

# Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength

Elodie Monsellier and Hugues Bedouelle<sup>1</sup>

Unit of Molecular Prevention and Therapy of Human Diseases (CNRS FRE 2849), Institut Pasteur, 28 rue Docteur Roux, 75724 Paris Cedex 15, France

<sup>1</sup>To whom correspondence should be addressed.  
E-mail: hbedouel@pasteur.fr

**The fluorescence of tryptophan is used as a signal to monitor the unfolding of proteins, in particular the intensity of fluorescence and the wavelength of its maximum  $\lambda_{\max}$ . The law of the signal is linear with respect to the concentrations of the reactants for the intensity but not for  $\lambda_{\max}$ . Consequently, the stability of a protein and its variation upon mutation cannot be deduced directly from measurements made with  $\lambda_{\max}$ . Here, we established a rigorous law of the signal for  $\lambda_{\max}$ . We then compared the stability  $\Delta G(\text{H}_2\text{O})$  and coefficient of cooperativity  $m$  for a two-state equilibrium of unfolding, monitored with  $\lambda_{\max}$ , when the rigorous and empirical linear laws of the signal are applied. The corrective terms involve the curvature of the emission spectra at their  $\lambda_{\max}$  and can be determined experimentally. The rigorous and empirical values of the cooperativity coefficient  $m$  are equal within the experimental error for this parameter. In contrast, the rigorous and empirical values of the stability  $\Delta G(\text{H}_2\text{O})$  generally differ. However, they are equal within the experimental error if the curvatures of the spectra for the native and unfolded states are identical. We validated this analysis experimentally using domain 3 of the envelope glycoprotein of the dengue virus and the single-chain variable fragment (scFv) of antibody mAbD1.3, directed against lysozyme.**  
*Keywords:* cooperativity/denaturant/dengue virus/envelope glycoprotein/free energy/scFv antibody fragment/unfolding

## Introduction

The possibility of measuring the stability of proteins with precision finds many applications in fundamental and applied research. It has allowed one to understand and quantify the forces that contribute to the conformational stability of proteins in their aqueous environment and the effects of sequence changes on this stability (Alber, 1989; Pace *et al.*, 1996). The data on the stability of proteins and their mutants are important for developing reliable energy functions for proteins (Guerois *et al.*, 2002; Bava *et al.*, 2004). These force fields are used in algorithms to predict the structure or docking of proteins and to design new proteins and stabilizing changes. The precise measurement of stability is also important to understand and describe the unfolding and folding of proteins at an atomic resolution, by a combination of experimental and theoretical approaches, i.e. the analysis of the  $\Phi$  values and molecular dynamics (Fersht and Daggett, 2002).

By definition, the thermodynamic stability  $\Delta G$  of a protein is equal to the variation of free energy between its native and

unfolded states. It can be deduced from the constant of equilibrium between these two conformational states and thus from the measurement of concentrations. The stability depends on the physico-chemical conditions and must therefore be given in standard conditions, e.g.  $\Delta G(\text{H}_2\text{O})$  in aqueous buffer at 20°C. The concentration of the unfolded state is usually very low in physiological conditions; therefore, the values of the stability are measured in variable physico-chemical conditions and extrapolated to the standard conditions. A physical quantity that is sensitive to the conformational state of the protein, is used for the measurement of concentrations.

The fluorescences of tryptophan and tyrosine residues are sensitive to their electronic environment. Therefore, the intrinsic fluorescence of proteins is widely used to measure the concentrations of their different molecular states in a reaction of unfolding. Only very low concentrations of protein are needed, which minimizes protein aggregation. The most useful fluorescence signals are the intensity  $Y$  of the emitted light and the wavelength  $\lambda_{\max}$  at which this intensity is maximal. These two parameters are usually measured after excitation at a fixed wavelength (Eftink, 1994).

The use of the fluorescence intensity  $Y$  as a signal to measure the stability of proteins may present difficulties. The  $Y$  signal is a function of the protein concentration and is therefore sensitive to volumetric errors. The  $Y$  signals of the native state N and of the unfolded state U of a protein generally vary with the concentration of the denaturant and the description of this variation requires at least two parameters (Santoro and Bolen, 1988). The precise determination of these parameters requires a large number of experimental points and thus large amounts of protein material. The  $Y$  signals of states N and U are not always sufficiently different for precise measurements (Tan *et al.*, 1998; Dumoulin *et al.*, 2002; Ewert *et al.*, 2003). For example, the denaturation of different domains in a protein can lead to variations of  $Y$  that compensate each other. The use of the  $\lambda_{\max}$  signal avoids many of the above difficulties. This signal does not depend on the concentration of protein and increases monotonically during the unfolding. The  $\lambda_{\max}$  signals of states N and U are often independent of the denaturant concentration. Therefore, the description of an unfolding profile requires less parameters and protein material when it is monitored with  $\lambda_{\max}$  as compared with  $Y$ .

The quantitative analysis of the unfolding profiles is easier when the recorded signal is a linear function of both concentrations and specific signals of the component molecular species. The  $Y$  signal satisfies these conditions of linearity because it depends only on the light absorbed by the molecules and on their quantum yields of fluorescence (emitted photons/absorbed photons). In contrast, there is no simple law for the composition of the  $\lambda_{\max}$  signals. Numerous authors ignore this physical difficulty, apply a linear law of additivity to  $\lambda_{\max}$  and attempt, by this empirical approach, to derive the stability  $\Delta G(\text{H}_2\text{O})$  of proteins or the concentration  $x_{1/2}$  of denaturant

that gives half unfolding. Some authors justify such an empirical approach by the observation that either the intensities or quantum yields of fluorescence for states N and U of the protein under study are identical (Tan *et al.*, 1998; Ewert *et al.*, 2003). A theoretical study has shown that the error on  $\Delta G(\text{H}_2\text{O})$ , calculated empirically from measurements of  $\lambda_{\text{max}}$ , can reach 50% (Eftink, 1994). Several experimental studies have compared the values for the thermodynamic parameters of unfolding at equilibrium, calculated rigorously from  $Y$  data and empirically from  $\lambda_{\text{max}}$  data. These values are close in some studies and differ widely in others (Jager and Pluckthun, 1999; Jung *et al.*, 1999; Martineau and Betton, 1999; Dumoulin *et al.*, 2002).

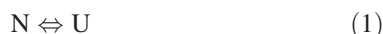
Hence the wavelength  $\lambda_{\text{max}}$  is a robust signal for monitoring the unfolding of proteins, but whether it allows one to derive reliable values of their stabilities remains to be demonstrated. In this study, we rigorously derived a law of composition for  $\lambda_{\text{max}}$  from that for  $Y$ . From this law, we could determine the correction that must be applied to the empirical value  $\Delta G'(\text{H}_2\text{O})$  of the stability, calculated by applying a linear law of the signal to  $\lambda_{\text{max}}$ . The corrective term depends on the curvatures of the emission spectra for states N and U at their respective  $\lambda_{\text{max}}$ . It can be easily determined and is not negligible in general.

We validated our theoretical analysis with two proteins. Domain 3 of the envelope glycoprotein E from serotype 1 of the dengue virus (E3.1, residues 296–400) has been implicated in the interactions between the virus and its cellular receptors (Mukhopadhyay *et al.*, 2005). The single-chain variable fragments scFv of antibodies are widely used in fundamental and applied research. Many studies aim at increasing the stability of scFvs, which is often limiting for applications. Such studies on scFv fragments require methods to compare precisely and reliably their stabilities and the recourse to the  $\lambda_{\text{max}}$  signal is often necessary and has been extensively used (Worn and Pluckthun, 2001). The scFv fragment of antibody mAbD1.3, directed against hen egg-white lysozyme, is a model system for fundamental studies and the development of new methodologies. Many structural and thermodynamic data are available on this system (Sundberg and Mariuzza, 2002).

## Theory

### Equilibrium of unfolding

Let P be a monomeric protein, N its native folded state and U its unfolded state. Let us assume that this protein unfolds according to the equilibrium



In physiological conditions, the protein is almost entirely in its native form N and the concentration of state U cannot be detected. To be studied, the equilibrium of unfolding is generally shifted with a denaturing agent, such as urea or guanidinium chloride (GdmCl). A new equilibrium forms for each concentration  $x$  of denaturant. An unfolding profile is obtained by measuring a signal of the protein, sensitive to its conformational state, as a function of  $x$ . The equations derived in this and the following paragraphs allow one to determine the concentrations of N and U for each value of  $x$ .

The laws of mass action and conservation give the two following equations, where  $K$  is an equilibrium constant and

$C$  (M) the total concentration of the protein:

$$[\text{U}]/[\text{N}] = K \quad (2)$$

$$[\text{N}] + [\text{U}] = C \quad (3)$$

Generally, it is more convenient to reason on molar fractions:

$$f_n = [\text{N}]/C \quad \text{and} \quad f_u = [\text{U}]/C \quad (4)$$

Equations 2 and 3 can be rewritten as

$$f_u/f_n = K \quad (5)$$

$$f_n + f_u = 1 \quad (6)$$

The variation of free energy  $\Delta G$  between states N and U is given by

$$\Delta G = -RT \ln(K) \quad (7)$$

where  $R$  is the gas constant and  $T$  is the temperature (K). By definition,  $\Delta G$  is the stability of protein N. Generally, one assumes that the variation of free energy between two conformational states is a linear function of  $x$  (Pace, 1986; Myers *et al.*, 1995):

$$\Delta G(x) = \Delta G(\text{H}_2\text{O}) - mx \quad (8)$$

Note that parameters  $f_n$ ,  $f_u$ ,  $K$  and  $\Delta G$  are functions of  $x$ . Let  $x_{1/2}$  be the concentration of denaturant that results in half-advancement of the unfolding reaction, i.e.  $f_n(x_{1/2}) = 0.5$ . Under these conditions, Equations 5–7 show that the stability  $\Delta G$  of the protein is zero and Equation 8 shows that the value of  $x_{1/2}$  is given by

$$x_{1/2} = \Delta G(\text{H}_2\text{O})/m \quad (9)$$

### Law of the signal: fluorescence intensity

Let us assume that the intensity of fluorescence, for a set excitation radiation, is used to monitor the unfolding equilibrium of Equation 1. If  $Y_t(\lambda, x)$  is the global signal of the unfolding mixture, the law of additivity of the signals applies:

$$Y_t(\lambda, x) = [\text{N}]Y_n(\lambda, x) + [\text{U}]Y_u(\lambda, x) + xY_d(\lambda) \quad (10)$$

where  $\lambda$  is the wavelength at which the fluorescence emission is measured and  $Y_n$ ,  $Y_u$  and  $Y_d$  are the molar signals of state N, state U and the denaturant, respectively. The signal of the denaturant alone is generally measured in a separate experiment and only the protein signal  $Y(\lambda, x)$  is considered:

$$Y = Y_t - xY_d = [\text{N}]Y_n + [\text{U}]Y_u \quad (11)$$

Equation 11 can be rewritten with molar fractions as follows:

$$Y = C(f_n Y_n + f_u Y_u) \quad (12)$$

Experimentally, one observes that  $Y(\lambda, x)$  is a linear function of  $x$  at low and high concentrations of denaturant (Santoro and Bolen, 1988) and therefore one can write for every  $x$

$$Y_n(\lambda, x) = y_n(\lambda) + xm_n(\lambda) = y_n(\lambda)[1 + xh_n(\lambda)] \quad (13)$$

$$Y_u(\lambda, x) = y_u(\lambda) + xm_u(\lambda) = y_u(\lambda)[1 + xh_u(\lambda)] \quad (14)$$

where  $h_n = m_n/y_n$  and  $h_u = m_u/y_u$  are intrinsic parameters of the protein. Generally, one monitors the unfolding reaction at the wavelength  $\lambda_\Delta$  such that

$$|Y(\lambda_\Delta, 0) - Y(\lambda_\Delta, x_{\max})| = \max_\lambda |Y(\lambda, 0) - Y(\lambda, x_{\max})| \quad (15)$$

where  $x_{\max}$  is the highest concentration of denaturant attainable.

