Order and Disorder in the Domain Organization of the Plasmid Partition Protein KorB

Karthik Rajasekar, Sidra Tul Muntaha, Jeremy R. H. Tame, Sireesha Kommareddy, Gordon Morris, Christopher W. Wharton, Christopher M. Thomas, Scott A. White, Eva I. Hyde, and David J. Scott

The plasmid partition protein KorB has a dual role: it is essential for the correct segregation of the low copy number broad host range RK2 plasmid while also being an important regulator of transcription. KorB belongs to the ParB family of proteins, and partitioning in RK2 has been studied as a simplified model of bacterial chromosome segregation. Structural information on full-length ParB proteins is limited, mainly due to the inability to grow crystals suitable for diffraction studies. We show, using CD and NMR, that KorB has regions of significant intrinsic disorder and hence it adopts a multiplicity of conformations in solution. The biophysical data are consistent with bioinformatic predictions based on the amino acid sequence that the N-terminal region and also the region between the central DNA-binding domain and the C-terminal dimerization domain are intrinsically disordered. We have used small angle x-ray scattering data to determine the ensemble of solution conformations for KorB and selected deletion mutants, based on models of the known domain structures. This conformational range of KorB is likely to be biologically required for DNA partitioning and for binding to a diverse set of partner proteins.

The partitioning of DNA between daughter cells upon cell division is critical for the survival of all organisms. Partitioning in most bacteria is dependent on two proteins from the ParA/ParB families (1). ParB binds to a specific centromere-like sequence (parS) on the chromosome forming a higher order nucleoprotein complex that is thought to pair the sister chromosomes/plasmids. ParA, an ATPase, binds to ParB and is thought to act as a motor, pulling or pushing two ParB-bound chromosomes apart to different poles of the bacterial cell (2–4).

One of the best characterized ParB proteins is KorB, encoded by the low copy number broad host range plasmid RK2, from the IncP-1 family (2–10). RK2 has been studied for several years as a small model genome and is also important as a carrier of multiple antibiotic resistance genes across a range of bacteria (11–15). KorB acts in conjunction with IncC (7, 16), a protein of the ParA family also encoded on the plasmid, and is more similar to chromosomal ParB homologues than those from P1 or F plasmids. This makes KorB a good candidate for structural studies to understand partitioning of bacterial chromosomes. In addition to its role in partitioning, KorB acts as a global repressor of transcription for at least six operons in RK2, where it also interacts with a number of other proteins (7). KorB binds simultaneously with RNA polymerase at promoters and interacts with it to prevent open complex formation (17). KorB can act upstream or downstream of the promoters and also at a distance (2–4, 9). At all promoters, KorB acts cooperatively with a second repressor, either KorA, TrbA, or KorC, forming a regulatory network that coordinates expression of the operons on the plasmid (3, 6, 7, 16). Therefore, KorB is of interest for understanding not only partitioning, but also for understanding cooperative protein-protein and protein-DNA interactions involved in gene regulation. KorB consists of three distinct segments (7) (Fig. 1) of which only two have been crystallized and their structures determined to atomic resolution. Firstly, the C-terminal 64 amino acid residue dimerization domain (residues 294–358) (18), which is structurally similar to SH3 domains, and consists of β-strands. Secondly, the central domain (residues 137–252), which interacts with operator DNA (19) and is entirely α-helical. The structure of the N-terminal 137 amino acids of KorB is unknown, however the corresponding region of a homologous protein, SpoOJ from Thermus thermophilus, together with its central domain, have been determined (20). It is of mixed secondary structure, and the N-terminal 20 amino acid residues of SpoOJ are not seen in this structure. KorB has an additional N-terminal 20 amino acids compared with SpoOJ; it also contains two additional predicted helices at the N terminus of the central domain, after the helix-turn-helix motif, which are not found in SpoOJ (Fig. 1B) (1). In KorB these helices contact DNA and are postulated to lead to sequence-specific binding. Thus, despite the sequence homology, the overall structure and DNA binding of SpoOJ may differ from that of KorB. To date there is no structural information about a full-length ParB protein.
Using a range of biophysical techniques we show the protein is modular in its domain organization and that the orientations of the domains of KorB within the protein are highly flexible due to zones of intrinsic disorder along the length of the protein. As the first structural description of a full-length ParB protein, we have reconstructed KorB from the known crystal structures using small angle x-ray scattering (SAXS)\(^5\) and calculated an ensemble of conformations that the protein adopts in solution. This flexibility will be important for the protein’s functional role in the cell enabling it to bind at different distances along DNA and to different binding partners (3–6, 9).

**EXPERIMENTAL PROCEDURES**

*Proteins KorB—*Deletion mutants were made by PCR amplification of the appropriate segment of the wild-type korB gene and inserted into the modified pET28a plasmid previously described (7). This yields proteins with an N-terminal 23-amino acid His tag. Proteins were overexpressed and purified on a nickel-agarose column as described (7), followed by gel filtration on a S200 column in 10 mM Tris HCl, pH 7.0, 100 mM NaCl, 0.1 mM EDTA, 10 mg/liter phenylmethylsulfonyl fluoride. They were concentrated to 5–10 mg/ml by ultrafiltration.

