Published in final edited form as:

Protein Pept Lett. 2014 ; 21(8): 752–765.

Thermodynamic stability and flexibility characteristics of antibody fragment complexes

Tong Li¹, Deeptak Verma¹, Malgorzata B. Tracka², Jose Casas-Finet³, Dennis R. Livesay^{1,*}, and Donald J. Jacobs^{4,*}

¹Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

²Department of Formulation Sciences, MedImmune, Cambridge, CB21 6GH, UK

³Analytical Biochemistry Department, MedImmune, Gaithersburg, MD 20878. USA, MedImmune Inc, Gaithersburg, MD 20878, USA

⁴Department of Physics and Optical Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

Abstract

Free energy landscapes, backbone flexibility and residue-residue couplings for being co-rigid or co-flexible are calculated from the minimal Distance Constraint Model (mDCM) on an exploratory dataset consisting of VL, scFv and Fab antibody fragments. Experimental heat capacity curves are reproduced markedly well, and an analysis of quantitative stability/flexibility relationships (QSFR) is applied to a representative VL domain and several complexes in the scFv and Fab forms. Global flexibility in the denatured ensemble typically decreases in the larger complexes due to domain-domain interfaces. Slight decreases in global flexibility also occur in the native state of the larger fragments, but with a concurrent large increase in correlated flexibility. Typically, a VL fragment has more co-rigid residue pairs when isolated compared to the scFv and Fab forms, where correlated flexibility appears upon complex formation. This context dependence on residue-residue couplings in the VL domain across length scales of a complex is consistent with the evolutionary hypothesis of antibody maturation. In comparing two scFv mutants with similar thermodynamic stability, local and long-ranged changes in backbone flexibility are observed. In the case of anti-p24 HIV-1 Fab, a variety of QSFR metrics were found to be atypical, which includes comparatively greater co-flexibility in the VH domain and less co-flexibility in the VL domain. Interestingly, this fragment is the only example of a polyspecific antibody in our dataset. Finally, the mDCM method is extended to cases where thermodynamic data is incomplete, enabling high throughput QSFR studies on large numbers of antibody fragments and their complexes.

Keywords

Antibody Structure; Computational Biology; Distance Constraint Model; Free Energy Landscape; and Quantitative Stability/Flexibility Relationships

*To whom correspondence should be addressed drivesa@uncc.edu and djacobs1@uncc.edu.

INTRODUCTION

Antibody drugs play a critical role in controlling many types of diseases such as infectious diseases, cancer, autoimmune diseases and inflammation. In fact, nearly thirty recombinant monoclonal antibodies (mAbs) are currently approved by the FDA [1]. Smaller recombinant antibody fragments like the classic fragment antigen binding (Fab) and single chain Fv fragment (scFv), and their variants, are emerging as credible alternatives due to specificity retention, plus increased tissue penetration and simplified expression [2, 3]. Unfortunately, antibody fragments have poor stability, which limits their utility [4, 5]. The role of domain-domain interactions on antibody stability has not been explored sufficiently. Moreover, how dimerization affects antibody dynamics is also an open question. A systematic investigation of these properties over a dataset of fragments and their complexes would thus provide invaluable insights that could be leveraged for the molecular design of antibody fragments suitable for therapeutic applications.

Unfortunately, predicting thermodynamic properties of proteins is especially difficult because many types of interactions, such as covalent bonding, hydrogen bonding, packing, solvation, hydrophobic effects and electrostatic effects play essential roles. Furthermore, a protein structure is comprised of heterogeneous microenvironments that change dynamically as the conformational state of the protein changes [6]. Thus, it remains a challenge to develop models that can accurately predict protein flexibility and stability in practical computing times. The stability of a folded protein is strongly dependent on the thermodynamic environment, under which the free energy difference between folded state and unfolded state determines the degree of stability. Despite continual improvements over decades of research on free energy calculations, the difficulty faced in obtaining sufficient sampling of conformational space using molecular dynamics (MD) or Monte Carlo simulation techniques continues to limit the applicability of these methods, especially for large systems. Moreover, even when accurate estimates for the free energy can be made through extensive conformational sampling, the required computer resources preclude high-throughput applications that systematically vary temperature, pressure, pH, and concentration of co-solvents [6]. As such, in high-throughput optimizing design applications it is not feasible to calculate free energy and other thermodynamic properties of proteins from brute force simulation methods.

To address the need for rapid free energy calculations, an Ising-like model called mDCM (minimal distance constraint model) [7] was developed. The conceptual idea behind mDCM is to apply constraint theory to a free energy decomposition scheme where enthalpy and entropy components are assigned to mechanical distance constraints that model microscopic interactions. In this way, a protein system is mapped into a mechanical framework. The total enthalpy of a given framework is calculated as a linear sum of enthalpy components. In contrast, the total entropy is nonadditive, and is calculated by employing network-rigidity as a long-range mechanical interaction [8]. Specifically, network rigidity is used to completely identify independent constraints and degrees of freedom (DOF) within the network. The mDCM defines a free energy functional in terms of three phenomenological parameters $\{u, v, \delta_n\}$ that respectively represent an average protein to solvent H-bond energy, the average

native-like torsion constraint energy and native-like torsion constraint entropy. The value of these parameters is readily determined by fitting to experimental heat capacity, usually obtained from differential scanning calorimetry (DSC). Even though fitting parameters are employed, it should be stressed that the mDCM is an all-atom statistical mechanical treatment of the system. The model predicts the free energy landscape (FEL), all relevant thermodynamic response functions and many mechanical properties such as local flexibility/rigidity profiles and cooperative correlations in protein motions. Previously, the mDCM has been successfully employed to gain insight into how protein stability and dynamics are interconnected, and how coupling between structural elements of a protein affect thermodynamic properties for a diverse set of proteins, such as lysozyme [9], thioredoxin [10, 11], ubiquitin [7], an orthologous RNase H pair [12] periplasmic-binding proteins [13], among others.

In this report, we consider several examples of antibody fragments that span three different size scales (VL, scFv and Fab). The experimental heat capacity curves are successfully reproduced across the entire dataset, and the corresponding quantitative stability/flexibility relationships (QSFR) are characterized. The salient features of how co-rigidity in residue couplings give way to an increase in co-flexibility within the VL domain upon complex formation supports the evolutionary hypothesis of antibody maturation [14, 15]. According to this hypothesis, mutations in the VH domain occurs first to rigidify the combining site to improve the binding specificity between antibody and antigen, and second, mutations in the VL domain increase the binding affinity. A necessary condition for this hypothesis to be valid is for VL to be capable to support dynamical motions prior to the mutations. Furthermore, a few localized regions are identified where significant change in backbone flexibility occurs between two single site scFv mutations, confirming that an increase of rigidity upon mutation is a viable mechanism. These scFv mutants have nearly the same melting temperatures. This observation demonstrates that a double mutation can modify flexibility/rigidity without inducing significant change in thermodynamic stability. Revealed by QSFR in the native state ensemble, a greater degree of correlated motions appear in larger size complexes with a concurrent but small decrease in global flexibility. The increase in correlated flexibility is usually prominent in the VL domain. However, an atypical case identified as the sole outlier in the exploratory dataset, corresponding to the anti-p24 HIV-1 Fab, shows the VH domain as having a greater degree of correlated motions compared to the VL domain. Interestingly, anti-p24 HIV-1 Fab is known to have polyspecificity [16].

