SUPPORTING MATERIAL

Scaling Properties of Glycine-Rich Sequences in Guanidine Hydrochloride Solutions

Michaela L. Finnegan and Bruce E. Bowler*
Department of Chemistry and Biochemistry and Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana

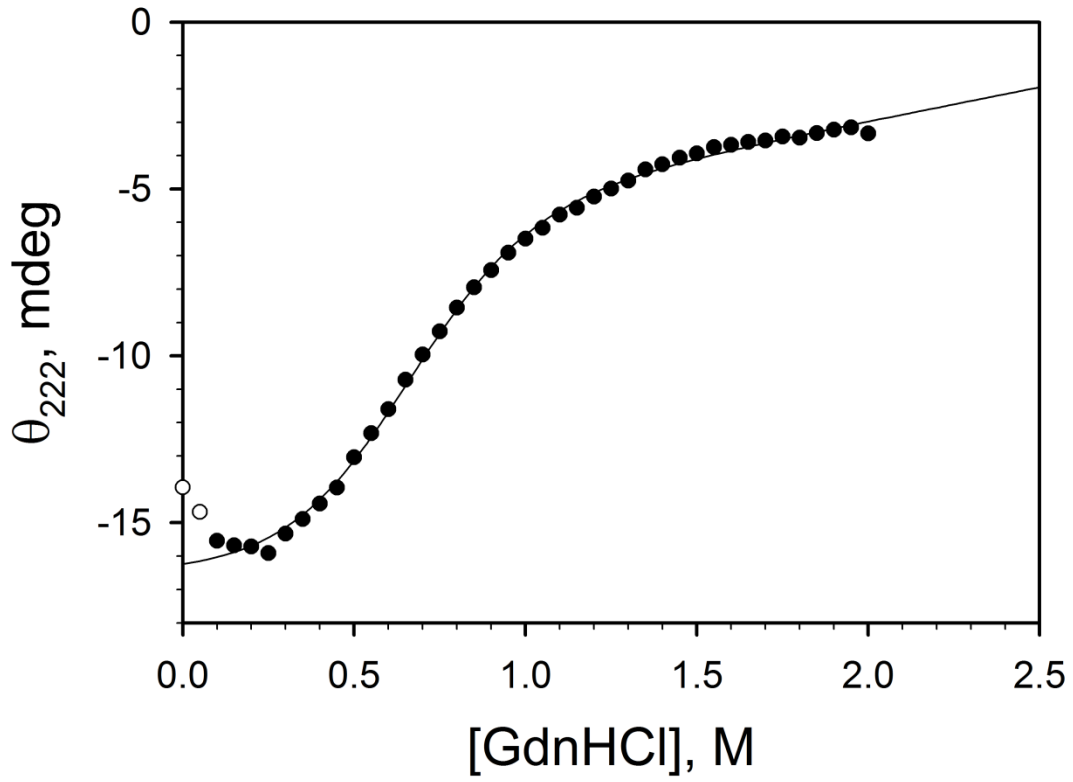

FIGURE S1 Plot of ellipticity at 222 nm vs. GdnHCl concentration for the NH5G-5 variant at 25 °C. The solid line is a fit to a two-state model which assumes a linear dependence of the free energy of unfolding, ΔG_u, on the concentration of GdnHCl. Only the closed circles are included in the fit.

