

# Hairpin folding rates reflect mutations within and remote from the turn region

Katherine A. Olsen, R. Matthew Fesinmeyer, James M. Stewart, and Niels H. Andersen\*

Department of Chemistry, University of Washington, Seattle, WA 98195

Edited by Alan R. Fersht, University of Cambridge, Cambridge, United Kingdom, and approved September 8, 2005 (received for review May 26, 2005)

Hairpins play a central role in numerous protein folding and misfolding scenarios. Prior studies of hairpin folding, many conducted with analogs of a sequence from the B1 domain of protein G, suggest that faster folding can be achieved only by optimizing the turn propensity of the reversing loop. Based on studies using dynamic NMR, the native GB1 sequence is a slow folding hairpin ( $k_F^{278} = 1.5 \times 10^4/s$ ). GB1 hairpin analogs spanning a wide range of thermodynamic stabilities ( $\Delta G_{U0}^{298} = -3.09$  to  $+3.25$  kJ/mol) were examined. Fold-stabilizing changes in the reversing loop can act either by accelerating folding or retarding unfolding; we present examples of both types. The introduction of an attractive side-chain/side-chain Coulombic interaction at the chain termini further stabilizes this hairpin. The 1.9-fold increase in folding rate constant observed for this change at the chain termini implies that this Coulombic interaction contributes before or at the transition state. This observation is difficult to rationalize by "zipper" folding pathways that require native turn formation as the sole nucleating event; it also suggests that Coulombic interactions should be considered in the design of systems intended to probe the protein folding speed limit.

$\beta$ -hairpin | exchange broadening | folding dynamics | loop search

Protein engineering experiments indicate that  $\beta$ -hairpins appear as transition-state features in numerous folding pathways. Hairpin redesign has resulted in changes in protein-folding mechanisms (1), and hairpin stabilization can accelerate (1–3) or retard (2, 4) protein folding. The protein-folding problem continues to be of high interest for at least three reasons: predicting structures from genome-derived sequences, improving *a priori* protein-fold design, and enhancing our understanding of the mechanisms of protein misfolding diseases (6, 7). Folding rates have been of particular interest, with more examples of redesigned proteins that fold near the calculated protein-folding speed limit (8) appearing regularly. For  $\beta$ -sheet proteins and  $\beta$  oligomers (as found in amyloid fibrils formed from misfolded protein states), hairpin dynamics play a key role.

Although there is an increasing body of data on hairpin folding dynamics (9–15), with one exception these have not included a set of probing mutations to address specific questions. That exception (14) suggests that loops with a greater turn preference accelerate folding to a much greater extent than the optimization of hydrophobic interactions in the folded state. Other data (12) suggest that the length of the loop connecting the hydrophobic residues that form hairpin-stabilizing cross-strand interactions is reflected in the folding rate. Throughout, hairpin folding has been modeled as a two-state equilibrium. Whether hairpin/coil transitions are 1- $\mu$ s versus 50- $\mu$ s events has significant consequences. Peptide helix nucleation is a sub- $\mu$ s event (10, 16, 17), which allows helix formation to be a preequilibrium event relative to the hydrophobic-collapse stage of protein folding. Can rapid formation of stable hairpins in the unfolded states of nascent proteins serve in a similar fashion? If so, hairpin optimization could be an evolutionary criterion for improving the efficiency of protein folding. Alternatively, is hairpin formation restricted to a transition state role in  $\beta$ -sheet folding?

Essentially all folding rate constants in the  $10^4$  to  $10^7/s$  range have been derived from fluorescence- or IR-monitored temperature jump experiments (12, 13, 18, 19). Two groups have recognized the potential of NMR spectroscopy to probe 10- to 50- $\mu$ s exchange phenomena (20–22). Raleigh and colleagues (22) extended lineshape analysis to protein-folding transitions with fast exchange on the chemical shift time scale; in a favorable case, rate constants as fast as  $2 \times 10^5/s$  were accessible. We now report that the large structuring shifts observed for hairpin peptides, in combination with an internal referencing method for extracting differential exchange-broadening measures ( $\Delta\Delta^{ex}$ ), allows NMR line broadening methods to quantitate the microsecond time constants needed for studying fast hairpin folding.

The second hairpin of the B1 domain of protein G (GB1p), residues 41–56 (23, 24), was the first hairpin to be examined extensively, and a folding time constant,  $(k_F)^{-1}$ , on the order of 6  $\mu$ s was reported in 1997 (9). We recently prepared a series of more stable GB1 hairpin analogs with melting temperatures in the 47–85°C range (25). Extensive spectral comparisons with the more stable analogs that retain the identical hydrophobic cluster revealed that GB1p is <30% folded at 25°C. These studies also established that the folding equilibria of these GB1 analogs, like that of the tryptophan zipper peptide (trpzip) analogs (26) studied previously, can be modeled as two-state processes.

