Supplementary Information for
Effects of cryo-EM cooling on structural ensembles

Lars V. Bock*[1] and Helmut Grubmüller[1]

[1]Theoretical and Computational Biophysics Department, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Supplementary Methods

Numerical solution of the heat equation for a water layer

To solve the heat equation, we used the Crank-Nicolson method (central difference in time and in space, CTCS)[1] with discretized time and space $T(i\Delta x, n\Delta t) = T_i^n$, with $\Delta x = 1$ nm and $\Delta t = 0.01$ ns:

$$\frac{T_i^{n+1} - T_i^n}{\Delta t} = \frac{1}{2}\left(\frac{[\alpha(x)\partial_x T(x,t)]_{x=(i+\frac{1}{2})\Delta x}^{t=(n+1)\Delta t} - [\alpha(x)\partial_x T(x,t)]_{x=(i-\frac{1}{2})\Delta x}^{t=(n+1)\Delta t}}{\Delta x} + \frac{[\alpha(x)\partial_x T(x,t)]_{x=(i+\frac{1}{2})\Delta x}^{t=n\Delta t} - [\alpha(x)\partial_x T(x,t)]_{x=(i-\frac{1}{2})\Delta x}^{t=n\Delta t}}{\Delta x} \right).$$

With $l_i = \frac{\Delta t}{4\Delta x^2}\alpha\left((i-\frac{1}{2})\Delta x\right)$ and $r_i = \frac{\Delta t}{4\Delta x^2}\alpha\left((i+\frac{1}{2})\Delta x\right)$, the discretized heat equation can be written as

$$-l_i T_{i-1}^{n+1} + (1 + r_i + l_i)T_i^{n+1} - r_i T_{i+1}^{n+1} = l_i T_{i-1}^n + (1 - r_i - l_i)T_i^n + r_i T_{i+1}^n.$$

For vectors $\boldsymbol{T}^{n+1} = (T_2^{n+1}, \cdots, T_{N-1}^{n+1})^\intercal$ and $\boldsymbol{T}^n = (T_2^n, \cdots, T_{N-1}^n)^\intercal$, this equation can be written in matrix form $A \cdot \boldsymbol{T}^{n+1} = B \cdot \boldsymbol{T}^n + \boldsymbol{b}$, where A and B and $(N-2) \times (N-2)$ matrices:

$$A = \begin{pmatrix} 1+r_2+l_2 & -r_2 & 0 & \cdots & & & & 0 \\ -l_3 & 1+r_3+l_3 & -r_3 & 0 & \cdots & & & 0 \\ 0 & -l_4 & 1+r_4+l_4 & -r_4 & 0 & \cdots & & 0 \\ \vdots & & & & \ddots & & & \vdots \\ & & & & & -l_{N-2} & 1+r_{N-2}+l_{N-2} & -r_{N-2} \\ 0 & \cdots & & & & 0 & -l_{N-1} & 1+r_{N-1}+l_{N-1} \end{pmatrix},$$

$$B = \begin{pmatrix} 1-r_2-l_2 & r_2 & 0 & \cdots & & & & 0 \\ l_3 & 1-r_3-l_3 & r_3 & 0 & \cdots & & & 0 \\ 0 & l_4 & 1-r_4-l_4 & r_4 & 0 & \cdots & & 0 \\ \vdots & & & & \ddots & & & \vdots \\ & & & & & l_{N-2} & 1-r_{N-2}-l_{N-2} & r_{N-2} \\ 0 & \cdots & & & & 0 & l_{N-1} & 1-r_{N-1}-l_{N-1} \end{pmatrix}.$$

The $N-2$ dimensional vector $\boldsymbol{b}$ contains the temperature $T_b = 90$ K at the boundary grid points $i = 1$ and $i = N$, $\boldsymbol{b} = \frac{T_b \Delta t}{2\Delta x^2}(\alpha(\Delta x), 0, \ldots, 0, \alpha(N\Delta x))^\intercal$. After multiplying with the inverse of matrix A, we get

$$\boldsymbol{T}^{n+1} = A^{-1}B \cdot \boldsymbol{T}^n + A^{-1}\boldsymbol{b}.$$

The starting temperature profile $\boldsymbol{T}^0$ was set to a temperature of 90 K for the ethane layer and to 277.15 K for the water layer (Fig. 1a, blue line). Next, equation was iterated until the temperature in the center of the profile dropped below 90.1 K (Fig. 1).

