Molecular Shapes of Transcription Factors TFIIB and VP16 in Solution: Implications for Recognition†

J. Günter Grossmann,‡ Andrew J. Sharff,§,|| Peter O'Hare,⊥ and Ben Luisi*,§

Synchrotron Radiation Department, CLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, U.K., Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, U.K., and Marie Curie Institute, The Chart, Oxted RH8 OTL, U.K.

Received December 20, 2000; Revised Manuscript Received March 20, 2001

ABSTRACT: The molecular shapes of transcription factors TFIIB and VP16 have been studied by small-angle X-ray scattering (SAXS). We interpret the shapes and discuss the implications for the specific recruitment of these proteins into regulatory assemblies. Human transcription factor TFIIB, a universal component of the transcription preinitiation complex, has a triangular form resulting from intramolecular associations between its two principal structural domains. A segment linking the two domains appears to be conformationally flexible. The solution shape of TFIIB can be well fitted with the crystal structure of the DNA-bound C-terminal domain together with the NMR structure of the N-terminal domain; however, the shape cannot accommodate the NMR structure of the isolated C-terminal domain. We discuss how the conformational differences between the solution structures of the isolated C-terminal domain and the intact protein might result from interdomain allostery. Docking the SAXS shape of intact TFIIB into the preinitiation complex suggests that the flexible linker region may contact the 3′ flanking region of the TATA element in the major groove. Transcription rates can be enhanced by activator proteins, and the classical example is the herpes simplex virus factor VP16 (α-TIF), which associates with cellular transcription factors, including TFIIB. The shape reconstruction of VP16 from its SAXS profile reveals a globular structural core that can be well modeled by the crystal structure of a conserved, central region of the protein. However, the carboxy terminus extends from this core and is essentially disordered. As it makes defined protein−protein interactions in the activation complex, the flexible segment is likely to condense upon assembly with its partners.

In eukaryotes and in archaebacteria, the site of transcription initiation is defined by the activities of complexes of regulatory proteins that assemble on promoter elements. These assemblies, known as preinitiation complexes, recruit and align the polymerase to transcribe the neighboring gene. The preinitiation complex for genes transcribed by polymerase II forms on a defined DNA site, the TATA element, and characteristically includes a number of general transcription factors. One of these, TFIIB, plays a role in defining the direction of transcription (*1*, *2*) and indirectly recruits polymerase. Not surprisingly, TFIIB is highly conserved in a wide diversity of organisms.

Human TFIIB is a 316-residue polypeptide and, in common with its homologues from other species, has two functionally distinct regions (Figure 1). The C-terminal domain contacts the DNA and the TATA-binding protein. Structurally, it comprises a direct repeat of two 10 kDa fragments that closely resemble the cell-cycle regulatory protein cyclin A (*3*, *4*). Preceding the C-terminal domain is a zinc-mediated fold that may interact indirectly with the

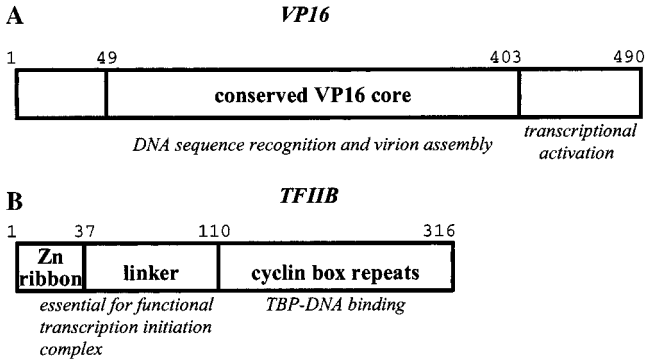

FIGURE 1: Schematic representation of the functional domains of TFIIB and VP16. The principal functional and structural domains are indicated. (A) VP16 from herpes simplex virus 1, showing the core globular region that is conserved among the herpes virus (bovine, equine, and human viruses HSV-1 and HSV-2). The crystal structure has been determined for this common core (*13*). The activation domain, which is enriched with acidic residues, is also denoted. (B) TFIIB, showing the putative zinc-ribbon domain, a flexible linker segment, and the cyclin A-related repeats. An NMR structure of the Zn ribbon is available for an archeon homologue (*P. furiosus*) (*6*), and the NMR structure is available for the repeat domain (*3*). The crystal structure of the ternary complex of TFIIB's cyclin A repeats, together with TBP on a TATA element, has also been determined (*4*, *7*).

† This study was supported by the Wellcome Trust.

* To whom correspondence should be addressed. E-mail: ben@cryst.bioc.cam.ac.uk. Fax: 44-1223-766002. Telephone: 44-1223-766019.

‡ CLRC Daresbury Laboratory.

§ University of Cambridge.

|| Current address: Astex Technology Ltd., Cambridge Science Park, Cambridge, U.K.

⊥ Marie Curie Institute.

10.1021/bi0028946 CCC: $20.00
Published on Web 05/04/2001

polymerase (*5*) (Figure 1). The structure of a portion of the N-terminal domain has been characterized for the protein from the archeon *Pyrococcus furiosus* by NMR (*6*). This segment forms an antiparallel Zn-ribbon.

The C-terminal cyclin-like domain of TFIIB contacts the DNA on the 5′ side of the TATA element with respect to the transcription start site through the domain's helix−turn−helix motif (*7*). The sequence at this site, known as the BRE[1] (for TFB recognition element), is found in a subset of TATA elements, and the specific protein−DNA interactions made here define the orientation of the complex and affect the direction of transcription for certain promoters (*2*, *8*). Molecular genetics studies have shown that the N-terminal domain interacts with the C-terminal domain. The interdomain association may be functionally important for interaction with transcription activators, such as VP16, and with the TBP−DNA complex (*9*−*12*).

