Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium

Malin J. Allert and Homme W. Hellinga

Department of Biochemistry, *Duke University Medical Center, Box 3711, Durham, NC, 27710, USA*

Correspondence to *Homme W. Hellinga. Fax: +919 684-8885. hwh@biochem.duke.edu.*
https://doi.org/10.1016/j.jmb.2019.12.039
Edited by Monika Fuxreiter

Abstract

One of the great ambitions of structural biology is to describe structure-function relationships quantitatively. Statistical thermodynamics is a powerful, general tool for computing the behavior of biological macromolecules at equilibrium because it establishes a direct link between structure and function. Complex behavior emerges as equilibria of multiple reactions are coupled. Analytical treatment of linked equilibria scales poorly with increasing numbers of reactions and states as the algebraic constructs rapidly become unwieldy. We therefore developed a generalizable, but straightforward computational method to handle arbitrarily complex systems. To demonstrate this approach, we collected a multidimensional fluorescence landscape of an engineered fluorescent glucose biosensor and showed that its features could be modeled with ten intricately linked ligand-binding and conformational exchange reactions. This protein represents a minimalist model of sufficient complexity to encompass fundamental biomolecular structure-function relationships: two-state and multistate conformational ensembles, conformational hierarchies, osmolytes, coupling between different binding sites and coupling between ligand binding and conformational change. The successful fit of this complex, multifaceted system demonstrates generality of the method.

Introduction

Statistical thermodynamics provides direct links between structure and function that are required to describe biomolecular behavior at equilibrium [1–4]. The immense versatility of biomolecular function arises from the exchange between multiple discrete states with distinct structural and functional properties in response to external conditions such as temperature or ligand concentrations (state variables). In statistical thermodynamics, structure is represented by computing the equilibrium populations of conformational states and ligand-binding site occupancies within the ranges of state variable values (phase space). Function is represented by associating numerical attributes such as reaction rates, or biophysical parameters (e.g. fluorescence) with discrete states. Structure-function relationships are then represented by the sum of these functional attributes weighted by the fractional population of their associated state. Complexity emerges by linking equilibria of the multiple reactions that underlie many fundamental processes in biology [5–7], including allosteric control of enzyme activity, conversion between different forms of energy, metabolite transport, and cellular signal transduction.

Classically, linkage relationships are analyzed using analytical treatments [3]. Describing linkage relationships between multiple reactions in this manner does not scale well with complexity because each situation typically involves case-specific constructs that rapidly become algebraically unwieldy as the number of reactions and states increases. We addressed this issue using numerical analysis and developed the nested equilibrated states tree (NEST) method to handle arbitrarily complex linkage relationships. The NEST method establishes generality by recognizing that linkage relationships are hierarchical in character and can be represented naturally by the tree abstract data type used in computer science [8–10]. This treatment enables (a) a straightforward syntax for describing the linkage relationships between reactions and states, (b) their visualization, (c) recursive construction of equation families and (d)

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

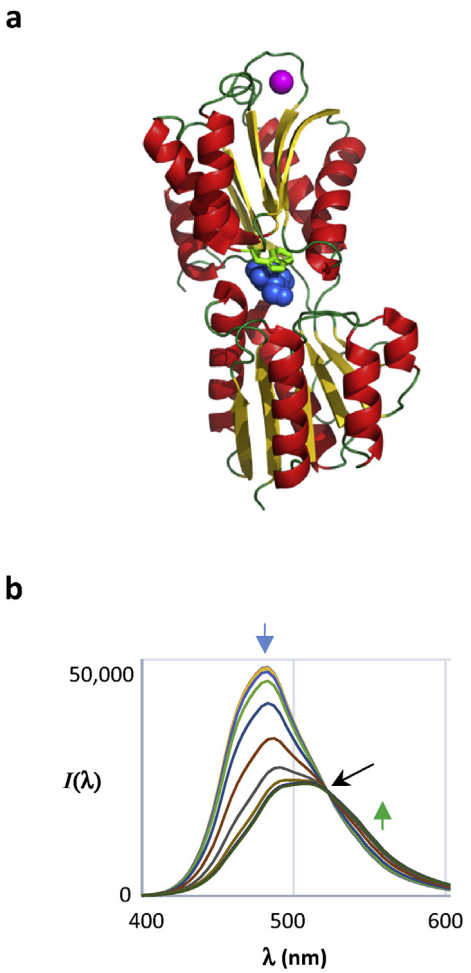


Fig. 1. Fluorescent glucose biosensor. (**a**) *E. coli* glucose-galactose binding protein [15–22] (ecGGBP) is a member of the periplasmic binding protein (PBP) superfamily [52,53]. The PBP fold comprises two α/β domains connected by a two- to three-stranded β hinge. Glucose (blue) binding is accompanied by a conformational change involving a bending motion at the hinge, which brings the two domains together such that ligand is enveloped within the interdomain interface [20–22]. The Ca^{2+}-binding site (magenta) is located in the C-terminal domain, ~30 Å from the glucose-binding pocket [15]. This loop adopts a conformation similar to the "EF-hand" motif. Tryptophan 183 (green) forms an extensive Van der Waals contact with the pyranose ring of the bound glucose. In ecGGBP183C·Acrylodan, W183 has been replaced with cysteine to which the thiol-reactive Acrylodan has been site-specifically conjugated [11]. (**b**) Glucose-dependent change in the emission spectrum of ecGGBP183-C·Acrylodan at 1 mM Ca^{2+}. Upon isothermal binding of glucose, the apoprotein shifts from a predominantly blue ($\lambda_{max} = 483$ nm) to a mixed blue/green emission spectrum (mixture of 483 and 550 nm peaks) with an isosbestic point at 524 nm (black arrow), consistent with conversion between two states (individual spectra shown for 0–500 mM glucose).

recursive computation of the distribution of discrete states. The object-oriented NEST implementation provides a powerful computational tool to model the experimental behavior of complex biomolecules and a common language to communicate their structure-function relationships easily, precisely, and quantitatively.

To demonstrate the capabilities of this method, we use NEST to quantitatively describe the complex experimental behavior of a fluorescent glucose biosensor derived from the *Escherichia coli* glucose-galactose binding protein (ecGGBP) [11], that shows promise for the management of diabetes, having been tested successfully for *in vivo* continuous glucose monitoring in animals [12] and humans [13]. In the ecGGBP183C•Acrylodan fluorescent conjugate, the Acrylodan fluorophore [14] is site-specifically attached to the thiol of the W183C mutant [11] (Fig. 1a). ecGGBP [15–19] binds both glucose and Ca^{2+}. Glucose but not Ca^{2+} binding is accompanied by a conformational change from an open to a closed state [20–22]. In addition to this glucose-mediated motion, protein (un) folding and conformational exchanges in the folded and unfolded state all cause the emission spectrum of the conjugated Acrylodan to switch between blue and green states (Fig. 1b). We have found that this switch is due to a two-state exchange between Acrylodan conformations corresponding to blue- and green-emitting forms, respectively [23,24]. This single fluorophore therefore simultaneously reports on the properties of multiple reactions, thereby providing detailed insights into the behavior of the protein. We used the ratio of the emission intensities at two distinct maxima that represent the blue and green states, respectively to minimize experimental noise [25,26]. We collected a ratiometric, high-precision, multidimensional landscape as a function of glucose, calcium, and temperature, comprising 29,184 data points. We showed that this finely sampled experimental landscape can be modeled by a NEST that encompasses multiple, intricately linked ligand binding, and conformational exchange reactions in the protein.

We interpreted this landscape by iteratively constructing and evaluating NEST models of increasing detail and complexity. The final model comprises four conformational exchange, and six ligand-binding reactions, encompassing 18 discrete states, eight of which exhibit distinct fluorescence signals. The model revealed that this seemingly simple protein encompasses many dominant effects that determine biological structure-function relationships: interactions between different ligand binding sites [5], two-state and multistate conformational ensembles [6], hierarchies of conformational exchange reactions [27,28], conformational coupling [7], and osmolyte effects on protein folding and conformational exchanges [29–32]. This protein therefore is a minimalist model of macromolecular structure-function relationships because it provides a single experimental probe that reports simultaneously on multiple, representative reactions. The successful fit of its complex behavior demonstrates generality of the NEST approach.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Theory

We first outline the NEST method qualitatively (Fig. 2), followed by a summary of the mathematical treatment (equations in Table 1; definition of symbols in Table 2; equations are referred to in bold numbers). The NEST method computes populations of discrete states representing the structure and function (fluorescence in this case) of the system over the entire experimental range of the state variables. The total fluorescence of the system is given by multiplying the intrinsic fluorescence signal associated with an individual state by the fractional population of that state and summing the total signal over all fluorescent states.

States comprise binding site occupancies and protein conformations

The concept of a "discrete state" is central to statistical thermodynamics. For ligand-binding reactions, the discrete states are the binding site occupancies (free or bound). Discrete conformational states are more challenging to define because spatial distributions are inherently continuous, and proteins and nucleic acids assume a large number of nearly isoenergetic conformations comprising correlated motions at multiple length scales [27,33]. Furthermore, ensembles of uncorrelated motions, such as an unfolded protein, or locally unfolded region, also can behave as discrete states, if these are associated with a uniquely identifiable property (e.g. fluorescence). Discrete conformational states therefore need not correspond to single spatial coordinates and are defined both by experimental or theoretical analyses of three-dimensional structures, and from experimentally distinguishable functional properties.

Distributions of states are calculated from pairwise free energies

The statistical thermodynamics of individual reactions are well-understood [1–4]. Within these, the *isolated* population of states is computed from the pairwise free energy of each excited state relative to a single ground state. In a two-state reaction, the ground state is the reactant, and the excited state is the product. For instance, in a protein (un)folding conformational reaction, $\mathbf{F} \rightleftharpoons \mathbf{U}$, ***F*** (folded) is the ground state, ***U*** (unfolded) the excited state (Fig. 2a); in a single-ligand binding reaction, $\mathbf{P} + \mathbf{L} \rightleftharpoons \mathbf{P{\cdot}L}$, the ground state is the apo protein, ***P***, and the excited state the ligand-bound complex, $\mathbf{P{\cdot}L}$ (Fig. 2b). In multistate conformational ensembles, $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_1 \rightleftharpoons \cdots \rightleftharpoons \mathbf{S}_{\mathbf{N}-1}$, such as used to describe thermal unfolding of the ecGGBP183C·Acrylodan conjugate, the free energy of each excited state, $\mathbf{S_i}$, is described by its pairwise relationship, $\mathbf{S}_0 \rightleftharpoons \mathbf{S_i}$, to a single ground state, (Fig. 2c).

State variables determine distributions of states

Free energies change in response to the effects of state variables on the intrinsic behavior of a reaction (ligand concentration on binding; osmolytes on conformational exchange reactions; temperature on both). All the states within a reaction are in exchange with each other: if the pairwise free energy of one state is altered, the fractional populations of all states change (redistribute).

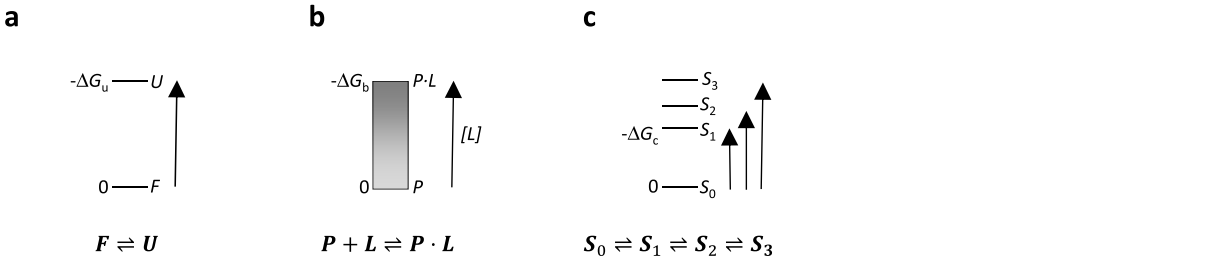
Ligand-binding reactions alter distributions of conformational states

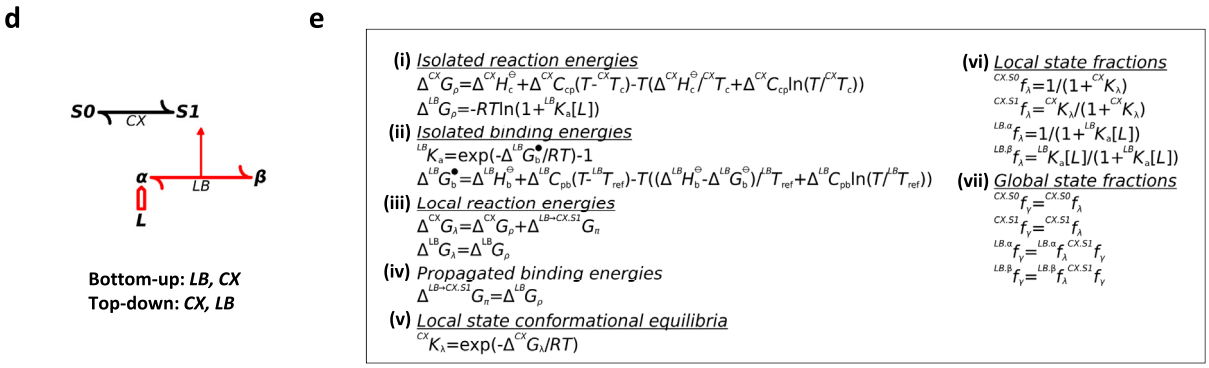
Pairwise free energies in conformational exchange reactions also can be altered through linkage to ligand-binding reactions. If a conformational state binds a ligand, then the distribution of all states in the conformational exchange reaction becomes responsive to the concentration of that ligand (i.e. the conformational exchange and ligand-binding reactions are linked through "conformational coupling"). This effect arises because binding of a ligand to a state in the macromolecule removes a particle from the surrounding "bath" of particles, thereby decreasing the inequality of the number of particles between these two compartments in the thermodynamic system. This equalization of particle numbers (material balance) is an "entropic driving force" that determines the equilibrium [1,2]. The population of the states within a conformational exchange reaction that has one or more of its states linked to a ligand-binding reaction therefore redistributes to maximize conformational states for which the linked ligand-binding reaction is in its ligand-bound state.

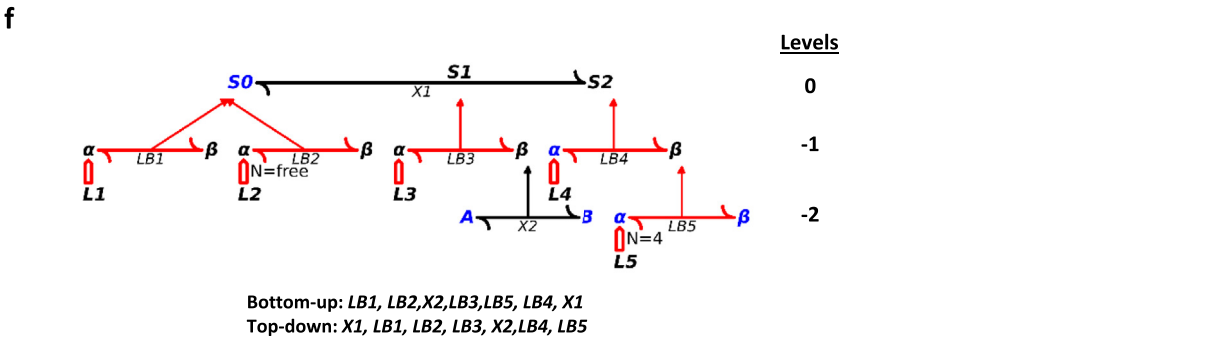
The redistribution of states is calculated using parent-child hierarchies

The "entropic driving force" for conformational coupling corresponds to the ligand-dependent free energy of the ligand-binding reaction. To model linkage between a ligand-binding and a conformational exchange reaction, we invoke the concept of a hierarchical relationship between the two reactions: for a "child" ligand-binding reaction linked to ("embedded in") a "parent" state in a conformational exchange reaction, the child ligand–binding free energy is added ("propagated") to the parent conformational free energy (Fig. 2d and e). This propagation alters the pairwise free energy of the parent state relative to its ground state (or that of all excited states, if the parent is the ground state). This alteration of the pairwise free energies initiates recalculation of the distribution of all the conformational states, redistributing these from their *isolated* fractional populations into their *local* distribution, thus taking into account the influence of ligand

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039







g

		Free Energies	Equilibrium constants	Partition function	State fractions
Conformational exchange	Two-state	1	2	$1+K_c$	3
	Multi-state	4	5	6	7
Ligand binding		9, 11, 12	10	8	13, 14

Fig. 2. NEST representation of coupled equilibria. The free energies of states are defined pairwise relative to a ground state: (a) two-state conformational exchange reaction; (b) ligand-binding free energy increases continuously with the logarithm of the concentration [6]; (c) multistate conformational ensemble. (d) Linkage of a conformational exchange reaction (*CX*) with a ligand-binding reaction (*LB*), that is, conformational coupling [6,7], is set up by embedding the child *LB* reaction into the parent state S_1 in *CX* (setup script shown in Fig. S1); (e) The NEST program automatically generates the equations for computing the global distribution of states in d (Table 2 for definition of symbols). These equations instantiate the general equations in Table 1, grouped as follows: (i) **1** and **9**; (ii) **10**–**12**; (iii) **15**; (iv) **20** and **21**; (v) **16** and **17**; (vi) **18** and **19**; (vii) **27**. The ligand-binding free energy of *LB* propagates into *CX* via $\boldsymbol{S_1}$, changing its isolated free energy to a local free energy, redistributing the states from isolated to local fractions (using the local equilibrium constant as a calculation

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

binding on the conformational exchange. This redistribution in turn alters the total population of all the states in the child reaction within the system considered as a whole (without affecting their internal relative distribution) because the global population of these states is directly proportional to the local population of the parent state within which the child reaction is embedded. This hierarchical parent/child representation of the linkage relationship between a conformational reaction and a ligand-binding reaction therefore comprises two components: redistributive effects in which propagation of the child ligand−binding free energy alters the parent distribution of states and hierarchical effects in which the redistributed fractional population of the parent state determines the total fraction of the child states in the entire system.

