University of Groningen

Size and shape of the repetitive domain of high molecular weight wheat gluten proteins. 1. Small angle neutron scattering

Egelhaaf, SU; van Swieten, E; Bosma, T; de Boef, E; van Dijk, AA; Robillard, GT; Egelhaaf, Stefan U.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Stefan U. Egelhaaf[1,*]
Eric van Swieten[2,*,†]
Tjibbe Bosma[2]
Esther de Boef[2]
Alard A. van Dijk[2,‡]
George T. Robillard[2,§]
[1] University of Edinburgh, Department of Physics and Astronomy, James Clerk Maxwell Building, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

[2] University of Groningen, Department of Biochemistry and the Groningen Biomolecular Sciences and Biotechnology Institute, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Size and Shape of the Repetitive Domain of High Molecular Weight Wheat Gluten Proteins. I. Small-Angle Neutron Scattering

Abstract: *The solution structure of the central repetitive domain of high molecular weight (HMW) wheat gluten proteins has been investigated for a range of concentrations and temperatures using mainly small-angle neutron scattering. A representative part of the repetitive domain (dB1) was studied as well as an "oligomer" basically consisting of four dB1 units, which has a length similar to the complete central domain. The scattering data over the entire angular range of both proteins are in quantitative agreement with a structural model based on a worm-like chain, a model*

frequently used in polymer theory. This model describes the "supersecondary structure" of dB1 and dB4 as a semiflexible cylinder with a length of about 235 and 900 Å, respectively, and a cross-sectional diameter of about 15 Å. The flexibility of both proteins is characterized by a persistence length of about 13 Å. Their structures are thus quantitatively identical, which implies that the central HMW domain can be elongated while retaining its structural characteristics. It seems conceivable that the flexible cylinder results from a helical structure, which resembles the β-spiral observed in earlier studies on gluten proteins and elastin. However, compared to the previously proposed structure of a (stiff) rod, our experiments clearly indicate flexibility of the cylinder.   © 2003 Wiley Periodicals, Inc. Biopolymers 69: 311–324, 2003

INTRODUCTION

Wheat gluten proteins are of considerable interest due to their functionality in bread. They form extensive insoluble protein networks in dough, which are stabilized by intermolecular disulfide bonds.[1,2] These networks contribute to the biomechanical properties, such as strength and elasticity. The functional behavior of this class of proteins is significantly determined by the high molecular weight (HMW) proteins.[3–9] HMW proteins have molecular masses from 68 to 89 kDa and are divided into three domains, a central repetitive domain flanked by short, nonrepetitive N- and C-terminal domains, as schematically shown in Figure 1A.[5,9,10] The terminal domains are suggested to have a globular structure with high contents of α-helix. They have several cysteine residues present, whose function in dough may lie in the formation of intermolecular disulfide bonds leading to cross-linked networks.[1,2,11] The large central repetitive domain consists of β-turns, and was proposed to form a "supersecondary structure" of an elongated, so-called, β-spiral.[5,7,8,9,12–16] The β-spirals may contribute to gluten elasticity. The precise molecular basis for the elastic properties of gluten and dough is not known, but is thought to involve contributions from the structures of the individual subunits and their interactions via noncovalent and covalent bonds. A thorough knowledge of the size and shape of these types of structures would considerably help to understand HMW proteins and their dominant role in gluten dough textures.

To our knowledge, no detailed three-dimensional (crystal) structure has been determined of the complete protein or the central domain. However, there are a few studies on the solution structure of the protein and the repetitive domain using different proteins and buffers as well as different measurement techniques, such as viscosimetry,[13] small-angle x-ray scattering,[17,18] and scanning tunneling microscopy.[14] They were interpreted in terms of a rodlike structure based on the β-spiral.

In this study we investigate two model proteins, both cloned from the central domain, produced in *Escherichia coli*, and purified with metal chelation chromatography and high-performance liquid chromatography (HPLC). The smaller protein, dB1, represents approximately the first quarter of the sequence of the repetitive central domain with a molecular mass of 16.9 kDa. The second protein, dB4, with a molecular mass of 63.8 kDa, is composed of four dB1 segments resulting in a molecular mass comparable to the full central domain (Figure 1A). The amino acid sequences of dB1 and dB4 are very similar to the central domains of HMW Dx5 and other x-type HMW subunits: The sequences are almost completely covered with three different consensus repeats, whose occurrences are 42% (PGQGQQ), 29% (PGQGQQGQQ), and 17% (GYYPTSPQQ) in dB1 and dB4, while the relative proportions in the central domain in HMW Dx5 are 45, 27, and 23%, respectively.[6,15,16,19–21] We thus consider dB1 and dB4 appropriate models for the wild-type domain of the HMW Dx5 protein. The solution structure of dB1 and dB4 has been studied using small-angle neutron scattering experiments. The data over the entire angular range quantitatively agree with a model of a flexible cylinder. The flexibility of the cylinder represents the new feature compared to models proposed earlier.[5,9,13,17,18] We discuss the consequences of the flexibility on the properties of the central domain and the network, such as the β-spiral structure, the formation of disulfide bonds, and the elasticity.

RESULTS

Cloning of dB1 and dB4

A representative part of the central repetitive domain of HMW Dx5 was amplified by polymerase chain reaction (PCR) and subcloned in pSk⁻ as previously described by van Dijk et al.[16] The location and length of dB1 in the DNA sequence is shown schematically in Figure 1A. To
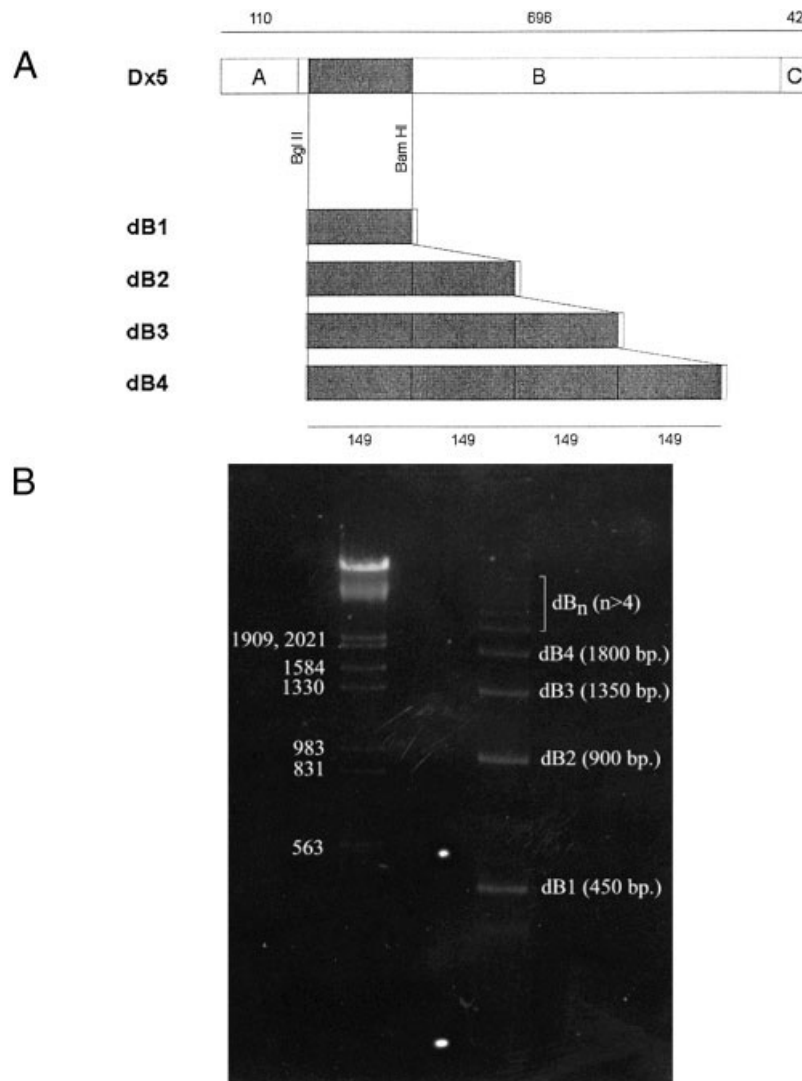
FIGURE 1 (A) Schematic drawing of HMW Dx5, divided into its domains. The width of the domains in the drawing is proportional to the length of the amino acid sequence, indicated in the figure. Only a limited number of extra, non-native, amino acids at the termini of dB1 and dB4 were introduced for cloning and purification purposes, being 7 and 2% of the total number of amino acids, respectively. The size of dB1 and dB4, and their locations in the sequence are indicated in the bottom of the figure. (B) The 1.5% agarose gel, DNA colored with ethidium-bromide (30 ppm) for fluorescent detection. Lane 1: marker, phage λ-DNA + Hind III/Eco RI; lane 2: dB1 concatamers up to 8 units. The number of base pairs are indicated in the figure.

