Kinetic Folding Mechanism of Erythropoietin

Douglas D. Banks,* Joanna L. Scavezze, and Christine C. Siska
Department of Analytical and Formulation Sciences, Amgen, Seattle, Washington 98119-3105

ABSTRACT This report describes what to our knowledge is the first kinetic folding studies of erythropoietin, a glycosylated four-helical bundle cytokine responsible for the regulation of red blood cell production. Kinetic responses for folding and unfolding reactions initiated by manual mixing were monitored by far-ultraviolet circular dichroism and fluorescence spectroscopy, and folding reactions initiated by stopped-flow mixing were monitored by fluorescence. The urea concentration dependence of the observed kinetics were best described by a three-state model with a transiently populated intermediate species that is on-pathway and obligatory. This folding scheme was further supported by the excellent agreement between the free energy of unfolding and *m*-value calculated from the microscopic rate constants derived from this model and these parameters determined from separate equilibrium unfolding experiments. Compared to the kinetics of other members of the four-helical bundle cytokine family, erythropoietin folding and unfolding reactions were slower and less susceptible to aggregation. We tentatively attribute these slower rates and protection from association events to the large amount of carbohydrate attached to erythropoietin at four sites.

INTRODUCTION

Erythropoietin (EPO) is a 30 kD hormone produced primarily in the kidneys and is responsible for regulating red blood cell production by stimulating late erythroid precursor cells (1). Recombinant EPO is produced by Amgen under the trade name Epogen and has reached prominent therapeutic importance for treating anemia resulting from numerous conditions including renal failure and the use of chemotherapy (2). Human EPO contains 165 amino acids and ~40% of its mass consists of carbohydrate with glycosylation occurring at three N-linked and one O-linked position (3). The NMR solution structure of a nonglycosylated mutant has been determined and revealed that EPO shares a high degree of structural similarity with other hematopoietic cytokines despite a lack of sequence homology (4). It is a monomeric left-handed four-helical bundle protein consisting of four long α-helices arranged in an up-up and down-down topology (Fig. 1). Cysteines 7 and 161 form a disulfide bond that connect the first and last antiparallel helices together and a second disulfide bond between cysteine 29 and 33 bridge the C-terminal portion of the end of the first helix with the first part of the long crossover loop leading to the second helix (4).

Since the cloning of the gene coding for human EPO some 23 years ago (5), numerous biophysical and structural characterizations of recombinant EPO have been made including determination of its equilibrium stability by thermal and solvent denaturation (3,6,7). To the best of our knowledge, no kinetic studies have yet been undertaken to elucidate the folding mechanism of EPO, or how the protein navigates the folding landscape to reach the final native state structure from an ensemble of unfolded states.

EPO is an attractive protein-folding model due to its intermediate size and high degree of folding reversibility. A wealth of information has already been learned studying the folding mechanisms of other single-domain proteins, some of which fold by strictly two-state mechanisms and others through transiently populated intermediates (8,9). Such information has led to the general understanding that proteins sharing similar native state topology, but not necessarily sequence homology, often fold with similar rates through similar mechanisms (10). Folding data have already been collected for some of the other members of the four-helical bundle cytokine family such as the bovine and human growth hormones (bGH and hGH) and more recently granulocyte stimulating factor (G-CSF) (11–13). The *Escherichia coli*-expressed growth hormones and G-CSF are not glycosylated and a comparison of their folding mechanism with the Chinese hamster ovary (CHO) cell derived glycosylated EPO may help explain what effect carbohydrate has on protein folding mechanisms and rates, an area of research in which little data have been collected (14,15).

MATERIALS AND METHODS

Materials

Ultra pure urea was purchased from MP Biomedicals (Solon, OH). Urea was prepared fresh for each experiment and concentrations were determined by refractive index (16). All other chemicals were of reagent grade and purchased from Sigma Aldrich (St. Louis, MO). Recombinant EPO was expressed in CHO cells and purified as a drug product (5,17). Protein concentrations were determined by absorbance at 280 nm using an extinction coefficient of 22,710 $M^{-1}cm^{-1}$.

Methods

Equilibrium denaturation

The urea-induced equilibrium unfolding of the disulfide intact EPO was achieved by equilibrating individual samples of EPO overnight in urea

Submitted November 17, 2008, and accepted for publication February 12, 2009.

*Correspondence: dbank@amgen.com

Editor: Heinrich Roder.

0006-3495/09/05/4221/10 $2.00

doi: 10.1016/j.bpj.2009.02.049

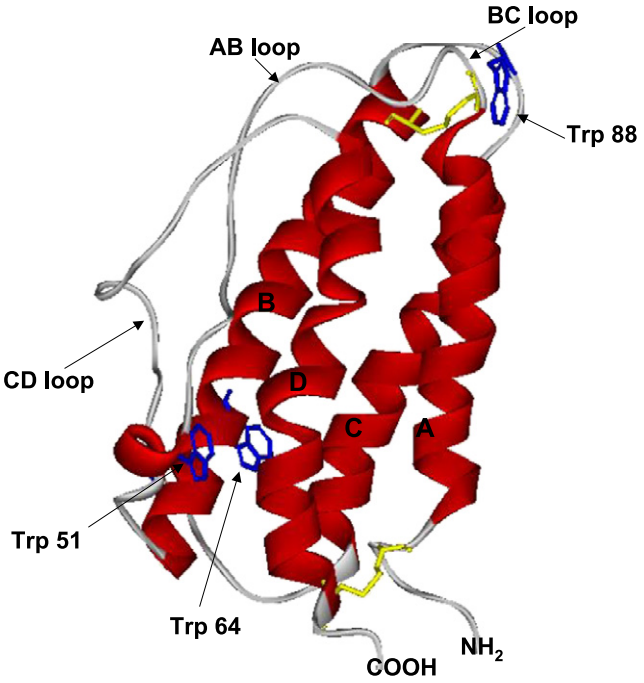

FIGURE 1 Ribbon diagram of EPO with helices lettered *A–D* from the N-terminus. Disulfide bonds are shown in yellow, and tryptophan residues are shown in dark blue. Trp-88 is located on the short loop connecting the antiparallel second and third helices (BC loop) and has been reported to be largely quenched by the disulfide bond bridging cysteines 29 and 33 within the N-terminal portion of the first crossover loop (27). The figure is derived from the structure of a nonglycosylated EPO mutant (1BUY.pdb) described elsewhere (4) and rendered using DSviewerPro 6.0.

concentrations ranging from 0 to 7.6 M in final conditions of 10 μM EPO, 140 mM NaCl, 20 mM sodium phosphate pH 6.9. Samples were analyzed the next day by far-ultaviolet (UV) circular dichroism (CD) and intrinsic tryptophan fluorescence (FL) spectroscopy at 25°C. Far-UV CD data were collected at 222 nm on a Jasco J-815 CD spectrophotometer using a pathlength of 0.2 cm. FL data were collected on both a T-format Quanta 6 PTI and Varian Carrey Eclipse spectrofluorometer using an excitation wavelength of 295 nm and monitoring emission from 305 to 400 nm. Two separate unfolding experiments and one refolding experiment were analyzed by CD and FL spectroscopy. The equilibrium unfolding transitions monitored by CD and at multiple FL emission wavelengths were globally fit to a two-state model using the data analysis program Savuka 6.2 described elsewhere (18,19). This model uses the linear extrapolation method between the free energy of unfolding and denaturant concentration (16) according to Eq. 1:

$$\Delta G^\circ = \Delta G^\circ(\mathrm{H_2O}) - m[\mathrm{Urea}]. \quad (1)$$

Here $\Delta G^\circ(\mathrm{H_2O})$ represents the free energy of unfolding in the absence of denaturant, and m represents the dependence of the free energy of unfolding on the denaturant concentration. For comparison of the data collected by the different optical probes, data were normalized to the fraction unfolded (F_{app}):

$$F_{\mathrm{app}} = \frac{(Y_F - Y_i)}{(Y_F - Y_U)}. \quad (2)$$

Here Y_i is the observed signal at a given urea concentration, and Y_F and Y_U are the extrapolated values of the folded and unfolded signal at the same urea concentration.

