Antigen Binding Reshapes Antibody Energy Landscape and Conformation Dynamics

Kazi Lutful Kabir
Department of Computer Science
George Mason University
Fairfax, VA, USA
kkabir@gmu.edu

Buyong Ma
Engineering Research Center of Cell & Therapeutic Antibody
School of Pharmacy, Shanghai Jiaotong University
Shanghai, China
mabuyong@sjtu.edu.cn

Ruth Nussinov
Computational Structural Biology Section
Laboratory of Cancer Immunometabolism
Frederick National Laboratory for Cancer Research
National Cancer Institute-Frederick
Frederick, MD, USA
nussinor@mail.nih.gov

Amarda Shehu
Department of Computer Science
George Mason University
Fairfax, VA, USA
amarda@gmu.edu

***Abstract*—This study elucidates the conformation dynamics of the free and antigen-bound antibody. Previous work has verified that antigen binding allosterically promotes Fc receptor recognition. Analysis of extensive molecular dynamics simulations finds that the energy landscape may play a decisive role in coordinating conformation changes but does not provide connections between the various conformational states. Here we provide such a connection. To obtain a detailed understanding of the impact of antigen binding on antibody conformation dynamics, this study utilizes Markov State Models to summarize the conformation dynamics probed in *silico*. We additionally equip these models with the ability to directly exploit the energy landscape view of dynamics via a computational method that detects energy basins and so allows utilizing detected basins as macrostates for the Markov State Model. Our study reveals many interesting findings and suggests that the antigen-bound form with high energy may provide many dynamic processes to further enhance co-factor binding of the antibody in the next step.**

***Keywords*—Antibody, Basin, Conformation, Energy landscape, Molecular dynamics, Markov state model.**

I. Introduction

The specificity and affinity of antibody-antigen recognition is mainly decided by the variable domains and, in particular, the complementarity-determining regions (CDRs). The recognition process involves conformation transitions mediated by the antibody's inherent flexibility [1], [2]. Recent studies suggest allosteric effects during antibody-antigen recognition [3], with both the variable and constant domains playing a role [4], [5]. For instance, antibodies with identical variable domains but different constant domains, may have significantly different affinities to Aβ [6]. Engineering CH and CL in Trastuzumab and Pertuzumab recombinant models also affect antigen-binding [7]. A previous antibody structural analysis finds that distant CH1−1 loops undergoing significant fluctuations upon antigen binding [8].

Many works have shown a two-way communication between the Fc and Fv domains; modifications of Fc can influence the Fv antigen recognition. Upon antigen binding, the C domains can affect the V region paratope conformation, as indicated by circular dichroism [9], NMR [10], and crystallography [11]. Simulations and experiments show that modification of the constant domain influences binding affinity [12]–[14] and specificity [15], [16] of the antibody-antigen interaction. For example, IgA Fc mutations reduce Her2 binding [17]. Our previous study has verified that antigen binding allosterically promotes Fc receptor recognition [1]. Analysis of molecular dynamics (MD) simulations suggests that energy landscape may play a decisive role in coordinating conformation changes. However, current works do not provide connections between the various conformational states.

In this study we aim to provide such connections. This study utilizes Markov State Models (MSMs) to summarize the conformation dynamics probed via MD simulation [18], [19]. Since Zwanzig's first report on short memory approximations in 1983 [20], MSMs have evolved from art to science [21], and now their theoretical underpinnings, as well as rigorous statistical analysis, make them reliable tools to describe conformation dynamics. The focus of this study is Immunoglobulin G (IgG), which binds to cognate antigens. IgGs are central to providing immunity to pathogens, and tumors and are a common template for antibody drugs [22]. MD simulation is utilized to navigate the energy landscape of the free antibody and the energy landscape of the antigen-antibody complex. As related in Section III, different random conformations are generated in each setting (free versus bound) to initialize various MD trajectories that seek collectively to provide a broad view of the energy landscape.

Separately for each setting, the free antibody and the antigen-antibody complex, the conformations accessed via MD simulation are collected and organized into macrostates. A computational method that takes into account both conformation geometry and energetics is able to detect energy basins and, making the connection between basins and macrostates, the method organizes collected conformations into

978-1-6654-0126-5/21/$31.00

macrostates. That means, this method handles the state space discretization. An MSM construction software is then utilized to reveal the inter-conversions between the so-identified macro-states and compute the precise state-to-state transition probabilities that provide a quantitative view of the conformation dynamics in each setting, allowing a more detailed understanding of the impact of the bound antigen on the conformation dynamics of the antibody.

II. Related Work

Various MSM packages are available, with the two most popular being MSMBuilder [23] and EMMA [24], [25]; PyEMMA provides a Python implementation of EMMA. Due to ease of use, rich analysis tools, and ability to control various steps of the computational process, we employ PyEMMA [25], as we have done in earlier work on peptide dynamics [26].

Both MSMBuilder and EMMA find wide applicability in computational molecular biology to uncover the multi-state folding kinetics of small proteins and peptides [27], estimate the kinetics of protein-ligand binding to determine drug efficacy [28] and the kinetics of protein-protein associations to reveal insights into macro-molecular recognition [29], as well as reveal antibody conformation dynamics to identify effective cancer therapies [30]. In particular, EMMA provides many statistical techniques to evaluate constructed MSMs [31].

At the core of any MSM construction package lies the ability to identify conformational states (macrostates). Theory dictates that conformations capable of rapid inter-conversion should be aggregated into the same macrostate, while conformations with slow inter-conversion rate should be separated into different macrostates [32]. In practice, MSM construction packages cluster conformations to organize them into macrostates. The inherent assumption is that a molecule converts/switches between geometrically-similar conformations more rapidly than between less similar conformations.

The above assumption does not consider energetics. Two conformations can be proximal in conformation space but separated by an energy barrier that slows the inter-conversion rate. This realization prompted us to evaluate an alternative approach to the identification of macrostates in recent work [26]; we demonstrate that taking into account energy in the identification of macrostates results in higher-quality MSMs. Our work on summarizing energy landscapes via energy basins [33] provides us with the tools to assign conformations obtained via MD simulation into energy basins.