*CD—*CD spectra were obtained on a JASCO 715 spectrometer using 2-mm path length cuvettes. The proteins (~0.2 mg/ml final concentration) were dialyzed into 10 mM sodium phosphate buffer, pH 7.0, 20 mM NaClO\(_4\). This allowed the observation of shorter wavelengths, as chloride ions used previously absorb highly below 210 nm. Spectra from 180–260 nm were scanned at 25 °C at a rate of 50 nm/min with a resolution of 0.2 nm. Spectra of buffer in the same cuvettes were taken previously absorb highly below 210 nm. Spectra from 180–260 nm were scanned at 25 °C at a rate of 50 nm/min with a resolution of 0.2 nm. Spectra of buffer in the same cuvettes were taken under identical conditions and subtracted from those of the protein. Thermal denaturation was monitored at 220 nm, using a temperature ramp between 10 ° and 80 °C over ~1 h. Secondary structure content was extracted from CD spectra using the program CDSTR (21, 22) within DICHROWEB (23, 24), using reference sets 4 and 7.

*NMR Spectroscopy—*One-dimensional \(^1\)H NMR spectra were taken at 25 °C on a Bruker 500-MHz NMR machine, using ~0.5 mg protein in 10 mM sodium phosphate buffer, pH 7.0, 100 mM NaCl, 0.1 mM EDTA. Data were collected using the program CDSSTR (21, 22) within DICHROWEB (23, 24), using reference sets 4 and 7.

Two-dimensional \(^{15}\)N-\(^1\)H heteronuclear single quantum coherence spectra of \(^{15}\)N-labeled (N31-C221)KorB and (N3297)KorB were taken at 30 °C on a Bruker 500-MHz NMR machine. WT KorB and (N150)KorB were both \(^{13}\)C/\(^{15}\)N/\(^2\)D-labeled, and transverse relaxation-optimized spectroscopy \(^{15}\)N-\(^1\)H heteronuclear single quantum coherence spectra were taken on a Varian 900-MHz machine.

*Analytical Ultracentrifugation—*All analytical ultracentrifugation data were obtained on a Beckman XL-I using absorbance optics and utilizing a protocol previously defined (25). Proteins (0.1–1 mg/ml) were in 10 mM Tris-HCl, pH 7.0, 100 mM NaCl, 0.1 mM EDTA. Sedimentation velocity experiments were carried out at 40,000 rpm using an AnTi60 rotor, at 20 °C, and at three loading concentrations to check for self-association. Cells were scanned every 10 min at 280 nm. All data were analyzed using SEDFIT (26).

*SEC-MALLS—*Analytical fractionation was carried out using a series of SEC columns TSK G6000PW and TSK G4000PW protected by a similarly packed guard column (Tosoh Bioscience, Tokyo, Japan) with on-line MALLS (Dawn HELIOS II, Wyatt Technology, Santa Barbara, CA) and refractive index (Optilab rEX, Wyatt Technology) detectors. The eluant was Tris- HCl, pH 7.0, 100 mM NaCl, 0.1 mM EDTA, 10 mg/liter phenylmethylsulfonyl fluoride pumped at 0.8 ml min\(^{-1}\) (PU-1580, Jasco, Great Dunmow, UK), and the injected volume was 100 µl (−1.0 × 10\(^{-3}\) g ml\(^{-1}\)) for each sample. Absolute weight average molecular weights (\(M_m\)) were calculated using ASTRA\(^6\) (Version 5.1.9.1) software (Wyatt Technology). Online viscometry measurements were made using the ViscoStar differential pressure viscometer (Wyatt Technology), and data were transformed using the software provided.

*Model of the N-terminal Region of KorB—*The model of the N-terminal region 54–154 of KorB was built using the program CPModels 2.0. (27) based on homology to SpoOJ (20). Molecular mass was determined from the forward scattering intensity of the sample due to photo-damage. Data were reduced to one dimension and displayed using the program PRIMUS (28). Molecular mass was determined from the forward scattering intensity of the sample due to photo-damage. Data were reduced to one dimension and displayed using the program PRIMUS (28).

Molecular mass was determined from the forward scattering intensity of the sample due to photo-damage. Data were reduced to one dimension and displayed using the program PRIMUS (28). Molecular mass was determined from the forward scattering intensity of the sample due to photo-damage. Data were reduced to one dimension and displayed using the program PRIMUS (28). Molecular mass was determined from the forward scattering intensity of the sample due to photo-damage. Data were reduced to one dimension and displayed using the program PRIMUS (28).

\(I(s) = \int P(r) \frac{\sin(2\pi sr)}{2\pi sr} dr\) (Eq. 1)

where \(r\) is the distance between scattering centers in the molecule, \(P(r)\) is the distance distribution function, and

\[ s = \frac{2\sin\theta}{\lambda} \]  
(Eq. 2)

and \(\theta\) is the scattering angle. Distance distribution data were derived from the scattering data using the indirect Fourier transform of Equation 1 where,

\[ P(r) = \frac{1}{\pi} \int \frac{I(s) \sin 2\pi sr}{2\pi sr} ds \]  
(Eq. 3)

using the program GNOM (29).

---

\(^5\) The abbreviations used are: SAXS, small angle x-ray scattering; SEC-MALLS, size exclusion chromatography coupled to multangle laser light scattering; WT, wild type; \(M_m\), weight average molecular weight; PDB, Protein Data Bank.

\(^6\) The abbreviations used are: SAXS, small angle x-ray scattering; SEC-MALLS, size exclusion chromatography coupled to multiangle laser light scattering; WT, wild type; \(M_m\), weight average molecular weight; PDB, Protein Data Bank.
**Intrinsic Disorder in KorB**

For potentially folded domains, data were reconstructed using the program Gasbor (30). Here, a simulated annealing method is used to minimize against $X^2$ discrepancy,

$$
X^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left[ \frac{(c(s_j))I_{DR}(s_j) - c_{exp}(s_j)}{\sigma(s_j)} \right]^2
$$

(Eq. 4)

where $c$ is a scaling function, $I_{DR}$ is the intensity from a bead representing an amino acid, and $\sigma$ is the standard deviation. The process is repeated 10–12 times, with resulting structures averaged.