Redistributive effects apply only to linkage relationships between conformational exchange and ligand-binding reactions. Hierarchical effects apply to linkage relationships between all pairs of reaction types. For instance, conformational exchange child reactions can be embedded either within conformational or ligand occupancy parent states to represent their subdivision into two or more conformations. This linkage relationship was used in ecGGBP183-C·Acrylodan to represent multiple conformations of its glucose complex. Ligand-binding reactions can be embedded within other ligand occupancy states to represent conditional binding events. This device was used in ecGGBP183C·Acrylodan to distinguish glucose binding in Ca^{2+}-bound and Ca^{2+}-free forms of the protein.

Parent-child hierarchies are represented with trees

The NEST methodology generalizes linkage relationships by taking advantage of recursion operations inherent to the tree abstract data type used to represent hierarchies in computer science [8−10]. In a NEST, nodes represent individual reactions; branches connected to states in a reaction represent embedded linkage relationships between a parent state and its child reaction(s) at a lower level in the tree (Fig. 2d−f, fig. S1 and S2). Distributive effects therefore are propagated upward along these connections. The resulting hierarchies can be arbitrarily complex; distributive effects therefore may propagate through several levels of reactions.

The distribution of states is calculated by three passes through a tree

The final distribution of states is computed by executing three consecutive passes through the reaction nodes in the NEST (Fig. 2d−f), successively calculating isolated, local, and global state distributions. In the first pass, each node is visited in arbitrary order, and pairwise free energy descriptions of individual conformational exchange or ligand-binding reactions are used to compute isolated fractional state distributions. These calculations transform free energies via equilibrium constants into state fractions (Fig. 2g). In two-state conformational exchange reactions, the free energy of the product relative to its reactant is described by the Gibbs-Helmholtz relationship (**1**). This free energy is converted into an equilibrium constant (**2**) from which the isolated state distribution is calculated (**3**). For multistate ensembles, individual pairwise Gibbs-Helmholtz relationships (**4**) encode the pairwise excited state free energy differences to a ground state; from these, pairwise equilibrium constants are calculated (**5**). These constants are summed to form a partition function (**6**) to calculate the *isolated* distribution of states (**7**). Binding polynomials (**8**) compute ligand concentration−dependent free energies of ligand-binding reactions (**9**). The temperature-dependent ligand affinity (**10**) is computed from the binding free energy at the standard state 1 M ligand concentration using a Gibbs-Helmholtz relationship (**11**), given a reference free energy at a standard temperature (**12**). The fractional occupancy (**13**) gives the *isolated* fractions of ligand-bound (**13**) and ligand-free states (**14**).

In the second pass, nodes are visited in bottom-up order (Fig. 2d−f), and the isolated state distributions of conformational exchange reactions are redistributed as local fractions by propagated ligand-binding free energies that originate from child reaction(s) at

intermediate). (f) NEST representation of multiple linked reactions (setup script and equations shown in Fig. S2). Reactions (black, conformational exchange; red, ligand binding) are shown by horizontal, named equilibrium arrows that connect named states in the reaction. Naming in ligand-binding reactions: α, apo state; β, ligand-bound state; ligand names are shown below α; for multiple, independent sites, the number of sites is indicated (if this is allowed to vary in a data fit, it is labeled "free"). Upward pointing vertical or slanted single arrows indicate embedded child→parent relationships in coupled equilibria. Red connections denote propagated ligand-binding energies. *X1* is an *n*-ary node (multiple branches); the others are binary (two branches). Blue states are associated with fluorescent signals. The *X1* reaction is conformationally coupled to all ligand-binding reactions. The hierarchical conformational exchange reaction *X2* is confined within $\boldsymbol{S_1}$ and its bound state of *LB3*. *LB1* and *LB2* are parallel reactions embedded within $\boldsymbol{S_0}$ (the ground state of *X1*). *L5* binds only in the presence of *L4* because *LB5* is embedded within the bound state of *LB4* (conditional binding). (g) The flow of calculations that convert free energies into state fractions in conformational exchange and ligand-binding reactions (numbers refers to equations in Table 1). NEST, nested equilibrated states tree.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Table 1. Equations used to build NEST models[a].

Equation number	Equation	Comments
1[b]	$\Delta G_c = \Delta H_c^{\ominus} + \Delta C_{pc}(T - T_c^{\ominus}) - T((\Delta H_c^{\ominus}/T_c^{\ominus}) + \Delta C_{pc}\ln(T/T_c^{\ominus})) + \sum_i m_i[O_i]$	Two-state conformational exchange: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_1$
2	$K_c = \exp(-\Delta G_c/RT)$	
3	${}^{\mathbf{S}0}f_\rho = 1/(1+K_c)$, ${}^{\mathbf{S}1}f_\rho = K_c/(1+K_c)$	
4	$\Delta^{\boldsymbol{\sigma}} G_c = \Delta^{\boldsymbol{\sigma}} H_c^{\ominus} + \Delta^{\boldsymbol{\sigma}} C_{pc}(T - {}^{\sigma}T_c^{\ominus}) - T((\Delta^{\boldsymbol{\sigma}} H_c^{\ominus}/{}^{\sigma}T_c^{\ominus}) + \Delta^{\boldsymbol{\sigma}} C_{pc}\ln(T/{}^{\sigma}T_c^{\ominus})) + \sum_i {}^{\sigma}m_i[O_i]$	Multistate conformational exchange: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_1 \ldots \mathbf{S}_0 \rightleftharpoons \mathbf{S}_{N-1}\boldsymbol{\sigma} \in \{\mathbf{S}_1, \cdots, \mathbf{S}_{N-1}\}$
5	${}^{\sigma}K_c = \exp(-\Delta^{\boldsymbol{\sigma}} G_c/RT)$	
6	$Q_c = 1 + \sum_{\boldsymbol{\sigma}} {}^{\boldsymbol{\sigma}}K_c$	
7	${}^{\mathbf{S}0}f_\rho = 1/Q_c$, ${}^{\sigma}f_\rho = {}^{\sigma}K_c/Q_c$	
8	$Q_b = (1 + K_a L)^n$	Independent ligand-binding sites
9	$\Delta G_b = -RT\ln Q_b$	
10	$K_a = \exp(-\Delta G_b^{\bullet}/RT)$	
11	$\Delta G_b^{\bullet} = H_b^{\ominus} + \Delta C_{pb}(T - T_b^{\ominus}) - T((\Delta H_b^{\ominus} - \Delta G_b^{\ominus})/T_b^{\ominus} + \Delta C_{pb}\ln(T/T_b^{\ominus}))$	
12	$\Delta G_b^{\ominus} = -RT_b^{\ominus}\ln(1+1/K_d^{\ominus}) \approx RT_b^{\ominus}\ln K_d^{\ominus}$	
13	$\overline{y} = {}^{\beta}f_\rho = K_a L/(1+K_a L)$	
14	${}^{\alpha}f_\rho = 1 - {}^{\beta}f_\rho$	
15[c]	$\Delta^{\sigma} G_{c,\lambda} = \Delta^{\sigma} G_c + \sum_{c(\boldsymbol{\sigma})} \Delta^{c(\sigma)\rightarrow\sigma} G_\pi - \sum_{c(\mathbf{S}0)} \Delta^{c(\mathbf{S}0)\rightarrow\mathbf{S}0} G_\pi$	Perturbation of conformational exchange reactions by propagated ligand-binding energies from embedded reactions. $\boldsymbol{\sigma} \in \{\mathbf{S}_1, \cdots, \mathbf{S}_{N-1}\}$
16	${}^{\sigma}K_{c,\lambda} = \exp(-\Delta^{\boldsymbol{\sigma}} G_{c,\lambda}/RT)$, ${}^{\mathbf{S}0}K_{c,\lambda} = 1$	
17	$Q_{c,\lambda} = 1 + \sum_{\boldsymbol{\sigma}} {}^{\boldsymbol{\sigma}}K_{c,\lambda}$	
18	${}^{\sigma}f_\lambda = {}^{\sigma}K_{c,\lambda}/Q_{c,\lambda}$ ${}^{\mathbf{S}0}f_\lambda = 1/Q_{c,\lambda}$	
19[d]	${}^{\alpha}f_\lambda = {}^{\mathbf{a}}f_\rho$, ${}^{\beta}f_\lambda = {}^{\beta}f_\rho$	Unperturbed ligand-binding state fractions
20[e]	$\Delta^{c(\mathbf{p})\rightarrow\mathbf{p}} G_\pi = \sum_{\boldsymbol{\sigma}\in c(\mathbf{p})} \left({}^{\sigma}f_\lambda \sum_{c(\boldsymbol{\sigma})} \Delta^{c(\boldsymbol{\sigma})\rightarrow\boldsymbol{\sigma}} G_\pi \right)$	Recursive weighted free energy propagation
21[f]	$\Delta^{c(\mathbf{p})\rightarrow\mathbf{p}} G_\pi = {}^{\alpha}f_\lambda \sum_{c(\boldsymbol{\alpha})} \Delta^{c(\boldsymbol{\alpha})\rightarrow\boldsymbol{\alpha}} G_\pi + {}^{\beta}f_\lambda \sum_{c(\boldsymbol{\beta})} \Delta^{c(\boldsymbol{\beta})\rightarrow\boldsymbol{\beta}} G_\pi + \Delta G_b$	
22	$\Delta H_{c,\rho} = \sum_{\boldsymbol{\sigma}} H_c^{\ominus} + \Delta^{\boldsymbol{\sigma}} C_{pc}(T - {}^{\boldsymbol{\sigma}}T_c^{\ominus})$	Recursive weighted enthalpy propagation in conformational exchange reactions $\boldsymbol{\sigma} \in \{\mathbf{S}_1, \cdots, \mathbf{S}_{N-1}\}$
23	$\Delta^{\boldsymbol{\sigma}} H_{c,\lambda} = \Delta^{\boldsymbol{\sigma}} H_{c,\rho} + \sum_{c(\boldsymbol{\sigma})} \Delta^{c(\boldsymbol{\sigma})\rightarrow\boldsymbol{\sigma}} H_\pi - \sum_{c(0)} \Delta^{c(\mathbf{S}0)\rightarrow\mathbf{s}} H_\pi$	
24	$\Delta^{c(\mathbf{p})\rightarrow\mathbf{p}} H_\pi = \sum_{\boldsymbol{\sigma}\in c(\mathbf{p})} \left({}^{\sigma}f_\lambda \sum_{c(\boldsymbol{\sigma})} \Delta^{c(\boldsymbol{\sigma})\rightarrow\boldsymbol{\sigma}} H_\pi \right)$	
25[g]	$\Delta H_{b,\lambda} = n\overline{y}(\Delta H_b^{\ominus} + \Delta C_{pb}(T - T_b^{\ominus}))$	Recursive weighted enthalpy propagation in ligand-binding reactions
26	$\Delta^{c(\mathbf{p})\rightarrow\mathbf{p}} H_\pi = {}^{\alpha}f_\lambda \sum_{c(\boldsymbol{\alpha})} \Delta^{c(\boldsymbol{\alpha})\rightarrow\boldsymbol{\alpha}} H_\pi + {}^{\beta}f_\lambda \sum_{c(\boldsymbol{\beta})} \Delta^{c(\boldsymbol{\beta})\rightarrow\boldsymbol{\beta}} H_\pi + \Delta H_{b,\lambda}$	
27	${}^{\sigma}f_\gamma = {}^{\sigma}f_\lambda{}^{p(\sigma)} f_\gamma$	Recursive global state populations
28	${}^{\sigma}R = {}^{\sigma}R_0 + \sum_i V_i(\partial^{\boldsymbol{\sigma}} R/\partial V_i)$	Fluorescence signals
29	$R_{tot} = \sum_{\boldsymbol{\sigma}} {}^{\boldsymbol{\sigma}}f_\gamma {}^{\sigma}R$	
30[h]	$K_c = \exp -((\Delta^{app} H_c/RT)((1/T) - (1/{}^{app}T_c)))$	van't Hoff analysis
31	$dR_{tot}/dT = {}^{\mathbf{S}1}fT^2(1 - {}^{\mathbf{S}1}f)A$	
32[i]	$r_{ij} = (\sum(p_i - \overline{p_i})(p_j - \overline{p_j}))/(\sqrt{\sum(p_i - \overline{p_i})^2}(p_j - \overline{p_j})^2)$	Pearson's correlation coefficient between parameters

lower level(s) in the NEST. We developed recursive equations to calculate these propagated energies, which are key elements in the NEST method. For conformational exchange reactions, bottom-up propagation of binding reaction free energies into a parent state (**15**) alters the equilibrium constant of that parent state (**16**), and the partition function of the reaction (**17**), from which the local state distribution is calculated (**18**). The local distribution of states in ligand-binding reactions is not altered by embedded binding reactions (**19**).

Intermediate-level conformational exchange reactions propagate embedded ligand-binding free energies upward to their parent state, weighted by the local fraction of that state (**20**). Intermediate-level ligand-binding reactions also propagate embedded ligand-binding free energies weighted by population of an entrant state and

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Table 1. (*continued*)

Equation number	Equation	Comments
33	$F_{2\|1} = ((S_1 - S_2)/S_2)/((n_2 - n_1)/n_2)$	Model comparison using extra sum of squares principle
34	$p_{2\|1} = (1/\sqrt{2\pi})\exp((F_{2\|1} - 1)/2)$	

NEST, nested equilibrated states tree.

[a] See Table 2 for definition of symbols. Equations **1**-**14** describe the statistical thermodynamics of individual reactions. **15**–**29** have been developed specifically for the NEST recursive algorithms, their weighted energy propagations, and treatment of attributes associated with a NEST state. **30** and **31** are used to fit experimental data to a two-state van't Hoff approximation of a conformational exchange reaction to extract apparent transition midpoint ${}^{app}T_c$ and $\Delta^{app}H_c$ values (John, D.M., Weeks, K.M. van't Hoff enthalpies without baseline. *Protein Sci*. **9**, 1416–1419, (2000)), which can be fit to a NEST model The origins of **1**–**3** are described in many sources, including [1]; **4**–**7** are extensions of **1**–**3** to build a multistate conformational ensemble from pairwise relationships between excited and ground states; **8**–**10**, **13**, and **14** originate from standard treatments of isothermal binding polynomials [1]; **11** and **12** are extensions for thermal binding polynomials.

[b] Two-state (un)folding reactions use the following synonyms: ΔG_u, $T_m^{\ominus}$, $\Delta H_u^{\ominus}$, ΔC_{pu}, K_c, ${}^{F}f_\rho$, and ${}^{U}f_\rho$.

[c] The localized pairwise conformational exchange free energy is obtained through readjustment of the isolated free energy, $\Delta^{\sigma}G_u$, by addition of the sum of the ligand-binding energies embedded within the excited state, $\sum_{c(\boldsymbol{\sigma})} \Delta^{c(\boldsymbol{\sigma})\rightarrow\boldsymbol{\sigma}}G_\pi$, given as a sum over the (parallel) child(ren) reaction(s) embedded (if any) within that state, $c(\boldsymbol{\sigma})$, and subtracting the energies of the child(ren) embedded (if any) within the ground state, S_0, of the ensemble, $\sum_{c(\mathbf{S}0)} \Delta^{c(\mathbf{S}0)\rightarrow\mathbf{S}0}G_\pi$. The last term is common to all members in a multistate conformational ensemble.

[d] The distribution of states in ligand-binding reactions are not perturbed by free energy propagation (a perturbation would be tantamount to changing ligand concentration). Accordingly, isolated and local fractions are identical.

[e] Free energy propagation through a conformational exchange reaction. The ligand-binding energy propagated from a conformational exchange child $c(\mathbf{p})$ into its parent state, $\boldsymbol{p}$, from its own embedded (if any) conformational exchange (parallel) child(ren) reaction(s), $c(\mathbf{p})$ is summed over all the states, $\boldsymbol{\sigma}$, in the (parallel) child(ren) reaction(s), $\boldsymbol{\sigma} \in c(\mathbf{p})$.

[f] Free energy propagation through a ligand-binding reaction. The ligand-binding energy propagated from a ligand-binding child into its parent passes on all the propagated energies (if any) that pass through apo and bound states, augmented by its own ligand-binding energy.

