

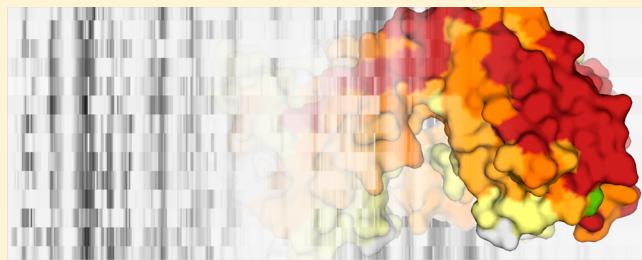
# Energetic and Dynamic Aspects of the Affinity Maturation Process: Characterizing Improved Variants from the Bevacizumab Antibody with Molecular Simulations

Dario Corrada and Giorgio Colombo\*

Istituto di Chimica del Riconoscimento Molecolare – Consiglio Nazionale delle Ricerche (CNR-ICRM), via Mario Bianco 9, 20131 Milano, Italy

## Supporting Information

**ABSTRACT:** Antibody affinity maturation is one of the fundamental processes of immune defense against invading pathogens. From the biological point of view, the clonal selection hypothesis represents the most accepted mechanism to explain how mutations increasing the affinity for target antigens are introduced and selected in antibody molecules. However, understanding at the molecular level how protein modifications, such as point mutation, can modify and modulate the affinity of an antibody for its antigen is still a major open issue in molecular biology. In this paper, we address various aspects of this problem by analyzing and comparing atomistic simulations of 17 variants of the bevacizumab antibody, all directed against the common target protein VEGF-A. In particular, we examine MD-based descriptors of the internal energetics and dynamics of mutated antibodies and their possible correlations with experimentally determined affinities for the antigens. Our results show that affinity improvement is correlated with a variation of the internal stabilization energy of the antibody molecule when bound to the antigen, compensated by the variation in the interaction energy between the antigen and the antibody, paralleled by an overall modulation of internal coordination within the antibody molecular structure. A possible model of the mechanism of rigidification and of the main residues involved is proposed. Overall, our results can help in understanding the molecular determinants of antigen recognition and have implications in the rational design of new antibodies with optimized affinities.



## INTRODUCTION

The interaction between an antibody (Ab) and its antigen (Ag) plays a critical role in immune defense against invading pathogens, and the high affinity and specificity observed in this interaction has made Ab::Ag complexes a paradigm for understanding molecular recognition. Antibodies are multi-chain proteins that consist of two identical light chains and two identical heavy chains. The overall protein structure is Y-shaped, in which three distinct regions are connected by flexible protease-sensitive linkers: two Ag binding regions (Fab fragments) and an effector region (Fc fragment). A Fab fragment is defined by four distinct immunoglobulin (Ig) domains: two variable domains  $V_L$  and  $V_H$  and two constant domains  $C_L$  and  $C_{H1}$ . The Ag binding site is restricted to variable domains, and in most cases, it is composed of six external loops termed complementarity determining regions (CDR, also called hypervariable regions).<sup>1</sup> The expression of the Ig chains involves an unusual arrangement of their related genes that lead to a random assembly of the codifying DNA segments, called somatic recombination. Such sequence recombination generates a diverse repertoire of germline receptors allowing peculiar specificity toward a wide range of Ag molecules that range from small ligands to large proteins.<sup>2</sup> In humoral (i.e., Ab-mediated) immunity, the activation of

naïve B cells is induced by contacts with the Ag, and it is followed by both a cellular response (proliferation and differentiation) and a molecular response (isotype switching and affinity maturation). In the molecular response, the immune system rapidly generates high affinity Abs with the introduction of extra somatic mutations in a process known as affinity maturation. In parallel, the functional optimization of these Abs proceeds through isotype switching.<sup>3–5</sup> In this phase, the affinity of an Ab can be improved up to  $10^3$  fold.<sup>6,7</sup>

In a previous study aimed at shedding light on the determinants of Ag recognition, we collected a data set of 28 Ab::Ag complexes, including Ags with different sizes and folds.<sup>8</sup> Through an analysis of the internal dynamics and energetics of the Ab molecules, we could define the global common features shared by distinct Ab::Ag complexes; in particular, the heavy and the light chain domains were shown to behave in different ways, irrespective of their high similarity in terms of primary sequence and folding. Next, we proposed a novel computational method for finding strong interaction centers—accounting for nonbonded interaction energy (IERPs) and mechanical coordination (CMRPs)—that accurately describes

Received: May 24, 2013

Published: October 29, 2013



the allosteric variations that involve Ab structures upon Ag binding.

In this work, we will focus on a group of Ab variants generated by an *in vitro* approach that mimics the affinity maturation process. Our data set comprises Fab molecules derived by introducing multiple mutations into the regions that compose the Ag binding site such as CDR-H1, CDR-H2, and CDR-H3. Furthermore, two specific studied variants show multiple mutations on the framework region HFR3 located on the heavy chain. In particular, our goal here is to exploit recently introduced methods for the analysis of protein energetics and internal dynamics<sup>8–16</sup> to investigate the effects of mutations in the Fab sequence on its stability and conformational properties and relate such effects to modulation of the affinity for the antigen. Importantly, mutations involve only the Ab molecule, while the Ag (and the related epitope) is invariant.

In this context, Chen and co-workers have adopted a phage display technique combined with off-rate selection of enhanced affinity variants.<sup>17</sup> The starting *wildtype* Fab molecule (Fab-12, ref 18) is bevacizumab, an IgG1 Ab approved for treating patients with certain types of colon, breast, and lung cancer.<sup>19</sup> This therapeutic Ab is directed against the vascular endothelial growth factor (VEGF-A), a cytokine whose increased activity and interactions with two tyrosine kinase receptors (Flt-1/VEGFR1 and Flk-1/KDR/VEGFR2) plays a key role in the neo-angiogenesis process of several solid tumors.<sup>20–22</sup> The Fab portion of the humanized Ab bevacizumab was solved in complex with VEGF-A (PDB ID: 1BJ1), indicating the importance of a selected group of residues along the primary sequence for efficient mutagenesis.<sup>17,18</sup> Indeed, mutation of buried or partially buried residues has been shown to affect binding by altering the Ag recognition mediated by CDR loops. The specificity of bevacizumab was revealed by alanine-scanning mutagenesis of VEGF-A.<sup>18,23</sup> Interestingly, experimental data have been generated on the relationships between the presence of specific mutations in the Ab and changes in affinity and binding kinetics with the Ag.<sup>17</sup>

In the first part of the paper, we will analyze internal stabilization energies in response to specific mutations of the Ab in the Ab::Ag complexes to define a quantitative relationship between the calculated modulation of Ab energetics and experimental dissociation constants  $K_d$ . In the second part, we will analyze the mutation dependent internal dynamics and coordination of each distinct Ig domain in the Ab mutants. The overall goal of these two analyses is to investigate to what extent energetic interaction networks and internal coordination patterns are affected during the affinity maturation process and in what relative amount.

Finally, in the third part, we will investigate the collective motions that connect the individual Ig domains of the related mutant antibodies and how they determine conformational changes upon Ag binding. We will adopt the so-called coordination propensity (CP) approach already introduced and validated by our group in other protein models.<sup>11</sup>

## METHODS

**Affinity Evaluations of the Ab::Ag Complexes.** Affinity data were experimentally determined in ref 17 using a random mutagenesis and affinity-based selection strategy. The Ag binding affinities were calculated experimentally by surface plasmon resonance (SPR). The assay consists of the direct measurement of association rate ( $k_{on}$ ) and dissociation rate

( $k_{off}$ ), and the equilibrium dissociation constant  $K_d$  is calculated from the ratio  $k_{off}/k_{on}$ .<sup>18</sup>

We will refer to  $pK_a$  as a measure of affinity ( $pK_a = -\log(1/K_d)$ ). The logarithmic transformation of  $K_d$  values has been adopted because this quantity is directly related to differences in free energy ( $\Delta G = -2.303RT \log K_a$ ), which we aim to qualitatively approximate using parameters extracted from the analysis of the main determinants of stabilization energy, through the calculation of  $E_{nb}$  via the energy decomposition approach and internal dynamics through the analysis of distance fluctuations,  $D_{ij}$  (vide infra).

The correlations between affinity values and our parameters have been calculated using either Pearson's and Spearman's rank coefficients. In case of linear correlations, the two numerical values are expected to be similar, otherwise the presence of outliers can influence the Pearson's  $r$  calculation. Logarithmic transformation is a widely accepted technique to dampen the background noise generated from outliers and the adoption of  $pK_a$  values yields comparable coefficients from Pearson's and Spearman's rank correlations (Table S1, Supporting Information).

### Structure Data Set and Molecular Dynamics (MD)

**Simulations.** The starting coordinates of the simulated systems were obtained from the X-ray crystal structures representing the Fab fragment of an anti-VEGF-A Ab<sup>18</sup> in the *wildtype* (PDB ID: 1BJ1) and affinity matured forms<sup>17</sup> (PDB ID: 1CZ8). Most of the mutants were modeled starting from 1BJ1, while the structure of variant Y0317 was obtained directly from 1CZ8. In both crystallographic structures, the amino acids of the AB loop in the C<sub>H1</sub> domain (residues 128–134) were missing.<sup>18</sup> Such residues were modeled using the Modeller 9v8 software, and the subsequent models were optimized on the basis of the best DOPE score.<sup>24</sup> From each of the starting crystallographic structures 10 loop refined models were generated, resulting in average DOPE scores of –506.63 and –550.12 for the 1BJ1 and 1CZ8 templates, respectively. The detailed list of DOPE scores is provided in Table S2 of the Supporting Information. The DOPE method is generally used to assess the quality of a structure model as a whole, and the model with the lowest score is most likely to be the best quality one. The models selected in this work show DOPE scores of –577.98 and –650.65 for the 1BJ1 and 1CZ8 templates, respectively.

