Symmetry and Secondary Structure of the Replication Terminator Protein of *Bacillus subtilis*: Sedimentation Equilibrium and Circular Dichroic, Infrared, and NMR Spectroscopic Studies†

Andrew V. Kralicek, Natalie A. Vesper, Greg B. Ralston, R. Gerry Wake, and Glenn F. King*

Department of Biochemistry, University of Sydney, Sydney, NSW 2006, Australia

Received January 22, 1993; Revised Manuscript Received June 14, 1993⊕

ABSTRACT: We have used analytical ultracentrifugation in combination with a number of spectroscopic techniques to analyze the symmetry and secondary structure of the DNA-binding replication terminator protein (RTP) of *Bacillus subtilis*. Sedimentation equilibrium studies confirm that RTP is a dimer in solution under the conditions used for spectroscopic analysis, whereas the number of cross peaks displayed in ^{1}H–^{15}N HSQC NMR spectra of uniformly ^{15}N-labeled RTP are consistent with the primary structure of the monomer. These two results in combination lead to the conclusion that RTP is a *symmetric dimer* in solution. Circular dichroic and Fourier-transform infrared spectra reveal, in contrast to the results obtained from a number of commonly used secondary structure prediction algorithms, that RTP contains 20–30% α-helical and 40–50% β-sheet/β-turn secondary structure and that the conformation of the protein remains unchanged over the pH range 5–8. It is proposed on the basis of protein folding-class prediction algorithms, in combination with various physical properties of RTP, that it belongs to the $\alpha + \beta$ protein-folding class.

Termination of replication of the circular bacterial chromosome involves the meeting and fusion of the two replication forks at the terminus region approximately opposite the origin. The process has been studied most intensively in *Bacillus subtilis* and *Escherichia coli* [see reviews by Kuempel et al. (1989), Lewis and Wake (1991), and Hill (1992)]. The first stage of termination in both organisms involves the arrest of one of the two replication forks at a specific DNA sequence (terminator) to which a terminator protein is bound. The terminator–protein complex is polar in its action, arresting a fork approaching from one direction, but not from the other. In the case of *E. coli* at least, it has been established that the ability to arrest a replication fork is due to a polar inhibition of the unwinding of the DNA double helix by the replicative DNA helicase (Lee et al., 1989; Khatri et al., 1989).

In *B. subtilis*, the replication terminator protein (RTP, 122 amino acid residues, 14.5 kDa) is encoded by the *rtp* gene which lies just downstream of two DNA terminators (imperfect inverted repeats) referred to as IRI and IRII (47 and 48 bp, respectively). IRI is the functional terminator *in vivo*; it acts to arrest the clockwise fork which arrives at the terminus, *terC*, just ahead of the anticlockwise fork. IRII is positioned to arrest the anticlockwise fork if it were to arrive at *terC* first. RTP is a dimer in solution at neutral pH (Lewis et al., 1990), and each of IRI and IRII contain two sites, designated "A" and "B", each of which binds a dimer of RTP (Lewis et al., 1990). It appears that both sites in each IR must be filled, with possible cooperative interaction between the two filled sites, to provide a functional terminator complex (Smith & Wake, 1992).

The *E. coli* replication terminator protein (Tus, for terminus utilization substance) binds to its DNA terminators as a monomer of 36 kDa [see Hill (1992)]. There is no sequence similarity between Tus and RTP, and neither of them appears to contain any of the structural motifs characteristic of various classes of DNA-binding proteins (Steitz, 1990). However, RTP exhibits sequence similarity over a 33-residue stretch with the *B. subtilis* DnaB protein involved in initiation of chromosome replication, implying that these proteins may share a common functional domain (Kralicek et al., 1991).

We are currently investigating the tertiary structure of RTP in order to gain a molecular understanding of its replication fork arrest function. In this paper we report on sedimentation equilibrium and heteronuclear magnetic resonance (NMR), Fourier-transform infrared (FTIR), and circular dichroic (CD) spectroscopic studies of RTP. We show that RTP is a symmetric dimer in solution, that it contains significant amounts of both α and β secondary structure, and that it probably belongs to the $\alpha + \beta$ protein-folding class.

EXPERIMENTAL PROCEDURES

Materials. Sodium 3-trimethylsilyl[2,2,3,3-^{2}H]propionate (d_4-TSP) was obtained from Fluka A.G. (Buchs, Switzerland), while $(^{15}NH_4)_2SO_4$ was from Cambridge Isotope Laboratories (Cambridge, MA). Fermentation grade D-[U-$^{13}C_6$]glucose was from Martek Corporation (Columbia, MD). Clostripain (EC 3.4.22.8), endoproteinase Arg-C from mouse submaxillary gland, endoproteinase Lys-C from *Lysobacter enzymogenes*, isopropyl β-D-thiogalactopyranoside (IPTG), and phenylmethanesulfonyl fluoride (PMSF) were purchased from Sigma Chemical Co. (St Louis, MO).

Purification of Replication Termination Protein. The *rtp* coding sequence was cloned into the pET-3c expression vector (Studier et al., 1990) to produce the plasmid pST2 (S. Thomson, G. F. King, and R. G. Wake, unpublished results). *E. coli* BL 21 cells, which also contain the plasmid pLysS, served as host for pST2. The cells were grown at 37 °C to mid-exponential phase in M9ZB medium, 50 mg/mL ampi-

† This work was supported by a grant to G.F.K. and R.G.W. from the Australian Research Council and by a University of Sydney Postgraduate Scholarship, an Australian Postgraduate Research Award, and a University of Auckland William Georgetti Scholarship awarded to A.V.K.

* Address correspondence to this author at the Department of Biochemistry, University of Sydney, Sydney, NSW 2006, Australia [tel: (61-2) 692-3902; fax: (61-2) 692-4726; email: gfking@extro.ucc.su.oz.au].

⊕ Abstract published in *Advance ACS Abstracts,* September 1, 1993.