[g] Unlike binding free energy (Fig. 2b), binding enthalpy saturates with site occupancy (as bond formation reaches an asymptote).

[h] Van't Hoff approximation is fit to thermal equilibrium of a two-state conformational change $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_1$.

[i] Statistics for analyzing models.

augmented by their own free energy (**21**). Ligand-binding free energies are cumulatively propagated upward, terminating at the main reaction of the tree ("root node").

In the third pass, nodes are visited in top-down order (Fig. 2d–f), and the final, global population of each state is calculated. In this process, the local fractions of each state are multiplied by the global fraction of the parent state within which they are embedded (**27**).

The distribution of states is used to calculate enthalpies

The three types of state distributions can be used to calculate reaction enthalpies. The enthalpy of conformational exchange reactions (**22**) is adjusted by the addition of ligand-binding enthalpies of embedded reactions (**23**). Ligand-binding enthalpies are propagated in accordance with the local fraction of the state through which they enter the reaction (**24**). As with free energies, the intrinsic enthalpies of ligand-binding reactions (**25**) are not perturbed by embedded reactions (**25**). Embedded enthalpies are propagated upward in a bottom-up pass, analogous to free energies (**26**).

The global distribution of state fractions is used to calculate molecular function

The fluorescence properties of a single state are represented by hyperplanes, where each dimension represents the intrinsic dependence of the signal on one of the state variables due to photophysical effects (**28**). The total fluorescence of the protein at any point in phase space is then the sum of the values of the points within these hyperplanes, weighted by the global population fraction of their associated states (**29**).

Complex landscapes can be analyzed with reduced representations

Rather than fitting a NEST model directly to fluorescent data points of the entire landscape derived thermodynamic parameters can be used, limiting the fit to a region in phase space corresponding to a conformational transition zone, such as thermal unfolding. In this approach, apparent transition mid-point temperatures, ${}^{app}T_m$ and corresponding enthalpies, at these temperatures, $\Delta^{app}H_u$, are extracted by fitting the temperature dependence of

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Table 2. Definition of symbols.

Symbol	Units	Class[a]	Equation		Definition
A	K^{-3}	dof	**31**		Arbitrary weighting factor
ΔC_{pb}	kcal/mol/K	dof	**11**, **25**	Heat capacity	Ligand binding
ΔC_{pc}		dof	**1**		Two-state conformational exchange
$\Delta^{\boldsymbol{\sigma}} C_{pc}$		dof	**4**, **22**		Multistate conformational exchange
${}^{\alpha} f_{\rho}$	unitless	ci	**14**, **19**	State population fraction	Isolated ligand binding, apo
${}^{\alpha} f_{\lambda}$		ci	**19**, **21**, **26**		Local ligand binding, apo
${}^{\beta} f_{\rho}$		ci	**13**, **14**,**19**,**26**		Isolated ligand binding, bound
${}^{\beta} f_{\lambda}$		ci	**19**, **21**		Local ligand binding, bound
${}^{\mathbf{S}0} f_{\rho}$		ci	**3**, **7**		Isolated conformational exchange, ground, state
${}^{\mathbf{S}1} f_{\rho}$		ci	**3**		Isolated conformational exchange, $\boldsymbol{S_1}$ state
${}^{\sigma} f_{\lambda}$		ci	**7**		Isolated conformational exchange, $\boldsymbol{\sigma}$ state
${}^{\sigma} f_{\lambda}$		ci	**18**, **24**, **27**		Local conformational exchange, $\boldsymbol{\sigma}$ state
${}^{\sigma} f_{\gamma}$		cv	**27**		Global fraction, state $\boldsymbol{\sigma}$
${}^{p(\sigma)} f_{\gamma}$		cv	**27**		Global fraction, parent of state $\boldsymbol{\sigma}$
ΔG_b	kcal/mol	ci	**9**	Binding free energy	Isolated (and local)
$\Delta G_b^{\ominus}$		dof	**11**, **12**, **21**		Reference, at $T_b^{\ominus}$, 1 M ligand
$\Delta G_b^{\bullet}$		ci	**10**, **11**		Standard, at *T,* 1 M ligand
ΔG_c	kcal/mol	ci	**1**, **2**	Conformational free energy	Two-state
$\Delta^{\boldsymbol{\sigma}} G_c$		ci	**4**, **5**, **15**		Isolated pairwise state $\boldsymbol{\sigma}$: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_{\boldsymbol{\sigma}}$
$\Delta^{\boldsymbol{\sigma}} G_{c,\lambda}$		ci	**15**, **16**		Local pairwise state $\boldsymbol{\sigma}$: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_{\boldsymbol{\sigma}}$
$\Delta^{c(\boldsymbol{\alpha})\rightarrow\boldsymbol{\alpha}} G_{\pi}$		ci	**21**	Recursive ligand-binding free energy propagation	Propagated from child of apo into apo
$\Delta^{c(\boldsymbol{\beta})\rightarrow\boldsymbol{\beta}} G_{\pi}$		ci	**21**		Propagated from child of bound into bound
$\Delta^{c(\mathbf{S}0)\rightarrow\mathbf{S}0} G_{\pi}$		ci	**15**		Propagated from child of ground into ground
$\Delta^{c(\boldsymbol{\sigma})\rightarrow\boldsymbol{\sigma}} G_{\pi}$		ci	**15**, **20**		Propagated from child of $\boldsymbol{\sigma}$ into $\boldsymbol{\sigma}$
$\Delta^{c(\mathbf{p})\rightarrow\mathbf{p}} G_{\pi}$		ci	**20**, **21**		Propagated from child of $\boldsymbol{\sigma}$ into parent
$\Delta H_b^{\ominus}$	kcal/mol	dof	**11**, **25**	Ligand-binding enthalpy	Reference, at $T_b^{\ominus}$
$\Delta H_{b,\lambda}$		ci	**25**, **26**		Local, at *T*
$\Delta H_c^{\ominus}$	kcal/mol	dof	**1**	conformational exchange enthalpy	Reference, at $T_c^{\ominus}$, two-state
$\Delta^{\boldsymbol{\sigma}} H_c^{\ominus}$		dof	**4**, **22**		Reference, at $T_c^{\ominus}$, pairwise state $\boldsymbol{\sigma}$: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_{\boldsymbol{\sigma}}$
$\Delta H_{c,\rho}$		ci	**22**		Isolated total reaction enthalpy
$\Delta^{\boldsymbol{\sigma}} H_{c,\rho}$		ci	**23**		Isolated reaction enthalpy of state $\boldsymbol{\sigma}$
$\Delta^{\boldsymbol{\sigma}} H_{c,\lambda}$		ci	**23**		Local reaction enthalpy of state $\boldsymbol{\sigma}$
${}^{app} H_c$		dof,ocv	**30**		van't Hoff enthalpy
$\Delta^{c(\boldsymbol{\alpha})\rightarrow\boldsymbol{\alpha}} H_{\pi}$	kcal/mol	ci	**26**	Recursive ligand-binding enthalpy propagation	Propagated from child of apo into apo
$\Delta^{c(\boldsymbol{\beta})\rightarrow\boldsymbol{\beta}} H_{\pi}$		ci	**26**		Propagated from child of bound into bound
$\Delta^{c(\mathbf{S}0)\rightarrow\mathbf{S}0} H_{\pi}$		ci	**23**		Propagated from child of ground into ground
$\Delta^{\boldsymbol{\sigma}\rightarrow\boldsymbol{\sigma}} H_{\pi}$		ci	**23**, **24**		Propagated from child of $\boldsymbol{\sigma}$ into $\boldsymbol{\sigma}$
$\Delta^{c(\mathbf{p})\rightarrow\mathbf{p}} H_{\pi}$		ci	**24**, **26**		Propagated from child of parent into parent
K_a	M^{-1}	ci	**8**, **10**, **13**		Ligand-binding association constant
K_c	unitless	ci	**2**, **3**	conformational exchange equilibrium constant	Two-state
${}^{\sigma} K_c$		ci	**6**, **7**		Isolated pairwise state $\boldsymbol{\sigma}$: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_{\boldsymbol{\sigma}}$
${}^{\sigma} K_{c,\lambda}$		ci	**16**, **17**, **18**		Local pairwise state $\boldsymbol{\sigma}$: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_{\boldsymbol{\sigma}}$
${}^{\mathbf{S}0} K_{c,\lambda}$			**17**		Local ground state
$K_d^{\ominus}$	M	dof	**12**	Reference dissociation constant, at $T_b^{\ominus}$	
L	M	sv	**8**, **13**	Ligand concentration	
m_i	kcal/mol/M	dof	**1**	Osmolyte *i*, linear free energy constant	Two-state
${}^{\sigma} m_i$		dof	**4**		Pairwise state $\boldsymbol{\sigma}$: $\mathbf{S}_0 \rightleftharpoons \mathbf{S}_{\boldsymbol{\sigma}}$
n	unitless	dof	**8**, **25**	Number of independent binding sites (stoichiometry)	
$[O_i]$	M	sv	**1**, **4**	Osmolyte *i* concentration	
Q_b	unitless	ci	**8**, **9**	Partition function	Ligand binding
Q_c		ci	**6**, **7**		Isolated conformational change
$Q_{c,\lambda}$		ci	**17**, **18**		Local conformational change
R	kcal/mol/K	pc	many	Gas constant	
${}^{\sigma} R$		ci	**28**, **29**	Fluorescence signal	Total signal for state $\boldsymbol{\sigma}$
${}^{\sigma} R_0$		dof	**28**		Constant
$\partial^{\boldsymbol{\sigma}} R / \partial V_i$		dof	**28**		First derivative of signal w.r.t. state variable
R_{tot}		ocv	**29**		Summed signal over all global states

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Table 2. (*continued*)

Symbol	Units	Class[a]	Equation		Definition
T	K	sv	many	Temperature	External
$T_b^{\ominus}$		pc	**11**, **12**, **25**		Reference for ligand-binding reaction
$T_c^{\ominus}$		dof	**1**		Midpoint ($\Delta G_c = 0$) conformational exchange
${}^{\sigma}T_c^{\ominus}$		dof	**4**, **22**		Pairwise definition in multistate ensemble
${}^{app}T_c$		dof,ocv	**30**		van't Hoff temperature at thermal equilibrium
V_i	various	sv	**28**	State variable	L, T, $[O_i]$
$\overline{y}$	unitless	cv	**13**, **25**		Fractional occupancy
r_{ij}	unitless		**32**	Pearson's correlation coefficient between two parameters	Pearson's correlation coefficient
p_i	various				Value of parameters i
p_j					Value of parameter j
$\overline{p_i}$					Mean value of parameter i.
$\overline{p_j}$					Mean value of parameter j
$F_{2\|1}$	unitless		**33**, **34**	Comparing model 2 to a simpler model 1 using extra sum of squares principle	F statistic
$S_1,\ S_2$			**33**		Sum of squares in model 1 and 2
n_1, n_2					Number of degrees of freedom in 1 and 2
$p_{2\|1}$			**34**		Probability of null hypothesis

NEST, nested equilibrated states tree.

[a] ci, calculation intermediate in the recursive buildup of a NEST model; cv, calculated value upon completion of the NEST computation (see also ocv); dof, degree of freedom (model parameter that is obtained from a fit to experimental data); ocv, observable calculated value (a calculated value, cv, that corresponds to an experimentally observable value such as fluorescence); pc, physical constant; sv, state variable.

the fluorescence signal to the van't Hoff approximation of the equilibrium constant of a two-state reaction near the midpoint temperature (**30**) [34]. The van't Hoff approximation also can be fit to the first derivative of the data (**31**). The NEST model generates these values using **1** and **22**. Consequently, the parameters in a NEST model can be fit by minimizing the least square differences between observed and calculated ${}^{app}T_m$ and $\Delta {}^{app}H_u$ values. This reduced representation approach was used at intermediate stages of model development.

Data fitting and model robustness

Models were fit to experimental data by least-square minimization using conjugate gradient or simplex methods [35] (singly or in succession). The robustness of a model is indicated by the degree of variation in the fit parameter set. We estimated precision in the model parameters using the bootstrap method [35] in which synthetic datasets were created by randomly duplicating a fraction of the original data points (typically 37%), and resolving the model fit (using the original best fit as a starting point). Errors were then estimated for each parameter as the standard deviation of their distribution of values in the resulting fits.

The bootstrap Monte Carlo calculations yield distributions for each parameter, which can be inspected for shape: Gaussian, as expected for uncorrelated, random error; asymmetric; bimodal; etc. Pairwise correlations between parameters p_i and p_j were calculated using Pearson's correlation coefficient, r_{ij} (**32**). The values of r_{ij} fall in the interval [-1,1] where 1 indicates positive correlation, -1 negative correlation, and 0 no correlation. Identification of correlated parameters provides insight into model design. Ideally, parameters are independent (uncorrelated). Correlated parameters may indicate an underlying molecular mechanism that is not accounted for in the model.

We developed our models by iteratively introducing increasing detail and complexity to account for the features in the experimental landscape. To determine whether additional complexity (degrees of freedom) was justified, we used the extra sum of squares (ESS) principle. The ESS principle measures the marginal reduction in the error sum of the squares between calculated and observed data points from the addition of degrees of freedom to the model. The fits of two models, M_1 and M_2, to the same data set then are compared by the ESS F test. The degrees of freedom are n_1 and n_2, respectively, where $n_1 < n_2$ (i.e. M_1 is simpler than M_2). The sum of the squares of the differences between the observed and calculated values are S_1 and S_2, where $S_2 < S_1$ (i.e. M_2 gives a better fit than M_1). The $F_{2|1}$ parameter is defined as the ratio of the relative improvement in fit to the relative change in the number of degrees of freedom (**33**). If the improvement in the fit scales directly with the change in the number of degrees of freedom, then $F_{2|1} = 1$. If the improvement is better than the change in the number of degrees of freedom, then $F_{2|1} > 1$. In the latter case, the $p_{2|1}$ value (the probability of the null hypothesis that $F_{2|1} = 1$) is calculated to determine the significance of this improvement using the standard normal distribution (**34**). If the value exceeds 0.05 (the confidence level), the more complex model is not justified.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

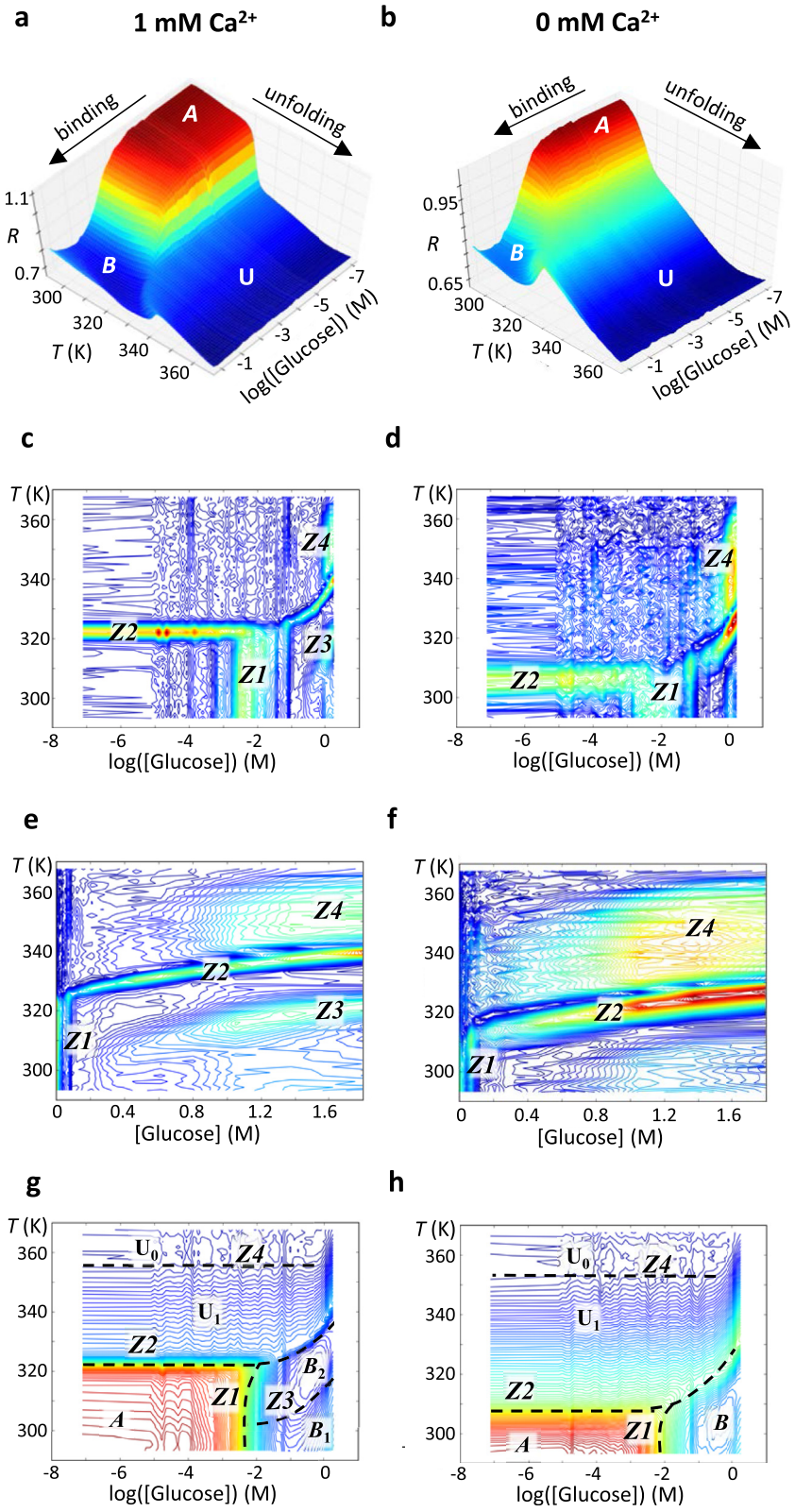


Fig. 3. Fluorescence landscapes of thermal glucose titrations of the ecGGBP183C·Acrylodan conjugate measured in the presence (left column) and absence (right column) of 1 mM Ca^{2+}. Each landscape comprises 3648 experimental observations: 48 glucose concentrations, 76 temperatures. Rainbow coloring indicates magnitude: blue, low; red, high. (**a**, **b**) Fluorescence ratio reports the exchange between three major states: ***A***, folded apoprotein; ***B***, glucose-bound complex; *U*, unfolded protein. (c-f) Absolute numerical total first derivative ($\delta R/\delta[Glc] + \delta R/\delta T$) of the ratiometric signal shown in **a** or

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Results

Temperature-dependent and ligand-dependent ratiometric fluorescent landscape at 1 mM Ca^2

Ratiometric fluorescent signals were collected for 48 glucose concentrations at 1 mM Ca^{2+} every 1 °C between 20 °C and 95 °C (Fig. 3a), using a Roche LightCyler with filters that coincide with the blue/green emission maxima of the ecGGBP183C•Acrylodan conjugate (Fig. 1b). This thermal glucose-binding landscape revealed that both glucose binding and thermal unfolding lower the ratiometric signal compared with the folded apoprotein. In the unfolded state, the ratiometric signal reaches approximately the same value as the saturated glucose complex, suggesting that (un)folding also causes Acrylodan to exchange between its blue (high ratiometric signal) and green (low ratiometric signal) conformational states. The green state conformational therefore occurs in both the glucose complex and the unfolded state [23,24]. The green state therefore forms in two distinct protein environments and in two distinct regions of phase space, separated by glucose or Ca^{2+} concentrations and temperature. The green state experiences distinct protein environments within the closed and unfolded states and therefore exhibit similar but not identical intrinsic fluorescence behaviors (see last section in Results). Accordingly, these green states are treated independently in the models described in the following paragraph.

