Implications of molecular heterogeneity for the cooperativity of biological macromolecules

SUPPLEMENTARY INFORMATION

Sergey V. Solomatin, Max Greenfeld, Daniel Herschlag

TABLE OF CONTENTS

SUPPLEMENTARY FIGURES

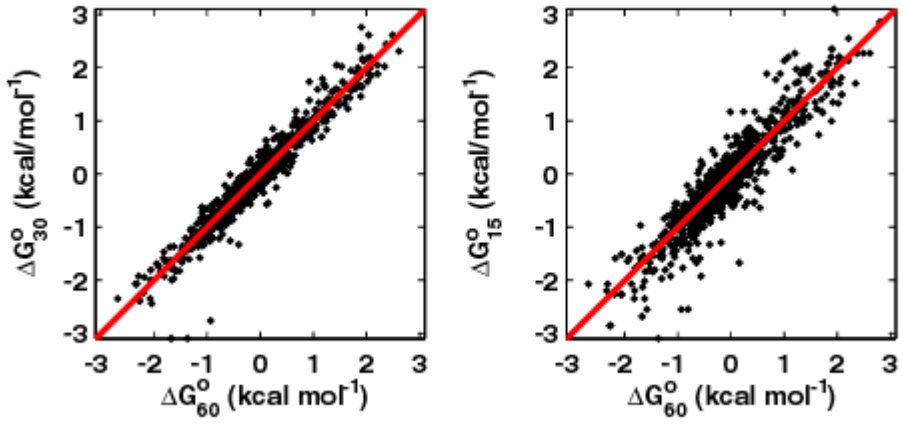

Figure 1: $\Delta G^{o}_{folding}$ measured from smFRET traces converge within 30 s of observation time. P4-P6 folding was observed in 15 mM Ba^{2+}. (left panel) Red diagonal line corresponds to $\Delta G^{o}_{folding}$ measured from 30 s traces (denoted as ΔG^{o}_{30}) being exactly equal to $\Delta G^{o}_{folding}$ measured from 60 s traces (denoted as ΔG^{o}_{60}) (right panel) Red diagonal line corresponds to $\Delta G^{o}_{folding}$ measured from 15 s traces (denoted as ΔG^{o}_{15}) being exactly equal to $\Delta G^{o}_{folding}$ measured from 60 s traces (in the right panel).

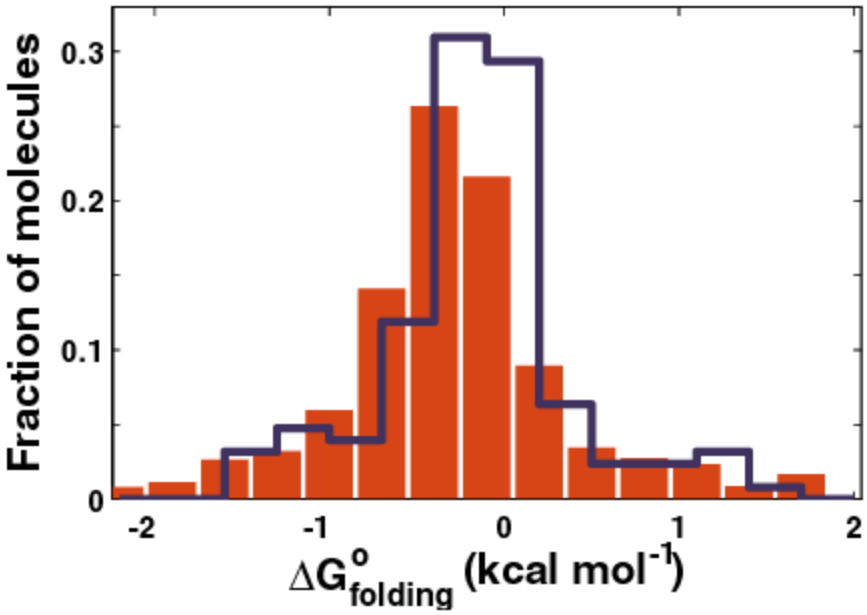

Figure 2: Heterogeneity of Ba^{2+}-dependent folding measured by smFRET is independent the ensemble size. For a large ensemble of P4-P6 molecules (red bars, N = 1241) at [Ba^{2+}] = 15 mM heterogeneity parameter $H = 0.7 \pm 0.1$ kcal mol^{-1} (errors are s.e.m.). For a small ensemble of P4-P6 molecules (blue line, N = 126) at [Ba^{2+}] = 15 mM heterogeneity parameter $H = 0.7 \pm 0.2$ kcal mol^{-1}.