0006-2960/93/0432-10216$04.00/0

cillin, and 25 mg/mL chloramphenicol or, to obtain ^{15}N-labeled RTP, in a defined minimal medium of 0.68% Na_2HPO_4, 0.3% KH_2PO_4, 0.05% NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 50 μg/mL ampicillin, 25 mg/mL chloramphenicol, and 0.1% $(^{15}NH_4)_2SO_4$ and 0.5% D-glucose as the sole carbon and nitrogen sources, respectively (Muchmore et al., 1989); in one additional experiment, the unlabeled glucose was replaced with D-[U-$^{13}C_6$]glucose in order to produce uniformly $^{15}N/^{13}C$-labeled protein. After induction with 0.5 mM IPTG, incubation was continued for either 2.5 h in rich medium, or 2 h in minimal medium, before harvesting the cells. The cells were resuspended in buffer A (25 mM Tris-HCl, pH 7.8, 250 mM sucrose, 1 mM PMSF, 2 mM EDTA), recentrifuged, snap frozen in liquid N_2, and stored at −80 °C. Cells (5 g) were resuspended in 45 mL of buffer B [50 mM Tris-HCl, pH 7.8, 1 mM EDTA, 1 mM dithiothreitol (DTT), 200 mM NaCl, 20 mM spermidine-HCl] and submitted to freeze-thawing until lysed. All subsequent steps prior to FPLC purification were performed at 4 °C. The cell debris was sedimented by centrifugation at 60000*g*, the supernatant was retained, and nucleic acids were removed as described previously (Lewis et al., 1989). This extract was dialyzed overnight against buffer C (50 mM KH_2PO_4/Na_2HPO_4, 1 mM DTT, pH 6.5). The sample was equilibrated with 20 mL 1 g dry weight) of preswollen CM-Sephadex C-25 (Pharmacia, Uppsala, Sweden), and then the supernatant was removed and the resin washed with 10 mL of buffer C. RTP was eluted from the resin with 12 mL of buffer C containing 0.7 M NaCl and then 4 mL of buffer C containing 1.5 M NaCl. The RTP-containing filtrates were pooled and dialyzed overnight against buffer C. The sample was then loaded via a superloop onto a cation-exchange FPLC column (Pharmacia Mono S, HR 10/10) equilibrated in buffer D (50 mM sodium phosphate, 1 mM DTT, pH 6.7) at room temperature and eluted with a linear 0–0.8 M NaCl gradient in buffer D. RTP eluted at 0.54 M NaCl and was collected as a 5.5-mL fraction. The RTP fraction was determined to be at least 98% pure as judged by SDS–PAGE.

Band Retardation Assays. Band retardation assays were performed as described previously (Lewis et al., 1989) except that vertical polyacrylamide gels were used instead of horizontal agarose gels. Each sample contained 0.109 pmol of a pPL-1 plasmid digest and appropriate amounts of RTP to give molar RTP/DNA ratios of 0, 1, 5, 10, 30, and 60. After 10 min was allowed for DNA–protein complexation, the samples were loaded onto a 1.5-mm 6% polyacrylamide vertical gel buffered with 36 mM Tris, 30 mM NaH_2PO_4, and 10 mM EDTA, pH 7.5. Gels were electrophoresed at 40 mA and 4 °C using a "Tall Mighty Small" apparatus (Hoefer Scientific Instruments, San Francisco) and then stained with ethidium bromide before photography.

Heteronuclear NMR Experiments. Two-dimensional (2D) NMR spectra of RTP were acquired on a Bruker AMX600 NMR spectrometer operating at 30 °C. Samples contained either 0.9–1.7 mM uniformly ^{15}N-labeled RTP in 100 mM NaCl, 2 mM DTT, 90% H_2O/10% D_2O, and 0.2 mM d_4-TSP, pH 5.9, or 1.0 mM uniformly $^{15}N/^{13}C$-labeled RTP in 100 mM NaCl, 9 mM DTT, 100% D_2O, and 0.2 mM d_4-TSP, pH 5.7. In one additional experiment, the 0.9 mM ^{15}N-labeled RTP sample was adjusted to pH 4.0, and a spectrum was acquired at 27 °C.

2D ^1H–^{15}N or ^1H–^{13}C heteronuclear single-quantum coherence (HSQC) spectra were acquired using published pulse sequences (Bax et al., 1990; Norwood et al., 1990); a single 1.5-ms spin-lock pulse was used immediately prior to the second ^1H $\pi/2$ pulse for water suppression (Messerle et al., 1989). No solvent presaturation was used during the intertransient delay period. The ^1H carrier frequency was always set to that of the water resonance, while in ^{15}N–^1H HSQC experiments the ^{15}N carrier frequency was positioned between the resonances arising from the six Arg ϵ-nitrogens and those from the backbone amide nitrogens, and in ^{13}C–^1H HSQC experiments the ^{13}C carrier frequency was positioned ~1000 Hz upfield of the tyrosine ϵ-carbon resonances. The GARP decoupling scheme (Shaka et al., 1985) was employed for ^{15}N or ^{13}C decoupling during acquisition. To minimize relaxation effects in ^{15}N–^1H HSQC experiments, the delay for evolution of ^{15}N–^1H one-bond couplings was set to 2.5 ms, slightly shorter than $1/4J_{NH}$ (Bax et al., 1990). For ^{13}C–^1H HSQC experiments this delay was set to 1.3 ms as a compromise between the one-bond ^{13}C–^1H couplings measured for the aromatic rings of Tyr/Phe residues ($^1J_{CH}$ ~ 160–165 Hz) and those measured for the imidazole ring of His residues ($^1J_{CH}$ ~ 200–205 Hz).

For ^{15}N–^1H HSQC spectra, the spectral widths were set to 11.1 ppm (6.67 kHz) for F_2 and 70 ppm (4.2 kHz) for F_1. For ^{13}C–^1H HSQC spectra, the spectral widths were set to 8.3 ppm (5 kHz) for F_2 and 73 ppm (11 kHz) for F_1; thus, the aliphatic ^1H–^{13}C correlations, which were not used for analysis, were severely aliased in the ^{13}C (F_1) dimension of this experiment. Time-proportional phase incrementation (Marion & Wüthrich, 1983) was employed to achieve quadrature detection. A total of 512–983 t_1 increments of 2048 complex data points were collected with a total of 32–64 transients per t_1 increment. The ^1H-frequency scale was referenced directly to d_4-TSP at 0.00 ppm, while the ^{15}N-frequency scale was indirectly referenced to liquid NH_3 by using the ^1H frequency of the d_4-TSP resonance (Live et al., 1984). The ^{13}C-frequency scale was indirectly referenced to d_4-TSP using the method of Bax and Subramanian (1986).

2D spectra were processed using UXNMR software running on either Bruker Aspect X32 or Silicon Graphics 4D/20 workstations. The digital resolution was improved by zero-filling to 4096 × 1024 ($F_2 \times F_1$) real data points. The spectral resolution was enhanced by apodization with a Lorentz–Gauss function in t_2 and a shifted sine bell function t_1 prior to Fourier transformation. After phase correction, baselines were corrected by subtracting a third-order polynomial fitted to the baseline.

Clostripain Digestion of RTP. Prior to use, clostripain was preactivated for 2 h at 30 °C in 50 mM Tris-HCl, pH 7.8, 2.5 mM DTT, and 1 mM $CaCl_2$ (Giles et al., 1979). Proteolytic cleavage of RTP was carried out at an RTP concentration of 0.5 mg/mL in the presence of 10% w/w clostripain, 50 mM Tris-HCl, pH 7.8, 10 mM DTT, and 1 mM $CaCl_2$ at 18 °C. The profile in Figure 4 was constructed from a 9-h digest by periodically removing aliquots and denaturing the protein by boiling. The samples were then electrophoresed on an 18% tricine–SDS–polyacrylamide gel (Schägger & von Jagow, 1987).

Individual RTP fragments were isolated from a 4-h digest of 300 μg of RTP by the following procedure: the digest was loaded onto a C18 reverse-phase column (Ultropak TSK ODS-120T 10 μm, 7.8 × 300 mm) equilibrated in 0.1% TFA/25% acetonitrile, and proteolytic fragments were eluted by applying a gradient of 25–50% acetonitrile in 0.1% TFA at 0.5 mL/min. Fractions containing RTP fragments were lyophilized, and these samples were subjected to five cycles of N-terminal sequencing on an Applied Biosystems 470A protein sequencer at the Brisbane Protein and Nucleic Acid Research Centre, University of Queensland. The molecular weights of each fragment were determined on a Sciex API III electrospray

mass spectrometer at Research Park, Bond University, Queensland.