*To whom correspondence should be addressed. Email: lbock@mpibpc.mpg.de

Metropolis sampling with Bayesian inference

The models of the cooling process model1, model2 and model3 have different sets of parameters p. Here, the aim is to obtain the probability distribution $P(\boldsymbol{p} \mid \boldsymbol{Q}, \boldsymbol{S})$ of the parameter set $\boldsymbol{p}$ given the rmsf quantiles obtained from the T-quench simulations $\boldsymbol{Q} = \{\boldsymbol{Q}_{\tau_c} \mid \tau_c \in \boldsymbol{\tau}_c\}$, where $\boldsymbol{\tau}_c$ is the set of cooling time spans used for training the model, e.g., $\boldsymbol{\tau}_c = [0.1\text{ns}, \ldots, 128\text{ns}]$. The standard deviations of the rmsf values are given by $\boldsymbol{S} = \{\boldsymbol{S}_{\tau_c} \mid \tau_c \in \boldsymbol{\tau}_c\}$ as described above. According to Bayes' theorem,

$$P(\boldsymbol{p} \mid \boldsymbol{Q}, \boldsymbol{S}) \propto P(\boldsymbol{Q}, \boldsymbol{S} \mid \boldsymbol{p}) \cdot P(\boldsymbol{p}), \tag{1}$$

with the likelihood $P(\boldsymbol{Q}, \boldsymbol{S} \mid \boldsymbol{p})$ and the prior probability of the parameters $P(\boldsymbol{p})$. The likelihood is given by

$$P(\boldsymbol{Q}, \boldsymbol{S} \mid \boldsymbol{p}) = \prod_{i=1}^{5} \prod_{\tau_c \in \boldsymbol{\tau}_c} P(\mathbf{Q}_{\tau_c,i}, \mathbf{S}_{\tau_c,i} \mid \boldsymbol{p}),$$

where $\mathbf{Q}_{\tau_c,i}$ and $\mathbf{S}_{\tau_c,i}$ are the i-th rows of $\mathbf{Q}_{\tau_c}$ and $\mathbf{S}_{\tau_c}$, respectively. For each quantile i and each cooling time span τ_c, the likelihood is given by

$$P(\mathbf{Q}_{\tau_c,i}, \mathbf{S}_{\tau_c,i} \mid \boldsymbol{p}) = \prod_{j=1}^{11} \frac{1}{\sigma_{i,\tau_c}(t_j)\sqrt{2\pi}} \exp\left(-\frac{(Q_{i,\tau_c}(t_j) - \text{rmsf}_{\text{model}}(\boldsymbol{p}))^2}{\sigma^2_{i,\tau_c}(t_j)} \right),$$

where $\text{rmsf}_{\text{model}}$ is taken from equation 5 (Methods), equation 7 (Methods), and equation 8 (Methods) for model1, model2 and model3, respectively.

For the prior distributions of the parameters, we used uniform distributions. For model1 and model3, uniform distributions between 0 MJ/mol/nm^2 and 30 MJ/mol/nm^2 and between between 0 Å and 10 Å were used for parameters c and d, respectively. For model2 and model3, uniform distributions between 0 Å and 10 Å, between 0 kJ/mol and 10 kJ/mol were used for parameters Δx and $\Delta G^\ddagger$. For ΔG (model2 and model3), we used a uniform distribution between 0 kJ/mol and $-k_b T_h \log(1/0.99 - 1)) \approx 10.6$ kJ/mol such that the occupancy of state A is at least 1% at T_h before cooling.

The probability distribution $P(\boldsymbol{p} \mid \boldsymbol{Q}, \boldsymbol{S})$ was sampled using the Metropolis Monte Carlo algorithm[2] with 10^6 steps. To that aim, we used the function $f(\boldsymbol{p}) = P(\boldsymbol{Q}, \boldsymbol{S} \mid \boldsymbol{p}) \cdot P(\boldsymbol{p})$ which is proportional to the probability distribution (compare equation 1).