obtain concatemers of dB1, *Bg*/II and *Bam*HI restriction sites were introduced, by PCR, at the 5′ and 3′ end of dB1, respectively. Self-ligation of the *Bg*/II- and *Bam*HI-digested dB1-monomers yielded a population of DNA fragments encoding various length dB1-multimers. Plasmids possessing fragments with the correct orientation were selected by restriction with *Bg*/II and *Bam*HI. These two restriction enzymes produce compatible cohesive ends. However, in dB1-concatemers with the correct orientation both restriction sites were lost due to the ligation of *Bg*/II ends to *Bam*HI ends. After self-

ligation and *Bg*/II/*Bam*HI selection dB1-multimers containing up to 8 dB1-monomers were observed, demonstrating an efficient way of genetic engineering of polymers (Figure 1B). The dB1 and dB4 were purified from agarose gel and ligated into the *Bg*/II/*Bam*HI-digested expression vector pQE60.

Expression and Purification of dB1 and dB4

The yield of dB1 after expression and purification on Ni^{2+}-agarose was approximately 20 mg/L cell cul-
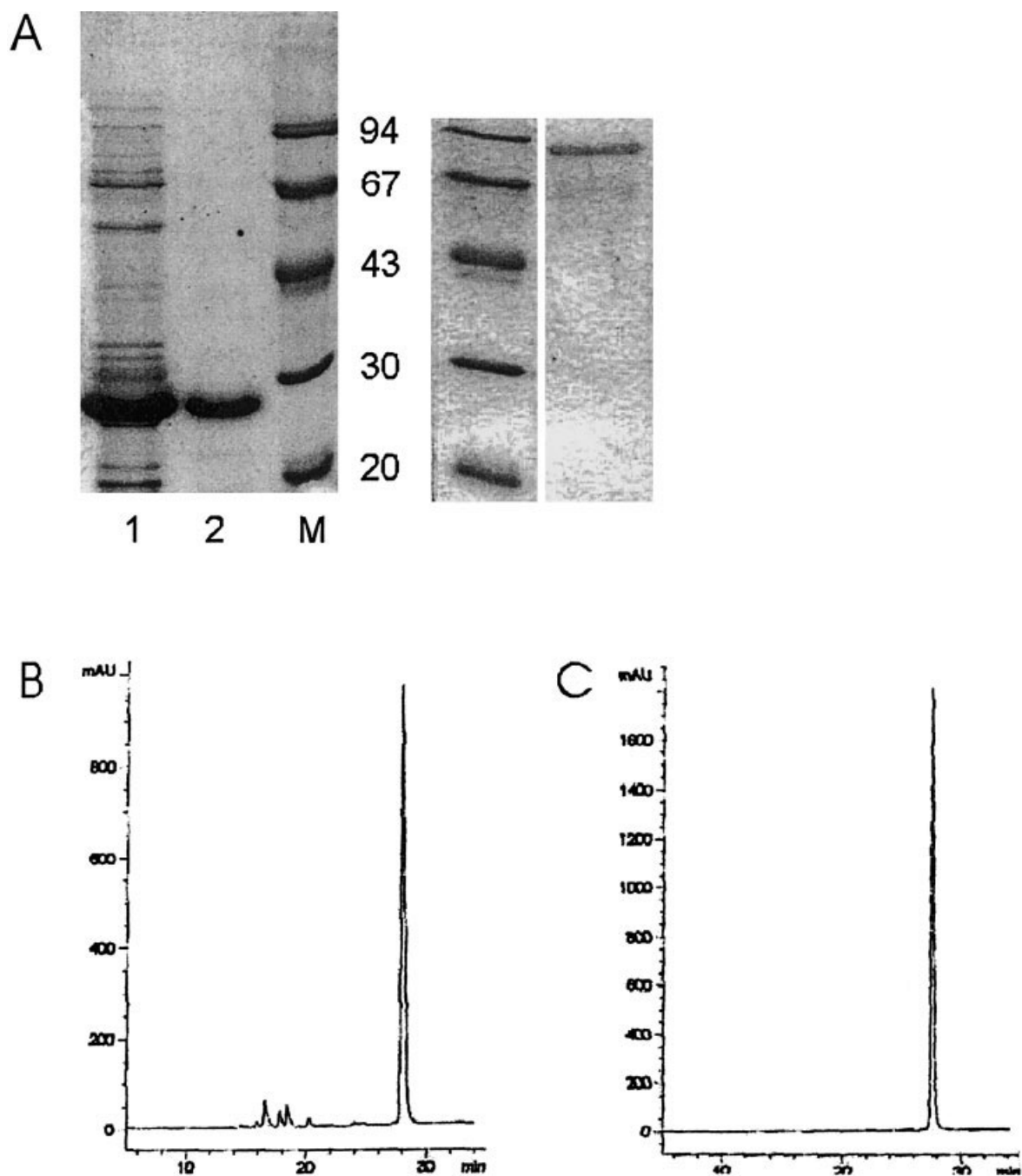
FIGURE 2 Purification of dB1. (A) SDS-PAGE. Lane 1: 100 mM imidazole fraction of Ni^{2+}-agarose dB1 purification, lane 2: HPLC fraction of dB1, M: marker: HPLC fraction of dB4. Analytical HPLC chromatograms of dB1 (B) after Ni^{2+}-agarose purification and (C) additional preparative reversed HPLC purification.

ture. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) demonstrated that a significant amount of other proteins was present in the purified fractions, as shown for dB1 in Figure 2A (lane 1). The estimated purity after the Ni^{2+}-agarose purification step was more than 95%, based on the optical density at 214 nm in an analytical high performance liquid chromatography (HPLC) chromatogram (Figure 2B). The second purification step, preparative (HPLC) gave essentially pure dB1 (>99%), as shown in Figures 2A (lane 2) and 2C. The SDS band was found at ≈26.5 kDa.[16] The yield of dB4

FIGURE 3 CD spectra of dB1 (+) and dB4 (●) at the same protein concentration of 0.10 mg/mL and a temperature of 25°C.