While these results from the comparative QSFR analysis are encouraging, large-scale application of the model will require a much larger dataset of fragments. However, as a practical matter, both structure and heat capacity data is often not available. For a missing structure, homology modeling is used to define the native state structure [manuscript in preparation]. For a given structure, the mDCM parameters are routinely obtained by fitting model predictions to experimental heat capacity data. In this report, we demonstrate for the first time that the mDCM is applicable across antibody fragment sizes, and the parameters are relatively conserved. In addition, parameterization is performed in the absence of experimental heat capacity curves assuming the melting temperature is known or estimated, which removes a major limiting factor. When experimental C_p curves are not available, an

iterative fitting approach is applied to ascertain the mDCM parameters starting from an ensemble of selected experimental C_p curves as an initial guess. The iterative procedure provides a narrow window of plausible parameters that can be used to complete the analysis within acceptable uncertainties. For the dataset under consideration here, as well as for a few other protein systems checked (unpublished results) this iterative procedure expands the utility of the mDCM to explore protein stability relationships across an entire protein family. In particular, going forward the mDCM can be employed to assess stability and flexibility properties of large numbers of antibody fragments and their complexes important to protein biologics.

MATERIALS AND METHODS

The minimal Distance Constraint Model

The first application of the DCM was to investigate helix–coil transitions using exact transfer matrix methods [8, 17]. Subsequently, a mean-field treatment was developed [7], making investigations of protein stability and flexibility computationally tractable [9, 11, 12, 18–22]. The model is based on a free energy decomposition scheme combined with constraint theory where structure is recast as a topological framework. Therein, vertices describe atomic positions and distance constraints that fix the relative atomic positions describe intramolecular interactions. From an input framework, a Pebble Game (PG) algorithm quickly identifies all rigid and flexible regions within structure [23, 24]. However, the PG does not model thermal fluctuations within the interaction network (i.e., the breaking and reforming of H-bonds). As such, the DCM was developed as a statistical mechanical model that introduces fluctuations into the network rigidity paradigm. Specifically, the DCM considers a Gibbs ensemble of network rigidity frameworks, each appropriately weighted based on its free energy. The free energy of each framework is calculated using free energy decomposition (FED). That is, each constraint is associated with a component enthalpy and entropy. The total enthalpy of a given framework is simply the sum over the set of distance constraints; however, as described below, the total entropy is calculated in a way that accounts for nonadditivity.

Within the mDCM applied to proteins, the number of native-like torsion constraints, N_{nat} , and number of H-bond constraints, N_{hb} , specifies a macrostate. Native torsions have reduced entropy relative to disordered states, meaning they correspond to good packing interactions. Together, the result of this dual view is that the model describes protein stability in terms of both packing and electrostatic interactions (note: salt bridges are considered to be a special case of H-bonds). The two variables define the FEL, where each node (N_{hb} , N_{nat}) specifies the topological macrostate of a protein. A free energy functional is defined for each macrostate by:

$$G(N_{hb}, N_{nat}) = U(N_{hb}) + u(N_{hb}^{max} - N_{hb}) + v(N_{nat}) - T \{ S_{conf}(N_{hb}, N_{nat}) + S_{mix}(N_{hb}, N_{nat}) \} \quad (1)$$

where U is the intramolecular H-bond energy, u is an average H-bond energy to solvent that occurs when an intramolecular H-bond breaks, v is the energy associated with a native-like torsion, $S_{conf}(N_{hb}, N_{nat})$ is the conformational entropy and $S_{mix}(N_{hb}, N_{nat})$ is the mixing entropy of the macrostate associated with the number of ways of distributing N_{nat} native-

torsions and N_{hb} H-bonds within the protein. To account for nonadditivity within entropy, the total conformational entropy, S_{conf} , is only calculated over the set of independent constraints [9] using:

$$S_{conf}(N_{nat}, N_{hb}) = R \left[\sum_{t \in hb} q_t \gamma_t + q_{nat} \delta_{nat} N_{nat} + q_{dis} \delta_{dis} (N_{tor} - N_{nat}) \right] \quad (2)$$

where the index t is over the full set of H-bonds that are identified from the input (crystal) structure, γ_t is the entropy of H-bond t , δ_{nat} and δ_{dis} respectively describe the entropy of a native-like and disordered torsion angle and N_{tor} is the total number of torsion angles. The values q_i are conditional probabilities for a constraint i to be independent when present, which is the attenuating factor that accounts for nonadditivity within free energy components. For a given framework, the PG is used to calculate the $\{q_i\}$ values therein, where constraints added to flexible regions are assigned $q_i = 1$ and constraints added to already rigid regions are assigned $q_i = 0$ because all the degrees of freedom have already been removed. The key to the mDCM is that it obtains the lowest possible upper bound limit on the conformational entropy because it preferentially places constraints with the lowest entropy first. In other words, it sums entropies only over a minimal set of the most constrictive yet independent interactions. Repeating this process many times, the mDCM Monte Carlo samples over many topological frameworks within a given macrostate, and the final $\{q_i\}$ are appropriately averaged. Because of the degeneracy in the native and disordered torsion parameters, it follows that for native torsions, $\sum_{i \in nat} q_i = q_{nat} N_{nat}$ and a similar simplification can be applied for the disordered torsions. These simplifications are reflected in Eq. (2), but the Monte Carlo sampling is done for each torsion constraint individually. Typically, less than 200 samples per macrostate are required for good statistics.

It is well understood that protein stability is dependent upon solvent condition (i.e., pH and ionic strength). These effects are modeled both explicitly and implicitly. First, using continuum electrostatics methods [25], the ionization state of the protein, which directly affects which interactions are present and which are missing, is explicitly accounted for. In addition, parameter values adjust to account for various solvent conditions. This is most clear in terms of u , which explicitly describes solvent H-bonding. In addition, the effects of solvent are also implicitly accounted for in the other two fitting parameters. The values of all energy terms (indicated by Roman characters) are kcal/mol, whereas all entropic terms (indicated by Greek characters) are unitless pure entropies. Hydrogen bond energies, U , are calculated from a modified [18] empirical potential [26]. The entropic cost associated with forming the H-bond, γ_t , is linearly related to its potential well depth. The value of δ_{dis} and the form of the linear relationships between H-bond energy and entropy have been established in prior works [7, 27], whereas u , v and δ_{nat} are fitting parameters determined by fitting to experimental heat capacity curves from DSC. Using simulated annealing, the three parameters are randomly guessed until the predicted heat capacity curve matches the experimental curve well. The least squares error (LSE) between the experimental and predicted heat capacity curves is used to determine the goodness of the fit.