The activity of the polymerase may be enhanced or suppressed by certain protein assemblies that bind specific sites which may be close to the promoter DNA in some cases but linearly distant in others. Examples are the complex formed by the virion protein VP16 from the herpes simplex virus. This protein is at the core of an assembly comprising DNA-binding transcription factor Oct-1, the human cellular factor (HCF-1), and defined target sites upstream of the early-intermediate genes in the herpes simplex virus genome. VP16 is found among herpes viruses HSV-1 and HSV-2 and equine and bovine herpes viruses. All these proteins share a common core region, and this region (corresponding to residues 47−402) has been structurally elucidated by X-ray crystallography for the HSV-1 protein (*13*).

In HSV-1, the VP16 protein is 490 residues in length (Figure 1). Outside the conserved core is a section that, in HSV-1 and HSV-2 VP16, acts as an activation domain. This domain has been shown to interact with TFIIB, and one proposed role for this domain is to catalyze preinitiation complex assembly (*11*, *14*).

We describe here the overexpression and purification procedures of the full-length HSV-1 VP16 and human TFIIB proteins. The procedures yield tens of milligrams of highly soluble purified protein, and this has allowed us to obtain shape information for the two proteins in solution using small-angle X-ray scattering (SAXS) experiments with synchrotron radiation. This technique provides a means of examining molecular interactions and conformational flexibility in solution (see refs *15* and *16* and references therein). Solution scattering in combination with high-resolution structural data can provide models for the interaction of individual components in molecular assemblies. The SAXS method is suitable for studying macromolecules and complexes with a wide range of molecular weights and dimensions (∼10−1000 Å).

The molecular shape of intact VP16 in solution was found to have characteristic features most recently deduced from the crystal structure of its conserved core (*13*). Shape analysis of VP16 also indicates that it has significant segmental flexibility, which is restricted to defined amino acid sections that are involved in protein−protein recognition. These observations are in agreement with NMR studies of the isolated activation domain (*17*) and suggest that the domain is segmentally mobile in the context of the full-length protein.

The shape of TFIIB is consistent with an intramolecular interaction between the N- and C-terminal domains. Such an interaction had been predicted on the basis of mutational effects on transcription and on protein−protein interaction assays (*9*, *12*). This shape can be fitted with the crystal structure of TFIIB in the ternary complex with TBP and the TATA element which includes the 5′ BRE that defines the direction of transcription (*7*). Surprisingly, the shape is not in good agreement with the NMR structure of the free C-terminal domain, where the cyclin A repeats pack more compactly. One interpretation of the SAXS, NMR, and mutagenesis data, taken together, is that the intramolecular interaction between the N- and C-terminal domains helps to appropriately present the molecule's recognition surfaces for ternary complex formation. It also suggests that any interaction that disrupts the intramolecular contact, such as interaction with VP16, could trigger allosteric transitions. In the context of the preinitiation complex, such an allosteric transition might affect the dissociation rates of TFIIB or trigger a change in promoter conformation.

MATERIALS AND METHODS

Preparation of TFIIB. Full-length human TFIIB was cloned into plasmid pGEX4T as previously described (*18*). This construct generates a fusion protein with glutathione *S*-transferase (GST) and a thrombin cleavage site preceding the N-terminus of TFIIB. For expression, the construct was transformed into *Escherichia coli* strain B834(DE3), which is a methionine auxiotroph. Cultures were grown in 2×YT medium at 30 °C and induced with 0.1 mM IPTG for approximately 4 h.

After being washed, the harvested cells were resuspended in a phosphate-buffered saline (PBS)/1% Triton mixture containing a cocktail of protease inhibitors. The cells were lysed in a French press, and the cell debris was removed by centrifugation. Polyethyleneimine was added to a final concentration of 0.1% to precipitate nucleic acid, which was then removed by centrifugation.

Prewashed glutathione Sepharose was added batchwise to the lysate and mixed gently for 1 h. The gel was then washed extensively with 50 mM Tris (pH 7.5), 0.3 M NaCl, 1 mM DTT, 1 mM sodium azide, and 10 μM zinc acetate, followed by a wash in the same buffer with only 0.15 M NaCl and 2.5 mM $CaCl_2$. For a 5 L culture, 100 units of thrombin was added to the gel and incubated overnight at 4 °C with gentle agitation. Benzamidine was then added to inhibit the thrombin. The gel was poured onto a column and washed with 50 mM Tris (pH 7.5), 0.1 M NaCl, and 1 mM sodium azide to recover TFIIB.

The eluant was loaded onto an S-Sepharose cation exchange column and eluted with a 0.1 to 1.0 M NaCl gradient. TFIIB eluted at ∼0.35 M NaCl and was determined to be pure by SDS−PAGE and amino acid composition analysis. The purified protein was shown to contain approximately stoichiometric quantities of zinc (1:1 Zn:protein ratio) by atomic absorption spectroscopy.