after Ni^{2+}-agarose purification was roughly 5 mg/L cell culture, as determined by absorption. The purities of dB4 after each purification step were comparable with those obtained for dB1. The first 11 N-terminal residues of dB4 were identified as GGSPGQASPQR, confirming the p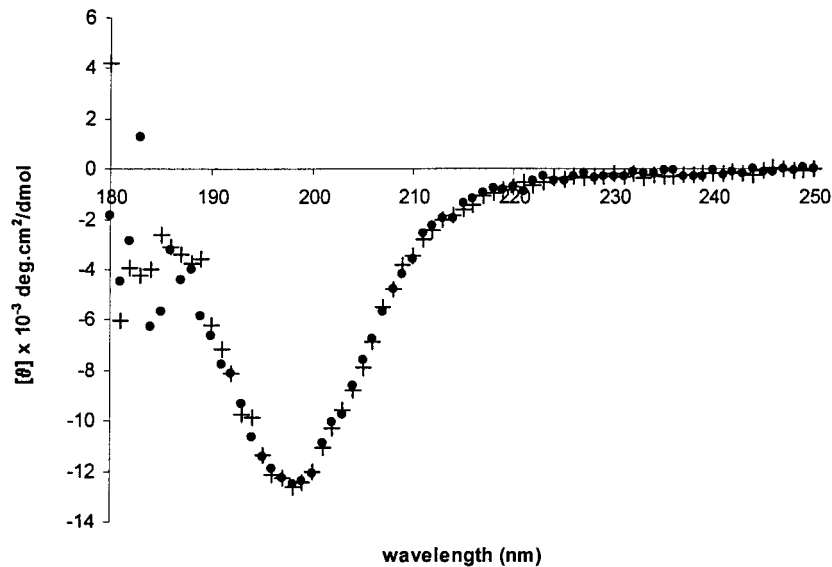rotein's correct identity. They are also equal to the first 11 residues of dB1, which is consistent with the modular sequence of dB4 (Figure 1A). The sequencing data also signified that the N-terminal methionine was cleaved off by post translational modification. The intactness of the C-terminus was ensured by the presence of the His tag used in the Ni^{2+}-agarose purification. The experimental molecular mass of 64 ± 1 kDa, obtained with mass spectroscopy, was consistent with the theoretical molecular mass of 63.8 kDa. The SDS band was found at 85–90 kDa (Figure 2A).

Small-Angle Neutron Scattering (SANS)

A series of scattering experiments was performed to obtain quantitative information on the solution structure of the central repetitive domain of HMW wheat gluten proteins. We investigated two model proteins, dB1 and dB4, both cloned from the central domain (Figure 1). The bigger protein, dB4, is composed of basically four dB1 units and has a molecular mass comparable to the full central domain. Only a limited number of extra, non-native, amino acids at the termini of dB1 and dB4 were introduced for cloning and purification purposes, being 7 and 2% of the total number of amino acids, respectively. Both proteins

are thus expected to have a similar structure, but to differ in size. This is supported by CD measurements. The CD spectra of dB1 and dB4 at the same concentration are identical within experimental accuracy, indicating preservation of secondary structure (Figure 3).

The scattering power of a particle increases dramatically with its molecular mass. Thus any aggregation has to be avoided, since even a small amount of aggregates would completely dominate the scattering behavior and render any meaningful data analysis impossible. Dynamic light scattering (DLS) was used to determine the most favorable conditions, in particular to find the best-suited buffer (acetate/EDTA). DLS was also applied to assess sample quality with respect to aggregation prior to the SANS experiments. Furthermore, care was taken to prepare protein solutions of high purity, especially to avoid contamination by other proteins (Figure 2). The buffer selection based on DLS is described in the second paper of this series.

SANS measurements were performed using a buffer with heavy water (D_2O) instead of light water (H_2O) in order to reduce incoherent background from hydrogen and maximize scattering contrast. SANS measurements were performed with dB1 and dB4 at different concentrations ranging from 1.8 to 9.9 mg/mL at 25°C. Figure 4A shows the resulting scattering intensities $l(q)$ normalized by the protein concentration as a function of the magnitude of the scattering vector $q = (4\pi/\lambda) \sin(\theta/2)$, where θ is the scattering angle and λ the neutron wavelength. When
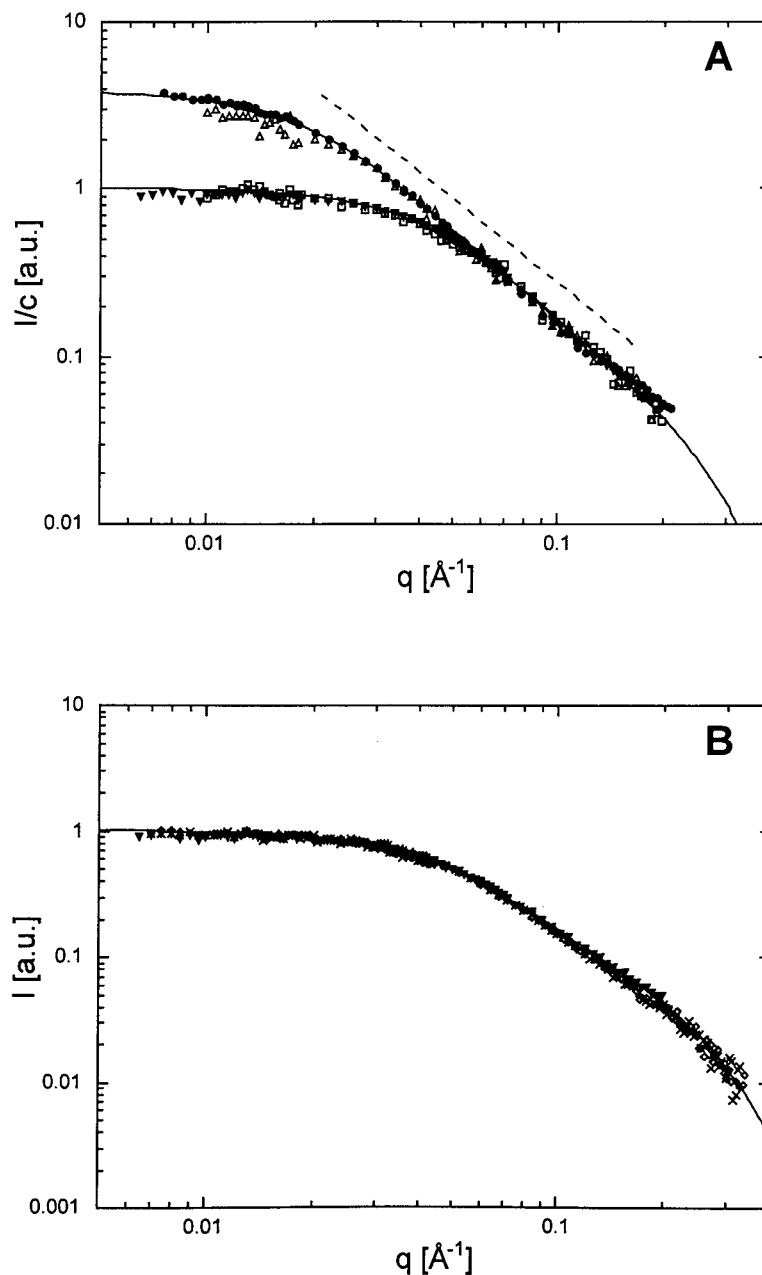
FIGURE 4   (A) SANS intensities $l(q)$ normalized by protein concentration c as a function of the magnitude of scattering vector q at different protein concentrations for dB1 (▼: 9.88 mg/mL; □: 1.82 mg/mL) and dB4 (●: 9.19 mg/mL; △: 1.91 mg/mL) at 25°C. The asymptotic $l(q) \sim q^{-5/3}$ behavior is shown as the dashed line and the solid lines are fits using a flexible cylinder model. (B) The $l(q)$ for dB1 at different temperatures (◇: 8°C, ▼: 25°C, ✕: 30°C) and constant concentration of 9.88 mg/mL. The solid line is calculated based on the values obtained from a fit to the data of dB1 and dB4 at 25°C.

normalized with respect to the concentration, the data do not show significant differences. This indicates that measurements are not affected by interaction effects within this concentration range. For the following detailed analysis we have, therefore, chosen

the samples with the higher concentrations of dB1 and dB4 to obtain better statistical accuracy. Experiments were also carried out at different temperatures, 8°C, 25°C, and 30°C (Fig. 4B). The data obtained are in perfect agreement and indicate no significant temper-