All the variants were generated through *in silico* mutagenesis using the PyMOL v1.3 software.<sup>25</sup> In total 17 Fab mutants were created. Each variant was created by selecting the rotamer of the mutated aminoacid that minimized the number of steric clashes with the surrounding neighbors. Upon visual inspection, in most cases, only one rotamer per mutant could be accommodated without creating excess strain and clashes. Such variant was chosen for simulations. For each variant, two independent systems were set up: the *holo* system containing the structure of the entire Fab::Ag complex and the *apo* system obtained after removing the Ag molecule. All the modeled structures were first energy minimized prior to perform molecular dynamics (MD) simulations with 2000 steps of the steepest descent method to remove bad contacts and nonoptimal lengths and angles we have applied.

For all of systems, 50 ns molecular MD simulations were performed. The simulation protocol adopted was extensively documented in ref 8. Briefly, every system is submitted to an unrestrained all-atom MD simulation in isothermal–isobaric (NPT) conditions, with explicit solvent and periodic boundary

conditions, using the GROMACS 4.0.7 software<sup>26</sup> with the GROMOS96 43A1 forcefield.<sup>27</sup> The box dimensions, number of water molecules, and counterions added for neutralization are listed in Table S3 of the Supporting Information.

The convergence of every simulation was evaluated by monitoring the atomic positional Root Mean Square Deviation (RMSD) of the backbone atoms with respect to the starting minimized structure. The simulations appear to be stable and the RMSD profiles (Figure S1, Supporting Information) are comparable among all the systems, revealing no major structural rearrangements. All analyses were carried out on the part of the trajectory obtained after discarding the first 10 ns to avoid equilibration artifacts.

**Cluster Analysis Techniques.** Trajectories were subjected to cluster analysis, using in house scripts developed in Perl and R languages. Cluster analysis is based on a distance matrix derived from paired RMSD information. First, for each snapshot of the trajectory the C $\alpha$  RMSD with respect to every other structure is calculated, and then the structural similarity between different snapshots is inferred applying the Manhattan distance function.<sup>28</sup>

The Manhattan distance between two points is the sum of the differences of their corresponding components. In our case, the different components of each data point are considered as a quantitative measure of the conformational differences between a reference snapshot and the others, as defined by the RMSD matrix. If we consider that a conformational transition is composed of a discrete number of smaller changes, the overall degree of dissimilarity between two points of the matrix could be described as a path composed of intermediate steps. In this context, the Manhattan distance function was considered as a suitable choice because it computes the distance that would be traveled to get from one point to the other if a grid-like path is followed.

An agglomerative hierarchical clustering strategy was adopted, namely, the complete linkage method. The most representative cluster of a trajectory is defined herein as one of the most populated clusters in which every component (i.e., the individual snapshot) is well aggregated with each other. As a measure of internal consistency of cluster, we have evaluated the average silhouette value. Silhouette measures how well an element belongs to its cluster, and the average silhouette measures the overall strength of cluster membership.<sup>29</sup> Let us consider  $a(i)$  as the average dissimilarity between object  $i$  and all the other objects of the cluster to which  $i$  belongs to. For all other clusters  $K$ ,  $d(i,K)$  can be considered as the average dissimilarity of  $i$  to all observations of  $K$ . The smallest  $d(i,K)$  value is called  $b(i)$ , and it describes the dissimilarity between  $i$  and its neighbor cluster. The silhouette width is then defined as

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

Observations with a large silhouette value are well clustered ( $s(i) > 0.5$ ); a small value means that the observation probably lies between two adjacent clusters ( $0.25 \leq s(i) \leq 0.5$ ). Observations with  $s(i)$  lesser than 0.25 are probably placed in the wrong cluster.<sup>30</sup>

**Nonbonded Interaction Energy Analysis.** The major contributions to energetic stability of the native structure from all-atom MD simulations were extracted for each simulated system using the energy decomposition method.<sup>8,10,12,16</sup> Ten structures were randomly sampled from the most representative

cluster and were energy minimized. Pairwise nonbonded energies for every residue are computed using the molecular mechanics generalized born surface area (MM-GBSA) method.<sup>31</sup> The complete network of interactions that defines the global stabilization energy of the protein is described by a matrix  $M_{ij}$

$$M_{ij} = E_{\text{elect},ij} + E_{\text{vdW},ij} + G_{\text{solv},ij} \quad (2)$$

For each pair of residues  $i$  and  $j$ , nonbonded energies were measured by summing electrostatic  $E_{\text{elect},ij}$  and van der Waals  $E_{\text{vdW},ij}$  contribution. In addition, solvation effects  $G_{\text{solv},ij}$  were taken into account.

The energy matrix  $M_{ij}$  is then decomposed through eigenvalue decomposition. Such a method allows identifying the principal components that define the total nonbonded energy of a system providing a simplified picture of the network of residues that contribute the most structural stabilization.<sup>10,14</sup> Therefore, the energy matrix can be re-expressed in the form

$$M_{ij} = \sum_{k=1}^N \lambda_k w_i^k w_j^k \quad (3)$$

where  $N$  is the number of amino acids;  $\lambda_k$  is the  $k^{\text{th}}$  eigenvalue with  $k$  ranging from 1 to  $N$ ;  $w_i^k$  and  $w_j^k$  are the  $i^{\text{th}}$  and  $j^{\text{th}}$  components of the associated normalized eigenvector.

The total nonbonded energy was further approximated by selecting a restricted subset of eigenvalues and related eigenvectors. In this way, the main energetic determinants of structural stabilization and/or adaptation to a specific state are emphasized, minimizing the possible noise present in the full energy matrix. Small or single-domain proteins can be accurately described considering only the eigenvector associated to the first eigenvalue  $\lambda_1$ , accordingly to what observed in.<sup>10</sup> In the case of multi-domain structures like the Abs, more eigenvectors are typically needed to represent all the main interactions that are fundamental for the protein stability. Thus, we have considered a set of essential eigenvectors which recapitulate a sufficient amount of energetic information, according to the procedure proposed in.<sup>9</sup> In this work the total nonbonded energy was approximated by

$$E_{\text{nb}} = \sum_{i,j=1}^N M_{ij} \approx \sum_{i,j=1}^N \sum_k \lambda^k w_i^k w_j^k \quad (4)$$

where  $k$  belongs to an ensemble of selected eigenvalues specific for each system. The selection strategy is based on the collection of the smallest set of essential eigenvectors able to describe the largest part of protein folding units (i.e., individual domains) with the minimum redundancy. Each essential eigenvector is characterized by a block of significant component values that correspond to the residues of the associated domain. Such blocks are obtained from the autocorrelation analysis of the entire set of eigenvectors. The assessment of the amount of information that each new essential eigenvector introduces and the rules establishing at what extent two adjacent blocks can overlap have been accurately described in ref 9.

**Coordinated Distance Fluctuations Analysis.** In order to analyze the dynamical and mechanical features of the different systems as a function of mutation from MD simulations, we used a recent approach developed by our group and validated in previous publications.<sup>8,11,13,15</sup> For each trajectory, all of the snapshots belonging to the most

Table 1. Substitutions Pattern and Binding Kinetics (at 25 °C) of the Variants<sup>a</sup>

variant	$K_d$ (nM)	CDR-H1			CDR-H2		HFR3		CDR-H3		
		T28	T30	N31	Y53	L71	T73	H97	Y99	G100	S100a
Y0244-1	9.00	T	T	N	Y	V	V	H	Y	G	S
Y0244-4	3.30	T	T	N	Y	I	K	H	Y	G	S
Y0192	2.90 <sup>b</sup>	T	T	N	Y	L	T	H	Y	G	S
Y0242-1	2.30	T	T	N	W	L	T	H	Y	G	S
Y0192 (T28D)	2.00	D	T	N	Y	L	T	H	Y	G	S
Y0192 (S100aT)	1.50	T	T	N	Y	L	T	H	Y	G	T
Y0238-5	1.40	T	T	N	Y	L	T	Y	I	A	K
Y0243-1	0.90	D	T	H	Y	L	T	H	Y	G	S
Y0192 (N31H)	0.80	T	T	H	Y	L	T	H	Y	G	S
Y0238-10	0.80	T	T	N	Y	L	T	H	R	N	T
Y0238-7	0.50 <sup>c</sup>	T	T	N	Y	L	T	Y	W	G	T
Y0238-1	0.40	T	T	N	Y	L	T	H	R	G	T
Y0268-1	0.38	D	T	H	W	L	T	Y	Y	G	T
Y0238-3	0.30 <sup>c</sup>	T	T	N	Y	L	T	Y	Y	G	T
Y0192 (H97Y)	0.20	T	T	N	Y	L	T	Y	Y	G	S
Y313-1	0.15 <sup>c</sup>	D	T	H	Y	L	T	Y	Y	G	T
Y0317	0.14 <sup>c</sup>	D	T	H	Y	L	T	Y	Y	G	T

<sup>a</sup>Sequence numbering is according to Kabat style.<sup>57</sup> <sup>b</sup> $K_d$  value for Y0192 variant is from ref 18. <sup>c</sup> $K_d$  values are calculated from the upper limits of detection, as mentioned in ref 17.

representative cluster were collected. Such subset was used for the calculation of a matrix of distance fluctuations

$$D_{ij} = \frac{\langle d_{ij} \rangle}{\sqrt{\langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle}} \quad (5)$$

For  $i$  not equal to  $j$ ,  $d_{ij}$  is the pairwise distance between the Cα atoms of amino acids  $i$  and  $j$ , and the brackets indicate the average over the subset of timeframes considered. As can be inferred from eq 5,  $D_{ij}$  values are expressed in nondimensional units. Higher  $D_{ij}$  values denote mechanical coordination between two residues because they move in a rigid-like fashion with their relative distance undergoing small fluctuations. Next, the distance fluctuation matrix was diagonalized to obtain the minimal group of eigenvalues and eigenvectors sufficient to describe and emphasize the most relevant distance fluctuation couplings. This eigendecomposition procedure adopted is the same as in ref 14. Eigenvalues and eigenvectors have been selected following the procedure proposed in ref 9.