In this paper, we collect conformations from various MD trajectories and group them into basins. We treat such basins as macrostates. We then make use of the capabilities in PyEMMA to compute state-to-state transition probabilities and so obtain an MSM summarizing the dynamics probed via the MD trajectories. We now proceed to relate further methodological details in Section III.

III. Methods

We first describe the methodology and then provide further details on the MD trajectories employed to obtain conformations of the free and antigen-bound antibody, as well as on the various settings employed to represent and prepare such conformations for analysis.

A. From MD Trajectories to an MSM of Dynamics

At its core, the pipeline takes in MD trajectories and returns an MSM transition probability matrix, where each entry specifies the probability of transition between a pair of identified macrostates. The first objective is to identify such macrostates. Once such states are identified, the second objective is to use the temporal information available in the MD trajectories to extract information on the accessibility of macrostates in terms of state-to-state probabilities of transition. The result is an MSM that summarizes the dynamics captured in the MD trajectories. Various statistical techniques then inform on the quality of the obtained MSM.

Each MD trajectory is a series of conformations accessed consecutively in an MD simulation. An important decision with implications for the ability of the pipeline to identify macrostates relates to the degree of the retained conformation representation detail for the purpose of assigning conformations to macrostates. This is known as featurization in MSM literature and is computationally ill-posed. We address it empirically, by considering several reasonable representations and analyzing their impact on the resulting MSMs.

The second important decision concerns the lag time. It is often not feasible to analyze all the conformations obtained via MD simulation. Most MSM construction software allow the user to select a lag time. This can be a multiple of the original time step between two successive conformations in an MD trajectory. Selection of the lag time is in its essence a data reduction strategy, as it allows skipping over conformations; as it can result in loss of temporal and spatial resolution of the constructed MSM, we analyze its impact on the quality of the resulting MSM, as well.

1) From MD Trajectories to Macrostates (Basins in the Landscape): As we have shown in earlier works [26], [34], a better strategy to identify macrostates is to utilize (rather than ignore) the energies of conformations obtained via MD simulation. The motivation is clear. Clustering only harnesses the geometrical similarity, ignoring energetic similarity. However, the landscape contains information on how conformations with similar energies inter-convert into one another and so provides an opportunity to quantitatively understand the underlying dynamics of a system of interest [35]. Our definition of macrostates utilizes this energy landscape view of dynamics. A macrostate, which is a thermodynamically-stable (or semi-stable) state does not directly rely on geometric similarity but instead corresponds to basins/wells in the energy landscape. So we leverage our earlier work on summarizing sampled conformation-energy pairs via energy basins [33].

Identification of basins relies on embedding the conformations collected over all MD trajectories into a nearest-neighbor graph [36], where each vertex is a conformation-energy pair. We skip over methodological details here in the interest of space, noting that the methodology to identify basins has

been utilized by our laboratories to support several molecular structure-function enquiries [26], [37]–[39].

In summary, the methodology groups conformations into distinct, non-overlapping energy basins. The conformation that sits at the very bottom of a basin is denoted as its *focal minimum* and is used as a unique identifier of a basin. Since basins contain actual conformations, a basin can be summarized in terms of energies of the conformations in it (minimum, mean, maximum energy), as well as the geometric dissimilarity among conformations, measured, for instance, via the maximum pairwise root-mean-squared-deviation (RMSD) [40] between conformations in a basin. We consider the identified basins to be the macrostates.

2) From Macrostates to State-to-State Transition Probabilities (The Construction of the MSM): State-to-state transition probabilities are now computed. The MD trajectories are utilized for this purpose. In a given MD trajectory, a conformation a is followed by a conformation b. Let us consider that the process above, which has assigned conformations into macrostates, has assigned some conformation a to some macrostate S_i and some conformation b to some macrostate S_j. The observed transition from a to b is the evidence of the transition from macrostate S_i to macrostate S_j and thus contributes one count to the total counts of transitions from S_i to S_j. In this way, the various MD trajectories contribute to "counts" of transitions between macrostates. The counts are normalized to turn them into probabilities.

Let us assume that the basin identification process above has resulted in M disjoint states S_1, S_2, ..., S_M. A matrix of conditional transition probabilities between these states is estimated from the simulation trajectories x_t. [19] The transition matrix, $\mathbf{T} \equiv (\ \mathrm{P}_{ij})$: $\mathrm{P}_{ij}(\tau) = Prob\ (\mathrm{x}_{t+\tau} \in \mathrm{S}_j \mid \mathrm{x}_t \in \mathrm{S}_i)$, where τ is the chosen lag time. This transition matrix is the tangible product of what is referred to as the construction of the MSM. It contains all the information needed, as every entry in the transition matrix specifies the probability with which two states inter-convert into each-other, thus summarizing the dynamics of the system under investigation.

The transition matrix can provide more information about the system under investigation through its eigen-decomposition into eigenvectors and eigenvalues. The highest eigenvalue (with a value of 1) and its corresponding eigenvector represents the equilibrium/stationary distribution. The higher the population of a macrostate, the more thermodynamically-stable the macrostate is.

