ABSTRACT: gp210 is a major constituent of the nuclear pore complex (NPC) with possible structural and regulatory roles. It interacts with components of the NPC via its C-terminal domain (CTD), which follows a transmembrane domain and a massive (∼200 kDa) N-terminal region that resides in the lumen of the perinuclear space. Here, we report the solution structure of the human gp210 CTD as determined by various spectroscopic techniques. In water, the CTD adopts an extended, largely unordered conformation, which contains a significant amount of left-handed polyproline type II (PII) helical structure. The conformation of the CTD is altered by high pH, charged detergents, and the hydrogen bond-promoting reagent trifluoroethanol (TFE), which decrease the PII fraction of the fragment. TFE also induces a conformational change in a region containing an SPXX motif whose serine becomes specifically phosphorylated during mitosis. We propose that PII elements in the CTD may play a role in its interaction with the NPC and may serve as recognition sites for regulatory proteins bearing WW or other, unknown PII-binding motifs.

Nuclear pore complexes (NPCs)[1] are large protein assemblies that span the nuclear envelope (NE) and mediate the bidirectional traffic of small molecules and macromolecules between the cytoplasm and the nucleus (*1*−*3*). The vertebrate NPC has a molecular mass of ∼125 MDa (*4*) and is made up of ∼30 distinct proteins (called nucleoporins) present in multiple copies (*5*). It consists of a symmetrical framework, measuring approximately 120 × 80 nm, which is made of two coaxial rings sandwiching a wheel-like array of eight spoke-shaped domains (*6*, *7*). The spoke-ring assembly embraces the central pore channel, which may be encompassed by a massive, hourglass-shaped structure termed transporter (*6*, *8*, *9*). In addition to the central framework, NPCs also contain extensive peripheral structures. These include eight globular particles and thin filaments that emanate toward the cytoplasm, and a fish trap-like structure, termed the nuclear basket, which extends into the nucleus (*8*, *10*−*12*).

Interestingly, the (vertebrate) NPC communicates with the surrounding pore membrane via only two integral proteins, POM121 and gp210 (*13*, *14*). gp210 was first identified in rat (*14*, *15*) and, following that, in *Drosophila* (*16*), *Xenopus* (*17*), and mouse (*18*). The rat gp210 is a relatively abundant, type I integral membrane protein estimated at 16−24 copies per NPC (*19*). Most of the protein's mass (∼200 kDa) is found in the NE lumen (*20*). This domain, comprising the N-terminus of the protein, is glycosylated in 9 out of 13 putative glycosylation sites (*14*). It also contains five EF-hand calcium-binding motifs (at residues 160−171, 424−435, 580−591, 1413−1424, and 1610−1621). The protein's C-terminal domain (CTD), which follows the protein transmembrane (TM) domain, is composed of ∼60 amino acids (*14*, *20*) and is phosphorylated on a single serine residue at position 1880 (in the rat) during mitosis (*21*). Newly synthesized gp210 arrives at its locale by first being transported into the ER via a leader sequence and then by a process of lateral diffusion through the membrane. The single TM of gp210 is sufficient to localize it to the pore membrane, as is the CTD, albeit with weaker sorting determinants (*22*). Spatially, gp210 has been localized to the radial arms, which interconnect the NPC spokes within the NE lumen, about 30 nm radially distant from the central axis of the pore, consistent with the luminal position of the protein's N-terminal domain (*23*).

Although gp210 was the first nucleoporin identified, as far back as 1982 (*15*), little is known about its structure and function. On the basis of the presence of two putative TM domains in the rat sequence, a role for gp210 has been suggested in the fusion of the nuclear membranes during pore formation (*14*). However, only one of the two TM domains, the C-terminal hydrophobic domain, actually crosses the pore membrane (*20*), and the second, putative TM, is not significantly conserved among species (*24*; our own unpublished results). In addition, gp210 is recruited to re-forming pores relatively late, at the end of telophase (*25*). gp210 may also function in the transmission of signals from the NE lumen to the NPC (*26*), perhaps in response to changes in perinuclear calcium levels (*27*). Given its locale and topo-

† This work was supported by the Israel Science Foundation (Grant 81/00).

* Corresponding author. E-mail: ziv.reich@weizmann.ac.il. Fax: ++972 8 934 6010.

‡ Department of Biological Chemistry, Weizmann Institute of Science.

§ Department of Plant Sciences,Weizmann Institute of Science.

[1] Abbreviations: CD, circular dichroism; CTD, C-terminal domain; FTIR, Fourier transform infrared spectroscopy; NE, nuclear envelope; NPC, nuclear pore complex; PII, polyproline II-like left-handed helix; TFE, trifluoroethanol; TM, transmembrane.

10.1021/bi0266176 CCC: $25.00
Published on Web 03/07/2003

logical characteristics, the most likely role of gp210 is, however, in the anchoring of the central framework of the pore complex to the nuclear membranes. Acting as such, it may also function in pore assembly and disassembly in the course of NE breakdown and re-formation during mitosis. A role for gp210 in pore formation has recently been demonstrated in reconstituted *Xenopus* nuclei (*28*). The same study also revealed that the gp210 CTD, together with some unknown binding partners, is directly involved in pore dilation.

Here, we have analyzed the structure of the C-terminal domain of human gp210 using various spectroscopic techniques. In aqueous solution, the CTD is largely unordered but still contains appreciable amounts of β-turns and left-handed polyproline type II-like (PII) structures. The conformation of the CTD is relatively insensitive to changes in the ionic strength and to low pH but can be altered by high pH, charged detergents, and TFE. TFE also induces a conformational change in a region containing an SPXX motif whose serine was shown to be phophorylated during mitosis. The implications of these observations to the interaction of the CTD with components of the NPC and with regulatory proteins containing PII-binding motifs are discussed.