FIGURE 5 SANS intensities $l(q)$ vs square of scattering vector q^2 for dB1 (▼) and dB4 (○). The solid lines are Guinier fits, which are only based on the data points indicated by filled symbols.
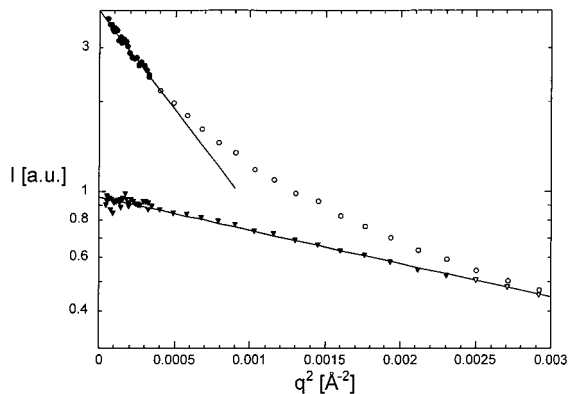
ature dependence of the structure within this range. A detailed investigation of the thermal behaviour of dB1 is treated elsewhere.[19]

The Radius of Gyration of dB1 and dB4. Without any assumptions concerning the structure of the proteins, we can obtain the radius of gyration, R_G that characterizes the overall size of the proteins R_G can be determined from the low q region ($q R_G < 1$) using the so-called Guinier approximation.[22]:

$$I(q) = I_0 e^{-q^2 R_G^2/3} \tag{1}$$

where l_0 is the forward scattering intensity at $q = 0$. Fitting $I(q)$ up to $q^2 < 0.0025$ Å^{-2} for dB1 and $q^2 < 0.0004$ Å^{-2} for dB4, we obtain $R_G \approx 28$ and 67 Å for dB1 and dB4, respectively (Figure 5). Based on these estimates of the R_G values we can already obtain a very rough idea on the structure of the proteins. Were the proteins globular, these R_G values would correspond to spheres of radii 36 and 87 Å, respectively. This is much larger than the 17 and 26 Å, respectively, estimated from their molecular masses and density (about 1.4 g/cm^3). This already indicates that the proteins are not compact, but have either an elongated or very open structure.

The Shape of dB1 and dB4. Detailed information on the shape of the protein can be obtained from the q dependence of the scattering intensity at intermediate and high q values. The magnitude of the scattering vector q determines the length scale, $L_{scatt} \sim 1/q$, probed in a scattering experiment. Different regions in the scattering pattern thus correspond to the behavior characteristic of the various length scales of the pro-

tein. In particular, higher q values correspond to smaller length scales; hence, the scattering data become sensitive to structural details and provide information on the more local structure of the protein. Beyond the Guinier region, which is associated with the overall size of the protein and has a rather small slope, the scattering intensity crosses over to a steep decay following a power law $l(q) \sim q^{-5/3}$ in the intermediate q range (Figure 4A, dashed line), before the slope changes again at $q \approx 0.13$ Å^{-1}. The exponent of $-5/3$ is frequently found for polymer solutions. It is characteristic for polymers in a good solvent, where the good solvent conditions imply excluded volume effects.[23–25] This model not only takes into account short-range interactions between neighboring segments, but also between segments widely separated along the chain. If the polymer is modeled as a connected path, the excluded volume effect will correspond to the condition that the path cannot pass through any sites that have already been traversed previously, which is called a "self-avoiding walk." (An "ideal chain" polymer only taking into account interactions with neighboring segments corresponds to a "random walk" without excluded volume effect and leads to a slope of -2).[23] It is important to note that this "polymer" does not necessarily correspond to the protein as a chain of polymerized amino acids, but is a more general model of a flexible cylinder describing the overall structure of the protein.

Persistence Lengths of dB1 and dB4. The cylinder nevertheless is not completely flexible, but only "semiflexible" It does not freely bend within very small distances but has a tendency to persist in some initial direction. Below a certain distance it thus behaves like a stiff cylinder. This defines the so-called persistence length l_p, which is used to characterize the flexibility of polymers. These distances over which the cylinder is locally stiff, represent a shorter length scale than the (contour) length L and hence control $l(q)$ at intermediate values of q. One thus expects a crossover to a scattering behavior characteristic for stiff cylindrical structures, which is an asymptotic q^{-1} dependence for $l(q)$.[23] This transition is predicted to be located at $q_p \approx 1.9$.[26,27] However, the transition region is often fairly broad. A particularly sensitive way to represent scattering data for semiflexible polymers is the so-called "Holtzer plot" (or "bending rod plot" shown in Figure 6, in which $ql(q)$ is plotted versus q.[26–28] In this plot the $l(q) \approx q^{-1}$ dependence for an infinitely thin cylinder results in a plateau.

The crossover from the asymptotic $q^{-5/3}$ to q^{-1} dependence is found at a similar q value for both proteins and occurs at about $q \approx 0.13$ Å^{-1}. This

results in an estimate of $l_p \approx 15$ Å. The Holtzer plot also shows a maximum toward smaller values of q. The position of the maximum $q_{max} \approx 0.045$ and 0.02 Å^{-1} for dB1 and dB4, respectively, is independent of the stiffness, i.e. l_p, and only depends on the radius of gyration, $q_{max} R_G \approx 1.4$.[26,27] This is in very nice agreement with the radii of gyration determined from the Guinier fit.

The height of the maximum depends on the number of persistence lengths l_p along the contour length L, $N_p = L/l_p$.[26] As expected, the maximum and thus N_p is higher for the longer dB4. In addition, the values of N_p estimated from the heights of the maxima agree with those calculated using the above values for R_G and l_p. The observed maxima are due to the stronger q dependence of a flexible coil ($q^{-5/3}$) when compared to a stiff cylinder (q^{-1}). In particular, no maximum is found for stiff cylinders. Our data are thus in clear disagreement with the structure of a stiff cylinder, which was proposed in earlier studies.[13,17,18] Our data also do not support a prolate shape, for which axial ratios much bigger than 1, i.e., an elongated prolate ellipsoid, show a similar q^{-1} behavior.

Contour Lengths of dB1 and dB4. Based on the radii of gyration R_G and the above estimate of the persistence length l_p we can now get an estimate of the (contour) length L of the cylinder. An analytical expression relating these quantities for semiflexible polymers without excluded volume effects is given in Eq. (2).[29]

$$R_{G,0^2}(L,l_p) = \frac{L\,l_p}{3}\left\{1 - 3\left(\frac{l_p}{L}\right) + 6\left(\frac{l_p}{L}\right)^2 \right.$$
$$\left. - 6\left(\frac{l_p}{L}\right)^3 (1 - e^{L/l_p})\right\} \quad (2)$$

The effect of excluded volume is accounted for by an expansion factor $\alpha = R_G/R_{G,0}$, which follows[24]:

$$\alpha(L,l_p) = \left\{1 + \left[\frac{L}{6.24\,l_p}\right]^2 + \left[\frac{L}{17.34\,l_p}\right]^3\right\}^{0.17/6} \quad (3)$$

A combination of these two equations gives R_G as a function of L and l_p and its application yields an estimate for the length of dB1 $L_{dB1} \approx 185$ Å and dB4 $L_{dB4} \approx 740$ Å. This corresponds to a ratio $L_{dB4}/L_{dB1} \approx 4$, which compares with a value of $607/160 = 3.79$ expected from the number of amino acids of the two proteins.