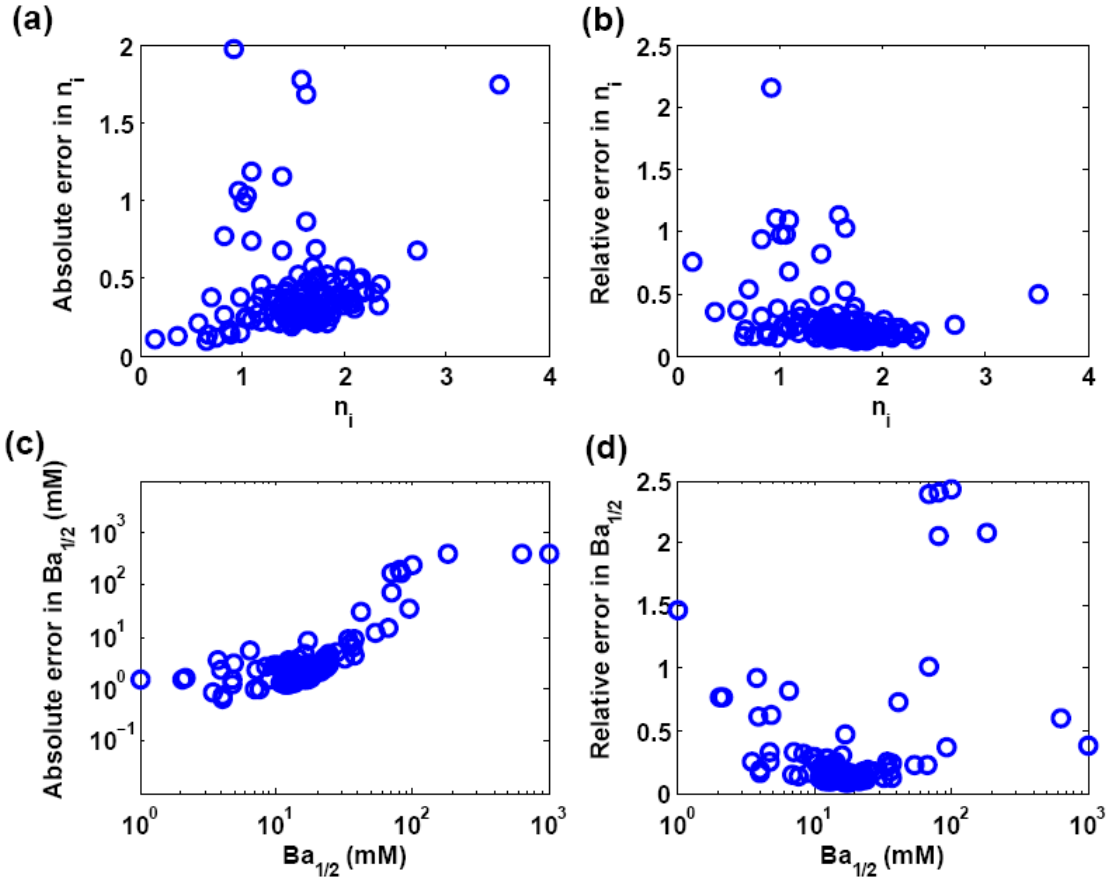

Figure 3: Hill parameters for 126 individual molecules and errors associated with each parameter. (a) The absolute errors in Hill cooperativity parameters, $\sigma(n_i)$. (b) The relative errors in Hill cooperativity parameters $\varepsilon(n_i) = \sigma(n_i)/n_i$. (c) The absolute errors in midpoints, $\sigma(Ba_{1/2})$. (d) The relative errors in midpoints $\varepsilon(Ba_{1/2}) = \sigma(Ba_{1/2})/Ba_{1/2}$.

