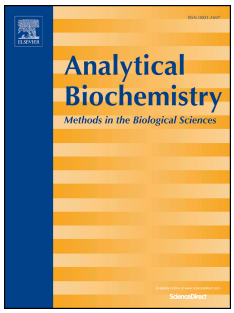
Accepted Manuscript



A new look at an old view of denaturant induced protein unfolding

Damien Hall, Akira Kinjo, Yuji Goto

PII: S0003-2697(17)30458-X

DOI: 10.1016/j.ab.2017.11.011

Reference: YABIO 12842

To appear in: *Analytical Biochemistry*

Received Date: 27 July 2017

Revised Date: 11 November 2017

Accepted Date: 16 November 2017

Please cite this article as: D. Hall, A. Kinjo, Y. Goto, A new look at an old view of denaturant induced protein unfolding, *Analytical Biochemistry* (2017), doi: 10.1016/j.ab.2017.11.011.

A new look at an old view of denaturant induced protein unfolding.

Damien Hall[#,+,*], Akira Kinjo[#] and Yuji Goto[#]

[#]*Institute for Protein Research. Osaka University, 3-1- Yamada-oka, Suita, Osaka 565-0871, Japan.* [+] *Research School of Chemistry, Australian National University. Acton ACT 2601, Australia.*

[*]Correspondence to: damien.hall@anu.edu.au; damien.hall@protein.osaka-u.ac.jp

Abstract: We re-examine a site-binding approach independently proposed by Schellman (Schellman, J.A. (1958) Compt. rend. Lab. Carlsberg Ser. Chim. 30, 439-449) and Aune and Tanford (Aune, K.C. and Tanford, D. (1969) Biochemistry, 8, 4586 - 4590) for explicitly including the denaturant concentration within the protein unfolding equilibrium. We extend and formalize the approach through development of a multi-dimensional analytical model in which the folding reaction coordinate is defined by the number of denaturant molecules bound to sites located on either the initially folded, or unfolded, states of the protein. We use the developed method to re-examine the mechanistic determinants underlying the sigmoidal shape of the unfolding transition. A natural feature of our method is that it presents a landscape picture of the denaturant induced protein unfolding reaction.

Characterizing the stability of a protein in response to an unfolding stress is important in a variety of diverse fields ranging from manufacturing of enzymes used in detergent [**1**], preventing spoilage in food processing [**2**], optimizing storage conditions for protein pharmaceuticals [**3**] and understanding the causes of protein misfolding diseases [**4**]. In general, proteins can be denatured by either changing the temperature (**5-9**), the pressure (**9-12**) or by addition of increasing concentrations of chemical denaturants (**9, 13**) such as urea or guanidine hydrochloride (**13-24**), various alcohols (**25-27**) and hydrogen/hydroxide ions (pH induced unfolding (**28-30**)). Whilst the thermodynamics of thermally induced unfolding is relatively conceptually straightforward (**6-9, 31-35**), denaturant based unfolding is somewhat more difficult to describe due to the diversity of denaturants and their different modes of action (**9, 17, 18, 35-40**). A key juncture in the history of denaturant induced protein unfolding was the quantitative functionalization of the equilibrium constant, K_{FU}, in terms of the denaturant concentration, C_D i.e. $K_{FU}\{C_D\}$. At an early stage in the development of the field, two alternative conceptual treatments for achieving this analytical functionalization competed for popular usage [**9, 17, 36**]. The m-value approach by Greene and Pace [**20**], which eventually achieved primacy, implicitly included the effects of denaturant within K_{FU} via a linear perturbation treatment of the free energy governing the unfolding equilibrium [**20, 41**] (see Appendix 1). A different approach that explicitly included the denaturant as a species in the unfolding reaction, was developed independently by Aune and Tanford [**15, 28**] and Schellman [**16, 36**][1]. Although the explicit approach has faded in popularity, it potentially possesses a more straightforward linkage to mechanistic reality than the m-value approach due to its formulation in terms of measurable species [**15, 16, 28, 36, 43, 44**]. In the next section we first present the original development, closely following Aune and Tanford's approach. We then extend their formalism through development of a multi-dimensional analytical model in which the unfolding reaction coordinate is defined in terms of the number of denaturant molecules bound to sites located on the initially folded or unfolded state. This model is then used to provide a fresh look at some basic aspects of the protein unfolding transition.

Approach of Schellman, Aune and Tanford

Both Aune and Tanford [**15, 28**] and Schellman [**16, 36**] considered folded, F, and unfolded, U, forms of the protein as respectively possessing N_{FD} and N_{UD} specific independent and equivalent sites to which denaturant could bind, with respective site binding constants of k_{FD} and k_{UD} (Eqn. 1).

$$F_{site} + D \overset{k_{FD}}{\rightleftarrows} FD \qquad [1a]$$

$$U_{site} + D \overset{k_{UD}}{\rightleftarrows} UD \qquad [1b]$$

[1] Interestingly the same equation was also provided by Brandts in 1964 but without any preceding derivation [**42**].

With this conceptualization, an operational unfolding equilibrium constant was defined based on an assumed stepwise equilibrium (Eqn. 2).

$$[..\rightleftharpoons FD_2 \rightleftharpoons FD_1 \rightleftharpoons F] \rightleftharpoons [U \rightleftharpoons UD_1 \rightleftharpoons UD_2 \rightleftharpoons..] \quad [2a]$$

$$K_{F\rightarrow U}\{C_D\} = \frac{\sum_{j-0}^{N_{UD}} C_{UD_j}}{\sum_{i-0}^{N_{FD}} C_{FD_i}} \quad [2b]$$

All possible denaturant bound forms of folded and unfolded protein were accounted for via binding polynomials calculated using the binomial theorem (Eqn. 3a) [**15, 41**]. This description was further simplified (to Eqn. 3b) by identification of the equilibrium relation $C_U = K_{F\rightarrow U}\{0\}C_F$ and invocation of the assumption that the site binding constant operating between the denaturant, the folded and unfolded forms of protein, was equal to a common value k_{PD} , i.e., $k_{PD} = k_{FD} = k_{UD}$.

$$K_{F\rightarrow U}\{C_D\} = \frac{C_U[1+k_{UD}C_D]^{N_{UD}}}{C_F[1+k_{FD}C_D]^{N_{FD}}} \quad [3a]$$

$$K_{F\rightarrow U}\{C_D\} = K_{FU}\{0\}[1+k_{PD}C_D]^{(N_{UD}-N_{FD})} \quad [3b]$$

To apply their formulation to protein unfolding data Aune and Tanford used the functionalization of the unfolding constant $K_{F\rightarrow U}\{C_D\}$ (shown by Eqn. 3b) to directly fit the total fraction of protein existing in the unfolded form, f_U, as a function of the denaturant concentration, C_D (Eqn. 4) [**15**].

$$f_U = \frac{\sum_{j=0}^{N_{UD}} C_{UD_j}}{\sum_{i=0}^{N_{FD}} C_{ND_i} + \sum_{j=0}^{N_{UD}} C_{UD_j}} = \frac{K_{F\rightarrow U}\{C_D\}}{(1+K_{F\rightarrow U}\{C_D\})} \quad [4a]$$

$$f_U = \frac{K_{F\rightarrow U}\{0\}[1+k_{PD}C_D]^{\Delta N}}{(1+K_{F\rightarrow U}\{0\}[1+k_{PD}C_D]^{\Delta N})} \quad [4b]$$

Taking account of the assumptions inherent within its derivation, Eqn. 4b provides a thermodynamically rigorous method for interrogating the dependence of the fraction of unfolded protein as a function of the denaturant concentration. We note that Eqn. 4b is based on just three[2] parameters, $K_{F\rightarrow U}\{0\}$, k_{PD} and ΔN.

Analog of the Aune-Tanford-Schellman Approach

Our re-visitation of the Aune and Tanford [**15**] amd Schellman [**16**] approach begins with a more formal description of the folding/unfolding transition (Fig. 1). A specific number, N_{FD}, of denaturant binding sites are considered to exist on the exterior of the folded protein, their interaction with denaturant regulated by a site binding constant, k_{1F}. The same number of N_{FD} sites also exist on the unfolded protein but are considered able to bind denaturant with a potentially different site binding constant, k_{1U} (Fig. 1b). Previously hidden regions of the protein interior are accessible to solvent upon protein unfolding, with these newly exposed regions presenting a specific number, N_{UD}, of binding sites for denaturant. Each interior site is considered to possess a common denaturant site binding constant, k_{2U}. This basic description can be more concisely expressed in two-dimensional matrix form (Fig. 1b) with each row /column entry defined by either an F or a U, with a two index subscript coordinate indicating the number of denaturant molecules bound to the initially folded or initially unfolded sites. In such an indexing scheme, $F_{i,0}$ represents a folded species with i denaturant molecules bound to the folded denaturant sites, whereas $U_{i,j}$ describes an unfolded species with i and j denaturant molecules respectively bound to the initially folded and unfolded sites. The final class of parameters described within the model are the denaturant loading dependent unfolding constants $(k_{FU})_{i,0}$ which describe the propensity of an $F_{i,0}$ folded species to transition to its $U_{i,0}$ state. As will be subsequently demonstrated, we need only specify the value for the unfolding constant at zero denaturant concentration, $(k_{FU})_{0,0}$, in order to calculate all other $(k_{FU})_{i,0}$ on the basis of known k_{1F}, and k_{1U}.

Combinatorial determination of the number of ways, W(X,Y), of arranging Y objects within X available positions (Eqn. 5a) [**45, 46**] was used to calculate appropriate statistical factors for modification of the site binding constant, k, governing the transition between two elements within the matrix (Eqns. 5b - e) [**46**][3].

[2] Upon replacing the difference between the two parameters, N_{UD} and N_{FD} with the single parameter ΔN.

[3] Note that in the case of the folding/unfolding equilibrium not involving denaturant, no statistical correction for denaturant binding is involved such that $K_{F0,0\rightarrow U0,0} = (k_{FU})_{0,0}$.

$$W(X,Y) = \frac{X!}{(X-Y)!Y!} \quad [5a]$$

$$K_{F_{0,0}\rightarrow U_{0,0}} = (k_{FU})_{0,0} \quad [5b]$$

$$K_{F_{i,0}\rightarrow F_{i+1,0}} = \left(\frac{W(N_{FD},i+1)}{W(N_{FD},i)}\right).k_{1F} \quad [5c]$$

$$K_{U_{i,j}\rightarrow U_{i+1,j}} = \left(\frac{W(N_{UD},i+1)}{W(N_{UD},i)}\right).k_{1U} \quad [5d]$$

$$K_{U_{i,j}\rightarrow U_{i,j+1}} = \left(\frac{W(N_{UD},j+1)}{W(N_{UD},j)}\right).k_{2U} \quad [5e]$$

With the equilibrium constants for each possible transition specified [4] the concentration of all species in the matrix can be calculated using equilibrium relations (Eqn. 6a-d) based on the concentration of the denaturant free folded state of the protein, $C_{F_{0,0}}$, and the free concentration[5] of denaturant at equilibrium, C_D.

$$C_{F_{m,0}} = C_{F_{0,0}}.\prod_{i=1}^{i=m}\left\{\left(\frac{W(N_{FD},i)}{W(N_{FD},i-1)}\right)k_{1F}C_D\right\} \quad [6a]$$

$$C_{U_{0,0}} = (k_{FU})_{0,0}.C_{F_{0,0}} \quad [6b]$$

$$C_{U_{0,n}} = (k_{FU})_{0,0}C_{F_{0,0}}\prod_{j=1}^{j=n}\left\{\left(\frac{W(N_{UD},j)}{W(N_{UD},j-1)}\right)k_{2U}C_D\right\} \quad [6c]$$

$$C_{U_{m,n}} = C_{U_{0,n}}.\prod_{i=1}^{i=m}\left\{\left(\frac{W(N_{FD},i)}{W(N_{FD},i-1)}\right)k_{1U}C_D\right\} \quad [6d]$$

For a given value of C_D, equation set 6 can be solved iteratively through incremental changes in $C_{F0,0}$, for a fixed value of the total protein concentration $(C_P)_{TOT}$, until an acceptable minimum of Eqn. 7 is achieved.

$$\min\left[(C_P)_{TOT} - \left(\sum_{i=0}^{N_{FD}} C_{F_{i,0}} + \sum_{i=0}^{i=N_{FD}}\sum_{j=0}^{j=N_{UD}} C_{U_{i,j}}\right)\right] \quad [7]$$

With each species so defined, we are able to calculate the effective value of the unfolding equilibrium constant as a function of denaturant loading according to Eqn. 8a. Use of Eqn. 8a allows for identification of the dominant equilibrium pathway of unfolding via Eqn. 8b.

$$K_{F_{m,0}\rightarrow U_{m,0}} = (k_{FU})_{0,0}.\left(\frac{k_{1U}}{k_{1F}}\right)^m \quad [8a]$$

$$C_{U_{m,0}} = K_{F_{m,0}\rightarrow U_{m,o}}C_{F_{0,0}}\prod_{i=1}^{i=m}\left\{\left(\frac{W(N_{FD},i)}{W(N_{FD},i-1)}\right)k_{1F}C_D\right\} \quad [8b]$$

Equations 5-8 represent the foundation of our multi- dimensional reformulation of the Aune and Tanford/ Schellman model with explicit inclusion of the denaturant species within the unfolding equilibrium reaction. The difference between the original model (Eqn. 4) and our multi-dimensional analog (Eqns. 5-8) can be best appreciated by examining the types of output afforded by each approach (Fig. 2). Fig. 2a shows the traditional description of a denaturant induced unfolding transition in terms of the fraction unfolded, f_U, versus the denaturant concentration, C_D, for both the Aune and Tanford/ Schellman model equation (black line - Eqn. 4) and our modern multi-dimensional analog (red circles - Eqn. 5-8). Fig. 2b describes the additional data afforded by the multi-dimensional model in terms of logarithmic occupation density, $\log_{10}(C_{P_{i,j}})$ where each P species (either F or U) is defined in terms of the denaturant loading of the folded or unfolded denaturant sites, $P_{i,j}$. Red dots represent the apex of each distribution for the given denaturant concentrations shown by the red circles in Fig. 2a. Fig. 2c represents the same information as Fig. 2b shown in a multi-dimensional heat map format – the form we employ in the presentation of results from hereon. Fig. 2d shows, in isolation, the distribution of protein states corresponding to a single denaturant concentration, the C_D equaling 8M point (labeled in Fig. 2A using a blue ring). Having introduced the characteristic outputs of the model we now use it to revisit the determinants of perhaps the most basic feature of the two-state protein folding reaction – its apparent cooperativity manifested as a sigmoidal form of the transition.

Results and Discussion

The multi-dimensional reformulated Aune and Tanford/ Schellman model has six parameters K_{FU}, k_{1F}, k_{1U}, k_{2U}, N_{FD} and N_{UD}. For cases involving proteins of known structure (or structural class) we can reduce the parameter space from six to five[6] by defining N_{FD} in terms of the folded structure and N_{UD}, on the basis of the number of amino acids in

[4] The interested reader may have noticed that we have not specified the unfolding equilibrium constants $K_{F_{i,o}\rightarrow U_{i,o}}$ for the case of i >0. We will address this point in the next section.

[5] This latter term can nearly always be substituted for the total concentration of denaturant within the limits of the approximation, $[N_{FD}+N_{UD}](C_P)_{TOT} << (C_D)_{TOT}$.

[6] On the condition of known mode of denaturant binding – otherwise the parameter number is again six.

the protein, N_{AA} (the basis of this assumption is derived in Appendix 2). Of this reduced model space we explore changes in each of the variable parameters in turn, examining five cases in all. For each case we compare the base simulation presented in Fig. 2 against two alternatives simulated using either a larger or smaller value of the parameter under examination. For the unit-less intrinsic stability partition constant, $(k_{FU})_{0,0}$, we simulate change over a million-fold range. For N_{AA} we examine a 2.25 fold range, whereas for the three association constants, k_{1F}, k_{1U} and k_{2U} (having units of M^{-1}) we assess the effect of a 25 fold range of values. To begin, we first study the effect of changes in the value of the intrinsic protein stability, $(k_{FU})_{0,0}$, i.e., the equilibrium unfolding constant operating at zero denaturant concentration.