Transformation of the ratiometric fluorescent signal into its total first derivative with respect to both temperature and glucose concentration (Fig. 3c and e) identified four transition zones (*Z1-Z4*) in which the signal changes nonlinearly, separating five planar regions, A, B_1, B_2, U_0, and U_1 (Fig. 3g). We used these first derivative plots to identify single reactions because these drive the exchange between two fluorescent states, which give rise to the transition zones.

Analysis of individual reactions in the 1 mM Ca^2 landscape

We postulated that the *Z1* transition corresponds to a glucose-binding reaction in the folded state, $\mathbf{A}+glucose \rightleftharpoons \mathbf{B}\cdot glucose$ (Fig. 4a and b). The glucose-dependent ratiometric signals corresponding to *Z1* and its associated signal planes (A, the folded apoprotein; B, the folded glucose-bound plane) fit well to individual ligand-binding isotherms, extracted from the experimental landscape (Fig. 4a). The temperature dependence of the resulting apparent glucose-binding affinities could be fit to the Gibbs-Helmholtz relationship **11** (Fig. 4b), consistent with the *Z1* transition zone corresponding to a single ligand-binding reaction.

We postulated that the *Z2* transition zone corresponds to a thermal (un)folding reaction, $\mathbf{F} \rightleftharpoons \mathbf{U}$. The midpoint temperatures (T_m) of the thermal transitions increase significantly at elevated glucose concentrations (>0.2 M). This stability increase could be modeled as a linkage between glucose binding and thermal (un)folding with a dominant stabilizing glucose-mediated osmolyte effect [30] (Fig. 4c), as postulated.

The *Z3* transition zone separates signal planes B_1 and B_2 within the folded, glucose-bound state (Fig. 3c, e, and g). We postulated that the *Z3* transition marks a thermally driven two-state conformational exchange in the glucose-bound complex, $\mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$. The ratiometric signals of thermal transitions were fit individually at each glucose concentration using the van't Hoff analysis (Fig. 4d). The resulting extracted midpoint transition temperatures, ${}^{app}T_c$, and corresponding enthalpies, $\Delta^{app}H_c$, could be fit to Gibbs-Helmholtz equation **1** with a glucose-mediated osmolyte effect (Fig. 4e and f), consistent with our postulate.

The *Z4* transition separates signal planes U_0 and U_1 (Fig. 3c, e, and g). The transition is comparatively broad with a near-linear increase in ratiometric signal at elevated glucose concentrations. We postulated that this transition corresponds to a thermally driven conformational change in the unfolded state, $\mathbf{U}_0 \rightleftharpoons \mathbf{U}_1$.

Linking all individual reactions constructs the global model for the 1 mM Ca^2 landscape

To obtain direct fits of the entire temperature- and glucose-dependent fluorescence landscape at 1 mM Ca^{2+}, we explored three NEST models that couple conformational changes to glucose binding with increasing levels of detail and complexity (Table 3). To fit the models to the experimental fluorescence

b. The derivatives are plotted as logarithmic (**c**,**d**) and linear (**e**,**f**) dependence on glucose concentration. The "ridges" of the first derivative plot visualize midpoints of four transition zones, *Z1-Z4*, and are used to identify the location of two-state reactions (see text): *Z1*, glucose binding; *Z2*; thermal (un)folding; *Z3*, conformational exchange in the glucose-bound complex (this transition is absent at 0 mM Ca^{2+}); *Z4*, conformational exchange in the unfolded state. (**g**,**h**) The four transition zones separate five states associated with unique ratiometric fluorescence signals (shown as projections down the ratiometric signal *z*-axis of panels **a** and **b**): ***A***, folded apoprotein; the B_1 and B_2 substates of ***B***, glucose-bound complex; the U_0 and U_1 substates of ***U***. ecGGBP, *Escherichia coli* glucose-galactose binding protein.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

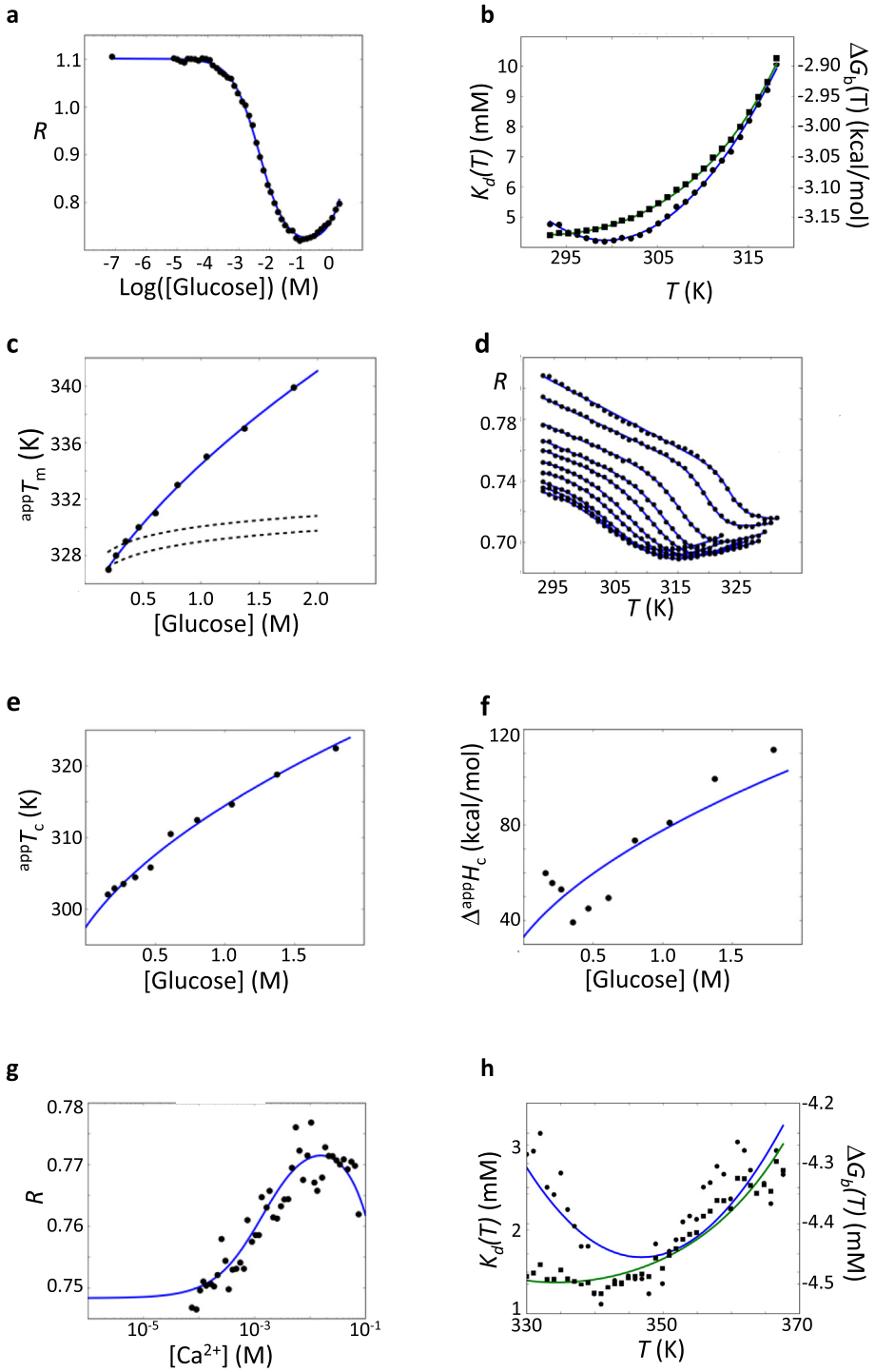


Fig. 4. Analysis of isolated two-state reactions in thermal glucose- and Ca^{2+}-binding landscapes. (**a**-**f**) Experimental data extracted from the glucose-binding landscape measured at 1 mM Ca^{2+} (Fig. 3); (**g** and **h**) Experimental data extracted from the Ca^{2+} -binding landscape measured in the absence of glucose (Fig. 6). (**a**) A single glucose-binding isotherm extracted at 20 °C in the transition zone *Z1*. Circles, fluorescent ratio data points; solid line, binding isotherm (K_d = 4.4 mM; constant apo base line: $A = 1.102$; linear, glucose-dependent bound base line: $A = 0.713$, $B = 0.057$). The ligand-binding isotherm was fit to **13** (Table 1). (**b**) Temperature dependence of all glucose binding isotherms extracted within *Z1*: squares, K_d values; circles, ΔG_b values extracted for each temperature in *Z1*); lines, Gibbs-Helmholtz fit ($T_b^{\ominus} = 298.15$K; $\Delta G_b^{\ominus} = -3.19$ kcal/mol; $K_d^{\ominus} = 4.6$ mM; $\Delta H_b^{\ominus} = -2.60$ kcal/mol; $\Delta C_{p,b} = -0.494$ kcal/mol/K). The free energies were fit to the Gibbs-Helmholtz relationships using equations **10** and **11**. (**c**) The effect of glucose on the midpoint of the (un)folding conformational transition temperature, T_m, in transition zone *Z2* determined at concentrations >0.2 M. The T_m

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

data, we applied the protocol detailed in the Theory section: the reaction parameters specify free energies that are converted into equilibrium constants, which are converted into fractions of reactant and product states, which are combined with baseplanes describing their intrinsic fluorescence properties, resulting in a calculated fluorescence landscape that is compared with the corresponding experimental values (Fig. 2g). Standard iterative optimization procedures were used to determine optimal parameter values for the free energies and baseplanes that minimize the difference between the calculated and experimental landscapes. The first and simplest model, NEST1 (Fig. 5a, Fig. S3), posits a two-state unfolding mechanism in which glucose binding is coupled to the folded (***F***) state. Fluorescence signals are assigned to the glucose-free ($^{\alpha}$***F***) and glucose-bound ($^{\beta}$***F***) states of ***F***, and the unfolded (***U***) state. This model broadly recapitulates the glucose binding (*Z1*) and thermal (un)folding (*Z2*) reactions but not *Z3* or *Z4*. Accordingly, the calculated and observed contour levels deviate visibly in the regions corresponding to the glucose complex and thermally unfolded states.

In the second model, NEST2 (Fig. 5b, Fig. S4), the $^{\beta}$***F*** state is split into the $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ states, which are in conformational exchange with each other, but not with ***F*** or ***U***, presenting an example of a conformational exchange reaction that is hierarchically nested within another state. These two new states each are connected to a distinct fluorescence signal. In this model, *Z3* is now accounted for resulting in good visible agreement between calculated and observed signals in the region of the glucose complex but still exhibits some deviation in the thermally unfolded region.

The third model, NEST3 (Fig. 5c, Fig. S5), accounts for *Z1*-Z4 by invoking a three-state conformational ensemble, in which ***U*** is split into $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$, each in exchange with ***F*** and connected to a distinct fluorescence signal. The glucose-binding reaction is embedded within the folded state, thereby coupling the distribution of ***F***, $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$ to glucose concentrations. Glucose also exerts osmolyte effects on all conformational exchange reactions, including the transition between $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$. A model in which $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$ are embedded in ***U*** and not in exchange with ***F*** does not fit, as well as NEST3 (not shown). Of all the landscapes tried, NEST3 fits the fluorescent landscape most precisely, visibly matching the experimental observations over the entire phase space.

To assess the statistical significance of improvements as the models progress from simple (NEST1) to complex (NEST3), we calculated the ESS $F_{x|1}$ statistic of model *x* (*x* equals 2 or 3) relative to NEST1 (Table 3) and where appropriate its corresponding $p_{x|1}$-value (Theory section). This $F_{x|1}$ statistic is the ratio of the relative improvement in least square differences of experimental and calculated fluorescence values to the relative increase in the number of degrees of freedom of the more complex model. If $F_{x|1} = 1$, the increased complexity did nothing (null hypothesis); if $F_{x|1} < 1$, the increased complexity overfit the model; if $F_{x|1} > 1$, the increased complexity improved the model. In the latter case, the likelihood that this improvement is a random deviation from the null hypothesis is given by the $p_{x|1}$ value. In light of this analysis, NEST2 is a slight overfit, even though it visually accounts for the $\mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$ transition (zone *Z3*). By contrast, the most complex model, NEST3, is a statistically significant improvement ($p_{3|1} = 7 \times 10^{-12}$);

values (circles) were extracted from individual thermal melts using the absolute partial derivative of the experimental ratiometric data and determining the temperature at which its value reaches a maximum (i.e. the midpoint of the ridge in the *Z2* zone Fig. 3c and e). Solid line, Gibbs-Helmholtz fit of the experimental T_m values with a linear free energy term for a glucose-dependent osmolyte effect ($T_m^{\ominus} = 324.5$ K; $\Delta H_u^{\ominus} = 115$ kcal/mol; $\Delta C_{pu} = 10.25$ kcal/mol/K; $m = 5.0$ kcal/mol/M), equation **1**. To illustrate the importance of the osmolyte effect, we simulated a hypothetical model (dashed black lines), in which the change in stability is due only to conformational coupling between binding of glucose to the folded state and a two-state (un)folding reaction (see NEST1 model in Fig. 5 for explanation), omitting the osmolyte term from the conformational exchange reaction and using glucose binding parameters calculated in panel **b**. Two lines are shown, with $K_d^{\ominus}$ values of 4.6 mM (top line) and 12.0 mM (bottom line), respectively. This simulation demonstrates that glucose binding alone cannot account for the increase in thermostability. (**d**) Conformational exchange reaction in transition zone *Z3*. Circles, experimental observations; lines, fit of two-state conformational exchange reaction directly to the experimental ratiometric signal *R* calculated by applying fractions of each of the two states (**3**) derived from the van't Hoff two-state model (30) to two fluorescence signal baselines corresponding to the $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ states, respectively (**28**). (e and f) The $^{app}T_c$ and $\Delta^{app}H_c$ values extracted at all glucose concentrations were fit jointly to the Gibbs-Helmholtz equation **1** with a glucose-dependent osmolyte effect ($T_c^{\ominus} = 297.5$ K; $\Delta H_c^{\ominus} = 33$ kcal/mol; $\Delta C_{pc} = 2.6$ kcal/mol/K; $m = 3.1$ kcal/mol/M). (**g**) Overt fluorescence response of Ca^{2+} binding to the unfolded state (transition zone *Z5*) illustrated by a binding isotherm extracted at 57 °C from the experimental thermal Ca^{2+} titration landscape (Fig. 6c). Circles, ratiometric data; blue line, isotherm fit ($K_d = 1.5$ mM; constant apo base line: $A = 0.748$; linear, Ca^{2+}-dependent bound base line: $A = 0.776$, $B = -0.137$). (**h**) K_d values (squares) were extracted for all binding isotherms measured in the *Z5* zone using equation **13**, and their temperature-dependent free energy values, ΔG_b (circles), were fit to the Gibbs-Helmholtz equations **10** and **11** ($T_b^{\ominus} = 323.15$ K): $\Delta G_b^{\ominus} = -4.16$ kcal/mol; $K_d^{\ominus} = 1.5$ mM; $\Delta H_b^{\ominus} = 4.03$ kcal/mol; $\Delta C_{p,b} = -0.358$ kcal/mol/K): K_d (squares, green), ΔG_b (circles, blue). NEST, nested equilibrated states tree.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Table 3. Comparison of the fits[a] of different NEST models to the thermal glucose titration landscape determined in the presence of 1 mM Ca^{2+}.