Mechanical response

Once the free energy landscape is complete, the mDCM calculates a wide variety of ensemble averaged mechanical properties. Two useful quantities are the backbone Flexibility Index (FI) and Cooperativity Correlation (CC) [9]. For a given framework, the PG algorithm will assign each backbone rotatable bond as: (i.) *flexible* if it is in an under-constrained region; (ii.) *isostatic* if it is in a just-enough-constrained region or (iii.) *redundant* if it is in an over-constrained region. Each bond is assigned a flexibility index, f_i , defined based on a single constraint network as follows. If the bond in question is part of an isostatically rigid region, $f_i = 0$. If the bond in question is part of a flexible region, the number of rotatable bonds within that flexible region is counted, and denoted as H . The number of independent disordered torsions within that same flexible region is also counted, and denoted as A . To represent the density of independent DOF within the flexible region, the value $f_i = A/H$ is assigned to all bonds within this cluster. Conversely, if the bond in question is found to be in an over-constrained region, the total number of rotatable bonds are counted and denoted as D , whereas the total number of redundant constraints within that region is counted and denoted B . The value $f_i = B/D$ represents the density of redundant constraints within this overconstrained region, and it is assigned to all the bonds within this cluster. Once this counting is complete for every rigid cluster, every rotatable bond in the protein will have a f_i value assigned to it. To distinguish between densities of DOF versus redundant constraints, the f_i values corresponding to flexible regions are positive, whereas the above f_i values in overconstrained regions are multiplied by -1 . Repeating this process many times over many topological frameworks and macrostates, the final $\{f_i\}$ are appropriately averaged based on their Boltzmann probabilities. Typically, which is the case herein, we only average over the macrostates corresponding to the native basin in order to focus on native structure equilibrium fluctuations.

CC is calculated by using a similar procedure; however, we are now tracking mechanical couplings. That is, within a given framework, the PG algorithm will determine if each pair of backbone torsions are: (i.) flexibly correlated, (ii.) mechanically independent, or (iii.) within the same rigid cluster. This process is repeated across the ensemble, and average properties are calculated and plotted on an $N \times N$ matrix, where N is the number of backbone torsions. Each pixel is colored by the ensemble-averaged descriptions defined by the three options. By convention, red = flexibly correlated, white = mechanically independent, and blue = co-rigid.

Dataset selection

Our dataset includes one VL, three scFv and five Fab antibody fragments ranging from one domain to four domains. Table 1 summarizes the antibody fragments and their properties. All the antibody fragments have their heat capacity curves determined experimentally except for the antiluorescein scFv, whose C_p curve was borrowed from another scFv [28]. The fragments in this dataset were selected based on three primary criteria. First, a good resolution X-ray crystal structure must be available, and second, the experimental heat capacity curves were available with the transition demonstrating two-state characteristics. Third, the scFv and Fab fragments belong to the same antibody family and share almost the same VH and VL domains. The number of H-bonds in these various fragments is an

important property that directly impacts protein stability. The H-bond number is calculated based on mDCM hydrogen bond criteria defined in mDCM, which includes very weak feeble H-bonds to avoid arbitrary cutoffs. Feeble H-bonds do not contribute to the folding/unfolding transition because they have virtually zero probability to be present. Nevertheless, the relative counts that include these feeble H-bonds provide a good sense of the connectedness of the HBN.

Structure preparation

The crystal structures of the two anti-LT β R scFv mutants do not exist, therefore, the structure of these anti-LT β R scFv structures were generated from the anti-LT β R Fab domain. Based on the sequence of the anti-LT β R scFv [29], the coordinates of the VH and VL domains of the scFv was obtained from the identical regions of the Fab and the two domains were joined via a (Gly₄Ser)_n linker (in this case n=3), which was modeled using the SWISS-MODEL server [30, 31]. The side chain prediction program SCWRL4 [32] was used to model the side chain of the two scFv mutants. For antifluorescein scFv, the missing loop in the crystal structure was also modeled using the SWISS-MODEL server. Except the specific building procedure used on the scFv structures, the remaining structures come from X-ray crystallography with the ligands, waters and other heteroatoms deleted. In all cases, hydrogen atoms were added using H++ server [25] to ensure proper ionization at the pH of the DSC experiments. H++ uses Poisson-Boltzmann continuum electrostatic theory to calculate the appropriate ionization state of the protein and performs an optimization of hydrogen positions. Subsequently, these structures were minimized using the Molecular Operating Environment software [33] with the Amber force field [34]. All H-bond, salt-bridge, torsion angle, and covalent bond constraints are included within the DCM input, which is used to construct the topological frameworks.

Approach to obtain mDCM parameter without experimental C_p values

Because there are many more proteins with known structures than known heat capacity curves, it generally happens that heat capacity curves will not be available to fit to. Therefore, as a practical matter, we applied an iterative fitting protocol to predict the C_p curve. This protocol starts from a collection of trial heat capacity curves, which were generated based on an existed experimental heat capacity curve from a similar sized protein. Firstly, we borrowed an experimental C_p curve from a variant of recombinant human scFv, FW1.4, which has one peak corresponding to the T_m at 342K [28]. Secondly, this C_p curve was rescaled in terms of its height by generating several more test C_p curves by multiplying a consecutive multiplier for each incremental change in temperature T of a couple of degrees in Kelvin, while keeping the baselines the same. This rescaling changes the maximum of the C_p curve covering a reasonable large range of possibilities, without affecting the baselines. Thirdly, two rounds of iterative fitting were performed to reproduce each C_p curve. The first round of iterative fitting started from obtaining the appropriate parameter values to our trial C_p curves. The initial ranges for u , v , and δ_{nat} are respectively $[-3.2, -1.0]$, $[-1.5, -0.1]$ and $[0.7, 2.5]$. After the first round fitting, we collected the best-fit curves and corresponding $\{u, v, \delta_{nat}\}$ values, which were next used as input for our next round of fitting after shifting T_m of the C_p curves to the melting temperature of the target protein. The final best-fit is selected as the one with the lowest LSE among all the good fits,

which have the LSE lower than the bottom 20%. The inputs for the range of the three parameters $\{u, v, \delta_{nat}\}$ are next narrowed down again according to the parameter range of all the best-fit curves from the first round best fit. During this process, not only the parameter ranges but also the maximum height of heat capacity is narrowed down. It is worth noting that as the shape of the heat capacity curves change, they produce different baselines, which also have to be teased out. To provide a systematic procedure, baselines were set to start from 0, which means the lowest values in the curves are reset to 0 and other values are shifted correspondingly. Before the second round fitting, the best fit curves from the first round fitting (which has slightly shifted predicted T_m) are shifted back to the T_m of the target protein. Meanwhile the baselines of the best-fit curves are also shifted to start from 0. The shift of T_m and baseline and also the narrowed parameter range help self-correcting the fitting. This protocol was benchmarked on the LT β R scFv P104D mutant.

RESULTS

Model parameters for antibody fragments

X-ray crystal structures and heat capacity data are available for several engineered antibody fragments. As an exemplar case, Fig. 1 shows experimental heat capacity curves from DSC with corresponding best fits for antibody fragments VL, scFv and Fab. Taken together, the three phenomenological parameters, $\{u, v, \delta_n\}$ effectively account for hydrophobic interactions, structural diversity and differentiation of solvent conditions. For our antibody fragment dataset, Table 1 summarizes molecular characteristics of the fragments and best-fit parameters. All best-fit parameters are physically reasonable and consistent across the different antibody fragments. The success of the model across the three fragment length scales further demonstrates the generality of the mDCM. During the process of finding the best mDCM parameters, multiple good fits are obtained corresponding to parameter variations that yield similar results. This plurality in parameter space is due to compensation among the three adjustable parameters as previously documented and quantified (12). This means that finding good fits to the heat capacity curve is not hypersensitive to the exact parameter values. In particular, more compensation is observed between the two energy parameters $\{u, v\}$ than what occurs between the entropy parameter, δ_n , and either energy parameter.