Manual mixing and stopped-flow kinetics

Kinetic data were collected at 25°C in the final buffer conditions of 10 μM EPO, 140 mM NaCl, 20 mM sodium phosphate, pH 6.9 after rapid dilution to various urea concentrations. Manually mixed folding and unfolding kinetic responses of the disulfide intact EPO were monitored at 222 nm using a Jasco J-815 CD spectrophotometer and at 360 and 325 nm by FL after excitation at 295 nm using and a Quanta 6 PTI and Varian Carey eclipse spectrofluorometers. Stopped-flow FL (SF-FL) data were collected on a Bio-logic MOS 450 spectrophotometer interfaced to a stopped-flow SFM 400 tower and equipped with a high-density mixer. SF-FL data were collected with an excitation wavelength of 295 nm and kinetic responses were detected using a 320 nm cutoff filter. Dead times for SF-FL of ~5.5 ms were calculated measuring the quenching of *N*-acetyltryptophanamide with increasing concentrations of *N*-bromosuccinimide as described elsewhere (20). The signal/noise ratios for SF-FL experiments were enhanced by averaging 10 to 15 shots. CD and FL kinetic traces were fit locally to a single or the sum of multiple exponential equations of the form shown in Eq. 3 with Kaleidagraph 3.5 (Synergy Software):

$$Y(t) = Y_\infty + \sum_{i=1}^{N} \Delta Y_i \exp(-t/\tau_i). \quad (3)$$

Here Y_∞ is the final signal after equilibrium has been reached, ΔY_i is the change in signal associated with a particular kinetic phase, and τ_i is the relaxation time. The data fitting program Savuka 6.2 (18,19) was used to globally fit the urea concentration dependence of unfolding and folding rate constants of multiple FL and CD kinetic traces within the linear regions of the chevron plot (see Fig. 4) to Eq. 4:

$$k_i = k^{\mathrm{H_2O}} \exp\left\{\frac{m^\ddagger [\mathrm{Urea}]_i}{RT}\right\}. \quad (4)$$

Here, $k^{\mathrm{H_2O}}$ is the rate constant in the absence of urea, and $m^\ddagger$ is proportional to the change in solvent accessible surface area between the initial and transition state species. These parameters were used as the initial microscopic rate constants for fitting the urea concentration dependence of the observed rates to the three three-state folding models.

Burst-phase folding analysis

The dead time of the SF-FL instrument was subtracted from the FL kinetic refolding traces and the signal at the time of mixing calculated using the fits of these adjusted traces. The signal of the kinetic trace after equilibrium had been reached was then normalized to an equilibrium unfolding curve generated using the same 320 nm cutoff filter used for the SF-FL experiments. The time zero signal was normalized by this same amount and compared with the extrapolated equilibrium unfolding signal at the final urea concentrations the protein was folded to.

Data analysis

The equation describing the on-pathway mechanism (21) is given below:

$$f(\lambda) = \lambda^2 - \lambda(k_{\mathrm{UI}} + k_{\mathrm{IU}} + k_{\mathrm{IN}} + k_{\mathrm{NI}}) + (k_{\mathrm{UI}}k_{\mathrm{IN}} + k_{\mathrm{UI}}k_{\mathrm{NI}} + k_{\mathrm{IU}}k_{\mathrm{NI}}) = 0. \quad (5)$$

The observed rate constants are represented as λ_1 and λ_2 and are the roots of $f(\lambda) = 0$. The analytical solutions to the off-pathway and parallel-path mechanisms can be found in Bachmann and Kiefhaber (21). Fitting was initially performed by fixing the microscopic rate constants and $m^\ddagger$-values to the results derived from the global fits of the linear regions of the chevron plot, Eq. 4, and optimizing the microscopic rate constant and associated $m^\ddagger$-value for the unfolding of the intermediate species. This approach resulted in agreement between the predicted and observed rate constants

only for the on-pathway model. For the off-pathway and parallel pathway models, fitting was performed following the methods described by Capaldi et al. (22), where each of the individual microscopic rate constants and $m^{\ddagger}$-values were varied manually until a satisfactory match between the observed and predicted data were found. Kinetic traces were simulated by calculating the amplitudes for each kinetic phase using the equations given in Bachmann and Kiefhaber (21), and the parameters that resulted in the best fits to the observed rates of the chevron plot for each of the three three-state models. The free energy of each kinetic species (see Fig. 7) were calculated using Eq. 6:

$$\Delta G^{\circ} = -RT\ln K_{\mathrm{eq}}, \tag{6}$$

where the equilibrium constant was expressed as a ratio of the microscopic rate constants given in Table 1. The unfolded state was arbitrarily set to zero, and the height of the energy barriers between the unfolded/intermediate and intermediate/native species were estimated using transition state theory based on the Erying equation, Eq. 7:

$$k = k_a e\left(\frac{-\Delta G^{\circ\ddagger}}{RT}\right). \tag{7}$$

The preexponential factor was set to 10^8 s^{-1} as described by Krieger et al. (23). The free energy diagram reaction coordinate represents the relative compactness of each kinetic species and transition state or β-value. The β-values were defined as given in Eq. 8:

$$\begin{aligned}\beta_{\mathrm{TS1}} &= m_{\mathrm{UI}}/m_{\mathrm{eq}};\ \beta_I = (m_{\mathrm{UI}} - m_{\mathrm{IU}})/m_{\mathrm{eq}};\ \beta_{\mathrm{TS2}} \\ &= (m_{\mathrm{UI}} - m_{\mathrm{IU}} + m_{\mathrm{IN}})/m_{\mathrm{eq}}.\end{aligned} \tag{8}$$

The kinetic $m^{\ddagger}$-values were taken from Table 1, and m_{eq} was determined from the equilibrium unfolding study.