Circular Dichroic Spectroscopy. A solution of 34 μM RTP in 100 mM NaF was used for CD studies; NaF was used rather than NaCl to maintain ionic strength because of its lower absorbance in the UV region. Circular dichroic spectra of the sample at pH 5.0, 6.0, 7.0, and 7.9 were collected over the absorbance range 195–260 nm using a 0.01-cm path length quartz cell on a Jobin Yvon Dichrographe III instrument under constant nitrogen flush at 20 °C. The instrument had been calibrated with epiandrosterone. Each spectrum was digitized every 0.5 nm and a background spectrum of 100 mM NaF was subtracted. The data were streamlined by averaging through a moving window of five data points before calculation of optical activity. Estimates of protein secondary structure at each pH were calculated using two methods: (i) fitting basis spectra derived from Chang et al. (1978) or (ii) fitting reference curves derived from convex constraint analysis (Lincomb method; Perczel et al., 1992).

Meniscus Depletion Sedimentation Equilibrium Experiments. Three different loading concentrations (approximately 0.25, 0.5, and 1.0 g/L in 100 mM NaCl, pH 5.9) of the ^{15}N-labeled RTP used in NMR experiments were centrifuged at an angular velocity of 32 000 rpm for up to 24 h in a Beckman-Spinco Model E analytical ultracentrifuge fitted with electronic speed control and a rotor temperature indicator control unit. The Rayleigh interference camera was focused on the $^2/_3$ plane of the cell to minimize errors from Weiner skewing. Sample volumes of 0.13 mL, giving a solution column of 3 mm, were routinely used. As An-F Ti rotor, a Yphantis 12-mm six-channel centerpiece, and sapphire windows were used. The use of silicone layering oil was avoided in sedimentation equilibrium experiments. The experiments were performed initially at a temperature of 20 °C for 18 h, after which time a photograph was taken and the temperature increased several degrees for an additional 4 h before taking another photograph. This procedure was repeated until a measurement had been made at 32 °C. The Rayleigh interference pattern was recorded photographically on Kodak-T-Max 35-mm film. The photographs were measured on a Nikon comparator, at 20× magnification, with the aid of an automated plate reader (De Rosier et al., 1972). A photograph was taken at 10 000 rpm during acceleration for measurement of baseline correction. A concentration of 1.0 g/L was assumed to be equivalent to a fringe displacement of 4.04 fringes.

To determine if all three samples in each experiment were homogeneous and had reached chemical equilibrium during the time course of the experiment, a reference concentration, $c(r_F)$, common to all three channels was chosen (usually 1.0 g/L), and plots of $\Omega(r)$ versus $c(r)$ were calculated from the experimental $c(r)$ data using

$$\Omega(r) = c(r) \exp[\phi M_1(r_F^2 - r^2)]/c(r_F)$$

where $\phi_1 = (1 - \bar{v}\rho)\omega^2$, $\bar{v}$ is the partial specific volume of RTP, ρ is the solvent density, ω is the angular velocity, and $c(r)$ is the total protein concentration at radial distance r (Milthorpe et al., 1975). M_1, the monomer molecular weight for RTP, was taken as 14 500, $\bar{v}$ was calculated to be 0.75 mL/g (Perkins, 1986), and ρ was calculated to be 1.002 g/mL. The square of the radial position, r_F, corresponding to the reference concentration in each channel was estimated by interpolation using a six-point quadratic. Coincidence of the $\Omega(r)$ versus $c(r)$ plots over their common concentration ranges was taken as an indication of the attainment of chemical equilibrium and the absence of contaminants not participating in chemical equilibria.

In addition, the distribution of the apparent weight average molecular weight (Teller, 1973) was calculated at the midpoints of a sliding, 11-point quadratic fit to ln $c(r)$ versus r^2 data:

$$M_{w,app} = (2RT \, d \ln c(r)/dr^2)/[(1 - \bar{v}\rho)\omega^2]$$

Again, overlap of the data from different initial loading concentrations was taken as an indication of the attainment of chemical equilibrium and the absence of contaminants.

Fourier Transform Infrared Spectroscopy. Four hundred microliters of a 7.7 mg/mL sample of RTP in 100 mM NaCl at pH 5.5 was dialyzed twice against 30 mL of 100 mM NaCl in D_2O at 5 °C. The sample was then split into 100-μL aliquots which were then dialyzed against 100 mM NaCl in D_2O at pD (uncorrected pH meter reading) 5.0, 6.1, 7.1, and 8.1. All four samples were then diluted with their respective dialysis solutions to give a concentration of 3.4 mg/mL. Infrared spectra were recorded at 20 °C for each pD value using a DIGILAB FTS-20/80 spectrometer equipped with a HgCdTe detector. Samples were placed in a demountable cell fitted with CaF_2 windows, a teflon spacer, and a path length of 0.05 mm. The sample compartment was continuously purged with dry N_2 to eliminate absorption of water in the spectral region of interest.

Data acquisition, storage, and processing was performed using the standard DIGILAB software supplied with the spectrometer. Absorption spectra were the result of 1024 scan interferograms measured at 2-cm^{-1} resolution that had been Fourier transformed and ratioed against a single beam background spectrum (collected in the absence of the cell). Spectra of the dialysis solutions were subtracted from the sample spectra at the appropriate pD to give infrared absorption spectra of RTP only. These were then truncated to the region of interest (i.e., the amide I band), deconvolved using triangular apodization with σ = 15 cm^{-1} (half-width at half-height) and k factor = 1.7–1.8, and baseline corrected using a spline function. Quantitative analysis of the secondary structure of RTP at each pD was performed using the JPLOT fitting program (Dr. Jeff Riemers, Department of Chemistry, University of Sydney), which employs a Gauss–Newton nonlinear least-squares minimization procedure.

RESULTS

Heteronuclear NMR Experiments. The ^{15}N–^{1}H HSQC experiment correlates the ^{1}H and ^{15}N chemical shifts of directly bonded ^{15}N–^{1}H pairs by selecting for the large ^{15}N–^{1}H one-bond scalar couplings of approximately 90 Hz (Bax et al., 1990; Norwood et al., 1990). We acquired ^{15}N–^{1}H HSQC spectra of RTP using a spectral width in the F_1 (^{15}N) dimension that would encompass both backbone amide and arginine $^{15}N_\epsilon$–^{1}H correlations. Previous ^{15}N–^{1}H HSQC experiments which employed spectral widths that encompassed the full range of direct correlations between protein nitrogen atoms and their covalently bonded protons showed an absence of correlations involving histidine side chain imino protons and lysine side chain amide protons, indicating that these protons are exchanging with the solvent too rapidly on the NMR timescale to be observed in this experiment.