The approximated distance fluctuation matrix generated from using the selected eigenvalues and eigenvectors is suitable to identify local patches of strong mechanical coordination.<sup>8</sup> In this work, the sum of the contributions of each amino acid is used as a quantitative reporter of the global flexibility of a protein as well as to describe the internal coordination of a specific domain. The trajectories were further analyzed using the concept of coordination propensity (CP), introduced in ref 11 and here defined as

$$CP_i = \sum_{j=1}^N f_i(j) \quad (6)$$

For each residue  $i$ , the coordination propensity value  $CP_i$  is the summation of the contributions given by all the other residues  $j$  of the protein. Such contributions are defined on the basis of discrete function  $f_i(j)$  as follows

$$f_i(j) = \begin{cases} 1, & \text{if } D_{ij} > thr \text{ and } \langle d_{ij} \rangle > dist \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

A threshold value  $thr$  was established by averaging distance fluctuation values of residues pairs consecutive along the primary sequence, a sliding window of  $I \pm 4$  is chosen to define the neighboring residues. We established a cutoff distance,  $dist = 40 \text{ \AA}$ , that resembles the average maximum length of an Ig domain for our data set. A residue  $j$ , separated from the amino acid  $i$  by a distance above  $dist$ , can be referred as dynamically coupled ( $f_i(j) = 1$ ), if it shows a distance fluctuation value  $D_{ij}$  higher than the threshold  $thr$ .

This approach allows shedding light on the regions whose dynamics is related and possibly underlies conformational changes. The profile of CP values for a particular protein could be conveniently represented with a heat map in order to highlight the most coordinated residues along the primary sequence. Moreover, in this work, CP values have also been also mapped on the 3D structure of the Abs through a color ramp scale; in such way, pathways of coordinated motions can be emphasized. For the sake of simplicity, the CP values are normalized with the total amount of residues; therefore, CP is expressed as a value varying from 0 to 1.

## RESULTS

This work aims to investigate the physicochemical bases of the modulation of affinity for a certain protein Ag determined by mutations in the sequence of an Ab. To this end, we have carried out computational analyses on 17 mutational variants of the Fab fragment of the humanized Ab bevacizumab<sup>32</sup> by analyzing the *apo* and *holo* forms. Affinity measures have been collected from experimental data provided in ref 17. A summary of the pattern of amino acid substitutions and related dissociation constant  $K_d$  is shown in Table 1. In the following paragraphs, we will refer to  $pK_a$  as measure of affinity (see Methods section for further explanations). Therefore, more negative  $pK_a$  values correspond to lower  $K_d$  values and to higher affinities. Correlations of the calculated quantities with

Table 2. Approximated Average Stabilization Energies for *apo* and *holo* Systems<sup>a</sup>

variant	$K_d$ (nM)	$pK_a$	$E_{nb}^{apo}$ (kJ/mol)	$E_{nb}^{holo}$ (kJ/mol)	$\Delta E_{nb}$ (kJ/mol)
Y0244-1	9.00	0.95	-25,296 (4413)	-30,695 (4699)	-3318
Y0244-4	3.30	0.52	-29,591 (6702)	-34,879 (5004)	-2496
Y0192	2.90	0.46	-26,269 (6826)	-32,434 (5522)	-2597
Y0242-1	2.30	0.36	-27,334 (4470)	-30,482 (5335)	-5621
Y0192 (T28D)	2.00	0.30	-28,708 (5899)	-28,754 (7537)	-5610
Y0192 (S100aT)	1.50	0.18	-30,479 (4984)	-28,058 (8136)	-4728
Y0238-5	1.40	0.15	-18,034 (9318)	-28,632 (8010)	-7329
Y0243-1	0.90	-0.05	-26,479 (3700)	-30,842 (6553)	-3922
Y0192 (N31H)	0.80	-0.10	-28,171 (7568)	-27,389 (3431)	-3717
Y0238-10	0.80	-0.10	-28,928 (5316)	-30,818 (4950)	-5953
Y0238-7	0.50	-0.30	-22,703 (1562)	-26,258 (5077)	-7544
Y0238-1	0.40	-0.40	-26,752 (6311)	-26,355 (7783)	-7236
Y0268-1	0.38	-0.42	-30,231 (4221)	-26,745 (8108)	-6995
Y0238-3	0.30	-0.52	-22,975 (7130)	-30,507 (3073)	-8308
Y0192 (H97Y)	0.20	-0.70	-26,487 (5728)	-29,217 (4623)	-5891
Y313-1	0.15	-0.82	-23,005 (6931)	-23,678 (8172)	-9450
Y0317 <sup>b</sup>	0.14	-0.85	-29,400 (5597)	-16,764 (4450)	-5116
Pearson's $r$			-0.08	-0.68	0.67
$p$ value			0.76	$\leq 0.01$	$\leq 0.01$

<sup>a</sup>Standard deviation are presented in brackets. The interaction energy  $\Delta E_{nb}$  was obtained as  $E_{nb}^{complex} - (E_{nb}^{antibody} + E_{nb}^{antigen})$ . The terms adopted for the calculation of  $\Delta E_{nb}$  have been calculated from the representative structure of the most populated cluster of each systems (specific data are provided in Table S4, Supporting Information). Correlations and related  $p$  values are calculated between  $E_{nb}$  (or  $\Delta E_{nb}$ ) and  $pK_a$  data points.

<sup>b</sup>Systems for Y0317 variant are obtained from the pdb file 1CZ8.

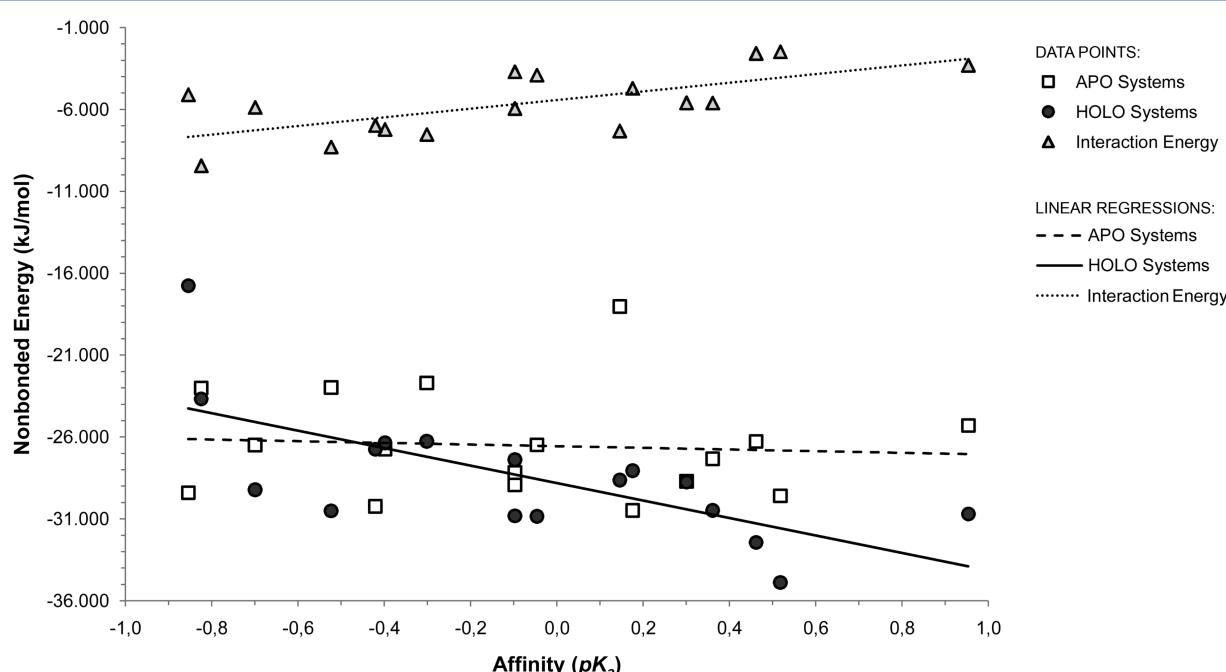


Figure 1. Scatter plot of interaction energies ( $\Delta E_{nb}$ , data points in triangles) and approximated nonbonded energies ( $E_{nb}$ ) for the antibody molecule in *holo* (filled circles) and *apo* systems (empty squares) of each variant. Continuous and dotted lines show the linear regressions.

the raw  $K_d$  values are reported in Figure S2 of the Supporting Information.