MATERIALS AND METHODS

TFE (99.5%, NMR grade) was purchased from Aldrich. Sodium dodecyl sulfate (SDS), dodecyl β-D-maltoside, Chaps, and routine chemicals were from Sigma. Water was ~18 MΩ/cm. Handling of TFE-containing samples was by glassware as TFE has been reported to extract CD-active component(s) from Eppendorf tubes (*29*).

Cloning and Sequencing of the Human gp210 CTD. The human gp210 CTD was cloned by RT-PCR with HeLa total RNA extract as template, using oligonucleotide primers based on a putative human cDNA sequence (GI: 4240301). For sequence determination, amplified DNA was ligated into pBluescriptKS (+) (Stratagene) as *Eco*RI−*Hin*dIII fragments. Sequencing was carried out in the Sequence Analysis Unit of the Weizmann Institute of Science.

Peptide Synthesis. Peptide chains were assembled by conventional solid-phase synthesis using an ABIMED AMS-422 automated solid-phase multiple peptide synthesizer (Langenfeld, Germany); Fmoc strategy was employed throughout the assembly. Crude peptides were purified with a prepacked LichroCart RP-18 column (Merck) employing a binary gradient formed from 0.1% TFA in water and 0.1% TFA in 75% acetonitrile in water. For purity evaluation, analytical reversed-phase HPLC was performed using a prepacked Lichrospher-100 RP-18 column (Merck). HPLC separations and analyses were performed using a Spectra Physics SP8800 liquid chromatography system equipped with an Applied Biosystems 757 variable wavelength absorbance detector. The column effluents were monitored by UV absorbance at 220 nm. The appropriate fractions were collected, lyophilized, and analyzed for amino acid composition (Waters 2690 separations Module, Milford, MA) and molecular weights (VG-platform-II electrospray single quadropole mass spectrometer; Micro Mass). The former analysis was also used to verify peptide concentration as determined by UV absorption at 280 nm. The purity of peptides was >95%.

CD Spectroscopy. CD spectra were recorded on AVIV 202 spectropolarimeter (Aviv Inc., Lakewood, NJ) equipped with a thermoelectric cell holder and calibrated with ammonium *d*-10-camphorsulfonate. Far-UV CD measurements were carried out in 0.1 cm path length quartz cuvettes, using a step size of 0.25 nm, a bandwidth of 1 nm, and a time constant of 1 s. The peptide was dissolved in 10 mM sodium phosphate buffer (pH 7.5) and was kept at 41.3 μM, with the exception of the TFE experiments where the peptide concentration was 44.6 μM. Far-UV CD data are given as mean residue ellipticity [θ]. The optical path length in the near-UV measurements was 1 cm, and the concentration of the peptide was 374 μM (392 μM when TFE was used). Spectra acquired (five scans) were averaged, baseline corrected, and smoothed using the Savitzky−Golay algorithm with a sliding window of 2.25 nm. The far-UV spectra were analyzed by Selcon 3 using data reference set 2. Thermal denaturation curves were extracted from the CD spectra recorded at different temperatures. Heating and cooling rates were 2 °C /min; the equilibration time was 5 min.

FTIR Spectroscopy. FTIR measurements were carried out with a Protege 460 FTIR system (Nicolet), using a DGTS detector with a KBr window paired with a KBr beam splitter. Samples (30 mg/mL) were placed between two CaF_2 windows separated by a homemade Mylar spacer (5 μm width) and examined in transmission. Spectra (512 scan) were averaged and corrected for background solvent components. Initial positioning of the Voigt curves used in band decomposition was made according to the Fourier-deconvoluted spectra in the amide I and amide II regions.

Fluorescence Spectroscopy. Fluorescence emission spectra were recorded on an SLM-Aminco 8000 fluorescence spectrophotometer. The excitation wavelength was 280 nm; the integration time was 0.3 s. Spectra shown are corrected for solvent contributions.

RESULTS

Sequence Analysis of Human gp210 CTD. We used a published EST sequence of the human homologue of gp210 (*30*) to plan primers to amplify a segment encoding for the last 420 C-terminal residues of the protein by RT-PCR, using HeLa cell total RNA as a template. The segment was then cloned and sequenced to verify the EST assignment, which was found to be correct. To define the protein's CTD, the sequence was aligned with the known sequences of the rat and mouse homologues. The domain was also assigned independently using the TMpred software, which had best predicted the experimentally determined TM segment of rat gp210 (*20*).

The sequences of three known mammalian CTDs are shown in Figure 1. Amino acids are numbered according to their position in the rat CTD and are referred to as such throughout the text. The CTD of all species is basic (pI_{calc} of 9.6 for mouse and human and 8.2 for the rat) and contains a high content of proline (17−20%), alanine (15−18%), serine (15−17%), and histidine (7−8%) and a low content (<20%) of "order"-conferring hydrophobic amino acids (Ile, Leu, Val, Trp, Phe, Cys). All sequences contain multiple putative phosphorylation sites, of which one, Ser^{53} (1880 in the rat sequence), was shown to be phosphorylated during mitosis (*21*). They also contain several SPXX motifs, which

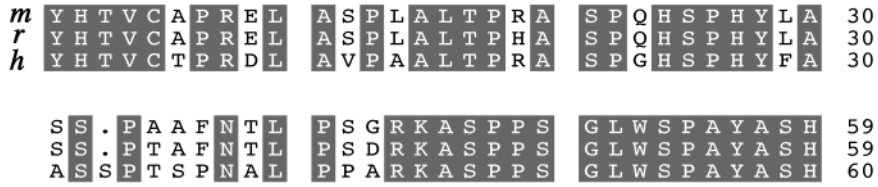

FIGURE 1: Multiple sequence alignment of mammalian gp210 CTD. Shown are the human (*h*), rat (*r*), and mouse (*m*) homologues.