On the basis of the primary structure of RTP (Carrigan et al., 1987), we would expect to observe 120 backbone amide correlations (122 residues less the N-terminal Met and Pro-52), eight pairs of side chain amide correlations (two Asn and six Gln residues), and six Arg $^{15}N_\epsilon$–^{1}H cross peaks. Figure 1A shows the backbone amide region of a ^{15}N–^{1}H HSQC spectrum of RTP; the cross peaks are well resolved in both dimensions, indicating that RTP is likely to be highly

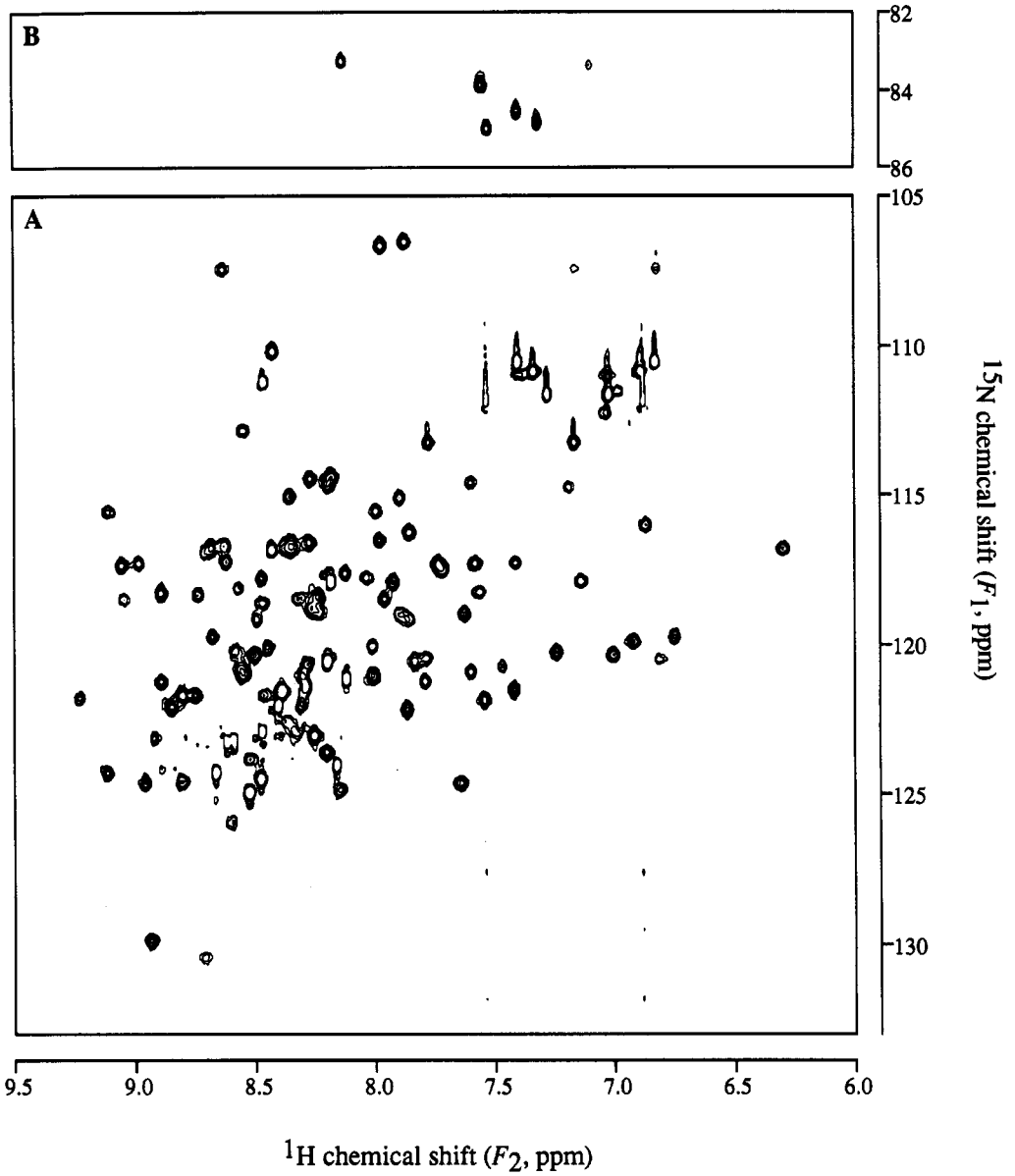

FIGURE 1: (A) Backbone amide region of the 600-MHz ^{15}N–^{1}H HSQC spectrum of uniformly ^{15}N-labeled RTP (1.7 mM, pH 5.9) recorded at 30 °C without solvent presaturation. (B) Arg $^{15}N_{\epsilon}$–^{1}H region from the same HSQC experiment.

structured. We have identified 116 backbone amide correlations and seven pairs of Asn and Gln side chain $^{15}NH_2$ correlations in this region of the spectrum; the latter cross peaks each have a weak partner 0.5–0.6 ppm upfield in the ^{15}N dimension resulting from the deuterium isotope effect produced by the 10% semideuterated NHD moieties present in solution (Bax et al., 1990). Figure 1B shows the Arg $^{15}N_{\epsilon}$–^{1}H region in which we can observe a set of six cross peaks, one of which has noticeably weaker intensity implying that it is more exposed to the solvent.

Following NMR experiments on the 0.9 mM ^{15}N-labeled RTP sample, a band retardation assay was immediately performed to ensure the protein remained functional throughout the NMR study. Figure 2 shows the results of a band retardation assay which compares the behavior of unlabeled RTP with that of the ^{15}N-labeled RTP. The two protein samples exhibited indistinguishable specificity and affinity for the four RTP binding sites present on the 219-bp fragment of pPL-I, with each producing the characteristic ladder of retarded bands with increasing concentrations of RTP (Lewis et al., 1989). This indicates that the NMR experiments were performed on a fully functional sample of RTP.

Meniscus Depletion Sedimentation Equilibrium Studies. Previous studies at 20 °C have reported that RTP exists as a dimer in solution at pH 7.8, whereas at pH 4.6 and low salt concentration there was evidence of dissociation to monomer (Lewis et al., 1990). In order to ascertain the oligomeric state of the RTP used in the NMR experiments, meniscus depletion sedimentation equilibrium studies were performed on the 0.9 mM ^{15}N-labeled RTP immediately following the NMR experiment at temperatures between 20 and 32 °C, in 100 mM NaCl, pH 5.9. Over the common concentration range, the data at each temperature showed good overlap of the $\Omega(r)$ distribution and of the apparent weight average molecular weight ($M_{w,app}$), indicating the attainment of both chemical and sedimentation equilibrium. At all temperatures, the $M_{w,app}$ extrapolated to approximately 32 kDa at zero concentration, demonstrating that RTP forms a stable dimer under the chosen range of experimental conditions. No evidence of dissociation to monomer, or association beyond dimer, was seen at the accessible concentrations between 20 and 32 °C. The $M_{w,app}$ showed a slight decrease with protein concentration, consistent with nonideality arising from the size and charge of RTP. Fitting the $M_{w,app}$ data obtained at