Case 1: Change in $(k_{FU})_{0,0}$

The blue, red and green lines shown in Fig. 3A span a million-fold increase in the intrinsic stability as manifested within the $(k_{FU})_{0,0}$ parameter (with lower values of $(k_{FU})_{0,0}$ reflecting greater stability). Changes in $(k_{FU})_{0,0}$ tend to displace the transition curve rather than modify its general slope characteristics (apparent cooperativity). From the trajectory plots shown in Figs. 3B-D we note that although the final state constitution (f_U ~ 1.0) of the three simulations featuring different intrinsic stability is much the same, the early stages of unfolding, i.e., $f_U = 0.1$, are significantly different in the sense that these unfolded states exhibit different numbers of bound denaturant molecules. This finding can be rationalized by the fact that the drive to populate denaturant sites on the protein is regulated via mass action principles such that lower concentrations of denaturant will lead to lower occupancy, with the converse case true for higher concentrations of denaturant. Lower intrinsic stabilities (higher values of $(k_{FU})_{0,0}$) facilitate exposure of the U state denaturant sites to smaller concentrations of denaturant, thereby allowing U states with less than 10 bound denaturant molecules to exist.

Case 2: Change in N_{AA}

Figure 4 shows the effect of either a 1.5 fold decrease, or a 1.5 fold increase, in N_{AA} (the number of amino acids in the protein chain) on the unfolding transition of a globular protein. Fig. 4A shows the change in stability with the change in polypeptide length over the range $N_{AA} \in [100, 150, 225]$. The total number of denaturant binding sites, which change in the manner prescribed by Eqns. A2b and d, is listed in the figure legend. What seems initially to be a paradoxical result i.e., increasing N_{AA} results in the destabilization of the protein, can be explained on the basis of the following two points relating to factors which either promote or retard protein unfolding,

(i) In the model, the driving forces promoting unfolding with increasing N_{AA}, derive jointly from both the opportunity for lowering the system potential energy (gained as a consequence of forming a greater number of productive bonds between the denaturant and the internal binding sites) and the gain in system entropy (obtained from the greater number of possible ways of distributing bound denaturant molecules within a larger number of available sites). For the globular geometry assumed in the present case (see Appendix 2) the number of internal denaturant sites is approximately linearly dependent on N_{AA}, whereas the number of external solvent exposed sites shows a $(N_{AA})^{2/3}$ dependence (c.f. Eqn. A2a and Eqn. A2c). As such both of these two 'forces' driving unfolding will, for globular proteins at least, increase with longer polypeptide chain length.

(ii) In the model, the features helping to prevent protein unfolding stem jointly from favorable denaturant interaction with the initially solvent exposed binding sites, i.e. $(k_{1F}, N_{FD}) > 0$, and the intrinsic stability of the protein at zero denaturant concentration, i.e. $(k_{FU})_{0,0} < 1$. As discussed in point (i) above, the relationship between N_{FD} and N_{UD} (determined by the compact shape adopted by globular proteins ($N_{UD} > N_{FD}$)) will produce a situation whereby increasing denaturant concentration ultimately promotes unfolding [7]. With regard to the effect of changes in N_{AA} on intrinsic protein stability, we note that the intrinsic stability is generally thought to reflect a balance between a combination of differential intra- and inter-molecular interaction potential energies

[7] Interestingly, in a manner that is dependent upon the relative values of N_{FD}, N_{UD}, k_{1F}, k_{1U} and k_{2U}, a case exists such that small concentrations of denaturant, below the T_D, will stabilize the protein against other types of denaturing pressure, such as temperature induced unfolding. Such denaturant induced stabilization behavior has been observed experimentally by Goto et al. (**47, 48**) and others (**49, 50**). As the relationship between N_{FD} and N_{UD} is also determined by protein shape there is also significant potential for antithetical denaturant action for proteins featuring relatively exotic, non-globular structure (**51, 52**). We will revisit this situation in a future publication.

between the folded and unfolded states (i.e. solvent-solvent, solvent-protein and amino acid-amino acid within the polypeptide chain), along with differential entropies reflecting the degeneracy/multitude of states for which a particular potential energy state can exist (i.e. number of states reflecting equivalent modes of solvent binding to the folded and unfolded states of the protein along with the number of iso-energetic configurational states in which the folded and unfolded protein states can separately exist - **53**). An equation for intrinsic stability as a function of amino acid number has been previously given by Dill (**35, 53**) and is included here in slightly modified form as Eqn. 9 (note this is discussed in more detail in Appendix 3).

$$(k_{FU})_{0,0} = e^{-\left(\frac{N_{AA}[\delta\mu_{AA,tr}(C_D)]-kTlnZ}{kT}\right)} \quad [9a]$$

$$Z \leq h^{(2N_{AA}-1)} \quad [9b]$$

Here $\delta\mu_{AA,tr}(C_D)$ is the average change in free energy (per mole of amino acid) upon partition of the amino acid from a hydrophobic phase (reflective of the interior of a protein) to an aqueous solution phase (containing a particular concentration of denaturant) and Z is the degeneracy of configurational states for a polypeptide given h different allowable orientations per amino acid ϕ and ψ angle in the polypeptide chain. Z is written as an inequality as it must be less than the equivalent value for a freely rotating chain composed of N_{AA} monomers (**35, 53**)[8]. Dill has suggested that increasing N_{AA} over the size range of 50 to 300 amino acids tends towards greater intrinsic stability (smaller $(k_{FU})_{0,0}$) (**53**)[9].

In order to isolate the effects of particular parameters, in the present simulation, only the size has been altered, with no inclusion of a compensatory change in the intrinsic stability parameter, $(k_{FU})_{0,0}$. As such the observation of destabilization accompanying increases in N_{AA}, can be taken as providing a window into the extent of factors promoting protein destabilization only, whereas in the majority of experimental cases, increases in N_{AA} will tend to stabilize proteins via concomitant changes in the $(k_{FU})_{0,0}$ parameter. In this light Figs. 4B-D show that the greater occupation of unfolded (but not folded) denaturant binding sites drives an apparent decrease in stability with increasing protein size.

Case 3: Change in k_{1F}

In Fig. 5A we note a rather modest change in protein stability [as judged by change in the $(C_D)_{MP}$], along with the negligible change to the apparent cooperativity (as judged by change in the steepness of the transition) when k_{1F}, the site binding constant for denaturant binding to solvent accessible regions in the folded state, is changed over a 25 fold range, in isolation from the other variable parameters. As is directly reconcilable from mass action principles, increases in k_{1F} tend to drag the equilibrium towards the folded state, thereby leading to denaturant induced stabilization (**47-50** – see also Footnote 4). In the present case, due to the greater number of binding sites in the unfolded state and the much greater multiplicity of states in which these U site denaturant molecules may be complexed, the protein still unfolds at some limiting denaturant concentration. This however, need not necessarily always be true and could be taken as marking the boundary of transition from destabilizing denaturant to stabilizing ligand (**43, 47-50**). One interesting point not seen in the other case examples is that as k_{1U} becomes larger we see a change in the F → U transition pathway (Fig. 6B-D) such that the transition proceeds via a more denaturant loaded F state. As for the initial case example considering changes in intrinsic stability, the unfolded denaturant sites of proteins with decreased stability (corresponding to lower k_{1F} values) exhibit lower extents of denaturant occupancy at fractional extents of unfolding (~f_U < 0.1). This behavior makes the unfolding path trajectories significantly different, from a chemical species perspective, for the different cases of k_{1F} (Figs. 5B-D).

Case 4: Changes in k_{1U}

A 25-fold change in k_{1U}, the association equilibrium constant governing binding of denaturant to the folded sites in the unfolded U state, produces significant changes in both the stability and the apparent cooperativity (Fig. 6A). It also produces an interesting effect not seen for the other cases in that increases in k_{1U} dramatically change both the extent of denaturant binding as

[8] A lower estimate of $Z \approx h^{(2N_{AA}-1)}.\exp[-(2N_{AA}-1)]$ has been suggested by Dill (**53**) for the representation of a polypeptide chain by an excluded volume chain with h energetic minima per bond rotation.

[9] Up to a certain size limit of ~300 residues at which point larger proteins tend to form multiple domains which then assemble as intact units via a different governing equation.

well as the direction of the unfolding trajectory (c.f. Fig. 6B and D). As the k_{1U} parameter determines the strength of the interaction between denaturant and the range of N_{FD} sites in the unfolded state of the protein (regions of the protein that were initially exposed to solvent in the folded state and continue to be exposed to solvent in the unfolded state). For a globular protein the number of N_{FD} sites should be much less than the number of N_{UD} sites (i.e. $N_{FD} < N_{UD}$). The system should therefore be less sensitive to the k_{1U} parameter for an identical increase in the numerical value of k_{1U} and k_{2U} and indeed this is what is seen (c.f. green line in Fig. 6A vs. Fig. 7A).

Case 5: Changes in k_{2U}

The blue, red and green lines in Fig. 7A describe the effects of a 25-fold increase in the value of k_{2U}, the association equilibrium constant governing the binding of denaturant molecules to the initially solvent inaccessible regions of the protein. Due to the U state possessing the largest proportion of denaturant binding sites, the unfolding transition is particularly sensitive to changes in k_{2U}. This sensitivity is seen both at the level of altering stability and cooperativity in the f_U vs. C_D plot (Fig.7A), as well as at the level of changing the pathway followed in the bound site trajectory plots shown as Fig. 7B-D. Interestingly, Fig. 7B presents a different perspective on the existence of very low levels of a denaturant populated state despite being majority folded ($f_U < 0.01$). Although this point will be further addressed in the discussion, the notional concentration independence of the protein unfolding reaction (on the protein concentration itself) potentially opens the door for examining unfolded state separate from the folded state by simply increasing $(C_P)_{TOT}$. If the well-separated nature of the unfolded and folded state ensembles shown in Fig. 7B can be captured via use of an appropriate experimental technique, then the relationship between the protein and denaturant concentrations may prove to be a more powerful tool for tracking the progress of the unfolding reaction than previously thought.

Conclusions

Summary of the model

In this paper we have extended an existing model of protein unfolding that explicitly includes denaturant in the unfolding reaction [**15, 16, 43**]. Our extension is based on consideration of each denaturant laden unfolded species as separate and countable – allowing their determination using a constraint based iterative solution method. The model produces a non-typical output stream that describes the unfolding reaction coordinate in terms of the density of protein states with denaturant bound to the initially solvent exposed and initially solvent hidden regions. Due to the additional dimensionality, these 'new look' unfolding trajectory plots contain more information than their f_U vs C_D forebears (c.f. Fig. 2A vs. 2B), potentially offering greater insight into the unfolding reaction. Amongst these improvements are the following;

(i) A more realistic accounting of the properties displayed by internal and surface located amino acid types by allowing for denaturant k_{1F} sites to be described by a k_{1U} binding constant in the unfolded state – where both k_{1F}, k_{1U} and k_{2U} may take on unique values.

(ii) Ability to identify the denaturant-laden folded state from which the unfolding trajectory is most likely to take place (see Eqn. 8b) due to absolute evaluation of the number of denaturant molecules bound rather than simply the difference (as prescribed by Eqn. 4.

(iii) Capability to evaluate a new type of unfolding trajectory through consideration of the weighted average of all species in the denatured ensemble for a series of denaturant concentrations (Fig. 2B-D).

In the following sections we compare our model against other explicit approaches and describe possible means for its experimental testing and extension.

Other analytical denaturant site-binding models

Our approach belongs to a class of mechanistic based models that feature independent and group-wise equivalent denaturant binding sites existing on the folded and unfolded states of the protein [**15, 16, 43, 46**]. Before discussing the literature associated with this class of models it is helpful to discuss what we actually mean when invoking the term 'denaturant binding site'. With reference to the denaturants urea and guanidine hydrochloride, results of molecular dynamics simulations (**54-62**) suggest that the energetic basis of these 'sites' derive from a mixture of;

(a) Specific direct interactions between the denaturant and protein driven by hydrogen bond formation between denaturant and backbone amide and side chain moieties (**55, 57, 59, 61**).

(b) Non-specific direct interactions between the denaturant and protein driven by a combination of short-range attractive dispersive interactions (**54, 57, 59, 60**).

(c) Non-specific indirect interactions between the denaturant and protein driven by disruption of water structure by the denaturant (**55, 56, 62** vs. **58**)[10].

Irrespective of the exact nature of the binding site their invocation is central to models seeking explicit inclusion of denaturant into the unfolding mechanism on the basis of proximal contact between the protein and the denaturant. Schellman [**36**] and Fersht [**9**] provide introductions to the quantitative description of denaturant induced protein unfolding in terms of site binding theory. Two general approaches have been adopted, models based on the existence of distinct binding sites displaying a positional energetic minimum (**15, 16, 43**) and models based on the concept of partition of denaturant into a protein surface phase (**36, 44, 57, 60, 63-67**).

(i) Models based on energetic sites: Schellman and Aune and Tanford (**15, 16**) independently produced binding models of the general type shown by Eqns. 1-4. Aune and Tanford (**16**) extended this approach to include displacement[11] type reactions and denaturation based on pH titration in which the binding ligand was a proton. In a series of publications from the 70s to the late 90s Schellman refined the site binding model to account for multiple adsorbing ligands [**36, 44**]. Although not directed at studying denaturant induced unfolding Miller and Dill [**43**] applied ligand binding theory to their two dimensional HP system [12] to generate a model of denaturant unfolding based on a two state equilibrium (F ⇌U)[13].

(ii) Models based on partition and surface phase: A different approach for mechanistic inclusion of the denaturant concentration into the quantitative formulation of denaturant unfolding based on the general concept of differential partition of solute between bulk and local surface-phase (immediately adjoint to the protein surface) has been variously advanced by the groups of Bolen [**63**], Parsegian [**64**], Winzor [**65**], Minton [**66**] and Schellman [**36, 44**] and Matubayasi [**67**]. Differential partition of solute into this local surface phase can signify an attractive (K_P > 1 more solute than expected) or a repulsive (K_P < 1 less solute than expected) potential of mean force (**67**) (Eqn. 10a). The overall extent of attraction can be measured by a partition excess, Γ_{32}, [**36, 44, 57, 60**] generalized here for a three component solution in which component 1 is water, component 2 is protein and component 3 is denaturant (Eqn. 10b).

$$C_3^{local} = K_P C_3^{bulk} \qquad [10a]$$

$$\Gamma_{32} = \langle N_3^{local} - \left(\frac{N_3^{bulk}}{N_1^{bulk}}\right) N_1^{local} \rangle \qquad [10b]$$

The designations local and bulk respectively refer to the volume regions defined by the normal projection from the surface a distance approximately equal to the molecular distance of a denaturant molecule [**60, 68-70**] and the volume regions starting from this end of the local phase extending further out into solution (below the protein-protein free volume approach distance defined by the protein concentration. Schellman has advanced this approach, extending it to take into account various forms of non-ideality in the bulk solution phase [**69**]. It is interesting to note that this partition/surface excess approach is also amenable to implementation in our multi-dimensional modelling strategy. A further point for consideration is that all chemical site binding models should be based on chemical activity, the mass action analog of concentration, which includes any additional non-deal work associated with maintaining the listed concentration in the specified environment[14]. In the present work we have assumed ideal behavior of the denaturant but note that Aune and Tanford [**15**] did present an empirical activity correction based on vapor pressure data. In future work we will explore improved experimental and theoretical methods for making this activity correction.

[10] Note that this may be extremely dependent upon solvent composition [see **58** vs. **62**]

[11] Referring to displacement of solvent molecules.

[12] Here two-dimensional refers to where H = hydrophobic amino acid and P = polar amino acid).

[13] We note that Dill and coworkers developed another class of models based on the non-mechanistic m-value formulation for description of both the denaturant and temperature dependence of the unfolding equilibrium constant [**35**].

[14] The chemical potential of an ideal species is defined as per Eqn. a1b, however for a non-ideal component an additional term is included $RT\log_e\gamma$ such that the chemical activity of species i , $a_i = C_i \gamma_i$.

Experimental testing

Used in combination with a non-linear least squares regression routine [**71**], the developed model may be used to fit denaturant induced protein unfolding curves and evaluate the variable parameters within the model. Experimental dependencies for each parameter could be constructed, upon such phenomenon as size, shape and internal secondary structure content [e.g. **18**]. The model could be used to fit standard f_U vs C_D type transition curves (e.g. Fig. 2A) or alternatively fit data of the type shown in Fig. 2B, gained from H/D exchange measurements, saturation transfer/spin relaxation NMR experiments [**22, 72-74**] or 2D IR studies [**24**].

