Role of N-Linked Glycan in the Unfolding Pathway of *Erythrina corallodendron* Lectin†

Nivedita Mitra,‡ Nathan Sharon,§ and Avadhesha Surolia*,‡

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India, and Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Received July 7, 2003

ABSTRACT: *Erythrina corallodendron* lectin (ECorL) exhibits an exquisitely structured oligosaccharide chain. Interestingly, the bacterially expressed, nonglycosylated counterpart, *r*ECorL, possesses an essentially identical carbohydrate specificity and agglutinating activity as the glycosylated lectin, thus suggesting that the overall structure of the two are identical. This paper reports the unfolding behavior of *E. corallodendron* lectin in its glycosylated (EcorL) and nonglycosylated (*r*ECorL) forms. ECorL shows a two-state unfolding pattern during isothermal melts and differential scanning calorimetry (DSC). The T_g of ECorL as obtained from isothermal melts is 74 °C at pH 7.4. The T_p obtained from DSC studies is between 74.8 to 68.1 °C in the pH range of 5.26−7.77. The recombinant lectin (*r*ECorL), which is devoid of carbohydrate, shows, in contrast to the glycosylated protein, a non-two-state unfolding profile as measured by both probes mentioned, but the number of intermediates during unfolding could not be ascertained. Simulated annealing on ECorL, with the sugars removed, reveals that the protein Cα backbones overlap, indicating that the overall structure, including the mode of dimerization, of *r*ECorL is insignificantly altered as compared to ECorL. The alterations in the folding behavior of *r*ECorL as compared to that observed in ECorL may be due to the fact that, unlike most other glycoproteins, one of the glycans in ECorL is unusually structured and forms many hydrogen bonds with the protein. It therefore appears that while the covalently linked sugar does not contribute appreciably to the final folded structure of ECorL, it does alter its folding process in a significant manner.

Protein glycosylation is one of the most common and most complex posttranslational processes, involving numerous enzymes and substrates, both donors and acceptors, and involves numerous processing steps. Databases show that as many as 70% of known protein sequences have potential glycosylation sites (*1*). A SWISS-PROT-based study suggested that more than half of all proteins are glycosylated (*2*). Many studies have been conducted to elucidate the contribution of sugar groups in protein folding, structure, and function. In some cases deglycosylation decreases protein stability or slows down or even completely inhibits its refolding (*3*, *4*). For example, gp120, which contains more than 50% carbohydrate by weight, does not reach its final conformation without the glycosylations being intact (*5*). In most cases where glycosylation leads to correct folding, removal of the sugar groups after folding does not affect activity significantly. In the recently solved structure of quercetin 2,3-dioxygenase from *Aspergillus japonicus*, carbohydrate chains seem to stabilize dimerization by interacting with the protein chain of the other monomer (*6*). In some cases removal of the covalently linked glycan seems to have no effect at all on the folding rates or stabilization of the protein (*7*, *8*).

Despite extensive studies in the literature that attempt to explain the role of glycosylation on proteins, to date, to our knowledge, there has been no investigation where the unfolding mechanism is altered, i.e., a change from a two-state mechanism to a multi-state unfolding mechanism or vice versa. The only exception is a report on the small proteoglycan biglycan, which follows a noncooperative unfolding profile, suggesting the presence of stable folding intermediates, but adopts a two-state unfolding pattern when its sugars are removed (*9*).

This paper explores the role of glycosylation on the folding pathway of the glycoprotein *Erythrina corallodendron* lectin (ECorL),[1] a legume lectin. Despite exhibiting nearly overlapping tertiary structures ("jelly roll" motif), these protein molecules show diverse modes of oligomerization (*1*, *10*, *11*).

ECorL is a dimeric protein with a subunit molecular mass of 28,750 Da. It possesses two N-linked oligosaccharide chains attached to Asn 17 and Asn 113 as determined by

† This work has been supported by grants from the Department of Biotechnology, Government of India, and Council of Scientific and Industrial Research, India. VP-DSC used in these studies has been supported by the Departments of Science and Technology (DST), Government of India, to A.S., under its Intensification of Research in High Priority Areas program.

* To whom correspondence should be addressed. Phone: 91-80-29342714. Fax: 91-80-3600535. E-mail: surolia@mbu.iisc.ernet.in.

‡ Indian Institute of Science.

§ Weizmann Institute of Science.

[1] Abbreviations: ECorL, *Erythrina corallodendron* lectin; *r*ECorL, recombinant EcorL; DSC, differential scanning calorimetry; Gdn-HCl, guanidine hydrochloride; WBA I, winged bean basic agglutinin; WBA II, winged bean acidic agglutinin; PNA, peanut agglutinin; GdnCl, guanidine hydrochloride.

10.1021/bi035169e CCC: $25.00
Published on Web 10/03/2003

mass spectrometric studies (*12*). Only the glycan attached to Asn 17 has been observed in the different X-ray crystal structures obtained to date (*13*, *14*). The oligosaccharides are largely in the form of the heptasaccharide Manα6-(Manα3)(Xylβ2)Manβ4GlcNAcβ4 (L-Fucα3)GlcNAcβ. Some of the protein molecules do not have fucose in the attached glycan, but the rest of the carbohydrate chain is the same as above. This contributes to heterogeneity in the glycan (*15*).

ECorL has one Mn^{2+} ion and one Ca^{2+} ion per subunit, which are required for sugar binding. The quaternary structure of ECorL, which exhibits a handshake mode of association, is different from that of the canonical dimer found in concanavalin A (Con A), lentil lectin, pea lectin, etc. The canonical dimer has an extensive intersubunit interface area of 1000 Å^2 as compared to only 700 Å^2 at the ECorL interface (*13*).

The bacterially expressed recombinant ECorL (*r*ECorL), is a dimer devoid of carbohydrate and has a monomeric molecular mass of 26,309 Da (*16*). However, the hemagglutinating activity and the carbohydrate specificity of *r*ECorL are essentially identical to that of ECorL (*16*, *17*).

This paper reports isothermal denaturation melts and differential scanning calorimetric studies on ECorL and *r*ECorL. The ΔC_p of unfolding of ECorL, which is a measure of the hydrophobic buried surface area in the protein, is comparable to that of winged bean acidic agglutinin (WBA II) (*18*). Both proteins show a high level of similarity in their tertiary and quaternary structures (*13*, *14*, *19*) and, thus, are expected to have similar folding characteristics. Yet we observe that the unfolding of ECorL in the presence of 5 mM Mn^{2+} and Ca^{2+} is a two-state process. In contrast, the recombinant protein unfolds via intermediate(s), even at high concentrations of Mn^{2+} and Ca^{2+}.

EXPERIMENTAL PROCEDURES

Materials. *N*-(2-Hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (HEPES) and 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) were from Sigma Chemical Co. Guanidine hydrochloride was from United States Biochemicals. The other reagents were of the highest purity available. Stock solutions of guanidine hydrochloride (Gdn-HCl) were prepared in 5 mM HEPES buffer, pH 7.4, and the molarity of Gdn-HCl was determined by refractive index (*20*).