Parameter class		Parameters		Models		
		Variable	Units	NEST1	NEST2	NEST3
Degrees of freedom				15	22	28
Bootstrap trials[b]				1688	2264	1262
Goodness of fit[c]		RMSD		$0.0087\pm7\times10^{-5}$	$0.0082\pm8\times10^{-5}$	0.0040±0.0002
		CoV	%	1.1	1.0	0.5
Model evaluation by ESS[d]		$F_{x\|1}$	unitless	(reference)	0.3	8.0
		$p_{x\|1}$		(reference)		7.2×10^{-12}
U ⇌ **F** or **U**$_0$ ⇌ **F**		$T_m^{\ominus}$	K	322.1±0.034	322.1±0.040	327.1±0.36
		$\Delta H_u^{\ominus}$	kcal/mol	183.9±3.61	185.1±3.72	−297.7±13.9
		ΔC_{pu}	kcal/mol/K	3.44±1.027	10.87±0.268	−9.81±0.418
		m	kcal/mol/M	4.74±0.272	6.79±0.216	−8.52±0.270
U$_0$ ⇌ **U**$_1$		$T_m^{\ominus}$	K			366.5±5.3
		$\Delta H_u^{\ominus}$	kcal/mol			−7.5±7.16
		ΔC_{pu}	kcal/mol/K			1.22±0.237
		m	kcal/mol/M			−0.57±0.054
B$_1$ ⇌ **B**$_2$		$T_c^{\ominus}$	K		295.6±0.80	293.2±2.40
		$\Delta H_c^{\ominus}$	kcal/mol		0.2±1.9	−5.1±5.9
		ΔC_{pc}	kcal/mol/K		3.77±0.423	3.45±0.500
		m	kcal/mol/M		2.62±0.233	2.49±0.298
Glucose binding[e]		$\Delta G_b^{\ominus}$	kcal/mol	−3.207±0.0029	−3.155±0.0027	−3.153±0.0027
		$\Delta H_b^{\ominus}$	kcal/mol	−2.32±0.129	−2.13±0.198	−2.41±0.210
		$\Delta C_{p,b}$	kcal/mol/K	−0.58±0.014	−0.45±0.026	−0.41±0.028
Fluorescent glucose complex states[f]	α	R_o		1.61±0.011	1.59±0.014	1.59±0.015
		$\partial R/\partial T$	K^{-1}	$-0.0017\pm4\times10^{-5}$	$-0.0016\pm5\times10^{-5}$	$-0.0016\pm5\times10^{-5}$
	B$_1$ or β	R_o		1.39±0.007	1.28±0.013	1.26±0.017
		$\partial R/\partial T$	K^{-1}	$-0.0023\pm2\times10^{-5}$	$-0.0018\pm5\times10^{-5}$	$-0.0018\pm7\times10^{-5}$
		$\partial R/\partial[Glc]$	M^{-1}	0.051±0.0005	0.036±0.0017	0.034±0.0022
	B$_2$	R_o			0.60±0.032	0.52±0.062
		$\partial R/\partial T$	K^{-1}		0.0002±0.0001	0.0005±0.0002
		$\partial R/\partial[Glc]$	M^{-1}		0.025±0.0011	0.024±0.0016
Fluorescent unfolded state(s)[g]	**U**$_0$	R_o				−0.97±0.377
		$\partial R/\partial T$	K^{-1}			0.0044±0.0010
		$\partial R/\partial[Glc]$	M^{-1}			−0.180±0.0321
	U or **U**$_1$	R_o		1.59±0.005	1.60±0.005	1.26±0.125
		$\partial R/\partial T$	K^{-1}	-0.0025 ± 10^{-5}	-0.0026 ± 10^{-5}	−0.0015±0.0004
		$\partial R/\partial[Glc]$	M^{-1}	0.022±0.0008	$0.022\pm8\times10^{-5}$	0.040±0.0007

NEST, nested equilibrated states tree.

[a] See Table 2 for definition of symbols. Least squares fit to the observed ratiometric signal using simplex and conjugate gradients minimizers in succession. Variation in parameters is estimated from bootstrap analysis.

[b] 37% of the data is randomly duplicated in each bootstrap trial.

[c] $RMSD=\sqrt{\frac{\sum_i^N({}^{abs}R_i-{}^{calc}R_i)^2}{N}}$, where ${}^{obs}R_i$ and ${}^{cal}R_i$ are the observed and calculated ratiometric signals, respectively, determined at the *i*th combination of temperature and glucose concentrations. The coefficient of variance is $CoV=100\times RMSD/\overline{R}$ where $\overline{R}$ is the average ratiometric signal in the landscape.

[d] Extra sum of squares (ESS) principle, equations 33 and 34, Table 1.

[e] $T_b^{\ominus}=298.15$ K.

[f] Parameters of the fluorescence signal planes attached to the states: F_o, constant; $\partial R/\partial T$, temperature-dependent slope; $\partial R/\partial[Glc]$, glucose-dependent slope (not used for the apo-glucose state, α).

[g] In the NEST3 model, U_0 instead of *F* was chosen as the reference state. Accordingly, all the parameters in the unfolded exchange reaction switch sign compared with their counterparts in NEST1 and NEST2.

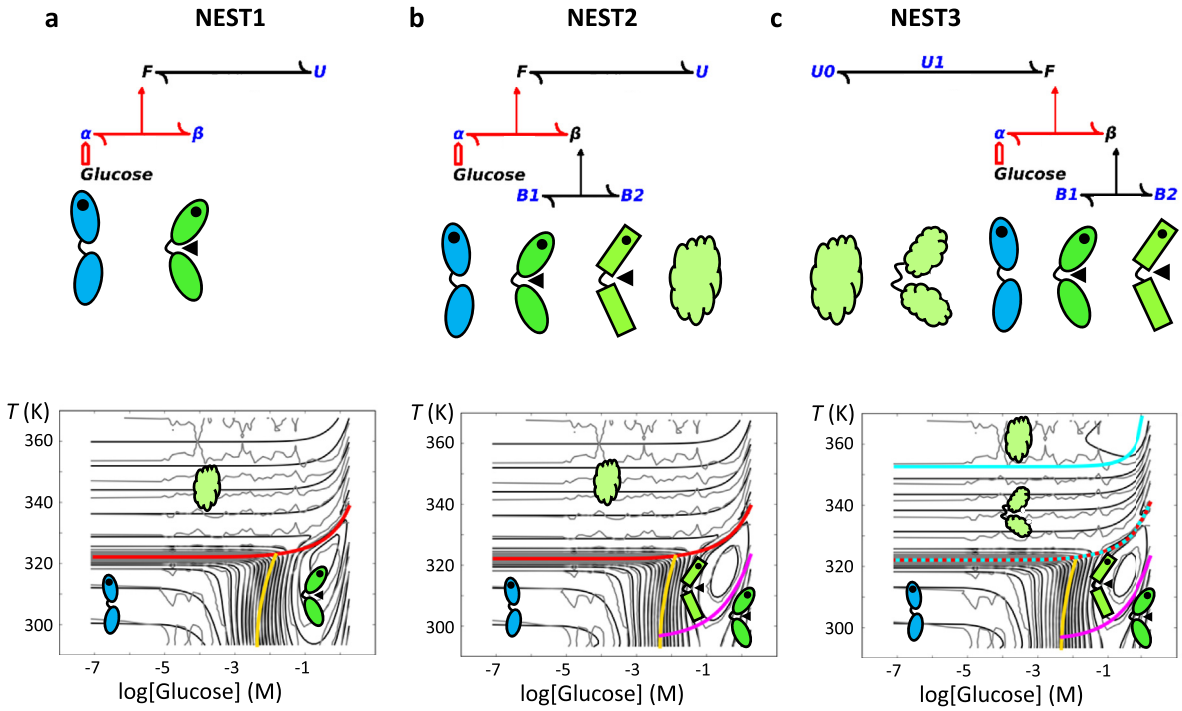


Fig. 5. Comparison of the global fits of three NEST models to the experimental thermal glucose-binding landscape determined at 1 mM Ca^{2+}. The top row shows the NEST models (Fig. 2 for explanation of representation; scripts and equations are shown in Fig. S3-S5). States are represented by cartoons (black triangle, glucose; black circle, Ca^{2+}; blue, open apostate; dark green, closed glucose-bound states with ovals corresponding to $\boldsymbol{B_1}$ and rectangles to $\boldsymbol{B_2}$ conformational states; light green, unfolded states). The bottom row shows comparisons of experimental (gray contours) and calculated (black contours) ratiometric signals. Each model was fit to the experimental data using a nonlinear least-squares difference solver. Cartoons indicate dominant states in regions of phase space. Colored lines indicate the transition midpoints for the reactions: gold, K_d for glucose binding; red, T_m for $\mathbf{F} \rightleftharpoons \mathbf{U}$ (**a** and **b**) or T_c for $\mathbf{U}_0 \rightleftharpoons \mathbf{F}$ (c); blue, T_c for $\mathbf{U}_0 \rightleftharpoons \mathbf{U}_1$ (solid line, "hot arm"; dashed, "cold" arm); magenta, T_c for $\mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$. (**a**) The NEST1 model couples the glucose-binding reaction to a two-state (un)folding reaction. (**b**) The NEST2 model introduces a two-state conformational exchange reaction ($\mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$) within the glucose-bound complex. (**c**) The NEST3 model elaborates NEST2 by replacing the thermal (un)folding reaction with a three-state conformational exchange comprising the folded state, ***F***, and two forms of the unfolded state, $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$. The T_c values for the $\mathbf{U}_0 \rightleftharpoons \mathbf{F}$ and the $\mathbf{U}_0 \rightleftharpoons \mathbf{U}_1$ cold arm overlap. NEST, nested equilibrated states tree.

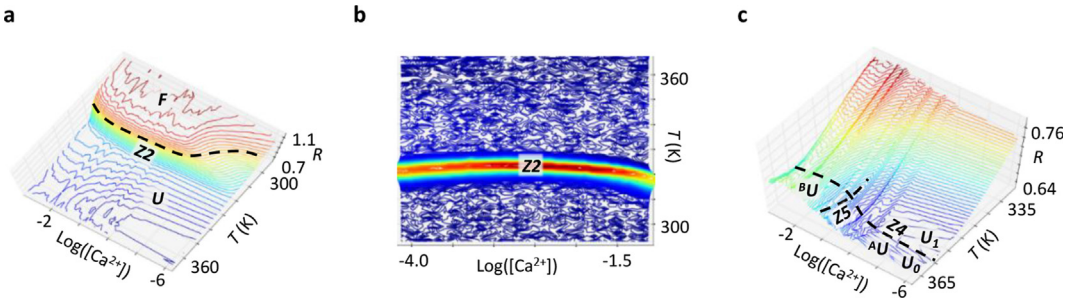


Fig. 6. Experimental thermal Ca^{2+}-binding landscape determined in the absence of glucose. (**a**) Thermal Ca^{2+}-binding landscape. At low concentrations (<3 mM), addition of Ca^{2+} enhances protein thermostability. (**b**) First derivative of the signal (rainbow coloring indicates magnitude) reveals that the midpoint temperature of the thermal unfolding reaction (*Z2* transition) decreases at elevated Ca^{2+} concentrations (>3 mM). (**c**) Fluorescence in the temperature range dominated by the unfolded state shows both a direct response to binding of Ca^{2+} (*Z5* transition), and an independent temperature-dependent conformational change that also is observed in the thermal landscape of the glucose titration (*Z4* transition).

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

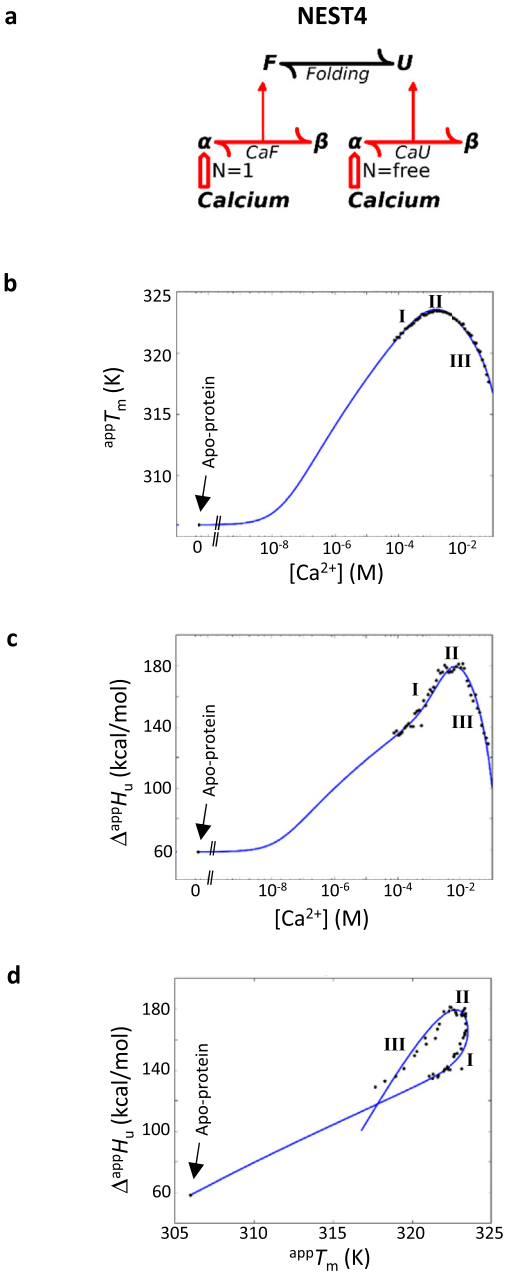


Fig. 7. The dependence of protein thermostability on Ca^{2+}. (**a**) NEST4 models the effect of Ca^{2+} on two-state thermostability as a balance of free energies of a single, high-affinity site binding to the folded state, *F*, and multiple, low-affinity sites binding to the unfolded state, *U* (setup script and equations shown in Fig. S6). (**b**-**c**) The experimental observations, pairs of ${}^{app}T_m$ and $\Delta^{app}H_u$ values (48 data points), were extracted for individual thermal unfolding transitions at each Ca^{2+} concentration (Fig. 6) using van't Hoff analysis (equation **30** and fit to NEST4 by two iterations of joint least-squares minimization of both values, successively weighting differences between calculated and observed values for $\Delta^{app}H_u$ by 0.01 and 0.05. In NEST4, equations **15**–**21** and **22**–**24** were used to propagate ligand-binding free energies and enthalpies, respectively from the two Ca^{2+}-binding reactions into the two-state unfolded reaction. The precision of the fits (blue lines) to experimental data points (circles) was assessed by bootstrapping of the extracted ${}^{app}T_m$ and $\Delta^{app}H_u$ data. **F**⇌**U**: $T_m^{\ominus} = 306.0 \pm 1.30$ K, $\Delta H_u^{\ominus} = 58.5 \pm 1.10$ kcal/mol, $\Delta C_{p,u} = 4.41 \pm 0.483$ kcal/mol/K; high-affinity Ca^{2+} binding site ($T_b^{\ominus}$=298.15 K): $\Delta G_b^{\ominus} = -9.3 \pm 2.47$ kcal/mol, $\Delta H_b^{\ominus} = 26.5 \pm 73.0$ kcal/mol, $\Delta C_{p,b} = -1.9 \pm 3.6$ kcal/mol/K; low-affinity Ca^{2+} binding site ($T_b^{\ominus}$=323.15 K): $\Delta G_b^{\ominus} = -3.4 \pm 0.07$ kcal/mol, $\Delta H_b^{\ominus} = 37.6 \pm 13.6$ kcal/mol, $\Delta C_{p,b} = 6.7 \pm 2.0$ kcal/mol/K, $N = 3.8 \pm 0.35$. The noninteger stoichiometry of the Ca^{2+}-binding site models the averaged behavior of an ensemble with different affinities. Three regions can be distinguished in the fits: (*I*) high-affinity Ca^{2+} binding to ***F*** dominates the thermodynamics of protein thermostability; (*II*) the Ca^{2+}-binding free energies of the folded and unfolded states are balanced; (*III*) the free energy contributions of multiple weak Ca^{2+}-binding sites in the unfolded state outweigh those of the high-affinity site by virtue of stoichiometry. (**d**) The "curl" observed for the dependence of $\Delta^{app}H_u$ values on their corresponding ${}^{app}T_m$ temperatures is characteristic for the presence of weak ligand-binding sites in the unfolded state with stoichiometries that exceed binding to the folded state.

modeling conformational transitions in the unfolded state therefore matters.