#### Law of the signal: $\lambda_{\max}$

For a given concentration  $x$  of denaturant and variable values of the wavelength  $\lambda$ ,  $Y(\lambda, x)$  represents the emission spectrum of protein P. The wavelength at which the intensity  $Y(\lambda, x)$  of the emitted light is maximum is denoted  $\lambda_{\max}(x)$ . Then, if  $Y(\lambda, x)$  is approximated by a continuous differentiable function of  $\lambda$ :

$$Y(\lambda_{\max}, x) = \max_\lambda Y(\lambda, x) \quad \text{and} \quad (\partial Y/\partial \lambda)(\lambda_{\max}, x) = 0 \quad (16)$$

The differentiation of Equation 12 gives

$$\partial Y/\partial \lambda = C(f_n \partial Y_n/\partial \lambda + f_u \partial Y_u/\partial \lambda) \quad (17)$$

Let  $\lambda_n$  and  $\lambda_u$  be the  $\lambda_{\max}$  values for states N and U of protein P, respectively. If both  $Y_n$  and  $Y_u$  are increasing functions of  $\lambda$  for  $\lambda < \lambda_n$  and decreasing functions for  $\lambda_u < \lambda$ , then Equations 16 and 17 imply

$$\lambda_n \leq \lambda_{\max} \leq \lambda_u \quad (18)$$

Hence  $\lambda_{\max}$  remains within the interval of wavelengths  $[\lambda_n, \lambda_u]$  for any mixture of states N and U.

The  $Y(\lambda, x)$  function can be written as a Taylor expansion about  $\lambda = \lambda_{\max}$  (Weisstein, 2002; <http://mathworld.wolfram.com/TaylorSeries.html>). For many proteins, the fourth-order remainder of the Taylor expansion is negligible and their fluorescence spectra can be approximated over a wide interval of wavelengths by the following cubic function (see Results):

$$\begin{aligned} Y(\lambda, x) \approx & Y(\lambda_{\max}, x) + [(\lambda - \lambda_{\max})/1!][(\partial Y/\partial \lambda)(\lambda_{\max}, x)] \\ & + [(\lambda - \lambda_{\max})^2/2!][(\partial^2 Y/\partial \lambda^2)(\lambda_{\max}, x)] \\ & + [(\lambda - \lambda_{\max})^3/3!][(\partial^3 Y/\partial \lambda^3)(\lambda_{\max}, x)] \end{aligned} \quad (19)$$

Equation 19 can be simplified if one takes Equation 16 into account:

$$Y(\lambda, x) \approx a(x) + [(\lambda - \lambda_{\max})^2/2!]b(x) + [(\lambda - \lambda_{\max})^3/3!]c(x) \quad (20)$$

with  $a(x) = Y(\lambda_{\max}, x)$ ,  $b(x) = (\partial^2 Y/\partial \lambda^2)(\lambda_{\max}, x)$  and  $c(x) = (\partial^3 Y/\partial \lambda^3)(\lambda_{\max}, x)$ . In particular, the fitting of Equation 20 to the spectra of states N and U enables one to determine precise values of  $\lambda_n$  and  $\lambda_u$ , respectively, based on an extended portion of each spectrum (see Results and Figure 2).

Once the values of  $\lambda_n$  and  $\lambda_u$  are known with precision, the molar spectra of states N and U can generally be approximated on the interval  $[\lambda_n, \lambda_u]$  by the following quadratic functions, obtained by neglecting the third-order remainder of a Taylor expansion (see Results and Figure 3):

$$Y_n(\lambda, x) = a_n(x) + 0.5(\lambda - \lambda_n)^2 b_n(x) \quad (21)$$

$$Y_u(\lambda, x) = a_u(x) + 0.5(\lambda - \lambda_u)^2 b_u(x) \quad (22)$$

With these approximations, which should be checked in each particular case and for  $\lambda$  belonging to  $[\lambda_n, \lambda_u]$ , Equations 12, 21

and 22 give

$$Y(\lambda, x) \approx C\{f_n[a_n + 0.5(\lambda - \lambda_n)^2 b_n] + f_u[a_u + 0.5(\lambda - \lambda_u)^2 b_u]\} \quad (23)$$

According to Equations 16, 18 and 23,  $\lambda_{\max}(x)$  is a solution of the equation

$$(\partial Y/\partial \lambda)(\lambda_{\max}, x) \approx C[(\lambda_{\max} - \lambda_n)b_n f_n + (\lambda_{\max} - \lambda_u)b_u f_u] = 0 \quad (24)$$

This solution is given by

$$\lambda_{\max}(x) \approx \lambda_n b_n f_n (b_n f_n + b_u f_u)^{-1} + \lambda_u b_u f_u (b_n f_n + b_u f_u)^{-1} \quad (25)$$

#### Comparison of the approximate and empirical equations

The stability of a protein P, unfolding according to Equation 1, is often deduced from the following set of empirical equations, drawn by homology with Equations 5–8 and 12 (see Introduction):

$$f'_u/f'_n = K' \quad (26)$$

$$f'_n + f'_u = 1 \quad (27)$$

$$\Delta G' = \Delta G'(\text{H}_2\text{O}) - m'x = -RT \ln K' \quad (28)$$

$$\lambda_{\max} = f'_n \lambda_n + f'_u \lambda_u \quad (29)$$

where the corresponding empirical (or apparent) parameters are labeled with a prime. Comparison of Equations 29 and 25 shows that the empirical parameters  $f'_n$  and  $f'_u$  are related to the molar fractions  $f_n$  and  $f_u$  of states N and U by

$$f'_n \approx b_n f_n (b_n f_n + b_u f_u)^{-1} \quad \text{and} \quad f'_u \approx b_u f_u (b_n f_n + b_u f_u)^{-1} \quad (30)$$

From Equations 26, 30 and 5, one deduces

$$K' = f'_u/f'_n \approx (b_u/b_n)(f_u/f_n) = (b_u/b_n)K \quad (31)$$

and from Equations 28, 31 and 7, for every value of  $x$ :

$$\Delta G'(x) = -RT \ln [K'(x)] \approx \Delta G(x) - RT \ln [b_u(x)/b_n(x)] \quad (32)$$

In particular, for  $x = 0$ :

$$\Delta G(\text{H}_2\text{O}) \approx \Delta G'(\text{H}_2\text{O}) + RT \ln [b_u(0)/b_n(0)] \quad (33)$$

Equation 33 shows that the stability of protein P is not equal to the empirical parameter  $\Delta G'(\text{H}_2\text{O})$  and there is a corrective term in general.

If  $\Delta G'(x)$  and  $\Delta G(x)$  in Equation 32 are replaced by their expressions in Equations 28 and 8 and  $\Delta G(\text{H}_2\text{O})$  in Equation 8 by its expression in Equation 33, one obtains

$$\begin{aligned} m = m' + (RT/x)\{ & \ln [b_n(x)/b_n(0)] \\ & - \ln [b_u(x)/b_u(0)] \} \quad \text{for every } x \end{aligned} \quad (34)$$

Equation 34 allows one to calculate the cooperativity of unfolding  $m$  from the empirical parameter  $m'$ . Let  $x'_{1/2}$  be the concentration of denaturant that gives  $f'_n(x'_{1/2}) = 0.5$ .

From Equations 26–28, it follows that

$$x'_{1/2} = \Delta G'(\text{H}_2\text{O})/m' \quad (35)$$

Equations 9, 33 and 35 allow one to compare  $x'_{1/2}$  with  $x_{1/2}$ :

$$x_{1/2} \approx (m'/m)\{x'_{1/2} + (RT/m')\ln[b_u(0)/b_n(0)]\} \quad (36)$$

Hence the empirical concentration  $x'_{1/2}$  is not equal to the concentration  $x_{1/2}$  of denaturant that results in half-advancement of the unfolding reaction.

### Geometric interpretation

The curvature  $\kappa$  of any curve  $Y(\lambda)$  in the plane is given by (Weisstein, 2002; <http://mathworld.wolfram.com/Curvature.html>)

$$\kappa = (\partial^2 Y/\partial \lambda^2)/[1 + (\partial Y/\partial \lambda)^2]^{3/2} \quad (37)$$

Equations 16 and 37 imply, if  $Y(\lambda, x)$  is a fluorescence intensity,

$$\kappa(\lambda_{\max}, x) = (\partial^2 Y/\partial \lambda^2)(\lambda_{\max}, x) \text{ for every } x \quad (38)$$

Equations 20 and 38 imply that parameters  $b_n(x)$  and  $b_u(x)$  are the curvatures of the fluorescence spectra for states N and U of protein P at their respective  $\lambda_{\max}$  and at a concentration  $x$  of denaturant. Therefore, the correcting factor in Equation 33 depends on the ratio of these curvatures in the absence of denaturant. Note that the ratio  $b_u/b_n$  does not depend on the concentration of the protein or the spectrofluorimeter or its setup. Below, we give methods for determining the values of  $b_n$  and  $b_u$  experimentally.

### Curvature versus concentration in urea

For simplicity, we can rewrite Equation 33 as follows:

$$\Delta G(\text{H}_2\text{O}) \approx \Delta G'(\text{H}_2\text{O}) - E \quad (39)$$

$$E = RT \ln[b_n(0)/b_u(0)] \quad (40)$$

The corrective term  $E$  involves the curvature  $b_u(0)$  of the spectrum for state U in the absence of denaturant, which cannot be measured directly. Equation 40 can be rewritten, for every  $x$ , as follows:

$$E = RT \ln[b_n(0)/b_u(x)] + RT \ln[b_u(x)/b_u(0)] \quad (41)$$

Let us assume that the curvature  $b_u(x)$  of the spectrum for state U varies linearly with  $x$  over the whole interval of the concentration in denaturant and that the curvature  $b_n(x)$  of the spectrum for state N varies linearly with  $x$  in the pre-transition region (see Results for justifications). Then

$$b_u(x)/b_u(0) = 1 + k_u x \text{ for every } x \quad (42)$$

$$b_n(x)/b_n(0) = 1 + k_n x \text{ for } x \text{ small} \quad (43)$$

where  $k_n$  and  $k_u$  are intrinsic parameters of the protein. The combination of Equations 41 and 42 gives an expression for  $E$  where  $x_{\max}$  is the maximum concentration of denaturant attainable (e.g. 8 M urea) and whose every term can be measured experimentally:

$$E = RT \ln[b_n(0)/b_u(x_{\max})] + RT \ln(1 + k_u x_{\max}) \quad (44)$$

The combination of Equations 34, 42 and 43 gives, for  $x$  belonging to the pre-transition region,

$$m = m' + (RT/x)[\ln(1 + k_n x) - \ln(1 + k_u x)] \quad (45)$$

As  $\ln(1 + kx) \approx kx$  in the neighborhood of  $x = 0$ , Equation 45 gives

$$m = m' + RT(k_n - k_u) \quad (46)$$

The combination of Equations 36, 40 and 46 gives

$$x_{1/2} = (x'_{1/2} - E/m')/[1 + (k_n - k_u)RT/m'] \quad (47)$$

Note that Equation 46 does not depend on the exact form of Equation 43.