The possible correlations between the main coupling energetic interactions in the Fab structure for each complex and the experimental affinity for the Ag (VEGF-A) were thus investigated. The principal energetic couplings were obtained from MD simulations applying the energy decomposition method (EDM), described in ref 33. The data were then used to define an approximate stabilization energy ( $E_{nb}$ , see eq 4) for the Fabs either in complex with the Ag (*holo* state) and in the

free form (*apo* state). The approximation has been shown to be effective in several instances. From the comparison with the complete nonbonded energy, the approximated energy term recapitulates the most relevant information on nonbonded energy for small proteins<sup>10</sup> as well as large and/or multichain proteins.<sup>8,9</sup>

In this part of the present work, our aim is to analyze whether the redistribution of the stabilization energy among antibody residues determined by the presence (or absence) of the antigen can be correlated to modulation of the Ab affinity

Table 3. Average Distance Fluctuation Values for the Individual Domains in *holo* Systems<sup>a</sup>

variant	$K_d$ (nM)	p $K_a$	V <sub>L</sub>	V <sub>H</sub>	C <sub>L</sub>	C <sub>HI</sub>
Y0244-1	9.00	0.95	42.05 (9.19)	31.63 (6.45)	42.56 (10.21)	34.28 (6.34)
Y0244-4	3.30	0.52	36.01 (9.63)	33.32 (7.15)	42.15 (9.35)	34.53 (6.29)
Y0192	2.90	0.46	37.38 (7.40)	36.91 (6.96)	44.37 (8.83)	31.74 (7.75)
Y0242-1	2.30	0.36	38.17 (9.01)	40.33 (7.33)	48.32 (9.41)	36.68 (9.70)
Y0192 (T28D)	2.00	0.30	43.81 (8.75)	40.25 (7.27)	42.66 (9.68)	36.68 (8.82)
Y0192 (S100aT)	1.50	0.18	46.21 (9.70)	37.63 (10.01)	45.61 (10.94)	38.03 (7.10)
Y0238-5	1.40	0.15	46.50 (9.36)	45.96 (11.76)	48.49 (9.24)	39.19 (7.69)
Y0243-1	0.90	-0.05	43.91 (8.89)	44.53 (10.33)	47.88 (10.21)	38.55 (8.80)
Y0192 (N31H)	0.80	-0.10	44.81 (9.87)	40.42 (7.92)	47.40 (10.18)	33.35 (8.93)
Y0238-10	0.80	-0.10	43.73 (9.21)	42.13 (8.32)	49.94 (9.28)	33.49 (9.26)
Y0238-7	0.50	-0.30	38.74 (9.58)	36.33 (7.38)	44.17 (10.60)	34.83 (6.92)
Y0238-1	0.40	-0.40	51.67 (9.84)	40.74 (9.01)	53.87 (10.52)	38.41 (8.99)
Y0268-1	0.38	-0.42	46.88 (9.53)	39.77 (9.86)	50.60 (12.31)	36.83 (8.05)
Y0238-3	0.30	-0.52	42.31 (9.38)	37.14 (7.82)	46.38 (7.50)	35.30 (7.40)
Y0192 (H97Y)	0.20	-0.70	48.04 (9.96)	39.30 (6.77)	49.08 (9.61)	38.52 (7.69)
Y313-1	0.15	-0.82	48.03 (8.05)	40.98 (8.03)	51.28 (10.50)	38.01 (8.43)
Y0317 <sup>b</sup>	0.14	-0.85	49.25 (7.45)	44.75 (9.56)	49.10 (8.93)	39.56 (6.86)
Pearson's <i>r</i>			-0.63	-0.49	-0.68	-0.49
<i>p</i> value			≤0.01	≤0.05	≤0.01	≤0.05

<sup>a</sup>Standard deviation are presented in brackets. Correlations and related *p* values are calculated between distance fluctuation and p $K_a$  data points. <sup>b</sup>the systems for Y0317 variant are obtained from the pdb file 1CZ8.

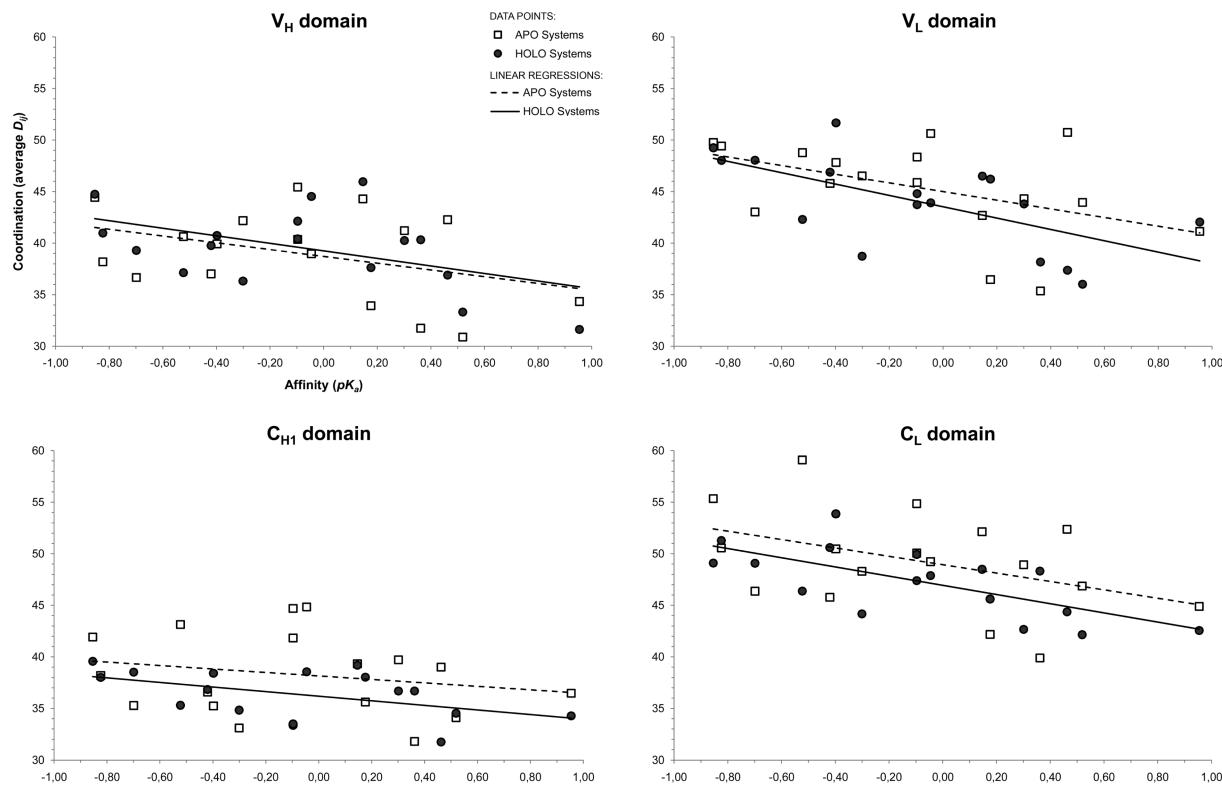
for the Ag. Approximated stabilization energy values are listed in Table 2 for both *apo* and *holo* systems of each variant. In this work, we have calculated the average energy value from 10 snapshot randomly selected from the cluster that hosts the representative structure of the most populated conformer (see Methods section for a more detailed discussion on cluster analysis method here adopted). It is worth noting that the stabilization energy of the same Fab variant in the free ( $E_{nb}^{apo}$ ) and in the bound form ( $E_{nb}^{holo}$ ) cannot be directly compared.  $E_{nb}^{holo}$  takes into account the energy coupling of the only Fab portion of the complex and is calculated in the presence of the antigen. As such, it reports on the organization/distribution of the Fab internal energetics under the influence of the antigen.  $E_{nb}^{apo}$ , on the other hand, analyzes these aspects for the different Fab mutants in isolation in solution. The overall goal of such analyses is to reveal the existence of a possible quantitative trend between p $K_a$  and  $E_{nb}$  values of the Fab in free or bound form separately in order to assess whether  $E_{nb}$  can be adopted as a simple energetic parameter accounting for the affinity of the matured forms of the antibody. In Figure 1, the linear regression between stabilization energy  $E_{nb}$  and p $K_a$  measures is plotted (also linear regression between  $E_{nb}$  and  $K_d$  values are provided in Table S1 and Figure S2, Supporting Information). The *holo* systems show a relevant (considering the simplified parameter) correlation ( $r = -0.68$ ), whereas no correlation can be found for *apo* systems ( $r = -0.08$ ). For what concerns the *holo* systems data points, Tukey mean difference test identifies the Y0317 variant as the only outlier (it falls over 1.96 standard deviations away the average difference). In contrast, the calculation of  $E_{nb}$  values for the Abs in free form (*apo* systems) does not yield any significant correlation. These data suggest that most of the relevant energetic determinants for Ag binding can be inferred from the analysis of the bound conformers (*holo* systems), but it is difficult to predict them from the analysis of the isolated Abs (free forms, *apo* systems). The availability of a structural model of an Ab::Ag complex can then be exploited to computationally evaluate/predict and

possibly design the effect of mutations of the relative affinities of homologous Abs.