Ca^2 binds to folded and unfolded states

A thermal Ca^{2+}-binding landscape recorded in the absence of glucose (Fig. 6) revealed that removal of Ca^{2+} substantially destabilizes the protein, consistent with previous observations [36–38]. Maximal stability is reached at ~3 mM Ca^{2+}. Above this concentration, Ca^{2+} destabilizes the protein.

In the NEST4 model, the Ca^{2+}-dependent thermostability was modeled with a two-state protein folding reaction, coupled to a single Ca^{2+}-binding site in the folded and multiple Ca^{2+} sites in the unfolded state (Fig. 7a, Fig. S6). ${}^{app}T_m$ and $\Delta^{app}H_u$ values were extracted for individual thermal transitions at each Ca^{2+} concentration using the van't Hoff analysis (not shown) and fit with NEST4 (Fig. 7b–d). The fit revealed the opposing effects of Ca^{2+} binding to a high-affinity site in the folded state, which increases thermostability, and to multiple low-affinity sites in the unfolded state, which destabilizes the protein. The differences in the stoichiometry of Ca^{2+} sites in these two states causes the weak destabilizing contributions to exceed the stabilizing effects at elevated Ca^{2+} concentrations as the Ca^{2+}-binding free energy contributed by multiple, low-affinity sites overwhelms that of a single, high-affinity site. Consequently, protein thermostability reaches a maximum at an intermediate Ca^{2+} concentration. These effects illustrate the interplay between binding affinity and site stoichiometry in conformational coupling involving multiple states that bind the same ligand.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

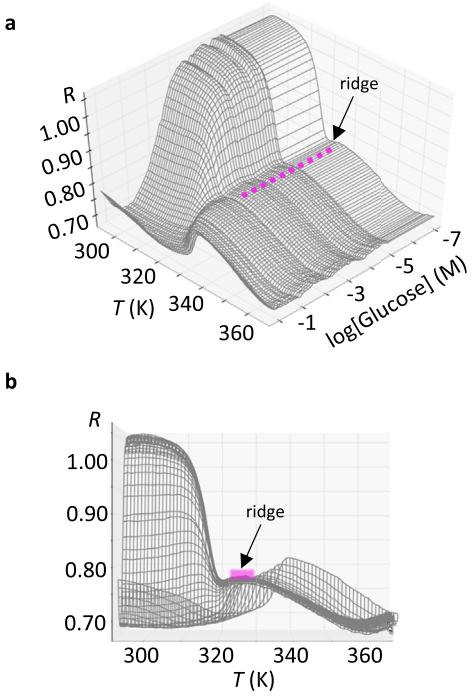


Fig. 8. The experimental thermal glucose-binding landscape determined in the presence of 100 mM Ca^{2+}. (**a** and **b**) At this Ca^{2+} concentration, the ratiometric signal displays a "ridge" (magenta line) in the unfolded state at ~330 K, which is absent at low Ca^{2+} concentrations (cf. Fig. 3).

Unlike glucose, the addition of Ca^{2+} does not evince a fluorescence change in the folded state. However, the unfolded state exhibits a small change in fluorescence, defining an additional transition zone *Z5* (Fig. 6c). Ca^{2+}-binding isotherms isolated within *Z5* (Fig. 4g) could be fit to Gibbs-Helmholtz equation **11** (Fig. 4h), consistent with *Z5* corresponding to a low-affinity Ca^{2+}-binding reaction that evokes a small fluorescence response higher than 3 mM.

The Ca^2-free protein adopts a unique conformational state

The thermal glucose-binding landscape measured in the absence of Ca^{2+} (Fig. 3b and h) shows no evidence for the conformational exchange reaction in its glucose complex (compare the left and right columns in Fig. 3). This landscape fits well to a model, NEST5 (Fig. S7), in which this reaction is absent but the three-state (un)folding ensemble is retained (Supplementary Table 1).

The distinctiveness of the Ca^{2+}-free compared with Ca^{2+}-bound landscapes reflects a change in the conformational state of these two forms. Spectroscopic studies have shown that in the absence of Ca^{2+}, ecGGBP adopts a distinct conformation, probably corresponding to a partially unfolded state of the Ca^{2+}-binding domain [36–38], consistent with the comparatively smaller ΔC_{pc} and m values observed in NEST5. The partially folded protein exhibits a weaker glucose affinity compared with the fully folded protein ($K_d^{\ominus}$ =9.6 mM and 4.9 mM at 0 mM and 1 mM Ca^{2+}, respectively), uncovering a linkage between Ca^{2+} and glucose binding.

Ca^2 binding alters the shapes of thermal glucose-binding landscapes

Landscapes were collected at eight Ca^{2+} concentrations. With the exception of the 0 mM Ca^{2+} landscape (see previously), the NEST3 model developed to describe the 1 mM Ca^{2+} landscape (Fig. 5c) also fits well to landscapes obtained with other Ca^{2+} concentrations. Each landscape was fit with unique NEST3 parameters (Supplementary Table 1). These fits are consistent with a Ca^{2+}-dependent change in thermostability (see previously). High Ca^{2+} concentrations introduce a "ridge" in the region of phase space where the unfolded state(s) dominate (Fig. 8), further demonstrating that Ca^{2+} binding visibly alters the distribution of states in the unfolded protein.

The entire experimental phase space can be reproduced with one comprehensive model

NEST6 models the ratiometric signal over the entire temperature-, glucose-, and Ca^{2+}-dependent phase space by incorporating the effects modeled in NEST3-5 for lower-dimensionality landscapes, with additional features that enable their combination (Fig. 9a, Fig. S8). The three-state conformational ensemble identified in NEST3 to model thermal (un) folding (Fig. 5c) is expanded with an additional state, the partially unfolded Ca^{2+}-free ***P*** state, corresponding to the Ca^{2+}-free state modeled in NEST5 (Fig. S7). Consequently, the distribution of all four states (***F***, ***P***, $\boldsymbol{U_1}$, $\boldsymbol{U_0}$) is affected by the differences in glucose- and Ca^{2+}-binding affinities between the states, as well as glucose-mediated osmolyte effects. ***P*** and ***F*** bind glucose with different affinities. Glucose binds only to the Ca^{2+} complex of ***F***; this condition is imposed by embedding glucose binding within the Ca^{2+}-bound state. The exchange between the $\mathbf{B}_1$ and states is present in ***F***, but not ***P***, and is hierarchically confined to the glucose complex of the Ca^{2+}-bound protein. Not all states are associated with fluorescent signals. Dark states nevertheless contribute to the total fluorescence signal by the effect of their distribution on the global fractions of the fluorescent states.

Combining the effects of Ca^{2+} on both protein thermostability (NEST4, Fig. 7) and the direct fluorescence response of the unfolded state (Fig. 4g and h) required that the multiple Ca^{2+}-

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

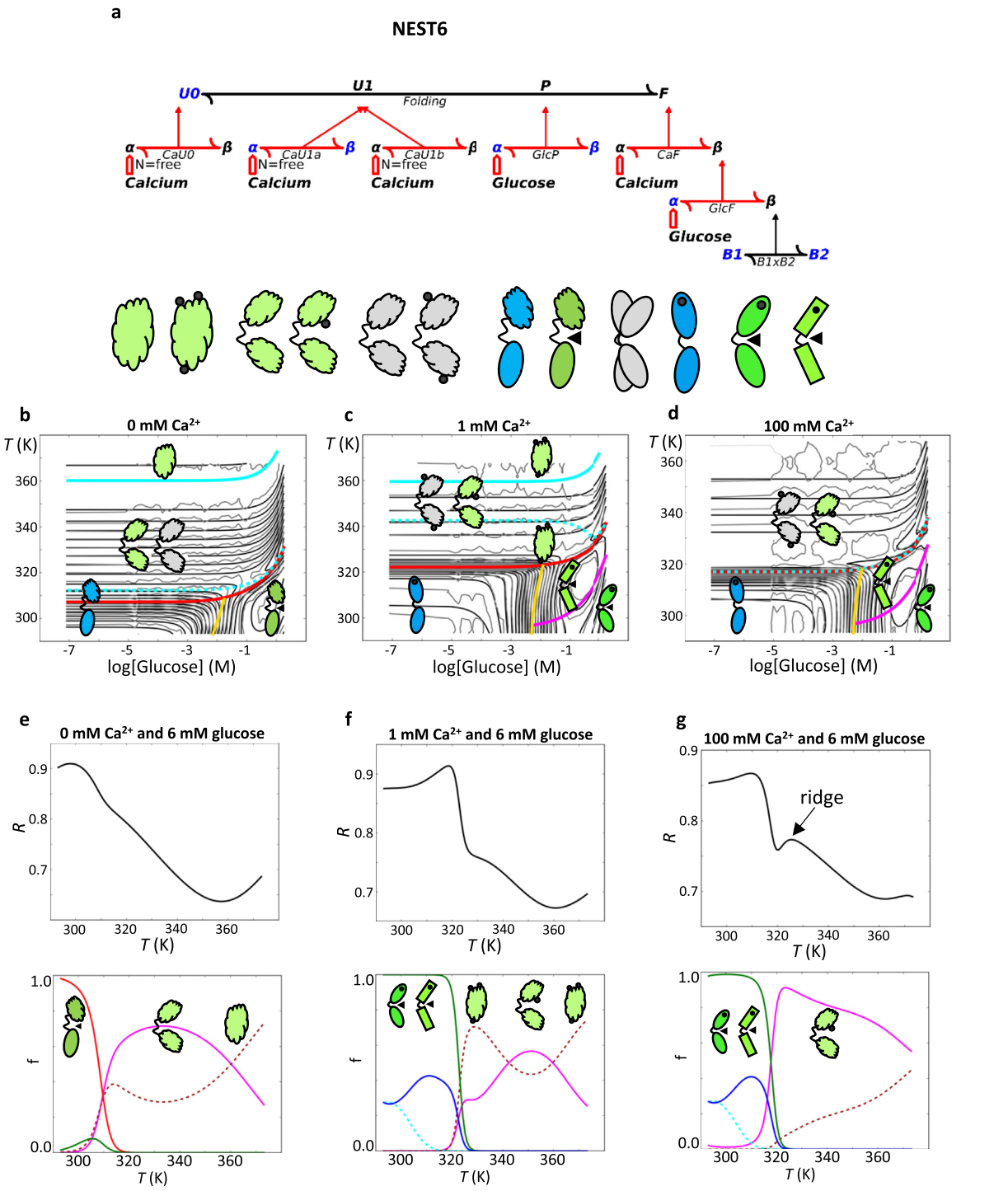


Fig. 9. Comprehensive fit of ecGGBP183C•Acrylodan fluorescence as a function of temperature, glucose, and Ca^{2+}. (**a**) The NEST6 model combines the effects modeled in NEST3-5 (setup script and equations shown in Fig. S8). The model was fit to the combined data of eight experimental thermal glucose titrations determined at difference Ca^{2+} concentrations (0.0, 0.1, 0.3, 1.0, 3.0, 10.0, 30.0, 100.0 mM). (**b-d**) Illustration of the global fit for individual experimental landscapes at three different Ca^{2+} concentrations (arranged in labeled columns), showing the positions of the midpoints of ligand-binding and conformational exchange reactions (coloring as in Fig. 5), with cartoons indicating the dominant state in each region of phase space. (**e-g**) Temperature dependence of fluorescence signals (top row) at 6-mM glucose (euglycemia concentration) and corresponding distribution of states (bottom row): green, ***F***; orange, ***P***; magenta, $\boldsymbol{U_1}$; dashed brown, $\boldsymbol{U_0}$; dashed cyan, $\boldsymbol{B_1}$; blue, $\boldsymbol{B_2}$ (Fig. S9 for more such plots at additional Ca^{2+} and glucose concentrations). NEST, nested equilibrated states tree; ecGGBP, *Escherichia coli* glucose-galactose binding protein.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Table 4. NEST6 global fit[a].

Parameter class		Parameters: Variable	Parameters: Units	Fit values[b]
$\mathbf{U}_0 \rightleftharpoons \mathbf{F}$		$T_c^{\ominus}$	K	304.3±0.2
		$\Delta H_c^{\ominus}$	kcal/mol	−54.8±1.3
		ΔC_{pc}	kcal/mol/K	−4.6±0.1
		m	kcal/mol/M	−5.6±0.1
$\mathbf{U}_0 \rightleftharpoons \mathbf{P}$		$T_c^{\ominus}$	K	309.1±0.1
		$\Delta H_c^{\ominus}$	kcal/mol	−85.2±0.9
		ΔC_{pc}	kcal/mol/K	−3.1±0.2
		m	kcal/mol/M	−4.5±0.1
$\mathbf{U}_0 \rightleftharpoons \mathbf{U}_1$		$T_c^{\ominus}$	K	359.7±0.2
		$\Delta H_c^{\ominus}$	kcal/mol	−16.7±0.1
		ΔC_{pc}	kcal/mol/K	−0.63±0.01
		m	kcal/mol/M	−0.70±0.01
$\mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$		$T_c^{\ominus}$	K	295.2±0.4
		$\Delta H_c^{\ominus}$	kcal/mol	0.0±1.4
		ΔC_{pc}	kcal/mol/K	4.3±0.2
		m	kcal/mol/M	3.8±0.1
Glucose binding	to ***P***	$\Delta G_b^{\ominus}$	kcal/mol	−2.74±0.006
		$\Delta H_b^{\ominus}$	kcal/mol	−10.6±0.3
		$\Delta C_{p,b}$	kcal/mol/K	0.10±0.02
	to ***F***	$\Delta G_b^{\ominus}$	kcal/mol	−3.11±0.001
		$\Delta H_b^{\ominus}$	kcal/mol	−2.77±0.07
		$\Delta C_{p,b}$	kcal/mol/K	−0.303±0.008
Ca^{2+} binding	to $\boldsymbol{U_0}$	$\Delta G_b^{\ominus}$	kcal/mol	−4.89±0.03
		$\Delta H_b^{\ominus}$	kcal/mol	29.2±0.5
		$\Delta C_{p,b}$	kcal/mol/K	−1.68±0.02
		N		1.6±0.02
	to $\boldsymbol{U_1}$ (fluorescent)	$\Delta G_b^{\ominus}$	kcal/mol	−2.79±0.03
		$\Delta H_b^{\ominus}$	kcal/mol	75.1±0.8
		$\Delta C_{p,b}$	kcal/mol/K	−3.02±0.03
		N		1.3±0.02
	to $\boldsymbol{U_1}$ (dark)	$\Delta G_b^{\ominus}$	kcal/mol	−3.45±0.01
		$\Delta H_b^{\ominus}$	kcal/mol	−33.4±0.3
		$\Delta C_{p,b}$	kcal/mol/K	0.92±0.01
		N		2.3±0.02
	to ***F***	$\Delta G_b^{\ominus}$	kcal/mol	−13.91±0.06
		$\Delta H_b^{\ominus}$	kcal/mol	−70.2±1.3
		$\Delta C_{p,b}$	kcal/mol/K	2.02±0.09
Fluorescent glucose complex states	α (GlcP)	R_o		0.275±0.003
		$\partial R/\partial T$	K^{-1}	0.00275±0.00001
	β (GlcP)	R_o		2.664±0.004
		$\partial R/\partial T$	K^{-1}	−0.00686±0.00001
		$\partial R/\partial[Glc]$	M^{-1}	0.033±0.0007
	$\boldsymbol{\alpha}$ (GlcF)	R_o		1.459±0.004
		$\partial R/\partial T$	K^{-1}	−0.00122±0.00002
		$\partial R/\partial[Ca^{2+}]$	M^{-1}	−0.208±0.009
	$\boldsymbol{B_1}$	R_o		1.23±0.01
		$\partial R/\partial T$	K^{-1}	−0.0017±0.00003
		$\partial R/\partial[Glc]$	M^{-1}	0.0428±0.0005
		$\partial R/\partial[Ca^{2+}]$	M^{-1}	−0.307±0.003
	$\boldsymbol{B_2}$	R_o		0.47±0.01
		$\partial R/\partial T$	K^{-1}	0.00066±0.00004
		$\partial R/\partial[Glc]$	M^{-1}	0.027±0.001
		$\partial R/\partial[Ca^{2+}]$	M^{-1}	0.082±0.009
Fluorescent unfolded state(s)	$\boldsymbol{U_0}$	R_o		0.58±0.02
		$\partial R/\partial T$	K^{-1}	0.00056±0.00005
		$\partial R/\partial[Glc]$	M^{-1}	0.019±0.001
		$\partial R/\partial[Ca^{2+}]$	M^{-1}	1.47±0.06
	α (CaU1a)	R_o		3.352±0.008
		$\partial R/\partial T$	K^{-1}	−0.00794±0.00002
		$\partial R/\partial[Glc]$	M^{-1}	0.0940±0.0006
	β (CaU1a)	R_o		2.551±0.007
		$\partial R/\partial T$	K^{-1}	−0.00548±0.00002

(continued on next page)

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

Table 4. (*continued*)

Parameter class	Parameters		Fit values[b]
	Variable	Units	
	$\partial R/\partial[Glc]$	M^{-1}	0.0830±0.0009
	$\partial R/\partial[Ca^{2+}]$	M^{-1}	0.032±0.01

NEST, nested equilibrated states tree.
[a] See Table 2 for definition of symbols and Table 3 for additional notes.
[b] Parameter variation is estimated using bootstrap analysis.