Preparation of VP16. Full-length HSV-1 VP16 was cloned into plasmid pGEX2T and expressed in *E. coli* strain B834-(DE3). Cultures were grown at 37 °C and induced with 0.1

[1] Abbreviations: SAXS, small-angle X-ray scattering; TBP, TATA-binding protein; HSV, herpes simplex virus; BRE, TFB recognition element; CD, circular dichroism.

mM IPTG. The cells were washed and lysed in 50 mM Tris (pH 7.5), 0.1 M NaCl, and 1 mM DTT. Glutathione Sepharose was added batchwise, as described for the preparation of TFIIB. VP16 was released from the glutathione Sepharose by treating the washed resin with thrombin. The liberated VP16 was washed from the matrix, then directly loaded onto a Q-Sepharose anion exchange column, and eluted with a 0.1 to 1.0 M NaCl gradient. VP16 eluted at ~0.3 M NaCl and was determined to be pure by SDS−PAGE, electrospray mass spectrometry, and amino acid composition analysis.

Far-UV Circular Dichroism. Circular dichroism data were collected at station 3.1 at the Daresbury Synchrotron Radiation Source using a 0.01 mm path length quartz cuvette. Data were collected from 170 to 260 nm. The VP16 sample was at a concentration of 8.5 mg/mL and the TFIIB sample at 1 mg/mL in buffer composed of 50 mM NaCl and 5 mM DTT. The spectra were background corrected, scaled according to the protein concentration, and processed to estimate the secondary structure components using the program SELCON (http://srs.dl.ac.uk/vuv/cd/selcon.html).

Synchrotron X-ray Scattering Experiments and Data Analysis. X-ray solution scattering data were collected with the low-angle scattering camera at station 2.1 (*19*) at the Daresbury Synchrotron Radiation Source using a position-sensitive multiwire proportional counter (*20*). At a sample-to-detector distance of 2 m (5.7 m) and an X-ray wavelength λ of 1.54 Å, a momentum transfer q interval of 0.04−0.35 Å^{-1} (0.01 Å^{-1} ≤ q ≤ 0.20 Å^{-1}) was covered. The modulus of the momentum transfer is defined as $q = 4\pi \sin \Theta/\lambda$, where 2Θ is the scattering angle. The q range was calibrated using an oriented specimen of wet rat tail collagen (based on a diffraction spacing of 670 Å). All samples were held in a temperature-controlled brass cell containing a Teflon ring sandwiched by two mica windows. The cell has a sample volume of 120 μL and a thickness of 1.5 mm. Buffer and sample were assessed in alteration, each in frames of 10−60 s (depending on the radiation sensitivity of the proteins). Scattering profiles were corrected for background scattering (subtraction of the scattering from the camera and a cell filled with buffer), sample transmission and concentration, and positional nonlinearities of the detector. Data reduction was performed with the OTOKO software package (*21*), modified, and adapted for the Daresbury NCD software suite.

For SAXS data collection, VP16 was prepared in 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 5 mM 2-mercaptoethanol, and 0.1% Genapol X-100 (Boehringer-Mannheim). TFIIB was in 50 mM Tris-HCl (pH 7.5) and 100 mM NaCl. VP16 and TFIIB samples were assessed at 6 °C with protein concentrations in the range between 0.1 and 3 mg/mL. To assess protein aggregation and the contribution of potentially unstructured polypeptide segments, scattering data were collected at a short (2 m) as well as a long (5.7 m) sample-to-detector distance. Scattering data analysis for TFIIB and VP16 followed standard procedures (*22*). Using the indirect Fourier transform method as implemented in the program GNOM (*23*), the radius of gyration R_g, the forward scattering intensity I_0, and the intraparticle distance distribution function $p(r)$ were extracted from the experimental scattering data. The calculation of molecular volumes was based upon Porod's invariant (*24*), bearing in mind the limited experimental scattering interval (*25*). The multipole expansion method proposed by Stuhrmann (*26*) and developed by Svergun and co-workers (*27*−*29*) was applied to restore the molecular shape of both transcription factor proteins. The smoothed scattering profiles were fitted ab initio by the scattering from an envelope function beginning with an ellipsoid as an initial approximation. The calculations provided consistent and stable shape boundaries as judged from several runs with altered starting conditions. The final protein shapes were described by spherical harmonics up to fourth-order assuming no molecular symmetry (i.e., 19 free parameters). This treatment is adequate in view of the negligible changes compared to envelopes determined for third-order harmonics (10 free parameters) and with respect to the information content of the experimentally available data range.

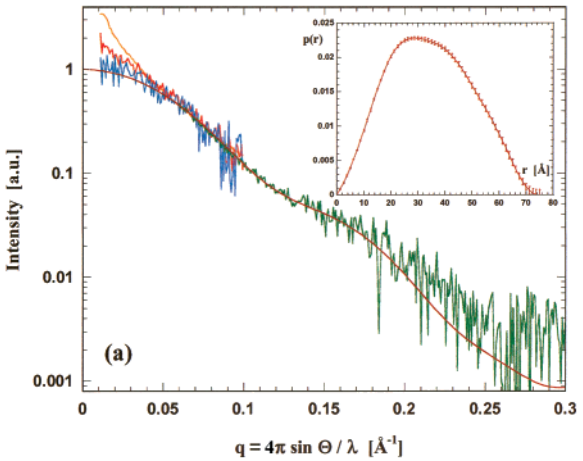

FIGURE 2: Small-angle X-ray solution scattering of full-length human TFIIB. The scattering profiles were recorded at a protein concentration of 0.75 mg/mL. The very small angle region (q < 0.07 Å^{-1}) indicates radiation-induced aggregation with time [portions of the low-angle scattering data are given after exposure to X-rays for 20 s (blue), 100 s (red), and 20 min (orange)]. The scattering data (green) used for shape analysis and extending to higher scattering angles represent data from eight freshly prepared samples. The calculated distance distribution function $p(r)$ is given in the inset. The deviation of the smooth curve (which represents the calculated scattering from the restored shape) from the experimental data for q > 0.2 Å^{-1} is the result of protein inhomogeneity which is contributing to the shape scattering at higher scattering angles (this is also observed in the case of VP16; see Figure 4).