MATERIALS

Chemicals

DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc., and RNA oligonucleotides were purchased from Dharmacon (Thermo Fisher Scientific, Inc.). Biotinylated bovine serum albumin (bBSA) and streptavidin (both from Sigma-Aldrich) were dissolved in storage buffer (50 mM NaMOPS, pH 7.0, 50% glycerol), and stored at –20 °C without freezing. Trolox [(±)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid] was from Aldrich (>97% purity) and was used without further purification. Glucose oxidase (Type VII from *Aspergilius niger*, Sigma-Aldrich) and catalase (Roche) were used as supplied. All standard chemicals were purchased from Sigma-Aldrich and used without further purification.

RNA constructs

Fluorescently labelled P4-P6 RNA of the *Tetrahymena* group I ribozyme, extended at the 3′-end by 26 nucleotides (ACCAAAAUCAACCUAAAACUUACACA, T2 "tail"), was synthesized by a five-piece splinted ligation, as described previously.[1] RNA fragments were synthesized by *in vitro* transcription from PCR-amplified DNA templates and purified by denaturing PAGE using UV shadowing (254 nm hand-held UV lamp). RNA oligonucleotides were labelled with the fluorescent dyes, as described,[2] and purified by denaturing PAGE.

Standard buffers

"Standard buffer" composition: 50 mM NaMOPS (pH 7.0), 80 mM NaCl.

"Annealing buffer" composition: 10 mM NaMOPS (pH 7.0), 10 μM $MgCl_2$.

"smFRET buffer" composition: 50 mM MOPS (pH 7.0), 80 mM NaCl, 1-500 mM $BaCl_2$, 2 mg/ml glucose, 1.8 mM Trolox, 100 units/ml glucose oxidase, 1000 units/ml catalase.

METHODS

Simulations

Binding isotherms of a heterogeneous ensemble of molecules were simulated using *Matlab R2008a* (Mathworks), as follows. A ligand concentration of 1 mM was chosen as a standard state, and standard state free energies of binding $\Delta G^{o}_{binding}$ for 3000 molecules were drawn from a normal distribution (generated by *randn* function in *Matlab R2008a*) with a mean of 5 kcal mol^{-1}, and a standard deviation H = 1.5 kcal mol^{-1}. For each i-th molecule, free energies of binding over a range of ligand concentrations from 1 to 500 mM were calculated using Equation S1 derived from the Hill equation:

$$\Delta G^{i}_{binding} = \Delta G^{o,i}_{binding} - n \times RT \times \ln[Ligand] \qquad \text{(S1)},$$

where n was equal to 3 for each molecule.

From these simulated data, the fraction bound for each molecule at each ligand concentration was calculated using Equation 2 from the main text. The bulk fraction bound at each ligand concentration was calculated by summing over the fractions bound for each molecule. The Hill equation (Eq. 1) was fit to the bulk fraction bound using a non-linear least square optimization routine implemented in *Matlab R2008a* (*lsqcurvefit* function). The simulations to obtain n_{bulk} over the range of n and H values were performed in the same way for each combination of n and H.

Single molecule FRET experiments

Sample preparation. A DNA tether (biotin-TGTGTAAGTTTTAGGTTGATTTTGGT) was bound to the P4-P6 RNA by annealing at 50 ^{o}C for 30 min in the annealing buffer, followed by 0.1 ^{o}C per s cooling to 4 ^{o}C. The annealed complex was then diluted to 30-100 pM for deposition on the slides.

Quartz slides (G. Finkenbeiner Inc.) were cleaned by treatment with Piranha solution for 1 hour, followed by KOH etching under sonication for 30 min and extensive rinsing with de-ionized water, and stored in lime glass beakers under de-ionized water before use. Flow chambers were constructed by attaching a glass coverslip to a quartz slide using double-sided Scotch tape. Each slide had two through-holes pre-drilled at the opposite ends, and a Luer-type connector was glued over one of the holes from the side opposite the coverslip. The flow chamber was filled with 50 mM NaMOPS buffer (pH 7.0) and coated with bBSA (10 min incubation of 1 mg ml^{-1} bBSA solution [50 mM NaMOPS, pH 7.0, 50% glycerol]), followed by rinsing with a 10x volume of standard buffer and coating with streptavidin (10 min incubation of 0.1 mg ml^{-1} streptavidin solution [50 mM NaMOPS, pH 7.0, 50% glycerol]).