Lectins. ECorL was isolated from *E. corallodendron* seeds by affinity chromatography (*21*). The *r*ECorL clone in the pET 3d vector was transfected into *Escherichia coli* BL21-(DE3)pLysS cells. The cells were grown as described, and the protein was purified from inclusion bodies by unfolding in 6 M urea followed by refolding (*16*). An additional step of affinity purification was done using a lactosyl Bio-Gel column to ensure the presence of only the active, refolded molecules in the final *r*ECorL preparation. Protein concentration was determined from its specific extinction coefficient, ECorL $A^{1\%}_{280} = 15$ (*22*).

Isothermal Guanidine Hydrochloride-Induced Denaturation. Equilibrium unfolding, as a function of Gdn-HCl concentration, was monitored by fluorescence spectroscopy and far-UV CD. Fluorescence measurements were done on a JASCO FP777 spectrofluorometer in a 1 cm cell connected to a Julabo (Germany) circulation water bath. The excitation and the emission wavelengths were fixed at 280 and 321 nm, respectively. The excitation and the emission slit widths were 3 and 5 nm, respectively. Each measurement was an average of three readings. Protein concentration used for isothermal melts in the fluorometer was 0.7 μM. Twelve isothermal Gdn-HCl-induced denaturation profiles were collected in the temperature range of 280−320 K, using a Julabo water bath to maintain the sample temperature. The buffer used was 5 mM HEPES, 5 mM Mn^{2+} and Ca^{2+}, and 150 mM NaCl, pH 7.4.

Far-UV CD measurements were made at 225 nm in a 0.1 cm path length cuvette on a Jasco spectropolarimeter 715 attached to a Peltier PTC-348 WI. The protein concentration used was 2 μM. Spectra were collected at a scan speed of 50 nm min^{-1}. Each data point was an average of four accumulations.

Reversibility was checked by fluorescence by refolding the protein that had been unfolded at 6.8 M Gdn-HCl. The final protein concentration after refolding was 0.7 μM, and that of Gdn-HCl was 0.4 M or less.

Differential Scanning Calorimetry. Differential scanning calorimetry (DSC) measurements were performed on a Microcal VP-DSC scanning microcalorimeter, which consists of two fixed 0.5073 mL cells, a reference cell, and a sample cell. The scan rate used was 20 K/h. ECorL scans were performed as a function of pH from 5.26 to 7.77. HEPES buffer was used at the higher pH range (6.5−7.77) and acetate buffer at the lower pH range (5.26−6.2). Both the acetate and the HEPES buffers had 150 mM NaCl and 5 mM Mn^{2+} and Ca^{2+} ions. The buffers in the sample and the reference cells were matched exactly by extensively dialyzing the protein against the required buffer and using the dialysate in the reference cell. Data were analyzed using the Origin version 5.0 software provided by Microcal Inc.

Light Scattering. Light scattering experiments were performed as a function of time, protein concentration, and guanidine concentration. Scattering experiments were done by measuring the absorbance of samples at 360 nm. The presence of aggregates would scatter the incident light, which would lead to an increase in absorbance. The experiments were done for both ECorL and *r*ECorL. Scattering was measured as a function of Gdn-HCl concentration (0, 2, 5, and 7 M) at two different protein concentrations (4 and 20 μM). The samples were incubated at 25 °C, and absorbance readings were taken at intervals of 0, 2, and 6 h.

Another experiment was done as a function of protein concentration (1, 5, 10, 15, and 20 μM) in the absence of any Gdn-HCl. The samples were again incubated at 25 °C, and absorbance readings were taken at intervals of 0, 2, 4, and 6 h.

Gel Filtration. Gel filtration studies were done using a Sephacryl S-100 10/30 column attached to a Pharmacia FPLC system. The sample volume injected into the column was 200 μL of 17 μM protein. The column was preequilibrated with 5 mM HEPES buffer with 150 mM NaCl and 5 mM Mn^{2+} and Ca^{2+} ions, containing the required concentration of Gdn-HCl, pH 7.4 at room temperature. The flow rate was 0.1 mL/min, and eluent was detected on-line by absorbance at 280 nm.

RESULTS

Isothermal Denaturation Studies. Equilibrium denaturation profiles, in which the Gdn-HCl concentration was raised from

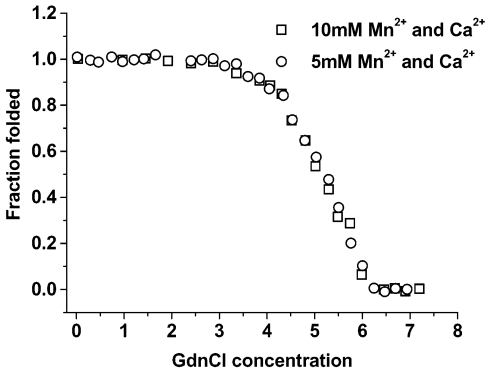

FIGURE 1: Gdn-HCl-induced unfolding of ECorL in the presence of 5 and 10 mM Mn^{2+} and Ca^{2+}. There is no difference in the stability profiles obtained at either ion concentration. The experiment was performed at 25 °C with 0.7 μM monomeric protein concentration. The unfolding was monitored using fluorescence.

0 to 8 M in increments of 0.2−0.4 M, were monitored using fluorescence and far-UV CD. When excited at 280 nm, native ECorL shows a fluorescence emission maximum at 328 nm, which shifts to 353 nm on denaturation with 8 M Gdn-HCl. The difference spectrum shows a maximum at 321 nm, and this was chosen as the emission wavelength in the isothermal denaturation melts. Also, the refolding efficiency after unfolding the protein with Gdn-HCl was close to 85−90% as measured by fluorescence emission scans.

In 0 and 1 mM Mn^{2+} and Ca^{2+}, pH 7.4, ECorL did not show a monophasic equilibrium denaturation profile at some of the temperatures examined. But when 5 mM Mn^{2+} and Ca^{2+} were added to the buffer, the curves became two-state at all temperatures at which unfolding was studied. Under these conditions the protein displayed reversible unfolding indicating that Mn^{2+} and Ca^{2+}, which are essential for the activity of the protein, are also involved in modulating its folding. Using 10 mM $MnCl_2$ and $CaCl_2$ did not show any significant difference in the equilibrium denaturation profile, and therefore further experiments were done in the presence of 5 mM $MnCl_2$ and $CaCl_2$ (Figure 1). It is also seen that the unfolding reaction is dependent on the concentration of the protein, with stability increasing with increase in concentration (Figure 2). Earlier DSC studies on ECorL showed an increase in T_p (peak maxima of the DSC scan) with an increase in protein concentration (*23*). Thus the protein displayed increased stability with increasing concentration in both thermal- and denaturant-induced unfolding, indicating that a dimer to monomer transition occurs during the unfolding reaction. Thus, in ECorL the folding reaction starts with folded dimers (A_2) and ends with unfolded monomers (U). Depending on the stabilities of the dimer and the folded monomer, there might be different pathways in this folding reaction. For example, the dimeric lectin (A_2) can initially give rise to the folded monomers which then unfold completely, i.e., $A_2 \leftrightarrow 2A \leftrightarrow 2U$, or it can result in the formation of unfolded dimers followed by their dissociation to the denatured constituent polypeptide chain, i.e., $A_2 \leftrightarrow U_2 \leftrightarrow 2U$.