The energy contributions that come from the $\{u, v\}$ parameters compensate one another through a linear relationship that has been noticed in all work to date involving the mDCM. A line in the $u - v$ plane forms by letting $\{u, v\}$ vary and considering all good fits while holding the best-fit value of δ_n fixed for each complex size. For all the fragments studied, Fig. 2 plots the compensating relationship between $\{u, v\}$ as determined by good-fit parameters, including the best-fit parameters. Consistent with the fact that smaller systems are typically easier to find good fits for, it is interesting to note that the scatter tightness is consistent fragment size, which is likely associated with the fact that the larger fragments undergo multiphasic transitions that are more difficult to describe by a single $\{u, v\}$. The slope of the linear regression line through the data is 0.58, which is consistent with previous findings. However, what makes this result particularly interesting is that this line describes energy compensation between the $\{u, v\}$ parameters that extend over different length scales

involving structures with different δ_n parameters. The best-fit values for δ_n are close for fragments of similar size, but they increase with size of the system, with approximate doubling the size from VL to scFv and then doubling again to the Fab complexes. Although there is an increase in δ_n as the complex size increases, it is remarkable that the increase in δ_n as a function of fragment size is actually quite small. Indeed, the surprising result is the small magnitude of the drift. This relative insensitivity of mDCM parameters across fragments was seen before and utilized in the study of proteolysis fragments within thioredoxin (11).

In previous works the role of the δ_n parameter has been attributed as an effective way to account for the architecture (shape and size) of a protein. As a phenomenological parameter it is difficult to tease out why δ_n increases as fragment size increases due to the mean field nature of the mDCM, surface to volume effects, self-avoidance and not explicitly modeling denatured structures. Specifically, no simple reasoning suggests there should be an increase (or decrease) in δ_n as the size of a protein increases as a general result, and other examples have shown there are no universal size dependent trends [27]. However, the net effect of a larger δ_n indicates there is more intrinsic entropy (disorder) in the native state, such that the entropy gain that results from a local environment changing from good to poor packing is reduced in the larger complexes. In the case of antibody fragments, larger complexes will be intrinsically more susceptible to local structural arrangements consistent with the constraints imposed by the H-bond network (HBN).

The HBN was found to be the major determinant affecting thermodynamic stability among proteins of approximately the same size, where the δ_n parameter was held fixed over all proteins (or fragments) in the dataset [13]. Here, the δ_n parameter range does not have sufficient plasticity to cover the length scale variation of all the fragments and their complexes simultaneously. Specifically, we were unable to simultaneously fit all the heat capacity curves in the dataset using a single δ_n parameter. Furthermore, there are systematic differences in the energy parameters even when δ_n is allowed to vary. With all energies in units of kcal/mol, the statistical average u value and corresponding standard deviation for VL, scFv and Fab from all good fits are -2.31 ± 0.15 , -2.34 ± 0.23 and -2.90 ± 0.26 respectively. The averages and standard deviations for v are respectively -0.32 ± 0.10 , -0.39 ± 0.11 and -0.77 ± 0.16 for VL, scFv and Fab. The average values clearly show the $\{u, v\}$ parameters undergo a large concerted change upon complex formation, and these changes reflect the change in solvent accessible surface to volume ratio as well as the additional H-bonds that occur at domain interfaces. While thermodynamic stability is governed primarily by the HBN, the systematic differences that must occur in all three parameters are ascribed to packing effects that vary with respect to size among the fragments and their complexes.

1D free energy landscapes and protein stability/flexibility

Once parameterization is achieved, thermodynamic properties and QSFR are calculated. An important QSFR descriptor is the one-dimensional FEL that tracks the free energy, $G(T, \theta)$, as a function of temperature and global flexibility order parameter. The global flexibility order parameter is defined as the average number of independent disordered torsion constraints divided by the number of protein residues. Fig. 3A,B,C shows the FEL for three

exemplar cases using VL, scFv and Fab respectively as a function of flexibility order parameter at their corresponding melting temperatures obtained, with best-fit parameters given in Table 1. Note that there is a sub-ensemble of constraint topologies that share the same number of independent DOF at each value of the order parameter due to fluctuations in how constraints are distributed within the structure. That is, H-bonds are forming and breaking, and torsion constraints fluctuate between native and disordered microstates, which represent locally good and poor packing respectively.

Fluctuations in the constraint network lead to different rigid cluster decompositions in a given sub-ensemble. As such, rigid clusters also form and break. As described previously [7, 27], an insightful quantity to characterize rigid cluster fluctuations is the rigid cluster susceptibility (RCS), which is shown in Fig. 3. The peak of a RCS curve identifies the rigidity percolation threshold, below which the protein is globally rigid with sparse regions of flexibility that usually support correlated motions. The flexibility order parameters at the percolation threshold, θ_p , are given as 1.30, 0.65, 0.60 for the VL, scFv and Fab systems. Table 2 list key values of the flexibility order parameter corresponding to the native, transition and denatured states, and the peak in rigid cluster susceptibility. Above the rigidity threshold, the protein is globally flexible, but contains a large number of rigid clusters that “flicker” in and out of the structure. As the RCS increases, fluctuations that involve mechanical transitions (i.e. rigid $\Leftrightarrow$ flexible) span over larger regions. Near the peak in the RCS curve, both rigidity and flexibility propagation will typically span the entire protein intermittently as the constraint network fluctuates intensely. The most interesting features from the RCS curves are: (1) There are much greater fluctuations in the rigid clusters as the size of the complex increases, where the peak heights increase much faster than a simple proportionality to the number of residues, which is attributed to surface volume effects, addition of a linker and additional domain interface interactions; and (2) The peak of the RCS curve shifts to lower global flexibility as the size of the complex increases, indicating that more constraints are necessary to make the protein globally rigid.

As the flexibility order parameter increases, the protein transits from a native to unfolded state, which often (but not necessarily) corresponds to a rigid and a flexible mechanical state. Each FEL has two minima labeled as θ_N and θ_U for the native and unfolded states respectively, and the presence of a straddling free energy barrier will cause two-state behavior that mimics a first-order phase transition. However, as we have demonstrated before [27], the mDCM can also describe multistate transitions when appropriate, which is important for this work since large multidomain systems like antibody fragments are typically not two-state. The difference in global flexibility between unfolded and native states at the transition temperature is given by $\theta = \theta_U - \theta_N$. The flexibility order parameter difference, θ , upon unfolding are 0.97, 0.54 and 0.25 for VL, scFv and Fab respectively, which indicates there is a release of roughly 4, 2 and 1 DOF for every four residues upon unfolding respectively. Although the change in DOF per residue dramatically decreases, the average number of DOF per residue in the native ensemble is nearly constant for the VL fragment and the scFv and Fab complexes. Thus, the unfolded basin becomes globally less flexible for larger complexes compared to VL fragment. Furthermore, taking the order parameter as a reaction coordinate, the transition state located at θ_T is identified by the peak of the free energy barrier. The greater free energy barrier that must be crossed to go from θ_N

to θ_U via a path through the transition state θ_T provides greater kinetic stability against unfolding. The free energy barriers defined as $\Delta G_{N \rightarrow U}^* = G(T_m, \theta_T) - G(T_m, \theta_N)$ for the exemplar three cases are found to be 2.39, 1.40 and 0.46 kcal/mol for the VL, scFv and Fab fragments at their respective melting temperatures. Surprisingly, the antibody fragments forming larger complexes are predicted to have lower free energy barriers, and the barrier heights become less symmetrical in the larger complexes, where the complementary barriers $\Delta G_{U \rightarrow N}^* = G(T_m, \theta_T) - G(T_m, \theta_U)$ are found to be 2.03, 0.86 and 0.01 kcal/mol respectively.