Data acquisition. A diode-pumped solid-state green laser (532 nm; Gem, Laser Quantum) and a diode red laser (635 nm; Hitachi HL6344G; maximum power, 10 mW) were combined using dichroic mirrors and focused through a prism onto the surface of a quartz slide below a critical angle. We adjusted the intensity of the green laser to achieve an average signal-to-noise ratio of >5, which typically required a power of ~20 mW at the laser aperture. Fluorescent images were collected using a ×60 water-immersion Nikon

objective (numerical aperture, 1.2), filtered through a 550-nm long-pass filter (Chroma Technology) to remove scattered excitation light, and chromatically separated using dichroic mirrors (635-nm cut-off) into a 'green' image and a 'red' image. We filtered the 'green' image through a 580/30-nm band-pass filter, and the 'red' image, through a 670/30-nm band-pass filter, and focused them, respectively, onto the left and right halves of a back-illuminated charge-multiplying charge-coupled device (CCD) (Cascade:128+, Photometrics, Roper Scientific). Full CCD images (128×128 pixels) were read out in 40-ms frames with a conversion gain of 3 and multiplication gain of 3,200 and saved as 16 bit data.

Design of single molecule titration experiments. A flow chamber with immobilized P4-P6 molecules was rigidly mounted on the microscope stage. A flexible polyethylene tube was connected to the outlet of the Luer connector to allow buffer injection without changing the position of the slide. A series of "smFRET buffers" with varied $BaCl_2$ concentrations was injected into the chamber. After each injection the system was equilibrated for 5 minutes with the laser switched off and the camera not recording. Following the equilibration period, the laser was switched on and the fluorescence data were recorded for 30 s. In a few experiments, data were recorded for 20 s instead of 30 s. After that measurement, a fresh buffer with different $BaCl_2$ concentration was injected, and the procedure repeated as above until all molecules photobleached.

Data analysis. Data collection and analysis were performed using home-written programs in C++ and Matlab (Mathworks Co.). To obtain FRET data, we first determined the positions of pixels that contained fluorescently active molecules by

averaging the first 30 image frames and locating the pixels on the 'red' side of the CCD with an intensity that exceeded background by a certain amount (typically a threshold of 5σ, where σ is the standard deviation of the background fluorescence). The corresponding positions on the 'green' side were determined by applying linear offsets determined independently from images of fluorescent beads that are visible on both sides of the CCD. The local background (7×7 pixel) was subtracted from each spot for each frame. Time traces of the fluorescence intensity of Cy3 (I_{green}) and the Cy5 (I_{red}) for each spot were recorded and used to calculate FRET traces:

$$FRET = \frac{I_{red} - I_{cross-talk}}{I_{green} + I_{red} - I_{cross-talk}} \quad \text{(S2)}$$

The cross-talk $I_{cross\text{-}talk}$ is the fluorescence in the red channel arising from imperfect chromatic separation of Cy3 fluorescence from Cy5 fluorescence. Background subtraction occasionally leads to negative fluorescence intensity values, which results in calculated values of FRET <0, or FRET >1. To improve visualization, FRET traces are truncated below FRET = –0.2 and above FRET = 1.2. This procedure does not affect the thermodynamic and kinetic analysis of FRET data.

Pixels that contained fluorescently active molecules were identified according to the criteria previously described,[3] and the coordinates of the pixels were recorded. To identify which pixels correspond to the same molecule in multiple movies of the titration series, we compared the coordinates of molecules in each movie to the coordinates in the first movie. The molecules that co-localized within a 1 pixel distance were accepted as identical. In a few instances, linear offsets of 1-2 pixels had to be applied to all molecules to obtain co-localization. Presumably, these instances correspond to occasional small displacement of the flow chamber relative to the objective at the time of the buffer

exchange operation. After co-localization of the two movies, molecules in the third and all subsequent movies were co-localized using the same procedure.