In both cases the final unfolded monomer is obtained through an intermediate, but in ECorL we see the presence of only two species in the unfolding reaction; i.e., its denaturation can be described by a two-state process consisting of dissociation of the folded dimer to the unfolded

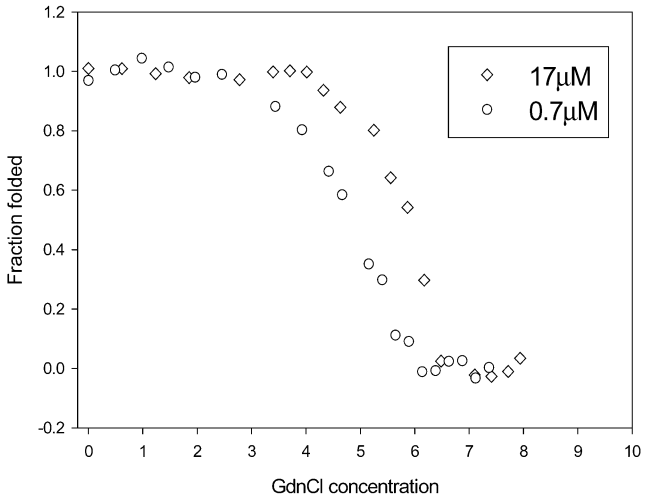

FIGURE 2: Effect of protein concentration on the unfolding of ECorL. The experiment was done at 25 °C with 5 mM Mn^{2+} and Ca^{2+}. The unfolding was monitored using fluorescence.

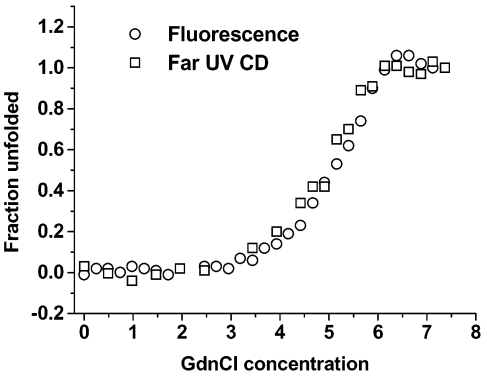

FIGURE 3: Profile showing the complete overlap of denaturation profiles of ECorL obtained by far-UV CD and fluorescence. The monomeric protein concentration used was 2 μM. The experiment was performed at 25 °C.

monomers, i.e., $A_2 \leftrightarrow 2U$. The superimposition of denaturant-induced ECorL unfolding profiles, obtained from measuring the intrinsic fluorescence change and the far-UV CD spectrum change, supports the existence of two species in the unfolding reaction of ECorL (Figure 3). Both spectroscopic probes, one of which is an indication of the secondary structure (far-UV CD) and the other that of the tertiary structure (fluorescence), show a similar loss in protein structure, thus ruling out the presence of an intermediate.

The equilibrium constant (K_u) of the reaction can be obtained from the equation:

$$K_u = [2U]^2/[A_2] = 2P_t[f_u^2/(1 - f_u)] \quad (1)$$

where f_u is the fraction of unfolded protein at a given Gdn-HCl concentration ([D]) and P_t is the total monomeric protein concentration. f_u is given by

$$f_u = Y_0 - (Y_f + m_f[D])/(Y_u + m_u[D]) - (Y_f + m_f[D]) \quad (1a)$$

Y_0 is the observed spectroscopic property (e.g., fluorescence intensity), Y_f and m_f are the intercept and the slope of the folded baseline of the unfolding curve, and Y_u and m_u are the intercept and the slope of the unfolded baseline.

The exposure of hydrophobic surfaces during the unfolding reaction leads to a large positive change in heat capacity (ΔC_p), which is independent of temperature in the range of measurement (*24*). This parameter provides the temperature dependence of ΔG, ΔH, and ΔS of unfolding:

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

$$\Delta H_T = \Delta H_g + \Delta C_p(T - T_g) \quad (3)$$

$$\Delta S_T = \Delta S_g + \Delta C_p \ln(T/T_g) \quad (4)$$

$$\Delta G_T = \Delta H_g(1 - T/T_g) + \Delta C_p[(T - T_g) - T \ln(T/T_g)] \quad (5)$$

T_g, the midpoint of thermal transition, corresponds to the temperature at which $\Delta G = 0$. ΔH_g and ΔS_g are the values of ΔH and ΔS at T_g.

We use the linear free energy model to calculate ΔG_{H_2O}, which is the free energy of unfolding in water, from the free energies of unfolding in the presence of Gdn-HCl (*25*−*28*). According to this model ΔG, ΔH, ΔS, and ΔC_p are linearly dependent on the molar denaturant concentration [D] in the transition region, i.e.

$$\Delta G_u = \Delta G_{H_2O} + m[D] \quad (6)$$

$$\Delta S_u = \Delta S_{H_2O} + s[D] \quad (7)$$

$$\Delta H_u = \Delta H_{H_2O} + h[D] \quad (8)$$

$$\Delta C_{pu} = \Delta C_{pH_2O} + c[D] \quad (9)$$

where m, h, s, and c are the slopes of the straight lines formed by the respective equations.

The ΔG_u of unfolding at each denaturant concentration in an isothermal denaturation profile is calculated using the equation:

$$\Delta G_u = -RT \ln K_u \quad (10)$$

where K_u is obtained from eq 1.

The linear region (transition zone of the melt) of the ΔG_u vs [D] plot is extrapolated to [D] = 0. The slope of this curve is m, and the Y intercept is ΔG_{H_2O}. The values of ΔG_{H_2O}, m, and C_m (the denaturant concentration at which the concentration of folded dimers is equal to the concentration of unfolded monomers) are shown in Table 1. ΔG_{H_2O} is plotted against temperature in Figure 4. The plot of ΔG vs T is defined as the stability plot of the protein (*28*). From Table 1 it is seen that the value of m is independent of temperature. The ΔC_p of unfolding, a parameter directly proportional to the amount of buried surface area exposed upon unfolding, was experimentally determined by isothermal melts to be 3 kcal mol^{-1} K^{-1}. This number was much smaller than the ΔC_p calculated from the change in accessible surface area (ΔASA) upon unfolding of the protein, which was found to be 7.9 kcal mol^{-1} K^{-1}. The PDB coordinates used were those of 1ax1 (*14*). The calculation of ΔC_p from ΔASA is done using the equation given by Spolar et al. (*29*) [$\Delta C_p = (0.32 \pm 0.04)\Delta ASA_{np} - (0.14 \pm 0.04)\Delta ASA_{pol}$], where the apolar and polar contributions to ΔASA are taken separately into consideration. ΔASA_{np} and ΔASA_{pol} were calculated using NACCESS (*30*). The ΔC_p values were close to those obtained for WBA II (3.7 kcal mol^{-1} K^{-1} experimentally and 7.3 kcal mol^{-1} K^{-1} from ΔASA), another protein closely related in structure to ECorL (*18*) but differing in the sites of glycosylation (*19*).