RESULTS

Equilibrium denaturation

Previously published data have shown that thermal and chemical induced equilibrium denaturation of EPO were reversible processes that were best fit to a two-state model and yielded a free energy of unfolding of 6–7 kcal mol^{-1} (3,7). We repeated these measurements using the same solvent conditions and temperature used for the kinetic folding studies. This was essential because moderately small differences in salt concentration were noticed to influence the equilibrium stability of EPO. Urea induced equilibrium unfolding was monitored by far-UV CD at 222 nm and tryptophan FL from residues 51, 64, and 88. Two CD and one FL unfolding data sets and one FL refolding data set were collected and globally fit to a two-state model yielding a free energy of unfolding value of 7.2 ± 0.1 kcal mol^{-1} and an m-value of 1.69 ± 0.03 kcal mol^{-1} M^{-1}. Representative equilibrium unfolding curves monitored by CD and FL are shown with their local fits in Fig. 2 *A*, and the unfolding data normalized to the fraction unfolded are shown with their global fit in Fig. 2 *B*.

Folding of EPO

Folding reactions were initiated by rapid dilution of an unfolded stock of EPO equilibrated overnight in 6 M urea to various final urea concentrations by manual and stopped-flow mixing methods. Folding was a slow process and required between 1 and 10 min before reaching equilibrium when refolding to final urea concentrations ranging from 1 to 4.5 M urea, respectively. For this reason, manual mixing (monitored by far-UV CD and FL) was utilized when refolding between final urea concentrations of 1 to 4.5 M, and stopped-flow mixing (monitored by FL) was utilized when refolding between 0.6 and ~2.5 M urea.

Folding reactions monitored by SF-FL were best fit to a sum of three exponential equations, Eq. 3, when refolding to low urea concentrations (Fig. 3 *A*). The fast phase showed an increase in FL intensity and was essentially complete within 300 ms (Fig. 3 *A*, *inset*). This phase only accounted for ~8% of the total change in FL signal or amplitude upon refolding to the lowest urea concentrations and disappeared when folding to final urea concentrations above 1.5 M (Fig. 3 *B*). The intermediate and slow kinetic phases both showed a decrease in FL intensity as the protein folded, with the intermediate folding phase carrying ~70% of the FL amplitude when folding to the lowest urea concentrations. The amount of protein folding through the slow phase increased as EPO was folded to higher final urea concentrations. By 1.8 M urea 80% of the total folding signal was contributed by the slow folding phase, and by 2.0 M urea the refolding of EPO monitored by SF-FL was best fit to

TABLE 1 Kinetic and equilibrium parameters of EPO

	Rate constant (s^{-1})	$m^{\ddagger}$-value (kcal mol^{-1} M^{-1})	ΔG° (H_2O) (kcal mol^{-1})	m-value (kcal mol^{-1} M^{-1})
(*U* to *I*)	0.25 (0.01)	−0.29 (0.02)		
(*I* to *U*)*	0.0008 (0.0001)	0.82 (0.06)		
(*I* to *N*)	0.063 (0.003)	−0.08 (0.02)		
(*N* to *I*)	$1.07E^{-4}$ ($2.0E^{-6}$)	0.458 (0.002)		
UN†			7.2 (0.2)	1.65 (0.07)
UN‡			7.2 (0.1)	1.69 (0.03)

Rate constants and $m^{\ddagger}$-values (except for *I* to *U*) for on-pathway folding scheme were derived from the global fits within the linear regions of the chevron plot to Eq. 4.

*Rate constant and associated $m^{\ddagger}$-value for *I* to *U* were derived from the fit of the chevron plot to an on-pathway model, Eq. 5.

†Equilibrium unfolding parameters calculated from the microscopic rate constants.

‡Equilibrium unfolding parameters determined from equilibrium unfolding experiments. The values in parentheses represent the standard deviation of the fits of the observed rate constants and equilibrium unfolding parameters. Conditions: 140 mM NaCl, 20 mM NaPi (pH 6.9), 25°C.

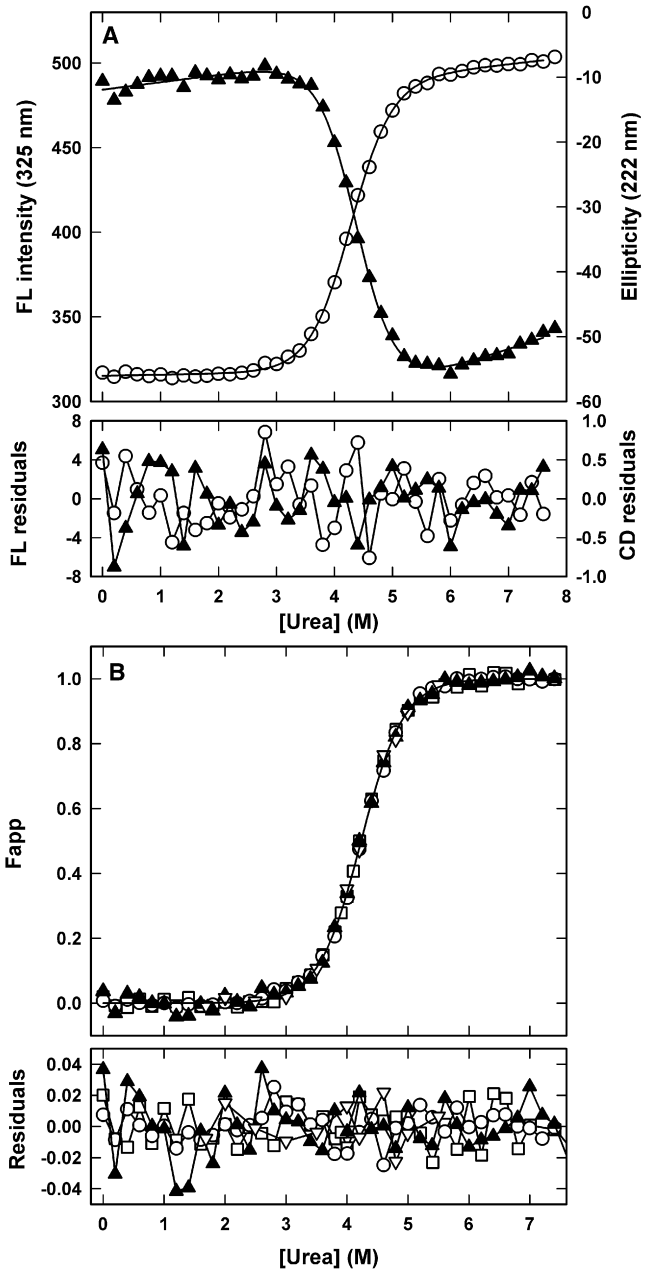

FIGURE 2 (*A*) Representative equilibrium unfolding of EPO monitored by far-UV CD at 222 nm (*open circles*) and intrinsic tryptophan FL at 325 nm (*solid triangles*) were best fit locally to two-state unfolding model (*solid lines*). (*B*) These two data sets, along with an additional CD unfolding data set and a FL refolding data set (*open boxes* and *triangles*, respectively), were fit globally to the same model and are shown normalized to the fraction unfolded. The solid line represents the global fit and the fitted parameters are given in Table 1. The residuals between the data and the fits are shown in the panels below. Conditions are described in the legend to Table 1.

a single exponential equation (Fig. 3 *B*). The protein concentration dependence of the folding reaction monitored by SF-FL was examined by refolding to final EPO concentrations of 1, 2, 5, and 10 μM. The fast kinetic phase disappeared when refolding to EPO concentrations below 2 μM. The intermediate and slow folding phases, however, did not change in either rate constant or amplitude with changing final EPO concentration.