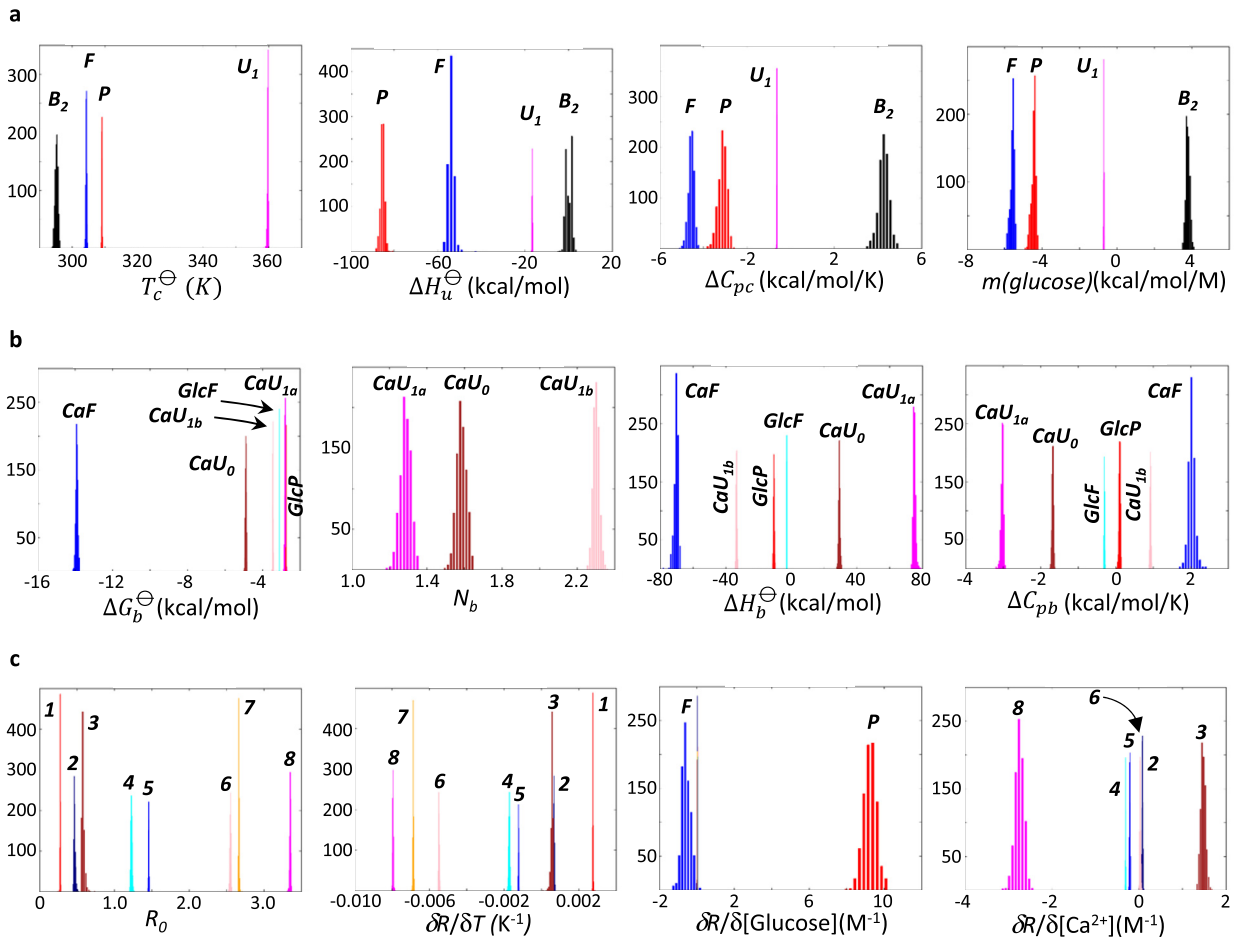


Fig. 10. Bootstrap analysis of NEST6 global fit parameters. View Table 4 for averaged values. (**a**) Parameters for the conformational exchange reactions (labels indicate the state for pairwise exchange reactions). (**b**) Parameters for ligand-binding reactions (labels correspond to reaction names in Fig. 9a). (**c**) Parameters for the fluorescence hyperplanes. Fluorescent states are numbered: 1, α of *GlcP*; 2, $\boldsymbol{B_2}$; 3, $\boldsymbol{U_0}$; 4, $\boldsymbol{B_1}$; 5, α of *GlcF*; 6, β of *CaU1a*; 7, β of *GlcP*; 8, α of *CaU1a*. NEST, nested equilibrated states tree.

binding sites in $\boldsymbol{U_1}$ were divided into fluorescently responsive and nonfluorescent (dark) sites. The $\boldsymbol{U_0}$ state also has multiple Ca^{2+} sites that do not evoke a direct fluorescent response.

The NEST6 model was fit to the combined 29,184 observations of the eight thermal glucose-binding landscapes with parameter variances similar to individually fit landscapes (Table 4; Table S2). The four-dimensional fluorescent landscape and distributions of underlying conformational states are visualized in a movie of the entire phase space (Supplementary Material).

Bootstrap analysis revealed that each of the 64 parameters in the model was well constrained by the data (Table 4; Fig. 10). For each reaction, parameter precision is correlated with its data coverage in

phase space, the magnitude of associated signal features, and evocation of direct or indirect effects on fluorescence. Precisions of conformational exchange reactions are consequently ordered as: $\mathbf{U}_0 \rightleftharpoons \mathbf{F} > \mathbf{U}_0 \rightleftharpoons \mathbf{P} > \mathbf{U}_0 \rightleftharpoons \mathbf{U}_1 > \mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$. The ligand-binding reactions are more precisely determined for glucose than Ca^{2+}. Binding of Ca^{2+} is more precisely determined in the folded than unfolded state. The comprehensive model faithfully reproduces features in individual landscapes (Fig. 9b–d), including the "ridge" observed at elevated Ca^{2+} concentrations (Figs. 8 and 9g).

In terms of ESS analysis (see previously), the improvements in the fit to the entire dataset as a consequence of the extra complexity of NEST6 relative to NEST3 is highly significant, with a $F_{6|3}$ statistic of 46.0, which corresponds to $p_{6|3} < 10^{-300}$.To test the independence of the NEST6 parameters, all pairwise Pearson correlation coefficients, r_{ij}, were calculated (Theory section). For pairs with $|r_{ij}| > 0.5$ with p values < 0.1, the pairwise distribution of values was inspected visually to determine whether the correlation was qualitatively significant. None of the pairs exhibited bimodal (or higher order) correlations, nor did the correlations extend over regions that exceeded the variance of each individual parameter. The NEST6 model parameters therefore are independent of each other.

Discussion

The redistribution of conformational states by differential ligand binding is complex

The temperature-dependent distributions of the ***P***, ***F***, $\boldsymbol{U_1}$, and $\boldsymbol{U_0}$ states exhibit complex responses to Ca^{2+} and glucose. The influence of embedded reactions on the redistribution of parent states is illustrated by the effect of Ca^{2+} binding on ***F*** and ***P***, as indicated by the cartoons of the dominant states (Fig. 9e–g). In the absence of Ca^{2+}, the partially folded ***P*** state dominates over ***F*** in the folded temperature range (Fig. 9e), whereas the reverse is true in the presence of Ca^{2+} (Fig. 9f and g). The partially unfolded protein, ***P***, is intrinsically ~5 K more thermostable than the folded protein, ***F***, in the absence of Ca^{2+} (Table 4). The high-affinity Ca^{2+}-binding site is embedded in ***F***, whereas ***P*** does not bind Ca^{2+}. Consequently, high-affinity Ca^{2+} binding to ***F*** increases the thermostability of this state, such that its population exceeds that of ***P*** in the presence of Ca^{2+}.

F is a member of the four-state ensemble ***F***, ***P***, $\boldsymbol{U_1}$, and $\boldsymbol{U_0}$. Changes in the stability of any one member affect the fractional distribution of all states. Consequently, $\boldsymbol{U_1}$ and $\boldsymbol{U_0}$ shift to elevated temperatures in the presence of Ca^{2+} (i.e. the protein is thermostabilized by Ca^{2+}). However, ***F*** is not the only state that binds Ca^{2+}, $\boldsymbol{U_1}$ and $\boldsymbol{U_0}$ also do. Their distribution therefore also shifts in response to Ca^{2+} binding. At high Ca^{2+} concentrations (Fig. 9g), this leads to the dominance of $\boldsymbol{U_1}$ to form the fluorescent ridge in the landscape, and for $\boldsymbol{U_1}$ and $\boldsymbol{U_0}$ to crowd out ***F*** at elevated temperatures (i.e. Ca^{2+}-mediated thermal *de*stabilization). The subtleties of these ligand-dependent redistributions are further illustrated in Fig. S9.

Solvent-accessible surface areas determine the intrinsic distribution of conformational states

The magnitude of the heat capacity, ΔC_{pc}, and osmolyte effect, m, of conformational exchange reactions is determined by the size of the change in solvent-accessible surface area (SASA) that accompanies the reaction. Their sign, which is the same for both, is determined by exposure (negative) or burial (positive) of hydrophobic residues in the SASA. The two quantities therefore are linearly correlated [39] (Fig. S10).

All the four states in the thermal (un)folding conformational ensemble have negative ΔC_{pc} values, indicating that their formation buries hydrophobic SASA relative to the U_0 ground state. Based on the ordering of the ΔC_{pc} values, the relative size of these buried areas ($\Delta SASA(\mathbf{F}) > \Delta SASA(\mathbf{P}) > \Delta SASA(\mathbf{U}_1)$ is consistent with the P state being a partially folded form of ***F***. Formation of $\mathbf{U}_1$ buries a small (8-fold less than ***F***) hydrophobic SASA, indicating that $\mathbf{U}_1$ has some residual structure in its thermally unfolded state, which is sufficient for Ca^{2+} binding to evoke a change in the conjugated Acrylodan fluorescence (Fig. 4g and h).

The exchange reaction between the $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ conformational states is an example of a hierarchical conformational change and is embedded within ***F***. The apparent m and ΔC_{pc} values combine the inherited values of the parent state (because the $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ substates are part of F) with the contribution of the child reaction. Consequently, the values of the child reaction considered in isolation are given by the differences in the absolute values [1] of the parent and child reaction (Table 4). In this case, the intrinsic ΔC_{pc} for the exchange between $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ is 0.3, indicating that the change in SASA is ~15-fold smaller for $\mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$ than for $\mathbf{U}_0 \rightleftharpoons \mathbf{F}$. It is therefore unlikely that this reaction involves a conformational exchange in the protein. Instead, the two states could correspond to different conformations of the Acrylodan conjugate, which alter protein-fluorophore interactions, changing the solvent accessibility of the protein surface. The $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ states are absent in the ***P*** state in which the C-terminal domain is partially unfolded. Therefore, the interactions between the postulated Acrylodan conformations and the protein

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

reside in this domain, where the fluorophore is attached to 183C (Fig. 1a).

Fluorescence emission spectra are intrinsically dependent on temperature and glucose

The hyperplanes describing the intrinsic ratiometric fluorescence of the fluorescent states each are linearly dependent on temperature, glucose, and Ca^{2+}. The absolute magnitudes of these dependencies are unique for each state, but their relative magnitudes all follow the same pattern (Table 4): $\partial R/\partial T < \partial R/\partial [Glc] < \partial R/\partial [Ca^{2+}]$. The change in the fluorescent ratio indicates a change in color. Such solvatochromic shifts are the consequence of changes in the dipolar relaxation of the fluorescent dipoles by the surrounding solvent [40,41]. Temperature and solutes (glucose, Ca^{2+}) alter the dielectric constant and refractive index of water, thereby changing the solvent dipolar relaxation, as described by the Lippert-Mataga model [40,41]. The temperature dependencies of the water refractive index [42] and dielectric constant [43] are much less than their dependence on glucose [43,44], consistent with the rank ordering of the slopes for the dimensions corresponding these state variables in the intrinsic fluorescence hyperplanes of the states.

The glucose-bound $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ states in ***F***, the glucose-bound ***P*** state, the $\boldsymbol{U_0}$ and the apo- and Ca^{2+}-bound states of $\boldsymbol{U_1}$ each exhibit similar green colors (and corresponding ratiometric signal values). However, these green states are distinct by virtue of the different protein environments surrounding their green Acrylodan conformation. Accordingly, each state is associated with a unique combination of temperature-, glucose-, and Ca^{2+}-dielectric dependencies (Table 4), the values of which are well constrained by the experimental data (Fig. 10c). Despite apparent similarity in their color, the NEST6 model therefore readily distinguishes between these green states in different locations of the phase space.

Conclusions

The NEST numerical method provides a general solution for building direct, quantitative links between biomolecular structure and function. Functional relationships are represented as a hierarchy of ligand-binding and conformational exchange reactions embedded within states of parent reactions. Individual conformational exchange and ligand-binding reactions are treated as building blocks that are linked together in tree representations, creating arbitrarily complex structure-function relationships. Despite the complexity that can be created, these relationships are syntactically straightforward to implement and interrogate. Graphical rendering of the corresponding tree clarifies the logic of equilibrium linkages and provides a visually intuitive representation of structure-function relationships. The method enables systematic buildup and evaluation of structure-function models, as increasing mechanistic detail is introduced and constrained by additional experimental data, systematically implementing Einstein's maxim [45] that "Everything should be made as simple as possible, but no simpler".

The four-dimensional landscape of the ecGGBP183C•Acrylodan conjugate provides an information-rich, minimalist experimental model that represents major macromolecular structure-function relationships. In this system, conformational coupling in which ligand binding alters the distribution of conformational states [6] is manifested as ligand-dependent changes in protein thermostability. The global fit of the complete phase space required four major states (two folded, ***F*** and ***P***, and two unfolded, $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$). The distribution of states in this ensemble in response to glucose, Ca^{2+} and temperature (Fig. 9 and supplementary movie) is highly complex because a relative change in the free energy of any one state owing to differences in temperature dependence, or binding of Ca^{2+} and glucose, affects the distribution of all. This experimental model system clearly demonstrates that multistate conformational ensembles in which all states are in exchange with each other exhibit behaviors that are both qualitatively different and more versatile than collections of two-state reactions.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.jmb.2019.12.039

The exchange reaction between the $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ conformational states in ecGGBP is an example of a hierarchical conformational change that is embedded within another state (the Ca^{2+} and glucose ternary complex of ***F***) and not in exchange with other conformational states. Such hierarchically organized conformational changes and motions are a common feature in proteins and arise from correlated motions on different length scales that range from single side chains to clusters of residues and domains [27,28]. Motions at smaller length scales need not be in exchange with each other if they are located in different larger scale units, and therefore are described by independent conformational exchange reactions. Changes in the distribution of such embedded conformations therefore do not alter the distribution of parent states. In this case, the $\mathbf{B}_1 \rightleftharpoons \mathbf{B}_2$ exchange reflects a motion in the folded state of the domain to which the fluorophore is covalently coupled. Changes in the distribution of the $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ states do not alter the distribution of states of their parental ensemble, ***F***, ***P***, $\boldsymbol{U_1}$, and $\boldsymbol{U_0}$.

The Ca^{2+}-binding sites in the four main conformational states of ecGGBP are remarkably diverse, ranging from none (***P***) to one high-affinity site (***F***), to

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

an ensemble of multiple sites with different low affinities ($\boldsymbol{U_0}$ and $\boldsymbol{U_1}$). The profound influence of the low-affinity sites on the behavior of the system (Fig. 7) emphasizes that conformational coupling arises not only from differences in ligand-binding affinities but also in stoichiometry.

Cooperativity between different ligands is revealed by the dependence of glucose on Ca^{2+}. This effect is also the consequence of conformational coupling as Ca^{2+} switches the protein from the ***P*** state (no Ca^{2+}) to the ***F*** state (single, high-affinity Ca^{2+}). Because these two states have different glucose affinities (Table 4), the apparent response to glucose is Ca^{2+}-dependent.

The effects of glucose in ecGGBP emphasize that all conformational exchange reactions are sensitive to osmolyte effects owing to changes in SASA between states. Consequently, multistate conformational ensembles encode more complex osmolyte responses than two-state reactions. Most thermodynamic treatments of osmolyte effects have been limited to protein folding reactions (denaturants, stabilizers) [29,46], but the biology of osmolytes reveals that many enzymes activities also are sensitive to osmolytes [47]. All proteins that integrate multiple functions or control elements are likely to exhibit osmolyte-sensitive conformational coupling. Osmolyte perturbation of the underlying conformational exchange reactions therefore can provide an experimental probe for these reactions, which has not yet been widely used.

Temperature is not used often as a state variable, limiting most experiments to isothermal observations. This is in part because of the need for cumbersome, nonlinear equations that complicate analytical treatments of thermal reactions. The NEST numerical method makes the treatment of thermal and isothermal observations equally accessible. Temperature dependencies are encoded by differences in SASA, and multistate ensembles can exhibit a greater variety of response patterns than two-state reactions, analogous to osmolyte sensitivity. Temperature dependence therefore can provide an information-rich experimental probe of mechanism.