RESULTS

TFIIB. TFIIB was sensitive to radiation-induced aggregation despite cooling the sample at 6 °C (the aggregates were most likely in the form of deposits on the sample cell window). Due to the radiation sensitivity, only concentrations of <1 mg/mL have been measured and the radiation exposure was limited to less than 30 s for data included in the analysis of the small-angle region. To improve the signal-to-noise ratio, data were accumulated from a series of freshly prepared samples (Figure 2). Extrapolation to the intensity at zero scattering angle (I_0) and comparison with that of bovine serum albumin at a known concentration as a molecular weight standard show that TFIIB is monomeric in solution. The conformational flexibility of full-length TFIIB (35 kDa) in the low concentration regime appears to be less significant (as can be seen from the very low angle region), and the analysis of the full scattering profile yields the following

values: $R_g = 25.0$ Å (±2%), $D_{max} = 75$ Å (±4%), and $V_{TFIIB} = 60\,000$ Å^3 (±5%). This volume is in good agreement with the known protein mass. The distance distribution function $p(r)$ for TFIIB is shown in the inset of Figure 2.

The molecular envelope of TFIIB is presented in stereoscopic view in Figure 3A. The shape suggests that full-length TFIIB comprises three subdomains which are arranged in a V-shaped pattern. This shape requires that the subdomains form a stable association, which is consistent with observations from molecular genetics (*9*) and from NMR (*11*). However, extended conformations (contributing to less than 10%) cannot be excluded with this SAXS analysis, and as we will discuss later, it is anticipated that a portion of the linker region between the subdomains is highly mobile.

One of the subdomains could be fitted with the NMR structure of the N-terminal zinc-ribbon domain of *P. furiosus* TFIIB (*6*). Our full-length TFIIB sample was found to contain Zn in a ~1:1 stoichiometry, which is in accord with the Zn-ribbon model. The shapes of the other two subdomains can accommodate the crystal structure of the C-terminal domain of TFIIB (TFIIBcore) as it is found in a ternary complex with the conserved core of TBP (TBPcore) and a TATA-containing oligonucleotide (*7*).

A NMR structure is also available for the free C-terminal domain of human TFIIB (*3*), and this differs from the DNA-bound form in that the cyclin A subdomains are more compactly associated. Intriguingly, the NMR structure of the isolated C-terminal domain does not fit as well in the calculated molecular shape of full-length TFIIB (Figure 3B). The structure of the DNA-bound form provides a superior fit.

Several criteria were taken into account in modeling the atomic structures of the separate N- and C-terminal domains into the molecular boundaries of the full-length protein. First, the conformation of the C-terminus could only be fitted into the shape in one orientation that would keep the DNA reading surfaces exposed. Second, the domain termini were kept in proximity, although it must be noted that the linker of approximately 60 residues, between the Zn ribbon and the first of the cyclin repeats, is missing from this model. We will return to the role of this linker in the Discussion. Third, the modeling was guided by knowledge of the contact regions between TFIIBn and TFIIBc. These are regions 24−61 in the N-terminal domain, which were reported to associate with the second cyclin repeat (*9*). The observations of Roberts and Green (*9*) are consistent with a discontinuous recognition surface, and our model requires that the segment of residues 24−61 contacts both the cyclin repeats. Last, conflicting overlaps were avoided between the Zn-ribbon motif and the DNA-binding region on the surface of the C-terminal domain (Figure 3A). For the constructed model, the structural parameters are calculated to be as follows: $R_g = 22.2$ Å, $D_{max} = 72$ Å, and $V = 50\,000$ Å^3. These values essentially agree with the experimental figures, keeping in mind that the calculated values are based on the unsolvated TFIIB model and exclude the linker peptide segment of 59 residues.

We evaluated the secondary structure content of our TFIIB sample using far-UV circular dichroism. The secondary structure composition from three separate runs at three different concentrations ranged from 68 to 72% α-helix, from 7 to 8% β-sheet, and from 21 to 24% turn and other. The C-terminal domain, residues 107−317, comprises 60% α-helix in the crystal structure (*7*). The CD composition would require that the linker component between the Zn ribbon and the cyclin repeats is almost entirely α-helix, but consensus secondary structure prediction suggests that it has a short helix and two β-strands. This portion was not modeled, and it is possible that it does have defined secondary structure. As it might move freely with respect to the remainder of the protein, only its average spatial conformation would contribute to the overall dimension of the shape.

It is possible that alternative assignments can be made for the individual subdomains (cyclin repeats, Zn ribbon, and the interdomain linker) into the shape features. However, alternative models are not consistent with the fitting criterion described above. We also note that the shape fits well onto the modeled ternary complex. Using the cyclin repeats as a reference, we docked the molecular shape and model of full-length TFIIB onto the crystal structure of Tsai and Sigler (*7*) of the TBPcore−TFIIBcore complex on a TATA element (Figure 3C). In this model, we have extended the DNA from the 5′ and 3′ directions of the TATA element using idealized B-DNA. Remarkably, the SAXS model fits compactly into the assembly.

As viewed from the approximate 2-fold symmetry axis of the TATA element and the TBP protein, the trajectory of the DNA takes different directions on the 5′ and 3′ flanking regions of the TATA element (Figure 3C). This asymmetry arises from the contacts made by the C-terminal cyclin unit of TFIIB with the BRE element. The protruding surface of the TFIIB N-terminal domain is on the correct side of the asymmetric joint to make a complementary match to the DNA.