Conformers for the intrinsically disordered domains were calculated using a bead for each amino acid, and the vector between the center of each bead was weighted by the known calculated using a bead for each amino acid, and the vector across residues 137–154. For WT and (NΔ150)KorB, dimers were constructed using the known structure of the dimerization domain as a template. Random conformers were selected from the pool of 10,000 generated monomer conformers and aligned using the CCP4 module LSQKAB (32). Each constructed dimer was checked for steric clashes; dimers without clashes were then collated until 10,000 models, each with non-symmetry-related monomers, had been constructed. The theoretical intensity of each calculated was used to minimize against $X^2$ discrepancy, as expected, full-length KorB elutes first from the SEC column, and (CA105)KorB last (supplemental Fig. S2), each giving a single peak. The constructs (NΔ297)KorB and (NΔ31-CΔ221)KorB are below the size that can be accurately determined by SEC-MALLS and, as such, were not subjected to this method of analysis. Analysis of the multiangle-dependent light scattering of these peaks showed that all the proteins were monodisperse; as judged by the ratio of the $Z$-averaged and the number averaged molecular masses, $M_z/M_n$ (Table 1). The weight $M_w$ values were determined for each of the protein samples from solution light scattering and are also shown in Table 1. The WT KorB and (NΔ150)KorB proteins are dimeric as expected, because they contain the C-terminal dimerization domain, whereas (CA105)KorB and (CA60)KorB are monomeric, due to the deletion of this domain.

**RESULTS**

**KorB Exhibits Zones of Intrinsic Disorder along Its Length**—Fig. 1A shows the domain organization in KorB. Bioinformatic analysis of the amino acid sequence using the program DISPROT (34) predicts that KorB protein has a zone of disorder for the first 64 amino acids, as well as between the DNA-binding domain and C-terminal dimerization domain (supplemental Fig. S1). These regions contain many charged residues, with only small hydrophobic residues that cannot form a compact hydrophobic core. Such disordered regions are predicted to be common in eukaryotic genomes and occur in several proteins with key functions in cell signaling and gene regulation, but few examples have been found in bacteria (35). Predictions of disorder in these regions of KorB correspond well with the known structural data (Fig. 1B). For instance the known structure of the DNA-binding domain (PDB code: 1R71) was determined from a KorB derivative that also included the linker region (residues 252–294), although none of the additional residues in this region were visible in the crystal structure (19). The C-terminal dimerization domain (residues 297–358), which has also been crystallized (18) (PDB code: 1IGU) was initially obtained on attempting to crystallize the full-length protein. This suggests that the wild-type protein is readily proteolyzed at residue 296, at the end of the linker region.

To characterize the structure of the protein in solution, we have studied wild-type KorB and a series of deletion mutants of KorB, which contain different combinations of predicted ordered and disordered domains (Fig. 1A). (CA60)KorB does not include the C-terminal dimerization domain, whereas (CA105)KorB does not include the dimerization domain and the putative flexible linker. (NΔ150)KorB includes the central, DNA-binding domain, and the C-terminal dimerization domain, with the putative flexible linker but does not include the first 150 amino acids of KorB, including the putative disordered N-terminal region; whereas (NΔ297)KorB contains primarily the C-terminal dimerization domain, crystallized previously (PDB code: 1IGU). To examine the N-terminal region of the protein, we constructed (NΔ31-CΔ221)KorB, which contains all the structurally uncharacterized N-terminal regions of KorB, apart from the first 31 amino acids: all of which are proposed to be disordered.

**Light Scattering and Analytical Ultracentrifugation Show Solution Homogeneity and Shape Asymmetry**—To determine sample homogeneities and solution molecular weights, the four largest proteins were analyzed by both SEC-MALLS (size exclusion chromatography coupled to multi-angle laser light scattering) and analytical ultracentrifugation. As expected, full-length KorB elutes first from the SEC column, and (CA105)KorB last (supplemental Fig. S2), each giving a single peak. The constructs (NΔ297)KorB and (NΔ31-CΔ221)KorB are below the size that can be accurately determined by SEC-MALLS and, as such, were not subjected to this method of analysis. Analysis of the multiangle-dependent light scattering of these peaks showed that all the proteins were monodisperse; as judged by the ratio of the $Z$-averaged and the number averaged molecular masses, $M_z/M_n$ (Table 1). The weight $M_w$ values were determined for each of the protein samples from solution light scattering and are also shown in Table 1. The WT KorB and (NΔ150)KorB proteins are dimeric as expected, because they contain the C-terminal dimerization domain, whereas (CA105)KorB and (CA60)KorB are monomeric, due to the deletion of this domain.

All sedimentation velocity profiles (Fig. 2A) showed a single symmetric peak, which did not vary with loading concentration of the sample (data not shown). This is also indicative of a single monodisperse species and complements the SEC-MALLS analysis. Molecular weights determined from the ratio of sedimentation and diffusion coefficient were close to those determined from SEC-MALLS (Table 1), again indicative of monodispersity. For the smaller proteins, (NΔ297)KorB and (NΔ31-
In contrast, the frictional ratios for the four largest proteins range from 1.6 to 1.9 (Table 1) are larger than expected for a "typical" globular protein (1.01–1.60) (25). This indicates that the molecules are quite extended in solution. The viscosity increments derived for the four largest proteins were found to fall between values of 11.0 and 14.1 ml/g. Intrinsic viscosity has a weak dependence on molecular mass but a strong dependence on shape and flexibility (36). These values fall closer to values in the literature for flexible and unfolded proteins (typically 13–15 ml/g), than for globular proteins (typically 2.5–5.0 ml/g). This situation is consistent with previous determinations of intrinsic viscosity on unstructured proteins (36, 38) and in-line with our bioinformatics predictions.