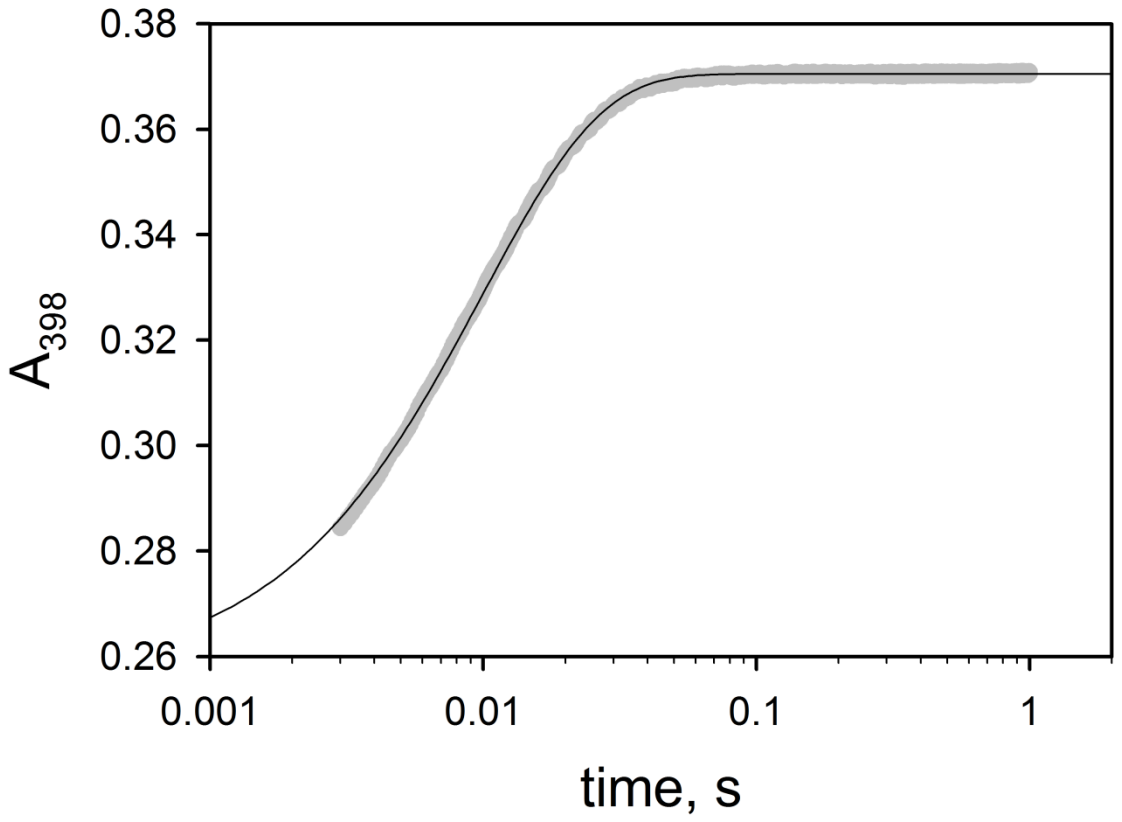

FIGURE S2 Plot of absorbance at 398 nm, A_{398}, versus time for NH5G-4 variant in 3.0 M GdnHCl after a pH jump at 25 °C. The solid line is a fit of the data to a single exponential rise to maximum equation. The starting pH was 6.2 and the ending pH was 3.24 ± 0.06.

TABLE S1 Primers used for insertion mutagenesis to produce the Gly-rich variants

Variant		Primer Sequence*
NH5G-2	Forward	5'-AAAGGGGGAGGGGGAGGGAAGGCCGGTTCTGCTAAG-3'
	Reverse	5'-CCCTCCCCCTCCCCCTTTACCGCCACCGCCACCGTG-3'
NH5G-3	Forward	5'-AAGGGCGGCGGTGGTGGCAGGGCCGGTTCTGCTAAG-3'
	Reverse	5'-GCCACCACCGCCGCCCTTCCCTCCCCCTCCCCCTTT-3'
NH5G-4	Forward	5'-AAAGGAGGAGGGGGGGGAAAGGCCGGTTCTGCTAAG-3'
	Reverse	5'-TCCCCCCCCTCCTCCTTTGCCACCACCGCCGCCCTT-3'
NH5G-5	Forward	5'-AAGGGTGGTGGAGGCGGCAAGGCCGGTTCTGCTAAG-3'
	Reverse	5'-GCCGCCTCCACCACCCTTTCCCCCCCCTCCTCCTTT-3'

*Underlined portions of each sequence are the sequence of the new insert. The remainder of each primer is used to anneal to the template DNA (NH5G-1 in the case of the NH5G-2 variant, etc.). Primers were designed so that the 5' end of each forward and reverse primer was complementary and contained the sequence of the new insert. All insertions were made so that the new insert was placed between the previous insert and the lysine next to Ala(−1) that was inserted in making the NH5G-1 variant. The portion of each forward primer highlighted in red anneals to the gene at the codon for the lysine next to Ala(-1) up through the codon for Lys4.