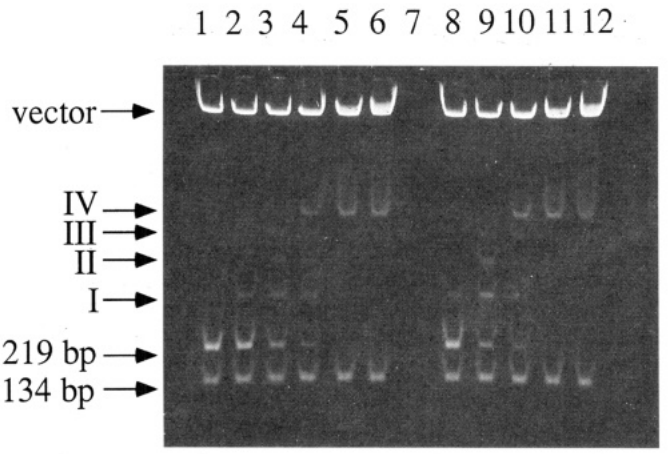

FIGURE 2: Band retardation assay comparing the DNA-binding activity of unlabeled RTP with ^{15}N-labeled RTP used in the NMR experiments. Lanes 2–6 and Lanes 8–12 show the effect of increasing concentrations of unlabeled and labeled RTP, respectively, on the mobility of a 219 bp fragment of DNA containing IRI and IRII. The molar RTP:DNA ratios for both lanes 2–6 and 8–12 are 1, 5, 10, 30, and 60. Lane 1 is a control which contains no RTP. The large band near the wells represents the bulk of the vector, and the 134 bp fragment is a region of DNA adjacent to the 219 bp fragment in the undigested plasmid. Bands I–IV represent the four protein–DNA complexes that result as the four RTP binding sites in IRI and IRII are filled (Lewis et al., 1989).

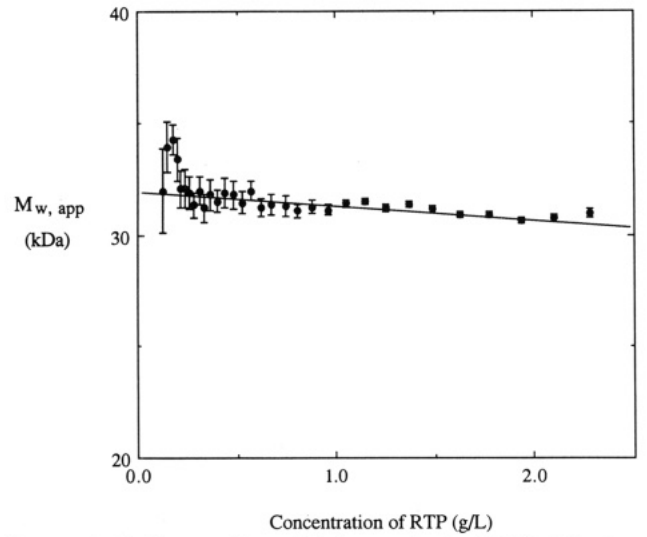

FIGURE 3: Sedimentation equilibrium analysis of RTP. The data from three individual samples of 0.25, 0.5, and 1.0 g/L were pooled, sorted, and averaged five points at a time to produce the mean apparent weight-average molecular mass corresponding to the mean concentration of the group of five points; the error bars show the standard error of this mean. The pre-averaged data were then fitted with a model for a single nonideal species to estimate the molecular mass and the second virial coefficient.

25 °C over the range 0.04–2.3 g/L (see Figure 3) yielded an estimate for the molecular weight of 31.9 kDa and an estimate for the second virial coefficient of 5.4×10^{-7} L mol g^{-2}.

It was recently reported that dimeric calcineurin B could be dissociated into monomers in the presence of the detergent 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) without altering the protein conformation (Anglister et al., 1993). However, we found no evidence of dissociation of dimeric RTP into monomers when we repeated the above sedimentation equilibrium studies with RTP dissolved in the presence of 5 mM CHAPS.

Correlation of Sedimentation Equilibrium and NMR Results. Whereas the ultracentrifugation studies indicate that RTP exists as a dimer under the conditions of the NMR experiment, the number of cross peaks seen in the ^{15}N-^{1}H HSQC experiment is consistent with what one would expect for monomeric RTP. These two results in combination lead to the suggestion that native RTP is a *symmetric* dimer in solution. However, the ^{15}N–^{1}H HSQC experiment needs to be interpreted cautiously since the backbone amide protons of proteins can exchange with protons of solvent water; if this exchange is rapid on the NMR time scale (i.e., exchange rate constants in the order of milliseconds) then a ^{15}N-^{1}H correlation will not be observed for the amide proton. The "disappearance" of rapidly exchanging amide protons is significantly exacerbated by radiofrequency irradiation of the water protons, and consequently this was completely avoided in our 2D HSQC NMR experiments; instead, spin-lock purge pulses (Messerle et al., 1989) were used for suppression of the solvent resonance.

However, in order to further address the possibility of cross peak "disappearance" due to rapid amide proton exchange, we repeated the ^{15}N–^{1}H HSQC experiment at pH 4.0 and 27 °C (compared to pH 5.9 and 30 °C in the original experiments), where the rate of exchange of RTP amide protons with those of solvent water would be expected to be reduced by some two orders of magnitude (Creighton, 1993). We found no significant difference in the number of cross peaks observed in this experiment (data not shown). This strongly argues that we are observing the full complement of amide protons at pH 5.9 and 30 °C and thus supports the conclusion that RTP is a symmetric dimer.

In order to further substantiate this conclusion, we acquired a ^{13}C–^{1}H HSQC spectrum of uniformly $^{15}N/^{13}C$-labeled RTP. The advantage of this experiment lies in the fact that the protons directly attached to the ^{13}C nuclei are not labile, and consequently all ^{13}C–^{1}H correlations should be observed. Unfortunately, the large number of correlations involving aliphatic carbons rendered this region of the spectrum unsuitable for analysis due to excessive cross peak overlap. However, we were able to analyze the aromatic region of the ^{13}C–^{1}H HSQC spectrum of RTP due to the small number of aromatic/heterocyclic residues in the protein (i.e., six Phe, six Tyr, two His). While cross peaks involving the Phe aromatic carbons were excessively overlapped, the number of ^{13}C–^{1}H correlations involving His γ- and ϵ-carbons was consistent with what one would expect for monomeric RTP (data not shown). Thus, the combined ^{15}N–^{1}H and ^{13}C–^{1}H HSQC experiments support the conclusion that RTP is a symmetric dimer; however, we cannot dismiss the possibility that there might be a small amount of local asymmetry at the site where the two monomers make intimate contact.

Clostripain Digestion. In order to further probe the conformation of RTP, we digested the protein with the enzyme clostripain, a cysteine protease which cleaves on the C-terminal sides of Arg residues (Mitchell & Harrington, 1970). RTP processes six Arg residues which are distributed throughout its sequence (i.e., residues 6, 16, 31, 59, and 109). Lys peptide bonds (of which there are 20 in RTP) are also cleaved by clostripain but generally one to two orders of magnitude less efficiently than Arg peptide bonds (Mitchell, 1968).