3) Statistical Evaluation: The MSM resulting from the computational process described above is subjected to rigorous analysis in order to evaluate whether the constructed MSM is reliable to utilize for making predictions regarding dynamics. As in other studies, we employ two main tests, the convergence analysis and the Chapman-Kolmogorov (CK) test. Both test for the Markov property that the MSM is memory-less; that is, the conditional probability distribution of future macrostates depends only upon the current macrostate and not on prior macrostates. [31] In our analysis, we conduct both evaluations whether the state-space decomposition (assignment of conformations into macrostates) results in a high-quality MSM.

a) Convergence Analysis: The convergence analysis tests whether the duration of the lag time is sufficient to guarantee that the state space discretization maintains the Markov property. If the state space decomposition is accurate, conformations within a state inter-convert on timescales faster than the lag time and transition to other states on slower timescales. It is standard practice to verify this property visually, via interpretation of the generated implied timescale plot of the model relaxation timescale versus model lag time. One expects to see an exponential decay in the plot to system equilibrium. With relaxation timescales being physical properties of the system, ideally, the implied timescales need to be independent of the lag time. According to the variational principle of conformation dynamics [41], it is desirable for the model to have a longer timescale. For an ideal model with good discretization, the implied timescales plot exhibit convergence within fewer steps.

b) CK Test: Discretization error can result in a deterministic fluctuation of the MSM dynamics from the actual dynamics that remains persistent even when excluding statistical error by means of excessive sampling. [42] The propagation error on the discrete space is measured by checking whether the approximation, $[\hat{\mathbf{T}}(\tau)]^k \approx \hat{\mathbf{T}}(k\tau)$ holds within statistical uncertainty where, $\hat{\mathbf{T}}(\tau)$ is the transition matrix estimated from the data at lag time τ, and $\hat{\mathbf{T}}(k\tau)$ is the transition matrix estimated from the same data at longer lag times kτ. The software we use for this purpose, PyEMMA, allows testing different models. Given a model estimated at lag time τ, a prediction can be made of a model quantity for lag time $k\tau$; the prediction can then be compared to an independently-estimated model at $k\tau$. The CK test computes the transition probability between meta-stable states for increasing lag times. The determination is made visually; ideally, plots show that the estimated and the predicted model exhibit negligible deviation.

B. Conformation Generation and Preparation

1) MD Simulations for Conformation Generation: The free antibody molecule (PDB ID: `1IGT`) contains $20,544$ atoms($1,322$ amino acids). The antigen-antibody complex contains $21,092$ atoms ($1,356$ amino acids). Initial antibody random conformations are generated by adjusting three sets of torsion angles: 231C-232N-232CA-232C, 232N-232CA-232C-233N, and 232CA-232C-233N-233CA (numbering in `1IGT`), each step with 60° rotation. During the conformation randomization, the `Fc` domain is kept fixed, whereas the `Fab` domains move freely, leading to 216 conformations. Excluding conformations with closed `Fab` domain or with `Fc` domain clashes, 12 conformations are selected as the starting points for the MD simulations. That is, we perform 12 independent MD simulations of the antibody with 12 different initial conformations. Each MD trajectory is $53-54$ns long, with a time step of $4ps$ between two consecutive frames/conformations. Thus, a total of $160,000$ conformations are generated for the free antibody and the antibody-antigen complex, respectively,

in this manner. The MD simulations are conducted using the NAMD software [43] with CHARMM force field [1]. Further details regarding the process of conformation generation can be found in Ref [1].

C. Preparation of Conformations for Analysis

From a given trajectory file (`.dcd` format), conformations are extracted using the *mdconvert* command-line script from the *MDTraj* python library. Considering a time lag of 128ps, which corresponds to selecting every 32^{nd} frame in a trajectory file, we obtain $5,000$ conformations for the free antibody and the antigen-antibody complex, respectively. We consider several options for representing conformations of each system.

1) Cartesian Coordinate-based Representations: **Setting 1:** In this setting, each conformation is simplified to a high-dimensional point ($CA_1.x$, $CA_1.y$, $CA_1.z$, $CA_5.x$, $CA_5.y$, $CA_5.z$, $CA_{10}.x$, $CA_{10}.y$, $CA_{10}.z$, ...), effectively skipping every 4 consecutive *CA* atoms. This makes it computationally feasible to compare conformations, which is a central step to embedding conformations in a nearest-neighbor graph for the purpose of identifying basins. **Setting 2:** In this setting, we consider all CA atoms; it is more computationally costly to compare conformations via this representation and identify basins but one does retain more detail.

2) Principal Component-based Representations: We rely on Principal Component Analysis (PCA) to identify a few variance-preserving dimensions along which to project collected conformations and obtain "reduced" coordinates for them. PCA is popular to analyze protein conformations [37], [44]–[46]. Specifically, we construct a matrix $A_{3k\times n}$, where $n = 5,000$ conformations (see above) and k is the number of CA atoms in the molecule of interest. All conformations are first optimally superimposed over the first conformation (chosen arbitrarily to be the reference conformation). The reference conformation is then "subtracted" from each of the conformations, and the matrix stores the resulting deviations. The purpose for this is to capture internal conformation changes rather than differences due to rigid-body motions (translations and rotations in three-dimensional space). A singular value decomposition of $\frac{1}{n-1}\cdot A$ is then carried out to obtain the (eigen-)decomposition $\frac{1}{n-1}\cdot A_{3k\times n} = U_{3k\times n}\cdot S_{n\times n}\cdot V^T_{n\times n}$. The procedure employed for this decomposition here is the `dgesvd` routine, which is available in Python (`scipy.linalg.lapack.dgesvd`) as part of the `scipy.linalg.lapack library`. While more information can be obtained in the corresponding documentation, in summary, the eigenvectors/PCs are $3k$-dimensional vectors stored in the columns of U, in order of corresponding highest-to-lowest singular values; these are stored in the diagonal of the S matrix. The singular values σ_i are square roots of the eigenvalues e_i, which provide the variance of the original (displacement) data over the corresponding eigenvector PC_i.

The main decision after utilizing PCA is to determine how many (projection) coordinates to employ to represent a conformation. Typically, an accumulation of variance analysis is conducted. After ordering the PCs by corresponding eigenvalue (highest to lowest), at each PC_i, the cumulative variance of $\{PC_1, \ldots, PC_i\}$ is plotted (not shown here in the interest of space). The individual variance of each PC is its eigenvalue, normalized over all eigenvalues (of all n obtained PCs).