TABLE S2. Thermodynamic parameters for equilibrium loop formation at 20 ± 2 °C in 1.5, 3.0 and 6.0 M GdnHCl

Variant	Loop Size	pK_a(obs)*	n_p*
		1.5 M GdnHCl	
NH5G-1	22	4.41 ± 0.01	1.024 ± 0.005
NH5G-2	28	4.60 ± 0.02	0.99 ± 0.06
NH5G-3	34	4.78 ± 0.02	1.1 ± 0.1
NH5G-4	40	4.87 ± 0.01	0.96 ± 0.06
NH5G-5	46	5.02 ± 0.03	1.04 ± 0.03
		3.0 M GdnHCl	
NH5G-1	22	4.59 ± 0.02	1.04 ± 0.03
NH5G-2	28	4.81 ± 0.01	1.11 ± 0.02
NH5G-3	34	4.895 ± 0.005	1.03 ± 0.02
NH5G-4	40	5.04 ± 0.01	1.03 ± 0.02
NH5G-5	46	5.15 ± 0.01	1.12 ± 0.07
		6.0 M GdnHCl	
NH5G-1	22	4.92 ± 0.02	1.15 ± 0.06
NH5G-2	28	4.97 ± 0.01	1.06 ± 0.04
NH5G-3	34	5.14 ± 0.03	1.06 ± 0.01
NH5G-4	40	5.31 ± 0.02	1.05 ± 0.06
NH5G-5	46	5.46 ± 0.02	1.05 ± 0.05

*Parameters are the average and standard deviation of three trials.

TABLE S3 Values of k_{obs} for loop breakage in 3.0 M and 6.0 M GdnHCl at 25 °C for the Gly-rich iso-1-cytochrome *c* variants

Variant	**Loop Size**	k_{obs}*			
		3.0 M GdnHCl		**6.0 M GdnHCl**	
		pH 3.24 ± 0.06[†]	**pH 3.57 ± 0.01**	**pH 3.22 ± 0.05**[‡]	**pH 3.5 ± 0.1**
NH5G-1	22	111 ± 1[§]	122 ± 2[§]	76.9± 0.7	75.3 ± 0.9
NH5G-2	28	103.8 ± 0.8	113 ± 6	75.8 ± 0.9	74.2 ± 1.0
NH5G-3	34	104.1 ± 0.9	110 ± 1	76.7 ± 0.6	74.0 ± 0.9
NH5G-4	40	102 ± 2	97.4 ± 0.6	77.0 ± 0.9	74.9 ± 0.7
NH5G-5	46	101 ± 1	94 ± 2	77.5 ± 0.9	74.9 ± 1.1

*Values are the average and standard deviation of eight trials.

[†]Using Eq. 4 from the main text, we calculate that the contribution of $k_{f,His}$ to k_{obs} is 1.3 to 2.8 s^{-1} at this pH, similar to the error in k_{obs}. Therefore, we have not corrected for this contribution and use k_{obs} at pH 3.24 as $k_{b,His}$ at 3.0 M GdnHCl.

[‡]Using Eq. 4 from the main text, we calculate that the contribution of $k_{f,His}$ to k_{obs} is 0.5 to 1.5 s^{-1} at this pH, similar to the error in k_{obs}. Therefore, we have not corrected for this contribution and use k_{obs} at pH 3.22 as $k_{b,His}$ at 6.0 M GdnHCl.

[§]Data are from supporting reference (1).