Due to the relatively long mixing time of between 8 and 10 s, all manual mixing folding data monitored by FL and CD only fit to a single exponential equation, with rate constants matching those of the slow SF-FL phase. SF-FL data were monitored using a 320 nm cutoff filter; however, manual mixing FL data were detected at different wavelengths using a monochrometer. When folding was monitored at 325 nm, the FL signal decreased after mixing to final urea concentrations between 1 and ~3 M. When folding to final urea concentrations >3 M, the FL intensity increased (Fig. 3 *C*). This interesting phenomenon may be indicative of a kinetic intermediate and is discussed later.

Unfolding of EPO

Unfolding reactions of EPO were also rather slow and required approximately between 3 and 10 min before reaching equilibrium when unfolding to final urea concentrations ranging from 7 to 4.5 M, respectively. Due to these slow unfolding times, manual mixing was used to jump from 0 M urea to the various final urea concentrations. The unfolding kinetic traces monitored by both CD and FL were best described by a single exponential equation at all final urea unfolding concentrations between 4.5 and 7.6 M. A representative manual mixing unfolding kinetic trace to 5.8 M urea monitored by CD is shown in Fig. 3 *D* demonstrating the appropriateness of this model and showing that even when unfolding to strong denaturant concentrations, >95% of the unfolding reaction was accounted for by manual mixing.

Urea concentration dependence of folding and unfolding reactions

The urea concentration dependence of the locally fitted SF-FL and manually mixed CD and FL folding and unfolding rate constants are shown in the chevron plot (Fig. 4 *A*). The logarithm of the slow refolding rate constants shows a linear dependence on the final urea concentration between 2.8 and 4 M urea. Below 2.8 M, the denaturant concentration dependence of the slow folding phase deviates from linearity and shows a pronounced curvature or ''roll-over''. Refolding data within the linear regions of the intermediate and slow kinetic folding phases of the chevron plot were fit globally to Eq. 4 to determine the rates of folding in the absence of denaturant and the kinetic $m^{\ddagger}$-values. Twenty-four FL and CD kinetic folding traces collected by manual mixing were fit between 2.8 and 4 M urea. A total of 29 SF-FL and manual mixed FL and CD folding traces were fit between 0.6 and 1.6 M urea for the slow folding phase, and 21 SF-FL folding traces were fit between 0.6 and 1.6 M urea for the intermediate folding phase. The fast kinetic folding phase (Fig. 4 *A*, *inset*) was not fit and is discussed later. The rate of unfolding in the absence of denaturant and associated $m^{\ddagger}$-value were determined from the global fit of

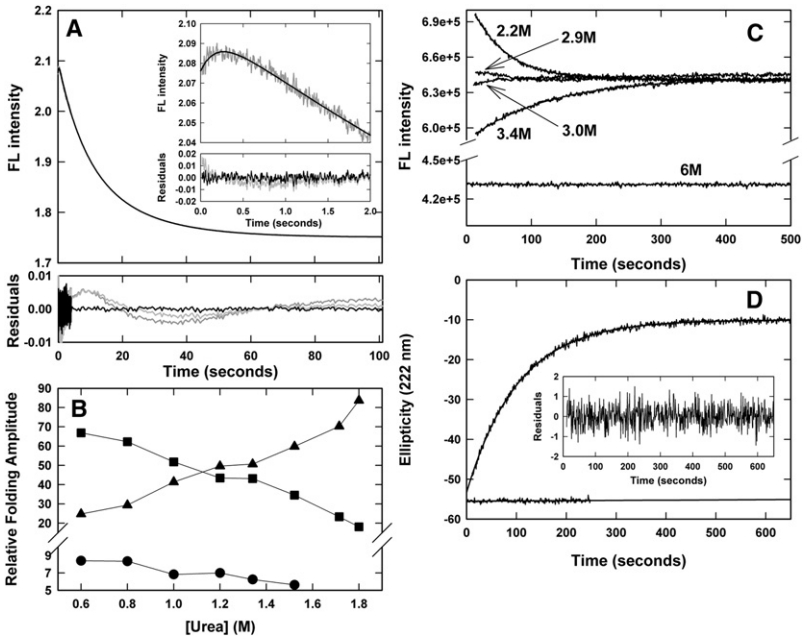

FIGURE 3 Representative stopped-flow and manual mixed EPO folding and unfolding kinetic traces monitored by far-UV CD at 222 nm and intrinsic tryptophan FL. (*A*) SF-FL data monitored with a 320 nm cutoff filter was best fit to a sum of three exponential equations. The residuals of this fit and of a single and double exponential fit are overlaid for comparison in the panel below as black, dark gray, and light gray lines, respectively. The inset is a magnification of the initial time points of the folding reaction. (*B*) The kinetic amplitudes of SF-FL folding. The solid circles, squares, and triangles represent folding from the fast, intermediate, and slow phases respectively. (*C*) Manual mixing folding monitored by FL at 325 nm was best fit to a single exponential equation and showed evidence of a hyperfluorescent intermediate populated within the time of mixing. The numbers corresponding to each trace indicate the final urea concentration EPO was folded to. (*D*) Unfolding of EPO to 5.8 M urea by manual mixing monitored by CD. The solid line represents the local fit of the data to a single exponential equation with residuals of the fit shown in the inset. The horizontal line is the folded signal at the same EPO concentration and an extrapolation of this signal to the end of the unfolding time. Conditions are described in the legend to Table 1.

22 manually mixed FL and CD kinetic unfolding traces to Eq. 4 within the linear region of Fig. 4 *A*, between the final urea concentrations of 5.2 and 7.4 M.

Burst-phase folding analysis

Burst-phase analysis was used to indirectly reveal the presence of a possible folding intermediate that cannot be detected during the dead time of the stopped-flow instrument. As shown in Fig. 4 *B*, a small amount of structure appears to form within the dead time of the SF-FL instrument. Unfortunately stopped-flow mixing was only possible when refolding to urea concentrations no higher than ~3 M or within the folded baseline region of the equilibrium curve. When trying to refold to higher urea concentrations, the relaxation times became too long, and the solution within the stopped-flow flow-cell would become unstable due to diffusion or convective flow.

DISCUSSION

Analysis of the chevron plot

The presence of the two faster folding phases and the pronounced roll-over or curvature in the folding limb of the chevron plot in Fig. 4 *A* indicate that the folding mechanism of EPO may deviate from simple two-state behavior. Several possibilities may be responsible for the curvature of the folding limb of the chevron plot including: heterogeneity of the unfolded state, the presence of transient aggregation, and the population of folding intermediates (21,24).