## Materials and Methods

**Materials.** Analogs GB1m2 and m3 were available from a previous study (25). GB1p and additional analogs were synthesized on an Applied Biosystem 433A synthesizer by using standard fluorenylmethoxycarbonyl solid-phase peptide synthesis methods and displayed the expected molecular weights by ion-trap MS (see *Supporting Text*, which is published as supporting information on the PNAS web site). Sequence and purity were verified by  $^1H$  NMR.

**NMR Spectroscopy.** Samples for 2D spectra consisted of  $\approx 1.5$  mM peptide in 20 mM (pH 6–7) phosphate buffer (in 10% or 99.96%  $D_2O$ ) with 2,2-dimethylsilapentane-5-sulfonic acid as the internal chemical shift reference. All NMR experiments were collected on either Bruker (Billerica, MA) DRX-500 or DMX-750 spectrometers. Peptide  $H_N$  and  $H_\alpha$  resonances could be assigned through a combination of 2D total correlation spectroscopy and NOESY experiments; the resulting backbone-resonance chemical shift deviations (CSDs) were concordant with the previous correlations (25) of these CSDs with fold population. 1D spectra used for extracting exchange-broadening data, with peptide concentrations 0.8–1.2 mM, were the average of 1,024 scans acquired at a resolution of 32,768 and 16,384 points for the 750- and 500-MHz spectrometers, respectively. As a check for self-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: GB1p, the second hairpin of the B1 domain of protein G; trpzip, tryptophan zipper peptide.

\*To whom correspondence should be addressed. E-mail: andersen@chem.washington.edu.

© 2005 by The National Academy of Sciences of the USA







judged to be better than  $\pm 10\%$  or 4 kJ/mol (whichever is larger). Even though the extraction of specific rate constants has some intrinsic limitations when  $k_{ex}$  exceeds  $6 \times 10^5/s$ , and obtaining accurate activation enthalpies by this method is problematic, the values in Table 2 do provide a basis for discussing structural effects on hairpin dynamics. The data for trpzip4 are included for comparison.

## Discussion

To date there have been very limited data on the relationship between hairpin fold stability and folding dynamics, in large part reflecting the lack of rate determinations for series of hairpin analogs with specific, but limited, mutations. A recent study focused on five trpzip (26) analogs that differ only in the reversing loop (turn) region and included the prior literature data for GB1p in the comparison (14). Du *et al.* (14) concluded that loop mutations that provide hairpin stabilization do so by accelerating folding, rather than retarding unfolding. The GB1p/trpzip4 dynamics comparison led them to conclude that improvements in the cross-strand hydrophobic cluster do not increase the folding rate constant; rather, they effect decreased unfolding. The latter conclusion remains intact with the present data (Table 2): the enhanced stability associated with incorporating additional indole rings in the hydrophobic cluster results from a 37-fold unfolding rate retardation.

The conclusion regarding loop mutations, however, needs to be modified. In the present series, the D7P mutation is a stabilizing loop mutation and results in a 5.5-fold decrease in the unfolding rate constant. A Pro insertion was also examined by Du *et al.* (14); replacing an EGNK loop with a hairpin stabilizing, and turn favoring, E-D-Pro-NK loop resulted in a 4-fold increase in the trpzip folding rate constant. This mutation replaces a flexible glycine with a conformation-restricting D-Pro, which is favored at this site in a type II'  $\beta$  turn. Du *et al.* argue that the rate acceleration is caused by the conformational rigidity of D-Pro (which reduces the entropic penalty for turn formation). In a computational folding simulation, the rigidity of a proline at a turn locus has resulted in folding frustration caused by kinetic traps (13). The GB1p to GB1m2 change also represents a stabilizing loop mutation (DDATKT  $\rightarrow$  NPATGK,  $\Delta\Delta G_U = 4.6$  kJ/mol) and includes the D7P mutation. In this case, the predominant effect is an increase (4-fold) in the folding rate constant. Does folding acceleration represent the greater compensating effect of a flexible glycine versus the restrictive proline, even though both are fold stabilizing? The conflicting results for a four-residue versus six-residue loop can be rationalized if the D-Pro-Asn locus reduces the configurational possibilities of the shorter loop to a much greater extent and biases the unfolded state toward conformations that are predisposed to turn and hairpin formation.