A key QSFR metric is the flexibility index that quantifies backbone flexibility, which is quantified by the FI, shown in Fig. 4 for the exemplar VL, scFv and Fab cases. Positive values of FI quantify the amount of excess DOF in flexible regions, and negative values quantify the amount of excess constraints in rigid regions. The flexibility indexes for the native states appear to be similar in appearance to the typical case found in other proteins. Backbone flexibility tends to be well conserved within protein families, and here, it is seen that backbone flexibility in the VL domain is well conserved as an isolated fragment or as part of a larger complex. Nevertheless, from the RCS curves it is known that dramatic changes are taking place in how the rigid clusters break and form. This means pairs of residues will intermittently belong to the same rigid cluster or to the same flexible region. Residue couplings are quantified with another important QSFR metric, referred to as a CC plot. In general, it is found that the CC plots are less conserved across protein families (relative to backbone flexibility), and are sensitive to the specific structural characteristics of the protein system being investigated.

The corresponding CC plots are also shown in Fig. 4 for the VL, scFv and Fab cases. The CC plots describe correlations between a pair of residues by tracking the frequency (in the sense of a statistical ensemble) that they are simultaneously part of the same rigid cluster, or part of the same flexible region. Note that a particular rigid cluster can itself be very mobile, meaning when two residues are members of a single rigid cluster their motion is highly correlated. When a pair of residues are co-flexible, the motion of either residue affects the other residue where flexibility propagates through. Two separate regions that are either rigid and/or flexible will not be correlated if there is no discernable mechanical coupling between the two residues. As such, white regions do not highlight the lack of being flexible or rigid, but rather, highlight independent atomic motions. The exemplar cases (cf. Fig. 4) show that a high degree of flexibility correlation can be present in a protein with low number of degrees of freedom per residue. This result indicates correlated motions propagate through the structure with a relatively small amount of independent DOF that drive these motions. The mechanisms responsible for these effects can now be discussed.

DISCUSSION

The consequence of free energy barriers lowering in larger complexes (on both sides) is that the folding to unfolding process, and the reverse process, is faster for the complexes on larger length scales than would otherwise be expected using the rule of thumb that larger proteins will have larger barriers. In particular, there is a bias toward sliding back to the native state of the protein once the unfolded state is achieved. As such, these large complexes find themselves most frequently populating the native ensemble even at the

melting temperature, in addition to having ample opportunity for crossing over the barrier into the unfolded state, as it is not particularly restrictive. It is insightful to compare where the peak in the RCS curve is located relative to the native, transition and unfolded states are located along the global flexibility order parameter. The VL domain (cf. Fig. 3A) exhibits typical behavior seen in single domain proteins, where the native state is globally rigid, the transition state exhibits the most fluctuation in rigid clusters, and the unfolded state represents a globally flexible structure. As the peak of the RCS curves moves to lower flexibility in the scFv complex (cf. Fig. 3B), the native state has the most fluctuation in rigid clusters. Although not an unprecedented situation, the native state in the Fab complex is located on the mechanically flexible side of the RCS curve, indicating that there is a high degree of susceptibility in structural rearrangements. The results are shown at the melting temperature, but the same characteristics remain at lower temperatures. Specifically, the peak in the RCS curve and the native state do not appreciably shift as temperature is lowered. Therefore, the FEL characteristic calculated is consistent with the known function of antibodies in the sense that large-scale conformational changes are needed to become commensurate with an antigen, while the native state is largely preserved.

Additional constraints from the domain-domain interface act upon the VL domain as it forms a complex, which causes a slight decrease in backbone flexibility and global flexibility order parameter in the native state ensemble. Surprisingly, these interfacial H-bonds cause correlations in flexibility to increase (not decrease). This is possible because correlated motions are sensitive both constraint location and number, the latter of which always reduces flexibility globally. Due to the remarkably high rigid cluster fluctuations that take place, correlated flexible motions survive the ensemble average in the VL domain because the intermittent rigid clusters that form do not extend far. In contrast, the VH domain typically does not acquire this enhanced increase in flexibility correlation, or at least not any notable increase in all Fab cases within the dataset, except for anti-p24 HIV-1 Fab that is discussed separately. The typical behavior in the CC plots shown in Fig. 4 is broadly consistent with the evolutionary hypothesis of the antibody maturation that the VH domain mutates first to rigid the combining site and finally VL domain mutates to increase the binding affinity between the antibody and antigen [14]. Compared to the VH chain, there is more opportunity for mutations in the light chain to rigidify the combining site for improving the antibody's specificity against antigen.

QSFR outlier identifies polyspecificity

The Fab anti-p24 HIV-1 was identified as the sole outlier in the dataset based on differences in QSFR characteristics. The free energy landscape shown in Fig. 3D exhibits a lower free energy in the unfolded basin at the melting temperature defined by the peak in the heat capacity. Normally, the folded state has a lower free energy (not equal) because the width of the unfolded basin is wider than the folded basin at the melting temperature. A reversal of roles occurs for this outlier at T_m , although the general shape and trends of the free energy landscape as a function of temperature is qualitatively similar, provided the melting temperature is not considered a special reference point. It is also noted from Table 1 that this outlier has the lowest δ_n across all the Fab complexes. Along the flexibility order parameter, Fig. 3D also shows that the RCS curve peaks below the native state of the protein. These

observations indicate more intrinsic flexibility resides in the anti-p24 HIV-1 Fab compared to all the others in the dataset. In comparison to the typical cases, the CC plot in Fig. 5 for the anti-p24 HIV-1 Fab shows that the VL domain does not possess a high level of co-flexibility, while the VH domain exhibits much more co-flexibility. In short, the co-rigidity and co-flexibility characteristics of the VL and VH domains have been reversed. There is insufficient statistics in our dataset to make hard conclusions based on these results, but it is intriguing to note that the anti-p24 HIV-1 Fab is known to exhibit polyspecificity [16]. Further work to determine if the QSFR properties can be used to identify functional characteristics will require expanding the dataset.

Parameter sensitivity against mutations

To meet the objective of substantially expanding the dataset requires benchmarking whether the calculated QSFR properties are robust against parameter sensitivity. Within the dataset, two scFv cases with similar QSFR properties differ by two mutations (V56G and P104D mutants). Based on prior works [9], there is a clear trend that backbone flexibility is selectively sensitive to mutations or larger sequence variations across families, and that application of the same set of $\{u, v, \delta_n\}$ model parameters is possible, and this robustness has been utilized on a dataset consisting of 14 point mutants in C-type lysozyme, relative to the human wild-type structure [12]. The rationale for why this is generally the case is because the flexibility profiles are a manifestation of constraint topology. To test whether parameter insensitivity holds true for the fragments studied here, we recalculated the QSFR properties for the two scFv mutants by crossing the best fit parameter sets and also by using the average parameter set from multiple good fits for these two mutants. Crossing the best-fit parameters, means applying the best parameter set for one protein to the other protein and vice versa. The average parameter set means using the average u , v and δ_n values respectively. Fig. 6 plots the backbone flexibility profile for scFv complex, and the V56G and P104D and mutants calculated from the best fit parameter sets, the average parameter set from all good fits, and the swapped best fit parameters from the V56G mutant. As expected, backbone flexibility is insensitive to the parameter set used. The insensitivity in parameters carries through to the CC plots as Fig. 7 shows for the scFv P104D and V56G mutants as calculated from the best fit parameter sets, average parameter set from all good fits, and the swapped best fit parameter sets from the V56G mutant.