Initial values of parameters d and Δx were set to 1 Å for model variants which have one value for all quantiles and to 1 Å, 2 Å, 3 Å, 4 Å, and 5 Å for variants which have one value per quantile. Initial values of parameters c were set to 5 MJ/mol/nm^2 and initial values of parameters ΔG and $\Delta G^\ddagger$ were set to 5 kJ/mol.

For each metropolis step, N_p substeps were carried out, where N_p is the number of free parameters of the model variant. In each substep, a new value for one of the parameters was drawn from a normal distribution centered on the current value with a standard deviation of σ_p. For parameter c, σ_p was set to 1 MJ/mol/nm^2. For d, σ_p was set to $2 \cdot 10^{-5}$ Å for variants with one value of d for all quantiles and to $2 \cdot 10^{-5}$ Å, $4 \cdot 10^{-5}$ Å, $6 \cdot 10^{-5}$ Å $8 \cdot 10^{-5}$ Å, and $10 \cdot 10^{-5}$ Å for variants with one value of d for each quantile. For Δx, σ_p was set to $1 \cdot 10^{-4}$ Å for variants with one value of Δx for all quantiles and to $1 \cdot 10^{-4}$ Å, $2 \cdot 10^{-4}$ Å, $3 \cdot 10^{-4}$ Å $4 \cdot 10^{-4}$ Å, and $5 \cdot 10^{-4}$ Å for variants with one value of Δx for each quantile. For model2, σ_p was set to 0.05 kJ/mol and 1 kJ/mol for parameters ΔG and $\Delta G^\ddagger$, respectively. For model3, σ_p was set to 0.2 kJ/mol and 2 kJ/mol for parameters ΔG and $\Delta G^\ddagger$, respectively. In each substep, the function f was evaluated with the new parameter, and the ratio α of the new and previous value of f was used as the acceptance ratio: If $\alpha > 1$, the new parameter was accepted. If $\alpha < 1$, a random number u between 0 and 1 was drawn and the new parameter was accepted if $u \leq \alpha$ and rejected otherwise.

Finding optimal model variants

The three models of the cooling process have different sets of parameters. When we use all model parameters for each quantile separately, model1, model2 and model3 have 10, 15, 25 free parameters, respectively. For different variants of the models, we set different parameters to be the same for all quantiles. To test how well a model variant reproduces and predicts the rmsf values obtained from the T-quench simulations, we used a cross-validation approach. For each model variant, first, we applied Metropolis sampling with Bayesian inference (as described above) where the likelihood function only depend on the rmsf values from the T-quench simulations with cooling time spans between 0.1 ns and 64 ns. Next, after omitting the first 20% of

the Metropolis steps, 1000 steps were randomly chosen, the values of the parameters were extracted and 1000 rmsf curves were calculated from the model for all cooling time spans. To check how well the model variant reproduces the rmsf values used for training, root mean square deviations (rmsds) of model rmsf values from those obtained from the T-quench simulations were calculated for cooling time spans 0.1–64 ns (Fig. S3, blue). To check how well the model variant predicts rmsf values not included in the training, rmsds of model rmsf values from T-quench simulation values were calculated for the cooling time span 128 ns (Fig. S3, red).

From the tested model variants, we chose the one with the lowest cross-validation rmsd for further analysis. For model1 and model2, we tested all possible variants (Fig. S3a,b) and we chose the model1 variant with five parameters for d and one fore c and the model2 variant with five parameters for Δx and $\Delta G^{\ddagger}$ and one for ΔG. For model3, several variants showed similar cross-validation rmsds and among these variants we chose the one with the fewest free parameters, namely five parameters for d and one for c, Δx, $\Delta G^{\ddagger}$, and ΔG.

Comparison of models

To check how well the model recreates and predicts the rmsf values obtained from the T-quench simulations, we trained the model on different sets of cooling time spans (Fig. 3e). Next for each set of cooling time spans, after omitting the first 20% of steps, from 1000 randomly chosen Metropolis steps, the values of the parameters were extracted and 1000 sets of rmsf curves for all cooling time spans τ_c were calculated from the model. Then, rmsds of the rmsf values obtained from the model from those obtained from the T-quench simulations were calculated for all cooling time spans (Fig. 3e). The rmsf curves obtained from the models trained on all cooling time spans are shown in (Fig. 3d, colored lines). Probability densities of the parameters obtained from two independent calculations are shown in Fig. S4a.