Interestingly, use of the model in an analytical capacity suggests a second line of experimental investigation based on a protein engineering /mutation approach [**75-77**]. Evaluation of a particular amino acid's contribution to one of the characteristic parameters (when located at a particular position) will provide mechanistic insight in regard to the mode of its effect upon stability [**75, 76**] thereby complementing modern techniques [**77, 78**]. For instance, with regard to the intrinsic stability parameter, $(k_{FU})_{0,0}$, a broad based alanine scan [**75, 78**] could identify key stabilizing amino acids, with their particular roles assigned through analysis of covariance with other parameters (k_{1F}, k_{1U}, k_{2U}, N_{FD} or N_{UD}). It is also conceivable, that designed mutation could be used to experimentally contrive a series of differently sized proteins for which intrinsic protein stability was kept constant with increasing size, thereby allowing the effects on apparent cooperativity predicted in this simulation (Fig. 4) to be tested (ref. **79** is interesting in regard to this point). Furthermore, identification of a strong/weak experimental correlation between a particular amino acid and any of the parameter set [$(k_{FU})_{0,0}$, k_{1F}, k_{1U}, k_{2U}, N_{FD} or N_{UD}] could assist in more judicious design of mutations chosen with alternative aims, of say increasing enzymatic catalytic efficiency (**80**) or selection of a non-structural amino acid as a site for covalent attachment of a useful molecule (such as a FRET donor/acceptor (**81**) or a PEG/biotin molecule for further stabilization/purification (**82**).

Possible Extensions to the Model

Despite (or perhaps because of) its relative simplicity we believe that the 'new look' model is extensible and there are a number of directions we would like to take it. Without being exhaustive we discuss a few of these areas below.

One future avenue would be the inclusion of more structural information in the assignment of the number of N_{FD} and N_{UD} sites. Appendix 2 describes this calculation for transition between a folded sphere (globular protein) and an unfolded thin rod (polypeptide chain) but it would be interesting to repeat this calculation for asymmetric proteins both at the mesoscopic level of approximation (**51**) as well as using structural information featuring atomic detail. Ultimately we wish to study site numbers as a function of shape and to use this model extension to examine how shape can determine stability (**18**).

Another area we would like to develop the model involves using it to examine competitive binding between two different types of denaturant, between denaturant and ligand and even that between denaturant and water (**15, 43, 47-50, 57, 60**). The model could be reformulated to accommodate displacement reactions occurring between two denaturants/ligands which could prove useful in addressing the dispute between models which assign differential weights to hydrophilic and hydrophobic forces [**83, 84**].

We believe that our model also holds great promise as the base vehicle for factoring in both denaturant and temperature dependence of the unfolding equilibrium – leading to a generalized analytical formulation. Similar to the approach attempted by Miller and Dill (**43**) we suggest that each site binding constant could be parameterized in terms of an Arrhenius type dependence.

Final Remarks

The original motivation underlying this study was the development of a model of denaturant induced protein unfolding into which we could introduce additional mechanistic factors affecting the denaturant protein interaction. Three such cases of particular interest to us are (i) role of denatured state in contributing to amyloid formation [**85-88**], (ii) effects of competition between denaturants and weakly stabilizing ligands, and (iii) competing effects of macromolecular crowding on protein folding and denaturant binding events [**89-91**]. As we realized the potential pedagogical power of our developed approach we decided to put these initial plans on hold and investigate the basic capabilities of the model to inform on the process of denaturant induced protein folding. The model provides a different perspective on the protein folding problem

and a fresh set of parameters with which to classify the process. It has (re)kindled our interest in the subject and we believe that it might provide others with a new viewpoint from which to ponder the protein folding problem – one of the biggest scientific questions of the last century [**92, 93**].

Acknowledgements: DH would like to acknowledge a lively/heated discussion with Dr. Allen P. Minton that prompted a deeper look at the equations used for characterizing the protein unfolding transition. DH would also like to thank Prof. Thomas Huber and Prof. Gottfried Otting for helpful discussions and comments on an early version of this paper. The work of DH was jointly funded by an Australian National University Senior Research Fellowship, an Osaka University Cross-Appointment as an Associate Professor. The work of AK was funded by the Protein Data Base of Japan (PDBj). The work of YG was funded under the Cooperative Research Program for the Institute for Protein, Osaka University, ICR-15-02 and was supported by JSPS KAKENHI Grant Number 15H04362.

Appendix 1: The m-value approach to the description of protein folding by Pace et al.

In a 1974 publication, Greene and Pace [**A1A**] showed that the measured free energy of unfolding was linear with respect to denaturant concentration for four different proteins. This empirical finding was the basis for the m-value approach which has later been rationalized by comparison against other theories by Pace [**A1B**] and in terms of thermodynamic theory by Schellman [**A1C**]. Using the formalism of the chemical potential we present a compact derivation of the m-value approach by Pace et al. [**A1A, A1B**]. From the lumped model of the reversible unfolding reaction shown in Eqn. 2a, in which C_U and C_F represents the summed concentration of all unfolded and folded species respectively, the condition for folding equilibrium at a certain denaturant concentration, C_D, can be operationally defined by Eqn. **a1a-c**.

$$\mu_{U\{C_D\}} = \mu_{F\{C_D\}} \quad [a1a]$$

$$\mu_{F\{C_D\}} = \left(\frac{\partial G}{\partial n_F}\right)_{n_D,n_U} = \mu^{\circ}_{F\{C_D\}} + RTlog_e\left(\frac{C_{F\{C_D\}}}{C^{\circ}_{F\{C_D\}}}\right) \quad [a1b]$$

$$\mu_{U\{C_D\}} = \left(\frac{\partial G}{\partial n_U}\right)_{n_D,n_F} = \mu^{\circ}_{U\{C_D\}} + RTlog_e\left(\frac{C_{U\{C_D\}}}{C^{\circ}_{U\{C_D\}}}\right) \quad [a1c]$$

In equation set A1a-e the symbol G refers to the free energy of the system at constant temperature and pressure, R is the universal gas constant and T is the absolute temperature. The superscript symbol ° describes the value of the respective standard state value which we define as a hypothetical one molar solution of the solute present as a dimensionless uncharged point particle [**A1D**]. The association constants defining this equilibrium condition can be operationally defined by Eqn. **a1d-e**.

$$\Delta\mu^{\circ}_{F\to U\{C_D\}} = \mu^{\circ}_{U\{C_D\}} - \mu^{\circ}_{F\{C_D\}} = -RTlog_e\left(\frac{C_{U\{C_D\}}}{C_{F\{C_D\}}}.\frac{C^{\circ}_{F\{C_D\}}}{C^{\circ}_{U\{C_D\}}}\right) \quad [a1d]$$

$$K_{F\to U,\{C_D\}} = \left(\frac{C_{U\{C_D\}}}{C_{F\{C_D\}}}.\frac{C^{\circ}_{F\{C_D\}}}{C^{\circ}_{U\{C_D\}}}\right) = exp\left(-\frac{\Delta\mu^{\circ}_{F\to U,\{C_D\}}}{RT}\right) \quad [a1e]$$

For a two-state folding equilibrium the value of the denaturant concentration at which $C_U = C_F$ (i.e. $K_{F\to U} = 1$) is termed the denaturant midpoint $(C_D)_{mp}$. If $(C_D)_{mp}$ is taken as a reference state then the difference in chemical potential between F and U at some value ΔC_D [where $\Delta C_D = C_D - (C_D)_{mp}$] can be calculated via a first order linear perturbation in free energy, with the constant derivative term named as the m value (Eqn. **a1f-g**).

$$\Delta\mu^{\circ}_{F\to U\{C_D\}} = \Delta\mu^{\circ}_{F\to U\{(C_D)_{mp}\}} + m.(\Delta C_D) \quad [a1f]$$

$$m = \left[\frac{d\left(\Delta\mu^{\circ}_{F\to U,\{(C_D)\}}\right)}{dC_D}\right]_{(C_D)_{mp}} \quad [a1g]$$

Functionally the m value can be used as fitting parameter [**A1B, A1E**] or estimated ab initio using various levels of statistical mechanics based approaches [**A1F, A1G**] to produce an equation of the form given by Eqn. **a1h-i**.

$$K_{F\to U\{C_D\}} = K_{F\to U\{(C_D)_{mp}\}}.exp(-m_{.}\,\Delta C_D/RT) \quad [a1h]$$

$$K_{F\to U\{(C_D)_{mp}\}} = 1.exp(-0/RT) = 1 \quad [a1i]$$

With only two variable parameters (m and $(C_D)_{mp}$), the first-order linear perturbation treatment (Eqn. **a1h**) has been shown to provide an excellent empirical description of the two state protein unfolding equilibrium for a large number of proteins [**A1a**].

Appendix 1 References

[A1A] Greene, R.F. and Pace, C.N., 1974. Urea and guanidine hydrochloride denaturation of

ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. *Journal of Biological Chemistry*, *249*(17), pp.5388-5393.

[A1B] Pace, C.N., 1986. Chapter 14. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in enzymology*, *131*, pp.266-280.

[A1C] Schellman, J.A., 2003. Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophysical Journal*, *85*(1), pp.108-125.

[A1D] Van Holde, K.E., Johnson, W.C. and Ho, P.S., 2006. Principles of physical biochemistry. Chapter 14 Chemical Equuilibria Invoving Macromolecules. 2nd Edition. Pearson Prentice Hall.

[A1E] Beg, I., Minton, A.P., Islam, A., Hassan, M.I. and Ahmad, F., 2017. The pH dependence of saccharides' influence on thermal denaturation of two model proteins supports an excluded volume model for stabilization generalized to allow for intramolecular electrostatic interactions. *Journal of Biological Chemistry*, *292*(2), pp.505-511.

[A1F] Tanford, C., 1970. Protein Denaturation: Part C. Theoretical Models for The Mechanism of Denaturation. *Advances in protein chemistry*, *24*, pp.1-95.

[A1G] Ghosh, K. and Dill, K.A., 2009. Computing protein stabilities from their chain lengths. *Proceedings of the National Academy of Sciences*, *106*(26), pp.10649-10654.

Appendix 2: Defining N_{UD} and N_{FD} in terms of N_{AA} for a particular protein shape

We demonstrate the basis of our simplifying assumption that N_{UD} and N_{FD} may be directly determined by a single parameter N_{AA} for a protein of known structure by treating the relatively simple example of a globular (and therefore approximately spherical) protein. To make our case we first assume that the total number of denaturant binding sites on the protein is proportional to the total normal projected surface area of the stretched polypeptide chain, A_{PP}, relative to the size of an adsorbing spherical denaturant of radius R_D [**A2A**]. Equation a2a, describing this total projected surface area, is realized by equating two different geometrical models of the polypeptide chain, corresponding to the linked bead model [**A2B**] (where the polypeptide chain is composed of a set of linked spheres) and a tube model [**A2C**] (where the extended chain is considered as a cylinder). Further required assumptions involve assigning all amino acids an average spherical size, R_{AA} [**A2D**] and defining the unfolded polypeptide contour chain length, L_U, in terms of the N_{AA} amino acids (where $L_U = 2N_{AA}R_{AA}$). The total number of denaturant binding sites on the unfolded polypeptide chain is calculated from this area via Eqn. a2b on the basis of two further assumptions, namely (i) an assumed area of projection of denaturant onto the surface of the cylinder, A_D (where $A_D = \pi R_D^2$) and (ii) a two dimensional packing fraction, β, reflecting fractional usage of the polypeptide surface area by denaturant.

$$A_U = \left[\sqrt{\frac{32(N_{AA}\pi)^2}{3}} + \left(\frac{4\pi}{3}\right)\right] {R_{AA}}^2 + 4\pi N_{AA} R_{AA} R_D \quad \text{[a2a]}$$

$$N_{TOT} = (N_{FD} + N_{UD}) = \beta A_U / A_D \quad \text{[a2b]}$$

In a similar fashion the normal projected surface area of the folded protein, A_F, relative to a spherical denaturant of radius R_D, can be determined on the basis of the assumptions of spherical geometry, volume additivity and set volume packing fraction, α, reflecting the degree of close packing of amino acids within the folded protein [see appendix of **A2E**] (Eqn. a2c). As for the preceding case, N_{FD} is considered proportional to A_F and is determined via Eqn. a2d on the basis of assumptions reflecting (i) an assumed area of projection of denaturant onto the spherical surface of the folded protein, A_D, and (ii) a two dimensional packing fraction, ε, reflecting fractional usage of the polypeptide surface area by denaturant.

$$A_F = 4\pi \left(\sqrt[3]{\frac{N_{AA}}{\alpha}}.R_{AA} + R_D\right)^2 \quad \text{[a2c]}$$

$$N_{FD} = \varepsilon A_F / A_D \quad \text{[a2d]}$$

The values of the additional supporting parameters, β and ε, reflecting the degree of two dimensional close packing of denaturant to the normal projected surface area of the unfolded polypeptide chain and the folded spherical protein, were set to their two dimensional random close packing limit of 0.7 [15] i.e. ε = β = 0.5. A numerical value for α

[15] The chosen fractional value should be lower than both that for random (0.53) and equilibrium (0.903) close packing of circles [**A2N**].

of 0.50 was calculated on the basis of Eqn. a2e where $N_{\mathcal{A}}$ is Avogadro's number, M_{AA} is the average mass of an amino acid (M_{AA} = 0.110 kg.mol^{-1} [**A2F**]), $\bar{v}_P$ is the partial specific volume of a protein in water ($\bar{v}_P$ = 7.3×10^{-4} m^3.kg^{-1} [**A2F**]) and R_{AA} is set equal to 0.4nm [**A2F**]. As denaturant molecules come in a range of sizes e.g. much smaller than an amino acid (such as guanidine hydrochloride and urea) or much larger than an amino acid (e.g. various long chain alcohols, sodium dodecyl sulfate) we have taken a conservative estimate and set the denaturant size equal to the amino acid size (i.e. $R_D = R_{AA}$)[16].

$$\alpha = \frac{3M_{AA}\bar{v}_P}{N_{\mathcal{A}}4\pi(R_{AA})^3} \qquad [a2e]$$

As the values of N_{UD} and N_{FD} are calculated on assumption of fractional attainment of the close packing relation of denaturant on the surface of the folded and unfolded structures (with no consideration paid to the location or geometrical relationship between particular chemical groups) it is expected that their values represent upper bounds. Appreciation of the importance of this parameter can be gained from inspection of the case (ii) simulations in which the protein size is varied. Within the confines of the model presented, this case is equivalent to varying the value of β and ε to produce more or less denaturant sites. Particular binding locations of various denaturant molecules have been considered previously via computer simulation based on molecular dynamics [**A2G**].

Although we have considered the relationship between protein size and denaturant binding site number (in the folded and unfolded states) for the particular case of spherical geometry, it is a trivial matter to recalculate this relationship for both

(i) Approximate regular shapes commonly used for general description of protein structures (such as ellipsoids or cylinders [**A2H**]),

(ii) Particular protein structures obtained from the protein data bank (e.g. PDBj [**A2I**]) when used in conjunction with a denaturant binding site model of the type described here.

Appendix 2 References

[A2A] Hall, D. 2008. Kinetic models describing biomolecular interactions at surfaces. In Handbook of Surface Plasmon Resonance (pp. 81-122). Royal Society of Chemistry.

[A2B] Tozzini, V., 2005. Coarse-grained models for proteins. Current opinion in structural biology, 15(2), pp.144-150.

[A2C] Nielsen, S.O., Lopez, C.F., Srinivas, G. and Klein, M.L., 2004. Coarse grain models and the computer simulation of soft materials. Journal of Physics: Condensed Matter, 16(15), p.R481.

[A2D] Hall, D., Li, S., Yamashita, K., Azuma, R., Carver, J.A. and Standley, D.M., 2014. A novel protein distance matrix based on the minimum arc-length between two amino-acid residues on the surface of a globular protein. Biophysical chemistry, 190, pp.50-55.

[A2E] Sasahara, K., Hall, D. and Hamada, D., 2010. Effect of lipid type on the binding of lipid vesicles to islet amyloid polypeptide amyloid fibrils. Biochemistry, 49(14), pp.3040-3048.

[A2F] Cantor, C.R. and Schimmel, P.R., 1980. Biophysical chemistry: Part III: the behavior of biological macromolecules. Macmillan.

[A2G] Stumpe, M.C. and Grubmüller, H., 2007. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *Journal of the American Chemical Society*, *129*(51), pp.16126-16131.