TABLE S4 Viscosity corrected loop breakage rate constants

Variant	Loop size	$k_b^{6M}{}_{corr}$*,† s^{-1}
NH5G-1	22	106.5 ± 1.0
NH5G-2	28	105.0 ± 1.2
NH5G-3	34	106.2 ± 0.8
NH5G-4	40	106.6 ± 1.2
NH5G-5	46	107.3 ± 1.2

*$k_b{}^{6M}{}_{corr}$ values were calculated using the uncorrected k_{obs} values in Table S3 (pH 3.22 data) at 6.0 M GdnHCl for $k_{b,His}$(6M), and the following equation:

$$k_{b,corr}^{6M} = k_{b,His}(6M)\left(\frac{\eta_{6M}}{\eta_{3M}}\right)$$

where η_{6M} and η_{3M} are the viscosities of 6.0 M GdnHCl and 3.0 M GdnHCl, respectfully. These values were obtained using a 4th order polynomial fit to the data in supporting reference (2). This equation is based on the observed effect of viscosity on loop formation in supporting reference (3).

†Error is propagated from the error in the uncorrected $k_{b,His}$(6M) values.

TABLE S5 Calculated rate constants for loop formation, $k_{f,His}$, in 3.0 M and 6.0 M GdnHCl at 25 °C for iso-1-cytochrome *c* variants

Variant	Loop size	$k_{f,His}$*	
		3.0 M GdnHCl	**6.0 M GdnHCl**
NH5G-1	22	$11.3 \pm 0.5 \times 10^{3\dagger}$	$3.7 \pm 0.1 \times 10^{3}$
NH5G-2	28	$6.4 \pm 0.2 \times 10^{3}$	$3.25 \pm 0.06 \times 10^{3}$
NH5G-3	34	$5.27 \pm 0.08 \times 10^{3}$	$2.2 \pm 0.1 \times 10^{3}$
NH5G-4	40	$3.7 \pm 0.1 \times 10^{3}$	$1.50 \pm 0.08 \times 10^{3}$
NH5G-5	46	$2.9 \pm 0.1 \times 10^{3}$	$1.07 \pm 0.06 \times 10^{3}$

*Loop formation rate constants, $k_{f,His}$, are calculated with the following equation: $k_{f,His} = k_{b,His} \times K_{loop}(His) = k_{b,His} \times 10^{-pK_{loop}(His)}$. The reported error is from standard propagation of the errors in $k_{b,His}$ and $pK_{loop}(His)$.

†Data from supporting reference (1).

.

Derivation of Eq 5.

Starting with the expression for $k_{f,His}$ based on transition state theory and assuming that the activation free energy of loop formation is entirely entropic, $\Delta G^{\ddagger}_{loop} = -T\Delta S^{\ddagger}_{loop}$, we have:

$$k_{f,His} = k^{o}_{f,His} e^{-\Delta G^{\ddagger}_{loop}/RT} = k^{o}_{f,His} e^{\Delta S^{\ddagger}_{loop}/R} \quad \text{(S1)}$$

$$\ln(k_{f,His}) = \ln(k^{o}_{f,His}) + \frac{\Delta S^{\ddagger}_{loop}}{R} \quad \text{(S2)}$$

In Eq. S1 and Eq. S2, $k^{o}_{f,His}$ is the rate constant for loop formation when $\Delta G^{\ddagger}_{loop} = -T\Delta S^{\ddagger}_{loop} = 0$.

We assume that the Jacobson-Stockmayer equation, Eq. 1 in the main text can be used for $\Delta S^{\ddagger}_{loop}$ in Eq. S2 yielding Eq. S3:

$$\ln(k_{f,His}) = \ln(k^{o}_{f,His}) + \ln\left(\left(\frac{3}{2\pi C_n \ell^2}\right)^{\nu_3} V_i\right) - \upsilon_3 \ln(n) \quad \text{(S3)}$$

Grouping terms and converting to common Log, we obtain Eq. S4

$$\mathrm{Log}(k_{f,His}) = \mathrm{Log}\left(k^{o}_{f,His}\left(\frac{3}{2\pi C_n \ell^2}\right)^{\nu_3} V_i\right) - \upsilon_3 \mathrm{Log}(n) \quad \text{(S4)}$$

Setting the first term on the right side of Eq S4 equal to $\mathrm{Log}(k_{f,His\_ref})$ converts Eq S4 to Eq 5 in the main text.

