Journal of Biomolecular NMR, **11:** 355–359, 1998.
KLUWER/ESCOM

ALASKA: A *Mathematica* package for two-state kinetic analysis of protein folding reactions

Rendall E. Burton, Richard S. Busby & Terrence G. Oas
Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

Received 5 September 1997; Accepted 17 November 1997

Key words: dynamic NMR, kinetics, line shape simulation, protein folding

Abstract

A *Mathematica* package (ALASKA) has been developed to simplify the measurement of protein folding kinetics by analysis of ^{1}H NMR lineshape analysis. This package reads NMR data in ASCII format and can simulate an aromatic ^{1} NMR spectrum with or without lineshape broadening from chemical exchange. We describe the analysis of a urea denaturation series of a fast-folding protein, the G46A/G48A variant of monomeric λ repressor.

The power of dynamic NMR in the study of rapid chemical reactions is well established (Rao, 1989). Rapid protein folding processes have received increasing attention because they represent the early events in the folding of all proteins. Fast chemical exchange has been observed for some protein folding reactions (Wemmer et al., 1981; Oas and Kim, 1988; Hoeltzli and Frieden, 1994), and stopped-flow kinetic measurements suggest that other proteins may fold and unfold rapidly enough to affect the line widths of proton NMR resonances (Jackson, et al., 1993; Kuszewski et al., 1994; Milla and Sauer, 1994; Kragelund et al., 1995; Schindler et al., 1995). We have recently reported the use of line-shape analysis to measure the folding kinetics of monomeric λ repressor, λ_{6-85} (Huang and Oas, 1995a; Burton et al., 1996, 1997). Line-shape analysis probes folding times as fast as as the microsecond timescale, a regime that is beyond most conventional stopped-flow devices.

To facilitate the use of line-shape analysis to measure the folding kinetics of small proteins, we have developed a group of *Mathematica* functions, organized in a notebook, that assist in the process. This package is referred to as ALASKA (Aromatic Lineshape Algorithm for 2-State Kinetic Analysis). We have included functions to simulate aromatic NMR spectra both with and without chemical exchange as well as an adaptation of the Levenberg–Marquardt nonlinear least-squares fitting algorithm (Marquardt, 1963). ALASKA can be used to fit any aromatic proton spectrum, limited only by the need to make sequence-specific assignments in both the native and denatured spectra. All of the functions are purely written in *Mathematica* code, so the package is completely portable to any machine that is capable of running *Mathematica*.

In addition to the ALASKA package, which fits spectra one at a time, a stand-alone C++ program (GlobalFit) has been developed that performs a 'global' fit of an entire urea denaturation series. In practice, the global fit is very difficult because of the large number of parameters that are simultaneously fit; initial parameter values have to be very close to optimal for the fit to succeed. Therefore, this program is used to refine parameters obtained by using the interactive ALASKA package.

In developing the package we assumed that the folding process is two-state, meaning that only completely folded or completely unfolded molecules are present at equilibrium. The two-state approximation has been used extensively in the thermodynamic and kinetic analysis of protein folding, and is supported by experimental evidence for several small proteins, including λ_{6-85} (Huang and Oas, 1995b). To provide kinetic information, an individual proton must resonate at one frequency in the folded protein structure, and at a significantly different frequency in the denatured protein. The protein must not sample any intermediate

environment for a significant fraction of time to allow the analysis described here. The resonance may appear at an intermediate chemical shift, but only as a consequence of ensemble averaging (Sandstrom, 1982). Under these circumstances, line shape and frequency reflect the relative populations of the two states and their rate of interconversion (Sandstrom, 1982).