The model predicts protein−DNA interactions on the 5′ side of the TATA element, in the vicinity of the BRE recognition element and centered 5 bp preceding it. These contacts would be made by the Zn-ribbon domain and would probably be entirely with the phosphate backbone. Additional contacts may also be made on the 3′ side of the TATA element by the first α-helix of the first cyclin unit together with residues in the linker region following the Zn-ribbon domain. The proposed contacts would occur in the major groove in the region corresponding to bases −22 to −18 upstream of the start site. These contacts are consistent with footprinting experiments showing major groove protection at these sites (S. Roberts, personal communication).

VP16. The small-angle X-ray scattering profile for VP16 in solution is shown in Figure 4 with the low-angle region highlighting the concentration dependence and significant contributions from mobile, unstructured chain segments in VP16. Even at the lowest concentration that was measured (0.13 mg/mL), where the contribution from possible protein aggregation is expected to be very small, the scattering profile still shows a significant upturn in intensity (for $q < 0.05$ Å^{-1}). Since neither radiation sensitivity nor protein impurity is responsible for this behavior, only highly mobile, random chainlike elements are consistent with such a scattering signal. However, they are expected to add only to the overall dimension of the particle's appearance but not to affect its molecular core structure, an effect similar to that of carbohydrate chains on the surface of a glycoprotein. Thus, it was assumed that the scattering contribution of the flexible polypetide segments becomes negligible when $q > 0.05$ Å^{-1} (which was acceptable in view of the subsequent results).