At each Ba^{2+} concentration, the thermodynamic and kinetic information about each molecule was extracted from FRET traces. The fraction of the time in the folded state was measured by thresholding. The fraction folded (*f*) was defined as a ratio of the time when FRET > $FRET_{threshold}$ (t_{high}) to the total length of the trace:

$$f = t_{high}/(t_{low} + t_{high}) \quad \text{(S3)}$$

The free energy of folding $\Delta G_{folding}$ was calculated from *f*:

$$\Delta G_{folding} = -RT \times \ln \frac{f}{1-f} \quad \text{(S4)}$$

Folding kinetics for each molecule were obtained by analyzing the fluctuations in I_{green} and I_{red} using a maximum likelihood Hidden Markov-type model[4,5] describing transitions between an unfolded state U (parameters I^{U}_{green}, $\sigma(I^{U}_{green})$, I^{U}_{red}, $\sigma(I^{U}_{red})$) and a folded state F (parameters I^{F}_{green}, $\sigma(I^{F}_{green})$, I^{F}_{red}, $\sigma(I^{F}_{red})$) with two rate constants, $k_{folding}$ *and* $k_{unfolding}$. Detailed description of the procedure is in preparation for publication (Greenfeld, M., Pavlichin, D., Mabuchi, H., Herschlag, D.).

The folding isotherm for each molecule was fit to the Hill equation:

$$f = \frac{([Ba^{2+}]/Ba^{2+}_{1/2})^{n}}{1+([Ba^{2+}]/Ba^{2+}_{1/2})^{n}} \quad \text{(S5)},$$

using the least square optimization routine *lsqcurvefit* as implemented in *MatlabR2008a* (Mathworks). The values of $Ba^{2+}_{1/2}$ and *n* for each molecule are plotted in **Fig. S2** along with the estimated errors obtained as described below.

Estimating variability in $\Delta G^{o}_{folding}$ ***arising from finite time sampling.*** The major source of errors in measuring $\Delta G^{o}_{folding}$ from single molecule FRET traces arises from finite time sampling. Individual molecules are observed over a limited time interval, and f measured over that time is only approximately equal to f_∞ measured over an infinitely long period of time. We determined the $\sigma_{sampling}$ error for each Ba^{2+} titration using simulations as follows. We simulated measurements of f multiple times for the same molecule, using the equilibrium and kinetic constants obtained for that molecule at a given [Ba^{2+}].

$$f_\infty = \frac{k_{fold} / k_{unfold}}{1 + k_{fold} / k_{unfold}} \qquad \textbf{(S6)}$$

This simulation gave a distribution of f values, from which we obtained a distribution of $\Delta G^{o}_{folding}$. The width of this distribution corresponds to the statistical error in measuring $\Delta G^{o}_{folding}$ arising from finite time sampling. For >90% of molecules, these errors were smaller than 0.3 kcal mol^{-1} (data not shown).

Furthermore, to establish validity of this computational approach by an empirical experimental approach, we measured $\Delta G^{o}_{folding}$ for individual molecules from 15 s, 30 s, and 60 s traces. $\Delta G^{o}_{folding}$ measured from 30 s and 60 s traces show excellent correlation (**Fig. S1**, left panel), as >98% of the points fall along the diagonal corresponding to $\Delta G^{o}_{folding}$(30s) = $\Delta G^{o}_{folding}$(60s). Spread of the points off of the diagonal corresponds to

the variability in $\Delta G^{o}_{folding}$. The average variability per molecule was estimated by calculating the root-mean-squared difference between $\Delta G^{o}_{folding}$ measured from 60 s and 30 s traces, as follows:

$$\sigma_{\Delta G} = \frac{1}{N}\sum\sqrt{(\Delta G_i^{30s} - \Delta G_i^{60s})^2} \qquad \text{(S7)}$$

The measured value $\sigma_{\Delta G}$ = 0.21 kcal mol^{-1} is in a good agreement with the computational estimates of the finite time sampling errors (0.25 kcal mol^{-1}).