The theory of two-site chemical exchange has been established for nearly four decades. The first general treatment is given by McConnell (1958)? which has been summarized by Sandstrom (1982). The derivation described here uses the terms defined by Sandstrom. The exchange process between states A and B is described by the following first-order rate equations:

$$A \xrightarrow{k_A} B$$
$$B \xrightarrow{k_B} A$$
$$\frac{d[A]}{dt} = -k_A[A]$$
$$\frac{d[B]}{dt} = -k_B[B]$$

In the absence of chemical exchange, the magnetization arising from states A and B (G_A and G_B, respectively) varies with time as described by the Bloch equations:

$$\frac{dG_A}{dt} = -\alpha_A G_A - iC_A$$
$$\frac{dG_B}{dt} = -\alpha_B G_B - iC_B$$
$$\alpha_x = \frac{1}{T_{2x}} - 2\pi i(\upsilon_x - \upsilon)$$
$$C_x = i\gamma B_1 M_0$$

Introducing chemical exchange between A and B adds terms describing the change in their contributions to the overall magnetization:

$$\frac{dG_A}{dt} = -\alpha_A G_A - iC_A - k_A G_A + k_B G_B$$
$$\frac{dG_B}{dt} = -\alpha_B G_B - iC_B - k_B G_B + k_A G_A$$

These equations can be solved for G_A and G_B using the steady-state approximation (dG/dt=0). The sum of these solutions gives the total magnetization, G:

$$G = -\frac{iC_0(k_A + k_B + \alpha_A p_A + \alpha_B p_B)}{\alpha_A k_B + \alpha_B k_A + \alpha_A \alpha_B}$$
$$p_A + p_B = 1$$
$$\frac{p_A}{p_B} = \frac{k_B}{k_A}$$
$$C_0 = C_A + C_B$$

The absorption-mode spectrum is described by the imaginary part of G. The final solution, which has been published elsewhere (Sandstrom, 1982; Huang and Oas, 1995a; Burton et al., 1996), is shown here:

$$I(\upsilon, k_f, k_u, C_O) =$$
$$C_O \frac{\{P[1 + \tau\left(\frac{p_N}{T_{2,D}} + \frac{p_D}{T_{2,N}}\right)] + QR\}}{p^2 + R^2}$$
$$P = \tau[\frac{1}{T_{2,N}T_{2,D}} - 4\pi^2\Delta\upsilon^2 + \pi^2(\delta\upsilon)^2]$$
$$+\frac{p_N}{T_{2,N}} + \frac{p_D}{T_{2,D}}$$
$$Q = \tau[2\pi\Delta\upsilon - \pi\delta\upsilon(p_N - p_D)]$$
$$R = 2\pi\Delta\upsilon[1 + \tau\left(\frac{1}{T_{2,N}} + \frac{1}{T_{2,D}}\right)] +$$
$$\pi\delta\upsilon\tau\left(\frac{1}{T_{2,D}} - \frac{1}{T_{2,N}}\right) + \pi\delta\upsilon(p_N - p_D)$$
$$\delta\upsilon = \upsilon_N - \upsilon_D$$
$$\Delta\upsilon = \frac{\upsilon_N + \upsilon_D}{2} - \upsilon$$
$$\tau = \frac{1}{k_u + k_f}$$
$$p_N = \frac{k_f}{k_u + k_f} \quad p_D = 1 - p_N$$

The line-shape fitting procedure requires high-quality NMR spectra. We collect 8192 real points per spectrum with a sweep width of 6500 Hz to maximize the resolution. Shimming is done manually for each sample to minimize line-shape distortions from magnetic field inhomogeneity. A recycle delay of 4.0 s is used between transients to prevent saturation of protons with long T_1 relaxation times. Selective suppression of individual peaks reduces the quality of the fits, as the simulation functions assume equal intensity for all protons. A total of 256 transients are collected and averaged to maximize the signal-to-noise ratio. The protein concentration is approximately 0.5 mM, which gives reasonable signal-to-noise and reduces aggregation.