Table 1: Thermodynamic Parameters Obtained for ECorL Unfolding from Isothermal Melts[a]

temp (K)	m (kcal mol^{-1} M^{-1})	C_m (M, Gdn-HCl)	ΔG_{H_2O} (kcal/mol)
280.0	−2.00	6.52	20.91
284.0	−2.09	6.54	21.64
290.0	−2.42	6.27	23.31
292.0	−2.70	6.11	24.69
297.0	−2.27	5.66	21.15
299.0	−2.16	5.62	20.53
301.7	−1.65	5.16	16.96
305.2	−1.44	4.27	14.69
307.0	−1.60	4.41	15.66
309.0	−1.81	3.98	15.88
310.0	−1.85	4.05	16.19
313.5	−1.54	3.12	13.60
314.0	−1.56	3.37	14.06
318.0	−1.83	2.67	13.80

[a] The table shows the m and C_m values obtained from different isothermal melts of ECorL. The melts were obtained by fluorescence measurements. As can be seen from the table, the m values do not vary much with temperature. The C_m values decrease with increase in temperature. The m values were calculated from the individual melts done at different temperatures. The average m calculated from the m values obtained was used for fitting the unfolding profile at each temperature. The ΔG_{H_2O} values reported were obtained from these fits.

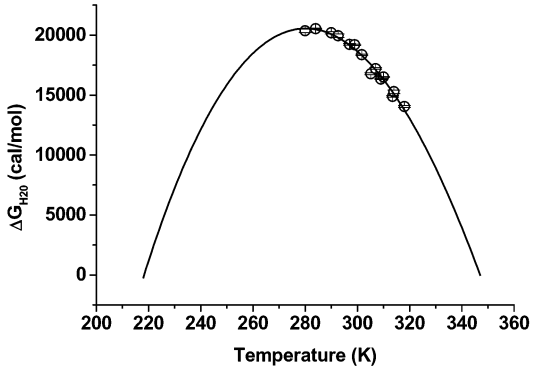

FIGURE 4: Stability plot of ECorL obtained from denaturant-induced isothermal unfolding experiments done at different temperatures as indicated in Table 1. The continuous line shows the least-squares fit of the points to eq 5. The data points (open circles) have been obtained using fluorescence signals. The parameters obtained from the fit are $T_g = 347 \pm 2.5$ K, $\Delta C_p = 2.8 \pm 0.4$ kcal mol^{-1} K^{-1}, and $\Delta H_g = 206.1 \pm 13.8$ kcal/mol. T_g is the higher temperature where $\Delta G_{H_2O} = 0$ and ΔC_p = difference in specific heat between the unfolded and the folded protein molecule.

The recombinant protein though showing fluorescence emission spectra in the native and the completely denatured conditions similar to that of ECorL; it exhibits major deviations in its folding pathway. The isothermal melts are completely noncooperative even in salt concentrations as high as 40 mM Mn^{2+} and Ca^{2+} (Figure 5). In the absence of any Mn^{2+} and Ca^{2+} ions a well-defined intermediate was observed. This non-two-state behavior prevents a thermodynamic evaluation of *r*ECorL unfolding.

Differential Scanning Calorimetry. ECorl and *r*ECorL were examined by DSC to determine differences in their

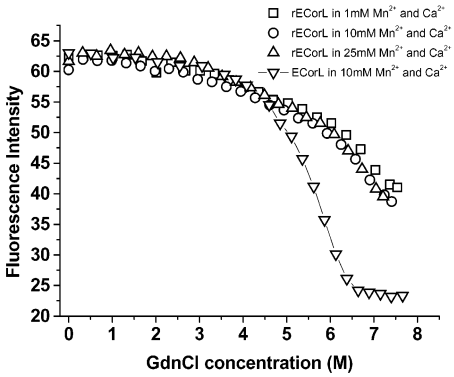

FIGURE 5: Unfolding profiles of *r*ECorL in various salt concentrations. The protein concentration used is 0.7 μM. The experiment was performed at 25 °C. The data points have been obtained using fluorescence. There is insignificant difference between the three profiles. Superimposed is the unfolding profile of ECorL at 24 °C.

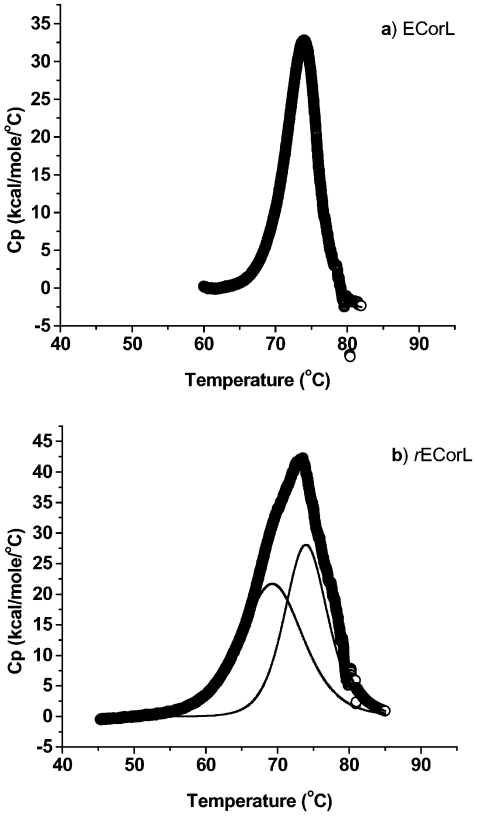

FIGURE 6: DSC unfolding profiles of ECorL (a) and *r*ECorL (b). ECorL shows a single peak. *r*ECorL shows at least two peaks. The protein concentration used is 15 μM for both proteins, and scans have been done at pH 7.4 at 20 K/h. Because of the aggregation accompanying DSC scans in both proteins, the data could not be evaluated. For the sake of representation, the DSC profile of ECorL was fitted using the dissociation two-state model, and that of *r*ECorL was deconvoluted using the non-two-state model. The raw data are represented by circles and the fits by lines. The T_{p1} and T_{p2} for *r*ECorL are 69.2 and 73.9 °C, and the T_p for ECorL is 73.8 °C.

thermal unfolding behavior (Figure 6). The scans of ECorL were performed as a function of pH to derive the ΔC_p of unfolding. It was observed that on decreasing the pH the stability of the protein increased as deduced by an increase in T_p. Thermal unfolding reactions were found to irreversibly aggregate the protein at all pH values tested. Hence, we decided not to evaluate any thermodynamic parameters for these experiments.