## Materials and methods

### Bacterial strains, plasmids and media

The *Escherichia coli* strains HB2151 (Carter *et al.*, 1985) and RZ1032 (Kunkel *et al.*, 1987) and plasmid pMR1 (Renard *et al.*, 2002) have been described. mAbD1.3 is a murine monoclonal antibody, directed against hen egg-white lysozyme. pMR1 codes for a single-chain scFv fragment of mAbD1.3, in the format NH<sub>2</sub>-V<sub>H</sub>-(Gly<sub>4</sub>Ser)<sub>3</sub>-V<sub>L</sub>-H<sub>6</sub>-COOH, where V<sub>H</sub> and V<sub>L</sub> are the variable domains of the heavy chain and light chain, respectively, and H<sub>6</sub> represents a hexahistidine tag. In pMR1, the expression of the scFvD1.3-H<sub>6</sub> gene is under control of the *tet* promoter and *ompA* signal sequence from *E.coli*. The sequence of the recombinant scFvD1.3-H<sub>6</sub> gene differed slightly from the published sequences at the 5'- and 3'-ends of the constitutive V<sub>H</sub> and V<sub>L</sub> genes, as a result of the cloning steps (Figure 1). Plasmid pLB11 is a derivative of the pET20b+ vector (Novagen) and codes for a hybrid E3.1-H<sub>6</sub> between domain 3 (residues 296–400) of the envelope glycoprotein E from the dengue virus (serotype 1) and a hexahistidine tag (Despres *et al.*, 1993; H.Bedouelle *et al.*, in preparation). The E3.1-H<sub>6</sub> domain comprises a unique disulfide bridge between residues Cys302 and Cys333. Buffer A was 50 mM Tris-HCl, pH 7.9, 150 mM NaCl. Ultrapure urea and guanidine hydrochloride (GdmCl) were purchased from MP Biochemicals. Solutions of urea and GdmCl were freshly prepared daily. The concentrations of urea or GdmCl were measured with a refractometer with a precision of 0.01 M.

### Proteins and general conditions

The E3.1-H<sub>6</sub> and scFvD1.3-H<sub>6</sub> recombinant proteins were produced from plasmids pLB11 and pMR1, respectively, in

	1	2	3	4	5	6	7	8
VH:	GAA	GTT	AAA	CTG	CAG	GAG	TCA	GGA
	109	110	111	112	113	114	115	116
VH:	GGG	ACC	ACG	GTC	ACC	GTC	TCC	TCA
	1	2	3	4	5	6	7	8
VL:	GAC	ATC	GAG	CTC	ACC	CAG	TCT	CCA
	101	102	103	104	105	106	107	108
VL:	GGG	ACC	AAG	CTC	GAG	ATC	AAG	CGG

**Fig. 1.** Modifications of the scFv genes in plasmid pMR1. The DNA sequences of the first eight residues and last eight residues of the V<sub>H</sub> and V<sub>L</sub> genes in plasmid pMR1, coding for scFvD1.3-H<sub>6</sub>, differed slightly from the published sequences as a result of the cloning steps (England *et al.*, 1999).

the periplasmic space of strain HB2151. They were purified by nickel ion chromatography as described (Renard *et al.*, 2002; H.Bedouelle *et al.*, in preparation). The protein fractions were analyzed by SDS-PAGE in denaturing conditions. The concentration of acrylamide-bisacrylamide (29:1) was 15% for scFvD1.3-H6 and 17% for domain E3.1-H6. The fractions that were homogeneous at >95% were pooled, dialyzed against buffer A, snap frozen in liquid nitrogen and stored at  $-70^{\circ}\text{C}$ . The concentration of protein in the purified preparations was measured by absorbance spectrometry. The extinction coefficients were calculated as described (Pace *et al.*, 1995):  $\epsilon_{280\text{nm}}(\text{E3.1-H6}) = 9530 \text{ mM}^{-1} \text{ cm}^{-1}$  and  $\epsilon_{280\text{nm}}(\text{scFvD1.3-H6}) = 51130 \text{ mM}^{-1} \text{ cm}^{-1}$ .

Unfolding with urea was performed as described (Pace, 1986). Each reaction mixture (1 ml) contained purified protein (10  $\mu\text{g/ml}$ ; 0.80  $\mu\text{M}$  for E3.1-H6 and 0.37  $\mu\text{M}$  for scFvD1.3-H6) and varying concentrations of urea (0–9 M) in buffer A. Control reactions were prepared by replacing the protein with buffer. The mixtures were incubated for 14 h at  $20^{\circ}\text{C}$  to enable the reactions of unfolding to reach equilibrium. To test the reversibility of the unfolding reaction, a protein sample (10  $\mu\text{g}$ ) was denatured in 7 M urea and buffer A for 4 h. The denatured protein was diluted with buffer A to reach a final concentration of urea between 7 and 1 M. The diluted mixture was then incubated for 14 h at  $20^{\circ}\text{C}$  to enable the reaction to reach equilibrium as above. The concentration of urea was measured in each reaction mixture after the completion of each experiment, as described above.

### Fluorescence measurements

Fluorescence experiments were performed at  $20^{\circ}\text{C}$  with a Perkin-Elmer LS-5B spectrofluorimeter. The proteins were excited at 278 nm and the amino acid tryptophan at 290 nm; the slit width was 2.5 nm for excitation and 5 nm for emission. The fluorescence spectra were recorded in the interval 320–370 nm for scFvD1.3-H6 and E3.1-H6 and 310–374 nm for tryptophan. The signal was acquired for 2 s at each wavelength and the increment of wavelength was 0.5 nm. The fluorescence signal for the protein or tryptophan was obtained by subtraction of the signal for the solvent alone. In a first step, each spectrum  $Y(\lambda, x)$ , where  $x$  was fixed and  $\lambda$  variable, was approximated over the whole interval of wavelength [ $\lambda_n - 20 \text{ nm}$ ,  $\lambda_u + 20 \text{ nm}$ ] by the fitting of Equation 20 to the experimental data, with  $a$ ,  $b$ ,  $c$  and  $\lambda_{\max}$  as floating parameters. In particular,  $\lambda_n = \lambda_{\max}(0)$  and  $\lambda_u = \lambda_{\max}(x_{\max})$  were determined in this way for  $x$  equal to 0 M and  $x_{\max}$  M of denaturant, respectively. In a second step, the  $Y(\lambda, x)$  spectrum was approximated on the narrower interval [ $\lambda_n - 2 \text{ nm}$ ,  $\lambda_u + 2 \text{ nm}$ ] by the fitting of Equation 21 or 22 with  $a$  and  $b$  as floating parameters and  $\lambda_{\max}(x)$  set to the value that had been determined in the first step. This procedure allowed us to optimize the  $b_n(x)$  and  $b_u(x)$  parameters in this narrower interval of wavelength to which  $\lambda_{\max}(x)$  necessarily belonged (Equation 18).

### Analysis of the unfolding profiles

The solution of Equations 5 and 6 is given by

$$f_n = 1/(1 + K) \quad \text{and} \quad f_u = K/(1 + K) = 1/(1 + K^{-1}) \quad (48)$$

The combination of Equations 6 and 12–14 gives

$$Y = C\{y_u + m_u x + [y_n - y_u + (m_n - m_u)x]f_n\} \quad (49)$$

The combination of Equations 7, 8 and 48 gives

$$f_n = 1/(1 + \exp\{[mx - \Delta G(\text{H}_2\text{O})]/RT\}) \quad (50)$$

Equation 49, where  $f_n$  is developed as in Equation 50 and which relates the intensity of fluorescence to the concentration  $x$  of urea, was fitted to the unfolding data with  $y_n$ ,  $m_n$ ,  $y_u$ ,  $m_u$ ,  $m$  and  $\Delta G(\text{H}_2\text{O})$  as floating parameters (see below).

Similarly, the combination of Equations 26–29 gives

$$\lambda_{\max} = \lambda_n + (\lambda_u - \lambda_n)f'_u \quad (51)$$

$$f'_u = 1/(1 + \exp\{[\Delta G'(\text{H}_2\text{O}) - m'x]/RT\}) \quad (52)$$

where  $\lambda_u$  is larger than  $\lambda_n$ . One generally observes experimentally that  $\lambda_n$  and  $\lambda_u$  do not vary with  $x$ . Equation 51, where  $f'_u$  is developed as in Equation 52 and which relates  $\lambda_{\max}$  to  $x$ , was fitted to the unfolding data with  $\lambda_n$ ,  $\lambda_u$ ,  $m'$  and  $\Delta G'(\text{H}_2\text{O})$  as floating parameters (see below).

### Calculations

The curve fits were performed with the Kaleidagraph program (Synergy Software), which uses a Levenberg–Marquardt algorithm. We used the general curve fit routine and the corresponding Pearson's coefficient of correlation,  $R_p$ . The three-dimensional structures of the variable fragment FvD1.3 (PDB 1vfa; Bhat *et al.*, 1994) and of domain E3.2 from serotype 2 of the dengue virus (PDB 1oan; Modis *et al.*, 2003) were analyzed with the WHAT IF program as described (Vriend, 1990; Renard *et al.*, 2002).