Local Cross-Sectional Structures. At even higher q values and thus smaller length scales, we can probe the local cross-sectional structure of the cylinder. Its size is described by the cross-sectional radius R_{cs}. This gives rise to a cross-section Guinier behavior at $q\,R_{cs} < 1$ and a strong decrease in the scattered intensity toward the highest q values measured ($q > 1/R_{cs}$). The data do not show a clear Guinier regime but only a broad crossover region. Possibly it even overlaps with the crossover region from the flexible to the semiflexible asymptotic behavior. This indicates that the persistence length l_p and cross-sectional radius R_{cs} are very similar. The same behavior can be observed for "real" polymers. Depending on the degree of deuteration of polystyrene (fully deuterated, phenyl ring deuterated, or backbone deuterated), its effective cross-sectional radius $R_{cs,eff}$ can be changed while l_p is kept constant. This results in different l_p–$R_{cs,eff}$ ratios. Accordingly, very different crossover regions are observed.[25,30] Due to the supposedly similar sizes, we do not attempt to obtain a value for R_{cs} from the asymptotic behavior only, but leave the determination of R_{cs} to model fitting.

Fitting the worm-Like Chain Model. Up to now we have been considering the scattering behavior in terms of asymptotic expressions only.[23] In order to take full advantage of the information content in the data over the entire q range, we also include the crossover regions. We thus need a complete, quantitative characterization of the model and its scattering function, based on which a least-squares analysis can then be performed. As mentioned above, the model of a semiflexible cylinder with a circular cross-section that does not penetrate itself is well studied in polymer theory. It is known as the worm-like chain model of Kratky and Porod[31] modified to include excluded volume effects and finite cross section.[25,32] Its scattering cross-section $l(q)$ can be described by[33]

$$l(q,L,l_p,R_{cs}) - l_0 S_{wc}(q,L,l_p)S_{cs}(q,R_{cs}) \quad (4)$$

where $S_{wc}(q,L,l_p)$ and $S_{cs}(q,R_{cs})$ are the normalized scattering functions of an infinitely thin worm-like chain and of the cross section, respectively. This decoupling approximation implies that the q dependence of the scattering intensity can be written as the product of the scattering function of an infinitely thin worm-like chain and of the cross-section, which is valid if the persistence length l_p of the chain is significantly larger than the cross-sectional radius R_{cs}. For $S_{wc}(q,L,l_p)$ a numerical parametrization based on Monte Carlo simulations is used (method 3 with excluded volume effect).[25] The finite size of the cross-section is included by $S_{cs}(q)$. For a circular cross-section of radius R_{cs} it is given by

$$S_{cs}(q,R_{cs}) = \left[2 \frac{J_1(qR_{cs})}{qR_{cs}} \right]^2 \qquad (5)$$

where $J_1(x)$ is the Bessel function of first order. This provides us with a complete description of the q dependence of the scattering intensity for a self-avoiding, semiflexible cylinder with circular cross-section, which depends on the length L, persistence length l_p cross-sectional radius R_{cs}, an overall scale factor, and a constant background, each for both proteins. The number of free parameters can be reduced using *a priori* knowledge on the proteins. Their amino acid sequence as well as our CD measurements (Figure 3) indicate that dB1 and dB4 have the same structure, and hence their persistence lengths and cross-sectional radii should be identical. Furthermore, based on the numbers of amino acid residues, the ratio of their lengths is given by $L_{dB4}/L_{dB1} = 3.79$. This leaves us with only seven fit parameters for both scattering curves; L_{dB4}, l_p, R_{cs}, two overall scalefactors and backgrounds. The starting values for L_{dB4} and l_p were chosen based on the above constraints and Eqs. (2) and (3) together with the radii of gyration determined from the Guinier fit. Throughout the data analysis corrections were made for instrumental smearing.[34,35] For each instrumental setting the ideal scattering curves were smeared by the appropriate resolution function when the model scattering intensity was compared to the measured $l(q)$ by means of least-squares methods.[36] The resulting fit yields good agreement with the data over the entire q range as shown in Figure 4A and supported by $\chi^2 = 2.6$. The resulting best-fit parameters are $L_{dB4} = 893 \pm 60$ Å, resulting in $L_{dB1} = 235 \pm 16$ Å, $l_p = 12.6 \pm 1.3$ Å, and $R_{cs} = 7.4 \pm 1.5$ Å. The errors are determined according to the procedure described by Pedersen.[36] The fitted values are similar to the values estimated based on the asymptotic behavior, which supports the chosen model function. Also the Holtzer plot (Figure 6), which is a very sensitive representation for the applied model, shows good agreement of the fit with the data. This demonstrates that the model provides a very good description of the scattering function and is capable of quantitatively reproducing the experimental features over an extended range of scattering vectors q. Hence we do not have to rely only on asymptotic expressions, such as a Guinier approximation for the overall size and cross-sectional radius.

The uncertainty in R_{cs} toward smaller values is rather large, while the possible error toward larger values is much smaller, only 0.1 Å. This large error is due to the similar size of R_{cs} and l_p. In the experiments both affect a similar q range. While the crossover to the asymptotic stiff cylinder behavior is located at q
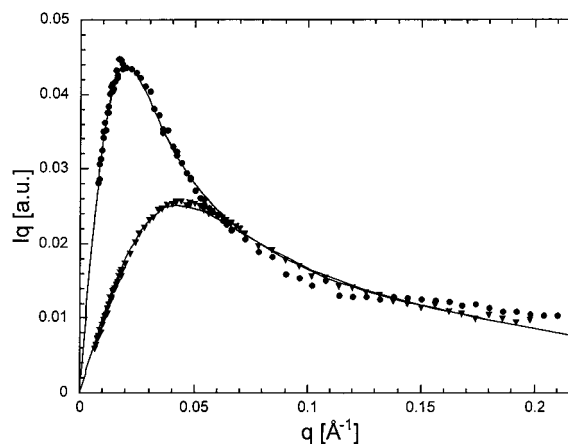

FIGURE 6 Holtzer plot $ql(q)$ vs q for dB1 (▼) and dB4 (●). The fit results for the flexible cylinder model are shown as the solid lines.

$\approx 1.9/l_p \approx 0.15$ Å^{-1}, the following cross-section Guinier region already ends at $q \approx 2/R_{cs} \approx 0.19$ Å^{-1}. This also makes it impossible to obtain detailed information on the shape of the cross-section. However, the temperature dependence of dB1 was studied at D22 up to slightly larger q values. Figure 4B shows this data together with the model scattering intensity calculated, based on the above given values. The very good agreement over the whole q range supports the values determined, especially R_{cs}.

Small-Angle X-Ray Scattering (SAXS)

We also performed SAXS experiments. While the low q region was strongly affected by aggregation, the scattering at higher q values shows the same behavior as observed by SANS. Furthermore, the SAXS experiments extend to higher q values than covered by SANS. Based on the above model and the values obtained by fitting the SANS data, the model scattering function can be calculated. The calculated intensities nicely agree with the measured SAXS data (Figure 7) and therefore confirm the values determined by SANS, in particular the cross-sectional radius R_{cs}.