Heterogeneity of the unfolded state leads to multiple folding phases or curvature in the folding limb of the chevron plot when the rate of interconversion of the unfolded species (for example *cis-trans* isomerization at Xaa-Pro peptide bonds) is slower or on the same timescale as the folding reactions leading to the native state, respectively (21). This possibility can be ruled out; unlike folding through intermediate states, the folding amplitudes from multiple unfolded states only depend on the relative population of each unfolded species present at the initiation of the folding reaction, and not on the final urea concentration EPO was folded to as shown in Fig. 3 *B* (25). Moreover, folding reactions that are limited by an isomerization events do not display a strong denaturant dependence on the folding rate constants or $m^{\ddagger}$-value (26), which is not the case as demonstrated in Fig. 4 *A*.

The formation of transient aggregate or an off-pathway intermediate species upon folding is another possible explanation for multiple kinetic phases and curvature within the refolding limb of the chevron plot (24). A good indication that this may be occurring is if the rates of folding increase with increasing final folding denaturant concentration (a positive $m^{\ddagger}$-value), signifying that the protein is dissociating or unfolding before folding to the native state (24). This appears to be the case for the fast folding phase of EPO monitored by SF-FL (Fig. 4 *A*, *inset*). To test whether this phase represents the unfolding of aggregate, the protein concentration dependence of the folding reaction was determined by folding to EPO concentrations between 1 and 10 μM. When folding at or below 2 μM EPO the fast kinetic phase disappeared and the folding reaction, even at the lowest final denaturant concentration, were best fit to the sum of two exponential equations. The protein concentration dependence of this phase indicates that it may represent the unfolding of transient aggregate; alternatively it may be reporting on the unfolding of an off-pathway intermediate

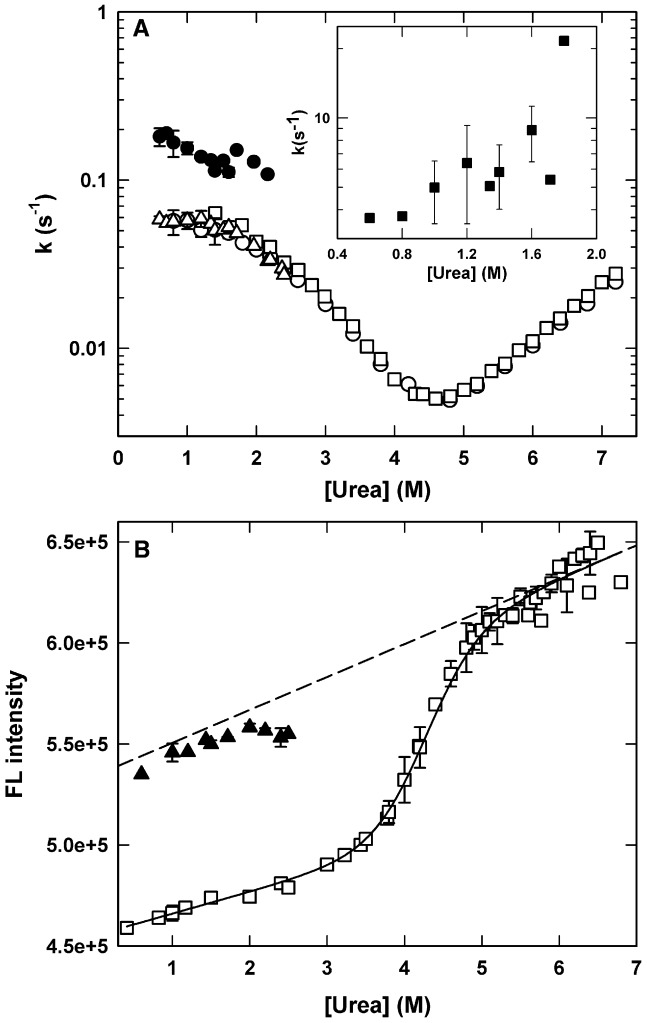

FIGURE 4 (*A*) Urea concentration dependence of the folding and unfolding reactions. Local fits of the SF-FL folding data are represented by solid squares (*inset*), solid circles, and open triangles. All manual mixing FL and CD data is shown by open circles and squares, respectively. Error bars represent the standard deviation of multiple kinetic traces and are smaller than all of the manual mixing data points. (*B*) SF-FL burst-phase analysis. The solid triangles are the FL signals calculated at the time of mixing from kinetic refolding traces that have been normalized to the folded baseline region of an equilibrium unfolding curve generated using a 320 nm cutoff filter (*open boxes*). This curve was fit to a two-state equilibrium model (*solid line*) to better define the unfolded baseline region (*dashed line*). Error bars represent the standard deviations of three refolding experiments and three equilibrium curves. Conditions are described in the legend to Table 1.

that simply is not observed at low protein concentrations due to its low amplitude and the limit of detection of the SF-FL instrument. Importantly, the rates of the two slower kinetic phases slow with increasing final urea concentration and their corresponding amplitudes did not change with protein concentration, suggesting that the intermediate and slow phases represent productive folding from monomer.

If the fast folding phase does represent the unfolding reaction of an off-pathway species, as its $m^{\ddagger}$-value suggests, then some amount of folding takes place within the dead-time of the SF-FL instrument. To test this hypothesis burst-phase analysis was used to indirectly monitor the formation of structure between the time of stopped-flow mixing and the first observable FL signal. Unfortunately, it was not possible to produce a complete equilibrium unfolding curve using the final SF-FL signals from the folding and unfolding reactions due to the long collection times within the folding/unfolding transition region, which were beyond the capabilities of the instrumentation. The SF-FL data were therefore normalized to an equilibrium curve generated using the same 320 nm cutoff filter used for the stopped-flow detection. This resulted in a relatively shallow unfolding curve with a poorly defined unfolded baseline with a large positive slope (Fig. 4 *B*). The unfolding curve was therefore fit to the two-state unfolding model, fixing the equilibrium parameters to those previously determined, to better define the unfolded baseline. The extrapolation of this baseline revealed a small difference between it and the FL signal calculated at the time of mixing, indicating the presence of an intermediate or collapsed state with a lower FL quantum yield than the unfolded state populated during the dead time of the instrument. We speculate that the fast folding phase represents the unfolding of this relatively unstable burst-phase species formed within the 5 ms dead time of the SF-FL instrument. This interpretation must be viewed with some caution, however, due to the difficulty in accurately defining the slope of the unfolded baseline.

Based on the previous arguments, the best explanation for the curvature of the folding limb of the chevron plot and the observed intermediate kinetic folding phase by SF-FL is the formation of a transient intermediate species during the folding of EPO. The manual mixing times required for refolding experiments were too long to observe this intermediate kinetic phase directly; therefore we are unable to comment on the formation of secondary structure of the intermediate species. Manual mixing folding experiments monitored by FL at 325 nm also revealed an interesting phenomenon not seen by SF-FL that used a 320 nm cutoff filter. When EPO was folded to low denaturant concentrations, the FL intensity decreased as the protein folded. However, when folding to urea concentrations above ~2.9 M and greater, the FL intensity increased as the protein folded (Fig. 3 *C*). This phenomenon is not due to the urea dependence on the native and unfolded state FL signals at 325 nm as demonstrated by the equilibrium unfolding curve in Fig. 2 *A* collected at this wavelength. These observations suggest that an intermediate with a greater FL quantum yield at 325 nm than either the unfolded state in 6 M urea or the folded state is formed within the time required for manual mixing. At a concentration of ~2.9 M urea, this intermediate is destabilized and folding proceeds directly from the unfolded state. Unfortunately these data alone give no indication of whether the intermediate is either on or off-pathway because there would be a net decrease in the FL signal at this wavelength for either mechanism.