The simplified energies calculated above are an approximation of the Fab stabilization in which the effect of the interaction with the antigen is only indirectly evaluated in the case of  $E_{nb}^{holo}$ . The observed negative correlation between p $K_a$  and  $E_{nb}^{holo}$  indicates that mutations may induce a partial destabilization of the Fab. In order to obtain productive and stable binding, this destabilization should be compensated by the interaction with the antigen. To quantitatively address this question, we analyzed the interaction energy  $\Delta E_{nb}$  between Fab and Ag through the calculation of the approximated stabilization energies of the representative structures obtained from the MD simulation of each mutant complex. The interaction energy has been calculated as the difference between the stabilization energy of the whole complex and the sum of the stabilization energies of the Fab and Ag molecules in free form, that is,  $\Delta E_{nb} = E_{nb}^{\text{complex}} - (E_{nb}^{\text{antibody}} + E_{nb}^{\text{antigen}})$ . The values for each system of the data set are provided in Table 2 and Table S4 of the Supporting Information. A positive correlation between  $\Delta E_{nb}$  and p $K_a$  is found ( $r = 0.67$ , Figure 1), actually indicating that a gain in stabilization of the entire complex is obtained through antigen binding (more negative  $\Delta E_{nb}$ ) and is related to the improvement of experimental affinity (more negative p $K_a$ ).

Besides energetic considerations, the selection of a specific Ab conformation apt to optimally recognize an Ag may be determined by the internal dynamics of the receptor systems and by the overall dynamics-related properties of the resulting complexes.

The former aspect is investigated through the analysis of mechanical inter-residue coordination. The distance fluctuation  $D_{ij}$  measurement reports the magnitude of variation from an averaged C $\alpha$ –C $\alpha$  distance between any two amino acids. Accordingly, pairs of distant residues showing high  $D_{ij}$  values move in a coordinated way, in spite of their physical separation. Localized ensembles of such pairs identify the regions whose coordinated motions may determine the overall dynamics of



**Figure 2.** Scatter plot of average distance fluctuations and  $pK_a$  measured for each Ig domain of each variant. Data points are depicted in filled circles or empty squares according to *holo* and *apo* systems, respectively. Continuous and dotted lines show the linear regressions. The correlation for  $C_L$  domains in *holo* systems can be regarded as significant ( $r = -0.68$ ).

the Fab.<sup>8</sup> On the other hand, the average  $D_{ij}$  value for a specific subset of localized residues or residues belonging to one domain can explain the general flexibility properties of that single domain compared to the overall structure of a protein. In Table S6 of the Supporting Information, the average distance fluctuation values are calculated for the whole Fab protein and for the light and heavy chains. Similarly to what previously observed for energetic values, distance fluctuation values do not correlate with  $pK_a$  values in *apo* systems. In the case of the *holo* systems, weak correlations are found, especially for the light chain ( $r = -0.46$ ). In light of such results, we pursued the measurement of average distance fluctuation by focusing on the individual domains. The average distance fluctuation values for *holo* system are shown in Table 3 (the same data for *apo* systems are provided in Table S7, Supporting Information). As depicted by the scatter plot in Figure 2, both the Ig domains composing the Fab molecule show a general trend in which an improvement in affinity (decreasing  $pK_a$ ) of a mutational variant is followed by an increasing coordination of the molecular architecture. The Pearson's correlation values are fair for the specific Ig domains ( $r_{VL} = -0.63$ ;  $r_{VH} = -0.49$ ; and  $r_{CH1} = -0.49$ ), except for the  $C_L$  domains showing a good correlation ( $r_{CL} = -0.68$ ) (linear regression between distance fluctuation and  $K_d$  values are provided in Table S1 and Figure S2, Supporting Information).

Increased internal coordination should not always be considered as an indicator of increased overall rigidity or of conformational entropy loss. Indeed, a large biomolecule (such as an antibody) may be characterized by parts of the protein showing high internal pair coordination (defining domains and subdomains) that can move independently of each other thus

giving the molecule a high overall conformational freedom, reflected in conformational entropy.

In order to obtain an independent evaluation of the relationships between the mutation-induced modulation of flexibility and changes in  $pK_a$ , we evaluated global dynamic parameters in the different complexes (*holo* systems). First, we evaluated the global flexibility of the antibodies according to the procedure described in ref 14. In this case, the pairwise covariance analysis of the atomic positional deviations is calculated for the trajectory of each complex and then simplified through principal component analysis or essential dynamics (ED) analysis. Such analysis allows extracting significant large-scale components and small amplitude vibrations from the MD sampling. In this context, we first calculated the eigenvectors recapitulating 90% of the total variance for each system. Next, we projected each trajectory on the subspace defined by the respective essential eigenvectors and calculated the residue based Root Mean Square fluctuations. The parameter representing global protein flexibilities for each *holo* system was then calculated by simply summing up the RMSF values. Calculation of the correlation between affinity and the global flexibility showed a rather poor correlation,  $r = -0.47$  (Table S5, Supporting Information). Analysis of the RMSF sums evaluated for the single Ig domains in each mutant shows that flexibility is unevenly distributed throughout the protein, as already observed with the analysis of dynamic pair-coordination, with highly coordinated groups of residues making some regions of the Fab more rigid than others (Table S5, Supporting Information).

To analyze the entropy-related aspects of Ab::Ag formation in a quantitative way, we calculated conformational entropies from the covariance matrices of the different mutants using the

Schlitter's approximation.<sup>34</sup> The values reported in Table 4 indicate a negative correlation between  $pK_a$  and the conforma-

**Table 4. Conformational Entropy of the Fab Proteins in *holo* Systems, Calculated According to Schlitter's Method<sup>34</sup>**

variant	$pK_a$	entropy (J/mol K)
Y0244-1	0.95	6405.52
Y0244-4	0.52	6674.21
Y0192	0.46	6549.78
Y0242-1	0.36	6736.21
Y0192 (T28D)	0.30	6611.38
Y0192 (S100aT)	0.18	6610.21
Y0238-5	0.15	6606.89
Y0243-1	-0.05	6837.33
Y0192 (N31H)	-0.10	6771.32
Y0238-10	-0.10	6819.07
Y0238-7	-0.30	6744.25
Y0238-1	-0.40	6814.13
Y0268-1	-0.42	6746.04
Y0238-3	-0.52	6848.51
Y0192 (H97Y)	-0.70	6729.80
Y313-1	-0.82	6618.82
Y0317 <sup>a</sup>	-0.85	6893.47
Pearson's $r$		-0.67
$p$ value		$\leq 0.01$

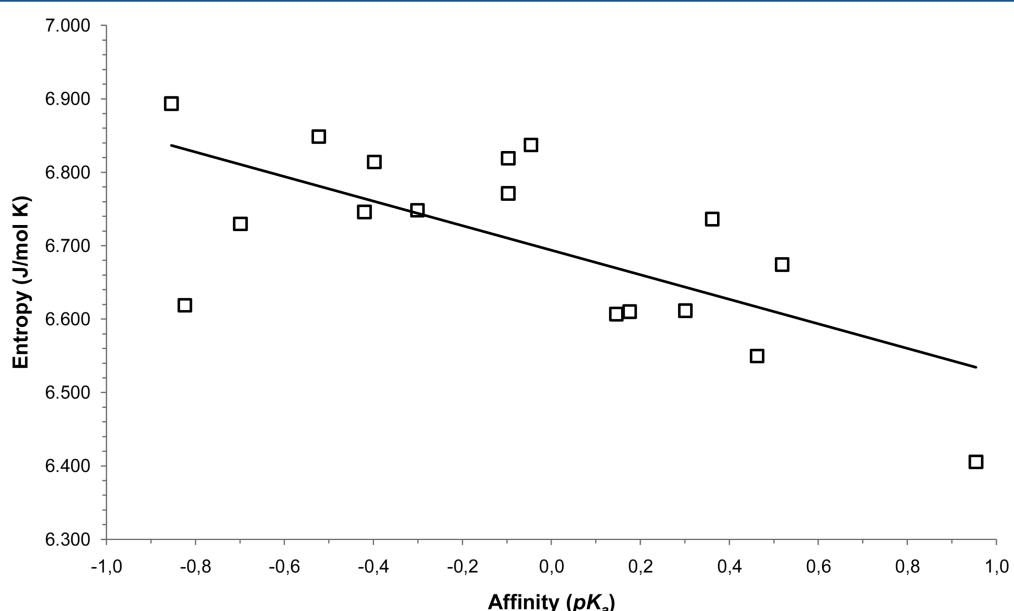
<sup>a</sup>Systems for Y0317 variant are obtained from the pdb file 1CZ8.

tional entropies,  $r = -0.67$ , suggesting that conformational entropy changes do not oppose improved complex formation upon mutation of the Ab (Figure 3).