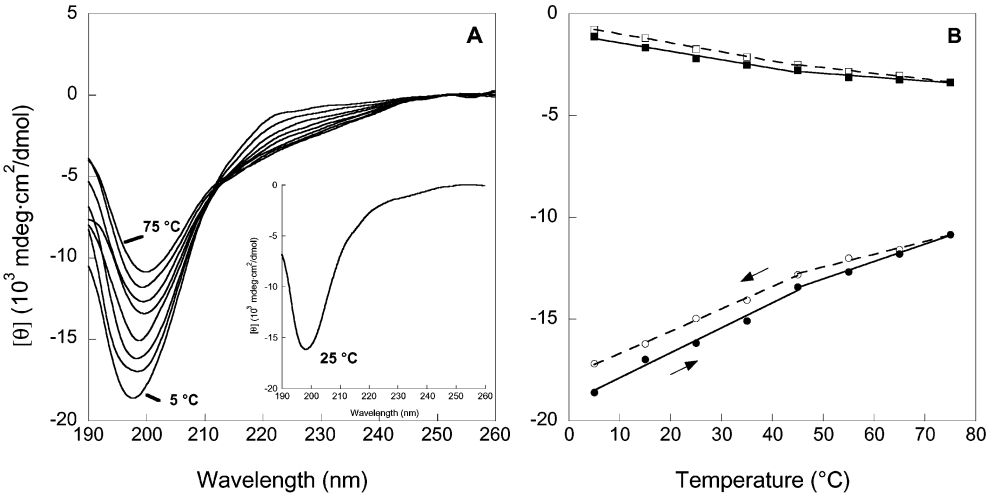

FIGURE 2: (A) Far-UV CD spectra of the CTD in water (10 mM sodium phosphate, pH 7.5) at different temperatures. Spectra taken in the reverse order (cooling) were essentially the same. (Inset) CD spectrum of the CTD at 25 °C. (B) Temperature dependence of the molar ellipticities at the spectra minima (bottom) and at 223 nm (top). Melting and annealing curves are indicated by solid and dashed lines, respectively.

Table 1: Percentage of Secondary Structures in the CTD[a]

	water (pH 7.5, 20 °C)	water (pH 7.5, 5 °C)	water (pH 12, 20 °C)	TFE (94%)	SDS (2 mM)	Chaps (2 mM)
α-helix	5.9	0	4.2	13.4	9.2	6
β-strand	14.7	16.1	25.6	14.9	16.5	16.7
β-turn	16.2	16.7	16.9	18.4	14.5	17.6
PII	13.5 (16.1)[b]	16.1 (25.9)[b]	10.1	9.9	12.1	11.2
unordered	46.4	49	30.4	39.3	43.5	43.7
3_{10} helix	3.8	3.6	9.7	4.5	3.8	4.4

[a] Derived by deconvolution of the CD data using Selcon 3 (data set 2). [b] Estimated by using eqs 1 and 2 in ref *40*. The PII content of the CTD calculated by this method for the conditions described in columns 3−6 was below 5%.

can adopt PII and *β*-turn conformations (*29*) and which, upon serine phosphorylation, may serve as recognition sites for type IV WW protein modules (*31*, *32*, and references cited therein). As can be expected from the sequences, secondary structure predictions for all CTDs indicated an almost completely unordered structure, with little or no defined secondary structural motifs (not shown).

Far-UV CD Measurements. Figure 2A (inset) shows the far-UV circular dichroic (CD) spectrum of the CTD, as obtained in buffered aqueous solution (pH 7.5) at 25 °C. The spectrum is characterized by a strong negative band at 199 nm and by an inflection point centered at 223 nm. These features are frequently observed in peptides containing a significant fraction of PII (see, e.g., ref *29*). To estimate the content of PII and other secondary structures in the CTD, we used Selcon 3 operating with a PII-containing reference data set (data set 2). This analysis yielded a PII fraction of about 14% (Table 1). Conventional methods of secondary structure analysis, however, often have difficulty in estimating contributions made by PII and other "nonclassical" structural components. We therefore used the method proposed by Park et al. (*33*) and later modified by Bienkiewicz et al. (*29*) to get an alternative estimate of the PII content in the CTD. For the CD data acquired at room temperature, this method gave a value of 16%, similar to the 14% yielded by the deconvolution analysis. In other cases, however, the values derived by the two methods differed significantly (Table 1). Besides PII, the CTD contains significant fractions of *β*-strands (15%) and *β*-turns (16%). The rest of the structure is mostly unordered (46%).

The effect of temperature on the CD spectrum of the CTD is shown in Figure 2A. Upon heating, the 199 nm band undergoes a red shift of 2−3 nm and becomes less negative. The signal around 220 nm, on the other hand, becomes more negative with increasing temperatures, which also leads to the disappearance of the inflection point at this region. Decreasing the temperature results in opposite effects. At 5 °C, a clear maximum can be seen at 223 nm. At this temperature, we estimate the PII content of the peptide at 26% (Table 1). The most notable feature of the spectra, however, is a well-defined isoelliptic point at 212 nm. Such isoelliptic points have been observed in the temperature-dependent CD spectra of several PII-containing peptides and are attributed to conformational equilibrium between PII and the unordered state (*34*−*36*).

The melting/annealing behavior of the CTD is shown in Figure 2B, which depicts the temperature dependence of the CD signals around the PII minima (the 199 nm band) and maxima (223 nm). In each case, the melting curve exhibits a biphase linear, rather than sigmoidal, dependence on the temperature, with a slight change in slope occurring at 45−48 °C. A similar biphasic behavior was observed in the