All Constructs Are Predominately Folded but Contain a Significant Unfolded Component—NMR and CD spectroscopy values were used to examine the secondary structure of the proteins. The one-dimensional NMR spectra of the proteins (supplemental Fig. S3A) showed some signatures of folded protein, with dispersed signals, including high field methyl groups (supplemental Fig. S3B) and low field amino protons. However, all the proteins also had a number of sharp, intense peaks corresponding to highly flexible segments.

Fig. 2B shows CD spectra for full-length KorB and the deletion mutants. All of the constructs showed some folded character; WT KorB, (N\Delta150)KorB, and (C\Delta60)KorB have similar CD spectra with a peak at 222 nm, consistent with some \(-\alpha\)-helical component, but the component at 210 nm is higher than that at 222 nm, concordant with a measurable component of random coil. (N\Delta31-C\Delta221)KorB shows a similar spectrum, but much less intense suggesting it contains less secondary structure. For (C\Delta105)KorB, the negative maximum is shifted to 206 nm, concordant with the protein containing a more random coil, whereas for (N\Delta297)KorB the spectrum is consistent with a \(\beta\)-sheet with a maximum at 200 nm and only a small peak at 217 nm. The thermal denaturation of the proteins was observed by monitoring the CD spectrum at 222 nm; all the constructs melted with a midpoint at \(40^\circ C\) (see supplemental Fig. S4), although the transitions did not appear to be cooperative, suggesting that the proteins are not folded as single globular domains.
TABLE 1

| Construct               | $s_{20,w}^0$ a | $\beta f_o^b$ | $M_w^c$ | $M_m^d$ | $M_r^e$ | Polydispersity ($M_r/M_m$) f | $\eta^g$ mL g$^{-1}$ |
|-------------------------|--------------|--------------|--------|--------|--------|-----------------------------|-----------------|
| WT KorB                 | 3.8 (±0.2)   | 1.9 (±0.1)   | 41.6   | 78.2   | 83.7   | 1.003 (±0.002)              | 14.1 (±0.01)   |
| NA150                   | 2.7 (±0.2)   | 1.6 (±0.1)   | 25.7   | 47.5   | 51.3   | 1.003 (±0.002)              | 11.6 (±0.07)   |
| CA60                    | 2.5 (±0.1)   | 1.7 (±0.1)   | 34.7   | 34.7   | 34.5   | 1.008 (±0.004)              | 13.6 (±0.01)   |
| CA105                   | 2.1 (±0.1)   | 1.6 (±0.1)   | 30.1   | 29.9   | 32.2   | 1.004 (±0.002)              | 11.0 (±0.01)   |
| (N31-CΔ221)KorB         | 1.5 (±0.1)   | 1.55 (±0.1)  | 18.1   | 19.9   | ND     | ND                          | ND              |
| (N297)KorB              | 1.5 (±0.1)   | 1.55 (±0.1)  | 18.1   | 19.9   | ND     | ND                          | ND              |

a Determined from sedimentation velocity.

b Determined from analysis of sedimentation coefficient using SEDNTERP.

c Calculated from the amino acid sequence.

d Determined from a direct fit to the sedimentation velocity data.

e Determined from multilangle light scattering.

f Determined from the molecular weight moments from light scattering analysis.

g Viscosity increment determined from differential pressure viscometry.

Analyzed using the CD spectra (23, 24) using the CDSTR method (21, 22) showed that each of the constructs contains 30–40% random coil. WT KorB, (N150)KorB, and (C60)KorB contain around 35% $\alpha$-helix with $\sim$15% $\beta$-sheet. (C105)KorB contains less $\alpha$-helix and more $\beta$-sheet, while (N31-CΔ221)KorB and (N297)KorB contain 30 and 45% beta sheet, respectively (see Table 2). From all these spectra, it is clear that each of the proteins contains some folded regions but also a significant amount of random coil structure, in line with the prediction of segments of intrinsic disorder.

The secondary structure content determined from the CD analysis for (N297)KorB agrees well with the crystal structure of this domain, whereas that of (N150)KorB agrees with that predicted from the two known crystal structures of KorB with random coil at residues 251–297. The other constructs contain the N-terminal region of KorB, the structure of which is currently not available. The secondary structural elements for these KorB constructs were estimated in two ways: either solely using the two crystal structures of KorB, with residues 1–137 and 258–297 of KorB unstructured (Table 2, column 5), or using, in addition, a homology model of the N-terminal region of KorB (Table 2, column 6) with residues 54–153 of KorB folded. This was built based on the homology of these residues of KorB with residues 24–120 of SpoOJ (homology level 35%; see Fig. 1B) using the program CPHmodels 3.0 (27). For WT KorB and (C60)KorB proteins the CD analysis shows that the secondary structure is similar to that predicted with the N-terminal region folded. For (C105) KorB and (N31-CΔ221)KorB there is less $\alpha$-helix than predicted for a folded N-terminal region, suggesting that one or more of the helices in the model may not be formed. In all cases, the CD suggests that the N-terminal 50 amino acids of KorB and the linker between the central and DNA-binding domains are unstructured.