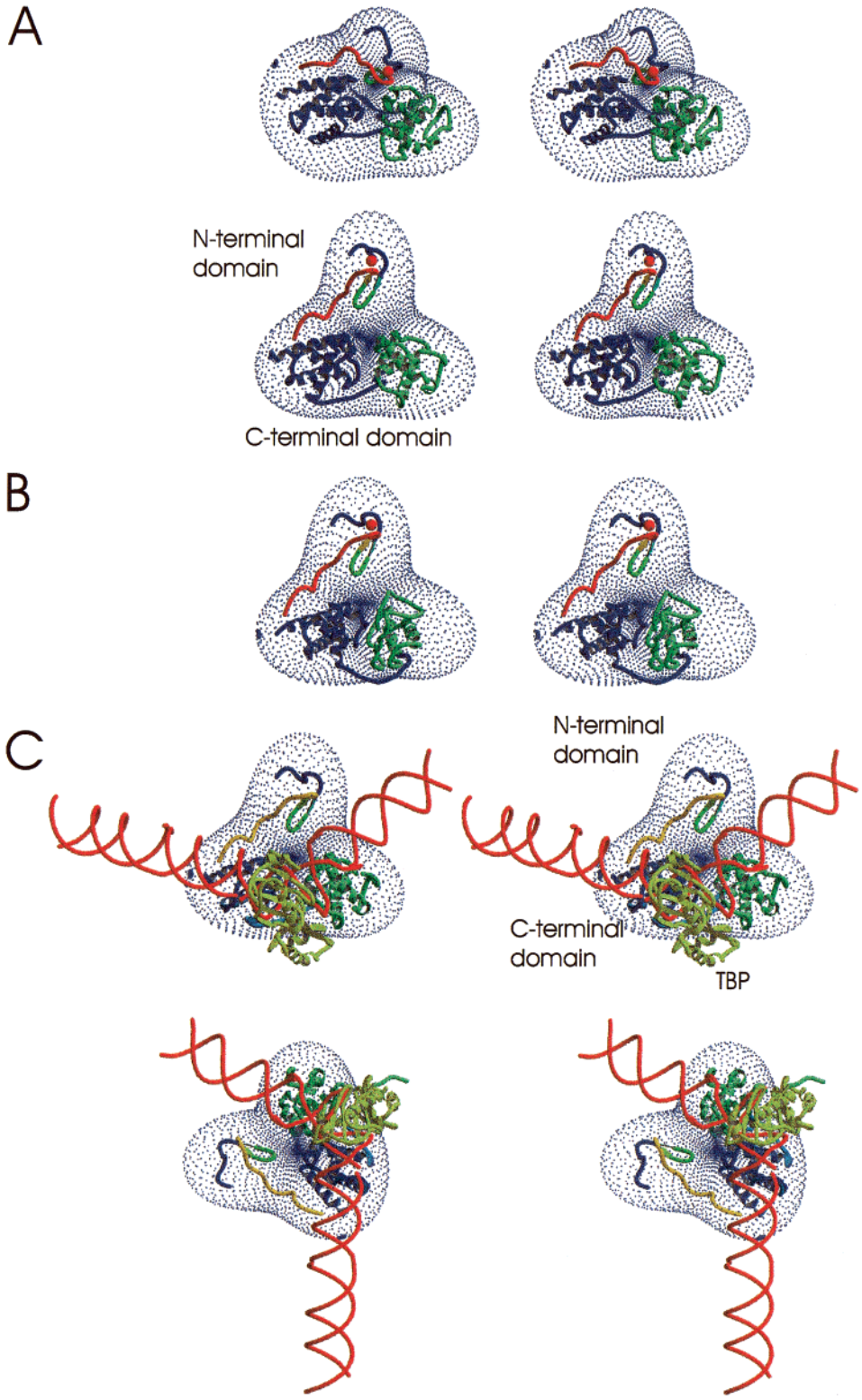

FIGURE 3: Molecular shape of TFIIB. (A) Two stereoscopic views of the low-resolution envelope of full-length TFIIB reconstructed using fourth-order harmonics. Superimposed are the NMR-derived N-terminal Zn-ribbon domain from *P. furiosus* (*6*) and the C-terminal repeat cyclin motifs from the X-ray structure taken from the ternary complex of the human TFIIB C-terminal domain with TBPcore on the TATA DNA element (*7*). The relative orientation of these two structures has been arranged manually to match the available envelope space. The zinc ion is represented by a sphere. The first of the cyclin motifs is in blue, and the second is in green. (B) The more compact NMR structure of the unbound TFIIB C-terminal domain (*3*) was also used for the modeling but leaves behind large gaps in the available envelope. (C) The model of the ternary complex of the full-length TFIIB−TBP complex and the TATA element, based on the SAXS model and the crystal structure of the TFIIBcore−TBPcore−TATA complex (*7*). The DNA in the crystal structure was extended on the 5′ side of the BRE and on the 3′ side of the TATA element using B-DNA models. This figure was prepared using MOLSCRIPT (*33*) and Raster3D (*34*).

The profile from this scattering angle onward still allowed a reliable data analysis with the $p(r)$ function vanishing at $r = D_{max} = 94 \pm 4$ Å (see the inset of Figure 4) and yielding a value for R_g of 28.7 ± 0.5 Å and a particle volume V_{VP16}

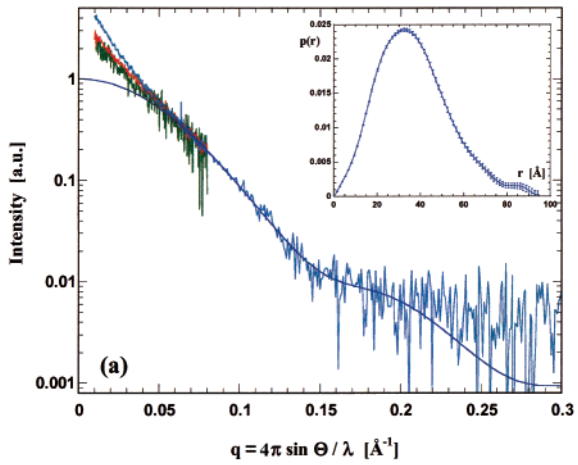

FIGURE 4: X-ray solution scattering of the herpes simplex virus 1 VP16 protein. Scattering profiles obtained for protein concentrations of 0.13 (green), 0.3 (red), and 1.3 mg/mL (blue). The latter was also recorded at the short sample-to-detector distance. When $q < 0.05$ Å^{-1}, the intensity increases noticeably toward zero scattering angle despite the use of a very dilute sample, indicating the presence of less-structured, nonglobular chain segments. The distance distribution function $p(r)$ calculated from the profile without the very small angle scattering contribution is shown as an inset. The smooth curve corresponds to the theoretical scattering from the restored shape.

of 115 000 ± 6000 Å^3. Moreover, the contribution of highly flexible polypeptide chain segments to the molecular shape is now restricted to those areas in which they are anchored to the protein's globular core. The analysis of the intensity of the forward scattered beam (I_0), including the low-angle signal (as recorded with the 5.7 m camera), shows that the VP16 sample is nonaggregated and that the protein is monomeric in solution, which is consistent with results from size exclusion chromatography (results not shown).

The envelope of VP16 determined ab initio from the scattering profile alone (see Figure 5) is reminiscent of the seatlike protein shape as reported in the crystal structure of the conserved central domain (*13*). The ribbon representation of this core domain has been superimposed on the molecular envelope to emphasize the agreement (Figure 5). Two areas to the left and right (which can be labeled as "armrests") project sideways from the "seat". These have no counterparts in the high-resolution structure. Interestingly, these two regions correlate well with the space most likely taken up by N-terminal residues 1−48 which were not included in the conserved VP16 core structure, and an apparent disordered segment in the crystal (residues 350−394; the boundaries correspond to the termini of the orange helix and strand, respectively). This segment is critical for VP16-induced complex assembly (*13*). In contrast, the area that the remaining C-terminal "domain" (residues 403−490) is anticipated to occupy (corresponding to the bottom part of the shapes in Figure 5) only shows a small spatial extension to the globular core; i.e., it would not be possible to accommodate all C-terminal residues within the envelope. It seems likely that these residues are responsible for the intensity increase (and indicative of random chainlike behavior) observed in the very low-angle scattering regime. Due to the precondition of molecular envelope reconstruction (i.e., the shape calculation assumes a globular, uniform particle), this scattering contribution was not included in the analysis. Thus, we conclude that the C-terminal activation