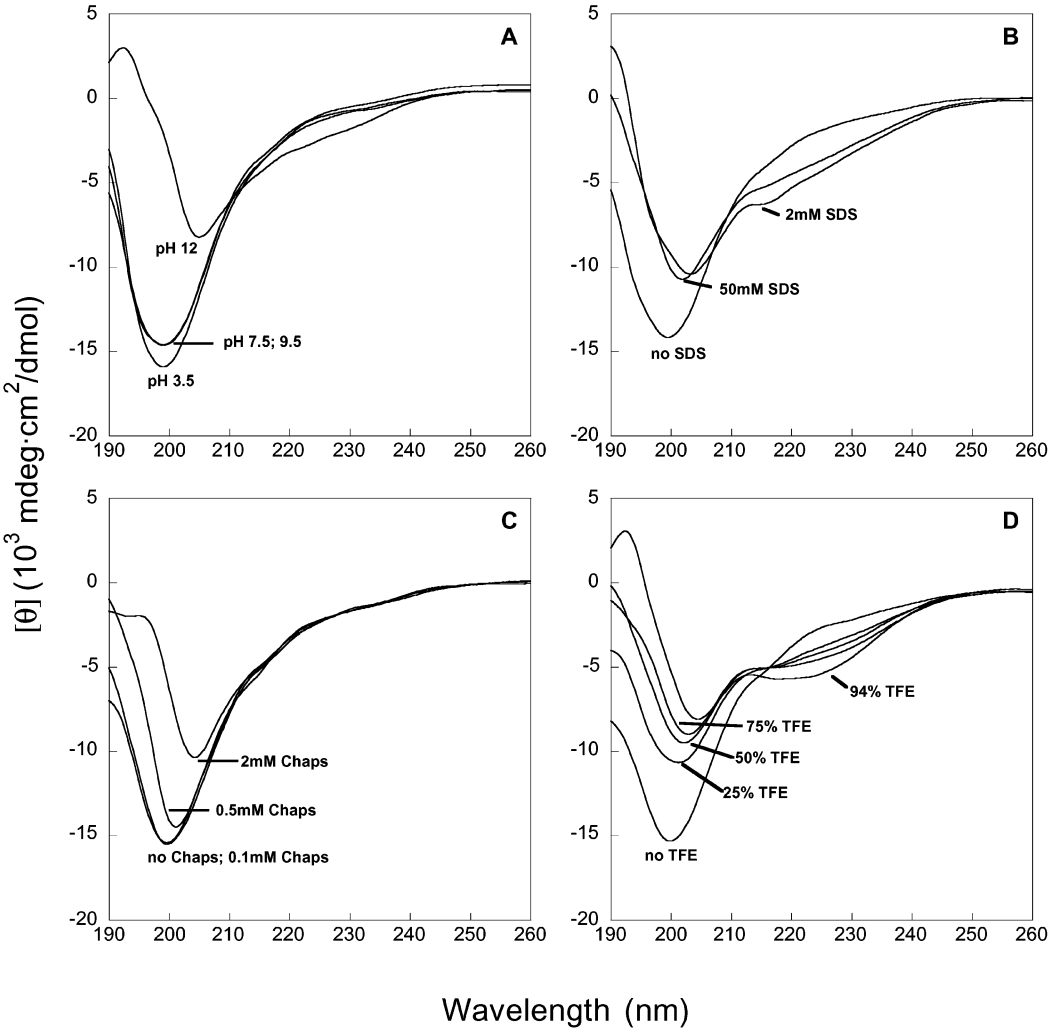

FIGURE 3: Effect of the solution parameters on the CD spectrum of the CTD: (A) pH; (B) SDS; (C) Chaps; (D) TFE. Addition of ammonium sulfate (≤0.4 M) or the nonionic surfactant, dodecyl β-D-maltoside (≤CMC), did not have any detectable effect on the spectrum.

melting of pure polyproline peptides (*37*), as well as of lysine homopolymers in a PII conformation (*38*). The characteristic hysteresis observed during the cooling phase of thermally denatured hydrogen-bonded secondary structures is missing from the curves.

Next, we investigated the effect of the solution parameters on the structure of the CTD. Increasing the ionic strength of the solution by raising concentrations (up to 400 mM) of ammonium sulfate had no effect on the CTD (not shown). This is consistent with the idea that electrostatics plays a relatively minor role in the stabilization of PII structures (*39*). Changing the pH between 3.5 and 9.5 likewise failed to affect the structure (Figure 3A). Only at the extreme pH of 12 could a significant effect be observed. At this pH, the 199 nm band undergoes a strong red shift, to 205 nm, and becomes significantly less negative. This red shift leads to the appearance of a positive peak at 192 nm. Past 212 nm, toward the longer wavelengths, the spectrum becomes more negative. Analysis of the data indicates that these spectral changes reflect primarily an increase of β-strand components at the expense of PII and the unordered conformation (Table 1).

The effect of the anionic detergent SDS on the CTD is shown in Figure 3B. In the presence of 2 mM SDS, the amplitude of the negative 199 nm band decreases while the spectrum beyond 212−213 nm becomes more negative. In addition, two strong inflection points appear, at 212 and 218 nm, while the one present in water (223 nm) is now absent. Analysis of the spectrum indicates an increase in α-helical content and a decrease in PII and unordered structures (Table 1). Raising the SDS concentration up to 50 mM resulted in similar spectra, indicating that no further structural changes had occurred. Similar effects were observed when the zwitterionic surfactant Chaps was used (Figure 3C). In contrast, no significant changes in the CD spectrum could be detected when the nonionic detergent, dodecyl β-D-maltoside, was applied at concentrations up to its CMC value (10 mM; not shown). On the basis of the above arguments, the inability of dodecyl β-D-maltoside to induce conformational changes in the CTD probably reflects a poor binding rather than a need for macroscopic screening of repulsive electrostatic interactions between side-chain charges.