$\Delta G^{o}_{folding}$ measured from 15 s and 60 s traces also show a good correlation (**Fig. S1**, right panel), but a larger spread from the diagonal. The variability in $\Delta G^{o}_{folding}$ is $\sigma_{\Delta G}$ = 0.32 kcal mol^{-1}. Increased variability is expected for shorter traces, and is consistent with the simulation results (data not shown).

Error analysis for individual Hill fits. The errors in the Hill parameters obtained for the Ba^{2+} titrations come from three principal sources: the error in measuring f (σ_f), the error from using f measured over a limited time interval ($\sigma_{sampling}$), and the error of the fit (σ_{fit}). The error σ_f was estimated as the total probability of observing FRET > $FRET_{threshold}$ when the molecule is in the low FRET state plus the probability of observing FRET < $FRET_{threshold}$ when the molecule is in the high FRET state. We used the maximum likelihood Hidden Markov model described above to compute these probabilities and found σ_f to be negligible (<0.3% for all molecules).

The $\sigma_{sampling}$ error arises from finite time sampling, as described above. We determined the $\sigma_{sampling}$ error for each Ba^{2+} titration using simulations as follows. We

simulated measurements of *f* multiple times for the same molecule, using the equilibrium and kinetic constants obtained for that molecule at a given [Ba^{2+}].

$$f_{\infty} = \frac{k_{fold} / k_{unfold}}{1 + k_{fold} / k_{unfold}} \qquad \textbf{(S6)}$$

This simulation gave a distribution of *f* values. We repeated this simulation at each Ba^{2+} concentration used in our titrations to obtain a simulated Ba^{2+} titration. For each molecule titrated, we carried out 1000 such simulations and obtained a set of n_i values by fitting each simulated titration to the Hill equation. The standard deviation of the n_i values obtained is the sampling error in n_i, $\sigma_{sampling}$, for that molecule.

We report the standard deviation of the distribution of *n* and $Ba_{1/2}$ as the absolute errors $\sigma(n_i)$ and $\sigma(Ba_{1/2})$, respectively, in **Fig. S3**. For the bulk isotherm, the sampling error was much smaller than the fit error, so the total error reported for n_{bulk} is comprised of only the fit error.

SUPPLEMENTARY REFERENCES

References 6-14 provide additional examples of heterogeneous molecular behavior.[6-14]

These are not intended as a comprehensive review of up to date literature on the subject.

1. Sattin, B.D., Zhao, W., Travers, K., Chu, S. & Herschlag, D. *J. Am. Chem. Soc.* **130**, 6085-6087 (2008).
2. Solomatin, S. & Herschlag, D. Methods Of Site-Specific Labeling Of RNA With Fluorescent Dyes. in *Methods Enzymol.*, Vol. 469 47-68 (Elsevier Academic Press Inc, San Diego, 2009).
3. Solomatin, S.V., Greenfeld, M., Chu, S. & Herschlag, D. *Nature* **463**, 681-684 (2010).
4. Cappe, O., Moulines, E. & Ryder, T. *Inference in Hidden Markov Models*, (Springer Science+Buisness Media, 2005).
5. Welch, L.R. *IEEE Information Theory Society Newsletter* **53** (2003).
6. Ha, T. et al. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9077-9082 (1999).
7. Tan, E. et al. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9308-9313 (2003).
8. Yang, H. et al. *Science* **302**, 262-266 (2003).
9. Rueda, D. et al. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10066-10071 (2004).
10. Xie, Z., Srividya, N., Sosnick, T.R., Pan, T., Scherer, N.F. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 534-539 (2004).
11. English, B.P. et al. *Nat. Chem. Biol.* **2**, 87-94 (2006).
12. Liu, S., Bokinsky, G., Walter, N.G. & Zhuang, X. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12634-12639 (2007).
13. Gell, C. et al. *J. Mol. Biol.* **384**, 264-278 (2008).
14. Li, Y. et al. *J. Phys. Chem. B* **113**, 7579-7590 (2009).