The data are converted to FELIX format and processed using the FELIX software (MSI, Inc.). The data are dc-offset corrected and zero-filled to 8192 complex points. No apodization function is applied to avoid distorting the line shapes. Each spectrum is manually phased and baseline corrected with a fifth-order polynomial. Correct phasing is essential for success in the later steps of the analysis. The spectra gre then referenced to TMSP. The aromatic region is isolated by left-shifting the data, then truncating. The data are reduced to real points only, then saved in FELIX ASCII

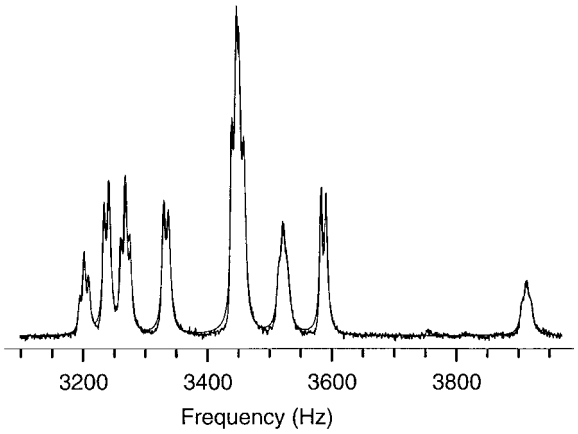

Figure 1. Use of the SpectrumNoEx function to simulate a spectrum of the G46A/G48A variant of λ_{6-85} at 30 °C (black line: experimental data; red line: best-fit simulation).

format. Any NMR processing package with an ASCII output capability is suitable for the line-shape analysis.

The notebook reads data from a list of user-supplied text files. The files can be either plain x, y data or in FELIX ASCII format. Once loaded, each spectrum is assigned as either native, denatured, or exchanging. The native and denatured spectra are fitted using the SpectrumNoEx function, which models the spectrum as a sum of Lorentzian peaks. An example of the results is shown in Figure 1. This process establishes the chemical shifts and intrinsic line widths of the native and denatured states, which are critical for simulating spectra perturbed by chemical exchange. We find that the chemical shifts of protons in both the native and denatured states have a linear dependence on urea concentration, analogous to the sloping baselines seen in equilibrium denaturant titrations followed by circular dichroism or fluorescence (Pace, 1986). An example of this type of chemical shift dependence is shown in Figure 2.

The SpectrumNoEx function simulates an aromatic ^{1}H spectrum by using a simple model. Phenylalanines are modeled as ABC spin systems, with only J_3 couplings. The *ortho* and *meta* protons are treated as groups of equivalent protons, each of which has twice the intensity of the *para* proton. Tyrosine resonances are modeled similarly at the *ortho* and *meta* positions. Histidines are modeled as two singlets, corresponding to the C2 and C4 protons. Tryptophan resonances are treated as non-equivalent, and only J_3, scalar couplings are considered. The relative heights of peaks within a multiples it; are determined by the ratio (Friebolin, 1993):

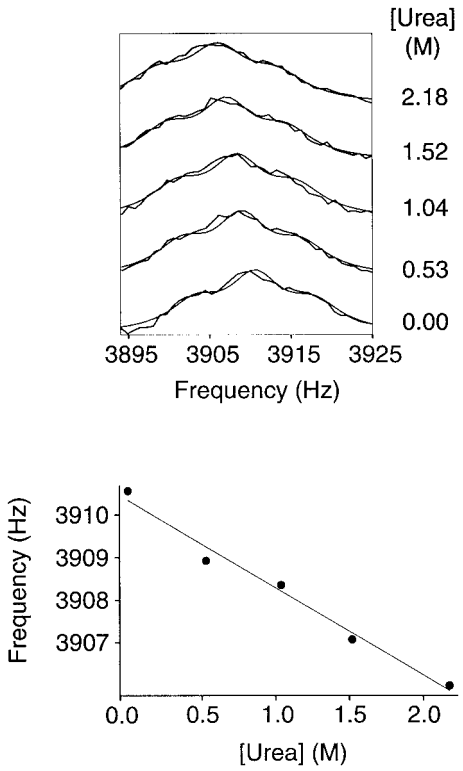

Figure 2. The urea dependence of the chemical shift of the *para* proton from Phe^{76} in the G46A/G48A variant of λ_{6-85}. Top panel: a stack plot of spectra acquired in 0.0–2.18 M urea, along with the best-fit simulated spectra. Bottom panel: a plot of the best-fit chemical shift for this proton *versus* urea concentration. The line is a least-squares fit, with a linear correlation coefficient of 0.96.