Figure 3D shows the effect of TFE on the CTD. TFE promotes the formation of intramolecular hydrogen bonds and is used to induce partly folded states of proteins. At rising concentrations, the minimum at 199 nm becomes less negative and is red shifted. At longer wavelengths, the spectra become increasingly more negative, to the point of displaying a broad minimum at 223 nm at 94% TFE. These spectral alterations mark a progressive increase in α-helical and β-turn

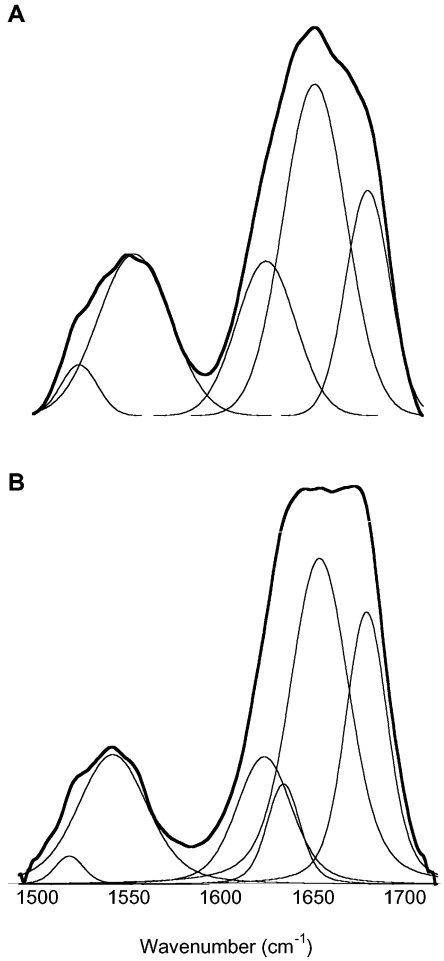

FIGURE 4: FTIR spectra of the CTD recorded in (A) water and (B) 94% TFE. Voigt components are shown as thin lines.

conformations at the expense of PII and unordered structures (Table 1). An isoelliptic point at 216 nm observed in the spectra obtained between 0 and 75% TFE indicates a two-state process in this range of TFE concentrations. This point, however, is not crossed by the spectrum recorded in 94% TFE. Thus, a structural transition not present in the dilute TFE mixtures may be taking place at this concentration of TFE. This transition is likely to begin already at 50−75% TFE, as suggested by the crossover observed in these TFE concentrations at ∼195 nm.

FTIR Measurements. The high solubility of the CTD (>30 mg/mL) allowed FTIR measurements to be made in concentrated solutions, circumventing the need for using thin films. Figure 4 shows the FTIR spectra of the CTD in water and in 94% TFE. In water, the amide I band is almost symmetrically distributed around 1647 cm^{-1} and the amide II band is centered at 1550 cm^{-1} (Figure 4A). TFE qualitatively changes the structure of the amide I band, which now appears to contain two components, and causes a shift of the amide II band to 1543 cm^{-1} (Figure 4B). Decomposition of the amide I band into Voigt components (Figure 4 and Table 2) reveals the presence of three peaks in the water spectrum (located at 1622, 1647, and 1676 cm^{-1}), of which, one, the 1622 cm^{-1} band, is indicative of PII (*40*−*42*). In TFE, the 1647 cm^{-1} band shifts to 1651 cm^{-1} and an additional component appears at 1632 cm^{-1}, apparently at the expense of the 1622 cm^{-1} band. The largely unordered

Table 2: Secondary Structure Analysis of Amide I Bands in the FTIR Spectra of the CTD

solvent	band position (cm^{-1})	relative area (%)	assignment
water			
	1622	25.2	PII
	1647	42	unordered
	1676	32.6	β-turn
TFE (94%)			
	1621	17.3	PII
	1631	8.4	β-strand
	1651	46.5	unordered
	1676	27.8	β-turn

nature of the fragment, in both solvents, is evident by the prominent peaks located at 1647 (water) and 1651 cm^{-1} (TFE). Note that, while giving values very close to those derived by the CD analysis for the content of the unordered conformation, the FTIR data indicate significantly higher fractions of PII and β-turn conformations (Tables 1 and 2, respectively).

Near-UV CD and Fluorescence Spectroscopy Measurements. The structure of the gp210 CTD in water and in 94% TFE has also been studied by near-UV CD and fluorescence spectroscopy. In buffer solution, the near-UV CD spectrum of the CTD is dominated by a broad, positive tyrosine peak centered at 270 nm (Figure 5A). A weak shoulder at 292−293 nm, assigned to tryptophan, is also apparent. In 94% TFE, the tyrosine peak decreased substantially, leading to accentuation of the tryptophan band, which remains relatively unaltered. The 280 nm-excited fluorescence emission spectra obtained for the CTD in the above solutions are shown in Figure 5B. The spectrum obtained in water is characteristic of water-exposed tryptophan residues, with a prominent peak at 353 nm. In 94% TFE, this peak is strongly quenched [as is also the case for *N*-acetyl-L-tryptophan (not shown)] due to an excited-state proton transfer from the TFE solvent (*43*). In addition, a shorter wavelength peak appears at 310 nm, signifying the presence of an (excitonic) uncoupled tyrosine.

DISCUSSION

While a substantial amount of information exists on the structure of intact vertebrate and invertebrate NPC, little is known about the structure and biophysical properties of the individual proteins that make these elaborate structures. Here, we employed spectroscopic techniques to acquire information about the solution structure of the C-terminal domain of one of these proteins, gp210. gp210 is a major mediator between the NPC and the surrounding NE lumen and, as such, may play a role in the anchoring of the NPC to the pore membrane, in pore assembly and disassembly during mitosis, and/or in the transmission of signals between the NE lumen and the pore. All of these functions critically depend on its CTD, which is the only region of the protein that is available for interaction with the central framework of the pore assembly.