## Results

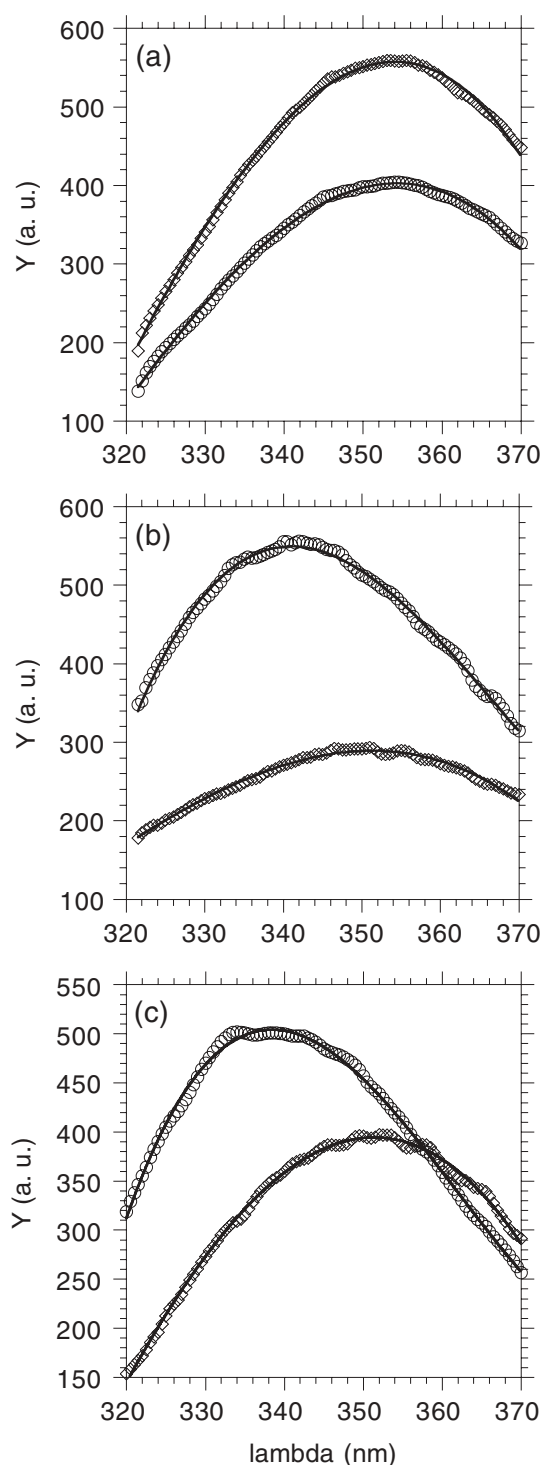
### Fluorescence of tryptophan in solution

The concentration of the unfolded state U of a protein is generally negligible and undetectable in the absence of a denaturing agent. Therefore, the properties of state U are extrapolated from measurements performed at high concentration of denaturant. The residues of tryptophan are exposed to the solvent in the U state of proteins. Therefore, we assumed that their properties of fluorescence in state U could be mimicked by those of the amino acid tryptophan in solution. We therefore determined the fluorescence properties of tryptophan and their variations with the concentration  $x$  of the denaturant, either urea or guanidine hydrochloride (GdmCl).

Solutions of the amino acid tryptophan were prepared in  $x$  M urea, with  $x$  varying between 0 and 8 M. Tryptophan was excited at 290 nm and its fluorescence emission spectrum was recorded at  $20^{\circ}\text{C}$  for each value of  $x$ . The maximal fluorescence emission intensity,  $\max_{\lambda} Y(x, \lambda) = Y[x, \lambda_{\max}(x)]$  and the wavelength  $\lambda_{\max}(x)$  of this maximum were determined by fitting the cubic function of Equation 20 to the spectrum on the interval of wavelengths 310–374 nm. The Pearson's coefficient for the fitting was  $R_p > 0.9985$  for every  $x$  (Figure 2a). We found that  $\lambda_{\max}(x)$  did not vary significantly with  $x$  and its value was equal to  $354.17 \pm 0.02 \text{ nm}$  (mean  $\pm$  SE) in these experiments with urea. In contrast,  $Y[x, \lambda_{\max}(x)]$  increased with the concentration of urea (see Figure 4) according to the linear law

$$Y[x, \lambda_{\max}(x)]/Y[0, \lambda_{\max}(0)] = 1 + h_{Wx} \quad (53)$$

with  $h_{W, \text{urea}} = 0.050 \pm 0.001 \text{ M}^{-1}$  (mean  $\pm$  SE in the curve fit). Such a linear variation of  $Y[x, \lambda_{\max}(x)]$  has already been reported for tryptophan and *N*-acetyl-L-tryptophanamide, with



**Fig. 2.** Determination of the  $\lambda_{\max}$  value by the fitting of Equation 20 to the emission spectra. Circles, 0 M urea; diamonds, 8 M urea; (a) amino acid tryptophan; (b) E3.1; (c) scFvD1.3. Excitation was at 290 nm for tryptophan and 278 nm for the proteins.

similar values of  $h_{W,\text{urea}}$  (Schmid, 1989; Eftink, 1994). This variation justifies the assumption of linear baselines for the transitions of unfolding, induced with urea and monitored by fluorescence (Equations 13 and 14). The value of the relative slope  $h_{W,\text{urea}}$  for the amino acid tryptophan was consistent with those for the baselines of the two proteins that we studied here (see the values of  $h_n$  and  $h_u$  in Table I). The quadratic function of Equation 22 was then fitted to each spectrum over the

**Table I.** Fluorescence parameters of states N (0 M urea) and U (8 M urea) of the proteins under study

Parameter	E3.1	scFvD1.3
$\lambda_{\Delta}$ (nm)	332	330
$Y_n(0)/Y_u(8)$	$2.5 \pm 0.2$	$1.5 \pm 0.1$
$h_n$ ( $M^{-1}$ )	$0.04 \pm 0.02$	$0.04 \pm 0.02$
$h_u$ ( $M^{-1}$ )	$0.12 \pm 0.05$	$0.07 \pm 0.02$
$\lambda_n$ (nm)	$340 \pm 1$	$338.6 \pm 0.2$
$\lambda_u$ (nm)	$350 \pm 1$	$350.7 \pm 0.2$
$b_n(0)/b_u(8)$	$2.5 \pm 0.2$	$0.96 \pm 0.06$
$k_n$ ( $M^{-1}$ )	$-0.02 \pm 0.04$	$0.01 \pm 0.01$

Excitation was at 278 nm and the temperature was 20°C. The subscripts n and u refer to states N and U, respectively.  $\lambda_{\Delta}$ , see Equation 15;  $h_n$  and  $h_u$ , see Equations 13 and 14. The parameters  $Y_n(0)/Y_u(8)$ ,  $h_n$  and  $h_u$  are related to the fluorescence intensity at the emission wavelength  $\lambda_{\Delta}$ .  $\lambda_n$  and  $\lambda_u$ ,  $\lambda_{\max}$  values of the emission spectra;  $b_n(0)$  and  $b_u(8)$ , curvatures of the emission spectra at  $\lambda_n$  and  $\lambda_u$ , respectively;  $k_n$ , see Equation 43. Mean and SE obtained from four to six independent experiments.

narrower interval 328–356 nm, with the value of  $\lambda_{\max}(x)$  fixed at 354.17 nm. We chose this interval because it contains the values of  $\lambda_{\max}$  for most proteins, whatever their folding state. The approximation of the tryptophan spectrum by Equation 22 on this interval was excellent for every  $x$  ( $R_P > 0.9985$ ; Figure 3a). We found that the curvature  $b_W(x)$  of the spectrum at  $\lambda_{\max}(x)$  varied significantly with the concentration  $x$  of urea (Figure 4), according to the linear law

$$b_W(x)/b_W(0) = 1 + k_W x \quad (54)$$

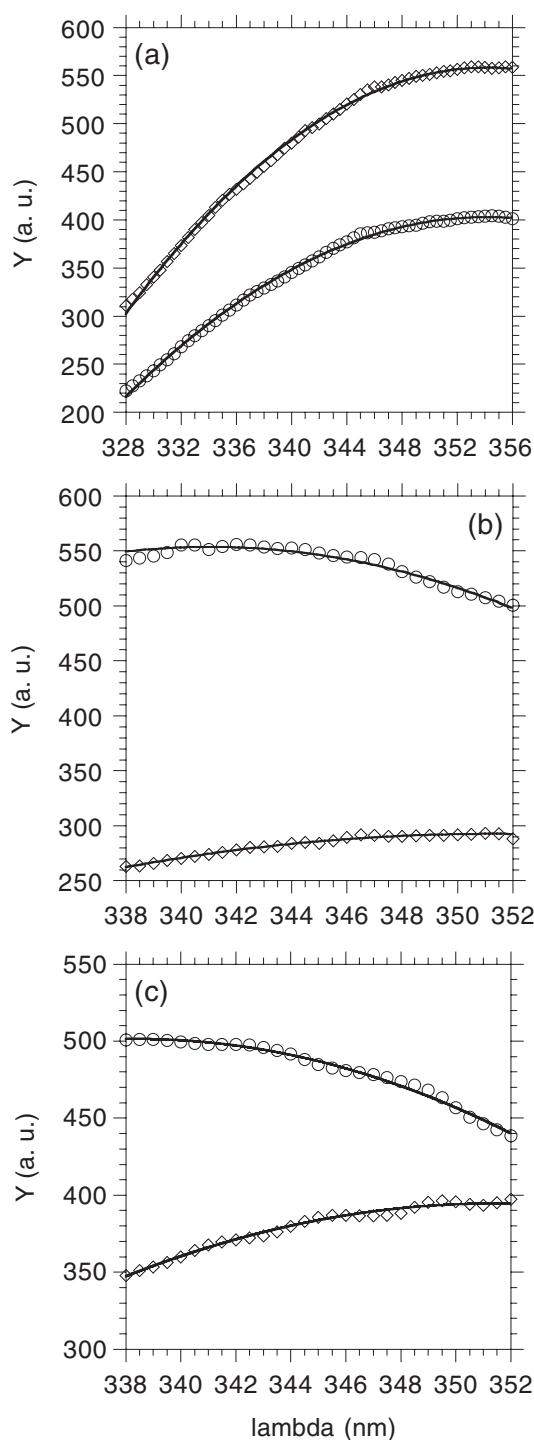
with  $k_{W,\text{urea}} = 0.0485 \pm 0.0011 M^{-1}$ .

Similarly, solutions of the amino acid tryptophan were prepared in  $x$  M GdmCl, with  $x$  varying between 0 and 6 M. We found that the maximal intensity  $Y[x, \lambda_{\max}(x)]$ , wavelength  $\lambda_{\max}(x)$  and curvature  $b_W(x)$  did not vary significantly with the concentration  $x$  of GdmCl ( $h_{W,\text{GdmCl}} = 0.0014 \pm 0.0009 M^{-1}$ ,  $\lambda_{\max} = 355.28 \pm 0.03$  nm and  $k_{W,\text{GdmCl}} = -0.001 \pm 0.002 M^{-1}$ , respectively). The negligible variation of  $Y[x, \lambda_{\max}(x)]$  with GdmCl has already been reported for both tryptophan and *N*-acetyl-L-tryptophanamide (Schmid, 1989; Eftink, 1994). *N*-Acetyl-L-tryptophanamide is sometimes used to avoid the charged  $\text{NH}_2$  and  $\text{COOH}$  groups that are present in the amino acid tryptophan but not in the corresponding protein residues.