[A2H] Taylor, W.R., Thornton, J.T. and Turnell, W.G., 1983. An ellipsoidal approximation of protein shape. *Journal of Molecular Graphics*, *1*(2), pp.30-38.

[A2I] Kinjo, A.R., Suzuki, H., Yamashita, R., Ikegawa, Y., Kudou, T., Igarashi, R., Kengaku, Y., Cho, H., Standley, D.M., Nakagawa, A. and Nakamura, H., 2012. Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. Nucleic acids research, 40(D1), pp.D453-D460.

Appendix 3: Combination of the Linear Extrapolation Method (LEM) and the Tanford Transfer Free Energy Method.

An anonymous reviewer of the manuscript has asked us to point out an alternative physical description of the two state protein unfolding reaction that is derived by combination of the m-value approach [**A3A**] and the Tanford transfer free energy method [**A3B**]. This can be realized by first extending Eqn. a1h to create Eqn. a3a which

[16] None of the conclusions drawn in this paper depend upon the exact size of the denaturant relative to the amino acid size.

yields the linear extrapolation of free energy to zero denaturant concentration, $\Delta\mu^{\circ}_{F\rightarrow U\{C_{D=0}\}}$.

$$\Delta\mu^{\circ}_{F\rightarrow U\{C_D\}} = \Delta\mu^{\circ}_{F\rightarrow U\{C_{D=0}\}} - \mathrm{m}.\,C_D \qquad \text{[a3a]}$$

A similar functional form can be determined for the two-state protein unfolding reaction when described in terms of the Tanford transfer free energy method which is based on the concept of a free energy cycle [a3b**]**. In this model the free energy per mole associated with the transfer of either the folded, or unfolded, species from solvent without denaturant to solvent with denaturant is respectively denoted as $\Delta\mu^{\circ}_{F,tr}$ and $\Delta\mu^{\circ}_{U,tr}$ [a3b-d]

$$\begin{array}{ccc} & \Delta\mu^{\circ}_{F\rightarrow U\{C_{D=0}\}} & \\ (F)_{C_{D=0}} & \rightleftharpoons & (U)_{C_{D=0}} \\ \Delta\mu^{\circ}_{F,tr}\ \Big\Updownarrow & & \Big\Updownarrow\ \Delta\mu^{\circ}_{U,tr} \\ (F)_{C_D} & \rightleftharpoons & (U)_{C_D} \\ & \Delta\mu^{\circ}_{F\rightarrow U\{C_D\}} & \end{array} \qquad \text{[a3b]}$$

$$\Delta\mu^{\circ}_{F\rightarrow U\{C_D\}} = \Delta\mu^{\circ}_{F\rightarrow U\{C_{D=0}\}} + \left(\Delta\mu^{\circ}_{U,tr} - \Delta\mu^{\circ}_{F,tr}\right) \text{[a3c]}$$

$$\Delta\mu^{\circ}_{F\rightarrow U\{C_D\}} = \Delta\mu^{\circ}_{F\rightarrow U\{C_{D=0}\}} + \textstyle\sum_{i=1} \alpha_i N_i \delta\mu_{i,tr} \quad \text{[a3d]}$$

In Eqn. a3d the difference in transfer free energy is approximated by a linear summation of the differential free energies for transferring a number N_i of a particular i type of amino acid (between the unfolded and folded states) from solvent without denaturant to solvent containing denaturant, $\delta\mu_{i,tr}$. An additional parameter, α_i, reflecting the fractional exposure of a particular amino acid in the folded and unfolded state was found necessary to reconcile constituent data with experiments on whole proteins [**A3C**]. In practical usage an averaged value, $\bar{\alpha}$, is adopted for the fractional contribution term allowing Eqn. a3d to be further simplified [a3e].

$$\Delta\mu^{\circ}_{F\rightarrow U\{C_D\}} = \Delta\mu^{\circ}_{F\rightarrow U\{C_{D=0}\}} + \bar{\alpha}\textstyle\sum_{i=1} N_i \delta\mu_{i,tr} \quad \text{[a3e]}$$

Bolen and colleagues [**A3D, A3E**] have used the apparent linear dependence of the unfolding free energy on the concentration of denaturant (Eqn. a3a) to demonstrate a direct equivalence for the transfer free energy term [a3f].

$$\bar{\alpha}\textstyle\sum_{i=1} N_i \delta\mu_{i,tr} = -m.\,C_D \qquad \text{[a3f]}$$

Such a formulation has allowed for electrostatic models of the free energy of transfer to be used in the reconciliation of experimental data of protein unfolding in solutions containing high concentrations of osmolytes such as trimethylamine oxide (TMAO) [**A3D**]. However due to its requirement for prior evaluation of the m value and its specification of a large number of individual transfer parameters the described method lacks the capabilities of the procedure described in the current paper with regard to,

(i.) Accounting for the form of the dependence of the unfolding curve on denaturant concentration when it differs from the ideal case.

(ii.) Providing useful information on the distributed nature of the unfolded state as a function of denaturant concentration.

Appendix 3 References

[A3A] Myers, J.K., Nick Pace, C. and Martin Scholtz, J., 1995. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science*, *4*(10), pp.2138-2148.

[A3B] Tanford, C., 1970. Protein Denaturation: Part C. Theoretical Models for The Mechanism of Denaturation. *Advances in protein chemistry*, *24*, pp.1-95.

[A3C] O'Brien, E.P., Ziv, G., Haran, G., Brooks, B.R. and Thirumalai, D., 2008. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. Proceedings of the National Academy of Sciences. 105, 13403-13408.

[A3D] Auton, M., and Bolen, D.W. (2005) Predicting the energetics of osmolyte-induced protein folding/unfolding. Proceedings of the National Academy of Sciences. 102, 15065-15068.

[A3E] Street, T.O., Bolen, D.W., and Rose, G.D. (2006) A molecular mechanism for osmolyte-induced protein stability. Proceedings of the National Academy of Sciences. 103, 13997-14002.

Appendix 4: Confidence of Parameters Determined Using the New Look Model.

An anonymous reviewer of our paper has asked us to describe the effect of normal error added to unfolding data on the capability of the model to accurately determine parameter values. Although we do not yet have any suitable multi-dimensional

data to directly compare against our model we have proceeded to answer this question by exploring the one-dimensional error surface associated with each parameter for a system of unknown geometry/site distribution. In this case (unknown geometry) we actually have six unknown parameters in the model, $(k_{FU})_{0,0}$, k_{1F}, k_{1U}, k_{2U}, N_{FD} and N_{UD}. Appendix 4 Fig. 1A-F shows the calculated confidence in each parameter estimated by sequential sampling from two different distributions of residuals. The first distribution, $\mathbf{R_{\lambda,\sigma}}$, reflects a set of absolute residuals produced by addition of error sampled from a Gaussian curve characterized by zero mean and standard deviation, σ, to an exact solution, $f_u(C_D;\lambda)$, determined using the parameter value λ. The second distribution, $\mathbf{R_{\lambda+\Delta\lambda,\sigma}}$, was constructed by histogram analysis of the absolute value of residuals existing between synthetic $f_u(C_D;\lambda)$ data sets with added normal error (zero mean and standard deviation, σ) and the set of $f_u(C_D;\lambda+\Delta\lambda)$ values determined by a slight perturbation in the parameter value, $\lambda + \Delta\lambda$ [**A4A**] [17]. In brief, this procedure examines the extent of coincidence of n sampled points to these two distributions of residuals, $\mathbf{R_{\lambda,\sigma}}$, and $\mathbf{R_{\lambda+\Delta\lambda,\sigma}}$, calculated over the transition region, $0.05 \leq f_u \leq 0.95$. Eqn. a4a closely approximates the probability of acceptance of the null hypothesis that the set of n residuals could belong to the joint area of overlap, $JA\,(\lambda_{true+\Delta\lambda}, \lambda_{true})$, of both $\mathbf{R_{\lambda,\sigma}}$ and $\mathbf{R_{\lambda+\Delta\lambda,\sigma}}$ distributions, whilst $Area_{\lambda_{true+\lambda}}$ is the integrated area of the $\mathbf{R_{\lambda+\Delta\lambda,\sigma}}$ distribution alone.

$$\mathrm{P}(\lambda + \Delta\lambda, n) \approx \left[\frac{(J\,A\,(\lambda_{true+\Delta\lambda}, \lambda_{true})}{Area_{\lambda_{true+\lambda}}}\right]^n \qquad \text{[a4a]}$$

A number of points can be noted from Appendix Figs. 1A-F and Eqn. a4a,

(i) Highest probability of acceptance of the null hypothesis is around the input value.

(ii) Acceptance probability for the null hypothesis broadens with regard to the magnitude of $\Delta\lambda$ for larger amounts of added normal error.

(iii) Determinability of the parameter is related to the number of data points (i.e. n) in the data set to be fitted.

(iv) Error distributions reflect the sensitivity of f_U, C_D plot results to change in the parameter. As such we note that the parameters relating to the number and affinity of the UD sites are the most sensitive to relative change in parameter value.

Typically, quoted errors in parameter values are given as the value of $\Delta\lambda$ at which $\mathrm{P}(\lambda + \Delta\lambda) = 0.05$. As noted from Appendix Fig. 1A-F, all parameter values have a discernible maxima at their true value. Furthermore, even with a very low number of data points (n = 5) all parameters are determinable to within 50% of their true value when extents of added normal error have a standard deviation below 0.075 (in relation to a range of f_u vaues of [0, 1.0]. This value of error is easily achievable in protein folding experiments providing sufficient care is taken (e.g. as for [**A4B**]). As such, we suggest that provided a suitably efficient non-linear regression routine is employed for the fitting cycle then our model will be be capable of determining parameter values with sufficient accuracy and acceptable levels of error for a data set with low noise and a large number of data points. In the next paper in this series we will explore this point in more detail.

Appendix 4 References

[A4A] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 1992. *Numerical recipes in C* 2nd Edition. Chapter 15 Modeling of Data. Cambridge university press.

[A4B] Sasahara, K., Demura, M. and Nitta, K., 2002. Equilibrium and kinetic unfolding of hen egg white lysozyme under acidic conditions. *Proteins: Structure, Function, and Bioinformatics*, *44*(3), pp.180-187.

Appendix 4 - Figure 1: Effect of added normal error on the determination of model parameter values. Three different extents of error, ξ, were added to the primary data, $[(C_D)_{TOT}, f_u \pm \xi]$, with ξ sampled from a normal distribution characterized by zero mean and a standard deviation of either 0

[17] By confidence we mean the probability P with which we can accept the results produced using the modified parameter value, $\lambda+\Delta\lambda$, as an equivalent approximation to the results produced using the true value of the parameter, λ_{true}. Typically parameter confidence is estimated using a large number of Monte Carlo cycles to add equivalent extents of normal error to the data, followed by non-linear regression to evaluate the best fit value of the parameter for evaluation. An frequency plot of obtained parameter values is then constructed and confidence values for the parameter are determined by integration of this distribution. For reasons relating to computational simplicity we have adopted a slightly different approach in which the calculated errors are derived from comparison of a distribution of absolute residuals calculated between the data and best fit (correct) solution and a distribution of residuals between with the correct values and solutions for each altered parameter value with added normal error.

(red lines) or 0.075 (blue lines). Error in each of the six model parameters was determined by Eqn. a4a for the three different extents of added error and set number of data points (n = 5) taken from the unfolding transition region. Appendix 4 Figs. 1 (A-F) respectively referring to $(k_{FU})_{0,0}$, k_{1F}, k_{1U}, k_{2U}, N_{FD} and N_{UD}.

References

[1] Souza, T.V., Araujo, J.N., da Silva, V.M., Liberato, M.V., Pimentel, A.C., Alvarez, T.M., Squina, F.M. and Garcia, W., 2016. Chemical stability of a cold-active cellulase with high tolerance toward surfactants and chaotropic agent. *Biotechnology Reports*, *9*, pp.1-8.

[2] Villamonte, G., Pottier, L. and de Lamballerie, M., 2016. Influence of high-pressure processing on the physicochemical and the emulsifying properties of sarcoplasmic proteins from hake (Merluccius merluccius). *European Food Research and Technology*, *242*(5), pp.667-675.

[3] Saito, S., Hasegawa, J., Kobayashi, N., Tomitsuka, T., Uchiyama, S. and Fukui, K., 2013. Effects of ionic strength and sugars on the aggregation propensity of monoclonal antibodies: influence of colloidal and conformational stabilities. *Pharmaceutical research*, *30*(5), pp.1263-1280.

[4] Hurle, M.R., Helms, L.R., Li, L.I.N., Chan, W. and Wetzel, R., 1994. A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proceedings of the National Academy of Sciences*, *91*(12), pp.5446-5450.

[5] Von Hippel, P.H. and Wong, K.Y., 1965. On the conformational stability of globular proteins-the effects of various electrolytes and nonelectrolytes on the thermal ribonuclease transition. *Journal of biological chemistry*, *240*(10), pp.3909-3923.

[6] Privalov, P.L. and Khechinashvili, N.N., 1974. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *Journal of molecular biology*, *86*(3), pp.665-684.

[7] Sturtevant, J.M., 1977. Heat capacity and entropy changes in processes involving proteins. *Proceedings of the National Academy of Sciences*, *74*(6), pp.2236-2240.

[8] McPhie, P., Ni, Y.S. and Minton, A.P., 2006. Macromolecular crowding stabilizes the molten globule form of apomyoglobin with respect to both cold and heat unfolding. *Journal of molecular biology*, *361*(1), pp.7-10.

[9] Fersht, A., 1999. A guide to enzyme catalysis and protein folding. *Structure and mechanism in protein science*, Chapter 17 'Protein Stability'. pp. 508-539.

[10] Mozhaev, V.V., Heremans, K., Frank, J., Masson, P. and Balny, C., 1996. High pressure effects on protein structure and function. *Proteins-Structure Function and Genetics*, *24*(1), pp.81-91.

[11] Sasahara, K., Sakurai, M. and Nitta, K., 2001. Pressure effect on denaturant□induced unfolding of hen egg white lysozyme. *Proteins: Structure, Function, and Bioinformatics*, *44*(3), pp.180-187.

[12] Pandharipande, P.P. and Makhatadze, G.I., 2016. Applications of pressure perturbation calorimetry to study factors contributing to the volume changes upon protein unfolding. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1860*(5), pp.1036-1042.

[13] Brandts, J.F., 1964. The thermodynamics of protein denaturation. I. The denaturation of chymotrypsinogen. *Journal of the American Chemical Society*, *86*(20), pp.4291-4301.

[14] Tanford, C., 1964. Isothermal unfolding of globular proteins in aqueous urea solutions. *Journal of the American Chemical Society*, *86*(10), pp.2050-2059.

[15] Aune, K.C. and Tanford, C., 1969. Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II. Dependence on denaturant concentration at 25. *Biochemistry*, *8*(11), pp.4586-4590.

[16] Schellman, J.A. (1956) The stability of hydrogen-bonded peptide structures in aqueous solution. *C. R. trav. lab. Carlsberg Ser. Chim.* 29:230-259

[17] Pace, C.N., 1986. Chapter 14. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in enzymology*, *131*, pp.266-280.

[18] Myers, J.K., Nick Pace, C. and Martin Scholtz, J., 1995. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science*, *4*(10), pp.2138-2148.

[19] Gupta, R., Yadav, S. and Ahmad, F., 1996. Protein stability: urea-induced versus

guanidine-induced unfolding of metmyoglobin. *Biochemistry*, *35*(36), pp.11925-11930.

[20] Greene, R.F. and Pace, C.N., 1974. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. *Journal of Biological Chemistry*, *249*(17), pp.5388-5393.

[21] Nozaki, Y. and Tanford, C., 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions establishment of a hydrophobicity scale. *Journal of Biological Chemistry*, *246*(7), pp.2211-2217.

[22] Roy, M. and Jennings, P.A., 2003. Real-time NMR kinetic studies provide global and residue-specific information on the non-cooperative unfolding of the β-trefoil protein, interleukin-1β. *Journal of molecular biology*, *328*(3), pp.693-703.

[23] Patra, M., Mukhopadhyay, C. and Chakrabarti, A., 2015. Probing conformational stability and dynamics of erythroid and nonerythroid spectrin: effects of urea and guanidine hydrochloride. *PloS one*, *10*(1), p.e0116991.

[24] Huerta-Viga, A. and Woutersen, S., 2013. Protein denaturation with guanidinium: a 2D-IR study. *The journal of physical chemistry letters*, *4*(20), pp.3397-3401.