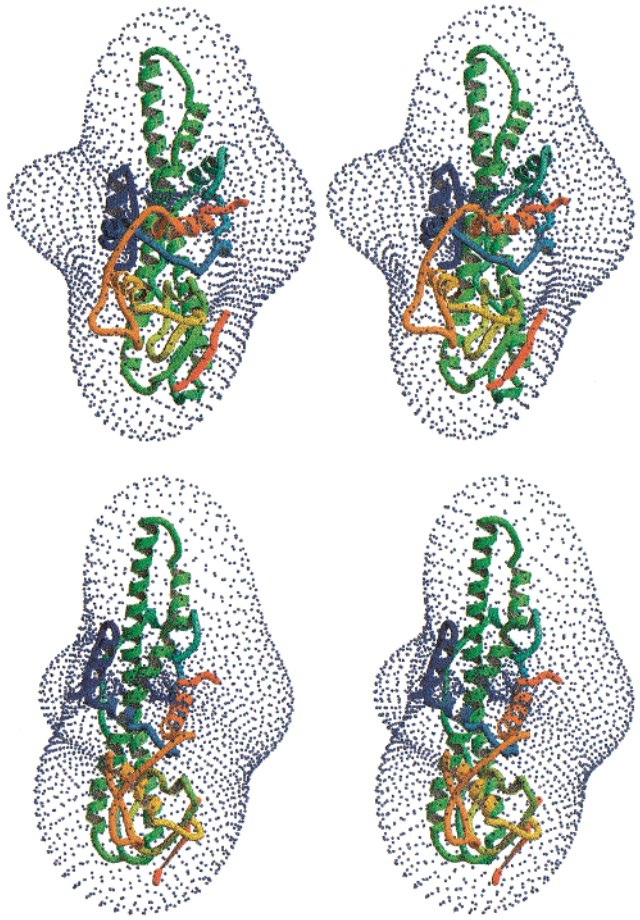

FIGURE 5: Molecular shape of VP16. Two orientations of the molecular shape of VP16 derived from the scattering data. The crystal structure of the central region of VP16 (*13*) is superimposed in the molecular shape. In the crystal structure, the peptide is disordered between the orange helix and orange strand (residues 350−394). A peptide segment might be accommodated in the space of the right "elbow" of the armchair-like molecular shape. The N-terminal residue of the orange strand is residue 395, and residue 349 is at the C-terminus of the helix. Residue 402 is at the C-terminus of the strand, and there appears to be no density to accommodate an ordered peptide corresponding to residues 403−490. The elbow on the left might accommodate a portion of amino terminal residues 1−48. This figure was prepared using MOLSCRIPT (*33*) and Raster3D (*34*).

domain is disordered and particularly nonglobular in its native solution state and does not contribute to the globular core but extends into the solvent. For comparison, the volume of the available crystal structure of the VP16 core structure (which includes only residues 47−349 and 395−402, i.e., 311 residues, corresponding to 63% of full-length VP16) is 58 000 Å^3. The calculation is based on the accessible surface area and does not include a hydration layer. This value corresponds to half the experimental value for the full-length protein in solution. The radius of gyration (22.7 Å) and maximum particle dimension (88 Å) of the crystallographic core structure also reflect the partial structure when compared to the experimental data of full-length VP16 in solution. However, it underlines our conclusion that mobile segments of the protein (e.g., disordered segments as well as omitted chain sections in the structural analysis of the crystalline state) expand the molecular boundary significantly.

We also evaluated the secondary structure of full-length VP16 by far-UV circular dichroism. Spectral deconvolution

suggests the following secondary structure composition: 62−69% α-helix, 0−10% β-sheet, and 20−33% other (data not shown). This is broadly consistent with the secondary structure content of the model based on the conserved core.

DISCUSSION

The purified recombinant VP16 protein analyzed here has been shown to be capable of stimulating transcription in vitro with Oct-1 from an octamer element (the ICPO promoter; L. Lee et al., in press in *Biochemistry*). While the recombinant VP16 is functionally competent to activate transcription, our SAXS data suggest that its activation domain does not assume a unique globular shape but rather behaves like a very flexible polypeptide chain with random conformations. In isolation, the activation domain has been shown to be flexible (*17*), and the SAXS shape suggests that the domain is likely to remain disordered (or unstructured) in the full-length protein. In our modeling of the molecular shape with the VP16 core crystal structure, the C-terminal residue (amino acid 402) is at the bottom of the shape shown in Figure 5 (the orange strand). There is no space here that could accommodate residues 403−490 if they were structured and associated with the surface of the core. Thus, as suggested previously, the acidic activation domain of VP16 may become structured when binding to its target protein (*30*).

The protruding left elbow of the VP16 shape (Figure 5) may correspond to portions of the N-terminus, as it is close to residue 47 of the core domain. The right elbow is near the segment between residues 349 and 395 which are poorly ordered in the crystal structure. This segment corresponds to an exposed region near residue 370 that is hypersensitive to proteases (*18*).

Our SAXS results for full-length TFIIB suggest that the Zn-ribbon domain at the N-terminus makes an intramolecular interaction with the C-terminal domain. The proximity of the N- and C-terminal domains is consistent with observations from NMR (*11*) and molecular genetics studies (*9*, *10*, *14*). The relative arrangement of these domains is modulated by the presence of the N-terminal zinc-binding domain. This latter domain therefore plays an important functional role in positioning and orienting the C-terminal domains which so far has not fully been appreciated due to a lack of a structure for the full-length TFIIB molecule.

Between the zinc ribbon at the N-terminus and the cyclin repeats at the C-terminus is a linker that has little predicted secondary structure. However, a 49-residue segment here has the highest degree of conservation in the TFIIB family (*31*). This linker has been shown to be functionally important for affecting the transcription start site (*10*, *32*). We speculate that the intramolecular interactions noted here may serve to present this linker for stereochemically defined interaction in the ternary complex with either the DNA or the polymerase.

The structure of the C-terminal domain of TFIIB has been elucidated in the free state (by NMR) and in the ternary complex on DNA with TATA-binding protein (by X-ray crystallography). The NMR structure has a different arrangement of cyclin A repeats, which form a more compact structure compared to the DNA-bound form (*3*). Intriguingly, the molecular shape obtained by SAXS for the full-length protein is in better agreement with the structure of the DNA-bound form, rather than the isolated form (Figure 3). One possible interpretation of this conformational difference is that the N-terminal Zn-ribbon domain, in interacting with the C-terminal domain, may preorganize the DNA recognition surface of the molecule for docking onto the target. It also suggests that the disruption of the intramolecular interaction between the N- and C-terminal domains may instigate an allosteric change in the C-terminal domain. The N-terminal domain interacts with transcriptional activators (*9*) and with polymerase II, via TFIIF (*5*). Perhaps these interactions activate such an allosteric change in the molecule. In the context of the preinitiation complex, an allosteric change could cause either the TFIIB to jettison from the assembly or a change in the superhelical tension of the DNA between the TATA element and the start site. This will remain speculation until the ternary complex of full-length TFIIB with TBP and the TATA element is structurally elucidated. Studies are in progress to test the inter- and intramolecular interactions in the ternary complex.