For the free antibody, the first five PCs cumulatively cover 80.89% of the total variance. The first eight PCs cover 90.51% of the variance, and the first 23 PCs cover 99.05% of the variance. Similarly, for the antigen-antibody complex, the first five PCs cumulatively cover 82.01% of the variance; the first nine PCs cover 91.77% of the variance, and the first 24 PCs cover 99.09% of the variance. **Setting 3:** We set our goal at capturing no lower than 90% of the total variance. This means that for the free antibody, each conformation is represented with 8 coordinates (projections of a conformation on the top 8 PCs); for the antigen-antibody complex, this threshold means that each conformation is represented with 9 coordinates. **Setting 4:** We set the goal at capturing no lower than 99% of the variance. This means that for the free antibody, each conformation is represented with 23 coordinates; for the antigen-antibody complex, this threshold means that each conformation is represented with 24 coordinates.

3) Time-lagged independent component-based Representations: TICA is a linear transformation method. In contrast to PCA, which finds coordinates of maximal variance, TICA finds coordinates of maximal auto-correlation at the given lag time. TICA is useful to find the slow components in a dataset and a reasonable choice to transform MD data. **Setting 5:** In this setting, our goal is to capture no lower than 90% of the total variance. This means that for the free antibody, each conformation is represented with 2273 co-ordinates (projections of a conformation on the top 2273 components); for the antigen-antibody complex, this threshold means that each conformation is represented with 2341 coordinates. **Setting 6:** We now set the goal at capturing no lower than 99% of the variance. This means that for the free antibody, each conformation is represented with 3189 coordinates; for the antigen-antibody complex, this threshold means that each conformation is represented with 3275 coordinates.

D. Evaluation Setup

Whether for the free antibody or the antigen-antibody complex, we construct six different MSMs corresponding to the settings described above. Convergence analysis and CK tests are utilized to evaluate the quality of each MSM. The best MSM (setting-1) obtained for each system is investigated in greater detail in Section IV, and comparisons are made to understand the main differences in the conformation dynamics between the free and antigen-bound antibody.

IV. Results

In the interest of space, the basins obtained in each of the six settings, as well as the resulting MSMs and the evaluation of their quality (in terms of the Markov property) for the free antibody and antigen-antibody complex are spared. According to the convergence analysis and the CK test (results not shown

in the interest of space), the best MSM obtained for the free antibody as well as for the antigen-antibody complex results from Setting 1. The rest of the analysis in this section focuses on the basins and MSMs resulting from this setting, respectively.

A. Visual Comparison of Embedded Landscapes

We first relate in Fig. 1 a two-dimensional embedding of the energy landscape of the free antibody and the antigen-antibody complex, respectively. The 5000 conformations sampled from the 12 MD trajectories for each system (as described in Section III), are subjected to PCA. Each dot shows the projection of a conformation on the top two (highest-variance) PCs; two PCs capture close to 60% of the total variance for the free antibody and more than 60% of the variance for the antigen-antibody complex. The color-coding relates the energies of the projected conformations, with a blue-to-red color-scheme denoting low-to-high internal energies.

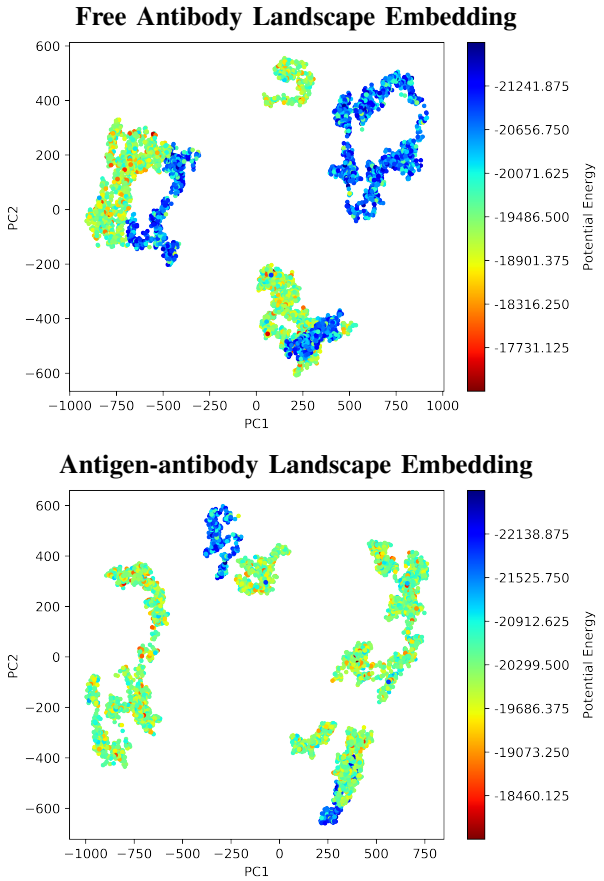

Fig. 1. Embedding of sampled conformations over the top two PCs; projections are color-coded by the energies of corresponding conformations (low-to-high in a blue-to-red color scheme).

The energy landscape of the free antibody contains four main clusters with a diffused distribution of low-energy conformations (see top panel). After antigen binding, the distribution of the clusters becomes more diffusive (see bottom panel), and the low-energy conformations are enriched in only two clusters. These observations hold even when more conformations are included in the analysis; adding $5,000$ more conformations for each system (selected every 16th frame over the MD trajectories) does not change the main features of the embedded landscapes; existing clusters do not merge, and no new clusters emerge (data not shown).

B. Basin Analysis

Thirteen basins are identified over the free antibody landscape; twelve basins are identified over the antigen-antibody conformations. We evaluate whether the identification of basins is biased by the starting conformation of the various MD trajectories as follows. The focal minimum conformation that sits at the bottom of a basin is considered as an identifier and representative conformation of a basin. The RMSD between this conformation and each conformation starting an MD trajectory is calculated. This calculation is repeated for each basin (representative conformation) against all starting conformations (of all MD trajectories), and related in Fig. 2.

Fig. 2 shows that the identification of basins is not biased by the conformations starting the MD trajectories. The method does not trivially assign the conformation that initiates an MD trajectory as the focal minimum of a basin. For the majority of the basins, the focal minimum conformation that uniquely represents a basin (its deepest point) resides on average more than 10Å away from the starting conformation of a trajectory. Some lower values are noted: between 5 and 10Å for 3 of the focal minima identified for the free antibody and for 2 of the focal minima identified for the antigen-antibody complex; in the latter case, one of the focal minima resides closer than 5Å to the starting conformation of an MD trajectory.