[25] Herskovits, T.T., Gadegbeku, B. and Jaillet, H., 1970. On the structural stability and solvent denaturation of proteins I. Denaturation by the alcohols and glycols. *Journal of Biological Chemistry*, *245*(10), pp.2588-2598.

[26] Hirota, N., Mizuno, K. and Goto, Y., 1998. Group additive contributions to the alcohol-induced α-helix formation of melittin: implication for the mechanism of the alcohol effects on proteins. *Journal of molecular biology*, *275*(2), pp.365-378.

[27] Amin, M.A., Halder, R., Ghosh, C., Jana, B. and Bhattacharyya, K., 2016. Effect of alcohol on the structure of cytochrome C: FCS and molecular dynamics simulations. *The Journal of Chemical Physics*, *145*(23), p.235102.

[28] Aune, K.C. and Tanford, C., 1969. Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. I. Dependence on pH at 25. *Biochemistry*, *8*(11), pp.4579-4585.

[29] Fink, A.L., Calciano, L.J., Goto, Y., Kurotsu, T. and Palleros, D.R., 1994. Classification of acid denaturation of proteins: intermediates and unfolded states. *Biochemistry*, *33*(41), pp.12504-12511.

[30] Sasahara, K., Demura, M. and Nitta, K., 2002. Equilibrium and kinetic folding of hen egg□ white lysozyme under acidic conditions. *Proteins: Structure, Function, and Bioinformatics*, *49*(4), pp.472-482.

[31] Lumry, R., Biltonen, R. and Brandts, J.F., 1966. Validity of the "two□state" hypothesis for conformational transitions of proteins. *Biopolymers*, *4*(8), pp.917-944.

[32] Privalov, P.L., 1979. Stability of proteins small globular proteins. *Advances in protein chemistry*, *33*, pp.167-241.

[33] Dill, K.A., 1985. Theory for the folding and stability of globular proteins. *Biochemistry*, *24*(6), pp.1501-1509.

[34] Becktel, W.J. and Schellman, J.A., 1987. Protein stability curves. *Biopolymers*, *26*(11), pp.1859-1877.

[35] Ghosh, K. and Dill, K.A., 2009. Computing protein stabilities from their chain lengths. *Proceedings of the National Academy of Sciences*, *106*(26), pp.10649-10654.

[36] Schellman, J.A., 2002. Fifty years of solvent denaturation. *Biophysical chemistry*, *96*(2), pp.91-101.

[37] Zweifel, M.E. and Barrick, D., 2002. Relationships between the temperature dependence of solvent denaturation and the denaturant dependence of protein stability curves. *Biophysical chemistry*, *101*, pp.221-237.

[38] Choi, H.S., Huh, J. and Jo, W.H., 2004. Comparison between denaturant-and temperature-induced unfolding pathways of protein: A lattice Monte Carlo simulation. *Biomacromolecules*, *5*(6), pp.2289-2296.

[39] Rocco, A.G., Mollica, L., Ricchiuto, P., Baptista, A.M., Gianazza, E. and Eberini, I., 2008. Characterization of the protein unfolding processes induced by urea and temperature. *Biophysical journal*, *94*(6), pp.2241-2251.

[40] Lim, W.K., Rösgen, J. and Englander, S.W., 2009. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proceedings of the National Academy of Sciences*, *106*(8), pp.2595-2600.

[41] Tanford, C., 1970. Protein Denaturation: Part C. Theoretical Models for The Mechanism of Denaturation. *Advances in protein chemistry*, *24*, pp.1-95.

[42] Brandts, J.F., 1964. The thermodynamics of protein denaturation. II. A model of reversible denaturation and interpretations regarding the stability of chymotrypsinogen. *Journal of the American Chemical Society*, *86*(20), pp.4302-4314.

[43] Miller, D.W. and Dill, K.A., 1997. Ligand binding to proteins: the binding landscape model. *Protein science*, *6*(10), pp.2166-2179.

[44] Schellman, J.A., 2003. Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophysical Journal*, *85*(1), pp.108-125.

[45] Steiner, E.S., 2008. The Chemistry Maths Book. 2nd Edition. Chapter 21 Probabilities and Statistics. Oxford University Press.

[46] Van Holde, K.E., Johnson, W.C. and Ho, P.S., 2006. Principles of physical biochemistry. Chapter 14 Chemical Equuilibria Invoving Macromolecules. 2nd Edition. Pearson Prentice Hall.

[47] Goto, Y., Takahashi, N. and Fink, A.L., 1990. Mechanism of acid-induced folding of proteins. *Biochemistry*, *29*(14), pp.3480-3488.

[48] Hagihara, Y., Tan, Y. and Goto, Y., 1994. Comparison of the conformational stability of the molten globule and native states of horse cytochrome c: effects of acetylation, heat, urea and guanidine-hydrochloride. *Journal of molecular biology*, *237*(3), pp.336-348.

[49] Mayr, L.M. and Schmid, F.X., 1993. Stabilization of a protein by guanidinium chloride. *Biochemistry*, *32*(31), pp.7994-7998.

[50] Bhuyan, A.K., 2002. Protein stabilization by urea and guanidine hydrochloride. *Biochemistry*, *41*(45), pp.13386-13394.

[51] Taylor, W.R., Thornton, J.T. and Turnell, W.G., 1983. An ellipsoidal approximation of protein shape. *Journal of Molecular Graphics*, *1*(2), pp.30-38.

[52] Pont, M.J. and Woods, E.F., 1971. Denaturation of tropomyosin by guanidine hydrochloride. *International journal of protein research*, *3*(1□4), pp.177-183.

[53] Dill, K.A., 1985. Theory for the folding and stability of globular proteins. *Biochemistry*, *24*(6), pp.1501-1509.

[54] Caflisch, A. and Karplus, M., 1999. Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure*, *7*(5), pp.477-S2.

[55] Bennion, B.J. and Daggett, V., 2003. The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences*, *100*(9), pp.5142-5147.

[56] Harano, Y. and Kinoshita, M., 2005. Translational-entropy gain of solvent upon protein folding. *Biophysical journal*, *89*(4), pp.2701-2710.

[57] Canchi, D.R. and García, A.E., 2013. Cosolvent effects on protein stability. *Annual review of physical chemistry*, *64*, pp.273-293.

[58] Bandyopadhyay, D., Mohan, S., Ghosh, S.K. and Choudhury, N., 2014. Molecular dynamics simulation of aqueous urea solution: is urea a structure breaker?. *The Journal of Physical Chemistry B*, *118*(40), pp.11757-11768.

[59] Paul, S. and Paul, S., 2015. Exploring the Counteracting Mechanism of Trehalose on Urea Conferred Protein Denaturation: A Molecular Dynamics Simulation Study. *The Journal of Physical Chemistry B*, *119*(30), pp.9820-9834.

[60] Mondal, J., Halverson, D., Li, I.T., Stirnemann, G., Walker, G.C. and Berne, B.J., 2015. How osmolytes influence hydrophobic polymer conformations: A unified view from experiment and theory. *Proceedings of the National Academy of Sciences*, *112*(30), pp.9270-9275.

[61] Stumpe, M.C. and Grubmüller, H., 2007. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *Journal of the American Chemical Society*, *129*(51), pp.16126-16131.

[62] Mountain, R.D. and Thirumalai, D., 2004. Alterations in water structure induced by guanidinium and sodium ions. *The Journal of Physical Chemistry B*, *108*(51), pp.19711-19716.

[63] Auton, M., Bolen, D.W. and Rösgen, J., 2008. Structural thermodynamics of protein preferential solvation: osmolyte solvation of proteins, aminoacids, and peptides. *Proteins: Structure, Function, and Bioinformatics*, *73*(4), pp.802-813.

[64] Colombo, M.F., Rau, D.C. and Parsegian, V.A., 1992. Protein solvation in allosteric regulation: a water effect on hemoglobin. *Science*, *256*(5057), p.655.

[65] Wills, P.R. and Winzor, D.J., 1993. Thermodynamic analysis of “preferential solvation” in protein solutions. *Biopolymers*, *33*(10), pp.1627-1629.

[66] Nagarajan, S., Amir, D., Grupi, A., Goldenberg, D.P., Minton, A.P. and Haas, E., 2011. Modulation of functionally significant conformational equilibria in adenylate kinase by high concentrations of trimethylamine oxide attributed to volume exclusion. *Biophysical journal*, *100*(12), pp.2991-2999.

[67] Shimizu, S. and Matubayasi, N., 2014. Preferential solvation: Dividing surface vs excess numbers. *The Journal of Physical Chemistry B*, *118*(14), pp.3922-3930.

[68] Schellman, J.A., 1987. Selective binding and solvent denaturation. *Biopolymers*, *26*(4), pp.549-559.

[69] Schellman, J.A., 2003. Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophysical Journal*, *85*(1), pp.108-125.

[70] Hall, D., Li, S., Yamashita, K., Azuma, R., Carver, J.A. and Standley, D.M., 2014. A novel protein distance matrix based on the minimum arc-length between two amino-acid residues on the surface of a globular protein. *Biophysical chemistry*, *190*, pp.50-55.

[71] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 1992. *Numerical recipes in C* 2nd Edition. Chapter 15 Modeling of Data. Cambridge university press.

[72] Smith, D.L., Deng, Y. and Zhang, Z., 1997. Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. Journal of mass spectrometry, 32(2), pp.135-146.

[73] Pintacuda, G. and Otting, G., 2002. Identification of protein surfaces by NMR measurements with a paramagnetic Gd (III) chelate. Journal of the American Chemical Society, 124(3), pp.372-373.

[74] Hajduk, P.J., Mack, J.C., Olejniczak, E.T., Park, C., Dandliker, P.J. and Beutel, B.A., 2004. SOS-NMR: a saturation transfer NMR-based method for determining the structures of protein− ligand complexes. Journal of the American Chemical Society, 126(8), pp.2390-2398.

[75] Fersht, A., 1999. A guide to enzyme catalysis and protein folding. *Structure and mechanism in protein science*, Chapter 15 'Protein Engineering'. pp. 508-539.

[76] Fersht, A.R., Matouschek, A. and Serrano, L., 1992. The folding of an enzyme: I. Theory of protein engineering analysis of stability and pathway of protein folding. *Journal of molecular biology*, *224*(3), pp.771-782.

[77] Bendl, J., Stourac, J., Sebestova, E., Vavra, O., Musil, M., Brezovsky, J. and Damborsky, J., 2016. HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. *Nucleic acids research*, *44*(W1), pp.W479-W487.

[78] Zou, J., Song, B., Simmerling, C. and Raleigh, D., 2016. Experimental and Computational Analysis of Protein Stabilization by Gly-to-D-Ala Substitution: A Convolution of Native State and Unfolded State Effects. *Journal of the American Chemical Society*, *138*(48), pp.15682-15689.

[79] Anwer, K., Sonani, R., Madamwar, D., Singh, P., Khan, F., Bisetty, K., Ahmad, F. and Hassan, M.I., 2015. Role of N-terminal residues on folding and stability of C-phycoerythrin: simulation and urea-induced denaturation studies. *Journal of Biomolecular Structure and Dynamics*, *33*(1), pp.121-133.

[80] Frushicheva, M.P., Mills, M.J., Schopf, P., Singh, M.K., Prasad, R.B. and Warshel, A., 2014. Computer aided enzyme design and catalytic concepts. *Current opinion in chemical biology*, *21*, pp.56-62.

[81] Cella, L.N., Biswas, P., Yates, M.V., Mulchandani, A. and Chen, W., 2014. Quantitative assessment of in vivo HIV protease activity using genetically engineered QD□based FRET probes. *Biotechnology and bioengineering*, *111*(6), pp.1082-1087.

[82] Lawrence, P.B., Gavrilov, Y., Matthews, S.S., Langlois, M.I., Shental-Bechor, D., Greenblatt, H.M., Pandey, B.K., Smith, M.S., Paxman, R., Torgerson, C.D. and Merrell, J.P., 2014. Criteria for selecting PEGylation sites on proteins for higher thermodynamic and proteolytic stability. *Journal of the American Chemical Society*, *136*(50), pp.17547-17560.

[83] Durell SR, Ben-Naim A. Hydrophobic-hydrophilic forces in protein folding. Biopolymers. 2017;107:e23020.

[84] Dill, K.A., Bromberg, S., Yue, K., Chan, H.S., Ftebig, K.M., Yee, D.P. and Thomas, P.D., 1995. Principles of protein folding—a perspective from simple exact models. *Protein Science*, *4*(4), pp.561-602.

[85] Hall, D. and Hirota, N., 2009. Multi-scale modelling of amyloid formation from unfolded

proteins using a set of theory derived rate constants. *Biophysical chemistry*, *140*(1), pp.122-128.

[86] Hall, D. and Edskes, H., 2012. Computational modeling of the relationship between amyloid and disease. *Biophysical reviews*, *4*(3), pp.205-222.

[87] Hall, D., Kardos, J., Edskes, H., Carver, J.A. and Goto, Y., 2015. A multi□pathway perspective on protein aggregation: Implications for control of the rate and extent of amyloid formation. *FEBS letters*, *589*(6), pp.672-679.

[88] Johnson, S.M., Connelly, S., Fearns, C., Powers, E.T. and Kelly, J.W., 2012. The transthyretin amyloidoses: from delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug. *Journal of molecular biology*, *421*(2), pp.185-203.

[89] Sasahara, K., McPhie, P. and Minton, A.P., 2003. Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *Journal of molecular biology*, *326*(4), pp.1227-1237.

[90] Hall, D. and Dobson, C.M., 2006. Expanding to fill the gap: a possible role for inert biopolymers in regulating the extent of the 'macromolecular crowding' effect. *FEBS letters*, *580*(11), pp.2584-2590.

[91] Hall, D., 2002. On the role of the macromolecular phase transitions in biology in response to change in solution volume or macromolecular composition: action as an entropy buffer. *Biophysical chemistry*, *98*(3), pp.233-248.

[92] Dill, K.A. and MacCallum, J.L., 2012. The protein-folding problem, 50 years on. *Science*, *338*(6110), pp.1042-1046.

[93] Finkelstein, A.V., Badretdin, A.J., Galzitskaya, O.V., Ivankov, D.N., Bogatyreva, N.S. and Garbuzynskiy, S.O., 2017. There and back again: Two views on the protein folding puzzle. *Physics of Life Reviews*. (in press).

Figure Legends

Figure 1: Multi-dimensional analogue of the Aune and Tanford denaturant site binding model for explicit inclusion of denaturant within the two-state protein unfolding reaction equilibria. (A) Schematic of the site binding model in which a particular sequence of denaturant equilibria is displayed with each equilibria governed by a particular site binding constant described using a lower case k (either k_{1F}, k_{1U} or k_{2U}). Blue and yellow parts of the polypeptide chain represent external (solvent exposed) and internal (hidden from solvent) segments. Red dots represent denaturant molecules bound to a region of the polypeptide chain that initially exists in an external region. Black dots represent denaturant molecules bound to polypeptide regions not initially exposed to the solvent. k_{1F}, k_{1U} and k_{2U} respectively represent site binding constants for denaturant binding to initially solvent exposed regions in the folded state, initially solvent exposed regions in the unfolded state and initially internalized regions in the unfolded state. Unfolding constants $(k_{FU})_{m,0}$ represent the site specific equilibrium constants for the unfolding of proteins loaded with m number of denaturant molecules bound to the initially solvent exposed polypeptide regions. (B) Representation of all chemical equilibria considered in the denaturant site binding model of protein unfolding. Folded and unfolded protein states are described in shorthand using an F or U with two subscripted indices m and n (i.e. $F_{m,0}$ or $U_{m,n}$) respectively referring to the number of denaturant molecules bound to the initially solvent exposed sites or the number bound to sites initially hidden from the solvent in the internal regions of the folded protein. The equilibrium constants featuring an upper case K to describe the equilibrium ratio of concentrations of F, U and D for the particular transitions indicated by the arrow between the subscripted structural states. In this formulation all N_F and N_U sites are considered as chemically equivalent and indistinguishable with respect to their location within their particular structural state.