The cleavage profile from a 9-h digest of RTP in 50 mM Tris-HCl, 10 mM DTT, and 1 mM $CaCl_2$, pH 7.8, with 10% w/w clostripain at 18 °C is shown in Figure 4. A total of four fragments are generated over time, with fragments III and IV being the final products. An identical profile is seen when RTP is digested with endoproteinase Lys-C, while digestion with endoproteinase Arg-C yields only fragment I (data not shown). All four fragments were successfully isolated from a 4-h clostripain digest of RTP by using C18 reverse-phase

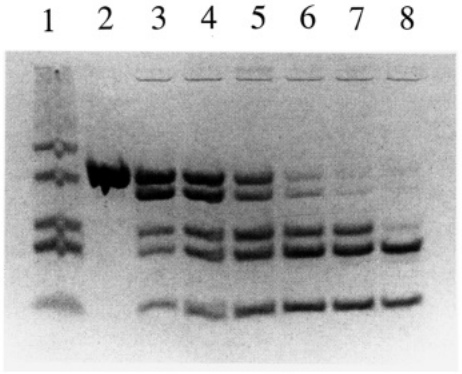

FIGURE 4: Profile from a 10% w/w clostripain digest of native RTP (0.5 μg/μL) in 50 mM Tris-HCl pH 7.8, 10 mM DTT, and 1 mM $CaCl_2$ at 18 °C. Over a period of 9 h, aliquots of digest were removed and denatured by boiling and then electrophoresed on an 18% tricine–SDS–polyacrylamide gel (Schägger & von Jagow, 1987). Lane 1 contains various myoglobin molecular weight standards: residues 1–153 (17.0 kDa), residues 1–131 (14.4 kDa), residues 56–153 (10.6 kDa), residues 56–131 (8.2 kDa), and residues 1–55 (6.2 kDa). Lane 2 contains an aliquot of the digest solution prior to addition of enzyme. Lanes 3–8, respectively, show samples taken at 30 min, 1 h, 2 h, 4 h, 6 h, and 9 h after the addition of clostripain. The slowest migrating band in lanes 3–8 corresponds to clostripain. See the text for an explanation of bands I–IV.

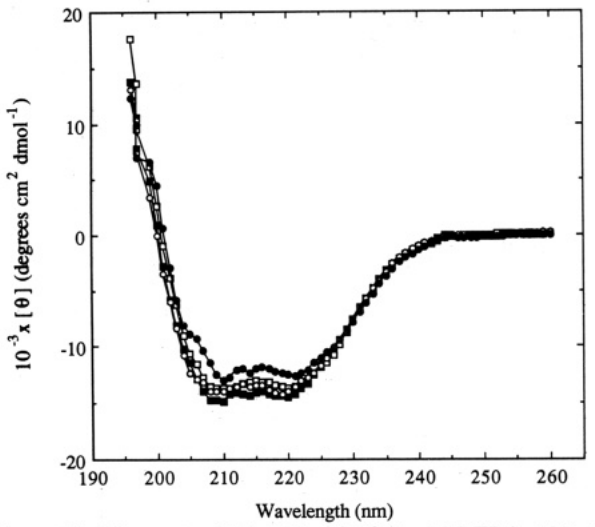

FIGURE 5: CD spectra (196–260 nm) of 34 μM RTP in 100 mM NaF, 20 °C, at pH 5.0 (■), 6.0 (□), 7.0 (●), and 7.9 (○). [θ] denotes the mean residue ellipticity.

HPLC, as the fragments were found to be associated under nondenaturing conditions. N-terminal sequencing (five cycles) and electrospray mass spectrometry enabled us to identify fragment I as residues 7–122, fragment II as residues 1–81, fragment III as residues 7–81, and fragment IV as residues 82–122. Thus, two clostripain cleavage sites can be identified: Arg-6 and Lys-81. These residues may be part of exposed loops within domains or the linking regions between domains. The other five Arg residues appear to be protected from proteolysis, possibly because RTP adopts a compact domain structure. This conclusion is supported by the ^{15}N–^1H HSQC NMR experiment, in which we observe one weak and five strong Arg $^{15}N_\epsilon$–^1H cross peaks. The weak cross peak possibly corresponds to the exposed Arg-6 residue which is accessible to clostripain.

Circular Dichroic Spectroscopy. CD studies of RTP were performed at pH 5.0, 6.0, 7.0, and 7.9 in 100 mM NaF, the same ionic strength as the previous NMR experiment and very similar to that of the band retardation assay (Lewis et al., 1989). The resulting CD spectra are shown in Figure 5; the four plots are very similar, with only the pH 7.9 spectrum showing a minor increase in ellipticity between 205 and 225 nm. In general, the CD spectra resemble the one presented by Mehta et al. (1992), which was collected from a sample of RTP in 5 mM KH_2PO_4, pH 7.0. Minor differences are that our spectra are not as positive in ellipticity between 195 and 200 nm, they are missing the "step" at 203 nm, and, unlike the spectrum of Mehta et al. (1992), our spectra coincide with zero ellipticity above 240 nm. These differences may result from the difference in ionic strength between the two studies.

Table I: Secondary Structure Percentages for RTP Determined from Deconvolution of the Amide I Region of FTIR Spectra or from Deconvolution of CD Spectra Using either the Lincomb Method of Fasman and Co-workers (Perzcel et al., 1992) or the Method of Yang et al. (1986)[a]

	CD (Yang et al., 1986)		CD (Lincomb)		FTIR	
pH	% α	% β	% α	% β	% α	% β
5.0	26	39	19	44	26	46
6.0	27	43	23	40	28[b,1]	46[1]
7.0	25	40	19	35	28[2]	51[2]
7.9	29	39	17	47	26[3]	50[3]
average	27	40	20	42	27	48

[a] % α indicates the calculated α-helical content of the protein, while % β represents the sum of the calculated β-sheet and β-turn components. [b] Actual pD values for FTIR spectra were [1]6.1, [2]7.1, and [3]8.1.

The CD spectra of RTP have the characteristic shape associated with a highly α-helical protein, namely, a double minimum at 222 and 208–210 nm, and a maximum below 200 nm (Yang et al., 1986). This is consistent with empirical secondary structure predictions obtained using the methods of Chou and Fasman (1978), Deléage and Roux (1987), Garnier et al. (1978), and Holley and Karplus (1989), which together predict that RTP is approximately 70% α-helical. We have applied two approaches to calculate the secondary structure of RTP from the CD spectra: (i) the Lincomb method which performs a fit of artificial pure component curves derived using convex constraint analysis of 25 CD spectra obtained from 25 different proteins (Perczel et al., 1992); and (ii) fitting pure component reference curves determined by Chang et al. (1978). In the case of the Lincomb method, the secondary structure percentages were calculated as the fraction of the contribution to total ellipticity after the deconvolved pure component curve attributed to aromatic residues and disulfides was removed. We chose to combine β-sheet and β-turn as the Lincomb method produces component curves which are a combination of the two and also because what constitutes representative CD spectra for β-turns is an open question (Yang et al., 1986). The results are shown in Table I, and two important points should be noted. First, there does not seem to be any major change in global secondary structure with pH changes in the range 5–8. Second, contrary to empirical secondary structure predictions, the fitting methods indicate RTP is approximately 40–45% β-sheet/β-turn and 20–30% α-helix. Mehta et al. (1992) only reported the α-helical content (32%) of RTP from their CD spectrum.