It is worth noting that the backbone flexibility profiles for the two scFv-mutant fragments are similar, with a few regions with significant change in flexibility. Fig. 4C shows the difference of the backbone flexibility of the two scFv fragments calculated using the average parameter set. The mutated residues P104D and V56G are located in region one (residue 53–54) and two (residue 99–105) respectively. Interestingly, both mutations increase backbone rigidity in their immediate surrounding residues, which provides a probable explanation why these two mutations both stabilize the wild-type scFv [35]. Region three (residue 171–172) is not close to the mutation site. The large and distant differences in flexibility suggest a possible route for allosteric response. These results establish assurance that any good fit to the heat capacity that is obtained by simulated annealing will provide robust QSFR predictions. Here, we have demonstrated that the mDCM parameters are transferable in the study of antibody fragments and their complexes over a wide range of different length

scales, which is of practical significance for predicting relative stability of antibody mutants going forward.

mDCM parameterization without experimental C_p curves

To broaden the utility of the mDCM, we applied an iterative fitting method to generate a set of reasonable $\{u, v, \delta_n\}$ parameters starting from initial C_p curves obtained by guessing. We tested this method on the scFv P104D mutant. Based on another scFv experimental C_p curve, a series of C_p curves (cf. Fig. 8A) are produced as described in the Methods section. The rationale for using this particular C_p curve as a reference is based on the fact that all scFv's tend to have conserved global structure. In addition, the VH and VL domains of these two scFvs both melt at their own temperature and display a single melting transition. The test C_p curves have their peaks shifted to match the known melting temperature, T_m , of the target protein, which in this case is 345K. The peak height of the initial C_p curve is rescaled to range from about 7 to kcal/mol, which extends beyond a reasonable range that could be expected for peak heights. The baselines of these curves are maintained in the initial trials. A total of 18 test C_p curves were generated starting from one initial curve. Two rounds of fitting are performed. In the first round, the structure is fitted with 5 different initial conditions using simulated annealing applied to each generated C_p curve from this set, yielding $18 \times 5 = 90$ fits. The top 25% of the ranked ordered cases based on the LSE of the predicted C_p curves (23 fits) are shown in Fig. 8B. After this first round, most of the fits have slightly shifted T_m . In the second round, the fitting procedure is repeated on the best set of predicted C_p curves after shifting the T_m back to the true T_m . The new fitting is performed with a narrowed range on the $\{u, v, \delta_n\}$ parameters based on the best-fits from round 1. Once the second round is completed, the top 25% of the ranked ordered cases according to the least squares error were retained (6 fits in this case) as shown in Fig. 8C leaving a degree of uncertainty in the peak heights within the range 23–54 kcal/mol.

This iterative method is not expected to find the perfect C_p curve, but it focuses peak height variation across an ensemble of trial C_p curves to be within a physically plausible range. Interestingly, the true experimental C_p curve is found to lie approximately in the middle of the collection of final trials, although the only expectation is it should lie within this range. Given the native state constraint topology, the underlying idea behind this iterative method relies on self-corrected fitting so that the mDCM parameters will retain relative proportions that are physically realizable. The parameters obtained from this method are also included in Table 1. It is found that δ_n falls within a narrow range, and Fig. 9A plots the relationship of $\{u, v\}$ pairs from the iterative fitting procedure. A strong correlation ($R = 0.998$) is observed between u and v . As discussed above there are multiple good fits for each antibody fragment, where the diversity is mainly captured through the u, v pair. A more negative u will drive native H-bonds to break more readily (a loss in constraints), which is compensated by a more negative v which drives more native torsion constraints to form (a gain in constraints). These compensating effects are explored when heat capacity is fitted. The backbone flexibility is shown in Fig. 9B for each set of good fit parameters to the known C_p curve. Similarly, the backbone flexibility for each parameter set obtained from the iteration method when the C_p curve was assumed to be unknown is shown in Fig. 9C. Although not shown, the difference in the averages of these two sets of data is virtually zero, meaning that

the iterative fitting procedure was able to produce a robust backbone flexibility profile. This high degree of agreement also carries over to the CC plots, for which a quantitative comparison is shown in Fig. 9D. These results demonstrate QSFR is largely insensitive to model parameterization. Thus, if the experimental heat capacity is missing, we can use this method to derive a reasonable $\{u, v, \delta_n\}$ parameter set.

CONCLUSIONS

A QSFR analysis on an exploratory dataset of antibody fragments using the mDCM has been obtained for the first time. Model parameters are determined to be remarkably consistent across length scales involving fragments and their complexes. The QSFR results are consistent with general trends observed in other protein systems previously studied, including single domain proteins. That is, apart from a small number of selected regions, backbone flexibility is well conserved across the Fab complexes, and molecular cooperativity plots are much more sensitive to sequence variation. The interesting distinguishing signature observed is the Fab antibody complexes possess large rigid cluster susceptibilities in the native state, indicating that rigidity percolation is nearly concurrent with the native state. This means large-scale rigidity/flexibility propagations are present. Moreover, free energy barriers between the folded and unfolded states decrease as the complex size increase. Despite being highly deformable, the actual number of free DOF within the unfolded state is much lower than is typically observed in other proteins. This result indicates a much higher degree of and flexibility correlation is present in the native and unfolded states than has been observed in previous works. In particular, a VL fragment has more co-rigid residue pairs when isolated compared to the scFv and Fab forms, where correlated flexibility appears upon complex formation. The general trend found in the context dependence on residue-residue couplings in the VL domain across length scales of a complex is consistent with the evolutionary hypothesis of antibody maturation. QSFR metrics for the anti-p24 HIV-1 Fab were found to be atypical, which includes a comparatively greater co-flexibility in the VH domain and less co-flexibility in the VL domain. Experimentally, the anti-p24 HIV-1 Fab is known to be polyspecific. While these results are encouraging, additional examples are needed to conclude this represents a distinguishing characteristic. As a practical matter, parameterizing the model becomes challenging in the absence of thermodynamic data, such as heat capacity curves. In this report, we have demonstrated that QSFR properties can be robustly calculated even in cases when the heat capacity curves are not available using a self-consistent iterative method. Therefore, expanding the studies to a larger dataset of antibody fragments is feasible, and is currently underway.

ACKNOWLEDGEMENTS

Support for Dr. Li is provided by MedImmune, LLC. Additional support for this work was provided by NIH grants R01 GM073082 and S10 SRR026514. Key to the distance constraint model is the use of graph-rigidity algorithms, claimed in U.S. Patent 6,014,449, which has been assigned to the Board of Trustees Michigan State University. Used with permission.