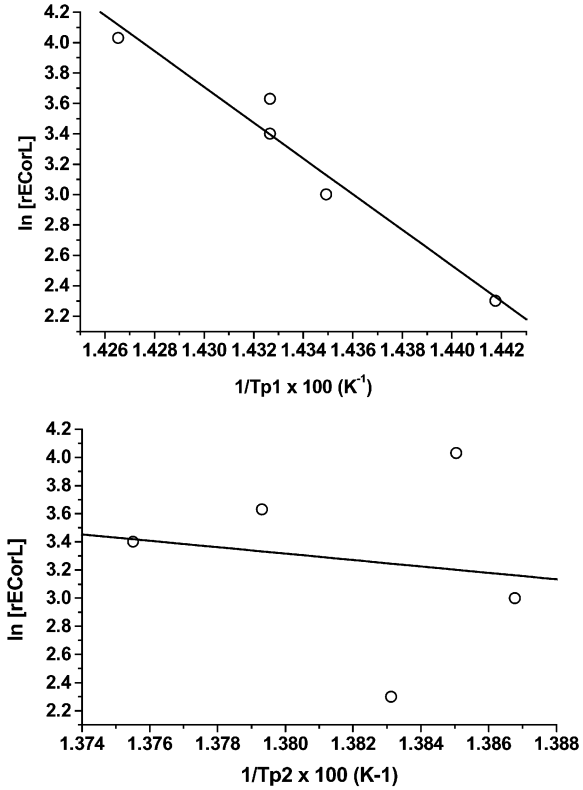

FIGURE 7: Plot of ln of protein concentration vs the reciprocal of the transition peak temperature obtained from DSC scans of *r*ECorL. Concentrations between 15 and 75 μM (monomeric) protein were used. In case of the first peak there is a significant change in T_{p1} with change in concentration of the protein. T_{p2} does not change much when the protein concentration is increased. This suggests that the first peak is a dimer to monomer transition and the second peak is a folded monomer to unfolded monomer transition. The data points have been represented by open circles and the linear regressions by straight lines.

To confirm the nature of the unfolding behavior of *r*ECorL, DSC scans were done on the protein at 0, 1, 5, and 10 mM Mn^{2+} and Ca^{2+}. All the scans showed two peaks, which further validated the absence of two-state unfolding behavior in *r*ECorL (Figure 6b). At 5 mM Mn^{2+} and Ca^{2+}, the first peak occurred nearly 4.7 K before the second one. The second peak occurred at a position close to that of the ECorL peak at the same pH. The fact that *r*ECorL begins melting before ECorL highlights that the former is less stable, again underscoring the differences in their thermal unfolding behavior. The two peaks probably corresponded to the two stages in the unfolding process of *r*ECorL. The first step could be dissociation of the dimers into two nearly completely folded monomers, followed by their dissociation or vice versa. To test this, DSC scans were done at different protein concentrations. It was observed that with increase in protein concentration the first peak (T_{p1}) occurs at increasingly higher temperatures but the second peak (T_{p2}) remained constant. This is depicted as a $1/T_p$ vs ln[protein] plot where the slope of the plot for the second peak is very low as compared to that for the first peak (Figure 7). Thus, the first transition is a concentration-dependent step, indicating the separation of the dimer into monomers, and the second peak is a concentration-independent step where the monomer unfolds. It should be pointed out here that the *r*ECorL DSC scans also showed irreversibility and aggregation, which might have led to underestimation of T_{p2} values.

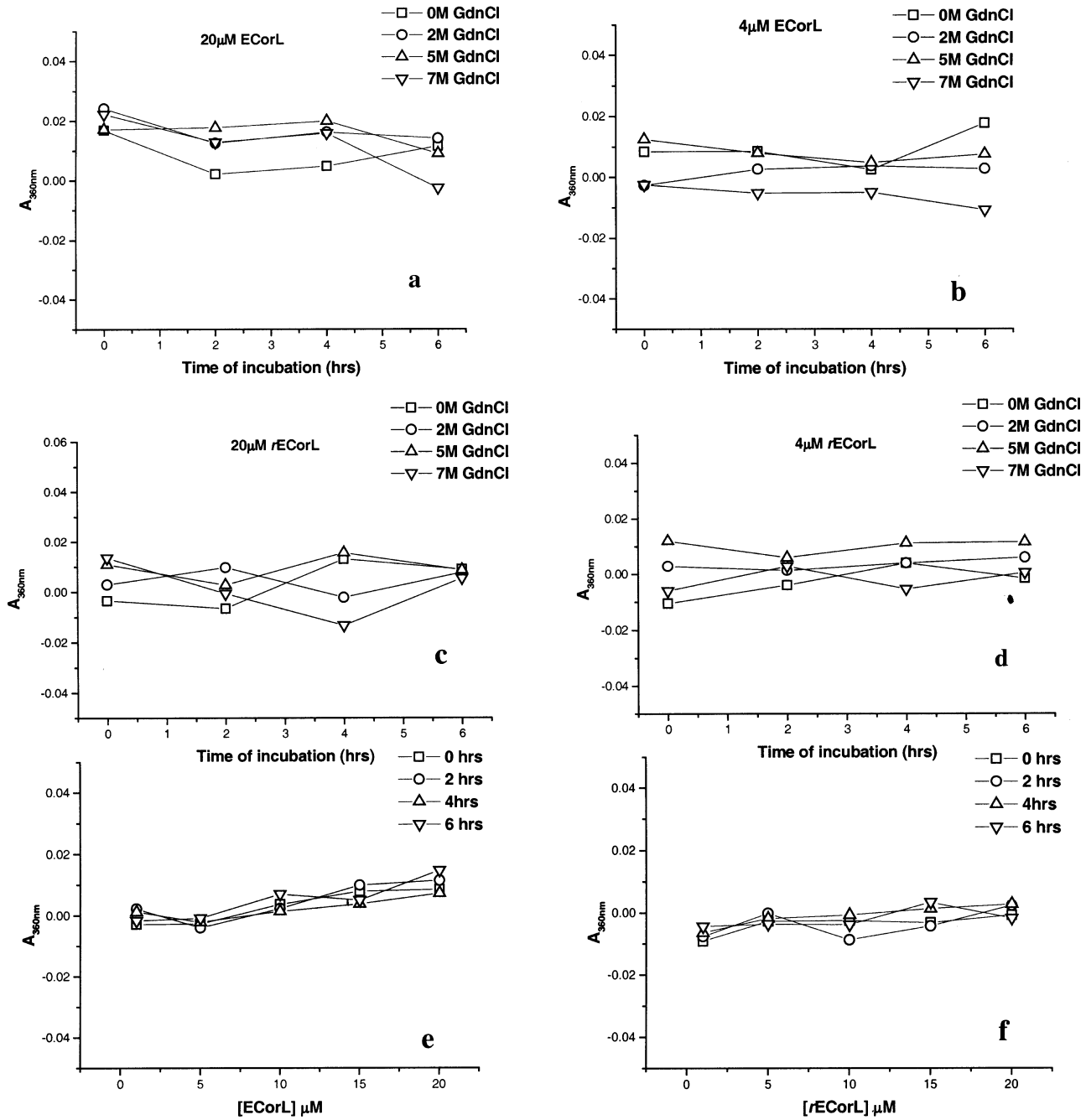

FIGURE 8: Panels a and b show the light scattering experiments done on ECorL as a function of Gdn-HCl concentrations. The panels show absorbance readings at 0, 2, 4, and 6 h after incubation at 25 °C. Panels c and d show similar readings for *r*ECorL. Panels e and f show absorbance readings at different concentrations of the protein (1, 5, 10, 15, and 20 μM) under native conditions taken at 2, 4, and 6 h after incubation at 25 °C.