Secondary structure analyses predict a mostly disordered conformation for the CTD, with little or no secondary structure. Given the sequence characteristics of this domain, such predictions are indeed quite expected. In fact, it bears many of the features that typify the sequence of intrinsically

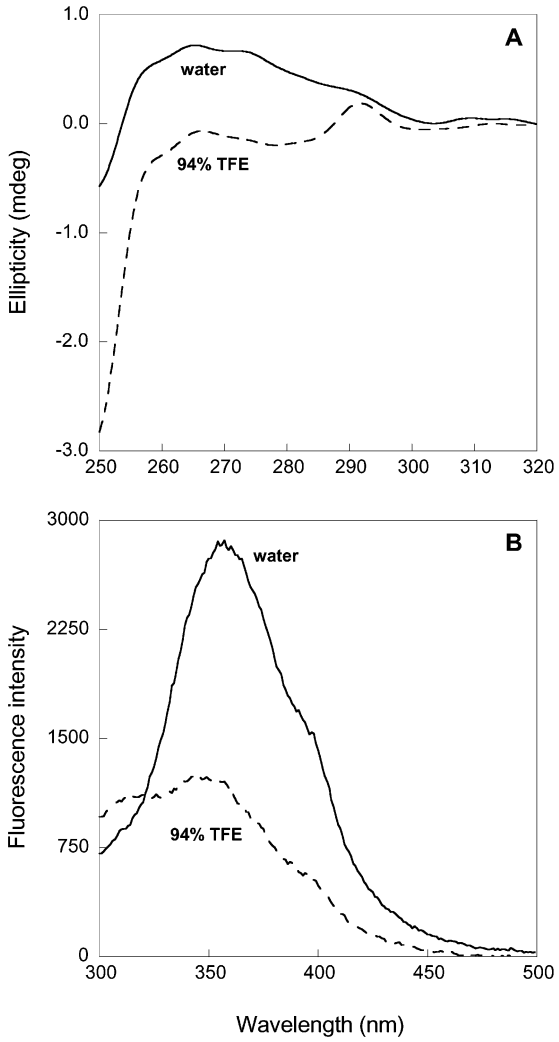

FIGURE 5: Near-UV CD (A) and fluorescence emission (B) spectra of the CTD in water (solid lines) and in 94% TFE (dashed lines). The excitation wavelength for the fluorescence measurements was 280 nm.

unordered ("natively unfolded") protein domains, i.e., low hydrophobicity, a far from neutral p*I* value, and a high content of amino acids such as proline, alanine, and serine, which are known to confer structural disorder (*44* and references cited therein). Natively unfolded proteins are disordered in their free (unbound) state and acquire their folded structure only when in complex with other proteins (*44*−*46*). Some structural determinants, however, may still be preserved to constrain configurational space and to allow for initial association with binding partners. Folding of the protein, induced by binding, may also start from nucleation sites contributed by these determinants.

Our results suggest that, for gp210 CTD, the aforementioned structural determinants may be, at least in part, PII helices. On average, every fifth amino acid in the sequence of all mammalian CTDs is proline. Given that each proline residue restricts the conformation of its preceding amino acid to the β-region of the (ϕ, ψ) space (*47*), this means that about 40% of the residues in these CTDs are predisposed to adopt a PII conformation. In addition, many of the prolines are followed by proline, alanine, glycine, or arginine residues, all of which have a strong propensity to form PII structures (*48*, *49*). Isoleucine and valine, which are particularly poor propagators of PII (*48*), are either absent (Ile) or present at only one copy (Val) in the sequence.

In accordance with the above considerations, analysis of the CD spectrum of the CTD fragment in water indicates a significant amount of PII, even at room temperature. Lower temperatures expectedly increase the PII content of the fragment, as evident by the continuous increase of the amplitude of the 199 nm band and the decrease in the CD signals at the 220−230 nm range. At 5 °C, a maximum in the 220−230 nm range, the CD hallmark of PII, begins to form. For a pure PII conformation, the CD signal in this region of the spectrum should be positive, which is not the case here. However, linear extrapolation of this band indicates positive values below ∼−70 °C, as has been observed for other PII-containing proteins and peptides (*29* and references cited therein).

The CD spectra observed for the CTD at different temperatures reveal a well-defined isoelliptic point at 212 nm, signifying an equilibrium between the PII conformation and the unordered conformation (*34*−*36*). Such a (predominantly) two-state process is further supported by the melting/annealing behavior of the CTD, which appears smooth and linear over the whole temperature range, and is devoid of the characteristic hysteresis. Given that amide groups in the PII helix form hydrogen bonds with the solvent rather than with each other, this lack of cooperativity during melting and annealing is fully expected. The change in the slope between the two linear regimes of the melting curves, which occurs at 45−48 °C, is also consistent with the temperature dependence of the CD spectra of PII-containing peptides (*37*, *38*).

The FTIR data provide additional support for the presence of PII in the CTD. Specifically, a band near 1620 cm^{-1} has been observed in the FTIR spectra of proline homo, poly (*40*), and oligo (*41*) peptides and has been attributed to hydrated carbonyls of proline residues within the PII helix (*42*). In accordance with the CD data, the 1622 cm^{-1} band becomes significantly weaker in 94% TFE, indicating a decrease in the PII content of the fragment. Replacing water by TFE leads to the appearance of a weak band at 1632 cm^{-1}, characteristic of β-strand conformation. It also induces a shift of the 1647 cm^{-1} band, which signifies unordered conformation, to 1651 cm^{-1}, indicating contribution of α-helical structures. The latter structures, however, are present at very low amount as evident by the absence of a distinct peak in the spectrum. As is the case for the CD measurements, the unordered conformation dominates the FTIR spectra, both in water and in TFE.

TFE also has a strong effect on the fluorescence spectrum of the peptide. Specifically, it leads to the appearance of a 310 nm peak in the 280 nm-excited emission fluorescence spectrum of the CTD. The CTD contains a single tryptophan residue in position 52, which precedes a SPXX motif located near the carboxyl end of the peptide. The last residue in this motif is tyrosine. Thus, the aforementioned spectral alterations are consistent with a decoupling of excitonic interaction between the two residues. Such a process may reflect the formation of a β-turn in this site. The data derived from the near-UV CD measurements (showing that TFE exposes a tryptophan band at 292 nm, which is masked by a broad tyrosine peak in the spectrum taken in water) are also in line with such a process. The strong quenching of the