REFERENCES

1. Igawa T, Tsunoda H, Kuramochi T, Sampei Z, Ishii S, Hattori K. Engineering the variable region of therapeutic IgG antibodies. *mAbs*. 2011; 3:243–252. [PubMed: 21406966]
2. Holliger P, Hudson PJ. Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* 2005; 23:1126–1136. [PubMed: 16151406]
3. Weisser NE, Hall JC. Applications of single-chain variable fragment antibodies in therapeutics and diagnostics. *Biotechnol. Adv.* 2009; 27:502–520. [PubMed: 19374944]
4. Demarest SJ, Glaser SM. Antibody therapeutics, antibody engineering, and the merits of protein stability. *Curr. Opin. Drug. Discov. Deve. I.* 2008; 11:675–687.
5. Wang W, Singh S, Zeng DL, King K, Nema S. Antibody structure, instability, and formulation. *J. Pharm. Sci.* 2007; 96:1–26. [PubMed: 16998873]
6. Jacobs, DJ. An Interfacial Thermodynamics Model for Protein Stability. In: Misra, AN., editor. *Biophysics, InTech*. 2012. p. 41
7. Jacobs DJ, Dallakyan S. Elucidating protein thermodynamics from the three-dimensional structure of the native state using network rigidity. *Biophys. J.* 2005; 88:903–915. [PubMed: 15542549]
8. Jacobs DJ, Dallakyan S, Wood GG, Heckathorne A. Network rigidity at finite temperature: relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems. *Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys.* 2003; 68:061109. [PubMed: 14754182]
9. Verma D, Jacobs DJ, Livesay DR. Changes in Lysozyme Flexibility upon Mutation Are Frequent, Large and Long-Ranged. *Plos Comput. Biol.* 2012; 8
10. Mottonen JM, Xu ML, Jacobs DJ, Livesay DR. Unifying mechanical and thermodynamic descriptions across the thioredoxin protein family. *Proteins: Struct. Funct. Bioinf.* 2009; 75:610–627.
11. Jacobs DJ, Livesay DR, Hules J, Tasayco ML. Elucidating quantitative stability/flexibility relationships within thioredoxin and its fragments using a distance constraint model. *J. Mol. Biol.* 2006; 358:882–904. [PubMed: 16542678]
12. Livesay DR, Jacobs DJ. Conserved quantitative stability/flexibility relationships (QSFR) in an orthologous RNase H pair. *Proteins*. 2006; 62:130–143. [PubMed: 16287093]
13. Livesay DR, Huynh DH, Dallakyan S, Jacobs DJ. Hydrogen bond networks determine emergent mechanical and thermodynamic properties across a protein family. *Chem. Cent. J.* 2008; 2
14. Zimmermann J, Oakman EL, Thorpe IF, Shi X, Abbyad P, Brooks CL 3rd, Boxer SG, Romesberg FE. Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proc. Natl. Acad. Sci. U S A.* 2006; 103:13722–13727. [PubMed: 16954202]
15. Thorpe IF, Brooks CL 3rd. Molecular evolution of affinity and flexibility in the immune system. *Proc. Natl. Acad. Sci. U S A.* 2007; 104:8821–8826. [PubMed: 17488816]
16. Keitel T, Kramer A, Wessner H, Scholz C, Schneider-Mergener J, Hohne W. Crystallographic analysis of anti-p24 (HIV-1) monoclonal antibody cross-reactivity and polyspecificity. *Cell*. 1997; 91:811–820. [PubMed: 9413990]
17. Jacobs DJ, Wood GG. Understanding the alpha-helix to coil transition in polypeptides using network rigidity: predicting heat and cold denaturation in mixed solvent conditions. *Biopolymers*. 2004; 75:1–31. [PubMed: 15307195]
18. Mottonen JM, Jacobs DJ, Livesay DR. Allosteric response is both conserved and variable across three CheY orthologs. *Biophys. J.* 2010; 99:2245–2254. [PubMed: 20923659]
19. Wood GG, Clinkenbeard DA, Jacobs DJ. Nonadditivity in the alpha-helix to coil transition. *Biopolymers*. 2011; 95:240–253. [PubMed: 21280020]
20. Jacobs DJ, Livesay DR, Mottonen JM, Vorov OK, Istomin AY, Verma D. Ensemble properties of network rigidity reveal allosteric mechanisms. *Methods Mol. Biol.* 2012; 796:279–304. [PubMed: 22052496]
21. Vorov OK, Livesay DR, Jacobs DJ. Helix/coil nucleation: a local response to global demands. *Biophys. J.* 2009; 97:3000–3009. [PubMed: 19948130]

22. Vorov OK, Livesay DR, Jacobs DJ. Nonadditivity in conformational entropy upon molecular rigidification reveals a universal mechanism affecting folding cooperativity. *Biophys. J.* 2011; 100:1129–1138. [PubMed: 21320459]
23. Jacobs DJ, Thorpe MF. Generic rigidity percolation: The pebble game. *Phys. Rev. Lett.* 1995; 75:4051–4054. [PubMed: 10059802]
24. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. *Proteins.* 2001; 44:150–165. [PubMed: 11391777]
25. Gordon JC, Myers JB, Foltz T, Shoja V, Heath LS, Onufriev A. H⁺⁺: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic acids research.* 2005; 33:W368–W371. [PubMed: 15980491]
26. Dahiyat BI, Gordon DB, Mayo SL. Automated design of the surface positions of protein helices. *Protein Sci.* 1997; 6:1333–1337. [PubMed: 9194194]
27. Livesay DR, Dallakyan S, Wood GG, Jacobs DJ. A flexible approach for understanding protein stability. *FEBS Lett.* 2004; 576:468–476. [PubMed: 15498582]
28. Borrás L, Gunde T, Tietz J, Bauer U, Hulmann-Cottier V, Grimshaw JP, Urech DM. Generic approach for the generation of stable humanized single-chain Fv fragments from rabbit monoclonal antibodies. *J. Biol. Chem.* 2010; 285:9054–9066. [PubMed: 20056614]
29. Michaelson JS, Demarest SJ, Miller B, Amatucci A, Snyder WB, Wu X, Huang F, Phan S, Gao S, Doern A, Farrington GK, Lugovskoy A, Joseph I, Bailly V, Wang X, Garber E, Browning J, Glaser SM. Anti-tumor activity of stability-engineered IgG-like bispecific antibodies targeting TRAIL-R2 and LTβR. *mAbs.* 2009; 1:128–141. [PubMed: 20061822]
30. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* 2006; 22:195–201. [PubMed: 16301204]
31. Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic acids research.* 2009; 37:D387–D392. [PubMed: 18931379]
32. Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins.* 2009; 77:778–795. [PubMed: 19603484]
33. Molecular Operating Environment (MOE), 2011.10. Vol. H3A 2R7. Montreal, QC, Canada: Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910; 2011.
34. Ponder JW, Case DA. Force fields for protein simulations. *Advances in protein chemistry.* 2003; 66:27–85. [PubMed: 14631816]
35. Miller BR, Demarest SJ, Lugovskoy A, Huang F, Wu X, Snyder WB, Croner LJ, Wang N, Amatucci A, Michaelson JS, Glaser SM. Stability engineering of scFvs for the development of bispecific and multivalent antibodies. *Protein Eng Des Sel.* 2010; 23:549–557. [PubMed: 20457695]
36. Tsybovsky Y, Shubenok DV, Kravchuk ZI, Martsev SP. Folding of an antibody variable domain in two functional conformations in vitro: calorimetric and spectroscopic study of the anti-ferritin antibody VL domain. *Protein Eng Des Sel.* 2007; 20:481–490. [PubMed: 17962224]
37. Midelfort KS, Hernandez HH, Lippow SM, Tidor B, Drennan CL, Wittrup KD. Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J. Mol. Biol.* 2004; 343:685–701. [PubMed: 15465055]
38. Welfle K, Misselwitz R, Hausdorf G, Hohne W, Welfle H. Conformation, pH-induced conformational changes, and thermal unfolding of anti-p24 (HIV-1) monoclonal antibody CB4-1 and its Fab and Fc fragments. *Biochim Biophys Acta.* 1999; 1431:120–131. [PubMed: 10209285]
39. Ionescu RM, Vlasak J, Price C, Kirchmeier M. Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies. *J. Pharm. Sci.* 2008; 97:1414–1426. [PubMed: 17721938]

Figure 1.
Best-fit heat capacity curves for a VL, scFv (anti-LB β R P104D mutant) and Fab (anti-LB β R) antibody fragments as an exemplar case respectively. Blue indicates fitting points for VL, Green is for scFv and Red is for Fab.