The GB1m2 to GB1m3 mutation ( $\Delta\Delta G_U = 1.4$ – $1.7$  kJ/mol) provides a previously unavailable probe of the effects of terminal Coulombic interactions upon dynamics. In this mutation, only the chain ends are altered: (GE—TE) is modified to (KK—QE). Viewed from the perspective of the folded state, this mutation replaces a potentially repulsive Coulombic interaction (the side chains of E2 and E16) with the possibility of additional attractive interactions. The net fold stabilization largely reflects a 1.9-fold increase in the folding rate constant. We suggest that this rate increase requires these Coulombic interactions be present before or at the folding transition state. There are two distinct views of the mechanism of hairpin formation. To be consistent with “zippering” hairpin folding models, with nucleation initiated by turn formation (31, 32), an extremely late transition state with ordering to the ends of the  $\beta$  strands would be required. In an early collapse folding scenario (12, 33, 34), the change in the Coulombic interactions

at the chain termini could exert an effect on the loop search required to establish the hydrophobic cluster: chain end repulsion in all of the analogs besides GB1m3 could be a factor in determining the folding rates. Because Coulombic interactions occur over longer distances than attractive van der Waals interactions, these could provide an effective barrier to visiting conformations that can lead to early hydrophobic cluster formation. Likewise, if the hydrophobic cluster forms before full definition of the turn conformation, the attractive interactions between the chain termini of GB1m3 could have the opposite effect, and thus increase the folding rate constant. In any case, the growing database of hairpin folding rates still does not provide a clear-cut distinction between the alternative folding mechanisms. Indeed, the mechanism may change both with loop size and for short loop sequences with unusually strong equilibrium preferences for turn-like conformations.

With the present article, the range of folding time constants for [4:4]-hairpins now spans from 3 to 20  $\mu s$ ; that for [2:2]-hairpin is even larger, from 0.83 (13) to 52  $\mu s$  (14). In the case of [2:2]-hairpins, it has been established that folding can be accelerated by prior collapse [starting from a cold-denatured state (15)] or cyclization (35). Less is known for the common [3:5]-hairpin motif of proteins. The single peptide model examined by temperature jump methods was a fast-folder [0.8  $\mu s$  (11)]. The FBP28 WW domain, with a five-residue loop, folds faster than comparable forms with a six-residue loop, and this folding acceleration can be moved to other sequences by loop substitution.<sup>†</sup> Fast folding may not, however, be a universal feature for these Gly bulge turns. In a study of trpzip analogs with Gly bulge turns, Blandl *et al.* (36) noted extreme exchange broadening of signals with large structuring shifts. It appears that fast folding requires not only a significant turn propensity but also a proper matching of turn geometry with the specific cross-strand interactions required for hairpin stabilization. To these loop optimization considerations we can now add mutations at the extreme termini of hairpins as another strategy for increasing the rate of hairpin formation. Properly positioned Coulombic attractions may represent an additional design element in protein folding optimization.

The internal referencing method for extracting exchange broadening presented herein extends the reach of dynamic NMR experiments to, and possibly beyond, the  $\mu s$  limit. This method should prove very useful because these NMR experiments are not technically demanding and less subject to instrumental artifacts than laser-induced temperature jump experiments. Their application requires NMR probes that display structuring shifts approaching (or greater than) 1 ppm and is limited to conditions in which the unfolded state is significantly populated ( $\chi_U \geq 0.1$ ). In addition, the extraction of folding and unfolding rate constants always requires a two-state folding assumption and relies on the equilibrium constant. For miniproteins and peptide models of secondary structure, the latter raises a caveat. Such systems display broad, low enthalpy melting transitions and complete sigmoidal melting curves are rarely available. Although statistical coil norms can be used for the expectation values for the unfolded state, the chemical shifts that represent 100% folding will be, in almost all cases, conjectures rather than firm experimental values. Even when chemical shifts appear to be leveling out at low temperatures for the most stable member of a series of analogs, one cannot eliminate the possibility that such an observation represents a broad inflection point, where the hot and cold denaturation transitions meet, and cannot be equated with  $>90\%$  folded. This potential source of error typically

<sup>†</sup>Gruebele, M., Tenth Annual Structural Biology Symposium, May 20–21, 2005, Galveston, TX.

leads to underestimation of  $k_F$ .<sup>‡</sup> However, for comparative studies within a series of analogs, so long as the structuring shifts are large and the mutations do not produce any signif-

icant change in  $\Delta\nu$ , linewidth observations over the  $\chi_U = 0.20$ – $0.65$  range will give relative folding rate constants suitable for the analysis of the sequence dependence of folding dynamics. We expect that this method will play a prominent role in the studies of the fast-folding limit (8) for proteins.

<sup>‡</sup>To include this potential source of error in the present study, we included  $\Delta\nu$  values and population estimates based on structuring shifts 23% larger than the most extreme values observed in our error propagation.

This work was supported by National Institutes of Health Grant GM59658.