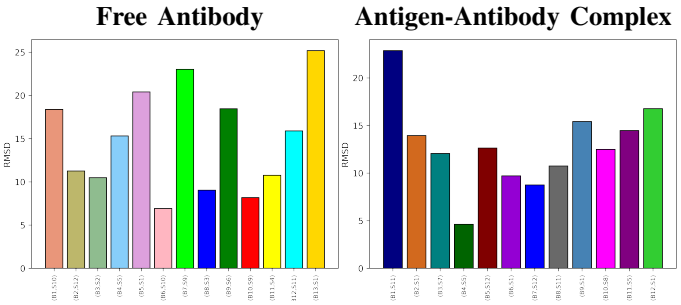

Fig. 2. The minimum RMSD (Å) (over backbone atoms) between the focal minimum conformation representing a basin and the conformations starting each of the MD trajectories is shown. The x axis shows the pairs corresponding to the minimum RMSD values. B* denotes (the focal minimum conformation of) a basin, numbered B1-13 for the free antibody and B1-12 for the antigen-antibody complex. S* denotes the conformation starting an MD trajectory, numbered S1-12.

C. Summarization and Comparison of Dynamics

The best MSMs constructed for each system, free antibody versus antigen-antibody complex, are related in Fig. 3. For each system, the equilibrium/stationary distribution is shown first in the top panel via a pie chart limited to the 6 most populous macrostates (labeled as B1-6 for basins; the numbering of basins observes the basin size). In each pie chart, the population of the remaining macrostates is accumulated

and labeled as B*. The bottom panel shows the transitions (again limited to the 6 most populous macrostates identified by the stationary distribution analysis).

Interesting observations can be drawn for the free antibody from the left panel of Fig. 3. One macrostate (largest basin), is significantly more thermodynamically-stable than the others, with an equilibrium population (31%) close to being 1.5 times than that of the next stable macrostate (23.95%). Three other macrostates have populations in the $8-16\%$ range, and the rest are 5.04% or lower. The self-transition probabilities are large. For four of the six basins, the self-transition probabilities are very high (above 0.94). Two exceptions are noted. Basin 1 has a lower self-transition probability of 0.8412. Basin 5 has an even lower self-transition probability of 0.623. While the cumulative out-of-basin transition probabilities for four basins (B2,B3,B4,B6) are just below 0.045; basin 5 transitions to basin 1 with a higher probability of 0.325, and basin 1 transitions to basin 5 with a probability of 0.1195.

A similar analysis can be carried out over the MSM shown in the right panel of Fig. 3 for the antigen-antibody complex. The stationary distribution is not as skewed as for the free antibody. There is no single macrostate with a population significantly higher than others; three macrostates (B1,B4,B6) have comparable populations in the $23-27\%$ range. The self-transition probabilities for the antigen-antibody complex are lower than those observed for the free antibody. Basins 1-3 have self-transition probabilities between 0.92 and 0.93. Basins 4-6 have much lower self-transition probabilities in the range 0.6-0.88. Basin 4 transitions to basin 1 with a probability of 0.11. Basin 5 transitions to basin 3 with a probability of 0.13 and to basin 2 with a probability of 0.1. Basin 6 transitions to basin 1 with a probability of 0.3. This suggests that the energy landscape of the antigen-antibody complex allows for more transitions among the various basins than the energy landscape of the free antibody.

D. Visualization of the Largest Basins.

Finally, we visualize the focal minima conformations corresponding to the 6 largest/most-populous basins for the free antibody and the antigen-antibody complex. Fig. 4 draws these conformations, using different colors for the various `IgG` domains and the antigen and the "*NewCartoon*" graphical representation in the VMD software. [47].

V. Conclusion

Based on the first two PCs and the conformational energies, we find that the conformational energy landscape of the free antibody mainly contains four clusters with a diffused distribution of low-energy conformers. After antigen binding, the distribution of the four cluster becomes more diffusive. However, the low-energy conformers' distributions narrowed and are enriched only in two of the four clusters. Such behavior provides new insights into previous analyses. Previous studies found that the free antibody has one major cluster that splits into two clusters after antigen binding. Both the current work and previous studies agree on the two major clusters of antigen-antibody complexes, but the analysis in this paper provides a more detailed view of the energy landscape.

The MSM-based analysis in this study shows that, with antigen binding, there are considerable conformation transitions among the different basins. These results suggest that the antigen-bounded form with high energy may provide many dynamic processes to further enhance co-factor binding of the antibody in the next step. We also observe that antigen binding causes reduction in the number of macrostates/basins across all the settings. Simulating the dynamics of large proteins and their complexes places large computational demands. Analyzing their conformation dynamics poses additional difficulties. Our current study indicates that rigorous MD simulations combined with a master equation framework represented by an MSM provides additional insights into antibody dynamics.

Acknowledgment

This work is supported in part by NSF Grant No. 1900061. Computations were run on ARGO, a research computing cluster provided by the Office of Research Computing at George Mason University, VA (URL:`https://orc.gmu.edu/`). This material is additionally based upon work by AS supported by (while serving at) the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261201500003I. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. This Research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and the Intramural Research Program of the NIH Clinical Center.

References

[1] J. Zhao, R. Nussinov, and B. Ma, "Antigen binding allosterically promotes Fc receptor recognition," *MAbs*, vol. 11, no. 1, pp. 58–74, 2019.

[2] Y. Chen, G. Wei, J. Zhao, R. Nussinov, and B. Ma, "Computational investigation of Gantenerumab and Crenezumab recognition of Abeta fibrils in Alzheimer's disease brain tissue," *ACS Chem Neurosci*, vol. 11, no. 20, pp. 3233–3244, 2020.

[3] I. Sela-Culang, V. Kunik, and Y. Ofran, "The structural basis of antibody-antigen recognition," *Front Immunol*, vol. 4, p. 302, 2013.