### Unfolding profiles of two model proteins

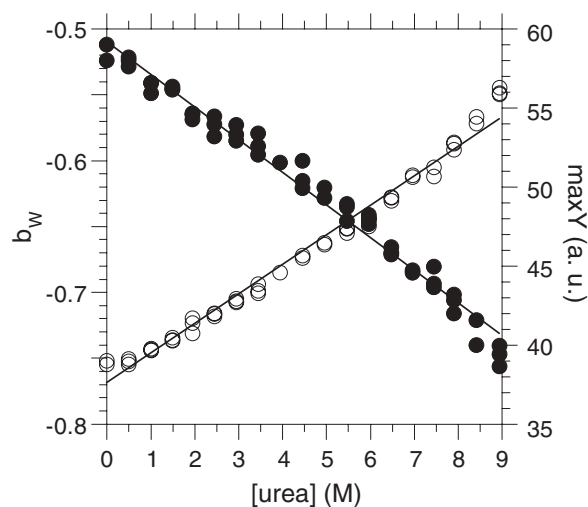
Domain E3.1 from serotype 1 of the dengue virus and the antibody fragment scFvD1.3 were produced in the periplasmic space of *E. coli* and purified by affinity chromatography on a nickel ion column, through a C-terminal hexahistidine tag (see Materials and methods). The purified preparations of proteins were homogeneous at >95%, as checked by SDS-PAGE. The proteins were incubated in increasing concentrations  $x$  of urea, used as a denaturant, and their fluorescence properties were characterized. Each experiment was repeated 4–6 times from independent preparations of protein. The reversibility of the unfolding reactions was verified.

We followed the unfolding with both  $Y(x)$ , the intensity of fluorescence emission by the reaction mixture at a fixed wavelength and  $\lambda_{\max}(x)$ , the wavelength of the maximal intensity (Figure 5). The  $Y(x)$  signal was measured at the emission wavelength  $\lambda_{\Delta}$  for which the difference between states N (0 M urea) and U (8 M urea) was maximal (Equation 15). The values of  $\lambda_{\max}(x)$  were determined by fitting Equation 20 to the emission

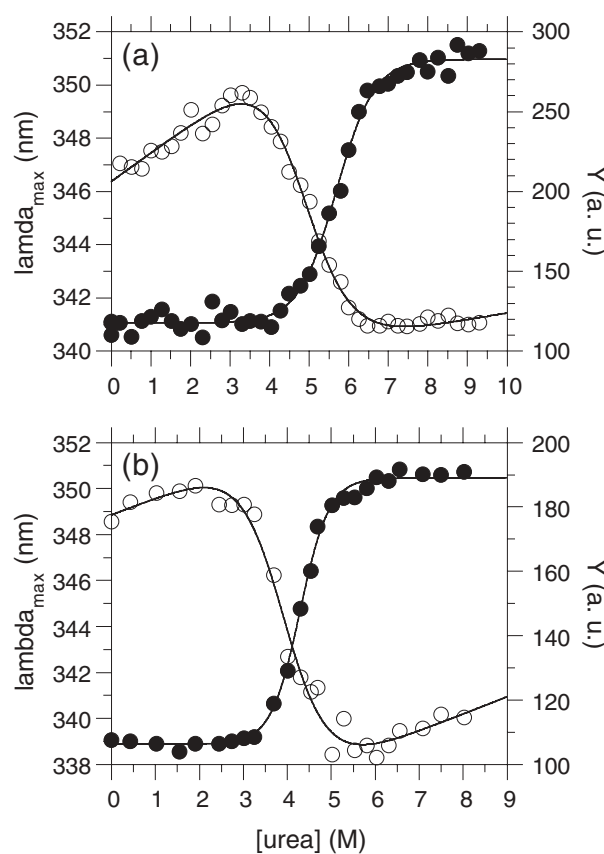


**Fig. 3.** Determination of the curvatures  $b_n(0)$  and  $b_u(8)$  by the fittings of Equations 21 and 22 to the emission spectra. Circles,  $b_n(0)$ ; diamonds,  $b_u(8)$ ; (a) amino acid tryptophan; (b) E3.1; (c) scFvD1.3. The fittings were performed on the restricted interval  $[\lambda_n - 2 \text{ nm}, \lambda_u + 2 \text{ nm}]$  with set values of  $\lambda_n$  and  $\lambda_u$  (see Materials and methods). Excitation was at 290 nm for tryptophan and 278 nm for the proteins.

spectra over the interval  $[\lambda_n - 20 \text{ nm}, \lambda_u + 20 \text{ nm}]$ , where  $\lambda_n$  and  $\lambda_u$  were the values of  $\lambda_{\max}(x)$  for  $x = 0$  and 8 M urea, respectively ( $R_p > 0.99$ ; Figure 2). The number of terms in Equation 20 and the interval of wavelengths that are used in the fitting should be adjusted for each particular protein. Here, we found that the use of wider or narrower intervals increased the error on  $\lambda_{\max}(x)$  in the fitting. The fitting of a second-power polynomial over the same interval of wavelength decreased the  $R_p$



**Fig. 4.** Dependences of the maximal intensity of fluorescence  $Y$  and curvature  $b_W$  on the concentration of urea for the spectra of the amino acid tryptophan. Open circles,  $\max Y$ ; closed circles,  $b_{W,\text{urea}}$ . Excitation was at 290 nm. The dependences followed the linear laws:  $Y[x, \lambda_{\max}(x)]/Y[0, \lambda_{\max}(0)] = 1 + h_W x$ , with  $R_p = 0.991$  and  $h_{W,\text{urea}} = 0.050 \pm 0.001 \text{ M}^{-1}$ ; and  $b_W(x)/b_W(0) = 1 + k_W x$ , with  $R_p = 0.990$  and  $k_W = 0.0485 \pm 0.0011 \text{ M}^{-1}$  (SE in the curve fit).



**Fig. 5.** Unfolding profiles of two proteins, monitored with both fluorescence intensity  $Y$  and wavelength  $\lambda_{\max}$ . Open circles,  $Y$ ; closed circles,  $\lambda_{\max}$ . (a) E3.1; (b) scFvD1.3. The curves result from the fittings of Equations 49 and 51 to the data points.

coefficient whereas that of a fourth-power polynomial left  $R_p$  unchanged but increased the errors on the fitting parameters. The characteristic parameters of states N and U are summarized in Table I. We found that  $\lambda_{\max}(x)$  remained constant outside the transition region for the two proteins under study,

i.e. its value remained equal to  $\lambda_n$  in the pre-transition region and to  $\lambda_u$  in the post-transition region (Figure 5).

The ratio  $Y_n(0)/Y_u(8)$  of the  $Y(x)$  signal for states N and U of domain E3.1 was important (2.5-fold) whereas the difference  $\lambda_u - \lambda_n$  of the  $\lambda_{\max}$  signals for the two states was moderate (10 nm). Both signals allowed us to monitor the unfolding of E3.1 with sensitivity. The wavelength  $\lambda_n$  of state N had a high value, 340 nm, and the molar intensity  $Y_n$  of this state increased strongly with the concentration of urea (Figure 5). The high value of  $\lambda_n$  was consistent with the known structural data. Indeed, domain E3.1 comprises only one Trp residue, which is conserved between the four serotypes of the dengue virus and partially exposed (20.7%) to the solvent in the crystal structure of glycoprotein E from serotype 2 (Modis *et al.*, 2003).

Both values of  $Y_n(0)/Y_u(8)$  and  $\lambda_u - \lambda_n$  for the scFvD1.3 fragment were moderate, 1.5-fold and 12 nm respectively. The wavelength  $\lambda_n$  of state N had also a high value, 339 nm. The scFvD1.3 fragment comprises six Trp residues. H-Trp52 in the variable domain  $V_H$  of the heavy chain and L-Trp92 in the variable domain  $V_L$  of the light chain are located in hypervariable loops and partially exposed to the solvent in the crystal structure of the free FvD1.3 fragment (38.8 and 39.8% exposure, respectively) (Bhat *et al.*, 1994). The four other Trp residues are conserved in all the molecules of immunoglobulins and buried in the structure (0.0, 0.2, 2.2 and 10.8% exposure). The high value of  $\lambda_n$  was thus consistent with the partial exposures of H-Trp52 and L-Trp92.

The rigorous Equation 49 and the empirical Equation 51 were fitted to the experimental values of  $Y(x)$  and  $\lambda_{\max}(x)$ , respectively, to obtain parameters  $\Delta G(\text{H}_2\text{O})$ ,  $m$  and  $x_{1/2}$  from  $Y(x)$  and  $\Delta G'(\text{H}_2\text{O})$ ,  $m'$  and  $x'_{1/2}$  from  $\lambda_{\max}(x)$  (Figure 5, Table II; see Materials and methods). The coefficients of cooperativity  $m$  and  $m'$ , determined from  $Y$  and  $\lambda_{\max}$ , respectively, were identical if the SE values were taken into account. The stabilities  $\Delta G(\text{H}_2\text{O})$  and  $\Delta G'(\text{H}_2\text{O})$  were significantly different for domain E3.1 but not for the scFvD1.3 fragment

**Table II.** Comparison of the thermodynamic parameters obtained with the  $Y$  and  $\lambda_{\max}$  signals for the two proteins under study

Parameter	Signal	Equation No.	E3.1	scFvD1.3
$\Delta G(\text{H}_2\text{O})$ (kcal/mol)	$Y$	12	$4.9 \pm 0.5$	$7.0 \pm 0.6$
$\Delta G'(\text{H}_2\text{O})$ (kcal/mol)	$\lambda_{\max}$	29	$6.6 \pm 0.3$	$7.4 \pm 0.4$
$\Delta G''(\text{H}_2\text{O})$ (kcal/mol)	$\lambda_{\max}$	25	$5.9 \pm 0.3$	$7.3 \pm 0.4$
$m$ (kcal/mol.M)	$Y$	12	$0.99 \pm 0.08$	$1.8 \pm 0.1$
$m'$ (kcal/mol.M)	$\lambda_{\max}$	29	$1.15 \pm 0.07$	$1.7 \pm 0.1$
$x_{1/2}$ (M)	$Y$	12	$4.9 \pm 0.1$	$4.0 \pm 0.2$
$x'_{1/2}$ (M)	$\lambda_{\max}$	29	$5.78 \pm 0.08$	$4.4 \pm 0.1$
$x''_{1/2}$ (M)	$\lambda_{\max}$	25	$5.14 \pm 0.08$	$4.28 \pm 0.08$
$R_p$ ( $\times 10^3$ )	$Y$	12	$996 \pm 1$	$991 \pm 3$
$R_p'$ ( $\times 10^3$ )	$\lambda_{\max}$	29	$994 \pm 1$	$999 \pm 0.3$

Unprimed parameters, obtained with the  $Y$  signal; primed parameters, obtained with the  $\lambda_{\max}$  signal and empirical equations; double-primed parameters, obtained with the  $\lambda_{\max}$  signal after correction. The second column gives the signal used to monitor the unfolding equilibrium and the third column gives the law of signal composition that was used to deduce the stability parameters of column 1 from the experimental data.  $\Delta G''(\text{H}_2\text{O})$  was calculated for each individual experiment according to Equations 39 and 44, with  $k_u = k_{W,\text{urea}} = 0.0485 \pm 0.0011 \text{ M}^{-1}$ ; this calculation is equivalent to Equation 57.  $x_{1/2}$ ,  $x'_{1/2}$  and  $x''_{1/2}$  were calculated for each individual experiment according to Equation 9, Equation 35 and  $\Delta G''(\text{H}_2\text{O})/m'$ , respectively. Mean and SE obtained from four to six independent experiments. SE on  $k_W$  was neglected in the calculation of SE on  $\Delta G''(\text{H}_2\text{O})$  and  $x''_{1/2}$ .  $R_p$ , Pearson's coefficient for the fitting of a two-state model to the experimental data.

if the SE values were taken into account. The values of  $x_{1/2}$  and  $x'_{1/2}$  were significantly different for E3.1 (Table II).