KorB Domain Structure Exhibits Modularity—WT KorB, (N150)KorB, (N31-CΔ221)KorB, and (N297)KorB were $^{15}$N-labeled, and the environments of the amide NH-groups in these proteins were examined by two-dimensional $^1$H-$^{15}$N heteronuclear single quantum coherence NMR. Despite its 84-kDa mass, WT KorB shows relatively sharp lines in the NMR spectrum (Fig. 3), probably due to independent motion between domains. There was also a large number of intense peaks at 8.0–8.5 ppm, characteristic of a random coil. Less overlap is seen in the spectrum of (N150)KorB, although it also has a number of sharp peaks in this region of the spectrum. The smaller domains show well resolved spectra. All of the peaks in the spectrum of (N297)KorB are observed in similar positions in the spectrum of (N150)KorB. Similarly, the resolved peaks observed in the spectrum of (N150)KorB are also seen in identical positions in the spectrum of WT KorB, whereas some of the additional peaks in the latter are observed in the spectrum of (N31-CΔ2121)KorB. One example is the group of peaks at 9.0–9.5 ppm in the $^1$H dimension and 127–130 ppm in the $^{15}$N dimension (box A). (N297)KorB shows four peaks in this region of the spectrum, also seen in (N150)KorB with one additional peak just resolved at 9.4 ppm and 127 ppm. WT KorB shows the same five peaks as (N150)KorB with an additional peak at 9.4 and 130 ppm. (N31-CΔ221)KorB shows only one peak, at 9.3 and 130 ppm, corresponding to this additional peak in the WT protein. Similarly in Box B 6.8–7.0 ppm and 116–121 ppm, (N297)KorB shows two peaks, (N150)KorB shows three, and WT KorB shows four, the additional peak being in a similar position to that in (N31-CΔ221)KorB. This clearly shows that the C-terminal domain is unaffected by the presence of the central domain or the N-terminal region of the protein.
and similarly the central domain is unaffected by the N-terminal amino acids. From this we can deduce that the free protein has a modular structure, whereby there is little or no protein-protein interaction between domains, so that the domains show independent orientations in solution.

SAXS Reveals the Extent of Multiple Conformers in Solution—To construct a model of the whole of KorB from the high resolution domain structures available, we used SAXS. Traditionally, this technique provides excellent low resolution data of the overall shape of proteins into which higher resolution structures can be placed (39). SAXS also has been proposed as a technique that is capable of characterizing intrinsically disordered domain structure due to its appropriate solution-based measurements of intermediate length scales (5–100 Å) (40, 41). SAXS data (Fig. 4A) were obtained at DESY in Hamburg and processed to produce Guinier plots (Fig. 4B), Kratky plots (Fig. 4C), and other analyses that reveal the extent of multiple conformers in solution.

### TABLE 2
Structure prediction from (i) CD and (ii) from crystallography and homology model of domains

The first four columns show the predictions from the CDSSTR algorithm contained within the DICHROWEB server. The differences in prediction reflect the use of different reference sets.

| Construct | Predicted % α-helical | Predicted % β-sheet | Predicted % turn | Random coil % | Secondary structure based on crystallography data of KorB | Secondary structure from crystallography and predicted homology model |
|-----------|-----------------------|---------------------|-----------------|--------------|----------------------------------------------------------|---------------------------------------------------------------|
| WT KorB   | 0.35                  | 0.14–0.17           | 0.16–0.19       | 0.28–0.36    | 21% H 7% S²                                                | 38% H 15% S                                                    |
| (NΔ150)KorB | 0.33                  | 0.15–0.17           | 0.17–0.20       | 0.28–0.34    | 30% H 12% S                                                | N/A                                                           |
| (CΔ60)KorB | 0.35–0.37             | 0.11–0.15           | 0.16–0.20       | 0.28–0.37    | 24% H 0% S                                                 | 45% H 9% S                                                    |
| (CΔ105)KorB | 0.19–0.22             | 0.18–0.31           | 0.16–0.21       | 0.33–0.46    | 28% H 0% S                                                 | 52% H 10% S                                                    |
| (NΔ31-CΔ221)KorB | 0.07–0.09             | 0.30–0.33           | 0.18–0.22       | 0.33–0.34    | N/A                                                       | 37% H 21% S                                                    |
| (NΔ297)KorB | 0–0.02                | 0.45                | 0.18–0.23       | 0.30–0.32    | N/A                                                       | N/A                                                           |

* H = α-helix; S = β-sheet; N/A = not applicable.
Intrinsic Disorder in KorB

4C), and distributions of distances (Fig. 4D). Guinier plots of the proteins were linear in $Q^2$ in the low angle region, showing that the proteins are monodisperse (Fig. 4B), in line with the measurements by light scattering and sedimentation velocity. Analysis of the Guinier region ($R_g Q < 1.3$) yielded the radii of gyration shown in Table 3, consistent with that found by second moment analysis of the distance distribution functions (Fig. 4D). The calculated molecular masses from the Guinier plots are similar to those from sedimentation analysis, again demonstrating the monodispersity of the samples. These are also consistent with the excluded (Porod) volumes from the distance distribution function (Table 3). WT KorB, (N$\Delta$150)KorB, (C$\Delta$60)KorB, and (C$\Delta$105)KorB all exhibited a definite peak in the Kratky plots, a signature of globular structure, although the data did not go to zero at high $Q$ value. This is indicative of some degree of random coil component (40), again in line with the predictions of intrinsic disorder and the CD (Fig. 2B) and NMR (Fig. 3) measurements.