Figure 2: Comparison of output from Aune and Tanford denaturant site-binding model for protein unfolding and its two-dimensional analogue. (A) Standard protein folding transition plot showing the dependence of the fraction of unfolded protein to total (f_U) as a function of the total denaturant concentration, C_D. In this plot the black line is generated using the original Aune and Tanford model (Eqn. 3 and 4) and the red filled circles were created by summing all unfolded protein states populated in the two-dimensional analogue (Eqns. 5 to 8). Parameters are listed at the end of the figure legend. (B) Non-standard protein folding transition plot generated from the three-dimensional viewpoint afforded by Eqns. 5-8. Plots of the base 10 logarithm of the concentration of protein states, $\log_{10}(C_{P_{i,j}})$, were made against

axes corresponding to the number of denaturant molecules bound to the accessible binding sites particular to the folded F and unfolded U states. A red dot was assigned to the concentration maxima for each distribution existing at a certain denaturant concentration. These distributions were then overlaid (without replacement) to create the landscape shown. Importantly, the apex of each distribution has its own red dot. (C) A planar heat map representation of the overlay transition plot shown in (B). Colour bar on the RHS describes the $\log_{10}(C_{P_{i,j}})$ above a cut off threshold of -10. Red dots as for (B). (D) An ensemble of unfolded states corresponding to the topmost denaturant concentration shown by a blue ring in (A). Base parameters for the simulation: $(k_{FU})_{0,0} = 1\times10^{-9}$; $k_{1U} = k_{2U} = k_{1F} = 0.01\ M^{-1}$; $N_{AA} = 150$ amino acids (M_W = 16,500 g/mole) ; $R_D = R_{AA} = 0.4$ nm ; $\alpha = 0.4976$; $\varepsilon = \beta = 0.5$; $(c_p)_{TOT} = 1$ mg/ml.

Figure 3: Effect of change in intrinsic stability reflecting partition between F and U states - $\Delta(\mathbf{k_{FU}})_{\mathbf{0,0}}$. (A) Base simulation case, $(k_{FU})_{0,0} = 1\times10^{-9}$ (red line), is contrasted against simulations reflecting a 1000 fold decrease in protein stability, $(k_{FU})_{0,0} = 1\times10^{-6}$ (blue line) and a 1000 fold increase in protein stability, $(k_{FU})_{0,0} = 1\times10^{-12}$ (green line) with all other parameters kept constant. (B – D) Continual overlay of the multi-dimensional heat map plots describing each ligand loaded ensemble along the transition path shown in A for the cases of (B) $(k_{FU})_{0,0} = 1\times10^{-6}$; (C) $(k_{FU})_{0,0} = 1\times10^{-9}$; and (D) $(k_{FU})_{0,0} = 1\times10^{-12}$. All other non-varied parameters were set to the values shown in Fig.2.

Figure 4: Effect of change in protein size - $\Delta\mathbf{N_{AA}}$. (A) Base simulation case, $N_{AA} = 150$ (red line), is contrasted against simulations reflecting a 2/3 decrease in amino acid number, $N_{AA} = 100$ (blue line) and a 3/2 fold increase, $N_{AA} = 100$ (green line) with all other parameters kept constant. (B – D) Continual overlay of the multi-dimensional heat map plots describing each ligand loaded ensemble along the transition path shown in (A) for the cases of (B) $N_{AA} = 100$; (C) $N_{AA} = 150$; and (D) $N_{AA} = 225$. All other non-varied parameters were set to the values shown in Fig.2. (In the text we discuss the apparent paradoxical increase in stability with a decrease in size).

Figure 5: Effect of change in native state affinity for denaturant - $\Delta\mathbf{k_{1F}}$. (A) Base simulation case, $k_{1F} = 0.01\ M^{-1}$ (red line), is contrasted against simulations reflecting a 5-fold decrease in binding strength, $k_{1F} = 0.002\ M^{-1}$ (blue line) and a 5-fold increase, $k_{1F} = 0.05\ M^{-1}$ (green line) with all other parameters kept constant. (B – D) Continual overlay of the multi-dimensional heat map plots describing each ligand loaded ensemble along the transition path shown in (A) for the cases of (B) $k_{1F} = 0.002\ M^{-1}$; (C) $k_{1F} = 0.01\ M^{-1}$; and (D) $k_{1F} = 0.05\ M^{-1}$. All other non-varied parameters were set to the values shown in Fig.2.

Figure 6: Effect of change in unfolded state affinity for denaturant at initially solvent exposed sites - $\Delta\mathbf{k_{1U}}$. (A) Base simulation case, $k_{1U} = 0.01\ M^{-1}$ (red line), is contrasted against simulations reflecting a five-fold decrease in denaturant binding strength, $k_{1U} = 0.002\ M^{-1}$ (blue line) and a five-fold increase, $k_{1U} = 0.05\ M^{-1}$ (green line) with all other parameters kept constant. (B – D) Continual overlay of the multi-dimensional heat map plots describing each ligand loaded ensemble along the transition path shown in (A) for the cases of (B) $k_{1U} = 0.002\ M^{-1}$; (C) $k_{1U} = 0.01\ M^{-1}$; and (D) $k_{1U} = 0.05\ M^{-1}$. All other non-varied parameters were set to the values shown in Fig.2.

Figure 7: Effect of change in unfolded state affinity for denaturant at initially internalized, non-solvent exposed sites - $\Delta\mathbf{k_{2U}}$. (A) Base simulation case, $k_{2U} = 0.01\ M^{-1}$ (red line), is contrasted against simulations reflecting a five-fold decrease in denaturant binding strength, $k_{2U} = 0.002\ M^{-1}$ (blue line) and a five-fold increase, $k_{2U} = 0.05\ M^{-1}$ (green line) with all other parameters kept constant. (B – D) Continual overlay of the multi-dimensional heat map plots describing each ligand loaded ensemble along the transition path shown in (A) for the cases of (B) $k_{2U} = 0.002\ M^{-1}$; (C) $k_{2U} = 0.01\ M^{-1}$; and (D) $k_{2U} = 0.05\ M^{-1}$. All other non-varied parameters were set to the values shown in Fig.2.

App. 4 - Fig. 1

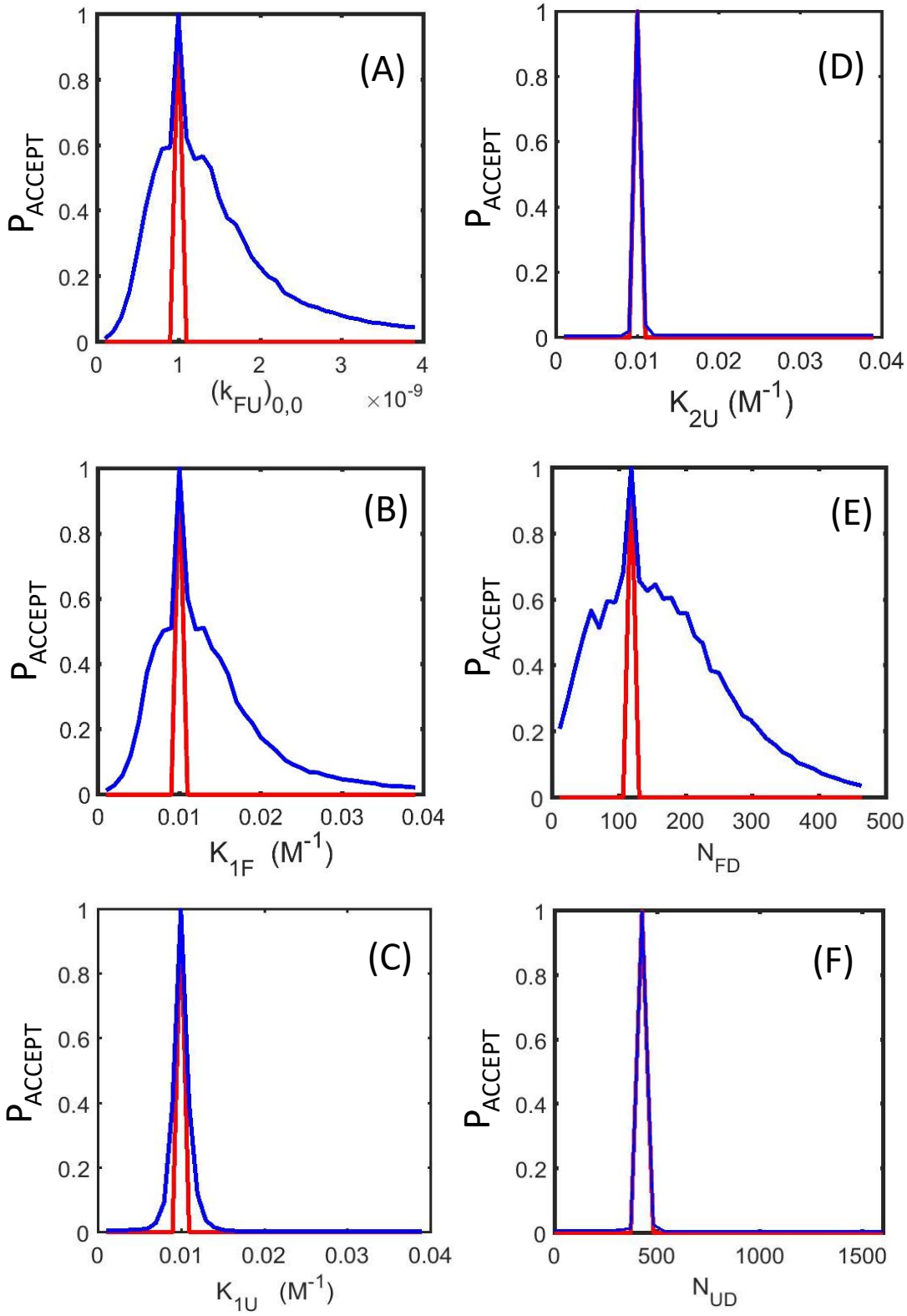
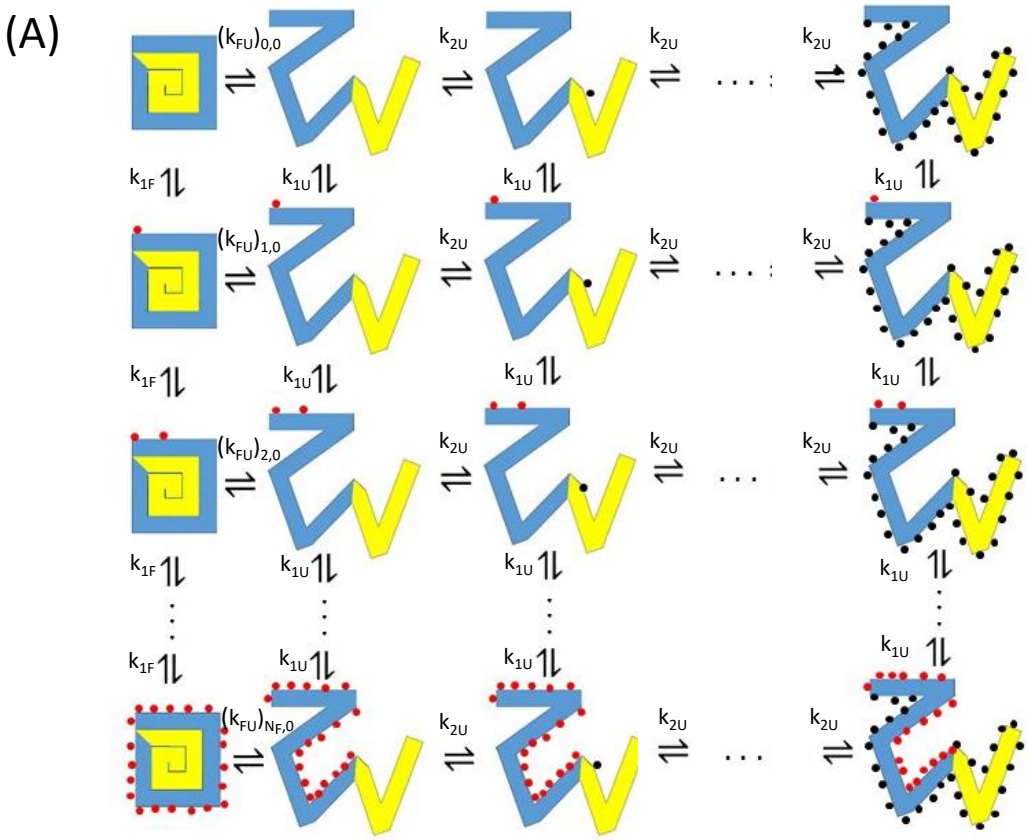


Fig. 1


(A)
$(k_{FU})_{0,0}$
$(k_{FU})_{1,0}$
$(k_{FU})_{2,0}$
$(k_{FU})_{N_F,0}$
k_{1F}
k_{1U}
k_{2U}