Supplementary Figures

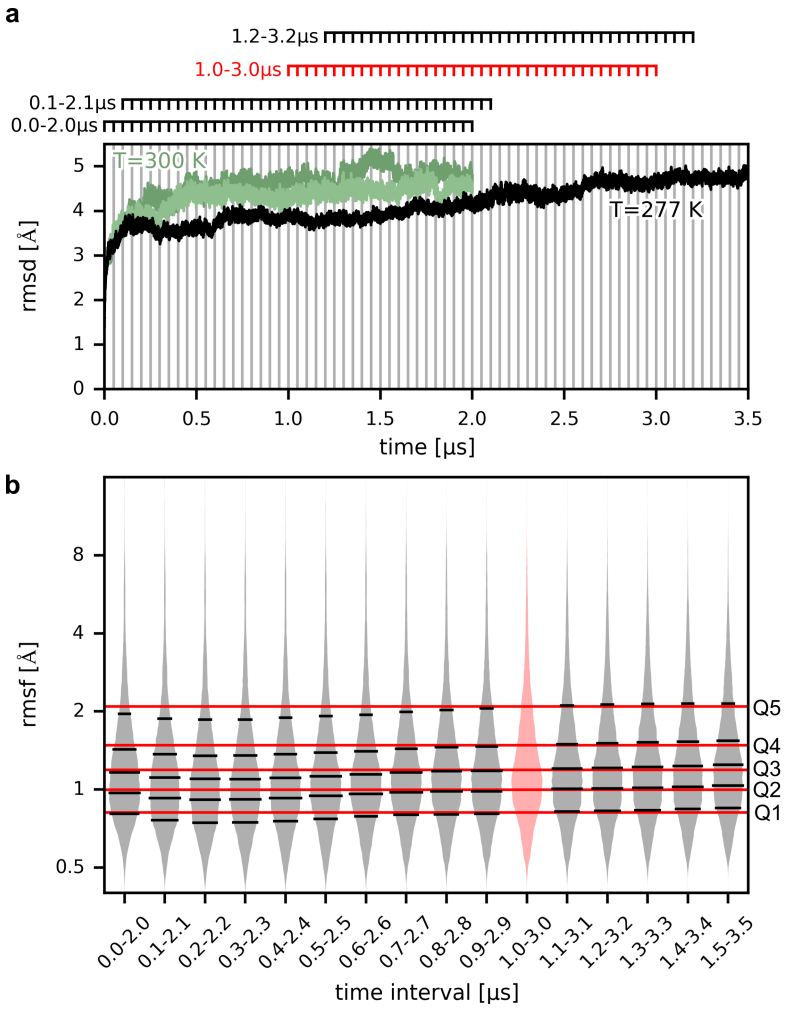

Figure S1: **Equilibration of the ribosome-EF-Tu complex at 277.15 K prior to cooling.** (**a**) Root mean square deviation (rmsd) as a function of simulation time (black line). Rmsd of two simulations of the same system at T=300 K (green lines) described earlier[3]. Time points of extracted snapshots are indicated by vertical lines (every 50 ns). Horizontal lines (top) indicate time intervals of 2 μs length. (**b**) The rmsf values of the ribosome-EF-Tu complex for ensembles of 41 structures obtained from different time intervals. Histograms of the rmsf values are shown as areas (grey, light red). The highlighted ensemble (red) was used for T-quench simulations. Horizontal lines (black, red) indicate the positions of the 6-quantiles (Q1–Q5) of each rmsf distribution.