Although Figure 6b showing the *r*ECorL scan was deconvoluted using the non-two-state method to show the two peaks obtained, it was not used to obtain thermodynamic data because of aggregation that occurred. Aggregation leads to incorrect estimation of the unfolded baseline. To rule out that the two peaks obtained in DSC scans of *r*ECorL were a result of refolding (a step in the purification process of the recombinant lectin from inclusion bodies), ECorL was unfolded and refolded and passed through an affinity column in a manner identical to *r*ECorL, and a DSC scan was performed. Refolded ECorL showed a DSC profile similar to that of native ECorL. Thus the non-two-state folding of *r*ECorL is intrinsic to the protein.

Light Scattering Experiments. Light scattering experiments were performed as mentioned in the Experimental Procedures section. These experiments were done to rule out the possibility that aggregation occurred during denaturant induced unfolding process of the recombinant protein. As a comparison, the experiments were also done on ECorL. No aggregation was observed under any of the experimental conditions. The results have been shown in Figure 8. As can be seen from Figure 8a−d, incubation with Gdn-HCl did not induce aggregation in either protein. Another experiment was done where aggregation was measured as a function of protein concentration (Figure 8e,f). No evidence of protein aggregation was seen in this case either. These experiments

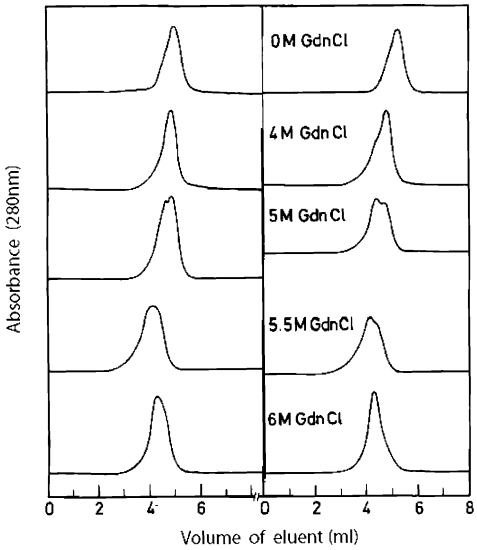

FIGURE 9: Gel filtration elution profiles of ECorL and *r*ECorL. The panel on the left shows the wild-type protein profile, and the right panel shows the recombinant protein. The protein concentration used is 17 μM, and a flow rate of 0.1 mL/min is used. The Gdn-HCl concentrations used are indicated in the figure.

thus rule out the possibility that any nonspecific aggregation occurs in *r*ECorL and ECorL.

Gel Filtration. Gel filtration of protein samples was carried out at different concentrations of Gdn-HCl (Figure 9). It is seen that the recombinant protein elutes after the wild-type protein. This probably results from the lowering of its hydrodynamic radius due to the absence of the carbohydrate, which protrudes out of the protein structure in the wild-type protein. Up to 4 M Gdn-HCl ECorL shows a single peak. At the same concentration of Gdn-HCl a slight shoulder, corresponding to the unfolded monomer, appears in the recombinant protein elution profile. Thus, ECorL displays greater stability as compared to the recombinant protein. Also, the folded peaks of the two proteins are much closer now, unlike that observed in the absence of denaturant.

At 5 M both proteins show two peaks, one corresponding to the folded dimer and the other to the unfolded monomer. The peaks in both proteins have similar elution volume. At 5 M and above a single peak is observed in the case of wild-type protein, but the recombinant protein continues to show two peaks. At 6 M Gdn-HCl both proteins show a single peak corresponding to the unfolded monomer. Thus we see that the wild-type protein shows just two kinds of species during unfolding, but the recombinant form shows three species detectable by gel filtration. But at no point did we see any peak that would suggest the presence of an aggregate.

DISCUSSION

One of the glycans in *E. corallodendron* lectin is unusually structured, being clearly visible in electron density maps of the protein (*13*). Moreover, on closer examination it is seen that these sugar groups form extensive hydrogen bonds within themselves and also to the main chain and side chains of the monomer to which the glycan is attached covalently. Using the program CONTACT (*31*) on the three-dimensional structure of the lectin (1ax1; *14*), six hydrogen bonds were calculated between the covalently bound heptasaccharide and the protein atoms. All of these interactions are shown in Figure 10. Although the initial paper describing the ECorL structure mentioned these and some more hydrogen bonds between the glycan and the protein, yet the detailed atom-wise description was not discussed (*13*).

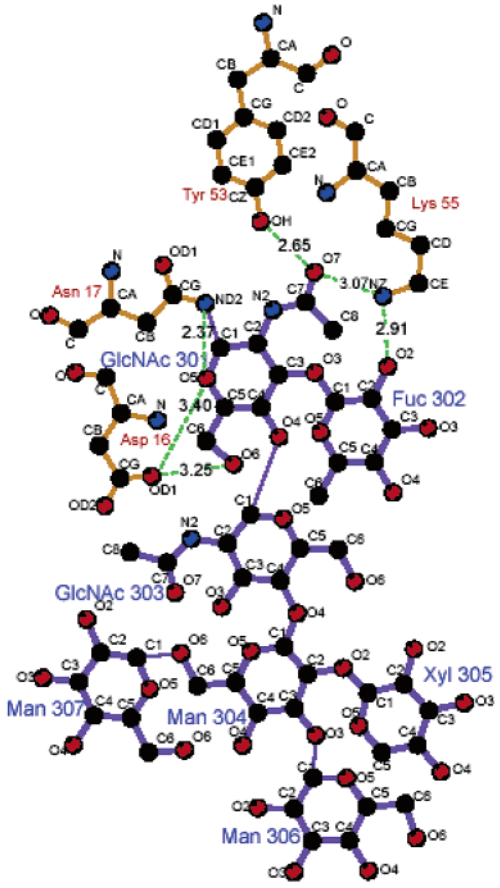

FIGURE 10: Representation of the possible hydrogen bonds (segmented lines) between the covalently bound heptasaccharide moiety and ECorL. This representation has been obtained using LIGPLOT. A similar hydrogen-bonding pattern was also obtained using CONTACT from the CCP4 suite of programs. The sugar names have been abbreviated as follows: *N*-acetylglucosamine, GlcNAc; mannose, Man; xylose, Xyl; fucose, Fuc. Although the initial paper describing the ECorL structure mentioned these and a few more hydrogen bonds between the glycan and the protein, yet the detailed atom-wise description was not discussed (*12*).