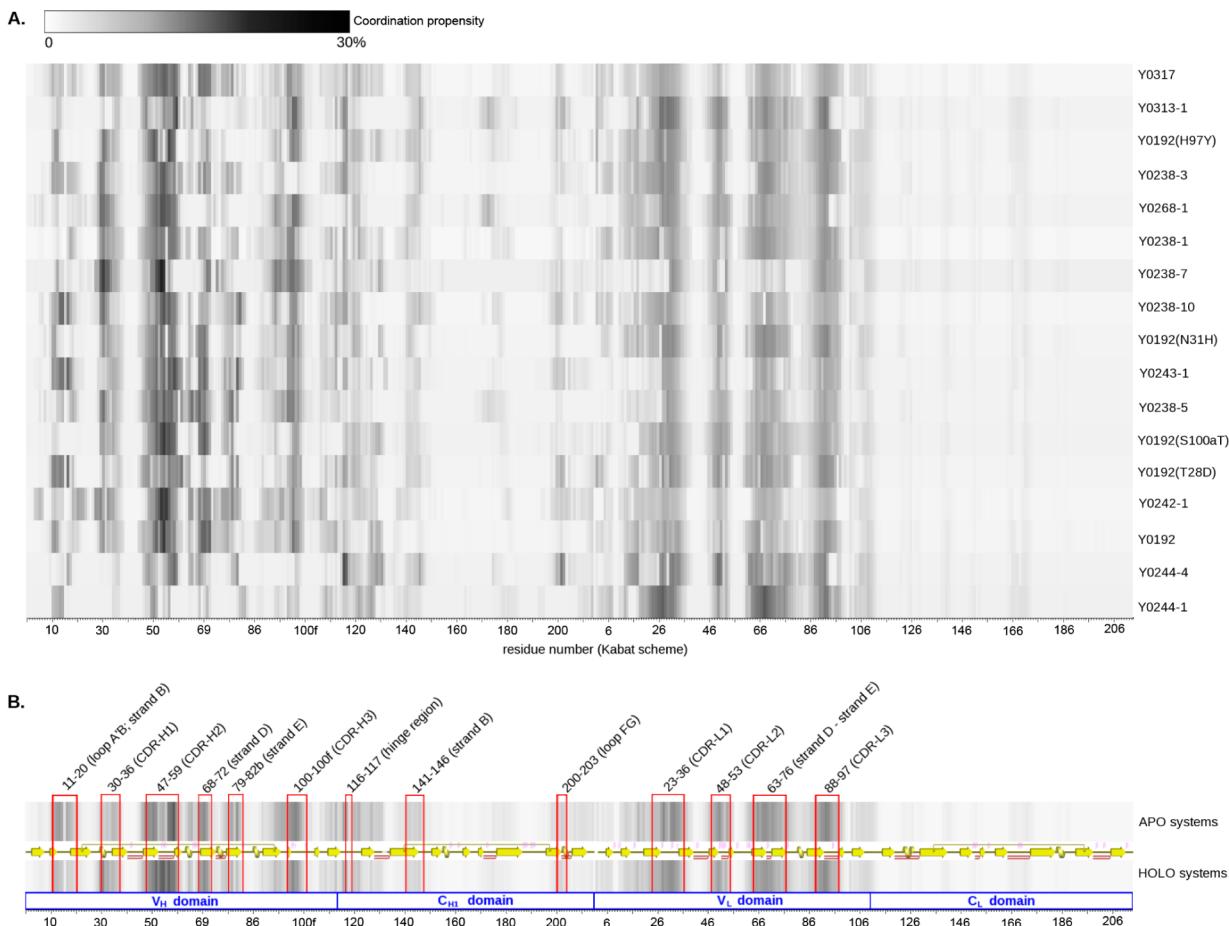
In our previous work, singular distance fluctuation values were collected from several systems of different Ab:Ag complexes with the aim of highlighting common dynamical hot spots referred as coordinated motion recurrent positions (CMRPs).<sup>8</sup> In this work, we aim to examine how long-range interdomain dynamical coordination patterns respond to mutations. We have calculated for each residue the amount

of all other protein residues with which it forms a coordinated pair, and we derived a measure of residue-based coordination propensity (CP). The average CP data are available in Table S8 of the Supporting Information; the values range from 0 to 26.98%, with means of 5.44% and 5.12% for *apo* and *holo* systems, respectively. This preliminary result suggests that only a restricted group of residues may be regarded as long-range coordinated with other residues across the whole Fab protein. In Figure 4A, the CP profile along the primary sequence is shown for the different variants in *holo* systems (CP profiles in *apo* system are supplied in Figure S3, Supporting Information). The darker spots identify clusters of residues that exhibit higher CP values (over 10%) and which then participate in long-range coordinated motions. In Figure 4B, the averaged CP profiles between *apo* and *holo* systems are compared. For both cases, we found 13 clusters, primarily located on the variable domains. A few interesting differences emerge between the two profiles, and they comprise the regions CDR-H2 and CDR-H3 for the heavy chain and CDR-L1, CDR-L2, LFR3, and CDR-L3 for the light chain near the Ag binding site and its surroundings.<sup>18</sup> A more convenient representation of how residues with high CP values are distributed along the protein structure is depicted in Figure 5. The binding site is composed of a shallow cleft where the hypervariable regions CDR-H3, CDR-L3 and CDR-H1, CDR-H2 constitute the opposite walls (green surface in Figure 5A). The cleft is narrow, and this constitutes a critical feature determining the high specificity of the Ab toward human VEGF-A. Indeed, site-directed mutagenesis demonstrated that Gly-88 of hVEGF-A is tightly packed deeply in the cleft such that there is not enough space for larger amino acids (such as the Ser residue mapped at equivalent position in the murine VEGF-A).<sup>18,35</sup> The mutations introduced by affinity maturation procedure (Table 1) do not impair the binding site/cleft structure and topology.<sup>17</sup>

Analysis of the coordination propensity profiles and their projection on the structures can be used to illustrate this concept. In the case of variable domains (Figure 5A), the CP profiles are in agreement with crystallographic data on the Fab-



**Figure 3.** Scatter plot of the conformational entropy of the Fab proteins in *holo* systems and  $pK_a$ . Entropy values have been calculated according to the Schlitter's method.<sup>34</sup> The conformational entropy of the antibody molecule increases as affinity for the antigen improves ( $r = -0.67$ ).



**Figure 4.** Coordination propensity profiles. Color ramp is scaled to the range of CP values found. (A) CP profiles for each variant of the *holo* systems. Variants show somewhat heterogeneous distribution of CP values along the primary sequence, with a restricted group of commonly shared bands. (B) Averaged CP profiles of *apo* and *holo* systems. CP profiles are really similar, and few differences can be identified looking at the CP values distribution. Differences in CP magnitude appear confined to the hypervariable regions CDR-H2, CDR-H3, CDR-L1, CDR-L2, and CDR-L3 and the framework region LFR3.

12 molecule, in which the lowest temperature factors (generally related to low flexibility of the residues) are located at the VEGF-A::Fab interface and propagate throughout the hypervariable regions.<sup>18</sup> Interestingly, the authors of the crystallographic structure suggest that the formation of the complex tends to stabilize the binding site and the consequent ordering of residues and interactions extend from the antigen molecule to the variable domain of the Fab.

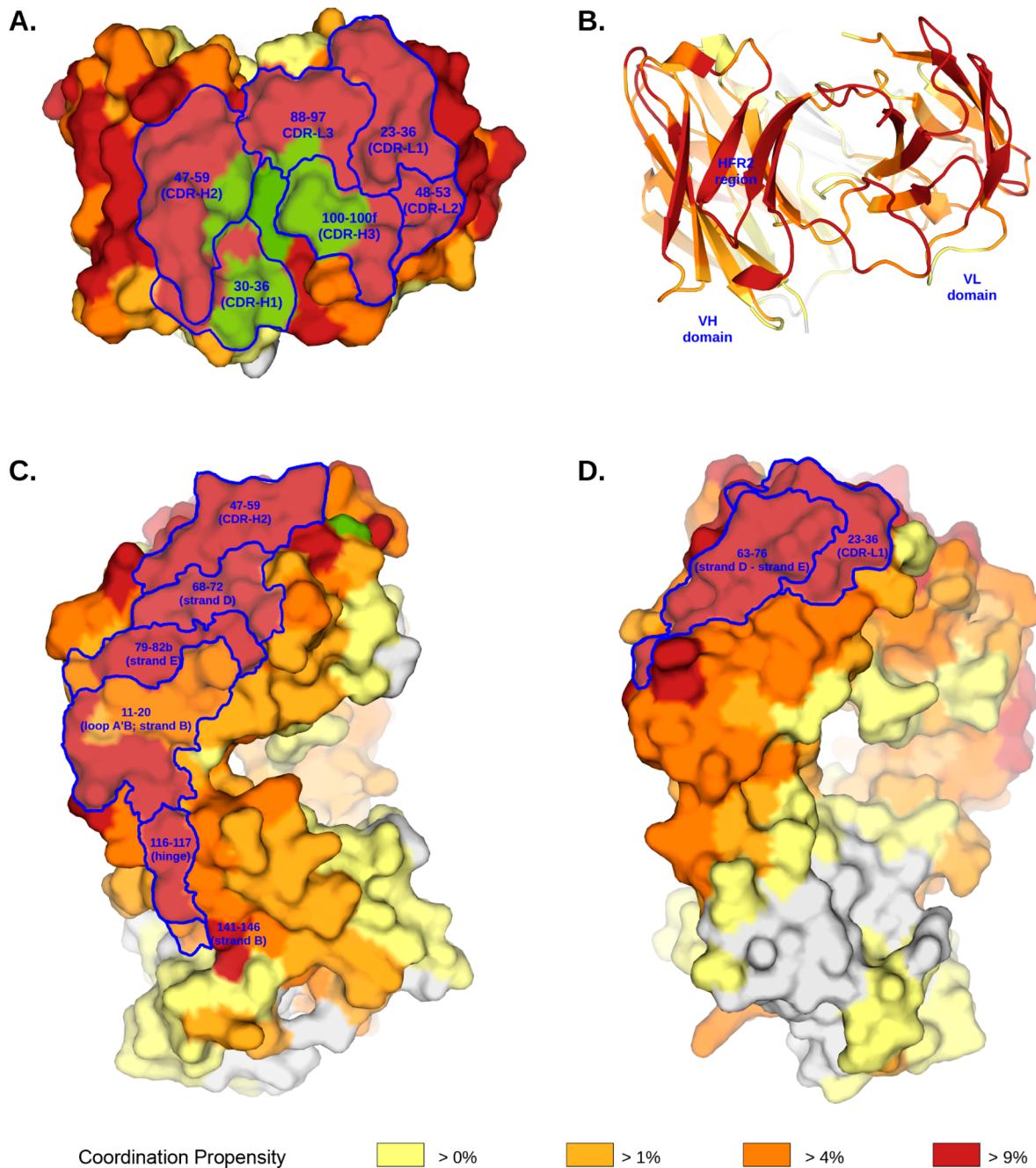
While the residues of the binding surface are all characterized by high CP values, the buried residues of the variable region show a diverse behavior. Figure 5B shows that high CP values characterize the internal sheets of the  $V_H$  domains (HFR2 region), while for the  $V_L$  domain, high CP values are limited to the surface. Strands C and C' (residues 36–49 of the heavy chain) are located immediately underneath the Ag binding, so that ligand binding can be immediately communicated to the HFR2 region. On the surface side, residues with high CP span the hypervariable CDR regions and extend along specific secondary structure elements that define the framework regions HFR3 and LFR3 (strands D and E of domains  $V_H$  and  $V_L$ ).