1. Nauli, S., Kuhlman, B. & Baker, D. (2001) *Nat. Struct. Biol.* **8**, 602–605.
2. Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y. & Baker, D. (2002) *J. Mol. Biol.* **315**, 471–477.
3. Boffill, R., Simpson, E. R., Platt, G. W., Crespo, M. D. & Searle, M. S. (2005) *J. Mol. Biol.* **349**, 205–221.
4. Platt, G. W., Simpson, S. A., Layfield, R. & Searle, M. S. (2003) *Biochemistry* **42**, 13762–13771.
5. Kuhlman, B., Gautam, D., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003) *Science* **302**, 1364–1368.
6. Fink, A. L. (1998) *Folding Des.* **3**, R9–R23.
7. Horwich, A. (2002) *J. Clin. Invest.* **110**, 1221–1232.
8. Kubelka, J., Hofrichter, J. & Eaton, W. A. (2004) *Curr. Opin. Struct. Biol.* **14**, 76–88.
9. Muñoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997) *Nature* **390**, 196–199.
10. Andersen, N. H., Dyer, R. B., Fesinmeyer, R. M., Gai, F., Maness, S. & Werner, J. H. (2001) in *Peptides 2000*, eds. Martinez, J. & Fehrentz, J.-A. (EDK, Paris), pp. 553–554.
11. Xu, Y., Oyola, R. & Gai, F. (2003) *J. Am. Chem. Soc.* **125**, 15388–15394.
12. Dyer, R. B., Manas, E. S., Peterson, E. S., Franzen, S., Fesinmeyer, R. M. & Andersen, N. H. (2004) *Biochemistry* **43**, 11560–11566.
13. Snow, C. D., Qiu, L., Du, D., Gai, F., Hagen, S. J. & Pande, V. S. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 4077–4082.
14. Du, D., Zhu, Y., Huang, C.-Y. & Gai, F. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 15915–15920.
15. Dyer, R. B., Maness, S. J., Franzen, S., Fesinmeyer, R. M., Olsen, K. A. & Andersen, N. H. (2005) *Biochemistry* **44**, 10406–10415.
16. Huang, C.-Y., Getahun, Z., Wang, T. & DeGrado, W. F. (2001) *J. Am. Chem. Soc.* **123**, 12111–12112.
17. Werner, J. H., Dyer, R. B., Fesinmeyer, R. M. & Andersen, N. H. (2002) *J. Phys. Chem. B* **106**, 487–494.
18. Dyer, R. B., Gai, F., Woodruff, W. H., Gilmanshin, R. & Callender, R. H. (1998) *Acc. Chem. Res.* **31**, 709–716.
19. Eaton, W. A., Muñoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R. & Hofrichter, J. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359.
20. Huang, G. S. & Oas, T. G. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6878–6882.
21. Myers, J. K. & Oas, T. G. (2001) *Nat. Struct. Biol.* **8**, 552–558.
22. Wang, M., Tang, Y., Sato, S., Vugmeyster, L., McKnight, C. J. & Raleigh, D. P. (2003) *J. Am. Chem. Soc.* **125**, 6032–6033.
23. Kobayashi, N., Endo, S. & Munekata, E. (1993) *Peptide Chem.* 278–280.
24. Blanco, F. J., Rivas, G. & Serrano, L. (1994) *Nat. Struct. Biol.* **1**, 584–590.
25. Fesinmeyer, R. M., Hudson, F. M. & Andersen, N. H. (2004) *J. Am. Chem. Soc.* **126**, 7238–7243.
26. Cochran, A. G., Skelton, N. J. & Starovasnik, M. A. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 5578–5583.
27. Piette, L. H. & Anderson, W. A. (1959) *J. Chem. Phys.* **30**, 899–908.
28. Kobayashi, N., Honda, S., Yoshii, H., Uedaira, H. & Munekata, E. (1995) *FEBS Lett.* **366**, 99–103.
29. Kobayashi, N., Honda, S., Yoshii, H. & Munekata, E. (2000) *Biochemistry* **39**, 6564–6571.
30. Honda, S., Kobayashi, N. & Munekata, E. (2000) *J. Mol. Biol.* **295**, 269–278.
31. Muñoz, V., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5872–5879.
32. Espinosa, J. F., Muñoz, V. & Gellman, S. H. (2001) *J. Mol. Biol.* **306**, 397–402.
33. Pande, V. S. & Rokhsar, D. S. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9062–9067.
34. Dinner, A. R., Lazaridis, T. & Karplus, M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9068–9073.
35. Maness, S. J., Franzen, S., Gibbs, A. C., Causgrove, T. P. & Dyer, R. B. (2003) *Biophys. J.* **84**, 3874–3882.
36. Blandl, T., Cochran, A. G. & Skelton, N. J. (2003) *Protein Sci.* **12**, 237–247.