Additionally, water-mediated interactions were also studied. Six out of the seven residues in the heptasaccharide *N*-glycan are involved in eight water-mediated interactions with the protein molecule as listed in Table 2. This suggests that the loss of this carbohydrate moiety might significantly affect the final structure and/or the folding pathway of the protein.

Retention of carbohydrate specificity in *r*ECorL, its dimeric nature and agglutinating activity similar to that of ECorL, reiterates the fact that the recombinant protein has tertiary and quaternary structure similar to that of its seed-derived counterpart (*16*).

When the structure of ECorL was first reported (*13*), it was suggested that the handshake association of the protein monomers, unlike the canonical structure found in concanavalin A, was due to its N-linked glycan which sterically prevented the canonical dimerization. Subsequently, structures of two legume lectins, winged bean acidic and basic agglutinins (WBA I and II) were reported, both of which

Table 2: Water-Mediated Contacts between the Heptasaccharide and ECorL Obtained Using CONTACT[a]

no. of the water molecule in the file 1ax1.pdb[b]	possible amino acid with which the water molecule interacts	possible sugar residues with which the water molecule interacts
500	Tyr 53, Lys 55	Fuc 302
501	Lys 55	GlcNAc 301, Fuc 302, GlcNAc 303, Xys 305
502	Asp 16	GlcNAc 301, Fuc 302
556	Lys 99, Asp 209	Man 306
557	Thr 19	Man 306
612	Lys 55, Asn 17	GlcNAc 301, Xys 305
613	Asp 16, Asn 17	GlcNAc 301,Xys 305
622	Asp 16	GlcNAc 301

[a] See ref *31*. [b] See ref *14*.

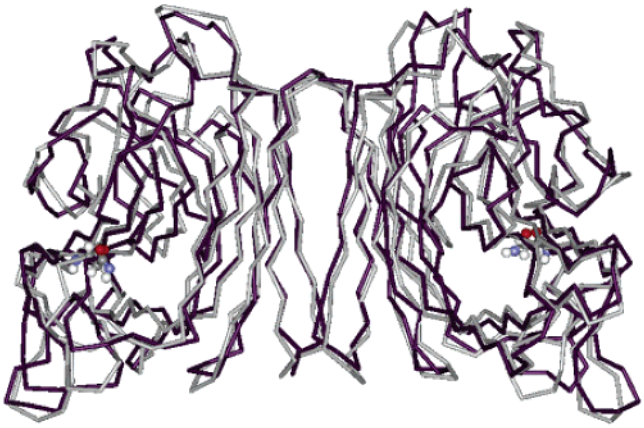

FIGURE 11: Overlay of the structures of ECorl and *r*EcorL derived from simulated annealing studies. The structure shown in gray represents ECorL, and that in black represents *r*ECorL. Asn 17, the site of glycosylation, is depicted as a ball-and-stick representation. The original structure used is that of ECorL (1ax1.pdb). Similated annealing was performed on the original structure and a structure in which the glycans were removed.

had the handshake type of association as seen in ECorL (*19*, *34*). However, in both WBA I and WBA II the glycan does not interfere with the canonical mode of dimerization. This indicated that the quaternary association was a result of the amino acid sequence and not due to location of the carbohydrate group.

To assess the change that occurred after removal of the glycan, the carbohydrate moiety was removed from the protein structure, and simulated annealing was performed on the original and the recombinant structure using the Molecular Operating Environment (MOE; *32*). The force field used was MMFF (*33*). Both the original and the recombinant molecules were heated to 1000 K from 300 K and then cooled slowly to 200 K. Between the two structures an rmsd of only 1.38 Å was observed (Figure 11). The preliminary X-ray data of *r*ECorL show its close similarity to ECorL (K. Kulkarni et al., unpublished data), which is in concurrence with the fact that structures of a number of nonglycosylated forms of proteins are nearly identical to the glycosylated forms (*35*).

Despite similar tertiary and quaternary structures, ECorL and *r*ECorl showed different unfolding pathways and stabilities. The different behaviors did not arise from nonspecific aggregation of the recombinant form as confirmed by the light scattering experiments and also supported by gel filtration studies. The N-linked glycans in ECorL probably play an important role in stabilizing a key segment of the secondary structure, through interactions such as hydrogen bonds, thus affecting the folding pathway.

It has also been speculated that the glycans can act as chaperones and assist protein folding (*36*). In absence of the N-linked glycan, refolding efficiency of soybean agglutinin (SBA) is drastically reduced (*37*). In recombinant SBA, the amount of protein obtained after refolding the inclusion bodies was very low, although the refolded protein showed activity similar to that of native SBA (*38*). Denatured, reduced ribonuclease B (RNase B), a glycosylated protein, regained activity much faster as compared to the nonglycosylated RNase A, which is similar in sequence to RNase B (*39*). In *r*ECorL, efficiency of folding is also affected as evidenced by the low yield obtained upon refolding of the protein from denatured inclusion bodies during purification. From the unfolded protein, the fraction that finally attains the folded structure is only about 10% (*16*) while that of ECorL is close to 85−90%. Most likely, this is attributable to a decrease in solubility of the folding intermediates due to the absence of sugar groups or due to a stabilizing effect of the carbohydrate chain(s) on one or more of the unfolding intermediates. This effect has also been described qualitatively in yeast invertase (*40*).

The first peak in the DSC profile of *r*ECorL, which appears before the ECorL peak, shows that the stability of the recombinant protein is reduced as compared to the glycosylated protein. This could be due to glycan-induced stabilization. One of the most remarkable drops in melting temperature of around 20 °C occurs in a 36-residue peptide PMP-C (*Pars intercerebralis* major peptide C) upon removal of the O-linked fucosyl group from threonine 9 of the peptide (*35*). Using NMR it was shown that, in the nonfucosylated form, PMP-C had a global enhancement of hydrogen exchange rates. In this peptide the fucose moiety is well ordered, being held in position by hydrophobic and hydrogen-bonding interactions. Increased stability has also been observed in the glycosylated RNase B in comparison to the nonglycosylated RNase A (*41*).

Several studies indicate that glycosylation is not essential for activity but is required for proteins to acquire the correct folded form and maintain stability; i.e., after the protein has folded, removal of the carbohydrate does not affect the protein activity but folding studies show reduced stability. This is also supported by the fact that N-linked glycosylation is a cotranslational reaction, thus suggesting a role of these sugar groups in early folding event(s) apart from those required for the quality control machinery of the endoplasmic reticulum.