Fourier Transform Infrared Spectroscopy. Figure 6 shows the amide I region from the deconvolved FTIR spectrum of RTP at pD 5.0. Line shape analysis of this spectrum enabled the identification of seven component bands: one at 1657 cm^{-1} that is associated with α-helix, two bands at 1679 and 1634 cm^{-1} that represent β-sheet, a band at 1669 cm^{-1} that corresponds to β-turn, a band at 1647 cm^{-1} that correlates with unordered structure, and two bands at 1615 and 1606 cm^{-1} that were assigned to side chain components. Similar FTIR spectra were obtained at the other pD values. Table I shows the calculated secondary structure percentages for

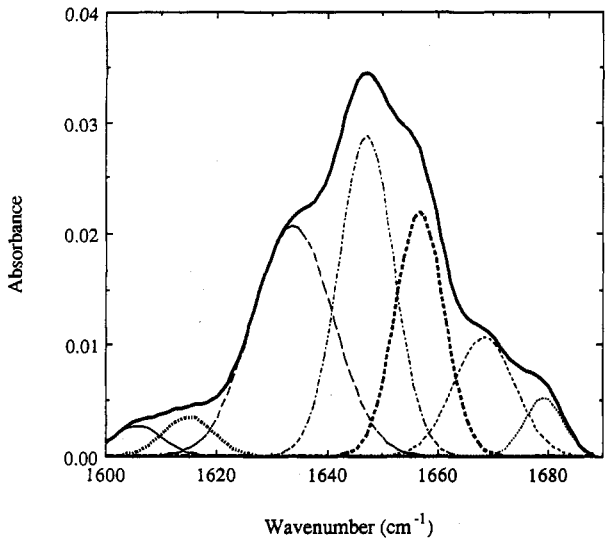

FIGURE 6: Amide I region of the deconvolved FTIR spectrum (—) obtained for 3.4 mg/mL RTP in 100 mM NaCl/D_2O at pD 6.0. The individual secondary structure components were calculated from a nonlinear least-squares fit in which all amplitudes and line widths were allowed to vary. The fitted curves correspond to α-helix (- - -), β-sheet (··· and – – –), β-turn (- · -), unordered structure (- · · -), and side chain contributions (— and • • •).

each pD as determined from the areas of the individual component bands; the side chain bands were excluded from these calculations. The secondary structure percentages determined from the FTIR spectra are very similar to those determined from deconvolution of CD spectra and further confirm that there is little, if any, change in the secondary structure of RTP over the pH range 5–8.

DISCUSSION

RTP displays no obvious sequence homology with any of the structural motifs which characterize various classes of DNA-binding proteins, including the helix–turn–helix, zinc finger, and leucine zipper proteins (Steitz, 1990). In this respect, the structure of RTP is of considerable interest as it may contain a previously uncharacterized DNA-binding motif. In this work we have probed the structure of RTP using ultracentrifugation in combination with a number of spectroscopic techniques.

Deconvolution of CD and FTIR spectra of RTP reveals that the protein contains ~20–30% α-helical structure and ~40–50% β-sheet/turn structure and that the conformation of the protein remains stable in the pH range 5–8. The secondary structure percentages derived from the two spectroscopic techniques are in good agreement with each other but vastly different from the predominantly α-helical secondary structure predicted by commonly used secondary structure prediction algorithms which are based solely on amino acid sequence. This result strongly supports the observation that secondary structure prediction algorithms are often able to predict the total extent of secondary structure, while failing to correctly analyze the individual amounts of the various types of secondary structure (Taylor & Thornton, 1984).

The spectroscopic observation that RTP is highly structured (i.e., only ~30% random coil) is consistent with the fact that only one of the six Arg residues in the protein (Arg-6) was accessible to the arginine-specific protease clostripain. The fact that Arg-6 is accessible to clostripain may indicate that the N-terminal region of the protein is relatively unstructured, which is consistent with the fact that the ^{15}N–^{1}H HSQC NMR spectrum of RTP indicates that one of the six Arg residues (presumably Arg-6) is more exposed to solvent than the other five.

The CD and FTIR data clearly indicate that RTP belongs to either the α/β or $\alpha+\beta$ protein folding class. Two recently proposed folding-class prediction methods were used in an attempt to discriminate between these possibilities on the basis of the amino acid sequence of RTP. The method recently introduced by Chou and Zhang (1992) enables calculation of a correlation coefficient for each of the four major protein-folding classes (i.e., α, β, α/β, $\alpha+\beta$). Using this method, we calculate a correlation coefficient of 0.85 for each of the four classes, indicating that this method does not enable a positive prediction of the protein-folding class for RTP. We also used the "greatest correspondence" approach introduced by Chou (1989). In this method, the percentage of each amino acid in a protein is compared with the percentage of that amino acid in each of the four protein-folding classes. The protein class with the best match is scored 1 and the others 0; a value of 0.5 is scored for classes which tie for the best match. Thus, if the amino acid composition of a protein exactly matched a particular protein-folding class, it would score a corresponding value of 20 for that class and 0 for the other three classes. We scored correspondence values of 7.5 ($\alpha+\beta$), 5.5 (β), 4 (α), and 3 (α/β) for RTP, thus giving a firm prediction that RTP belongs to the $\alpha+\beta$ protein-folding class. This result is reinforced by the correspondence between the size of the RTP polypeptide (122 residues) and the average chain length for $\alpha+\beta$ class proteins (135 residues; Chou, 1989), which is significantly below the average chain length for α/β class proteins (271 residues; Chou, 1989). Furthermore, there are numerous examples of $\alpha+\beta$ DNA-binding proteins which recognize specific DNA sequences [e.g., Arc repressor (Breg et al., 1990), Met repressor (Rafferty et al., 1989), and catabolite gene activator protein (Steitz, 1990)], whereas there are fewer examples of α/β DNA-binding proteins, and these often bind DNA in a largely nonspecific manner [e.g., DNase I (Suck et al., 1988)].

Sedimentation equilibrium studies (see Figure 3) confirm that RTP is a dimer in solution under the conditions used for spectroscopic analysis, whereas the number of cross peaks displayed in ^{15}N–^{1}H HSQC NMR spectra of RTP (see Figure 1) are consistent with the primary structure of the monomer. These two results in combination lead to the conclusion that RTP is a *symmetric dimer* in solution. This 2-fold symmetry is a property common to many dimeric DNA-binding proteins. For the large majority of these proteins, the 2-fold symmetry enables identical binding sites (usually an α-helix) on each monomer to be presented to each half-site of a (partially) symmetric recognition site, which usually spans consecutive major grooves of DNA; prokaryotic proteins exhibiting these DNA-binding properties include λ phage Cro repressor (Ohlendorf et al., 1982), phage 434 repressor (Anderson et al., 1987), and *E. coli trp* repressor (Otwinowski et al., 1988).