DISCUSSION

Flexibility of the Cylinder

The scattering data are in agreement with a structural model of a semiflexible cylinder. The asymptotic behavior of the scattering data in limited q ranges was analyzed but, in addition, we applied a single model
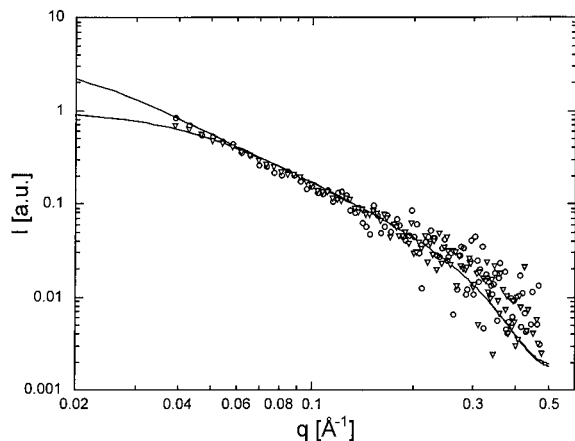
FIGURE 7 SAXS intensities $l(q)$ vs magnitude of the scattering vector q for dB1 ($\triangledown$) and dB4 ($\bigcirc$). The solid lines are calculated based on the SANS results. The samples were prepared by dialyzing the purified protein solutions in a buffer containing 20 mM acetate and 3 mM EDTA. The protein concentration and pH were 14.6 mg/mL and 4.6 (dB1), and 0.9 mg/ml and 4.4 (dB4), respectively.

function capable of describing the scattering data over the entire q range, which corresponds to a broad range of length scales. The best fit was obtained for flexible cylinders with length 235 and 893 Å for dB1 and dB4, respectively, a cross-sectional radius $R_{cs} = 7.4$ Å, and a flexibility quantified by a persistence length $l_p = 12.6$ Å. The scattering data of both, dB1 and dB4, can be fitted with identical l_p and R_{cs}, which reflects the fact that they both have the same structure, except that dB4 consists of basically four dB1 units and is thus longer. This implies that the structure of the repetitive domain is maintained and suggests that the central domain can be elongated while the basic structural properties are retained. The structure of a flexible cylinder is an extension of the (stiff) rod model, which was applied in earlier solution studies to analyze scattering and viscosimetric data.[13,17,18] On one hand, the cylindrical, rodlike shape is retained, while on the other hand, our scattering data clearly indicate a high degree of flexibility of the cylinder. This is most evident from the Holtzer plot (Figure 6), where a stiff cylinder would not show a maximum. The fitted values, especially L and R_{cs}, indicate that the flexibility is not associated with the protein backbone, but the supersecondary structure of the protein.

Length of the Cylinder

A quantitative comparison with previous studies is not straightforward, because different proteins and buffers were used. Nevertheless, there are two other stud-

ies where the lengths of the repetitive domains can be compared. SAXS measurements on HMW Ax1 in 0.1M acetic acid buffer, reported a length of 786 Å.[17] For an estimation of the length of the repetive domain of Ax1, the N- and C-terminal domains were assumed to be globular. Taking a partial specific volume of 0.73 mL/g, their theoretical diameters are 31.2 and 21.6 Å, respectively. With a repetitive domain of 681 residues,[37] the average length per amino acid, L_{AA}, is 1.1 Å. Secondly, viscosimetric studies on HMW Bx7 in 0.05M acetic acid/0.01M glycine reported a length of 504 Å.[13] Taking in account the terminal domains, and a repetitive domain of 645 amino acids,[20] we get here a L_{AA}, of 0.7 Å. Our value of $L_{AA} \approx 1.5$ Å for both dB1 and dB4 is significantly higher and indicates that the central domain, and thus HMW proteins have longer contour lengths than previously reported. A flexible cylinder instead of the assumption of a stiff rod is largely responsible for this difference. Although variation exists between the repetitive domains of Ax1, Bx7, and Dx5, we do not think that they are large enough to explain the differences.

Diameter of the Cylinder

Earlier work using SAXS focused on the low and high q region only. Based on the assumption of a completely stiff rod, the overall size in terms of the radius of gyration R_G or length L of the rod as well as the cross-sectional radius R_{cs} was obtained from Guinier fits to the low and high q data, respectively.[17,18] The values for R_{cs} are, however, significantly different. They observed a three to four times larger R_{cs}, which they attributed to side-to-side aggregation. Our explanation for this discrepancy would be the high flexibility and correspondingly small persistence length l_p, which is close to R_{cs}. As mentioned above, this results in a single, broad crossover region so that the determination of R_{cs} by a Guinier analysis is strongly influenced by l_p and prone to an apparently larger R_{cs}.

β-Spiral

It was proposed that the central repetitive domain of the HMW gluten protein adopts a supersecondary structure of a linear helical organization, called a β-spiral, which is thought to be similar to those suggested for elastin.[5,9,12,13,38,39] The formation of this spiral in gluten has been suggested by modeling[40] as well as by scanning tunneling microscopy (STM) observations.[14] The β-spiral could be stabilized by hydrogen bonds of the glutamine side chains from different windings, but has never been demonstrated
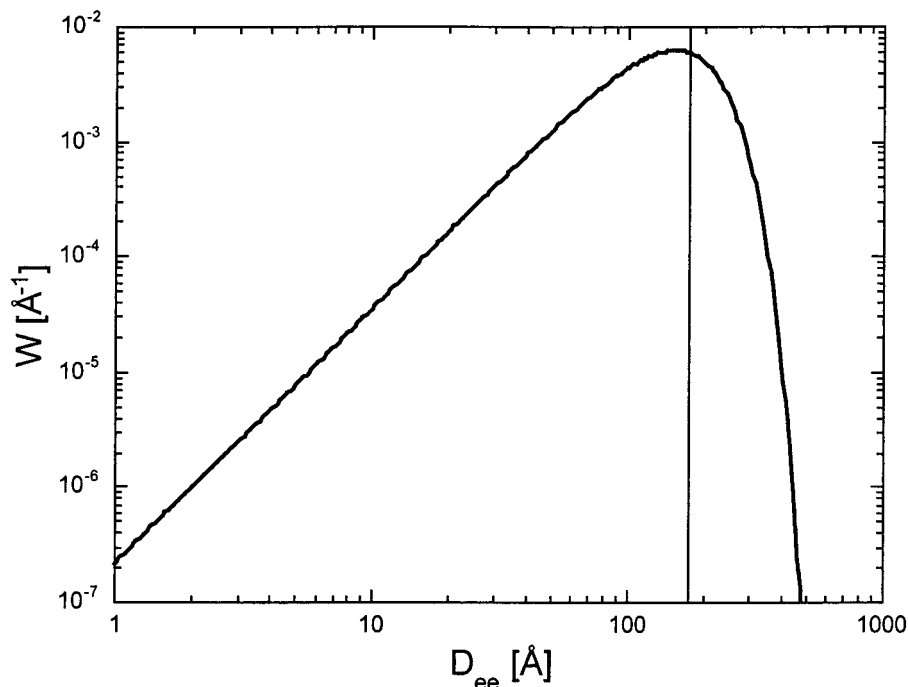
FIGURE 8   Distribution function $W(D_{ee})$ of the end-to-end distance D_{ee} based on the parameters determined for dB4. The vertical line indicates the average end-to-end distance $\{D_{ee}\}$.

experimentally.[15,16,21] Their existence was suggested for peptide analogues of the central domain that were found to stabilize a β-spiral consisting of at least 30 residues.[15] In addition, it is well-established that the repetitive domains consist of β-turns.[12,13,15,16,19,21,41] The number of β-turns present was estimated to be 27 in dB1, and thus 108 in dB4.[16] The flexibility of the cylinder, in combination with the abundant β-turns, makes a highly ordered, spiral structure with uniform radius and pitch very unlikely.