The pathway of residues with high CPs for the heavy chain continues from the variable domain  $V_H$  through the constant domain  $C_{HL}$ , including the solvent exposed region (Figure 5C). This region, mainly defined by strands F and G, was found to be critical for the signal transduction of the Fab molecules

because perturbations in such region may heavily affect the antigen binding performance.<sup>36,37</sup> The stretch of residues with efficient CP along the light chain stops near the hinge region (Figure 5D), defined by residues 107–112, 138–141, and 166–172 of the light chain. Strikingly, the  $C_L$  domain emerges as the most coordinated domain of the Fab molecule (Table 3) while hosting very low or zero CP values, as depicted by yellow and white surfaces at the bottom side of Figure 5D. These observations, combined with the results of the comparative study of 28 different Ab::Ag complexes (with nonhomologous Abs) we previously presented,<sup>8</sup> support a scaffolding function for the  $C_L$  domain, as also suggested by Feige and co-workers.<sup>38</sup>

## ■ DISCUSSION

Investigating the molecular origins of the mutation-dependent optimization of Ab affinity for an Ag is an important issue for fundamental and practical reasons. From the fundamental point of view, the comparison between sequence variants of Abs targeting the same molecule can help in understanding the molecular determinants of ligand recognition and binding in a key immunological process. From the practical point of view, this could be translated in molecular design rules for the development of new Abs with affinity characteristics optimized for a particular application (e.g., in the field of immunodiagnostics or of therapeutic antibodies).



**Figure 5.** Coordination propensity distribution along the Fab protein. Location of the hallmarks isolated from the CP profiles (Figure 4) have been shaded. (A) Antigen binding site view. Ag footprint has been highlighted by green surface. It is a shallow cleft defined by the opposite walls composed of CDR-H3, CDR-L3 and CDR-H1, CDR-H2 regions. (B) Antigen binding site view. Cartoon representation. Such a view distinguishes the two variable domains  $V_H$  and  $V_L$ , in which the whole internal sheet of  $V_H$  domain shows higher values of coordination propensity (i.e., strands are fully colored in red). (C) Heavy chain side. Both variable  $V_L$  and constant  $C_{H1}$  domains seem connected by a CP pathway that spreads from the N-terminal side (top) to the C-terminal side (bottom). (D) Light chain side. Constant domain  $C_L$  seems not affected by coordination propensity (bottom part of figure). Structural core motif of the immunoglobulin fold shows null CP values (regions colored in white) irrespective of the higher internal coordinated motions revealed in Table 3.

In nature, optimized Ab variants can be generated through affinity maturation.<sup>3–7</sup> *In vitro*, this process has been mimicked by phage display and selection techniques. In this paper, we have addressed the problem of describing the sequence-dependent modulation of Ab affinity for a specific protein Ag via a computational approach, based on all-atom MD simulations. In particular, we have used recently developed approaches for the analysis of protein energetics and internal dynamics and coordination<sup>8–16</sup> to get a molecular picture of

the differential binding properties of homologous Abs. One of the aims was to generate simple atomic-resolution “empirical” descriptors that can complement experimental approaches in the rational design of Abs. To this end, we have considered a data set of 17 mutational variants of the Fab of an Ab elicited against VEGF-A. Variants Abs of improved affinity were experimentally selected by phage display and off-rate selection.<sup>17</sup>

First, we examined possible correlations between the energy accounted for by the pair-interaction networks most responsible for the stabilization of the Ab and the experimentally determined activities. The calculation was carried out for the *apo* and *holo* forms of the Abs, through eigenvalue decomposition of the matrix of nonbonded interactions, using the energy decomposition method.<sup>8,9</sup> Figure 1 shows that a significant (anti)correlation can be obtained only when considering the variation in stability of the Ab structure in the *holo* forms (around  $-0.7$ ). These results suggest that affinity-increasing mutations may act by inducing a partial destabilization of the Ab structure in the bound state. Such destabilization can be exploited in increasing the probabilities of visiting conformational states that are still close to the native/active one but that are structurally more fit to recognize and bind the Ag.<sup>39,40</sup> This is qualitatively consistent with experimental results by Adhikary and co-workers,<sup>41</sup> who elicited 8 Ab variants against the Ag 8-me-thoxypyrene-1,3,6-trisulfonate (MPTS) and evaluated the correlations between measured  $pK_a$  and  $\Delta H^0$  via isothermal titration calorimetry. Interestingly, they found correlations similar (Pearson's  $r = -0.67$ ) to the ones we observe here (Table 2 and Figure 1). In our case, the partial destabilization observed for the antibody is compensated by the interaction with the antigen molecule. Indeed, we observe that the stabilization energy of the whole complex is more negative than the stabilization energy of the Ab plus Ag molecules in free form, in all cases (Table 2 and Table S4, Supporting Information). Interestingly, the interaction energies  $\Delta E_{nb}$  show a good positive correlation with  $pK_a$  ( $r = 0.67$ ), indicating that the more stable is the complex, the better the affinity. This may be seen as an expected result, but to our knowledge, this is the first case in which this is quantitatively calculated for such a complex and large system using a large data set and with a simple strategy amenable to medium-scale applications.

The effects of affinity-modulating mutations may also reverberate on the conformational dynamics of the Ab. We analyzed these aspects in terms of global flexibility and internal coordination of the Abs in the *apo* and *holo* states. The distance fluctuation matrix for each mutant was calculated and subjected to eigenvalue decomposition, according to a procedure previously used investigating how affinity is related to allosteric changes in zinc-finger proteins.<sup>14</sup> Interestingly, the affinity maturation process is accompanied by an increase in coordination among the individual Ig domains and long-range coordination effects that extend far from the binding site. In general, different Ig domains show different pair-fluctuation profiles (Figure 2), reinforcing the hypothesis that they may have different functional roles.<sup>42</sup> Structural adaptation occurs in the CDR regions (hosting the Ag binding site) and the dynamic consequence of Ag binding at this region is accompanied by the rigidification of remote regions such as, for example, the constant Ig domains of the Fab.<sup>43,44</sup> Such behavior, illustrated by the correlations reported in Figure 2 qualitatively agrees with experimental data obtained through three-pulse photon echo peak shift (3PEPS) spectroscopy.<sup>41,45</sup>

Long-range coordinated regions have been detected among different domains of the Fab protein (Figure 4), but this phenomenon is not uniformly diffused. Only a limited number of strong interaction centers (with CP values over 10%) are distributed along an apparently uncoordinated protein structure (average CP value of about 5%). As a result, the emergence of coordination patterns among different substructures does not

necessarily translate into an overall rigidification of the whole antibody protein with a consequent destabilizing entropic effect. Indeed the quantitative evaluation of the conformational entropy with the Schlitter's approach shows an increase that parallels affinity improvements (Table 4 and Figure 3).

