Supporting Information for

A Growth Type Pathway for Improving the Folding of BPTI

Yingsong Wang, Shweta Mandumula, Watson J. Lees

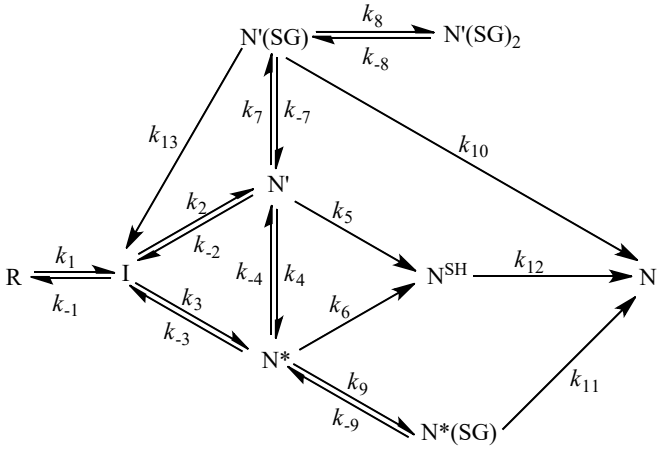

Fig. S1. Kinetic scheme with all rate constants used for modelling of BPTI folding. R, I, and N stand for reduced BPTI, BPTI with one disulfide or mixed disulfides, and native BPTI, respectively.

Description of calculations using the folding model.

The reverse kinetic problem was solved by fitting rate constants of all the reaction steps involved to reproduce experimental concentration profiles using Euler's method (15) (Δt = 0.1 s for reactions less than 6 h and 1 s for reactions more than 6 h). We used similar methodology to that described in the supporting information of reference 15. An excel spreadsheet was set up with the following columns: time, R, I, N', N*, N, N'(SG), N'$(SG)_2$, N^{SH}, and N*(SG), GSSG and GSH. R is reduced protein, I is considered as the concentration of intermediates containing one disulfide bonds or other unspecified mixed disulfides. The initial concentrations of all species at t=0 were then entered. In some cases, the experimentally determined concentrations at t=15 minutes were used as a

starting point. The rate constants k_1, k_{-1}, k_2, k_{-2}, k_{-3}, k_3, k_4, k_{-4}, k_5, k_6, k_7, k_{-7}, k_8, k_{-8}, k_9, k_{-9}, k_{10}, k_{11}, k_{12}, and k_{13} were then added to the file along with initial concentrations of the protein species and GSSG and GSH or the experimentally determined concentrations of the protein species at 15 min. If the Δt was 0.1 s, see above, the concentrations of all species were then calculated at 0.1 s based on the kinetic scheme, the rate constants, the concentration of species at t=0, and the time jump (0.1 s). For example, the concentration of $N'(SG)_2$ at 0.1 seconds would be

$$[N'(SG)_2]_{0.1} = [N'(SG)_2]_0 - k_{-8} * [N'(SG)_2]_0 * [GSH]_0 * 0.1 + k_8 * [N'(SG)]_0 * [GSSG]_0 * 0.1$$

where $[N'(SG)_2]_{0.1}$ is the calculated concentration of $[N'(SG)_2]$ at 0.1 seconds, $[N'(SG)_2]_0$ is the concentration of $[N'(SG)_2]$ at 0 seconds, k_{-8} is the rate constant for the reaction of $[N'(SG)_2]$ with GSH, see scheme above, $[GSH]_0$ is the concentration of GSH at 0 seconds, 0.1 is the time jump in seconds, k_8 is the rate constant for the reaction of [N'(SG)] with GSSG, $[N'(SG)]_0$ is the concentration of [N'(SG)] at 0 seconds, $[GSSG]_0$ is the concentration of GSSG at 0 seconds and 0.1 is the time jump in seconds. The process was repeated for all species at 0.2 seconds, then 0.3 seconds etc.:

$$[N'(SG)_2]_{0.2} = [N'(SG)_2]_{0.1} - k_{-8} * [N'(SG)_2]_{0.1} * [GSH]_{0.1} * 0.1 + k_8 * [N'(SG)]_{0.1} * [GSSG]_{0.1} * 0.1$$

$$[N'(SG)_2]_{0.3} = [N'(SG)_2]_{0.2} - k_{-8} * [N'(SG)_2]_{0.2} * [GSH]_{0.2} * 0.1 + k_8 * [N'(SG)]_{0.2} * [GSSG]_{0.2} * 0.1$$

.

.

.

.

The results predicted from these calculations were then used to predict the folding yields or to compared with the experimentally obtained concentration values.

For folding of reduced protein

For predicting the folding yields at 12 h from reduced protein a time step of 1 second was used to limit the file size. As the rate constants determined herein were for the later stages of folding, the experimentally obtained 15 min folding time point was used as a starting point in the model, see reference 15. The concentration of native protein at 12 h was then calculated. To maximize the production of native protein in 12 h the concentrations of GSSG and GSH as a function of time were varied. The concentrations of GSSG and GSH were limited to 50 mM for practical reasons. The highest percent of native protein at 12 h was calculated to occur when 35 mM GSH was added at 15 min and 50 mM GSSG was added at 1 h.

Determining rate constants

For a certain folding reaction, only the relevant rate constants involved in the reaction steps were used. The following reactions were used to determine the rate constants: rearrangement of N' (k_4 and k_5), reaction of N'$(SG)_2$ with GSH (k_{-8}), reaction of N* with high GSSG concentration with and without GSH (k_{-9} and k_{11}), and reaction of N'(SG) with GSH (k_{-7} and k_{13}). Rate constants k_{-2} (disappearance of N'), k_{-3} (disappearance of N*), k_7 (disappearance of N'), k_8 (disappearance of N'(SG)), k_9 (disappearance of N*), and k_{10} (formation of N) were determined using pseudo first order kinetics.

The model shown above in Figure S1 was simplified to determine specific reaction rates. For example, for N* reacting with GSSG at 50 mM, the concentrations of N*, N*(SG), N and GSSG and the rate constants k_9 and k_{11} were used in the model. In the model at t=0, the

concentration of N* was set to 30 micromolar, the concentration of GSSG to 50 mM and the rest set to 0. The experimentally obtained values were then compared to the calculated values and the sum of the differences squared calculated. The program Solver then varied k_9 and k_{11} to minimize the sum of the differences squared between the calculated and experimental data. The resulting values of k_9 and k_{11} were then reported, see Figure 2E.