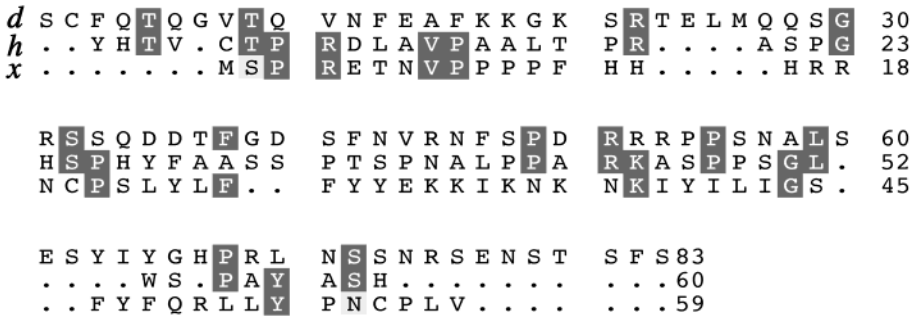

FIGURE 6: Sequences of gp210 CTD from human (*h*), *Xenopus* (*x*), and *Drosophila* (*d*).

tryptophan fluorescence, induced by the TFE, indicates that this residue (Trp^{52}) is largely solvent exposed in this conformation (*43*), as can be expected.

Besides PII, the CTD also contains significant fractions of β-strands and β-turns. The latter structures probably populate some of the CTD SPXX sites, which can readily adopt this conformation (*50*). As discussed above, a segment containing one of these motifs undergoes a conformational change into a β-turn in the presence of TFE. Serine 53, which is included in this region, was shown to be phosphorylated during mitosis, possibly by cyclin B−$p34^{CDC2}$ or a related kinase (*21*). Phosphorylation of SPXX sites within the C-terminal domain of the largest subunit of RNA polymerase II was recently shown to promote transition from β-turn to PII conformation, in the presence of TFE (*29*).

Our data thus show that the CTD largely adopts an unordered, probably extended conformation, interspersed with short secondary structural motifs, one of which comprises PII sites. The latter may play structural roles by restricting the conformational space of the fragment and by providing sites for interactions with components of the NPC. gp210 CTD also contains a significant fraction of β-turns, which may compete with the PII conformation over some sites in the fragment. Like PII, β-turns are often involved in protein−protein interactions. In addition, they can readily modulate the length of the CTD, by breaking or forming. We show further that the conformation of the CTD is sensitive to macroscopic or local changes in hydrophobicity, which act to decrease solvent−backbone interactions. Such changes in the local dielectric may be effected, in situ, by neighboring proteins of the pore.

As discussed above, the CTD has many of the features characterizing natively unfolded proteins or protein domains. Notably, the yeast nucleoporin Nup2p has recently been assigned a member of this group (*51*). It was proposed that other proteins of the NPC might also be intrinsically unstructured (*51*). Unlike truly natively unfolded proteins, the CTD, however, still contains a significant amount of secondary structures. It also responds to an increase in temperature by undergoing partial unfolding, in contrast to natively unfolded proteins, which acquire structural complexity at elevated temperatures (*44*). Thus, the CTD appears to possess a conformation that is an intermediate between the (native) unfolded and folded states and which is distinct from that of a molten globule (that contains a well-developed secondary structure). A high degree of structural disorder may allow the CTD to simultaneously interact with multiple binding partners, which might be important in the assembly of the CTD onto the NPC during pore formation (*51*). Localized, dispersed secondary structural motifs, on the other hand, may serve as recognition sites for some of these interactions and may facilitate folding of the CTD consequent to binding.

PII-structured SP motifs in the CTD may also serve as recognition sites for regulatory proteins bearing type IV WW domains. WW domains, named after a pair of conserved tryptophans, are compact protein modules found in a diverse set of proteins involved in cell signaling or regulation (*31*, *32*, *52*). On the basis of their binding specificity, they are conventionally classified into five groups, of which group IV specifically recognizes phosphorylated serine/threonine-proline (pSP/pTP) motifs generated by cyclin-dependent protein kinases (*53*). This group includes two proteins: the ubiquitin ligase Nedd4 and the mitotic peptidyl-prolyl *cis*−*trans* isomerase Pin1. *Cis*−*trans* isomerization of proline residues by the latter enzyme may result in several diverse outcomes, including dephosphorylation, transcriptional activation, modulation of intracellular localization, and ubiquitylation (*54*). A potential target for Pin1 in the CTD might be serine 53, which is contained in the last SPXX motif in the sequence and which was shown to be specifically phosphorylated during mitosis.

Finally, while highly conserved among mammals, gp210 CTD is poorly conserved between phyla (Figure 6). Indeed, no apparent homology can be observed in the sequences of gp210 CTD orthologues from human, *Xenopus*, and *Drosophila*. However, like the mammalian CTDs, the CTDs of both *Xenopus* and *Drosophila* are highly basic (p*I* 9.5 and 10.2, respectively) and poorly hydrophobic. The *Xenopus* CTD also contains a large number of proline residues (eight), two of which form a PII-forming PXXP motif, which is preceded by a tyrosine residue that might serve as a potential phosphorylation site. The *Drosophila* CTD contains only four prolines. However, three of them are clustered together within an eight amino acid-containing region that includes PP and SPXX motifs. While these observations suggest that PII structures may also be present in the gp210 CTD of nonmammalian species, experimental verification is required to confirm this notion.

ACKNOWLEDGMENT

We thank Dr. Robert Woody for help with the CD analysis software.

REFERENCES

1. Macara, I, G. (2001) *Microbiol. Mol. Biol. Rev. 65*, 570−594.
2. Görlich, D., and Kutay, U. (1999) *Annu. Rev. Cell Dev. Biol. 15*, 607−660.