### Determination of the spectral curvatures

The high standard errors on the values of  $\Delta G(\text{H}_2\text{O})$  and  $x_{1/2}$ , deduced from the  $Y$  signal, and the differences between these values and those of  $\Delta G'(\text{H}_2\text{O})$  and  $x'_{1/2}$ , deduced empirically from the  $\lambda_{\max}$  signal, stressed the importance of having a rigorous method for the calculation of the thermodynamic stability from  $\lambda_{\max}$ . We showed in the Theory section that such a method exists when the emission spectra of the protein under study can be approximated by a quadratic function on the interval of wavelength  $[\lambda_n, \lambda_u]$ . It requires the determination of parameters  $b_n(0)$  and  $b_u(0)$ , which are the curvatures of the emission spectra for state N at wavelength  $\lambda_n$  and state U at  $\lambda_u$ , respectively, in a medium without any denaturant.

We fitted the quadratic function of Equation 21 (or equivalently Equation 22) to the spectra of domain E3.1 and fragment scFvD1.3 in  $x$  M urea on the interval  $[\lambda_n - 2 \text{ nm}; \lambda_u + 2 \text{ nm}]$  and found that the fittings were excellent for every value of  $x$  ( $R_p > 0.95$ ; Figure 3b and c; see Materials and methods). From these fittings, we determined the values of  $b_n(x)$  for  $x$  in the region of pre-transition and the values of  $b_u(x)$  for  $x$  in the region of post-transition. The values of the ratio  $b_n(0)/b_u(8)$  were very different for the two proteins under study (Table I) and the difference in curvature between the spectra of states N and U for domain E3.1 are clearly visible in Figure 2. We observed that  $b_n(x)$  varied linearly with the concentration of urea in the pre-transition region but with a proportionality factor  $k_n$  whose value was low in each individual experiment and not significantly different from zero on average (Equation 43; Table I). We observed that  $b_u(x)$  also varied with  $x$ . The relation of dependence was imprecise because of the small number of experimental data points in the post-transition region, but consistent with that of the amino acid tryptophan. To obtain greater precision, we assumed that the curvature  $b_u(x)$  followed the same variation as that of tryptophan, i.e. that  $b_u(x)$  followed Equation 42 with a factor of proportionality  $k_u = k_{W,\text{urea}}$  (Equation 54).

### Quantitative parameters of stability obtained from $\lambda_{\max}$

From Equation 46 and the factors of proportionality given above,  $k_n = 0$  and  $k_u = k_{W,\text{urea}}$ , we evaluated the corrective term for  $m'$ . This term was equal (in kcal/mol.M) to  $-0.026$  for E3.1 and  $-0.020$  for scFvD1.3. It was substantially below the SE value on  $m'$  in every case (Table II). From Equation 44, the values of  $b_n(0)/b_u(8)$  and  $k_u = k_{W,\text{urea}}$ , we calculated the corrected value  $\Delta G''(\text{H}_2\text{O})$  of  $\Delta G'(\text{H}_2\text{O})$ . Finally, we calculated the corrected value  $x''_{1/2}$  of  $x'_{1/2}$  as  $\Delta G''(\text{H}_2\text{O})/m'$  since we found that  $m \approx m'$ . Table II gives the values of  $\Delta G(\text{H}_2\text{O})$ ,  $m$  and  $x_{1/2}$ , calculated from the  $Y$  signal, those of  $\Delta G'(\text{H}_2\text{O})$ ,  $m'$  and  $x'_{1/2}$ , calculated empirically from the  $\lambda_{\max}$  signal and those of  $\Delta G''(\text{H}_2\text{O})$  and  $x''_{1/2}$ , obtained after correcting the values of  $\Delta G'(\text{H}_2\text{O})$  and  $x'_{1/2}$ . The corrections brought the empirical values obtained from  $\lambda_{\max}$  closer to those obtained from  $Y$  in every instance.

If the standard errors were taken into account, the rigorous value  $m$  and the empirical value  $m'$  were equal in our two examples. Similarly, the rigorous value  $x_{1/2}$  and its corrected value  $x''_{1/2}$  were equal in our two examples. The values  $\Delta G(\text{H}_2\text{O})$  and  $\Delta G''(\text{H}_2\text{O})$  were equal for scFvD1.3; they were very close for E3.1, with intervals of error within



0.2 kcal/mol. The remaining differences might be due to the theoretical and experimental approximations that we performed. Alternatively, the experiments that used the intensity  $Y$  might be theoretically more rigorous but experimentally less precise.

## Discussion

The theory and results presented here allowed us to obtain rigorous values of the stability from unfolding profiles, monitored with the wavelength  $\lambda_{\max}$ , for a domain of a viral protein and the scFv fragment of an antibody. We discuss the validity of the theory and its application, in general and for the two studied proteins.

### Precise determination of $\lambda_{\max}$

The use of the wavelength  $\lambda_{\max}$  as a signal to monitor the unfolding of proteins requires a precise method to determine its value. This determination is not trivial because the fluorescence emission spectrum  $Y(\lambda)$  of proteins is complex in nature. Many methods have been proposed. The fitting of a polynomial, written as a Taylor expansion about  $\lambda_{\max}$  (Equation 20), has the following advantages. Such a function is continuous and differentiable and its fitting avoids the smoothing of the experimental data. It enables one to obtain directly the value of  $\lambda_{\max}$  as a fitting parameter and the SE value on  $\lambda_{\max}$  in the fitting, whatever the number of terms in the polynomial. We found an excellent fitting of a cubic function to our experimental data over a wide interval of wavelengths, 320–370 nm, with residuals lower than 1% of  $Y(\lambda_{\max})$  on average. The SE value on  $\lambda_{\max}$  in the fitting was typically 3–6% of the  $\lambda_{\max}$  value. However, the number of terms in the Taylor expansion and the interval of wavelengths that is used for the fitting should be optimized for each particular protein.

### Composition of the $\lambda_{\max}$ signals

We showed that the global  $\lambda_{\max}$  signal for a mixture of unfolding is a linear function of the specific  $\lambda_{\max}$  signals,  $\lambda_n$  and  $\lambda_u$ , for the constitutive states N and U, if their individual spectra can be represented by quadratic functions (Equation 25). The  $\lambda_{\max}$  signal of the mixture is not a linear function of the molar fractions  $f_n$  and  $f_u$  of states N and U as for the  $Y$  signal. However, we showed that it is possible to define apparent molar fractions  $f'_n$  and  $f'_u$  such that the  $\lambda_{\max}$  signal of the reaction mixture is a linear function of both  $f'_n$  and  $f'_u$  and the wavelengths  $\lambda_n$  and  $\lambda_u$  (Equation 30). We also showed that  $\lambda_{\max}$  for the unfolding mixture is between  $\lambda_n$  and  $\lambda_u$  if the spectra of N and U show regular behaviors (Equation 18). Therefore, the above theoretical treatment still applies if the spectra of N and U can be approximated by quadratic functions only over the interval  $[\lambda_n, \lambda_u]$  and not over the whole scale of wavelengths (Figure 3). First, we determined precise values of  $\lambda_n$  and  $\lambda_u$  with cubic functions (see previous paragraph). Then, we fitted the quadratic function that is constituted by the first three terms of the Taylor expansion of  $Y$  to the spectrum of N (or U) over the  $[\lambda_n - 2 \text{ nm}, \lambda_u + 2 \text{ nm}]$  interval and found that the fittings were excellent in our two experimental examples ( $R_p \geq 0.95$ ). Three parameters characterize the portion of spectrum that is approached by a parabola: the value of  $\lambda_{\max}$ , the intensity of fluorescence at  $\lambda_{\max}$  and the curvature of the spectrum at  $\lambda_{\max}$  (Equations 21, 22 and 38).

Is it always possible to approximate the spectra of states N and U by quadratic functions over the interval of wavelengths  $[\lambda_n, \lambda_u]$ ? The protein spectra that we report here and those that are available in the literature indicate that such an approximation is possible in many cases: for proteins that contain only one Trp residue as domain E3.1 or several Trp residues as fragment scFvD1.3; and for proteins that have various folds, including all  $\beta$ -proteins (E3.1 and scFvD1.3, this work; the *E.coli* CspA protein, Vu *et al.*, 2001) and  $\alpha/\beta$ -proteins (barnase, Sancho and Fersht, 1992; the *E.coli* CheY protein, Filimonov *et al.*, 1993; Protein L, Scalley *et al.*, 1997). We found that  $\lambda_n$  and  $\lambda_u$  did not vary as a function of the denaturant concentration for the two proteins under study and mentioned that this behavior is quite general. However, such variations of  $\lambda_n$  and  $\lambda_u$  have been reported in a few cases (Ewert *et al.*, 2002). The above theoretical treatment remains valid in such cases if quadratic functions can be fitted to the spectra over the interval  $[\min(\lambda_{\max}), \max(\lambda_{\max})]$  described by the  $\lambda_{\max}$  signal of the reaction mixture during the unfolding.

### Implications for the empirical use of $\lambda_{\max}$

Our theoretical analysis showed that the use of an empirical law of additivity for the  $\lambda_{\max}$  signal leads to inexact values of the stability parameters  $\Delta G'(\text{H}_2\text{O})$ ,  $m'$  and  $x'_{1/2}$  for a monomeric protein that unfolds according to a two-state equilibrium. We gave the corrective terms that allow one to obtain the rigorous values  $\Delta G(\text{H}_2\text{O})$ ,  $m$  and  $x_{1/2}$  from the empirical values (Equations 33, 34 and 36). For  $\Delta G(\text{H}_2\text{O})$  and  $x_{1/2}$ , the corrective terms involved the curvatures  $b_n(0)$  and  $b_u(0)$  of the spectra for states N and U at their respective  $\lambda_{\max}$  and a concentration of denaturant  $x = 0$  (Equations 33 and 36). For  $m$ , the corrective term involved the laws of variation for the curvatures  $b_n(x)$  and  $b_u(x)$  as a function of  $x$  (Equation 34). We found that the curvature  $b_n(x)$  varied linearly with  $x$  in the pre-transition region for the two proteins studied, with very small coefficients  $k_n$  of linear variation (Equation 43; Table I) and that  $b_u(x)$  varied linearly with  $x$  in the post-transition region, with coefficients  $k_u$  (Equation 42). We also found that the curvature  $b_W(x)$ , at  $\lambda_{\max}(x)$ , of the spectrum for the amino acid tryptophan varied linearly with  $x$  over the whole range of denaturant concentration, with a well-defined coefficient  $k_W$  of linear variation (Figure 4; Equation 54), and we proposed to use this linear law of variation for the U state of any protein.