$$\frac{I_1}{I_2} = \frac{I_4}{I_3} = \frac{f_2 - f_3}{f_1 - f_4}$$

where resonances 1 and 2 form one doublet and 3 and 4 form another. I_n is the intensity of peak n, and f_n is its frequency. Triplets are modeled as doublets of doublets, with the inner half of each doublet superimposed.

The fitting of the fully native and fully denatured spectra is the most challenging aspect of the analysis, especially in crowded regions. The package offers two primary tools for dealing with this difficulty. First, the fitting can be done 'by hand', adjusting the parameters, re-evaluating the simulated spectrum and comparing it to the observed spectrum until a match is achieved. Second, the fitting functions allow certain parameters to be held fixed during the fit. It is helpful to alternate between fixing the chemical shifts and fixing the line widths, because the algorithm often tries to correct a poorly fit peak by broadening it.

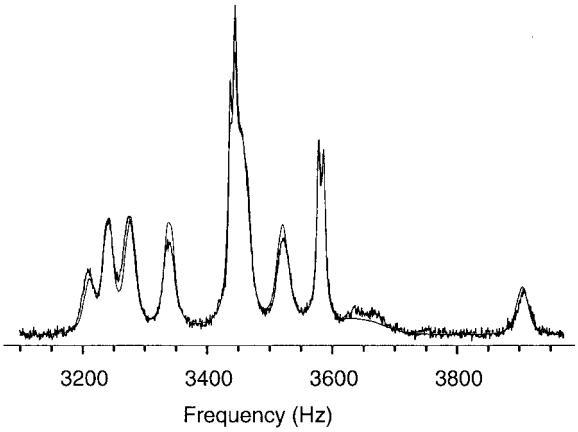

Figure 3. Fit to a spectrum influenced by chemical exchange due to the two-state folding reaction. Black line: data for a spectrum of the G46A/G48A variant of λ_{6-85} collected in 4.14 M urea at 30 °C, where both the native and denatured states are significantly populated and interconvert on a millisecond timescale. Red line: best-fit simulated spectrum in which the only parameters allowed to vary are the folding and unfolding rate constants. The chemical shifts and intrinsic line widths of the native and denatured states are held fixed at the values found for spectra outside the transition region.

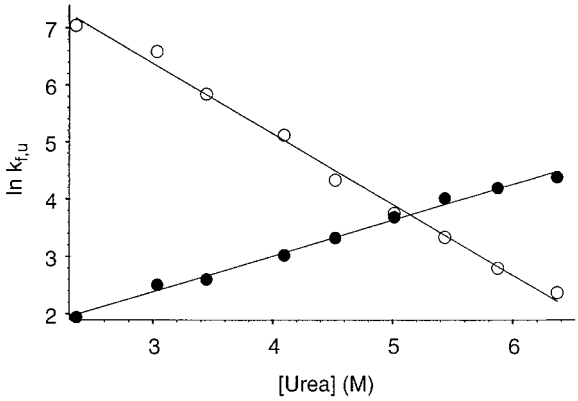

Figure 4. Plot of the natural logarithms of the rate constants (open circles: folding rate; filled circles: unfolding rate) versus denaturant concentration for the G46A/G48A variant of λ_{6-85} at 30 °C.

Once the parameters for the native and denatured spectra are set, the spectra collected at urea concentrations within the transition region of the denaturation curve can be fitted using the chemical exchange equations for two-site exchange (Sandstrom, 1982) found in the SpectrumEx function. None of the chemical shifts or intrinsic line widths of the end states are allowed to change at this point. The only parameters that vary are the folding rate constant (k_f), the unfolding rate constant (k_u), and an intensity parameter proportional to protein concentration (C). An example of a SpectrumEx fit is shown in Figure 3. Once the exchange-broadened spectra are fit, output functions tabulate the rate constants as a function of denaturant concentration into a matrix suitable for plotting or further analysis (Figure 4).