The cause for the increased FL of the intermediate species is likely due to tryptophan 88. In the native state, this tryptophan is adjacent to the disulfide bond connecting cysteines 29 and 33 and has been reported to be largely quenched (27) (Fig. 1). In the unfolded state, this should no longer be the case, and all three tryptophan residues would be more or less equally quenched by solvent relative to the folded state (at 325 nm). We speculate that upon folding, Trp-88 is partially buried in the structure of the intermediate species, away from the disulfide connecting cysteines 29 and 33, thereby increasing its quantum yield. As the protein folds to the native species, Trp-88 would move into range of the disulfide and the total FL intensity would be expected to decrease.

Determination of the three-state folding model

There are three possible folding mechanisms with a single intermediate species that may account for the intermediate and slow observable rate constants. They are the sequential, off-pathway and parallel folding schemes (Fig. 5 *A*). Discrimination between these three mechanisms is not possible using the folding kinetics of EPO alone (21). The analytical solutions of the differential equations describing each of these three mechanisms were therefore used to distinguish which mechanism best described the observed urea concentration dependence of the folding and unfolding rates. The logarithm of each of the folding and unfolding rate constants were assumed to change linearly with urea concentration (28). The results of the global fits within the linear regions of the chevron plot to Eq. 4 were therefore used as the initial parameters (the microscopic rate constants) for determining which of the three possible kinetic schemes (Fig. 5 *A*) best described the urea concentration dependence of the two observed rate constants. These parameters were well constrained by the observed rates due to the stability of the intermediate species. The only parameters not defined by the observed kinetic rates were the microscopic rate constant and associated $m^{\ddagger}$-value for the unfolding of the intermediate species. However, these values were also well constrained because they define, together with k_{UI} and $m^{\ddagger}_{UI}$, the extent of curvature observed for the slower folding rate constant dependence on final urea concentration. The exact parameters determined from the global fits of the observed data (Table 1) were used first in the on-pathway model, Eq. 5, to calculate k_{IU} and $m^{\ddagger}_{IU}$ and were able to adequately reproduce the urea concentration dependence of the two observed rate constants (Fig. 5 *B*, *red lines*). In contrast, the urea concentration dependence of the observed kinetic rates could not be reproduced using either the off-pathway or parallel folding models no matter how the microscopic rate constants were varied (Fig. 6).

The kinetic parameters derived from the best fits to each of the three three-state models were used to calculate the amplitudes of the observed folding reactions and simulate the

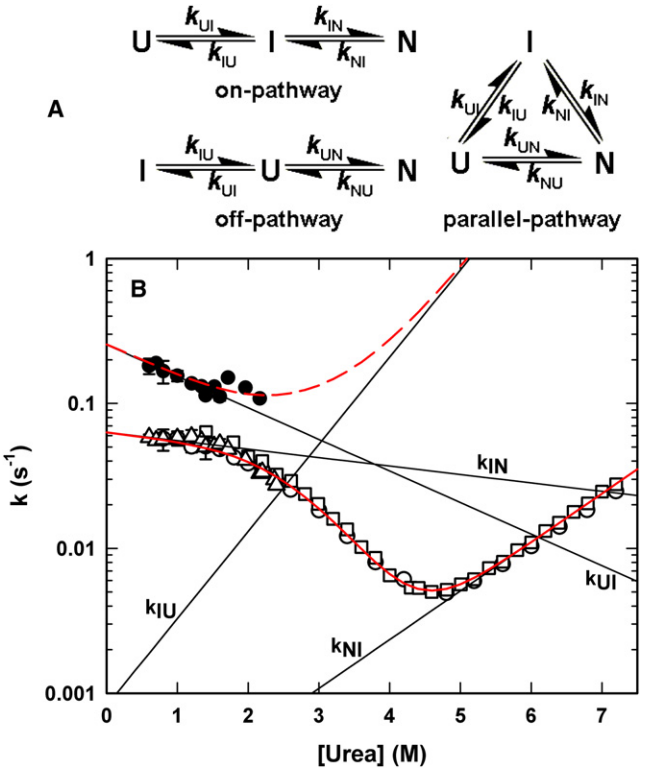

FIGURE 5 (*A*) Three possible three-state folding mechanisms that describe a two exponential process. The intermediate (*I*) may be either on-pathway, off-pathway, or folding may proceed directly to the native state (*N*) from the unfolded ensemble (*U*), as well as through an intermediate in a parallel pathway mechanism. The microscopic rate constants are shown above and below the arrows. (*B*) Urea concentration dependence of the folding and unfolding reactions fit to a three-state on-pathway model. The microscopic rate constants (*solid black lines*) were calculated using Eq. 4 and used as initial parameters for fitting the two observed rate constants (*red dashed* and *solid lines*) using the on-pathway mechanism, Eq. 5. All data points are the same as those described in Fig. 4 *A*. Fitted values of the rate constants and $m^{\ddagger}$-values are given in Table 1. Conditions are described in the legend to Table 1.

folding traces. Only the microscopic rate constants from the on-pathway model were able to simulate the folding kinetics using realistic values for the FL intensities of the unfolded and intermediate species. For example, the on-pathway microscopic rate constants simulated the folding trace shown in Fig. 3 *A* using FL intensity values for the unfolded and intermediate species of 2.1 and 1.83, respectively. The parameters from the off and parallel pathway models could only simulated the folding trace in Fig. 3 *A* using illogical FL intensity values for the unfolded and intermediate species such as 4.9 and −0.63 for the off-pathway model and 1.89 and 1.78 for the parallel pathway model. Further support for the validity of the on-pathway three-state folding mechanism was the excellent agreement between the free energy and *m*-values for the equilibrium unfolding of EPO calculated from the kinetic data assuming the on-pathway mechanism (7.2 ± 0.2 kcal mol^{-1} and 1.65 ± 0.07 kcal mol^{-1} M^{-1}), and these values determined directly