The following relations then result from Equation 46 and the value  $k_u = k_W$ . If the denaturant is urea,  $k_{W,\text{urea}} = 0.0485 \pm 0.0011 \text{ M}^{-1}$  at 20°C and

$$-0.1 \text{ M}^{-1} < k_n < 0.2 \text{ M}^{-1} \Rightarrow |m - m'| \leq 0.15RT \quad (55)$$

If the denaturant is GdmCl,  $k_{W,\text{GdmCl}} = -0.001 \pm 0.002 \text{ M}^{-1}$  at 20°C and

$$-0.15 \text{ M}^{-1} < k_n < 0.15 \text{ M}^{-1} \Rightarrow |m - m'| \leq 0.15RT \quad (56)$$

As  $0.15RT \leq 0.1 \text{ kcal/mol.M}$  for  $T \leq 338 \text{ K}$  (65°C), Equations 55 and 56 show that the difference between the empirical value  $m'$  and rigorous value  $m$  of the cooperativity parameter is lower than the experimental error on  $m'$  if the variation of the curvature  $b_n(x)$  in the pre-transition region remains within wide limits. We found that such was the case for the two proteins studied here (Table I).

The following relations result from Equation 44 and the value  $k_u = k_W$ . If the denaturant is urea and the protein is

unfolded in 8 M urea:

$$\Delta G(\text{H}_2\text{O}) = \Delta G'(\text{H}_2\text{O}) - RT\{\ln[b_n(0)/b_u(8)] + 0.328\} \text{ kcal/mol} \quad (57)$$

Equation 57 shows that  $\Delta G'(\text{H}_2\text{O})$  is generally different from  $\Delta G(\text{H}_2\text{O})$ , but in excess of only 0.19 kcal/mol relative to  $\Delta G(\text{H}_2\text{O})$  at 20°C if the spectra of state N in 0 M urea and state U in 8 M urea have the same curvature at their respective  $\lambda_{\text{max}}$ . Likewise, if the denaturant is GdmCl and the protein is unfolded in 6 M GdmCl:

$$\Delta G(\text{H}_2\text{O}) = \Delta G'(\text{H}_2\text{O}) - RT\ln[b_n(0)/b_u(6)] \text{ kcal/mol} \quad (58)$$

Equation 58 shows that  $\Delta G'(\text{H}_2\text{O})$  and  $\Delta G(\text{H}_2\text{O})$  are generally different and that they are equal at 20°C if the spectra of state N in 0 M GdmCl and state U in 6 M GdmCl have the same curvature at their respective  $\lambda_{\text{max}}$ .

Some authors are aware of the non-linear behavior of  $\lambda_{\text{max}}$  and choose not to calculate the value of  $\Delta G'(\text{H}_2\text{O})$  from the unfolding profile monitored with this signal. They restrict themselves to the empirical concentration  $x'_{1/2}$  of half unfolding. The theory that we present here shows that both  $\Delta G'(\text{H}_2\text{O})$  and  $x'_{1/2}$  require corrections. Moreover, if  $m \approx m'$  (Equations 55 and 56), then Equations 9 and 35 imply that the relative differences between the exact and empirical values are identical for  $\Delta G(\text{H}_2\text{O})$  and  $x_{1/2}$ :

$$[\Delta G(\text{H}_2\text{O}) - \Delta G'(\text{H}_2\text{O})]/\Delta G(\text{H}_2\text{O}) = (x_{1/2} - x'_{1/2})/x_{1/2} \quad (59)$$

### Validity of a two-state model of unfolding

The profiles of unfolding with urea were reversible, cooperative and showed only one visible transition for the two proteins under study, whether they were monitored with the intensity of fluorescence  $Y$  or the wavelength  $\lambda_{\text{max}}$ . These profiles were approximated satisfactorily by a two-state model of unfolding (Table II). They did not allow one to define the characteristic parameters of an intermediate state in a three-state model. The recombinant domain E3.1 of the envelope protein E from serotype 1 of the dengue virus comprises one disulfide bond and 113 residues, including the C-terminal hexahistidine. Its sequence is similar to those of the E3 domains from other flaviviruses (Bhardwaj *et al.*, 2001). Its fold is globular, compact and similar to that of the constant domain of immunoglobulins (Modis *et al.*, 2003). The profiles of unfolding for the E3 domains from serotype 2 of the dengue virus or other flaviviruses, monitored by circular dichroism (CD), fluorescence or gel filtration, show a single transition in every case (Yu *et al.*, 2004). These recombinant domains E3 are in a monomeric state, even at a high concentration of protein (Wu *et al.*, 2003; Volk *et al.*, 2004; Yu *et al.*, 2004). The value of the cooperativity coefficient  $m = 1.1 \pm 0.1$  kcal/mol.M that we determined for domain E3.1 was close to the value, 1.2 kcal/mol.M, that could be predicted from its numbers of residues and disulfide bonds (Myers *et al.*, 1995). This set of data is consistent with a two-state equilibrium of unfolding.

The unfolding of the scFv fragments, which are monomeric proteins, does not always occur according to a two-state equilibrium and its mechanism depends on the relative stabilities of the constitutive  $V_H$  and  $V_L$  domains and of their interface. Given the positions of the conserved Trp residues in the

structures of the scFv fragments, the dissociation of the two variable domains before their unfolding or the unfolding of one domain before the other one generally gives two clear transitions (Worn and Pluckthun, 1999). The existence of a unique transition for scFvD1.3 indicated that an unfolding intermediate, if it existed, would be in a very low concentration at equilibrium. Yasui *et al.* (1994) have reported experiments of heat denaturation that were performed on the FvD1.3 variable fragment, which is a heterodimeric protein, and its two isolated  $V_H$  and  $V_L$  domains. The denaturation was monitored by CD in the far-UV region. These experiments showed that FvD1.3 denatures at a temperature at which the isolated  $V_H$  and  $V_L$  domains are already fully denatured. Therefore, the dissociation of the FvD1.3 fragment into its two domains,  $V_H$  and  $V_L$ , is coupled with the denaturation of each domain and the denaturations of  $V_H$  and  $V_L$  are delayed when they are associated together. Also, we found that the introduction of stabilizing mutations into either  $V_H$  or  $V_L$  led to an overall stabilization of scFvD1.3 in every case (E.Monsellier *et al.*, in preparation). This observation is not consistent with a mechanism in which one of the two domains would unfold before the other. Size-exclusion chromatographic experiments have shown that scFvD1.3 is mainly in a monomeric state, with a very small proportion of dimeric molecules, most likely in the form of diabodies (Renard *et al.*, 2002). Together, these data indicate a two-state equilibrium of unfolding.

The value of the cooperativity parameter  $m$  that we determined for the scFvD1.3 fragment, 1.8 kcal/mol.M, was 35% lower than the value that could be predicted from its 245 residues and two disulfide bonds, 2.8 kcal/mol.M (Myers *et al.*, 1995). A low value of  $m$  is often considered as the sign of an unfolding equilibrium that comprises several states, in particular for the scFv fragments (Worn and Pluckthun, 1999). The experimental data that we recalled in the preceding paragraph show that scFvD1.3 unfolds without intermediate and that the above implication is not general. Otherwise stated, the predicted value of  $m$  might be too high for some scFvs and its experimental value might be the correct one. Similar observations have been reported for the *E.coli* CheY protein (Filimonov *et al.*, 1993). The predictive calculation of  $m$  relies on a correlation between the number of residues in a protein and the variation of its solvent accessible surface area (ASA) during the unfolding from state N to state U. Moreover, the solvent ASA of state U is calculated from an extended model of the polypeptide chain (Myers *et al.*, 1995). In the case of scFv fragments, the presence of the peptide linker (Gly<sub>4</sub>-Ser)<sub>3</sub> between  $V_H$  and  $V_L$ , the six hypervariable loops and the hexahistidine tag could increase the solvent ASA of state N. Both experimental and theoretical approaches have demonstrated the existence of residual interactions and structures in the unfolded state of proteins (Clarke *et al.*, 2000; Fersht and Daggett, 2002). The presence of two disulfide bonds and two folding domains in the scFv fragments could favor the formation of these residual interactions or structures in the U state and decrease its solvent ASA. These two causes, acting on states N and U, could decrease the real variation of solvent ASA and value of parameter  $m$ , relative to the predicted ones.

### Implications for the proteins under study

The recombinant domain E3.1 of the dengue virus could have applications in diagnosis, as a component of recombinant vaccines, as an inhibitor of the interactions between the

virus and its cellular receptors and as a tool in fundamental research. The knowledge of the determinants for the stability of E3.1 would enable one to manipulate this stability without interfering with the immunological and functional properties of this domain. Moreover, the conformational stability of the E3 domain might be correlated with the pathogenicity and with the specificities of vector and host for the flaviviruses (Yu *et al.*, 2004). Our results showed that the recombinant domain E3.1 was in a folded state and that it was stable, with  $\Delta G(\text{H}_2\text{O}) = 5.9 \pm 0.3$  kcal/mol and  $x_{1/2} = 5.14 \pm 0.08$  M urea at 20°C. These results were consistent with the structural properties of domains E3 from serotype 2 of the dengue virus or other flaviviruses, as determined by other biophysical and NMR methods (Wu *et al.*, 2003; Volk *et al.*, 2004; Yu *et al.*, 2004).

Antibody mAbD1.3, directed against hen egg-white lysozyme, has been widely used as an experimental system because the structure of the complex is known at high resolution (Bhat *et al.*, 1994). In particular, it has been used to analyze the interactions between antibodies and antigens and the role of the somatic maturation of antibodies from the thermodynamic, kinetic and structural viewpoints (England *et al.*, 1997, 1999; Dall'Acqua *et al.*, 1998). It has been used to validate experimental strategies aiming at transforming antibodies into reagentless fluorescent biosensors (Renard *et al.*, 2002, 2003; Renard and Bedouelle, 2004). Its hypervariable loops were grafted on to other polypeptide scaffolds for its humanization and stabilization (Foote and Winter, 1992; Donini *et al.*, 2003). However, the stability of scFvD1.3 and its consequences on the above properties have never been studied. Our results showed that scFvD1.3 had an average stability for a scFv fragment, with  $\Delta G(\text{H}_2\text{O}) = 7.3 \pm 0.4$  kcal/mol at 20°C. They provide the basis for a thorough mutational analysis of the relations between structure and stability for this antibody fragment (E.Monsellier *et al.*, in preparation).

## Conclusions

Most studies on the stability of proteins that unfold according to a two-state equilibrium use the following equation (Gittelman and Matthews, 1990):

$$F_{\text{app}} = (S_{\text{obs}} - S_{\text{nat}})/(S_{\text{unf}} - S_{\text{nat}}) \quad (60)$$

where  $F_{\text{app}}$  is the apparent molar fraction of state U and  $S$  is the measured signal. Equation 60 can be rearranged as

$$S_{\text{obs}} = (1 - F_{\text{app}})S_{\text{nat}} + F_{\text{app}}S_{\text{unf}} \quad (61)$$

The signal  $S$  enables one to deduce an apparent equilibrium constant  $K_{\text{app}}$  between states N and U:

$$K_{\text{app}} = F_{\text{app}}/(1 - F_{\text{app}}) \quad (62)$$

Equations 61 and 62 are identical with Equations 26 and 29 in the Theory section. When the observed signal  $S$  is the intensity of fluorescence  $Y$ , the law of the signal is linear with respect to the concentrations of the reactants N and U and therefore  $F_{\text{app}} = f_{\text{u}}$  and  $K_{\text{app}} = K$ , where  $f_{\text{u}}$  and  $K$  are the real molar fraction and equilibrium constant, respectively. When the signal is  $\lambda_{\max}$ , the law of the signal is not linear with respect to concentration and therefore  $F_{\text{app}} \neq f_{\text{u}}$ ,  $K_{\text{app}} \neq K$  and  $\Delta G_{\text{app}}(\text{H}_2\text{O})$  is not equal to the stability of the protein (see Equations 30, 31 and 33).