We initially derived a low resolution model of the overall shape of the dimeric C-terminal domain, (N$\Delta$297)KorB, from the SAXS data using GASBOR (30) and showed that it was consistent in shape with the crystal structure (PDB code: 1IGU). This shows that the solution structure is similar to that in the crystal structure (see supplemental Fig. S5A). The model of the N-terminal of KorB fits inside the envelope derived from the scattering data (see supplemental Fig. S5B). However, the previous CD data (Table 2) suggest that the long $\alpha$-helix is largely absent, casting doubt as to whether this region is fully folded. Therefore we constructed models of WT KorB, (C$\Delta$60)KorB, and (C$\Delta$105)KorB with region 31–136 folded and unfolded, so as to compare both of these models with our scattering data. The WT KorB, (N$\Delta$150)KorB, (C$\Delta$60)KorB, and (C$\Delta$105)KorB constructs containing regions of intrinsic disorder do not adopt a single conformation in solution and, as such, the above method of structure determination from SAXS is not appropriate. We therefore adopted the ensemble optimization method of Svergun and co-workers (31). Firstly, two sets of 10,000 conformers, based on the known domain structures with flexible amino acid chains between the domains, were generated (see “Experimental Procedures”). In the first set the N-terminal domain up to residue 137 is modeled as a flexible chain, whereas in the second residues 54–137 are assumed to have the structure based on the SpoOJ homology model. From each pool of structures, we selected distributions of 15–20 conformers that together fit the raw x-ray scattering data, using the genetic algorithm optimization method contained in the ensemble optimization method software (31) (supplemental Fig. S6 and Table 3). Fig. 5 shows the distributions of the radii of gyration (left-hand column) and maximum distance ($D_{max}$) of each construct (right-hand column) for the con-

TABLE 3

| Molecule               | $R_g^a$  | $R_g^b$  | Molecular mass$^c$ | Porod volume | $D_{max}$ | $(R_g)^a$ from selected distributions$^d$ | $(D_{max})$ from selected distributions$^d$ | $\chi^2$ of fit to distribution$^d$ |
|------------------------|----------|----------|--------------------|--------------|----------|------------------------------------------|------------------------------------------|-------------------------------------|
| WT KorB                | 55 (±0.2)| 55 (±0.2)| 82.7 (±5.5)        | 154          | 180      | 59/53                                    | 196/198                                  | 1.186/1.423                         |
| (N$\Delta$150)KorB    | 42 (±0.2)| 42 (±0.2)| 55.9 (±3.5)        | 118          | 160      | 35/33                                    | 114                                    | 2.688                              |
| (C$\Delta$60)KorB     | 39 (±0.2)| 38 (±0.2)| 38.0 (±5.1)        | 85           | 155      | 35/33                                    | 114/104                                 | 2.985/5.135                        |
| (C$\Delta$105)KorB    | 36 (±0.1)| 37 (±0.2)| 35.4 (±7.1)        | 67           | 125      | 33/30                                    | 98/80                                   | 2.994/4.112                        |
| (N$\Delta$31–C$\Delta$221)KorB | 21 (±0.1)| 21 (±0.2)| 14.1 (±3.7)        | 35           | 7.3      | ND$^*$                                   | ND$^*$                                 | ND$^*$                             |
| (N$\Delta$297)KorB    | 24 (±0.1)| 24 (±0.1)| 16.1              | 49           | 8.5      | ND$^*$                                   | ND$^*$                                 | ND$^*$                             |

$^a$ Determined from second moment of the distance distribution function from SAXS data using GNOM (27).

$^b$ Determined from Guinier analysis of the SAXS data.

$^c$ Determined from zero angle intensity of scattering: I(0).

$^d$ Calculated for unfolded N-terminal regions (left) and the folded N-terminal regions (right).

$^*$ ND, not determined.

FIGURE 4, A, SAXS data of WT KorB and deletion mutants. WT KorB (black line), (N$\Delta$150)KorB (red line), (C$\Delta$60)KorB (green line), (C$\Delta$105)KorB (yellow line), (N$\Delta$31–C$\Delta$221)KorB (blue line), and (N$\Delta$297)KorB (pink line). Data have been offset by one log unit for clarity. B, Guinier plots for each of the data; as can be seen they are all linear indicating that the samples are all monodisperse, in line with the other hydrodynamic data. C, Kratky plots of the scattering data. Both data sets for the (N$\Delta$31–C$\Delta$221)KorB and (N$\Delta$297)KorB constructs have been omitted for clarity, because both of these data sets were considerably more noisy in the high angle region, thus obscuring the other data sets in the plot. As can clearly be seen, all constructs exhibit a characteristic Gaussian shape due to folded domains of KorB, but also a significant rise at higher angle regions indicating a random coil component. D, the distance distribution functions calculated using GNOM for each of the KorB constructs.

15446 JOURNAL OF BIOLOGICAL CHEMISTRY
structs calculated with the N-terminal region fully unfolded (Fig. 5A) and with residues 54–137 folded (Fig. 5B). In *black* are the data for the pool of all calculated 10,000 random conformers and in *red* are the data for the distributions selected by the ensemble optimization method algorithm for each of the constructs (supplemental Fig. S7 shows the 20 selected conformers selected).

Although there is conformational heterogeneity, similar structural themes emerge. For the wild-type protein, a dimer has two flexible regions per monomer; there appears to be virtually no selection of preferred conformers from the generated random pool with the N-terminal domain fully unfolded (the *red* and *black* distributions overlay). With calculations based on a partially folded N-terminal domain, the peak for the selected distributions (*red*) is shifted toward the longer conformers in the pool of random conformers (*black*).

To assess this further, we took 9,920 conformers of the WT KorB from the initial pool of 10,000 conformers with residues 54–137 folded and only 80 conformers from the unfolded pool and repeated the selection procedure. Despite the unselected pool containing only 0.8% unfolded conformers, the selected pool contained 14.5% of these conformers, indicating that the unfolded proteins describe the conformation in solution better than the folded N termini. This result is in concert with the CD data shown previously where there is a loss of expected secondary structure with respect to that predicted from the homology model for the N-terminal domain. The NMR data for the N-terminal domain does show that there is some structure present (Fig. 3), however, the peaks between 8.0 and 8.5 ppm could arise from amide protons in either unstructured or α-helical regions. This domain may exist in an equilibrium between a folded and unfolded state, and the folded and unfolded conformations of this domain we have generated represent limiting cases.