One critical assumption of this analysis is that the observed line widths of peaks in the native and denatured states do no change significantly with denaturant concentration. The viscosity of a solution can change dramatically if chemical denaturants are added, or the temperature is changed. This will affect the intrinsic line widths because the T_2 relaxation time is sensitive to the correlation time of the molecule, which in turn is related to the tumbling time of the molecule. In principle, this effect could contribute to the line broadening observed in the presence of denaturant, since the viscosity of an 8.66 M urea solution is nearly double that of pure water (Weast, 1974). However, the observed line width, $T_{2,obs}$, equals $1/T_{2,intrinsic} + R_{ex}$. For a typical aromatic proton at 37 °C, $1/T_{2,intrinsic} \approx 20\ s^{-1}$. The effect of doubling the viscosity on $T_{2,intrinsic}$ depends on the details of possible relaxation pathways for the proton, but is unlikely to be more than twofold (Cavanagh, et al., 1996), which would increase it to 40 s^{-1}. If both the native and denatured state are more than 5% populated, the R_{ex} dominates the expression for $T_{2,obs}$. For example, if 5% of the molecules are in the denatured state, given an exchange rate $k_{ex} = (k_f + k_u)/2 = 1000\ s^{-1}$ and a typical difference between the native and denatured state resonant frequencies of 500 Hz, the R_{ex} term equals 250 s^{-1}. However, viscosity effects must be taken into account under conditions where the R_{ex} term is comparable to $1/T_{2,intrinsic}$ (e.g. smaller frequency differences between states, slower exchange rates, or very small populations of the native or denatured states).

After independent fitting of each exchange-broadened spectrum in a denaturation series, it is not unusual for some spectra to be poorly fitted. This is usually a result of poor baseline parameters. If the protein is too stable or unstable, it may be impossible to collect fully native or fully denatured spectra over a wide range of denaturant concentrations. Additional spectra within a tight range (1 M) of denaturant can help, but with diminishing returns. Another good strategy is to fit all of the spectra, simultaneously varying the baseline parameters and the rate constants. This is time-consuming, and has therefore been implemented in a stand-alone C++ program (GlobalFit). Because of the large number of parameters (usually > 80) and

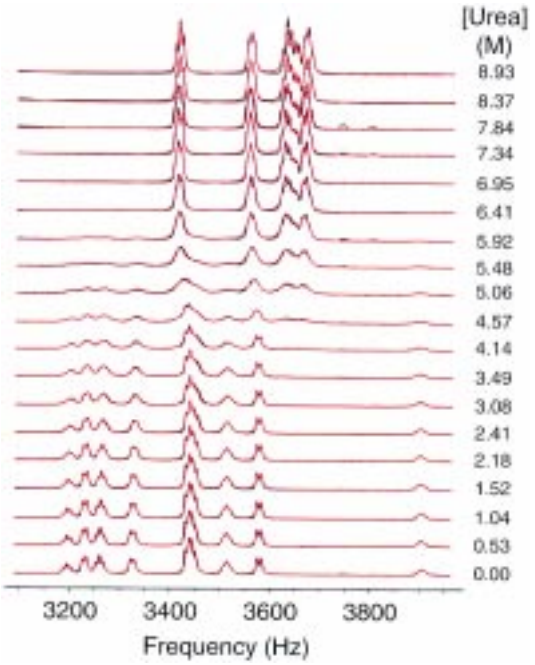

Figure 5. Results of a GlobalFit analysis of the entire urea denaturation series for the G46A/G48A variant of λ_{6-85} at 30 °C (black lines: observed data; red lines: best-fit simulated spectra). The trace peaks observed in 6–8 M urea have been seen for all mutants of λ_{6-85} studied so far, and appear to reflect the presence of small amounts of aggregated or chemically modified protein.

the presence of multiple minimain the χ^2 function, it is important that the initial values of each parameter are almost correct. Starting values obtained from a round of fitting with the ALASKA *Mathematica* package are usually sufficient. GlobalFit reads a parameter-list file with the initial values for each parameter and the filenames of the FELIX ASCII files with the NMR data, and then performs a Levenberg–Marquardt minimization to refine the parameters. There is a command-line option to dump the results of the fit in *Mathematica* text format that can be used to generate graphs Figure 5).