As demonstrated here, statistical thermodynamic models can link a large number of binding and conformational exchange reactions, with a concomitantly large number of degrees of freedom (thermodynamic and signal parameters). Consequently, care has to be taken to evaluate whether experimental data sufficiently constrain the model or whether the model "overfits" the data. We use four methods to assess the robustness of the relationships between our models and experimental data. First, we carry out a visual inspection to determine whether a model reproduces key features in the data. Second, we use the bootstrap method to calculate the variation (precision) of each parameter. Third, we calculate the pairwise correlation between parameter values in this bootstrap data set to determine whether the parameters are independent or correlated. Fourth, as more complex models are developed to represent the same data, the validity of the increase in complexity is assessed using the ESS principle. NEST2 and NEST3 are both elaborations of NEST1. NEST2 is not a statistically significant improvement over NEST1, whereas NEST3 is, even though it is more complex than NEST2. This case illustrates how complexity can be a necessary condition for building appropriate models. The NEST6 model meets all four criteria; its complexity is supported by the experimental observations.

Sophisticated statistical thermodynamics has been critical to understand the biology of important proteins such as hemoglobin, F_1F_o ATPase, molecular motors, and membrane channels, in which coupling of multiple conformational states to ligand binding is the essence of their function. NEST methodology uses two-state and multistate building blocks, seamlessly incorporates isothermal or thermal treatments, and straightforwardly builds complex linkage relationships between reactions. The NEST software is intended to make sophisticated statistical thermodynamic analysis of biomolecular structure-function relationships available to a wide community of life scientists.

Materials and methods

Protein expression, purification, and labeling

A full-length gene encoding a synthetic open reading frame for the *E. coli* GGBP amino sequence, lacking its leader peptide [11], flanked by 5' (122 bp) and 3' (103 bp) regulatory regions encoding the T7 promoter and transcription terminator respectively was assembled from oligonucleotides (80−100 bases) synthesized in-house (Mermade 192 DNA Synthesizer, BioAutomation) using an automated PCR-mediated gene assembly procedure [48] and cloned into the *Eco*RI and *Pst*I restriction sites of pUC19 (Invitrogen).

The expression plasmids were transformed into KRX-competent cells (Promega), and grown overnight on Luria Broth (LB) agar (100 mg/ml ampicillin) at 37 °C. A single colony was picked and grown in terrific broth (TB) media (100 mg/ml ampicillin) overnight at 37 °C. The culture was diluted 1:20 into 500 ml of fresh TB (100 mg/ml ampicillin, 1 mM $CaCl_2$), grown to an optical density of $A_{600} = 0.5$ at 37 °C in vigorously aerated shaker flasks, induced by the addition of 2.5 ml of rhamnose 20% (w/v), grown for a further 3−4 h. The cells were harvested by centrifugation (5000 rpm for 10 min) after decanting the supernatant, the cell pellet was stored at −80 °C until needed. The cell pellets were thawed and resuspended in 10 ml binding buffer (20 mM imidazole, 20 mM MOPS, 500 mM NaCl, 1 mM $CaCl_2$, pH 7.8). After resuspension, cells were lysed with 20-ml BugBuster HT (EMD Millipore), incubating at room temperature for 20 min on a slowly shaking platform. Insoluble cell debris was removed by centrifugation (15,000 rpm for

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

20 min at 4 °C) producing a clarified lysate from which the recombinant protein was purified by batch nickel-charged immobilized ion affinity chromatography (IMAC). Resuspended IMAC agarose beads (5 ml; Sigma-Aldrich P6611) were added to the lysate. After incubation at 4 °C in a Mini LabRoller (Labnet International) for 1 h, the beads were washed at least five times with binding buffer. The immobilized protein beads were resuspended in labeling buffer (20 mM MOPS, 100 mM NaCl, 1 mM $CaCl_2$, pH 6.9) and labeled overnight (4 °C, rotating end-over-end) with Acrylodan (5-fold stoichiometric excess over protein). After two rinses with labeling buffer to remove unincorporated label, the proteins were eluted from the beads. To elute labeled protein from the IMAC beads, 6 ml of elution buffer (400 mM imidazole, 500 mM NaCl, 1 mM $CaCl_2$, 20 mM MOPS, pH 7.8) was added, incubated for 30 min (4 °C, rotating end-over-end), and the beads removed by centrifugation. After dialysis of the eluate against three changes of assay buffer (20-mM MOPS, 20-mM KCl, with or without 1-mM $CaCl_2$ depending on experiment, pH 7.4), using 10 kDa semipermeable membrane (Snakeskin tubing, Thermo Scientific), the fluorescent conjugates were concentrated in a 10 kDa cutoff spin concentrator (Vivaspin, GE Healthcare). Protein purity was assessed by protein purity was assessed by gel electrophoresis (SDS/PAGE). Labeled protein concentrations were determined spectroscopically (Nanodrop1000) at 280 nm (using extinction coefficients calculated from their sequence [49,50]), or at the fluorophore absorbance peak (ϵ_{391nm} (Acrylodan) = 20,000 $M^{-1}cm^{-1}$).

Determination of temperature- and ligand-dependent fluorescent landscapes

Temperature-dependent ligand titrations were measured in a real-time PCR machine (Roche LightCycler 480), using manufacturer-provided filter sets to control excitation and emission wavelengths: excitation at 440 nm (half bandwidths (HBW) 35 nm, Acrylodan excitation optimum is 391 nm); emission 488 nm, and 510 nm (HBW 20 nm). Temperatures were varied from 293 K to 368 K in one-degree steps. At each temperature, data was collected at 1-s intervals for 60 s at which point the signal had relaxed to a steady value associated with the new temperature. Collection times were adjusted to minimize photobleaching (which was not detectable under these experimental conditions). The signal was recorded as the ratio of the intensities at 488 nm and 510 nm. Ratiometry [25,26] removes well-to-well and time-dependent variations that arise from fluctuations in fluorophore concentration (volumetric error and photobleaching), excitation light intensity, and detection efficiencies (microtiter plate and instrumentation). Data reduction software was developed to convert these observations into time-independent datasets that record fluorescence as a function of temperature for each well and associate wells with their concentration of titrants. Management tools were developed to maintain a database of titrations and their analyses.

Titration series were set up in 384-well microtiter plates, by mixing 5-μL protein conjugate to a final protein concentration of 10 μM with 15-μL titrant solution (varying concentrations of glucose and Ca^{2+}, in 20 mM MOPS, 20 mM KCl, pH 7.4) by hand. Glucose concentrations were varied from 0 to 1.7 M and calcium concentrations were varied from 0 to 100 mM. EGTA (1 mM) was added to chelate adventitious metal for experimental data points that contained 0 mM $CaCl_2$. Titration series (12-, 24-, and 48-points) were created using a Tecan Genesis liquid-handler robot programmed with software developed by us to construct logarithmic, linear, or logarithmic-linear concentration intervals.

Determination of emission intensity spectra

Glucose- and wavelength-dependent emission intensities were recorded on a Nanodrop3300 (Thermo Scientific) at room temperature using its 365-nm LED, which is close to the optimal excitation wavelength of Acrylodan (391 nm).

Software and data

The NEST method is encoded in an eponymous software package written in Python [51], the source code and documentation for which can be downloaded from <https://doi.org/10.5061/dryad.rr4xgxd5b>. The experimental data, fit models, and scripts used for fits or analysis can be downloaded from <https://doi.org/10.5061/dryad.rr4xgxd5b>.

Acknowledgements

The authors thank Drs. T.G. Oas, M. Boyce, P. Zhou and B.E. Coggins for many helpful discussions. This work was supported in part by a Corporate Research Agreement between Becton, Dickinson and Company (Research Triangle Park, North Carolina, USA) and Duke University (contract ARG20112333), and by the Department of Biochemistry

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.12.039.

Author contributions

H.W.H. contributed to conceptualization, software, investigation, visualization, writing the original draft, and in reviewing and editing. M.J.A. contributed to conceptualization, investigation, visualization, writing the original draft, and in reviewing and editing.

Received 17 September 2019;
Received in revised form 3 December 2019;
Accepted 17 December 2019
Available online xxxx

Keywords:
statistical thermodynamics;

periplasmic binding proteins;
glucose biosensor;
NEST

Abbreviations

NESTnested equilibrated states tree; EGTAethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N'-tetraacetic acid

References

[1] S. Dill Kab, Molecular Driving Forces, in: Garland Science, second ed., Taylor & Francis Group, LLC, 2011.
[2] R. Phillips, J. Kondev, J. Theriot, H.G. Garcia, Physical Biology of the Cell, second ed., Garland Science, 2013.
[3] J. Wyman, S.J. Gill, Binding and Linkage. Functional Chemistry of Biological Macromolecules: Mill Valley, University Science Books, CA, 1990.
[4] I. Bahar, R.L. Jernigan, K.A. Dill, Protein Actions, Principles & Modeling, Garland Science, 2017.
[5] H.N. Motlagh, J.O. Wrabl, J. Li, V.J. Hilser, The ensemble nature of allostery, Nature 508 (2014) 331–339.
[6] E. Freire, Statistical thermodynamic linkage between conformational and binding equilibria, Adv. Protein Chem. 51 (1998) 255–279.
[7] V.J. Hilser, J.O. Wrabl, H.N. Motlagh, Structural and energetic basis of allostery, Annu. Rev. Biophys. 41 (2012) 585–609.
[8] D.E. Knuth, The Art of Computer Programming, third ed., Addison Wesley, 1997.
[9] R. Sedgewick, K. Wayne, Algorithms, fourth ed., Addison Wesley, 2011.
[10] S.S. Skiena, Algorithm Design Manual, Springer, 2010.
[11] R.M. de Lorimier, J.J. Smith, M.A. Dwyer, L.L. Looger, K.M. Sali, C.D. Paavola, et al., Construction of a fluorescent biosensor family, Protein science : Publ. Protein Soc. 11 (2002) 2655–2675.
[12] K. Weidemaier, A. Lastovich, S. Keith, J.B. Pitner, M. Sistare, R. Jacobson, et al., Multi-day pre-clinical demonstration of glucose/galactose binding protein-based fiber optic sensor, Biosens. Bioelectron. 26 (2011) 4117–4123.
[13] K. Judge, L. Morrow, A.G. Lastovich, D. Kurisko, S.C. Keith, J. Hartsell, et al., Continuous glucose monitoring using a novel glucose/galactose binding protein: results of a 12-hour feasibility study with the becton dickinson glucose/galactose binding protein sensor, Diabetes Technol. Ther. 13 (2011) 309–317.
[14] F.G. Prendergast, M. Meyer, G.L. Carlson, S. Iida, J.D. Potter, Synthesis, spectral properties, and use of 6-acryloyl-2-dimethylaminonapthalene (acrylodan), J. Biol. Chem. 258 (1983) 7541–7544.
[15] N.K. Vyas, M.N. Vyas, F.A. Quiocho, A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis, Nature 327 (1987) 635–638.
[16] N.K. Vyas, M.N. Vyas, F.A. Quiocho, Sugar and signal-transducer binding sites of the Escherichia coli galactose chemoreceptor protein, Science 242 (1988) 1290–1295.
[17] J.Y. Zou, M.M. Flocco, S.L. Mowbray, The 1.7 A refined X-ray structure of the periplasmic glucose/galactose receptor from Salmonella typhimurium, J. Mol. Biol. 233 (1993) 739–752.
[18] M.N. Vyas, N.K. Vyas, F.A. Quiocho, Crystallographic analysis of the epimeric and anomeric specificity of the periplasmic transport/chemosensory protein receptor for D-glucose and D-galactose, Biochemistry 33 (1994) 4762–4768.
[19] S. Sooriyaarachchi, W. Ubhayasekera, W. Boos, S.L. Mowbray, X-ray structure of glucose/galactose receptor from Salmonella typhimurium in complex with the physiological ligand, (2R)-glyceryl-beta-D-galactopyranoside, FEBS J. 276 (2009) 2116–2124.
[20] L.A. Luck, J.J. Falke, Open conformation of a substrate-binding cleft: 19F NMR studies of cleft angle in the D-galactose chemosensory receptor, Biochemistry 30 (1991) 6484–6490.
[21] M.M. Flocco, S.L. Mowbray, The 1.9 A x-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from Salmonella typhimurium, J. Biol. Chem. 269 (1994) 8931–8936.
[22] M.J. Borrok, L.L. Kiessling, K.T. Forest, Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures, Protein science : Publ. Protein Soc. 16 (2007) 1032–1041.
[23] M.J. Allert, S. Kumar, A. Bergeron, J. Miriyala, L.S. Beese, H.W. Hellinga, Conformational coupling between ligand binding and internal fluorophore twisting in reagentless fluorescent glucose biosensors derived from periplasmic binding proteins, ——, 2020. In preparation.
[24] M.J. Allert, H.W. Hellinga, Glucose Biosensors and Uses Thereof. Patent Application, 2016.
[25] A.P. Demchenko, The concept of lambda-ratiometry in fluorescence sensing and imaging, J. Fluoresc. 20 (2010) 1099–1128.
[26] A.P. Demchenko, Practical aspects of wavelength ratiometry in the studies of intermolecular interactions, J. Mol. Struct. 1077 (2014) 51–67.
[27] H. Frauenfelder, S.G. Sligar, P.G. Wolynes, The energy landscapes and motions of proteins, Science 254 (1991) 1598–1603.
[28] J.R. Lewandowski, M.E. Halse, M. Blackledge, L. Emsley, Protein dynamics. Direct observation of hierarchical protein dynamics, Science 348 (2015) 578–581.
[29] J.A. Schellman, Fifty years of solvent denaturation, Biophys. Chem. 96 (2002) 91–101.
[30] T. Arakawa, S.N. Timasheff, Stabilization of protein structure by sugars, Biochemistry 21 (1982) 6536–6544.
[31] M. Auton, D.W. Bolen, Predicting the energetics of osmolyte-induced protein folding/unfolding, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15065–15068.
[32] T.O. Street, D.W. Bolen, G.D. Rose, A molecular mechanism for osmolyte-induced protein stability, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 13997–14002.
[33] H. Fraunfelder, F. Parak, R.D. Young, Conformational substates in proteins, Annu. Rev. Biophys. Biophys. Chem. 17 (1988) 451–479.
[34] D.M. John, K.M. Weeks, van't Hoff enthalpies without baselines, Protein Sci. : Publ. Protein Soc. 9 (2000) 1416–1419.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039

[35] W.H.T. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes. The Art of Scientific Computing, third ed., Cambridge University Press, 2007.
[36] S. D'Auria, A. Ausili, A. Marabotti, A. Varriale, V. Scognamiglio, M. Staiano, et al., Binding of glucose to the D-galactose/D-glucose-binding protein from Escherichia coli restores the native protein secondary structure and thermostability that are lost upon calcium depletion, J. Biochem. 139 (2006) 213–221.
[37] P. Herman, J. Vecer, I. Barvik Jr., V. Scognamiglio, M. Staiano, M. de Champdore, et al., The role of calcium in the conformational dynamics and thermal stability of the D-galactose/D-glucose-binding protein from Escherichia coli, Proteins 61 (2005) 184–195.
[38] L.A. Luck, J.J. Falke, 19F NMR studies of the D-galactose chemosensory receptor. 2. Ca(II) binding yields a local structural change, Biochemistry 30 (1991) 4257–4261.
[39] J.K. Myers, C.N. Pace, J.M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, Protein Sci. 4 (1995) 2138–2148.
[40] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Springer, New York, 2006.
[41] B. Valeur, M.N. Berberan-Santos, Molecular Fluorescence. Principles and Applications, Wiley, Weinheim, 2012.
[42] G. Abbate, U. Bernini, E. Ragozzion, F. Somma, The temperature dependence of the refractive index of water, J. Phys. D Appl. Phys. 11 (1978) 1167–1172.
[43] CRC, Handbook of Chemistry and Physics, 86th ed, CRC Press, 2005.
[44] X. Liao, V.G. Raghavan, V. Meda, V.A. Yaylayan, Dielectric properties of supersaturated alpha-D-glucose aqueous solutions at 2450 MHz, J. Microw. Power Electromagn. Energy 36 (2001) 131–138.
[45] A. Einstein, On the method of theoretical physics, Philos. Sci. 1 (1934) 163–169.
[46] T. Arakawa, Protein-solvent interaction, Biophys Rev 10 (2018) 203–208.
[47] P.H. Yancey, M.E. Clark, S.C. Hand, R.D. Bowlus, G.N. Somero, Living with water stress: evolution of osmolyte systems, Science 217 (1982) 1214–1222.
[48] J.C. Cox, J. Lape, M.A. Sayed, H.W. Hellinga, Protein fabrication automation, Protein Sci. : A Publ. Protein Soc. 16 (2007) 379–390.
[49] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319–326.
[50] P. Artimo, M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, et al., ExPASy: SIB bioinformatics resource portal, Nucleic Acids Res. 40 (2012) W597–W603.
[51] A. Martelli, Python in a Nutshell, O'Reilly, Beijing, 2006.
[52] R. Tam, M.H. Saier Jr., Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria, Microbiol. Rev. 57 (1993) 320–346.
[53] R.P. Berntsson, S.H. Smits, L. Schmitt, D.J. Slotboom, B. Poolman, A structural classification of substrate-binding proteins, FEBS Lett. 584 (2010) 2606–2617.

Please cite this article as: M. J. Allert and H. W. Hellinga, Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.12.039