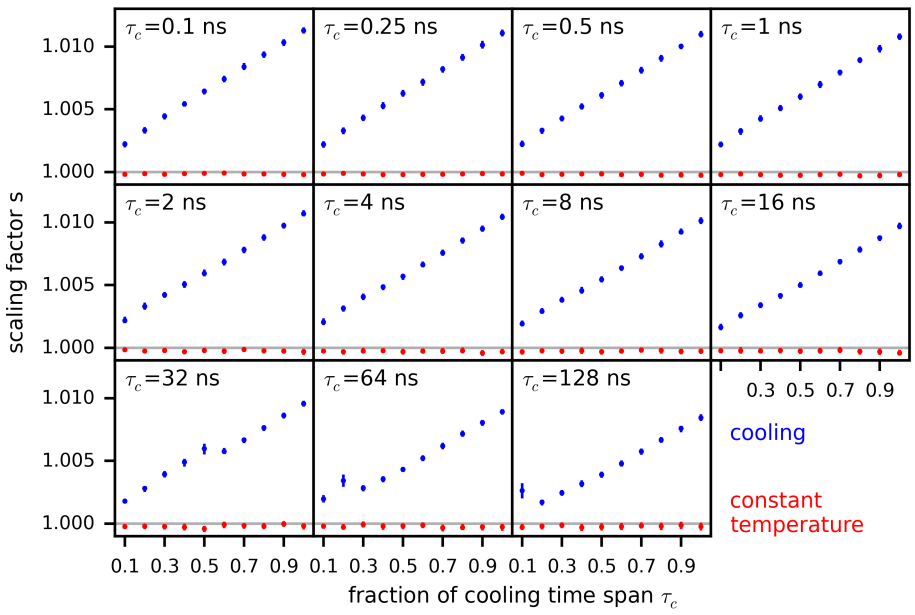

Figure S2: **Volume reduction during cooling.** For each cooling time span and each time point of the cooling trajectories, the scaling factor s which minimized the average rmsd between the scaled conformations and the starting conformations of the respective simulations is shown (mean: blue circles, standard deviation: blue vertical lines). As a control, the same analysis was carried out for conformations extracted from the 277.15-K simulation (red circles and lines).