Derivation of Eq 6.

Starting with Eq S4 we can write expressions for the loop formation rate constants for Ala-rich and Gly-rich sequences, $k_{f,His}(\mathrm{Ala})$ and $k_{f,His}(\mathrm{Gly})$, respectively:

$$\mathrm{Log}\left(k_{f,His}(\mathrm{Ala})\right) = \mathrm{Log}\left(k^{o}_{f,His}\left(\frac{3}{2\pi C_n(\mathrm{Ala})\ell^2}\right)^{\nu_{3,\mathrm{Ala}}} V_i\right) - \upsilon_3 \mathrm{Log}(n) \quad \text{(S5)}$$

$$\mathrm{Log}\left(k_{f,His}(\mathrm{Gly})\right) = \mathrm{Log}\left(k^{o}_{f,His}\left(\frac{3}{2\pi C_n(\mathrm{Gly})\ell^2}\right)^{\nu_{3,\mathrm{Gly}}} V_i\right) - \upsilon_3 \mathrm{Log}(n) \quad \text{(S6)}$$

When Eq S5 is subtracted from Eq S6, the $\nu_3 \mathrm{Log}(n)$ terms cancel and we obtain Eq S7:

$$\mathrm{Log}\left(k_{\mathrm{f},His}(\mathrm{Gly})\right) - \mathrm{Log}\left(k_{\mathrm{f},His}(\mathrm{Ala})\right) = \mathrm{Log}\left(\left(\frac{3}{2\pi C_n(\mathrm{Gly})\ell^2}\right)^{\nu_{3,\mathrm{Gly}}}\left(\frac{2\pi C_n(\mathrm{Ala})\ell^2}{3}\right)^{\nu_{3,\mathrm{Ala}}}\right) \quad \text{(S7)}$$

With $\nu_{3,\mathrm{Gly}} \neq \nu_{3,\mathrm{Ala}}$, there is no straightforward way to simplify Eq S7. Therefore, we make the simplifying assumption that the scaling exponent can be approximated by the average value of the scaling exponent for the Ala-rich and Gly-rich sequences, $\bar{\upsilon}_3 = (\nu_{3,\mathrm{Gly}} + \nu_{3,\mathrm{Ala}})/2$ at a given GdnHCl concentration. With this simplification, we can write Eq S8:

$$\mathrm{Log}\left(k_{\mathrm{f},His}(\mathrm{Gly})\right) - \mathrm{Log}\left(k_{\mathrm{f},His}(\mathrm{Ala})\right) = \bar{\nu}_3 \,\mathrm{Log}\left(\frac{C_n(\mathrm{Ala})}{C_n(\mathrm{Gly})}\right) \quad \text{(S8)}$$

Dividing through by $\bar{\upsilon}_3$ and taking the antilog of both sides of the equation, we obtain Eq. S9, which is Eq. 6 in the main text

$$\frac{C_n(\mathrm{Ala})}{C_n(\mathrm{Gly})} = 10^{\left(\frac{\mathrm{Log}\left(k_{\mathrm{f},His}(\mathrm{Gly})\right) - \mathrm{Log}\left(k_{\mathrm{f},His}(\mathrm{Ala})\right)}{\bar{\nu}_3}\right)} \quad \text{(S9)}$$

SUPPORTING REFERENCES

1. Tzul, F. O., and B. E. Bowler. 2009. Importance of contact persistence in denatured state loop formation: kinetic insights into sequence effects on nucleation early in folding. J. Mol. Biol. 390:124-134.
2. Kawahara, K., and C. Tanford. 1966. Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. J. Biol. Chem. 241:3228-3232.
3. Möglich, A., F. Krieger, and T. Kiefhaber. 2005. Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. J. Mol. Biol. 345:153-162.