The (NΔ150)KorB mutant, a dimer with one flexible region in each monomer but lacking the N-terminal domain, exhibits a small bias of conformers toward compact structures, as judged by both the $R_g$ and $D_{max}$ distributions. The deletion of the N-terminal region is enough to reduce the available number of conformers compared with the wild-type protein from a random to a slightly more defined set. The linker region is slightly more compact than the random coil calculations. The (CΔ60)KorB mutant, a monomer with two flexible regions, shows a similar biasing of selected conformers toward compact structures; irrespective of whether the N-terminal domain is modeled as folded or unfolded; compared with the wild-type protein. Finally, for (CΔ105)KorB mutant, a monomer with only one intrinsically disordered domain, there is a strong bias in the selection toward conformers with smaller $R_g$ values and $D_{max}$ than the random pool, indicating that the protein has a more compact structure than a random selection of conform-
Intrinsic Disorder in KorB

ers. However, in both of these constructs the distributions calculated for the unstructured N-terminal region fit better than for the partially folded N-terminal region of KorB (Table 3, column 9).

DISCUSSION

The solution data obtained on KorB are consistent with sequence predictions of zones of order and intrinsic disorder along the length of protein with disordered regions at the N terminus of the protein and between the central DNA-binding domain and the C-terminal dimerization domain. The CD and NMR data are all consistent with the proteins containing both folded and disordered domains that do not interact. The sedimentation and SEC-MALLS data show that the proteins are monodisperse and extended in solution (as judged by the measured $f_f/s$), with large Stokes radii. This result explains the previous difference between the stoichiometry of KorB estimated from gel filtration (17) and from analytical ultracentrifugation (Fig. 2A): an extended conformation will sweep out a large hydrodynamic volume and as such will elute earlier on the column and be assigned an anomalously larger molecular weight. The sedimentation velocity profiles of all of the proteins were symmetrical and did not show either bimodality or anomalous frictional ratios, both previously postulated to be signatures of intrinsic disorder (40, 43). This indicates that analysis of the sedimentation coefficient may not be a sensitive enough probe of solution conformational heterogeneity for intrinsically disordered systems; viscosity increment may be a more sensitive probe (36).

For WT KorB and the deletion mutants containing more than one domain, the regions of intrinsic disorder mean that a structural model at atomic resolution is presently beyond our technical grasp (44). However, SAXS, allied with other hydrodynamic techniques, is well placed for answering questions about the gross shape and extent of conformational flexibility adopted by the protein (40). As such, we have shown that the flexible linker region in (NΔ150)KorB adopts a slightly more compact conformations rather than the full range of random conformers available. In contrast the N-terminal region is quite compact, as seen with the (CΔ105)KorB construct. It is, however, somewhat difficult to judge the degree of folding of the N-terminal domain, and the folded and unfolded conformers we have generated most probably represent limiting cases. Overall, it is intuitive that the greater the number of intrinsically disordered domains, the greater the number of conformers the protein adopts.

Interestingly, the ordered and previously crystallized portions of KorB are associated with a single function (i.e. DNA binding for the central domain and dimerization for the C-terminal domain). In contrast, the disordered regions are implicated in many functions of KorB. The N-terminal region of KorB modulates DNA-binding strength and site selectivity, repression at a distance, as well as the localization of KorB in the cell. The region between the C-terminal domain and the central domain (residues 252–294) is implicated in several functions of KorB, including repression of transcription and cooperativity with other proteins. Such multifunctionality is a hallmark of intrinsically disordered domains (45).

We propose that the intrinsically disordered domain provides KorB a wide set of possible conformers available as templates for binding to a diverse set of proteins such as KorA, IncC, and RNA polymerase. Such binding events will have an effect on the ensemble of conformations KorB can adopt in solution and thus mediate the set of conformers available for interaction to a second binding partner, e.g. DNA or a third protein. The flexibility allows binding to occur at different distances between KorB and a second repressor while the conformational changes give rise to the cooperative effects.

The disordered regions are also likely to be important for DNA partitioning, as many of the partition proteins need to span two DNA duplexes (or several proteins) to carry out their function. Flexible interactions between the central and C-terminal domains of a ParB protein have been seen in the crystal structures of central and C-terminal domains of ParB from P1 plasmid of *Escherichia coli* bound to different target sequences (37, 42). In ParB, there is a 4-amino acid linker between the central and C-terminal domains and the domains are entirely distinct and rotate 60–160 degrees relative to each other in the asymmetric ParB dimer, allowing it to contact different DNA duplexes. Although KorB and ParB from P1 show only limited structural homology and appear to bind DNA differently, flexibility arising from intrinsic disorder appears to be an essential requirement for partitioning.

The multifunctional role of KorB, involving multiple interacting partners as well as roles in both gene regulation, and partition, makes it an important network hub in the plasmid maintenance and partition process, and an excellent model for understanding other proteins with such domains.

Acknowledgments—D. J. S. and E. I. H. are grateful for Beamtime awards at Deutsches Elektronen-Synchrotron, Hamburg, and we are grateful to Peter Konarov of European Molecular Biology Laboratory, Hamburg, for help with data collection. We also thank Sara Whittaker and Christian Ludwig of the Henry Wellcome Building NMR Centre for support. NMR spectra were collected at the Henry Wellcome Building NMR Centre at the University of Birmingham, funded by the Higher Education Funding Council for England and the Wellcome Trust.