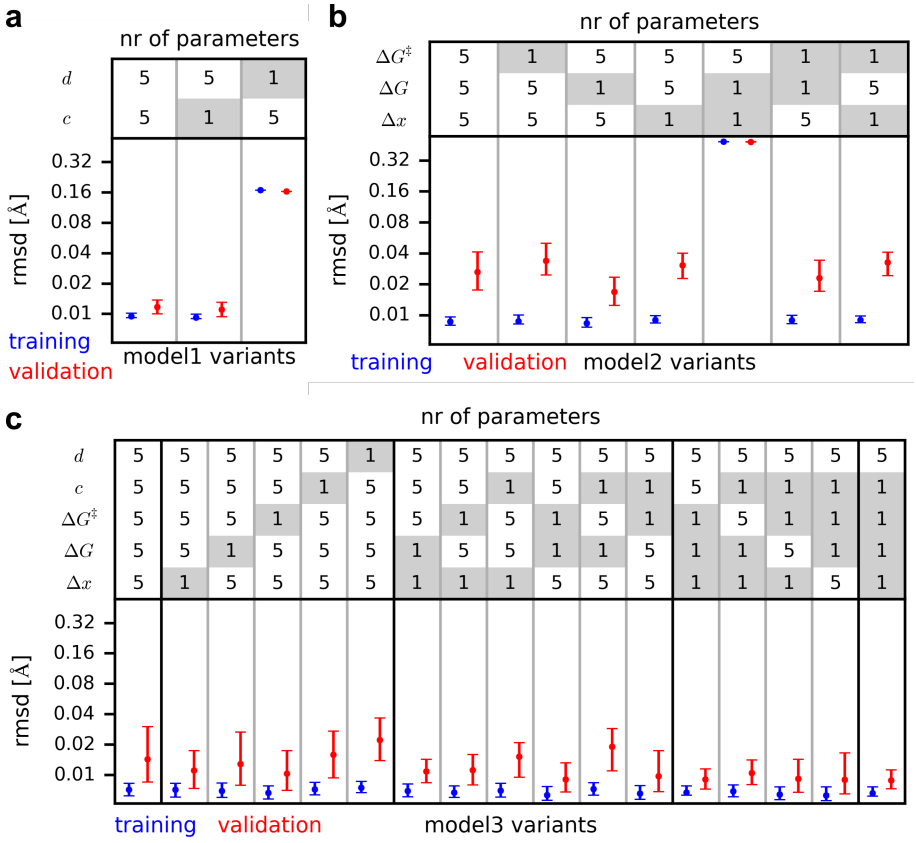

Figure S3: **Model parameter optimization.** Different variants of model1 (**a**), model2 (**b**), and model3 (**c**) with different numbers of the parameters Δx, ΔG, $\Delta G^{\ddagger}$, c, and d were trained on the rmsf values for cooling time spans between 0.1 ns and 64 ns (Fig. 2c, black lines). If the number of parameters is 5, there is one parameter for each quantile. If the number is 1, there is one parameter for all quantiles. The rmsds between the rmsf values calculated from the model and the training set are shown in blue (bars indicate the 95 % confidence interval). For validation, the rmsds between the rmsf values predicted by the model for cooling time span 128 ns and the rmsf values from the respective MD simulations are shown in red.

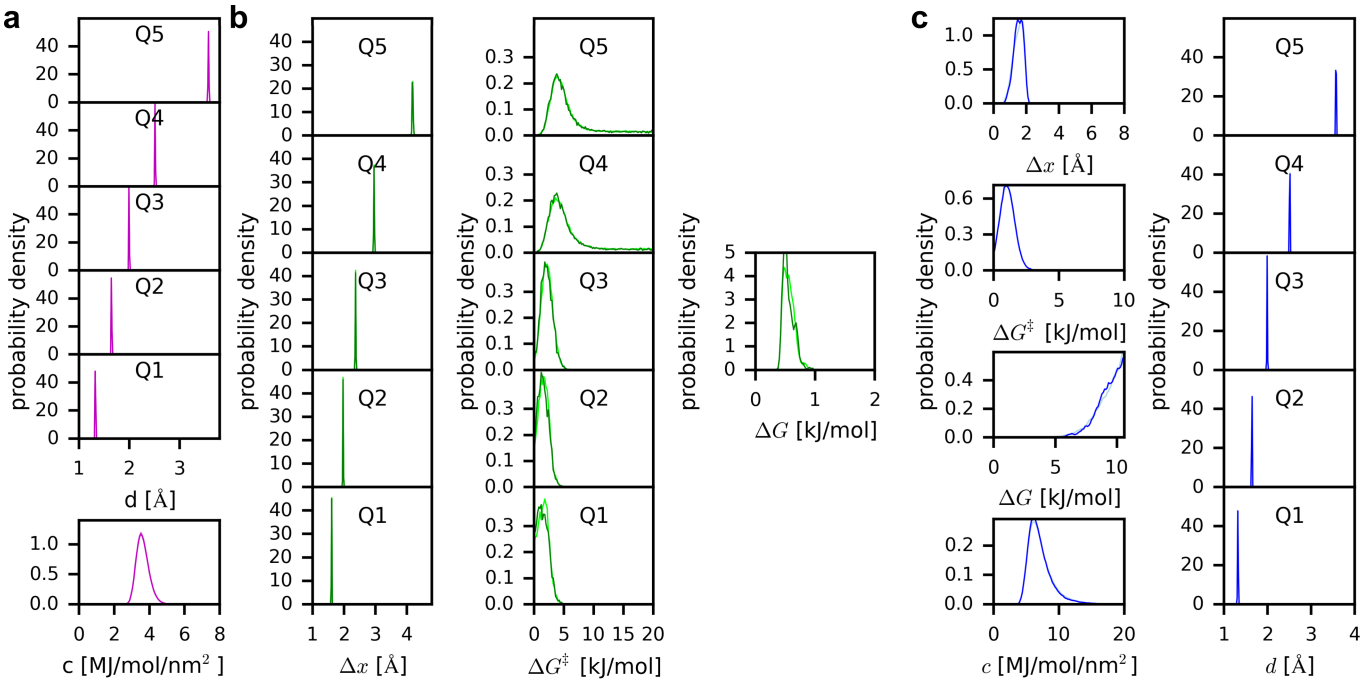

Figure S4: **Parameters obtained for thermodynamic and kinetic models.** Probability densities of the model parameters obtained via Bayesian statistics from two independent calculations for model1 (**a**), for model2 (**b**), and for model3 (**c**). The two shades of magenta, green, and blue denote two independent calculations.