Figure 2. Comparison of best-fit u , v pairs for the dataset. Blue indicates fitting points for VL, Green is for scFv and Red is for Fab. $R = 0.84$.

Figure 3.

Solid lines in panels A, B, C, D show the free energy landscape as a function of global flexibility corresponding to VL, scFv (anti-LB β R P104D mutant), Fab (anti-LB β R) and Fab (anti-p24 HIV-1). The locations of the native and unfolded basins are (A) $\theta_N = 0.84$, $\theta_U = 1.80$ (B) $\theta_N = 0.75$, $\theta_U = 1.28$ (C) $\theta_N = 0.70$, $\theta_U = 0.94$ (D) $\theta_N = 0.75$, $\theta_U = 1.03$. The corresponding rigid cluster susceptibility curves are shown as dashed lines.

Figure 4.

The top, middle and bottom rows plot backbone flexibility (left panels) and cooperativity correlation plots (right panels) for VL, scFv (anti-LTβR P104D mutant) and Fab (anti-LTβR) respectively. The (black, red) lines in the lowest left panel show the backbone flexibility for the Fab (light, heavy) chains. In the CC plots red regions indicate co-flexible regions, blue indicates co-rigid regions and white designates no correlations.

Figure 5. The CC plot for the anti-p24 HIV-1 Fab complex, showing that the VL domain consist of much greater co-rigidity and the VH domain consist of much greater co-flexibility than is typically found within the dataset as seen in Figure 4.

Figure 6. Backbone flexibility characteristics for anti-LT β R scFv P104D and V56G mutants. (A) Flexibility index (FI) comparison of the scFv P104D mutant calculated from best fit parameter, average parameter from all good fits and swapped best fit parameter from V56G mutant respectively. Red line is from best fit parameter, green is from average parameter and Blue is from swapped parameter. (B) The FI for V56G. (C) Comparison of FI calculated using average parameter for the scFv mutant pair. Orange is for P104D mutant and cyan is for V56G mutant. The scFv structure is shown above the FI. Magenta pentagrams denote the locations of the two mutations. Three regions with significant difference between the FI for the two mutants are highlighted in green and called R1, R2 and R3.

Figure 7.

Cooperativity correlation plots for the two scFv mutants, where red regions indicate co-flexible couplings and blue indicates co-rigid couplings. (A) (B) (C) are CC plots for scFv P104D mutant calculated from best fit parameter, average parameter from all good fits and swapped best fit parameter from V56G mutants respectively. (D) (E) (F) are corresponding plots for V56G mutant.

Figure 8.

Blind predicted heat capacity curves for scFv P104D mutant. (A) Generated trial C_p curves, where red denotes the real experimental C_p curve, green denotes randomly selected experimental C_p curve from another scFv and black denotes a series of generated scaled C_p curves. The iterative approach starts from fitting these trial curves. (B) The top 25% predicted C_p curves from the best-fit parameters after the first round of fitting. (C) The top 25% predicted C_p curves from the best-fit parameters after the second round of fitting.

Figure 9.

(A) Relationship between u , v parameters from the iterative fitting method. Red rectangle denotes the u , v pair from best fitting to its own experimental heat capacity curve. Blue circles denote the u , v pairs from the top 25% best fitting cases after the first round run. Green stars denote the u , v pairs for the top 25% best fitting cases after the second round run. $R = 0.998$. (B) The flexibility index (FI) for scFv P104D is shown for each good fit the known experimental data. (C) The blind predicted FI of scFv P104D is shown for each good iterative fit, assuming the C_p curve is not known. (D) The flexibility correlation (CC plot) using the best-fit parameters to the known C_p curve is compared to the average CC plot over all iterative fits without using the known C_p curve by plotting corresponding pixel values.

Table 1
Antibody dataset and mDCM parameters obtained from best fitting to heat capacity.

Frag	Antibody	PDB entry	Residue number	T _m (K)	HB number ^b	δ _{nat}	u _{sol}	v _{nat}
VL	antiferritin	1F6L	106	335 [36]	138	1.62	-2.45	-0.40
	anti-LBβR (mutant V56G)	3HC0 ^a	238	341 [35]	325	1.82	-2.31	-0.38
	anti-LBβR (mutant P104D) ^c	3HC0 ^a	238	345 [35]	346	1.80	-2.31	-0.36
scFv	antifluorescein	1X9Q	255	322 [37]	390	1.86	-2.90	-0.48
Fab	anti-p24 HIV-1	1CFQ	427	340 [38]	623	1.86	-3.03	-0.67
	anti-LBβR	3HC0	429	351 [29]	598	1.98	-2.36	-0.58
	anti-HER2	1N8Z	434	355 [39]	625	1.90	-3.03	-1.05
	anti-VEGF	1BJI	437	347 [39]	592	1.98	-2.60	-0.69
	antifluorescein	1FLR	437	326 [37]	658	1.96	-3.09	-0.72

^aThe two scFv fragments were modeling based on this crystal structure.
^bHB number means hydrogen bonding number, which is obtained from the mDCM.
^cThe iterative fitting mDCM parameters (six sets) were listed below the best fitting parameter.

Table 2

Key values of the global flexibility order parameter (θ) and barrier heights.^a

Fragment	θ_N	θ_I	θ_U	θ_I	θ_2	θ_P	Barrier ^a (N and T)	Barrier ^b (N and U)
VL	0.84	1.25	1.80	0.71	1.04	1.30	2.39	2.03
scFv (P104D)	0.75	1.00	1.28	0.67	0.87	0.75	1.40	0.86
scFv (V56G)	0.77	1.07	1.41	0.68	0.92	0.75	1.57	0.87
scFv (1X9Q)	0.80	1.00	1.17	0.73	0.90	0.80	0.56	0.24
Fab (1CFQ)	0.75	0.87	1.03	0.71	0.81	0.60	0.30	0.46
Fab (3HC0)	0.70	0.87	0.94	0.64	0.78	0.65	0.46	0.01
Fab (1N8Z)	0.50	0.68	0.82	0.44	0.59	0.40	1.01	0.21
Fab (1BJ1)	0.48	0.62	0.71	0.43	0.55	0.40	0.77	0.11
Fab (1FLR)	0.61	0.77	0.87	0.56	0.69	0.60	0.63	0.08

^a Ensemble averaging within the native state basin is performed over the range between θ_1 (value lower than θ_N and having equal free energy with θ_2) and θ_2 (middle value between θ_N and θ_T). N stands for native state, T for transition state, U for unfolded state, P for percolation threshold.

^b Barrier between the native basin (N) and transition state barrier (T).

^c Barrier between native (N) and unfolded (U) basins.