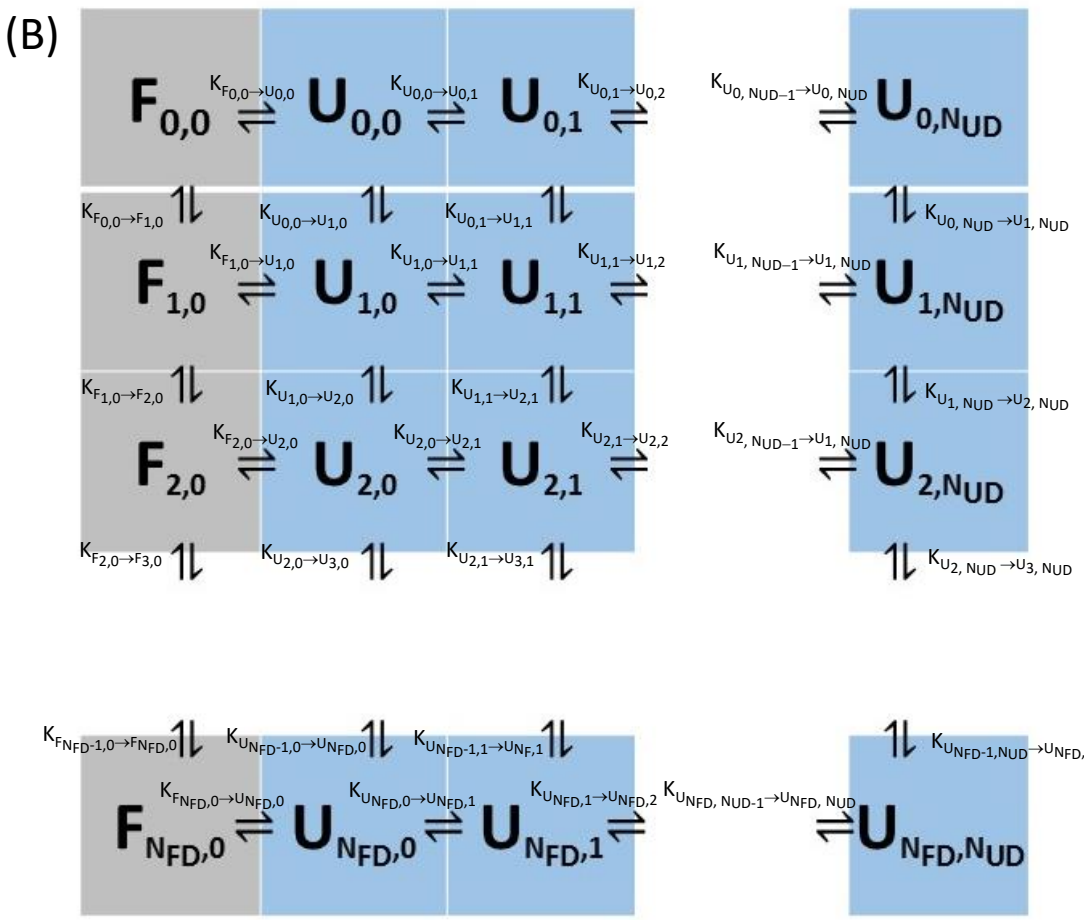
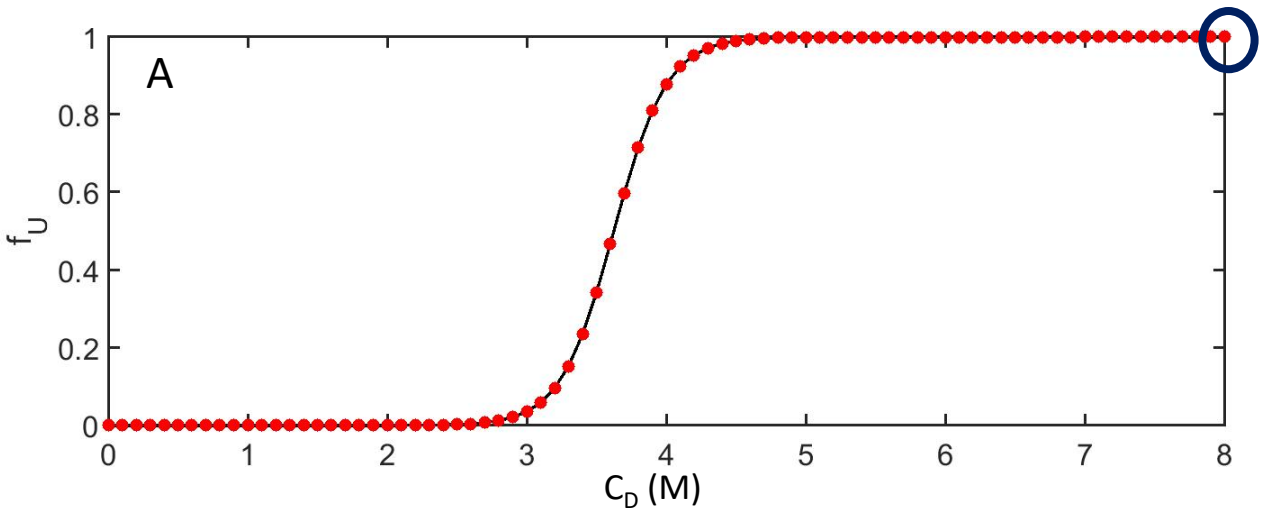
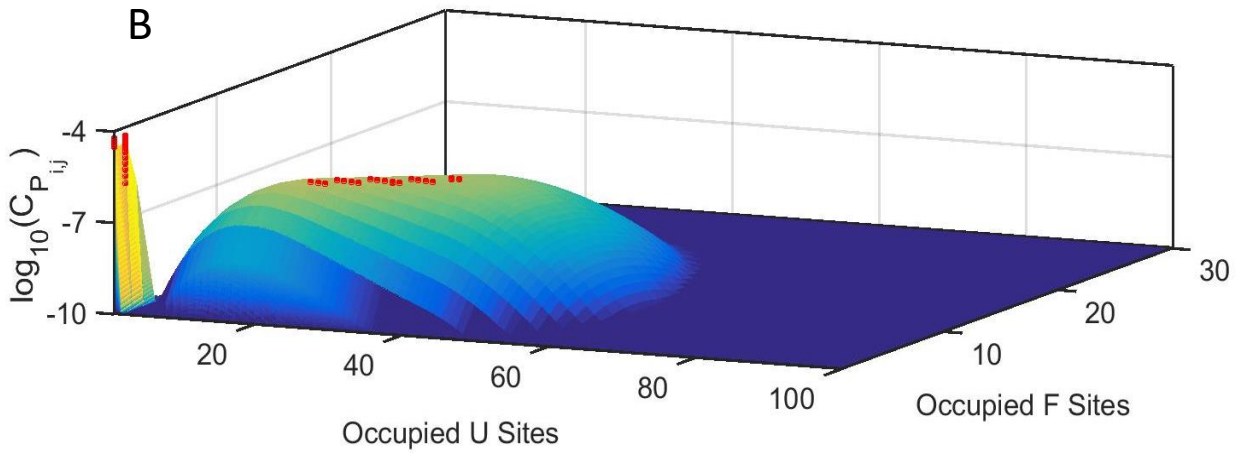
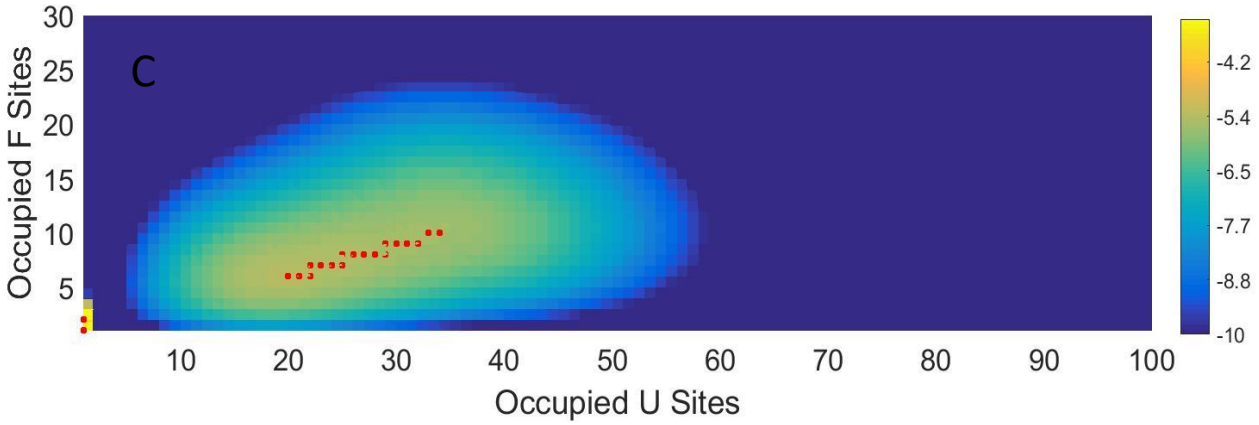

(B)
$F_{0,0}$ $U_{0,0}$ $U_{0,1}$ $U_{0,N_{UD}}$
$F_{1,0}$ $U_{1,0}$ $U_{1,1}$ $U_{1,N_{UD}}$
$F_{2,0}$ $U_{2,0}$ $U_{2,1}$ $U_{2,N_{UD}}$
$F_{N_{FD},0}$ $U_{N_{FD},0}$ $U_{N_{FD},1}$ $U_{N_{FD},N_{UD}}$
$K_{F_{0,0}\rightarrow F_{1,0}}$
$K_{U_{0,0}\rightarrow U_{1,0}}$
$K_{U_{0,1}\rightarrow U_{1,1}}$
$K_{U_{0,N_{UD}-1}\rightarrow U_{0,N_{UD}}}$
$K_{U_{0,N_{UD}}\rightarrow U_{1,N_{UD}}}$
$K_{F_{1,0}\rightarrow F_{2,0}}$
$K_{U_{1,0}\rightarrow U_{2,0}}$
$K_{U_{1,1}\rightarrow U_{2,1}}$
$K_{U_{1,N_{UD}-1}\rightarrow U_{1,N_{UD}}}$
$K_{U_{1,N_{UD}}\rightarrow U_{2,N_{UD}}}$
$K_{F_{2,0}\rightarrow F_{3,0}}$
$K_{U_{2,0}\rightarrow U_{3,0}}$
$K_{U_{2,1}\rightarrow U_{3,1}}$
$K_{U_{2,N_{UD}-1}\rightarrow U_{2,N_{UD}}}$
$K_{U_{2,N_{UD}}\rightarrow U_{3,N_{UD}}}$
$K_{F_{N_{FD}-1,0}\rightarrow F_{N_{FD},0}}$
$K_{U_{N_{FD}-1,0}\rightarrow U_{N_{FD},0}}$
$K_{U_{N_{FD}-1,1}\rightarrow U_{N_F,1}}$
$K_{U_{N_{FD}-1,N_{UD}}\rightarrow U_{N_{FD},N_{UD}}}$
$K_{F_{N_{FD},0}\rightarrow U_{N_{FD},0}}$
$K_{U_{N_{FD},0}\rightarrow U_{N_{FD},1}}$
$K_{U_{N_{FD},1}\rightarrow U_{N_{FD},2}}$
$K_{U_{N_{FD},N_{UD}-1}\rightarrow U_{N_{FD},N_{UD}}}$

Fig. 2


A
f_U
C_D (M)


B
$\log_{10}(C_{P_{i,j}})$
Occupied U Sites
Occupied F Sites


C
Occupied F Sites
Occupied U Sites

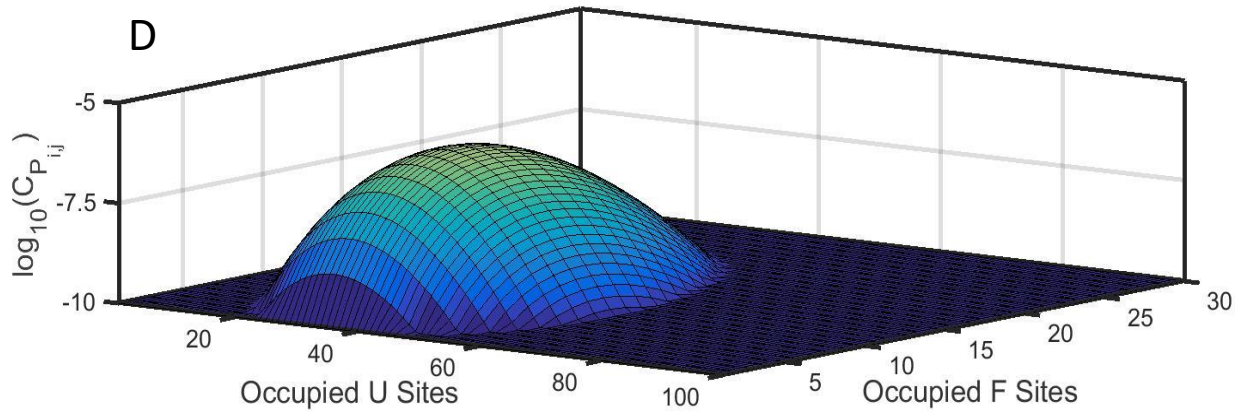

D
$\log_{10}(C_{P_{i,j}})$
Occupied U Sites
Occupied F Sites

Fig. 3

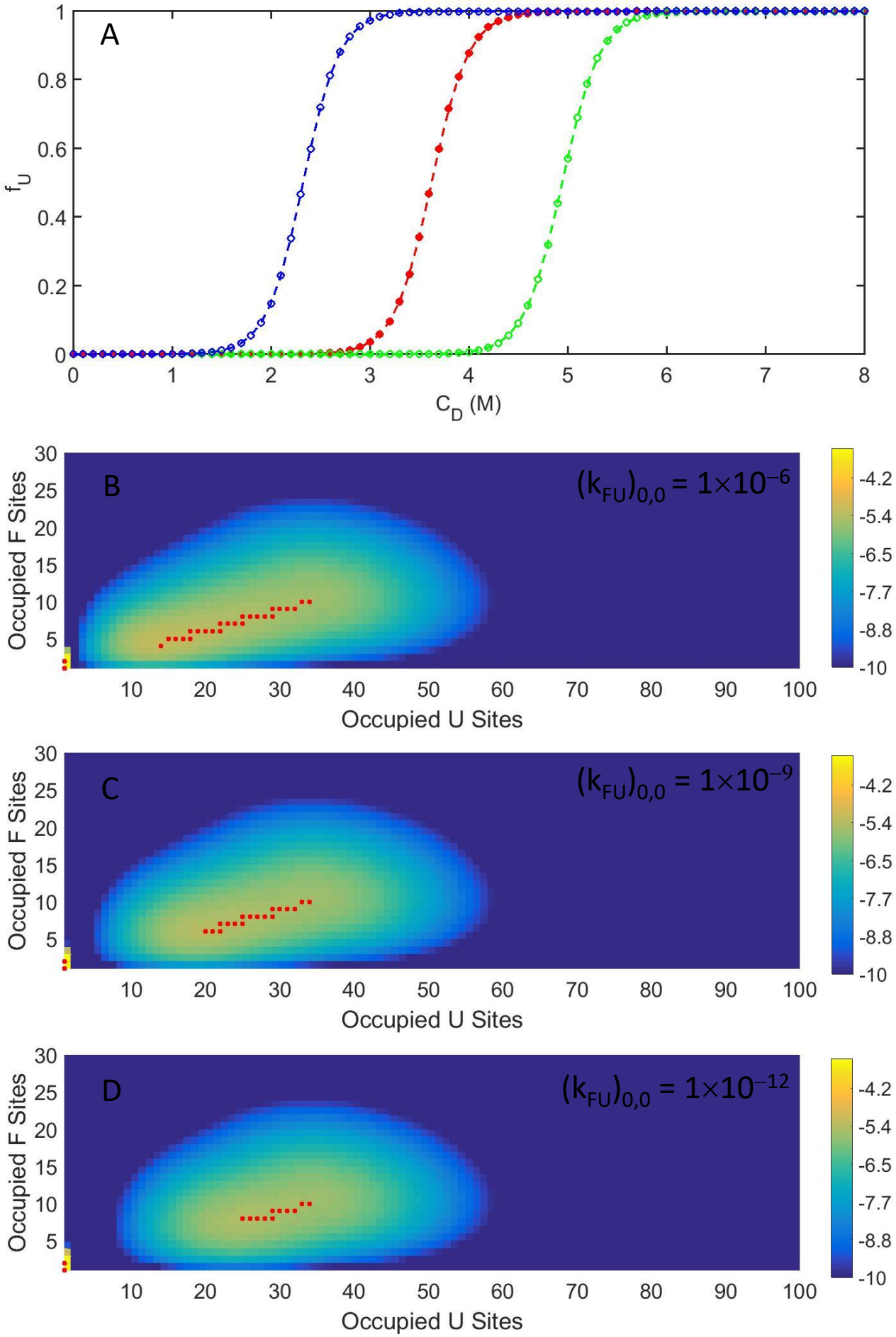

A
f_U
C_D (M)
B
$(k_{FU})_{0,0} = 1\times10^{-6}$
C
$(k_{FU})_{0,0} = 1\times10^{-9}$
D
$(k_{FU})_{0,0} = 1\times10^{-12}$
Occupied F Sites
Occupied U Sites

Fig. 4

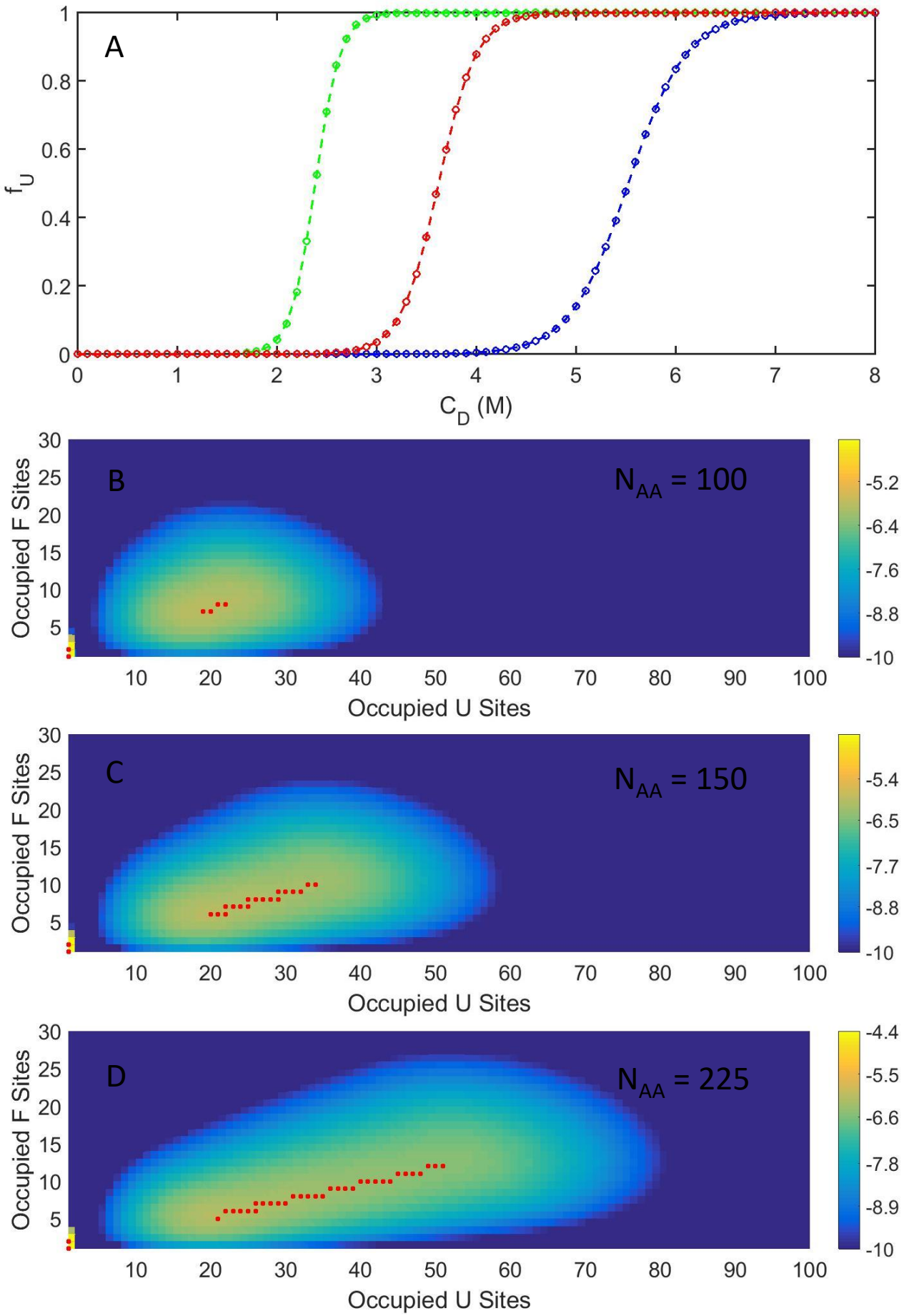

A
f_U
C_D (M)
B
N_{AA} = 100
C
N_{AA} = 150
D
N_{AA} = 225
Occupied F Sites
Occupied U Sites