ACKNOWLEDGMENT

We are very grateful to Stefan Roberts and Tali Haran for many stimulating discussions and helpful comments. We thank Dr. C. Hill for oligonucleotide synthesis, M. Weldon for protein N-terminal sequencing, P. Sharratt for amino acid composition analysis, and J. Lester for DNA sequencing. We thank David Clarke and the staff of the Daresbury Synchrotron Radiation Source for help collecting the circular dichroism data.

REFERENCES

1. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. (1998) *Genes Dev. 12*, 34−44.
2. Tsai, F. T. F., Littlefield, O., Kosa, P. F., Cox, J. M., Schepartz, A., and Sigler, P. B. (1998) *Cold Spring Harbor Symp. Quant. Biol. 63*, 53−61.
3. Bagby, S., Kim, S., Maldonado, E., Tong, K. I., Reinberg, D., and Itura, M. (1995) *Cell 82*, 857−867.
4. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995) *Nature 377*, 119−128.
5. Hahn, S., and Roberts, S. (2000) *Genes Dev. 14*, 719−730.
6. Zhu, W., Zeng, Q., Colangelo, C. M., Lewis, M., Summers, M. F., and Scott, R. A. (1996) *Nat. Struct. Biol. 3*, 122−124.
7. Tsai, F. T. F., and Sigler, P. B. (2000) *EMBO J. 19*, 25−36.
8. Bell, S. D., Kosa, P. L., Sigler, P. B., and Jackson, S. P. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 13662−13667.
9. Roberts, S. G., and Green, M. R. (1994) *Nature 371*, 717−720.
10. Bangur, C. S., Pardee, T. S., and Ponticelli, A. S. (1997) *Mol. Cell. Biol. 17*, 6784−6793.
11. Hayashi, F., Ishima, R., Liu, D., Tong, K. I., Kim, S., Reinberg, D., Bagby, S., and Ikura, K. (1998) *Biochemistry 37*, 7941−7951.
12. Hawkes, N. A., Evans, R., and Roberts, S. G. E. (2000) *Curr. Biol. 10*, 273−276.
13. Liu, Y., Gong, W., Huang, C. C., Herr, W., and Cheng, X. (1999) *Genes Dev. 13*, 1692−1703.
14. Roberts, S. G., Ha, I., Maldonado, E., Reinberg, D., and Green, M. R. (1993) *Nature 363*, 741−744.
15. Grossmann, J. G., and Hasnain, S. S. (1997) *J. Appl. Crystallogr. 30*, 770−775.
16. Trewhella, J. (1997) *Curr. Opin. Struct. Biol. 7*, 702−708.
17. O'Hare, P., and Williams, G. (1992) *Biochemistry 31*, 4150−4156.
18. Hayes, N., and O'Hare, P. (1993) *J. Virol. 67*, 852−862.

19. Towns-Andrews, E., Berry, A., Bordas, J., Mant, P. K., Murray, K., Roberts, K., Sumner, I., Worgan, J. S., and Lewis, R. (1989) *Rev. Sci. Instrum. 60*, 2346−2349.
20. Lewis, R. (1994) *J. Synchrotron Radiat. 1*, 43−53.
21. Boulin, C., Kempf, R., Koch, M. H. J., and McLaughlin, S. M. (1986) *Nucl. Instrum. Methods Phys. Res. A249*, 399−407.
22. Grossmann, J. G., Crawley, J. B., Strange, R., Patel, K. J., Murphy, L. M., Neu, M., Evans, R. W., and Hasnain, S. S. (1998) *J. Mol. Biol. 279*, 461−472.
23. Semenyuk, A. V., and Svergun, D. I. (1991) *J. Appl. Crystallogr. 24*, 537−540.
24. Porod, G. (1951) *Kolloidzeitschrift 124*, 83−114.
25. Feigin, L. A., and Svergun, D. I. (1987) *Structure Analysis by Small-Angle X-ray and Neutron Scattering*, Plenum Press, New York and London.
26. Stuhrmann, H. B. (1970) *Acta Crystallogr. A26*, 297−306.
27. Svergun, D. I., and Stuhrmann, H. B. (1991) *Acta Crystallogr. A47*, 736−744.
28. Svergun, D. I., Volkov, V. V., Kozin, M. B., and Stuhrmann, H. B. (1996) *Acta Crystallogr. A52*, 419−426.
29. Svergun, D. I., Volkov, V. V., Kozin, M. B., Stuhrmann, H. B., Barberato, C., and Koch, M. H. J. (1997) *J. Appl. Crystallogr. 30*, 798−802.
30. Uesugi, M., Nyanguile, O., Lu, H., Levine, A. J., and Verdine, G. L. (1997) *Science 277*, 1310−1313.
31. Na, J. G., and Hampsey, M. (1993) *Nucleic Acids Res. 21*, 3413−3417.
32. Hawkes, N. A., and Roberts, S. G. E. (1999) *J. Biol. Chem. 274*, 14337−14343.
33. Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
34. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D50*, 869−873.

BI0028946