[4] T. Li, M. B. Tracka, S. Uddin, J. Casas-Finet, D. J. Jacobs, and D. R. Livesay, "Rigidity emerges during antibody evolution in three distinct antibody systems: Evidence from QSFR analysis of Fab fragments," *PLoS Comput Biol*, vol. 11, no. 7, p. e1004327, 2015.

[5] A. Janda, A. Bowen, N. S. Greenspan, and A. Casadevall, "Ig constant region effects on variable region structure and function," *Front Microbiol*, vol. 7, p. 22, 2016.

[6] J. Zhao, R. Nussinov, and B. Ma, "Mechanisms of recognition of amyloid-β (aβ) monomer, oligomer, and fibril by homologous antibodies," *J Biol Chem*, vol. 292, no. 44, pp. 18 325–18 343, 2017.

[7] W. Lua, W. Ling, J. I. Yeo, J. Poh, D. P. Lane, and S. K. Gan, "The effects of antibody engineering CH and CL in trastuzumab and pertuzumab recombinant models: Impact on antibody production and antigen-binding," *Sci Rep*, vol. 8, p. 718, 2018.

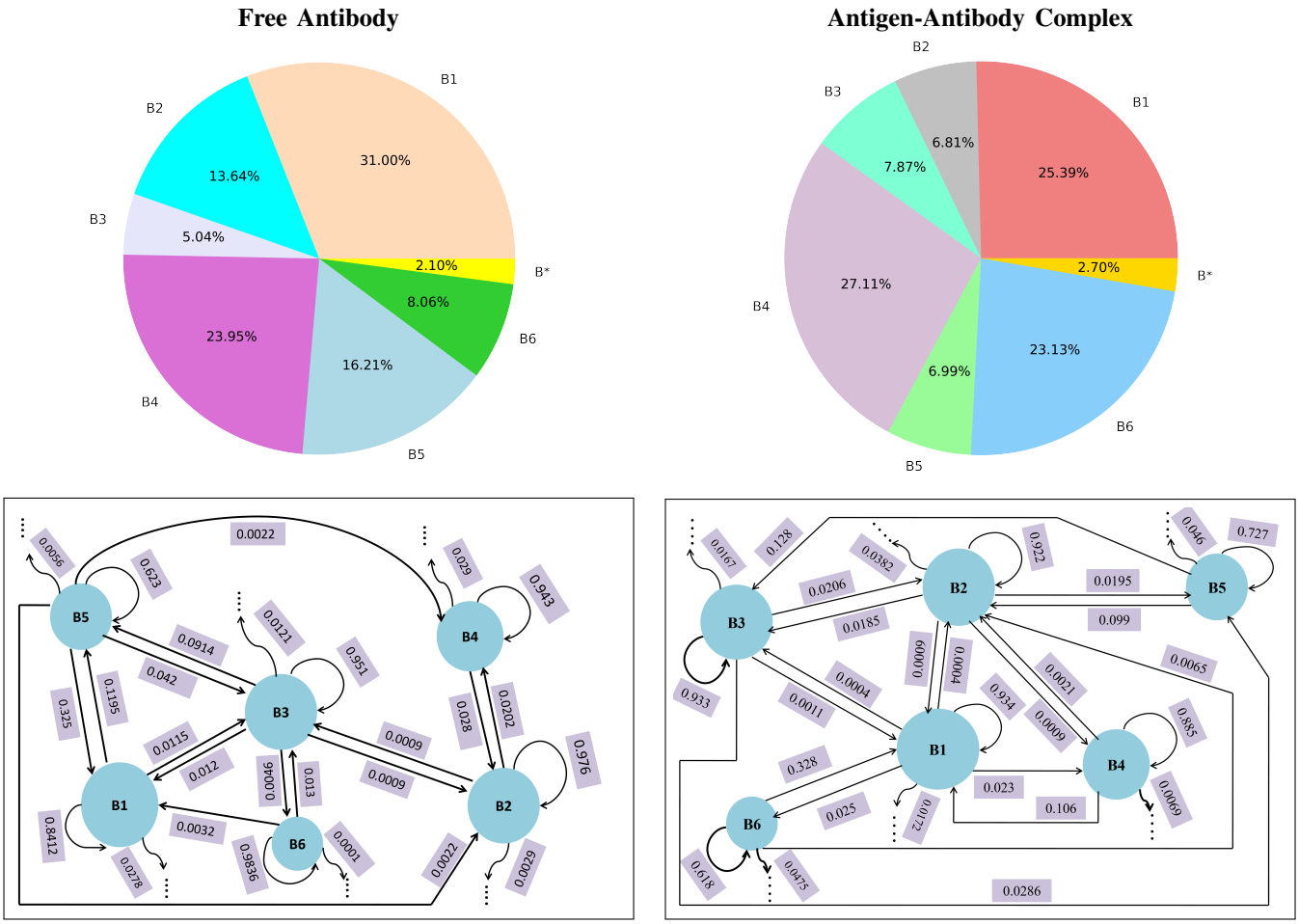

Fig. 3. Top Panel: Pie chart of adjusted state populations, showing the stationary distribution for the 6 top-populated macrostates/basins, with the other states accumulated in B*. Bottom Panel: MSM schematic. Basins are drawn as disks, with radii proportional to size (number of conformations). Transitions between basins are drawn as arrows, and transition probabilities are shown. The visual summary is restricted to the six top-populated states. Trailing arrows indicate transitions to other states.