Intramolecular Disulfide Bonds

In HMW proteins, the central domain is flanked by N- and C-terminal domains, which contain several cysteine residues. Intermolecular disulfide bonds between cysteines from different subunits allow the formation of a protein matrix and are thus important in determining their properties.[1,2,42] They compete with intramolecular disulfide bonds that may interfere with the polymerization. Intramolecular disulfide bonds between N- and C-terminal domains were indeed found.[2] This, however, cannot be understood based on the rod-like structure of the central domain, where the N- and C-terminal cysteines are far apart from each other. In contrast, flexibility of the central domain as suggested by our SANS experiments allows the N- and C-terminal domains to approach each

other and therefore intramolecular disulfide bonds can be formed. The proposed model of a flexible cylinder can thus resolve this contradiction.[2]

We can quantify this argument by considering the end-to-end distance D_{ee} of the flexible cylinder. Since it is a very dynamic structure and assumes steadily changing random configurations, one can only specify an average value for the end-to-end distance $\langle D_{ee} \rangle$ and a probability for a certain end-to-end distance $W(D_{ee})$.[24] The distribution $W(D_{ee})$ is shown in Figure 8 for dB4, whose length roughly corresponds to the length of the complete central domain. While the average is $\langle D_{ee} \rangle = 174$ Å, Figure 8 shows that there is also a considerable probability to find small distances D_{ee}. This probability can then be compared with the (very small) Boltzmann factor $\exp(-E_{S\text{-}S}/k_B T)$, where $E_{S\text{-}S} = 86\, k_B T$ is the energy gain in forming a disulfide bond.[43]

Elasticity

It has been proposed that elasticity is related to the formation of long chains of glutenins, which are connected by intermolecular disulfide bonds between the N- and C-terminal domains.[42] In addition, the β-spiral structure of the central domain with its intrinsic elasticity is also suggested to play an important role.[12,44] It is believed that stretching the β-spiral disrupts its

conformation, which is stabilized by hydrogen bonding between glutamine pairs,[16] and results in an energetically less favorable state. After cessation of stress, the stable conformation is restored and results in elastic recoil.

A new component is added by the flexibility of the cylinder. In response to an external force acting on its ends, it stretches and each segment tends to align with the force. It thus no longer has a random configuration, but is restricted to a certain direction and longer end-to-end distances D_{ee} (Figure 8). However, thermal agitation tries to restore disorder of the segments and hence opposes the stretching force. This elastic response is purely entropic. Stress-strain curves of single molecules have been obtained experimentally, as demonstrated by direct mechanical measurements of the elasticity of single DNA molecules.[45]

The proposed model allows us to take full profit of results from polymer theory. The dependence of the end-to-end distance D_{ee} on the external force has been analytically derived.[46–48] and computer simulations have been performed to investigate the full force regime, including the crossover between the linear force and Pincus-scaling regime.[49–51] The consequences for the elasticity of polymer networks have also been investigated theoretically and experimentally.[52] These models assume a constant length of the polymer, which corresponds to the flexible cylinder in our case. The deformability under stress and extensibility of the β-spiral is thus not taken into account. In particular, the intrinsic elasticity along the β-spiral axis adds an extra mechanism, which results in an enthalpic as well as entropic contribution to elasticity.

CONCLUSION

We have performed small-angle scattering experiments to investigate the solution structure of representative parts of the repetitive domain of HMW wheat gluten proteins, dB1 and dB4. The data over the entire q range and for both proteins are in agreement with a structural model of a flexible cylinder, whose parameters were determined. The supersecondary structure of this cylinder is consistent with a β-spiral.[12,13] The fact that both proteins can be described by the same structural model indicates that the structural properties are maintained upon "polymerization", which is corroborated by CD experiments. Compared to the rod-like structure proposed earlier.[13,17,18] the present model adds the important features of flexibility, and a more extended structure for the protein. This has substantial consequences. It allows for intramolecular disulfide bonds between the

N- and C- terminal domains and thus resolves a former contradiction.[2] Furthermore, it introduces a new structural element bearing elasticity and readily provides the theoretical background due to its relation to polymer theory.

MATERIALS AND METHODS

Materials

Restriction enzymes and T4-DNA ligase were obtained from Boehringer. Gel extraction kit was obtained from QIAGEN. Ni^{2+}-agarose was obtained from QIAGEN; isopropyl-D-thiogalactopyranoside (IPTG) was obtained from Boehringer. Tryptone and yeast extract were obtained from Difco. Buffer components were obtained from Sigma.

Cloning of dB1 and dB4

The pQE60dB1 and pSk-dB1 were constructed as described previously by van Dijk et al.[16] The dB1-encoding DNA was purified on 1.5% agarose gel after *Bgl*II/*Bam*HI digestion of pSKdB1. Multimers of the dB1 DNA were obtained after self-ligation of the dB1 DNA monomers. After heat inactivation of T4-DNA ligase, *Bgl*II and *Bam*HI were added to the ligation mixture, selecting dB1 DNA multimers with the correct orientation. The *Bgl*II/*Bam*HI-digested dB1 DNA multimers were purified from agarose gel and ligated into previously cut expression vector pQE60.

Expression and Purification

The dB1 was harvested from the *E. coli* strain SG 13009 (pREP4), transformed with the pQE60dB and induced with isopropyl-D-thiogalactopyranoside.[16] The first purification step was performed by using Ni^{2+}-agarose chelation chromatography; a second purification step involved a preparative HPLC C4 reversed phase column (type Vydac CAT 214TP1022). The column was equilibrated with a 0.1% *TFA* solution for 20 min. The sample was applied and eluted in 0.1% TFA with a 0–40% gradient, increasing at 1% per min, of a solution containing 90% acetonitrile in 0.1% TFA; the elution rate was 8 mL/min; dB1 eluted from the preparative column at approximately 32 min. The purity was checked on an analytical HPLC C4 reversed phase column (type Vydac CAT 214TP54), using the same elution conditions and gradient and an elution rate of 1 mL/min; dB1 eluted from the analytical column at approximately 28 min (see also Figure 2). The conservation of secondary structure of dB1 after HPLC purification was checked with CD in acetate buffer. The methods to express and purify dB4 were similar to those used for dB1.

Sample Preparation

The protein concentrations for the scattering experiments were attained by freeze-drying to the appropriate volumes.

The concentrated samples were dialyzed in demineralized water for removal of the acetonitrile/TFA solution, introduced by the HPLC purification. It was followed by dialysis in the standard acetate/EDTA buffer (20 m*M* acetic acid, 0.5 g/L sodium acetate, 3 m*M* EDTA, pH 4.2) for at least two days and by replacing the dialysis solution two times [1 mL sample in 3×15 mL D_2O (SANS) or $3 \times 1L$ H_2O (SAXS)]. After dialysis, the samples were checked again by analytical HPLC to rule out loss or breakdown of protein. In addition, DLS was used to verify that no aggregates are formed. After the small-angle scattering experiments, samples were collected and a determination of the concentration was made using CD.

Circular Dichroism

All CD measurements were performed in protein solutions with buffer concentrations lower than 2 m*M,* in order to enhance the sensitivity in the wavelength region from 180 to 200 nm. The temperature was 25°C, the bandwidth 2 nm, and the path length 1 mm. The concentration of the samples was determined from their minimum intensity at 198 nm, averaged over 120 s. The values were corrected for the buffer. Wavelength spectra were recorded from 180 to 250 nm, with steps of 1 nm, each averaged over 10 s. Three such spectra were averaged and the data were not smoothed.

Small-Angle Neutron Scattering

SANS experiments were performed at D11 and D22 at the Institut Laue-Langevin in Grenoble, France (http://www. ill.fr/YellowBook/D11/, http://www.ill.fr/YellowBook/D22/). A neutron wavelength of 8 Å and sample-to-detector distances of 1.2, 4, and 13 m at D11 and 2.0 and 14.0 m at D22 were used. All experiments at D22 were performed with the detector offset by 40 cm. The wavelength resolution for both instruments was 10% (full-width-at-half-maximum). The water for normalization and samples were kept in stoppered quartz cells with a path length of 1 and 2 mm, respectively. The raw spectra were corrected for background from the buffer, sample cell, and other sources by conventional procedures. The two-dimensional isotropic scattering spectra were corrected for dead time effects, azimuthally averaged, and corrected for detector efficiency by dividing with the incoherent scattering spectra of water.[53–55]

Small-Angle X-Ray Scattering

SAXS data were collected on the X33 camera of the European Molecular Biology Laboratory on the storage ring DORIS III of the Deutsches Elektronen Synchrotron at Hamburg using multiwire proportional chambers with delay line readout.[56] A wavelength of 1.5 Å was used and a combination of two sample-detector distances allowed us to cover a range of scattering vectors $0.04 < q < 0.5$ Å^{-1}. The raw data were normalized to the intensity of the incident beam, corrected for detector response, and the scattering of the buffer was subtracted.