A different mode of DNA-binding is presented by the structurally homologous Met and Arc prokaryotic repressor proteins, which also form symmetric dimers (Breg et al., 1990). These proteins dimerize via formation of an antiparallel β-sheet, to which each monomer contributes one strand. *Two* dimers are bound in adjacent major grooves, with the β-sheet of each dimer interacting with a *repeated* DNA recognition sequence (Somers & Phillips, 1992). It was observed in previous studies of RTP that each of IRI and IRII contain a singly repeated 8 bp asymmetric sequence approximately centered within each binding site of the IR (Lewis et al., 1990). The possibility was raised that this 8 bp sequence might act as a recognition sequence in the binding of RTP dimer to each

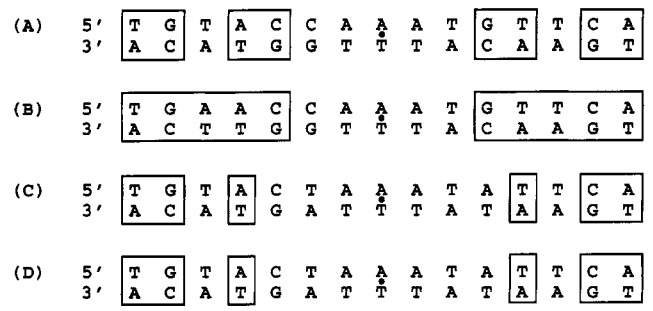

FIGURE 7: Partially symmetric DNA sequence located within the "B" site of the IRI and IRII regions of *terC*. This sequence is completely protected by RTP in DNase I footprinting assays (Lewis et al., 1990). Sequences are given for the IRI and IRII of *B. subtilis* strains 168 (A and C, respectively) and W23 (B and D, respectively). The small solid circle represents the center of symmetry.

of the two adjacent RTP-binding sites ("A" and "B") in IRI and IRII. This is reminiscent of the situation described above for the Met and Arc repressors (although the repeats in these cases are symmetrical), and the possibility that the mode of RTP binding to DNA may be similar to that exhibited by these two proteins is somewhat supported by a weak sequence homology we have detected between the Arc repressor and the N-terminal 50 residues of RTP and the fact that the data presented in this paper suggest that RTP, like the Arc and Met repressors, belongs to the $\alpha+\beta$ protein-folding class.

However, further analysis of the IRI and IRII has revealed that the "B" site encompasses a 15 bp DNA sequence of considerable symmetry (see Figure 7). This segment is contained within the 21 bp "core" IRI sequence described by Lewis et al. (1990) which can sustain binding of a single RTP dimer in the absence of the remainder of the IRI DNA. Thus, we cannot presently exclude the possibility that at least some RTP binding occurs in the "classical" manner described above for the Cro and *trp* repressor proteins. Resolution of the exact mode of RTP binding to DNA awaits a more detailed analysis of the DNA base pairs which directly interact with RTP and elucidation of the tertiary structure of RTP using either NMR spectroscopy or X-ray diffraction techniques.

ACKNOWLEDGMENT

We thank Prof. Gerald Fasman for providing the Lincomb program, Dr. Bill Bubb for his expert help with the heteronuclear NMR experiments, and Dr. Michael Morris and Dr. Ross Smith for their invaluable help with the recording and analysis of FTIR and CD spectra, respectively. We also acknowledge the contributions of Mr. Scott Thomson and Ms. Alison Franks in the early stages of this project, and Dr. Wayne Fairbrother and Dr. Wolfgang Bermel for illuminating discussions.

REFERENCES

Anglister, J., Bax, A., Delaglio, F., Grzesiek, S., & Vuister, G. W. (1993) *J. Cell. Biochem. 17C*, 243.
Anderson, J. E., Ptashne, M., & Harrison S. C. (1987) *Nature 326*, 846–852.
Bax, A., & Subramanian, S. J. (1986) *J. Magn. Reson. 67*, 565–569.
Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304–318.
Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R., & Kaptein, R. (1990) *Nature 346*, 586–589.
Chang, C. T., Wu, C.-S. C., & Yang, J. T. (1978) *Anal. Biochem. 91*, 13–31.
Chou, K.-C., & Zhang, C.-T. (1992) *Eur. J. Biochem. 207*, 429–433.
Chou, P. Y. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 549–586, Plenum Press, New York.
Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.
Creighton, T. E. (1993) *Proteins: Structure and Molecular Properties*, 2nd ed., pp 282–285, W. H. Freeman and Company, New York.
Deléage, G., & Roux, B. (1987) *Protein Eng. 1*, 289–294.
De Rosier, D. J., Munk, P., & Cox, D. J. (1972) *Anal. Biochem. 50*, 139–153.
Garnier, J., Osguthorpe, J. D., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.
Giles, A. M., Imhoff, J. M., & Keil, B. (1979) *J. Biol. Chem. 254*, 1462–1468.
Holley, L. W., & Karplus, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 152–156.
Khatri, G. S., MacAllister, T., Sista, P. R., & Bastia, D. (1989) *Cell 59*, 667–674.
Kralicek, A. V., Day, A. J., Wake, R. G., & King, G. F. (1991) *Biochem. J. 275*, 823.
Kuempel, P. L., Pelletier, A. J., & Hill, T. M. (1989) *Cell 59*, 581–583.
Lee, E. H., Kornberg, A., Hidaka, M., Kobayashi, T., & Horiuchi, T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9104–9108.
Lewis, P. J., & Wake, R. G. (1991) *Rev. Microbiol. 142*, 893–900.
Lewis, P. J., Smith, M. T., & Wake, R. G. (1989) *J. Bacteriol. 171*, 3564–3567.
Lewis, P. J., Ralston, G. B., Christopherson, R. I., & Wake, R. G. (1990) *J. Mol. Biol. 241*, 73–83.
Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1939–1941.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Mehta, P. P., Bussiere, D. E., Hoffman, D. W., Bastia, D., & White, S. W. (1992) *J. Biol. Chem. 267*, 18885–18889.
Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608–613.
Milthorpe, B. K., Jeffery, P. D., & Nichol, L. W. (1975) *Biophys. Chem. 3*, 169–176.
Mitchell, W. M. (1968) *Science 162*, 374–375.
Mitchell, W. M., & Harrington, W. F. (1970) *Methods Enzymol. 19*, 635–642.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44–73.
Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., & Campbell, I. D. (1990) *J. Magn. Reson. 87*, 488–501.
Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature 298*, 718–723.
Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmostein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321–329.
Perczel, A., Park, K., & Fasman, G. D. (1991) *Anal. Biochem. 203*, 83–93.
Perkins, S. J. (1986) *Eur. J. Biochem. 157*, 169–180.
Rafferty, J. B., Somers, W. S., Saint-Girons, I., & Phillips, S. E. V. (1989) *Nature 341*, 705–710.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368–379.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.
Smith, M. T., & Wake, R. G. (1992) *J. Mol. Biol. 170*, 648–657.
Somers, W. S., & Phillips, S. E. V. (1992) *Nature 359*, 387–393.
Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205–280.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorf, J. W. (1990) *Methods Enzymol. 185*, 60–89.
Suck, D., Lahm, A., & Oefner, C. (1988) *Nature 322*, 465–468.
Taylor, W. R., & Thornton, J. M. (1984) *J. Mol. Biol. 173*, 487–514.
Teller, D. C. (1973) *Methods Enzymol. 27*, 346–441.
Yang, J. T., Wu, C.-S. C., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208–269.