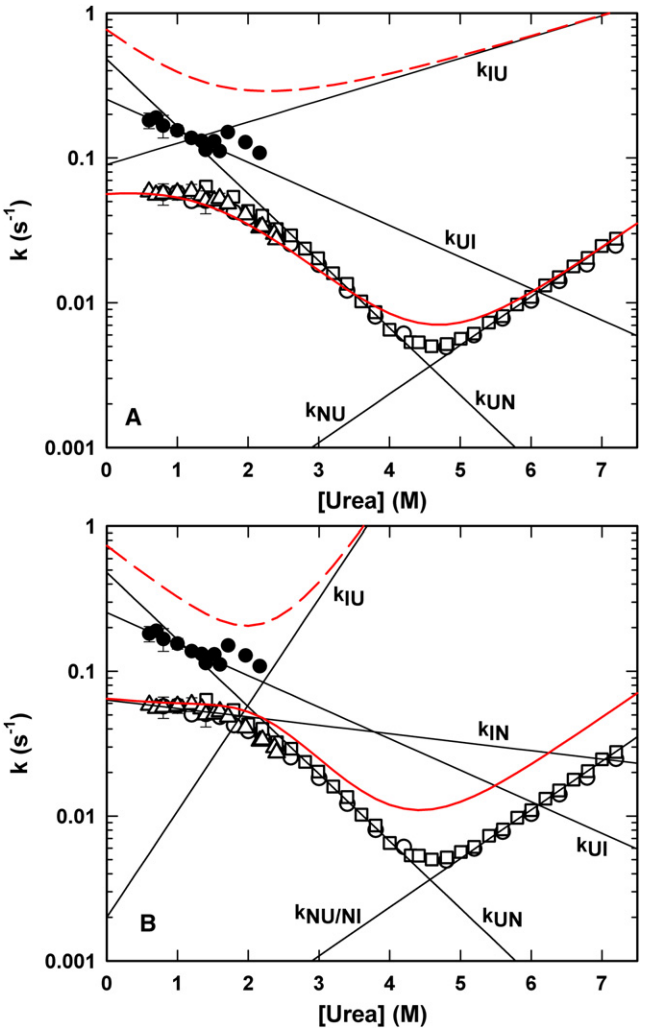

FIGURE 6 Urea dependence of the two observable rate constants fit to an off-pathway model (*A*) and to a parallel or triangle mechanism (*B*). Fits of the two observable rate constants are shown as red dashed and solid lines. All data points are the same as those described in Fig. 4 *A*.

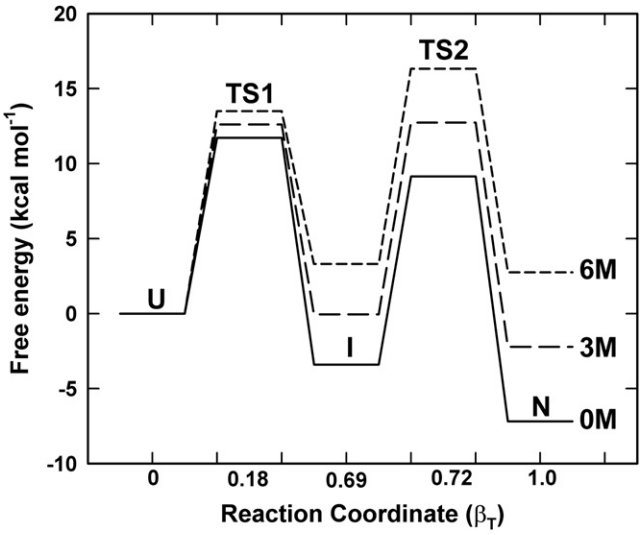

FIGURE 7 Free energy diagram of EPO folding and unfolding at 0 M, 3 M, and 6 M urea concentrations. Energy levels were calculated using the parameters in Table 1 and Eqs. 6 and 7. The reaction coordinate describes the amount of burial of surface area relative to the native state or β_T value calculated using Eq. 8.

from the equilibrium unfolding experiments (7.2 ± 0.1 kcal mol^{-1} and 1.69 ± 0.03 kcal mol^{-1} M^{-1}). Whether or not a small fraction of EPO may fold directly to the native state is difficult to rule out at this time, however, given the ability of the on-pathway model to reproduce the urea concentration dependence of the observed folding rates, folding amplitudes, and calculate the equilibrium unfolding parameters, we believe the folding mechanism of EPO to be predominately on-pathway.

The free energy profile of EPO folding

The observed and microscopic rate constants for the on-pathway three-state folding mechanism (Table 1) were used to calculate the change in free energy during the course of EPO folding/unfolding at three different urea concentrations using Eq. 6 (Fig. 7). The stability of the transition states between each of these kinetic species were estimated by transition state theory using a preexponential factor that takes into consideration intrachain diffusion processes, Eq. 7 (23). The progress of the folding reaction is plotted as a function of β. The values of β for each species were calculated using Eq. 8 and the kinetically derived $m^{\ddagger}$-values and depicts the relative compactness of each of the kinetic species (8). The values of β vary from 0 to 1 for kinetic species resembling the unfolded and native species in their degree of compaction or burial of surface area, respectively.

Based on the kinetic and equilibrium data, we propose that EPO folds initially to an off-pathway intermediate species (which may be aggregate) of limited stability within the 5 ms dead time of the stopped-flow instrument. This species unfolds within the first 300 ms of the observable reaction, and the folding of EPO proceeds predominately to an intermediate ensemble through a loosely packed transition state, which occludes only 18% of the total surface area buried in the native state. The intermediate state buries ~70% of the total surface area buried in the native state. Based on the manual mixing refolding data monitored by FL emission at 325 nm, we speculate that the compaction of the remaining 30% of surface area from the intermediate to native state involves a reconfiguration of the BC and AB loops containing tryptophan 88 and the disulfide bond connecting cysteines 29 and 33, respectively.

Comparison to the folding of G-CSF

Statistically significant correlations have been found between protein folding rates and mechanisms of relatively small single domain proteins to both protein stability and topological complexity, measured by a number of parameters such as contact order and the number of sequence-distant native

pairs (10,29). Proteins with more local contacts in the native structure were shown to fold at faster rates (10,29). We therefore thought it would be informative to compare the folding mechanism of EPO to other members of the four-helical bundle cytokine family that share a high degree of structural similarity to EPO and have been previously studied by kinetic folding methods. Comparisons to the folding mechanisms of the bovine and human growth hormones were difficult to make due to aggregation and precipitation observed when the *E. coli* derived growth hormones were refolded to the low denaturant concentrations where the intermediate folding phase of EPO was observed (11,12). This observation was likely due to the lack of glycosylation of the growth hormones, which has been noted to increase protein stability and solubility and minimize aggregation upon refolding (30,31). These observations were also made when EPO was expressed in *E. coli*; the nonglycosylated EPO was 1.4 kcal mol^{-1} less stable than CHO expressed EPO and was prone to aggregation and precipitation (3).

Aggregation when folding to low denaturant concentrations was not as problematic for the *E. coli* expressed nonglycosylated G-CSF. The denaturant dependence of the folding and unfolding rates are comparable to those observed for EPO, and a similar sequential folding mechanism with a populated intermediate was recently proposed (13). Interestingly, the folding rate constants were ~2 orders of magnitude faster than observed for EPO despite the studies being conducted at 10°C. It is therefore somewhat surprising that EPO and G-CSF, which share the same fold (the four-helical bundle of EPO superimposes on G-CSF with a root mean-square deviation of only 1.6 Å (4)) and similar stabilities, should have such a large discrepancy in their folding and unfolding rates.