3. Mattaj, I. W., and Englmeier, L. (1998) *Annu. Rev. Biochem. 67*, 265−306.
4. Reichelt, R., et al. (1990) *J. Cell Biol. 110*, 883−894.
5. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, M. J. (2002) *J. Cell Biol. 158*, 915−927.
6. Akey, C. W., and Radermacher, M. (1993) *J. Cell Biol. 122*, 1−19.
7. Hinshaw, J. E., Carragher, B. O., and Milligan, R. A. (1992) *Cell 69*, 1133−1141.
8. Goldberg, M. W., and Allen, T. D. (1993) *J. Cell Sci. 106*, 261−274.
9. Akey, C. W. (1989) *J. Cell Biol. 109*, 955−970.
10. Kapon, R., Nevo, R., Shahar, D., and Reich Z. (2001) *Single Mol. 2*, 85−89.
11. Ris, H. (1991) *EMSA Bull. 21*, 54−56.
12. Jarnik, M., and Aebi, U. (1991) *J. Struct. Biol. 107*, 291−308.
13. Hallberg, E., Wozniak, R. W., and Blobel, G. (1993) *J. Cell Biol. 122*, 513−521.
14. Wozniak, R. W., Bartnik, E., and Blobel, G. (1989) *J. Cell Biol. 108*, 2083−2092.
15. Gerace, L., Ottaviano, Y., and Kondor-Koch, C. (1982) *J. Cell Biol. 95*, 826−837.
16. Berrios, M., Victoria, H. M., McConnell, M., and Fisher, P. A. (1995) *Eur. J. Cell Biol. 67*, 1−7.
17. Gajewski, A., Lourim, D., and Krohne, G. (1996) *Eur. J. Cell Biol. 71*, 14−21.
18. Olsson, M., Ekblom, M., Fecker, L., Kurkinen, M., and Ekblom, P. (1999) *Kidney Int. 56*, 827−838.
19. Gerace, L., and Burke, B. (1988) *Annu. Rev. Cell Biol. 4*, 336−374.
20. Greber, U. F., Senior, A., and Gerace, L. (1990) *EMBO J. 9*, 1495−1502.
21. Favreau, C., Worman, H. J., Wozniak, R. W., Frappier, T., and Courvalin, J. C. (1996) *Biochemistry 35*, 8035−8044.
22. Wozniak, R. W., and Blobel, G. (1992) *J. Cell Biol. 119*, 1441−1449.
23. Grote, M., Kubitscheck, U., Reichelt, R., and Peters, R. (1995) *J. Cell Sci. 108*, 2963−2972.
24. Cohen, M., Wilson, K. L., and Gruenbaum, Y. (2001) *Gene Ther. Mol. Biol. 6*, 47−55.
25. Bodoor, K., Shaikh, S., Salina, D., Raharjo, W. H., Bastos, R., Lohka, M., and Burke, B. (1999) *J. Cell Sci. 112*, 2253−2264.
26. Greber, U. F., and Gerace, L. (1992) *J. Cell Biol. 116*, 15−30.
27. Greber, U. F., and Gerace, L. (1995) *J. Cell Biol. 126*, 5−14.
28. Drummond, S. P., and Wilson, K. L. (2002) *J. Cell Biol. 158*, 53−62.
29. Bienkiewicz, E. A., Woody, A.-Y. M., and Woody, R. W. (2000) *J. Mol. Biol. 297*, 119−133.
30. Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. (1998) *DNA Res. 5*, 355−364.
31. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., and Noel, J. P. (2000) *Nat. Struct. Biol. 7*, 639−643.
32. Kay, B. K., Williamson, M. P., and Sudol, M. (2000) *FASEB J. 14*, 231−241.
33. Park, S.-H., Shalongo, W., and Stellwagen, E. (1997) *Protein Sci. 6*, 1694−1700.
34. Toumadje, A., and Johnson, W. C., Jr. (1995) *J. Am. Chem. Soc. 117*, 7023−7024.
35. Drake A. F., Siligardi, G., and Gibbons, W. A. (1988) *Biophys. Chem. 31*, 143−146.
36. Woody, R. W. (1992) *Adv. Biophys. Chem. 2*, 37−79.
37. Ma, K., Kan, L.-S., and Wang, K. (2001) *Biochemistry 40*, 3427−3438.
38. Makarov, A. A., Adzhubei, I. A., Protasevich, I. I., Lobachov, V. M., and Fasman, G. D. (1994) *Biopolymers 34*, 1123−1124.
39. Rucker, A. L., and Creamer, T. P. (2002) *Protein Sci. 11*, 980−985.
40. Swenson, C. A., and Formanek, R. (1967) *J. Phys. Chem. 71*, 4073−4077.
41. Dukor, R. A., Keiderling T. A., and Gut, V. (1991) *Int. J. Pept. Protein Res. 38*, 198−203.
42. Mantch, H. H., Percel, A., Hollósi, M., and Fasman, G. D. (1993) *Biopolymers 29*, 1801−1807.
43. Chen, Y., Liu, B., and Barkley, M. D. (1995) *J. Am. Chem. Soc. 117*, 5608−5609.
44. Uversky, V. N. (2002) *Eur. J. Biochem. 269*, 1−10.
45. Uversky, V. N. (2002) *Protein Sci. 11*, 739−756.
46. Dyson, H. J., and Wright, P. E. (2002) *Curr. Opin. Struct. Biol. 12*, 54−60.
47. MacArthur, M. W., and Thornton, J. M. (1991) *J. Mol. Biol. 218*, 397−412.
48. Kelly M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F., and Creamer, T. P. (2001) *Biochemistry 40*, 14376−14383.
49. Stapley B. J., and Creamer, T. P. (1999) *Protein Sci. 8*, 587−595.
50. Suzuki, M. (1989) *J. Mol. Biol. 207*, 61−84.
51. Denning, D. P., Uversky, V., Patel, S. S., Fink, A. L., and Rexach, M. (2002) *J. Biol. Chem.* (in press).
52. Zarrinpar, A., and Lim, W. A. (2000) *Nat. Struct. Biol. 7*, 611−613.
53. Lu P.-J., Zhou, X. Z., Shen, M., and Lu, K. P. (1999) *Science 283*, 1325−1328.
54. Shaw, P. E. (2002) *EMBO Rep. 3*, 521−526.

BI0266176