Fig. 5

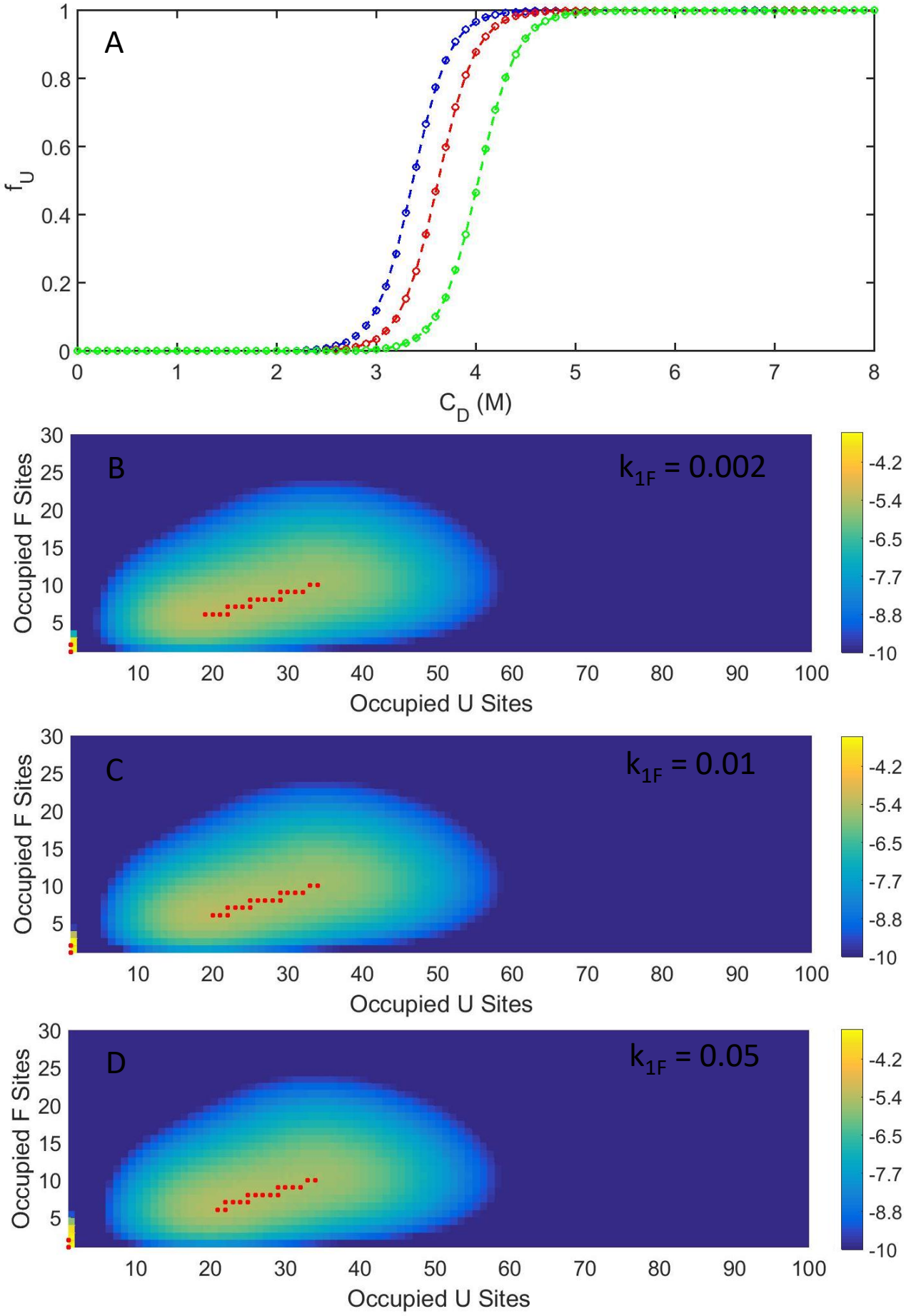
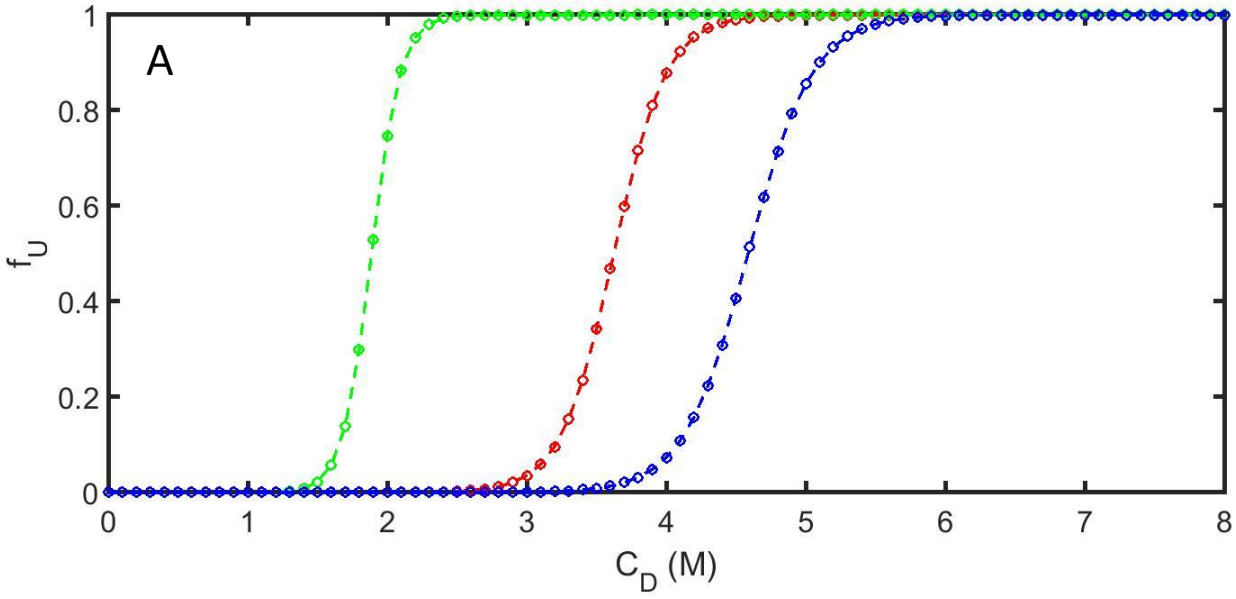
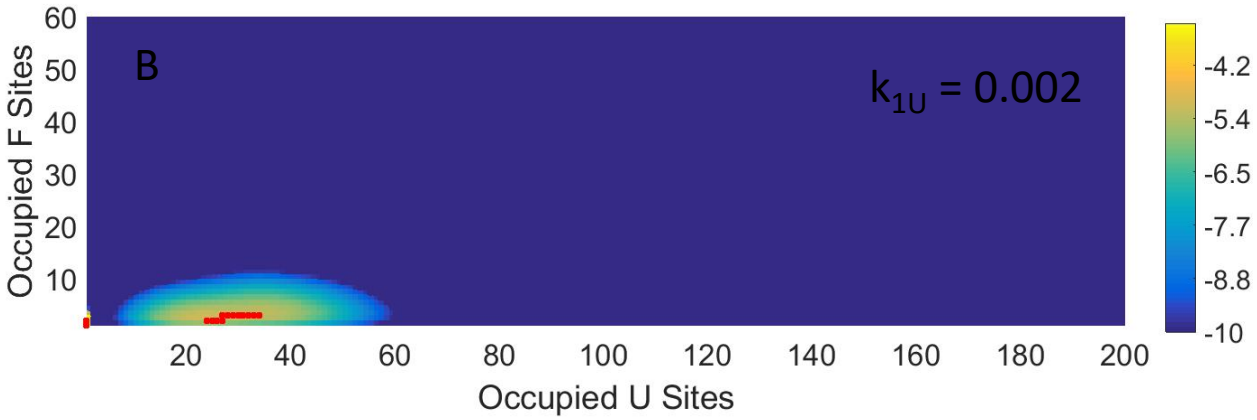
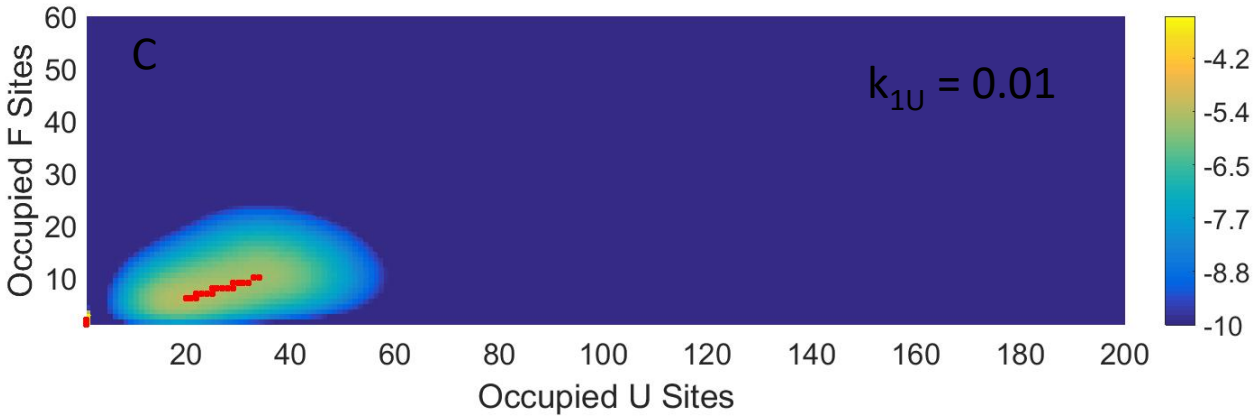

A
f_U
C_D (M)
B
k_{1F} = 0.002
C
k_{1F} = 0.01
D
k_{1F} = 0.05
Occupied F Sites
Occupied U Sites

Fig. 6


A
f_U
C_D (M)


B
k_{1U} = 0.002
Occupied F Sites
Occupied U Sites


C
k_{1U} = 0.01
Occupied F Sites
Occupied U Sites

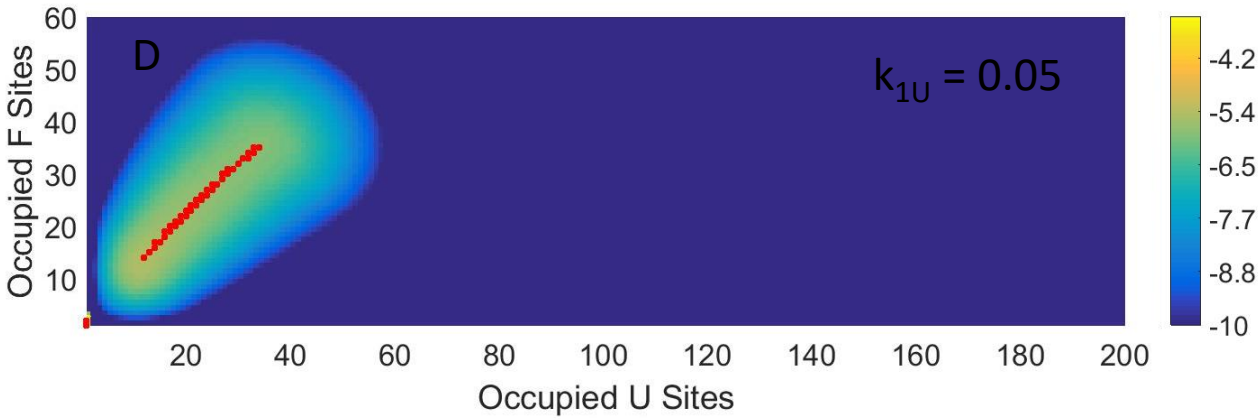
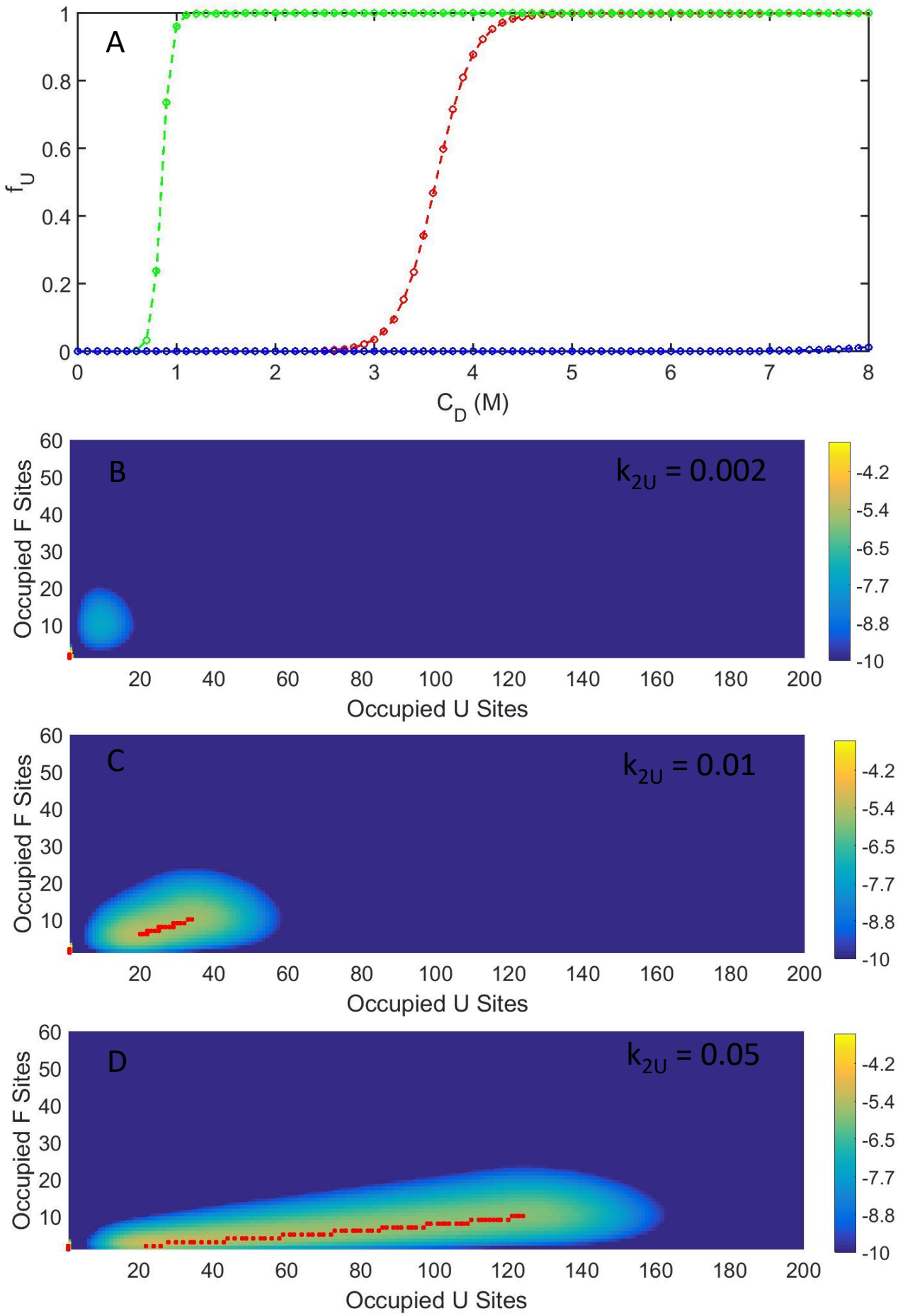

D
k_{1U} = 0.05
Occupied F Sites
Occupied U Sites

Fig. 7


A
f_U
C_D (M)
B
k_{2U} = 0.002
C
k_{2U} = 0.01
D
k_{2U} = 0.05
Occupied F Sites
Occupied U Sites