REFERENCES

1. Shani, N.; Steffen-Campbell, J. D.; Anderson, O. D.; Greene, F. C.; Galili, G. Plant Physiol 1992, 98, 433–441.
2. Shimoni, Y.; Blechl, A. E.; Anderson, O. D.; Galili, G. J Biol Chem 1997, 272, 15488–15495.
3. Payne, P. I.; Corfield, K. G.; Blackman, J. A. Theor Appl Genet 1979, 55, 153–159.
4. Payne, P. I.; Ann Rev Plant Physiol 1987, 38, 141–153.
5. Shewry, P. R.; Halford, N. G.; Tatham, A. S. In Oxford Surveys of Plant Molecular and Cell Biology, Vol. 6; Miflin, B. J., Miflin H. J.; Eds. University Press: Oxford, 1989; pp 163–219.
6. Shewry, P. R.; Halford, N. G.; Tatham, A. S. J Cereal Sci 1992, 15, 105–120.
7. Shewry, P. R.; Miles, M. J.; Tatham, A. S. Prog Bioph Mol Biol 1994, 61, 37–59.
8. Shewry, P. R.; Tatham, A. S.; Barro F.; Barcelo, P.; Lazzeri, P. Biotechnology 1995, 13, 1185–1190.
9. Shewry, P. R. In Seed Development and Germination; Kigel J., Galili, G., Eds.; Marcel Dekker: The Netherlands, 1995; pp 45–72.
10. Sugiyama, T.; Rafalski, A.; Peterson, D.; Söll, D. Nucleic Acid Res 1985, 13, 8729–8737.
11. van Dijk, A. A.; van Swieten, E.; Kruize, I. T.; Robillard, G. T. J Cereal Sci 1998, 28, 115–126.
12. Tatham, A. S.; Shewry, P. R.; Miflin, B. J. FEBS 1984, 177, 205–208.
13. Field, J. M.; Tatham, A. S.; Shewry, P. R. Biochem J 1987, 247, 215–221.
14. Miles, M. J.; Carr, H. J.; McMaster, T. C.; I'Anson, K. J.; Belton, P. S.; Morris, V. J.; Field, J. M.; Shewry, P. R.; Tatham, A. S. Proc Natl Acad Sci 1991, 88, 68–71.
15. van Dijk, A. A.; van Wijk, L. L.; van Vliet, A. A.; Haris, P.; van Swieten, E.; Tesser, G. I.; Robillard, G. T. Protein Sci 1997, 6, 637–648.
16. van Dijk, A. A.; de Boef, E.; Bekkers, A.; van Wijk, L. L.; van Swieten, E.; Hamer, R. J.; Robillard, G. T. Protein Sci 1997, 6, 649–656.
17. Matsushima, N.; Danno, G.; Sasaki, N.; Izumi, Y. Biochem Biophys Res Commun 1992, 186, 1057–1064.
18. Thomson, N. H.; Miles, M. J.; Popineau, Y.; Harries, J.; Shewry, P. R.; Tatham, A. S. Biochim Biophys Acta 1999, 1430, 359–366.
19. van Swieten, E. Ph.D. thesis, Groningen University: The Netherlands, 2001.
20. Anderson, O. D.; Greene, F. C. Theor Appl Genet 1989, 77, 689–700.

21. Tatham, A. S.; Drake, A. F.; Shewry, P. R. J Cereal Sci 1990, 11, 189–200.
22. Guinier, A. Ann Phys (Paris) 1939, 12, 161–237.
23. Kirste, R. G.; Oberthür, R. C. In Small Angle X-Ray Scattering; Glatter, O.; Kratky, O., Eds.; Academic Press: London, 1982; pp 387–431.
24. Pedersen, J. S.; Laso, M.; Schurtenberger, P. Phys Rev E 1996, 54, R5917–R5920.
25. Pedersen, J. S.; Schurtenberger, P. Macromolecules 1996, 29, 7602–7612.
26. Schmidt, M.; Paradossi, G.; Burchard, W. Makromol Chem, Rapid Commun 1985, 6, 767–772.
27. Denkinger, P.; Burchard, W. J Polym Sci B 1991, 29, 589–600.
28. Holtzer, A. J Polym Sci 1955, 17, 432–434.
29. Benoit, H.; Doty, P. J Phys Chem 1953, 57, 958.
30. Rawiso, M.; Duplessix, R.; Picot, C. Macromolecules 1987, 20, 630–648.
31. Kratky, O.; Porod, G. Rec Trav Chim Pays-Bas 1949, 68, 1106–1122.
32. Yoshizaki, T.; Yamakawa, H. Macromolecules 1980, 13, 1518–1525.
33. Pedersen, J. S.; Egelhaaf, S. U.; Schurtenberger, P. J Phys Chem 1995, 99, 1299–1305.
34. Pedersen, J. S.; Posselt, D.; Mortensen, K. J Appl Cryst 1990, 23, 321–333.
35. Pedersen, J. S. J Phys IV 1993, 3, 491–498.
36. Pedersen, J. S. Adv Coll Interface Sci 1997, 70, 171–210.
37. Halford, N. G.; Field, J. M.; Blair, H.; Urwin, P.; Moore, K.; Robert, L.; Thompson, R.; Flavell, R. B.; Tatham, A. S.; Shewry, P. R. Theor Appl Genet 1992, 83, 373–378.
38. Venkatachalam, C. M.; Urry, D. W. Macromolecules 1981, 14, 1225–1229.
39. Urry, D. W. Methods Enzymol 1982, 82, 673–716.
40. Matsushima, N.; Creutz, C. E.; Kretsinger, R. H. Proteins 1990, 7, 125–155.
41. Pézolet, M.; Bonenfant, S.; Dousseau, F.; Popineau, Y. FEBS 1992, 299, 247–250.
42. Ewart, J. A. D. J Sci Food Agric 1977, 28, 191–199.
43. Atkins, P. W. Physical Chemistry, 2nd ed.; University Press: Oxford, 1982.
44. Tatham, A. S.; Miflin, B. J.; Shewry, P. R. Cereal Chem 1985, 62, 405–411.
45. Smith, S. B.; Finzi, L.; Bustamante, C. Science 1992, 258, 1122–1126.
46. de Gennes, P. G. Scaling Concepts in Polymer Physics. Cornell University Press, Ithaca, NY, 1979
47. Pincus, P. Macromolecules 1976, 9, 386–388.
48. Oono, Y.; Ohta, T.; Freed, K. F. Macromolecules 1981, 14, 880–881.
49. Webman, I.; Lebowitz, J. L.; Kalos, M. H. Phys Rev A 1981, 23, 316–320.
50. Wittkop, M.; Sommer, J. U.; Kreitmeier, S.; Göritz, D. Phys Rev E 1994, 49, 5472–5476.
51. Pierleoni, C.; Arialdi, G.; Ryckaert, J. P. Phys Rev Lett 1997, 79, 2990–2993.
52. Cohen Addad, J. P. Physical Properties of Polymeric Gels; John Wiley: New York, 1996
53. Jacrot, B.; Zaccai, G. Bioplolymers 1981, 20, 2413–2426.
54. Wignall, G. D.; Bates, F. S. J Appl Cryst 1987, 20, 28–40.
55. Cotton, J. P. In Neutron, X-Ray and Light Scattering: Introduction to an Investigative Tool for Colloidal and Polymeric systems; Lindner, P., Zemb, T., Eds.; Elsevier/North Holland: Amsterdam, 1991; pp 19–31.
56. Koch, M. H. J.; Bordas, J. Nucl Instrum Methods 1983, 208, 461–469.