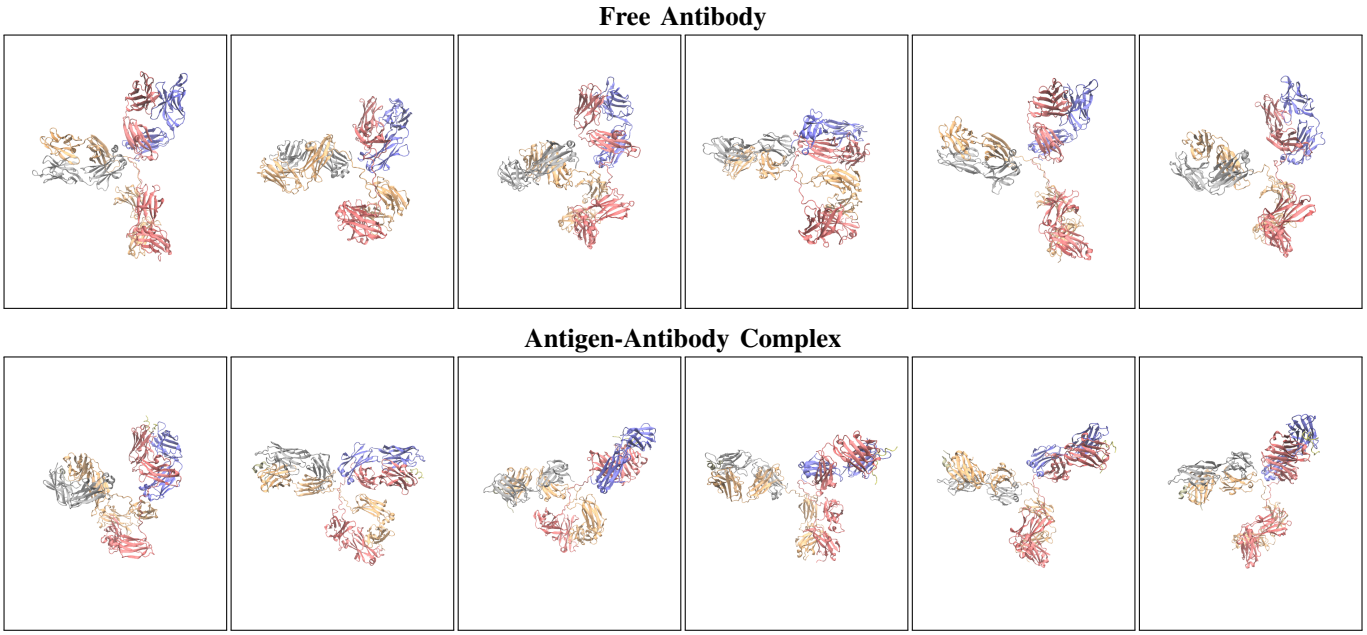

Fig. 4. Focal minima conformations corresponding to basins B1-6 in the MSMs shown schematically above are drawn here with VMD [47] for (top panel) the free antibody and the (bottom panel) antigen-antibody complex, respectively. Chains are drawn in different colors.

[8] I. Sela-Culang, S. Alon, and Y. Ofran, "A systematic comparison of free and bound antibodies reveals binding-related conformational changes," *J Immunol*, vol. 189, no. 10, pp. 4890–4899, 2012.

[9] A. Janda and A. Casadevall, "Circular dichroism reveals evidence of coupling between immunoglobulin constant and variable region secondary structure," *Mol Immunol*, vol. 47, no. 7-8, pp. 1421–1425, 2010.

[10] A. Janda, E. Eryilmaz, A. Nakouzi, D. Cowburn, and A. Casadevall, "Variable region identical immunoglobulins differing in isotype express different paratopes," *J Biol Chem*, vol. 287, no. 42, pp. 35 409–35 417, 2012.

[11] A. Correa, F. Trajtenberg, G. Obal, O. Pritsch, G. Dighiero, P. Oppezzo, and A. Buschiazzo, "Structure of a human IgA1 Fab fragment at 1.55Å Resolution: Potential Effect of the Constant Domains on Antigen-affinity Modulation," *Acta Crystallogr D Biol Crystallogr*, vol. 69, no. Pt 3, pp. 388–397, 2013.

[12] L. J. Cooper, A. R. Shikhman, D. D. Glass, D. Kangisser, M. W. Cunningham, and N. S. Greenspan, "Role of heavy chain constant domains in antibody-antigen interaction. apparent specificity differences among streptococcal IgG antibodies expressing identical variable domains," *J Immunol*, vol. 150, no. 6, pp. 2231–2242, 1993.

[13] A. Torosantucci, P. Chiani, C. Bromuro, F. De Bernardis, A. S. Palma, Y. Liu, G. Mignogna, B. Maras, M. Colone, A. Stringaro, S. Zamboni, T. Feizi, and A. Cassone, "Protection by anti-beta-glucan antibodies is associated with restricted beta-1,3 glucan binding specificity and inhibition of fungal growth and adherence," *PLoS ONE*, vol. 4, no. 4, p. e5392, 2009.

[14] G. D. Tomaras, G. Ferrari, X. Shen, S. M. Alam, H. X. Liao, J. Pollara, M. Bonsignori, M. A. Moody, Y. Fong, X. Chen, B. Poling, C. O. Nicholson, R. Zhang, X. Lu, R. Parks, J. Kaewkungwal, S. Nitayaphan, P. Pitisuttithum, S. Rerks-Ngarm, P. B. Gilbert, J. H. Kim, N. L. Michael, D. C. Montefiori, and B. F. Haynes, "Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG," *Proc Natl Acad Sci USA*, vol. 110, no. 22, pp. 9019–9024, 2013.

[15] K. Kato, C. Matsunaga, A. Odaka, S. Yamato, W. Takaha, I. Shimada, and Y. Arata, "Carbon-13 NMR study of switch variant anti-dansyl antibodies: antigen binding and domain-domain interactions," *Biochemistry*, vol. 30, no. 26, pp. 6604–661, 1991.

[16] M. Torres, R. May, M. D. Scharff, and A. Casadevall, "Variable-region-identical antibodies differing in isotype demonstrate differences in fine specificity and idiotype," *J. Immunol*, vol. 174, no. 4, pp. 2132–2142, 2005.

[17] C. T. Su, W. Lua, W. Ling, and S. K. Ga, "Allosteric effects between the antibody constant and variable regions: A study of IgA Fc mutations on antigen binding," *Antibodies*, vol. 7, no. 2, p. 20, 2018.

[18] V. J. Pande, K. Beauchamp, and G. R. Bowman, "Everything you wanted to know about markov state models but were afraid to ask," *Methods*, vol. 52, no. 1, pp. 99–105, 2010.