Thermodynamic parameters of ECorL unfolding were obtanied from isothermal denaturation data. The two-state unfolding pattern of ECorL was confirmed using spectroscopic methods as well as gel filtration experiments. The difference in ΔC_p of ECorL obtained experimentally and theoretically using buried surface area indicated that unfolding did not go to completion. Such differences are usually observed in oligomeric proteins. They probably arise due to incomplete unfolding even at highly denaturing conditions (*42*). DSC scans showed a single peak for ECorL consistent with the mechanism for the folding reaction for this protein. Isothermal denaturation studies on *r*ECorL did not show any well-defined intermediates but displayed distinctly nonsig-

moidal profiles (non-two-state), thus suggesting the presence of several poorly populated intermediates. DSC scans of the recombinant protein show two peaks at all protein and salt concentrations tested, suggesting the presence of at least one intermediate, thus confirming the conclusion of isothermal denaturation studies. Gel filtration studies also support the non-two-state nature of *r*ECorL unfolding

In conclusion, simulated annealing, fluorescence spectra of the proteins, their dimeric nature, and identical biological activities are consistent with insignificant structural differences between ECorL and *r*ECorL in their native states. However, isothermal melts and DSC studies show that absence of the glycan drastically changes the folding pathway of the protein, from a two-state unfolding pathway to a non-two-state unfolding pathway with one or more intermediates. This probably arises due to the role that sugars play during the initial folding of the molecule. The fact that recovery upon refolding is low in *r*ECorL suggests that carbohydrate groups play a role in maintaining the folding intermediates of the protein suitable for attaining the final folded dimeric state of the protein. This is also consistent with the reduced refolding efficiency of *r*ECorL upon unfolding with denaturant compared to the glycosylated protein.

REFERENCES

1. Vijayan, M., and Chandra, N. (1999) *Curr. Opin. Struct. Biol. 9*, 707−714.
2. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) *Biochim. Biophys. Acta 1473*, 4−8.
3. Grafl, R., Lang, K., Vogl, H., and Schmid, F. X. (1987) *J. Biol. Chem. 262*, 10624−10629.
4. Mathieu, M. E., Grigera, P. R., Helenius, A., and Wagner, R. R. (1996) *Biochemistry 35*, 4084−4093.
5. Li, Y., Luo, L., Rasool, N., and Kang, C. Y. (1993) *J. Virol. 67*, 584−588.
6. Fusetti, F., Schroter, K. H., Steiner, R. A., van Noort, P. I., Pijning, T., Rozeboom, H. J., Kalk, K. H., Egmond, M. R., and Dijkstra, B. W. (2002) *Structure 10*, 259−268.
7. Schülke, N., and Schmid, F. X. (1988b) *J. Biol. Chem. 263*, 8832−8837.
8. Tams, J. W., and Weilender, K. G. (1998) *FEBS Lett. 421*, 234−236.
9. Krishnan, P., Hocking, A. M., Scholtz, J. M., Pace, C. N., Holik, K. K., and McQuillan, D. J. (1999) *J. Biol. Chem. 274*, 10945−10950.
10. Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E., and Suddath, F. L. (1986) *J. Biol. Chem. 261*, 16518−16527.
11. Hardman, K. D., and Ainsworth, C. F. (1972) *Biochemistry 11*, 4910−4919.
12. Martin Young, N., Watson, D. C., Yaguchi, M., Adar, R., Arango, R., Rodriguez Arango, E., Sharon, N., Blay, P. K. S., and Thibault, P. (1995) *J. Biol. Chem. 270*, 2563−2570.
13. Shaanan, B., Lis, H., and Sharon, N. (1991) *Science 254*, 862−866.
14. Elgavish, S., and Shaanan, B. (1998) *J. Mol. Biol. 277*, 917−932.
15. Ashford, D. A., Dwek, R. A., Rademacher, T. W., Lis, H., and Sharon, N. (1991) *Carbohyr. Res. 213*, 215−227.
16. Arango, R., Adar, R., Rozenblatt, S., and Sharon, N. (1992) *Eur. J. Biochem. 205*, 575−581.
17. Adar, R., Moreno, E., Streicher, H., Karlsson, K. A., Angstrom, J., and Sharon, N. (1998) *Protein Sci. 7*, 52−63.
18. Mitra, N., Srinivas, V. R., Ramya, T. N. C., Reddy, G. B., Ahmad, N., and Surolia, A. (2002) *Biochemistry 41*, 9256−9263.
19. Manoj, N., Srinivas, V. R., Surolia, A., Vijayan, M., and Suguna, K. (2000) *J. Mol. Biol. 302*, 1129−1137.
20. Pace, C. N. (1990) *Trends Biochem. Sci. 15*, 14−17.
21. Lis, H., Joubert, F. J., and Sharon, N. (1985) *Phytochemistry 24*, 2803−2809.
22. De Bock, H., Loontiens, F. G., Lis, H., and Sharon, N. (1984) *Arch. Biochem. Biophys. 234*, 297−304.
23. Surolia, A., Sharon, N., and Schwarz, F. P. (1996) *J. Biol. Chem. 271*, 17697−17703.
24. Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167−241.
25. Schellman, J. A. (1990) *Biophys. Chem. 37*, 121−140.
26. Schellman, J. A. (1987) *Biopolymers 26*, 549−559.
27. Nicholson, E. M., and Scholtz, J. M. (1996) *Biochemistry 35*, 11369−11378.
28. Becktel, W. J., and Schellman, J. A. (1987) *Biopolymers 26*, 1859−1877.
29. Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
30. Hubbard, S. J. (1996) NACCESS, Version 2.1.1, Computer Program, Department of Biomolecular Sciences, UMIST, Manchester, U.K.
31. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
32. Molecular Operating Environment (MOE 2001.01), Chemical Computing Group, Inc., 1255 University St., Suite 1600, Montreal, Quebec, Canada H3B 3X3.
33. Halgren, T. A. (1996) *J. Comput. Chem. 17* (5 and 6), 490.
34. Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M., and Suguna, K., (1998) *J. Mol. Biol. 276*, 787−796.
35. Mer, J., Hietter, H., and Lefevre, J. F. (1996) *Nat. Struct. Biol. 3*, 45−54.
36. Yamaguchi, H. (2002) *J. Biochem. 14*, 139−151.
37. Nagai, K., and Yamaguchi, H. (1993) *J. Biochem. 113*, 123−125.
38. Adar, R., Streicher, H., Rozenblatt, S., and Sharon, N. (1997) *Eur. J. Biochem. 249*, 684−689.
39. Yamaguchi, H., and Uchida, M. (1996) *J. Biochem. 120*, 474−477.
40. Schülke, N., and Schmid, F. X. (1988a) *J. Biol. Chem. 263*, 8827−8831.
41. Arnold, U., and Ulbrich-Hofmann, R. (1997) *Biochemistry 36*, 2166−2172.
42. Backmann, J., Schafer, G., Wyns, L., and Bonisch, H. (1998) *J. Mol. Biol. 284*, 817−833.

BI035169E