The range of kinetic processes that influence NMR peak shapes is determined by two factors: the relative populations of the two interconverting states, and their rate of exchange. We find that line-shape analysis only works when both the native and denatured states are at least 5% populated. When one state is $< 5\%$ populated, other influences on line widths, such as magnetic field inhomogeneity, begin to dominate. The slowest rate measured for λ_{6-85} by this method is $\sim s^{-1}$, the fastest, $\sim 2000\ s^{-1}$. However, this rate is not the upper limit of the technique. The largest rate constants are measured at low urea concentrations and so far have been limited only by the denatured state population, not the interconversion rate. Rates as high as $10^4\ s^{-1}$ would be measurable by line-shape analysis of the most shifted resonances in the λ_{6-85} spectrum (Phe^{51} *para* proton, $\delta\upsilon = 573$ Hz at 600 MHz), if a denatured state population of 5% could be attained. Such measurements will require λ_{6-85} variants that are destabilized through increased unfolding rates. A search for such variants is currently underway.

A sample ALASKA notebook containing all of the functions necessary for line-shape analysis, along with NMR data in ASCII format and fitted parameters, is available from the authors. On an SGI Indigo^2 with 200 MHz MIPS, 4400 CPU and 64 MB RAM, the notebook can be evaluated in about one hour. The source code for GlobalFit is also available upon request.

References

Burton, R.E., Huang, G.S., et al. (1997) *Nat. Struct. Biol.* **4**, 305–310.

Burton, R.E., Huang, G.S. et al. (1996) *J. Mol. Biol.*, **263**, 311–322.

Cavanagh, J., Fairbrother, W.J. et al. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, New York, NY.

Friebolin, H. (1993) *Basic One- and Two-dimensional NMR Spectroscopy*, VCH Publishers, New York, NY.

Hoeltzli, S.D. and Frieden, C. (1994) *Biochemistry*, **33**, 5502–5509.

Huang, G.S. and Oas, T.G. (1995a) *Proc. Natl. Acad. Sci. USA*, **92**, 6878–6882.

Huang, G.S. and Oas, T.G. (1995b) *Biochemistry*, **34**, 3884–3892.

Jackson, S.E., Elmastry, N. et al. (1993) *Biochemistry*, **32**, 11270–11278.

Kragelund, B.B., Robinson, C.V. et al. (1995) *Biochemistry*, **34**, 7217–7224.

Kuszewski, J.j Clore, G.M., et al. (1994) *Protein Sci.*, **3**, 1945–1952.

Marquardt, D.W. (1963) *J. Soc. Industr. Appl. Math.*, **11**, 431–441.

McConnell, H.M. (1958) *J. Chem. Phys.*, **28**, 430–431.

Milla, M.E. and Sauer, R.T. (1994) *Biochemistry*, **33**, 1125–1133.

Oas, T.G. and Kim, P.S. (1988) *Nature*, **336**, 42–48.

Pace, C.N. (1986) *Methods Enzymol.*, **131**, 266–280.

Rao, B.D.N. (1989) *Methods Enzymol.*, **176**, 279–311.

Sandstrom, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press, London.

Schindler, T., Herrler, M., et al. (1995) *Nat. Struct. Biol.*, **2**, 663–673.

Weast, R.C. (Ed.) (1974) *Handbook of Chemistry and Physics*, CRC Press, Cleveland, OH.

Wemmer, D., Ribeiro, A.A., et al. (1981) *Biochemistry*, **20**, 829–833.