[19] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," *Curr Opin Struct Biol*, vol. 25, pp. 135–144, 2014.

[20] R. Zwanzig, "From classical dynamics to continuous time random walks," *J Stat Phys*, vol. 30, pp. 255–262, 1983.

[21] B. E. Husic and V. J. Pande, "Markov state models: From an art to a science," *J Am Chem Soc*, vol. 140, no. 7, pp. 2386–2396, 2018.

[22] M. J. H. Ratcliffe, Ed., *Encyclopedia of Immunobiology*, 2016, vol. 7.

[23] M. P. Harrigan, M. M. Sultan, C. X. Hernánndez *et al.*, "MSMBuilder: Statistical models for biomolecular dynamics," *Biophys J*, vol. 112, no. 1, pp. 10–15, 2017.

[24] M. Senne, B. Trendelkamp-Schroer, A. S. Mey, C. Schutte, and F. Noé, "EMMA: A software package for markov model building and analysis," *J Chem Theory Comput*, vol. 8, no. 7, pp. 2223–2238, 2012.

[25] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J. Prinz, and F. Noé, "PyEMMA 2: A software package for estimation, validation, and analysis of markov models," *J Chem Theory Comput*, vol. 11, pp. 5525–5542, 2015.

[26] K. L. Kabir, N. Akhter, and A. Shehu, "From molecular energy landscapes to equilibrium dynamics via landscape analysis and markov state models," *J Bioinf and Comput Biol*, vol. 17, no. 6, p. 1940014, 2019.

[27] C. R. Schwantes and V. S. Pande, "Improvements in markov state model construction reveal many non-native interactions in the folding of ntl9," *Journal of chemical theory and computation*, vol. 9, no. 4, pp. 2000–2009, 2013.

[28] N. Plattner and F. Noé, "Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and markov models," *Nature communications*, vol. 6, p. 7653, 2015.

[29] N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, "Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and markov modelling," *Nature chemistry*, vol. 9, no. 10, p. 1005, 2017.

[30] B. C. Taylor, C. T. Lee, and R. E. Amaro, "Structural basis for ligand modulation of the ccr2 conformational landscape," *Proceedings of the National Academy of Sciences*, vol. 116, no. 17, pp. 8131–8136, 2019.

[31] R. D. Malmstrom, C. T. Lee, A. T. Van Wart, and R. E. Amaro, "Application of molecular-dynamics based markov state models to functional proteins," *J Chem Theory Comput*, vol. 10, no. 7, pp. 2648–2657, 2014.

[32] G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, "Progress and challenges in the automated construction of markov state models for full protein systems," *The Journal of chemical physics*, vol. 131, no. 12, p. 124101, 2009.

[33] N. Akhter and A. Shehu, "From extraction of local structures of protein energy landscapes to improved decoy selection in template-free protein structure prediction," *Molecules*, vol. 23, no. 1, p. 216, 2018.

[34] K. L. Kabir, N. Akhter, and A. Shehu, "Connecting molecular energy landscape analysis with markov model-based analysis of equilibrium structural dynamics," in *Proceedings of 11th International Conference on Bioinformatics and Computational Biology*, vol. 60, 2019, pp. 181–189.

[35] R. Nussinov and P. G. Wolynes, "A second molecular biology revolution? the energy landscapes of biomolecular function," *Phys Chem Chem Phys*, vol. 16, no. 14, pp. 6321–6322, 2014.

[36] K. L. Kabir, L. Hassan, Z. Rajabi, N. Akhter, and A. Shehu, "Graph-based community detection for decoy selection in template-free protein structure prediction," *Molecules*, vol. 24, no. 5, p. 854, 2019.

[37] N. Akhter, W. Qiao, and A. Shehu, "An energy landscape treatment of decoy selection in template-free protein structure prediction," *Computation*, vol. 6, no. 2, p. 39, 2018.

[38] N. Akhter, G. Chennupati, K. L. Kabir, H. Djidjev, and A. Shehu, "Unsupervised and supervised learning over the energy landscape for protein decoy selection," *Biomolecules*, vol. 9, no. 1, p. 607, 2019.

[39] N. Akhter, G. Chennupati, H. Djidjev, and A. Shehu, "Decoy selection for protein structure prediction via extreme gradient boosting and ranking," *BMC Bioinf*, vol. 21, no. Suppl 1, p. 189, 2020.

[40] A. D. McLachlan, "A mathematical procedure for superimposing atomic coordinates of proteins," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 28, no. 6, pp. 656–657, 1972.

[41] F. Noé and F. Nuske, "A variational approach to modeling slow processes in stochastic dynamical systems," *Multiscale Modeling & Simulation*, vol. 11, no. 2, pp. 635–655, 2013.

[42] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: Generation and validation," *The Journal of chemical physics*, vol. 134, no. 17, p. 174105, 2011.

[43] J. C. Phillips, D. J. Hardy, J. D. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang *et al.*, "Scalable molecular dynamics on cpu and gpu architectures with namd," *The Journal of chemical physics*, vol. 153, no. 4, p. 044130, 2020.

[44] L. Orellana, O. Yoluk, O. Carrillo, M. Orozco, and E. Lindahl, "Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations," *Nat Commun*, vol. 7, p. 12575, 2016.

[45] R. Clausen, B. Ma, R. Nussinov, and A. Shehu, "Mapping the conformation space of wildtype and mutant H-Ras with a memetic, cellular, and multiscale evolutionary algorithm," *PLoS Comput Biol*, vol. 11, no. 9, p. e1004470, 2015.

[46] W. Qiao, N. Akhter, X. Fang, T. Maximova, E. Plaku, and A. Shehu, "From mutations to mechanisms and dysfunction via computation and mining of protein energy landscapes," *BMC Genomics*, vol. 19, no. Supp 17, p. 671, 2018.

[47] W. Humphrey, A. Dalke, and K. Schulten, "VMD - Visual Molecular Dynamics," *J. Mol. Graph. Model.*, vol. 14, no. 1, pp. 33–38, 1996, http://www.ks.uiuc.edu/Research/vmd/.