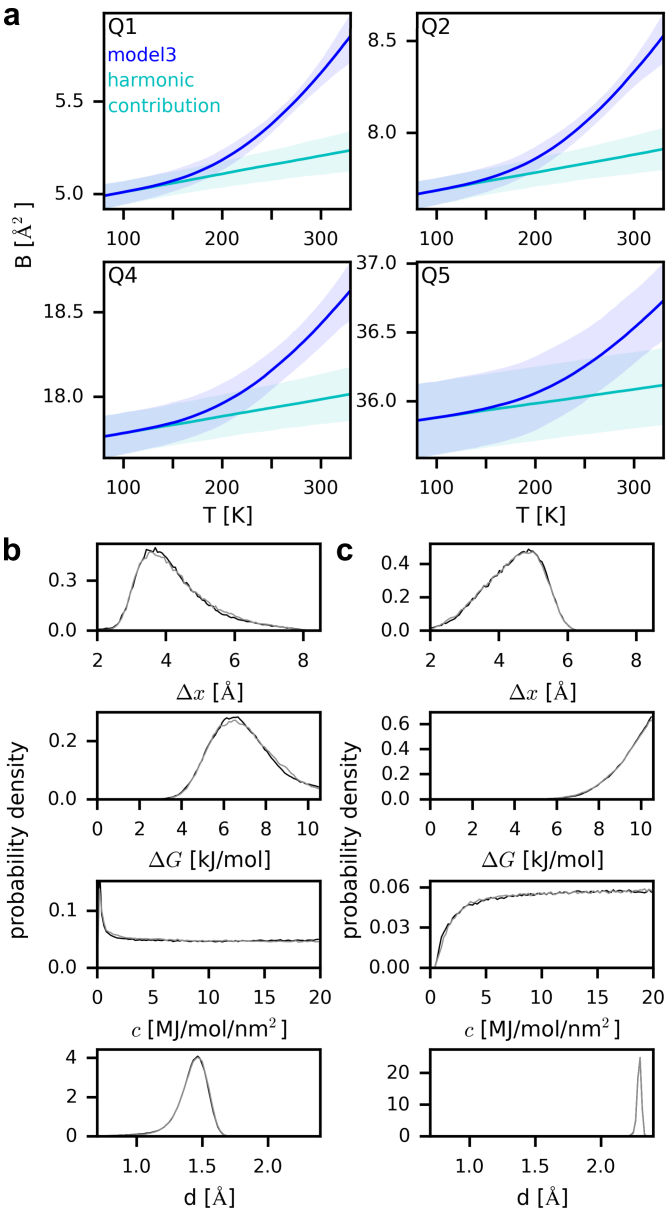

Figure S5: **B-factor dependency on temperature and model3 parameters.** (**a**) B-factors as a function of temperature calculated from model3 applied to ribosome T-quench simulations under equilibrium conditions (upper panel; expected values: blue lines, 95 % confidence interval: light blue area). The B-factors of the harmonic potentials are shown in cyan. (**b**,**c**) Probability densities of the parameters for the equilibrium model3 applied to experimental data. The B-factors obtained from x-ray crystallography of thaumatin [4] (b) and ribonuclease-A [5] (c) at different temperatures were used to train equilibrium model3.

References

[1] Crank, J. & Nicolson, P. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Mathematical Proceedings of the Cambridge Philosophical Society* **43**, 50–67 (1947).

[2] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092 (1953).

[3] Warias, M., Grubmüller, H. & Bock, L. V. tRNA Dissociation from EF-Tu after GTP Hydrolysis: Primary Steps and Antibiotic Inhibition. *Biophysical Journal* **118**, 151–161 (2020).

[4] Warkentin, M. & Thorne, R. E. Slow cooling of protein crystals. *Journal of Applied Crystallography* **42**, 944–952 (2009).

[5] Tilton, R. F., Dewan, J. C. & Petsko, G. A. Effects of Temperature on Protein Structure and Dynamics: X-ray Crystallographic Studies of the Protein Ribonuclease-A at Nine Different Temperatures from 98 to 320 K. *Biochemistry* **31**, 2469–2481 (1992).