The effect of glycosylation on the detailed kinetic folding mechanisms of proteins is not well understood (14,15). Studies of small synthesized glycopeptides have led to the hypothesis that glycosylation of proteins may influence their rates of folding by limiting or restricting conformational space, thereby biasing particular folding pathways and speeding up the folding process (15). It may also be argued that such conformational restrictions or rigidity may increase the height of transition state energy barriers for steric reasons and slow the folding or unfolding processes as proposed for horseradish peroxidase (32). Which scenario is more likely may be protein specific, depending on the location of glycosylation and these regions involvement in the rate limiting step of folding. The folding mechanism of EPO and G-CSF appear to be conserved; however, their folding and unfolding rates are dissimilar. We speculate that the relatively slow EPO kinetics may be due to the significant amount of glycosylation on four residues compared to the nonglycosylated G-CSF, although other factors may contribute to these slower rates such as differences in the disulfide bond connectivity. This large amount of carbohydrate may not change the folding pathway of EPO, but could be enough to increase the energetic roughness of the folding landscape and slow down the folding and unfolding reactions while minimizing the potential for off-pathway aggregation events of folding intermediate(s).

The authors thank Lisa M. Gloss, Michael Treuheit, and David Brems for critical review, helpful comments, and discussions.

REFRENCES

1. Graber, S. E., and S. B. Krantz. 1978. Erythropoietin and control of red-cell production. *Annu. Rev. Med.* 29:51–66.
2. Markham, A., and H. M. Bryson. 1995. Epoetin alfa. A review of its pharmacodynamic and pharmacokinetic properties and therapeutic use in nonrenal applications. *Drugs.* 49:232–254.
3. Narhi, L. O., T. Arakawa, K. H. Aoki, R. Elmore, M. F. Rhode, et al. 1991. The effect of carbohydrate on the structure and stability of erythropoietin. *J. Biol. Chem.* 266:23022–23026.
4. Cheetham, J. C., D. M. Smith, K. H. Aoki, J. L. Stevenson, T. J. Hoeffel, et al. 1998. NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nat. Struct. Biol.* 5:861–866.
5. Lin, F. K., S. Suggs, C. H. Lin, J. Browne, R. Smalling, et al. 1985. Cloning and expression of the human erythropoietin gene. *Proc. Natl. Acad. Sci. USA.* 82:7580–7584.
6. Arakawa, T., J. S. Philo, and Y. Kita. 2001. Kinetic and thermodynamic analysis of thermal unfolding of recombinant erythropoietin. *Biosci. Biotechnol. Biochem.* 65:1321–1327.
7. Lah, J., I. Prislan, B. Krzan, M. Salobir, A. Francky, et al. 2005. Erythropoietin unfolding: thermodynamics and its correlation with structural features. *Biochemistry.* 44, 13883–13892.
8. Jackson, S. E. 1998. How do small single-domain proteins fold? *Fold. Des.* 3:R81–R91.
9. Gunasekaran, K., S. J. Eyles, A. T. Hagler, and L. M. Gierasch. 2001. Keeping it in the family: folding studies of related proteins. *Curr. Opin. Struct. Biol.* 11:83–93.
10. Plaxco, K. W., K. T. Simons, I. Ruczinski, and D. Baker. 2000. Topology, stability, sequence, and the length: defining the determinants of two-state protein folding kinetics. *Biochemistry.* 39:11177–11183.
11. Brems, D. N., S. M. Plaisted, J. J. Dougherty, and T. F. Holzman. 1987. The kinetics of bovine growth hormone folding are consistent with a framework model. *J. Biol. Chem.* 262:2590–2596.
12. Youngman, K. M., D. B. Spencer, D. N. Brems, and M. R. DeFelippis. 1995. Kinetic analysis of the folding of human growth hormone. *J. Biol. Chem.* 270:19816–19822.
13. Brems, D. N. 2002. The kinetics of G-CSF folding. *Protein Sci.* 11:2504–2511.
14. Hackenberger, C. P. R., C. T. Friel, S. E. Radford, and B. Imperiali. 2005. Semisynthesis of glycosylated Im7 analogue for protein folding studies. *JACS.* 127:12882–12889.
15. Live, D. H., R. A. Kumar, X. Beebe, and S. J. Danishefsky. 1996. Conformational influences of glycosylation of a peptide: a possible model for the effect of glycosylation on the rate of protein folding. *Proc. Natl. Acad. Sci. USA.* 93:12759–12761.
16. Pace, C. N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131:266–280.
17. Takeuchi, M., N. Inque, T. W. Strickland, M. Kubota, M. Wada, et al. 1989. Relationship between sugar chain structure and biological activity of recombinant human erythropoietin produced in Chinese hamster ovary cells. *Proc. Natl. Acad. Sci. USA.* 86:7819–7822.
18. Zitzewitz, J. A., O. Bilsel, J. Luo, B. E. Jones, and C. R. Matthews. 1995. Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry.* 34:12812–12819.
19. Bilsel, O., J. A. Zitzewitz, K. E. Bowers, and C. R. Matthews. 1999. Folding mechanism of the alpha-subunit of tryptophan synthase, an

alpha/beta barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry.* 38:1018–1029.

20. Jemth, P., C. M. Johnson, S. Gianni, and A. R. Fersht. 2008. Demonstration by burst-phase analysis of a robust folding intermediate in the FF domain. *Protein Eng. Des. Sel.* 21:207–214.
21. Bachmann, A., and T. Kiefhaber. 2005. Kinetic mechanisms in protein folding. *In* Protein Folding Handbook. J. Buchner and T. Kiefhaber, editors. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. 379–410.
22. Capaldi, A., M. C. Ramachandra Shastry, C. Kleanthous, H. Roder, and S. E. Radford. 2001. Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat. Struct. Biol.* 8:68–72.
23. Krieger, F., B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber. 2003. Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* 332:265–274.
24. Silow, M., and M. Oliveberg. 1997. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. USA.* 94:6084–6086.
25. Balbach, J., and F. X. Schmid. 2000. Proline isomerization and its catalysis in protein folding. *In* Mechanisms of Protein Folding. R. H. Pain, editor. Oxford University Press, New York. 212–249.
26. Wallace, L. A., and C. R. Matthews. 2002. Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys. Chem.* 101–102:113–131.
27. Kerwin, B. A., K. H. Aoki, M. Gonelli, and G. B. Strambinin. 2008. Differentiation of the local structure around tryptophan 51 and 64 in recombinant human erythropoietin by tryptophan phosphorescence. *Photochem. Photobiol.* In press.
28. Tanford, C. 1970. Protein denaturation. *Adv. Protein Chem.* 24:1–95.
29. Maki, K., K. Kamagata, and K. Kuwajima. 2005. Equilibrium and kinetically observed molten globule states. *In* Protein Folding Handbook. J. Buchner and T. Kiefhaber, editors. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. 856–883.
30. Wang, C., M. Eufemi, C. Turano, and A. Giartosio. 1996. Influence of carbohydrate moiety on the stability of glycoproteins. *Biochemistry.* 35:7299–7307.
31. Mitra, N., S. Sinha, T. N. C. Ramya, and A. Surolia. 2006. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends Biochem. Sci.* 31:156–163.
32. Tams, J. W., and K. G. Welinder. 1998. Glycosylation and thermodynamic versus kinetic stability of horseradish peroxidase. *FEBS Lett.* 421:234–236.