REFERENCES

1. Bignell, C., and Thomas, C. M. (2001) *J. Biotechnol.* 91, 1–34
2. Kostelidou, K., and Thomas, C. M. (2000) *J. Mol. Biol.* 295, 411–422
3. Zatyka, M., Bingle, L., Jones, A. C., and Thomas, C. M. (2001) *J. Bacteriol.* 183, 1022–1031
4. Bhattacharrya, A., and Figurski, D. H. (2001) *J. Mol. Biol.* 310, 51–67
5. Chiu, C. M., Manzoor, S. E., Batt, S. M., Muntaha, S., Bingle, L. E., and Thomas, C. M. (2008) *Plasmid* 59, 163–175
6. Bingle, L. E., Macartney, D. P., Fantozzi, A., Manzoor, S. E., and Thomas, C. M. (2005) *J. Mol. Biol.* 349, 302–316
7. Lukaszewicz, M., Kostelidou, K., Bartosik, A. A., Cooke, G. D., Thomas, C. M., and Jagura-Burczy, G. (2002) *Nucleic Acids Res.* 30, 1046–1055
8. Kostelidou, K., Jones, A. C., and Thomas, C. M. (1999) *J. Mol. Biol.* 289, 211–221
9. Jagura-Burczy, G., Macartney, D. P., Zatyka, M., Cunliffe, L., Cooke, D., Huggins, C., Westblade, L., Khanim, F., and Thomas, C. M. (1999) *Mol. Microbiol.* 32, 519–532
10. Williams, D. R., Macartney, D. P., and Thomas, C. M. (1998) *Microbiol.* 144, 3369–3378
11. Batt, S. M., Bingle, L. E., Daftorn, T. R., and Thomas, C. M. (2009) *J. Mol. Biol.* 385, 1361–1374
12. Macartney, D. P., Williams, D. R., Stafford, T., and Thomas, C. M. (1997) *Microbiology* 143, 2167–2177
13. Jagura-Burdzy, G., and Thomas, C. M. (1997) *J. Mol. Biol.* 265, 507–518
14. Zatyka, M., Jagura-Burdzy, G., and Thomas, C. M. (1994) *Microbiology* 140, 2981–2990
15. Pansegrau, W., Lanka, E., Barth, P. T., Figurski, D. H., Guiney, D. G., Haas, D., Helinski, D. R., Schwab, H., Stanisich, V. A., and Thomas, C. M. (1994) *J. Mol. Biol.* 239, 623–663
16. Jagura-Burdzy, G., Kostelidou, K., Pole, J., Khare, D., Jones, A., Williams, D. R., and Thomas, C. M. (1999) *J. Bacteriol.* 181, 2807–2815
17. Williams, D. R., Motallebi-Veshareh, M., and Thomas, C. M. (1993) *Nucl. Acids Res.* 21, 1141–1148
18. Delbrück, H., Ziegelin, G., Lanka, E., and Heinemann, U. (2002) *J. Biol. Chem.* 277, 4191–4198
19. Khare, D., Ziegelin, G., Lanka, E., and Heinemann, U. (2004) *Nat. Struct. Mol. Biol.* 11, 656–663
20. Leonard, T. A., Butler, P. J., and Löwe, J. (2004) *Mol. Microbiol.* 53, 419–432
21. Srerama, N., and Woody, R. W. (2000) *Anal. Biochem.* 287, 252–260
22. Srerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) *Anal. Biochem.* 287, 243–251
23. Whitmore, L., and Wallace, B. A. (2008) *Biopolymers* 89, 392–400
24. Whitmore, L., and Wallace, B. A. (2004) *Nucleic Acids Res.* 32, W668–673
25. Scott, D. I., Harding, S. E., and Rowe, A. J. (2005) *Analytical Ultracentrifugation: Techniques and Methods*, pp. 1–26, Royal Society of Chemistry, Cambridge, United Kingdom
26. Schuck, P. (2000) *Biophys. J.* 78, 1606–1619
27. Lund, O., Nielsen, M., Lundegaard, C., and Worning, P. (2002) CPHmodels: 3M, a computer program to extract 3D models in CASPS Conference
28. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* 36, 1277–1282
29. Svergun, D. (1993) *J. Appl. Crystallogr.* 26, 258–267
30. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) *Biophys. J.* 80, 2946–2953
31. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) *J. Am Chem. Soc.* 129, 5656–5664
32. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002) *Acta Crystallogr. D. Biol. Crystallogr.* 58, 1955–1957
33. Svergun, D., Barberato, C., and Koch, M. H. (1995) *J. Appl. Crystallogr.* 28, 768–773
34. Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z., and Dunker, A. K. (2007) *Nucleic Acids Res.* 35, D786–793
35. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) *J. Mol. Biol.* 337, 635–645
36. Harding, S. E. (1997) *Prog. Biophys. Mol. Biol.* 68, 207–262
37. Schumacher, M. A., Mansoor, A., and Funnell, B. E. (2007) *J. Biol. Chem.* 282, 10456–10464
38. Chenal, A., Guijarro, J. I., Raynal, B., Delepierre, M., and Ladant, D. (2009) *J. Biol. Chem.* 284, 1781–1789
39. Petoukhov, M. V., and Svergun, D. I. (2007) *Curr. Opin. Struct. Biol.* 17, 562–571
40. Receveur-Bréchot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) *Proteins* 62, 24–45
41. Bernado, P. (2010) *Eur. Biophys. J.*, DOI: 10.1007/s00249-0009-00549-00243
42. Schumacher, M. A., and Funnell, B. E. (2005) *Nature* 438, 516–519
43. Szilágyi, A., Györrfy, D., and Závodszky, P. (2008) *Biophys. J.* 95, 1612–1626
44. Uversky, V. N. (2002) *Protein Sci.* 11, 739–756
45. Haynes, C., Oldfield, C. J., Ji, F., Kligord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., and Iakoucheva, L. M. (2006) *PLoS Comput. Biol.* 2, e100