Here, we established a rigorous law of the signal for  $\lambda_{\max}$ . This law is valid if the spectra of states N and U can be approximated by quadratic functions on the interval of wavelengths  $[\lambda_{\text{n}}, \lambda_{\text{u}}]$ , included between their respective values of  $\lambda_{\max}$ . This condition should be checked for each individual protein and will likely be verified for most of them. We showed that the characteristic parameters of the unfolding equilibrium can then be deduced from the values of  $\lambda_{\max}$  by the same equations (Equations 26–29) as those in use for  $Y$ , provided that the following corrections are performed. (i) The exact stability  $\Delta G(\text{H}_2\text{O})$  is deduced from the empirical stability  $\Delta G'(\text{H}_2\text{O})$  by Equation 33 (see also Equations 44, 57 and 58). (ii) The coefficient of cooperativity  $m$  is identical with its empirical value  $m'$  within experimental error. (iii) The concentration  $x_{1/2}$  of denaturant for the half-advancement of the unfolding reaction is calculated as the ratio  $\Delta G(\text{H}_2\text{O})/m$ . The corrective factor of Equation 33 is not zero in general. It involves the ratio of the curvatures  $b_{\text{n}}(0)$  and  $b_{\text{u}}(0)$  for the spectra of the native and unfolded proteins, at their respective  $\lambda_{\max}$  (i.e.  $\lambda_{\text{n}}$  and  $\lambda_{\text{u}}$ ) and a zero concentration of denaturant. The curvatures  $b_{\text{n}}(0)$  at zero concentration of denaturant and  $b_{\text{u}}(x_{\max})$  at maximal concentration of denaturant can be determined precisely by fitting the quadratic functions of Equations 21 and 22 to the emission spectra on the reduced interval  $[\lambda_{\text{n}} - 2 \text{ nm}; \lambda_{\text{u}} + 2 \text{ nm}]$ . The curvature  $b_{\text{u}}(0)$  can be obtained from the curvature  $b_{\text{u}}(x_{\max})$  by using the case of the amino acid tryptophan, i.e. with Equation 54 (see also Equations 57 and 58). The corrective term is negligible and hence  $\Delta G(\text{H}_2\text{O})$  and  $\Delta G'(\text{H}_2\text{O})$  are equal within the experimental error when the curvatures  $b_{\text{u}}(x_{\max})$  and  $b_{\text{n}}(0)$  are identical. We have validated our theoretical analysis by determining the stabilities of two proteins that have fundamental or applied interest. The results were excellent.

Often, it is not the intrinsic value of  $\Delta G(\text{H}_2\text{O})$  which is important, but its variation  $\Delta\Delta G(\text{H}_2\text{O}) = \Delta G(\text{H}_2\text{O}, \text{wt}) - \Delta G(\text{H}_2\text{O}, \text{mut})$  upon mutation. Our theoretical analysis shows that if the mutation does not modify the curvatures of the spectra for states N and U at their respective  $\lambda_{\max}$  or at least their ratio, then the real value  $\Delta\Delta G(\text{H}_2\text{O})$  should be equal to the empirical (or apparent) value  $\Delta\Delta G'(\text{H}_2\text{O})$ . We hope, by this study, to make the use of the  $\lambda_{\max}$  signal for the determination of protein stability more rigorous. Our approach could be useful for other spectral signals that are not linear, e.g. the intensity-averaged emission wavelength (Royer *et al.*, 1993). We are working on its extension to other mechanisms of unfolding and types of proteins (Park and Bedouelle, 1998).

## Acknowledgements

We thank Shamila Naïr for critical reading of the manuscript. This work was supported by grants from CNRS (Program on Proteomics and Protein Engineering), the French Ministry of Defense (DGA, contracts 01 34 062 and 04 34 025) and the European Commission (INCO-DEV, contract DENFRAME 517711).

## References

- Alber, T. (1989) *Annu. Rev. Biochem.*, **58**, 765–798.
- Bava, K.A., Gromiha, M.M., Uedaira, H., Kitajima, K. and Sarai, A. (2004) *Nucleic Acids Res.*, **32**, D120–D121.
- Bhardwaj, S., Holbrook, M., Shope, R.E., Barrett, A.D. and Watowich, S.J. (2001) *J. Virol.*, **75**, 4002–4007.
- Bhat, T.N. *et al.* (1994) *Proc. Natl Acad. Sci. USA*, **91**, 1089–1093.
- Carter, P., Bedouelle, H. and Winter, G. (1985) *Nucleic Acids Res.*, **13**, 4431–4443.

- Clarke,J., Hounslow,A.M., Bond,C.J., Fersht,A.R. and Daggett,V. (2000) *Protein Sci.*, **9**, 2394–2404.
- Dall'Acqua,W. *et al.* (1998) *Biochemistry*, **37**, 7981–7991.
- Despres,P., Frenkiel,M.P. and Deubel,V. (1993) *Virology*, **196**, 209–219.
- Donini,M., Morea,V., Desiderio,A., Pashkoulou,D., Villani,M.E., Tramontano,A. and Benvenuto,E. (2003) *J. Mol. Biol.*, **330**, 323–332.
- Dumoulin,M., Conrath,K., Van Meirhaeghe,A., Meersmen,F., Heremans,K., Frenken,L.G.J., Muyldermans,S., Wyns,L. and Matagne,A. (2002) *Protein Sci.*, **11**, 500–515.
- Eftink,M.R. (1994) *Biophys. J.*, **66**, 482–501.
- England,P., Bregegere,F. and Bedouelle,H. (1997) *Biochemistry*, **36**, 164–172.
- England,P., Nageotte,R., Renard,M., Page,A.L. and Bedouelle,H. (1999) *J. Immunol.*, **162**, 2129–2136.
- Ewert,S., Cambillau,C., Conrath,K. and Pluckthun,A. (2002) *Biochemistry*, **41**, 3628–3636.
- Ewert,S., Honegger,A. and Pluckthun,A. (2003) *Biochemistry*, **42**, 1517–1528.
- Fersht,A.R. and Daggett,V. (2002) *Cell*, **108**, 573–582.
- Filimonov,V.V., Prieto,J., Martinez,J.C., Bruix,M., Mateo,P.L. and Serrano,L. (1993) *Biochemistry*, **32**, 12906–12921.
- Foote,J. and Winter,G. (1992) *J. Mol. Biol.*, **224**, 487–499.
- Gittelman,M.S. and Matthews,C.R. (1990) *Biochemistry*, **29**, 7011–7020.
- Guerois,R., Nielsen,J.E. and Serrano,L. (2002) *J. Mol. Biol.*, **320**, 369–387.
- Jager,M. and Pluckthun,A. (1999) *FEBS Lett.*, **462**, 307–312.
- Jung,S., Honegger,A. and Pluckthun,A. (1999) *J. Mol. Biol.*, **294**, 163–180.
- Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
- Martineau,P. and Betton,J.M. (1999) *J. Mol. Biol.*, **292**, 921–929.
- Modis,Y., Ogata,S., Clements,D. and Harrison,S.C. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 6986–6991.
- Mukhopadhyay,S., Kuhn,R.J. and Rossmann,M.G. (2005) *Nat. Rev. Microbiol.*, **3**, 13–22.
- Myers,J.K., Pace,C.N. and Scholtz,J.M. (1995) *Protein Sci.*, **4**, 2138–2148.
- Pace,C.N. (1986) *Methods Enzymol.*, **131**, 266–280.
- Pace,C.N., Vajdos,F., Fee,L., Grimsley,G. and Gray,T. (1995) *Protein Sci.*, **4**, 2411–2423.
- Pace,C.N., Shirley,B.A., McNutt,M. and Gajiwala,K. (1996) *FASEB J.*, **10**, 75–83.
- Park,Y.C. and Bedouelle,H. (1998) *J. Biol. Chem.*, **273**, 18052–18059.
- Renard,M. and Bedouelle,H. (2004) *Biochemistry*, **43**, 15453–15462.
- Renard,M., Belkadi,L., Hugo,N., England,P., Altschuh,D. and Bedouelle,H. (2002) *J. Mol. Biol.*, **318**, 429–442.
- Renard,M., Belkadi,L. and Bedouelle,H. (2003) *J. Mol. Biol.*, **326**, 167–175.
- Royer,C.A., Mann,C.J. and Matthews,C.R. (1993) *Protein Sci.*, **2**, 1844–1852.
- Sancho,J. and Fersht,A.R. (1992) *J. Mol. Biol.*, **224**, 741–747.
- Santoro,M.M. and Bolen,D.W. (1988) *Biochemistry*, **27**, 8063–8068.
- Scalley,M.L., Yi,Q., Gu,H., McCormack,A., Yates,J.R.,III and Baker,D. (1997) *Biochemistry*, **36**, 3373–3382.
- Schmid,F.X. (1989) In Creighton,T.E. (ed.), *Protein Structure: a Practical Approach*. IRL Press, Oxford, pp. 251–285.
- Sundberg,E.J. and Mariuzza,R.A. (2002) *Adv. Protein Chem.*, **61**, 119–160.
- Tan,P.H., Sandmaier,B.M. and Stayton,P.S. (1998) *Biophys. J.*, **75**, 1473–1482.
- Volk,D.E., Beasley,D.W., Kallick,D.A., Holbrook,M.R., Barrett,A.D. and Gorenstein,D.G. (2004) *J. Biol. Chem.*, **279**, 38755–38761.
- Vriend,G. (1990) *J. Mol. Graph.*, **8**, 52–56.
- Vu,D.M., Reid,K.L., Rodriguez,H.M. and Gregoret,L.M. (2001) *Protein Sci.*, **10**, 2028–2036.
- Weisstein,E.W. (2002) *CRC Concise Encyclopedia of Mathematics*, 2nd edn. CRC Press, Boca Raton, FL.
- Worn,A. and Pluckthun,A. (1999) *Biochemistry*, **38**, 8739–8750.
- Worn,A. and Pluckthun,A. (2001) *J. Mol. Biol.*, **305**, 989–1010.
- Wu,K.P., Wu,C.W., Tsao,Y.P., Kuo,T.W., Lou,Y.C., Lin,C.W., Wu,S.C. and Cheng,J.W. (2003) *J. Biol. Chem.*, **278**, 46007–46013.
- Yasui,H., Ito,W. and Kurosawa,Y. (1994) *FEBS Lett.*, **353**, 143–146.
- Yu,S., Wu,A., Basu,R., Holbrook,M.R., Barrett,A.D. and Lee,J.C. (2004) *Biochemistry*, **43**, 9168–9176.

Received June 27, 2005; accepted July 1, 2005

Edited by André Menez