In this picture, actual rigidification can in fact be limited to specific parts of the antibody endowed with specific Ag-recognition functions. In a previous study, Thorpe and co-workers considered a set of three matured Abs plus the germline one<sup>46</sup> binding to a small molecule antigen and observed a decrease of RMSF of the binding site residues as the affinity of the variants improves. Repeating this calculation on our data set of 17 complexes, by considering the RMSF of residues that are located within 6 Å from the ones defining the Ag footprint, we observe similar results. A modest positive correlation ( $r = 0.54$ ) between the flexibility parameter was obtained by summing the RMSF values of the binding site residues and the  $pK_a$ , indicating that rigidification of the binding site accompanies affinity maturation. The local trend is opposite to the observed global increase of general flexibility measured for the whole Ig domains (Table S5, Supporting Information). In this context, the data and characterizations we presented in this study help identify the network of coordinated residues that extends from the binding site to remote regions of the Ab and may help preorganize the binding site for efficient Ag recognition. Similar phenomena have been experimentally observed also in other antibodies through NMR-based amide hydrogen exchange techniques.<sup>47,48</sup>

In the biological context, the clonal selection hypothesis offers a widely accepted model to explain how the molecular recognition mechanism of antibodies evolves from germline Abs that are polyspecific into more Abs that more specifically recognize their Ags.<sup>49</sup> In this framework, the model emerging from our energetics and internal dynamics analyses suggests a mechanism, whereby mutations determine a partial destabilization of the Ab, which may allow it to efficiently sample closely related conformations from which the one with the best fit to the Ag can be selected. From the energetic point of view, stabilization is then achieved through the interaction with the antigen. In parallel, the dynamics of the different antibody mutants adapt to the presence of the antigen through differential modulations of internal coordination patterns and flexibility. Once the optimal Ab structure is selected and bound to the Ag, the Ab binding site is locked in the correct conformation, and the molecular signal can be transmitted through the Ab structure for subsequent immunological functions. It is important to underline, at this point, that what we propose here is a qualitative and partial picture of the molecular mechanisms of Ab:Ag binding. Indeed, effects such as solvent displacements or the quantitative evaluation of entropic effects related to side chain reorganization are not included in our simplified model and out of the reach of the present investigation. Therefore, how and to what extent the conformational landscape of the Abs is remodeled by the mutations and their influence on stability and protein dynamics cannot be quantitatively determined. To achieve such an ambitious goal through computational approaches, methods that allow accurate profiling of the free energy of the Ab:Ag complex formation reaction should be used. These may entail the use of enhanced sampling methods ranging from accelerated MD<sup>50</sup> to metadynamics,<sup>51</sup> which would require the definition of a suitable set of collective variables that describe the process.

A possible experimental approach to accomplish such task is based on three-pulse photon echo peak shift (3PEPS) spectroscopy, taking advantage from the availability of Abs that bind chromophoric antigens. In different papers, Romeseberg and colleagues show how this technique can offer a quantitative measure of the conformational heterogeneity.<sup>41,45,52</sup> Interestingly, the authors have tested the hypothesis that affinity maturation modulates Ab dynamics and flexibility in order to obtain a limited number of closely related conformations from which the one with the best fit to the Ag can be selected.<sup>52</sup> Other possible experimental-structural approaches that may be used for analyzing differential conformational dynamics could make use of novel techniques such as double electron-electron resonance (DEER). The latter offers the possibility to measure distances between two paramagnetic centers at a range of 18–80 Å. Rather than calculating an average distance, DEER measurements produce a distance distribution that allows for the measurement of simultaneous states within a bulk sample.<sup>53</sup>

Finally, two more points deserve consideration. First, the present study further supports the hypothesis that also residues that do not interact directly with the ligand may have relevant functional roles in Ag binding. Indeed, the variants used in our study in which mutations are located on HFR3 (Y0244-1 and Y0244-4) show affinities that are worse than the *wildtype* variant Y0192 (Table 1). Second, the 3D distribution of CP values mapped over the Ab structure clearly shows that the network of coordinated motions across distinct Ig domains is asymmetric and principally involves the heavy chain, while the light chain does not show relevant CP values along the constant domain C<sub>L</sub>. Such observations are consistent with the suggestions proposed in our previous work, in which the two chains composing the Ab structure share a high similar 3D structure but carry out functionally distinct roles.<sup>8</sup>

## CONCLUSIONS

In this work, we have investigated the variations of internal energetics and dynamics of the Ab structures that occur during the process of affinity maturation. In particular, we noticed a partial destabilization of the structure of the Ab in the bound state upon mutation, accompanied by diffuse rigidification of the structure. It is worth noting that even if most of the mutations considered affect the hypervariable region, they induce structural and dynamic rearrangements throughout the entire Fab, perturbing the light and heavy chains in a sizable different manner. These observations suggest the possibility that allosteric mechanisms may control the transmission of the signal encoded by the Ag to the effector domains of the Abs. A caveat must be added here because the effector part of the Ab molecule (Fc fragment) is not present in our simulations. Nevertheless, our data show interesting and nontrivial correlations with experimentally determined affinity data.

In an applicative perspective, one could evaluate the effects of mutations on the energetics of the bound Ag and on the internal dynamic coordination patterns using the approach we have presented here to rationally select initial Ab variants as starting leads for optimization of binding properties. This would allow focusing on specific regions of the Ab for modification, which could be experimentally targeted by saturation or random mutagenesis, followed by affinity screening. Different computational approaches to this goal have recently been proposed,<sup>54–56</sup> and could be integrated with or used to complement what we have presented here.

## ASSOCIATED CONTENT

### S Supporting Information

Figure S1: Root Mean Square Deviation (RMSD) profiles of the Fab fragments. Figure S2: Scatter plots of stabilization energy ( $E_{nb}$ ) or distance fluctuations versus experimental  $K_d$  values. Figure S3: Coordination propensity profiles of the *apo* systems. Table S1: Coefficient values according to Pearson's correlation and Spearman's rank correlation. Table S2: DOPE scores of the refined loop models generated by Modeller software. Table S3: Solvation box settings. Table S4: Approximated stabilization energies for the complex, antibody, and antigen. Table S5: Sum of RMSF values of the Fab proteins in *holo* systems. Table S6: Average distance fluctuation values for whole Fab molecules, heavy and light chains. Table S7: Average distance fluctuation values for the individual domains in *apo* systems. Table S8: Average coordination propensity values for *apo* and *holo* systems. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: giorgio.colombo@icrm.cnr.it. Phone: +39 02 28500031. Fax: +39 02 28901239.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by CARIPLO “From Genome to Antigen: A Multidisciplinary Approach Towards the Development of an Effective Vaccine Against *Burkholderia pseudomallei*, the Etiological Agent of Melioidosis” (contract number 2009-3577). Support was also received from AIRC (Associazione Italiana Ricerca sul Cancro), Grant IG.11775 to G.C. and by the Italian Ministry of Education and Research through the Flagship (PB05) “InterOmics”.

## REFERENCES

- (1) Schroeder, H. W.; Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **2010**, *125*, S41–S52.
- (2) Oda, M.; Uchiyama, S.; Robinson, C. V.; Fukui, K.; Kobayashi, Y.; Azuma, T. Regional and segmental flexibility of antibodies in interaction with antigens of different size. *FEBS J.* **2006**, *273*, 1476–1487.
- (3) Torres, M.; Casadevall, A. The immunoglobulin constant region contributes to affinity and specificity. *Trends Immunol.* **2008**, *29*, 91–97.
- (4) Tudor, D.; Yu, H.; Maupetit, J.; Drillet, A.-S.; Bouceba, T.; Schwartz-Cornil, I.; Lopalco, L.; Tuffery, P.; Bomsel, M. Isotype modulates epitope specificity, affinity, and antiviral activities of anti-HIV-1 human broadly neutralizing 2F5 antibody. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12680–12685.
- (5) Neuberger, M. S.; Ehrenstein, M. R.; Rada, C.; Sale, J.; Batista, F. D.; Williams, G.; Milstein, C. Memory in the B-cell compartment: antibody affinity maturation. *Philos. Trans. R. Soc., B* **2000**, *355*, 357–360.
- (6) Wedemayer, G. J.; Patten, P. A.; Wang, L. H.; Schultz, P. G.; Stevens, R. C. Structural insights into the evolution of an antibody combining site. *Science* **1997**, *276*, 1665–1669.
- (7) Furukawa, K.; Akasako-Furukawa, A.; Shirai, H.; Nakamura, H.; Azuma, T. Junctional amino acids determine the maturation pathway of an antibody. *Immunity* **1999**, *11*, 329–338.
- (8) Corrada, D.; Morra, G.; Colombo, G. Investigating allostery in molecular recognition: Insights from a computational study of multiple antibody-antigen complexes. *J. Phys. Chem. B* **2013**, *117*, 535–552.

- (9) Genoni, A.; Morra, G.; Colombo, G. Identification of domains in protein structures from the analysis of intramolecular interactions. *J. Phys. Chem. B* **2012**, *116*, 3331–3343.
- (10) Morra, G.; Colombo, G. Relationship between energy distribution and fold stability: Insights from molecular dynamics simulations of native and mutant proteins. *Proteins* **2008**, *72*, 660–672.
- (11) Morra, G.; Verkhivker, G.; Colombo, G. Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer. *PLoS Comput. Biol.* **2009**, *5*, e1000323.
- (12) Colacino, S.; Tiana, G.; Colombo, G. Similar folds with different stabilization mechanisms: The cases of prion and doppel proteins. *BMC Struct. Biol.* **2006**, *6*, 17.
- (13) Mollica, L.; Morra, G.; Colombo, G.; Musco, G. HMGB1-carbenoxolone interactions: dynamics insights from combined nuclear magnetic resonance and molecular dynamics. *Chem. Asian J.* **2011**, *6*, 1171–1180.
- (14) Torella, R.; Moroni, E.; Caselle, M.; Morra, G.; Colombo, G. Investigating dynamic and energetic determinants of protein nucleic acid recognition: analysis of the zinc finger zif268-DNA complexes. *BMC Struct. Biol.* **2010**, *10*, 42.
- (15) Morra, G.; Potestio, R.; Micheletti, C.; Colombo, G. Corresponding functional dynamics across the Hsp90 chaperone family: Insights from a multiscale analysis of MD simulations. *PLoS Comput. Biol.* **2012**, *8*, e1002433.
- (16) Ragona, L.; Colombo, G.; Catalano, M.; Molinari, H. Determinants of protein stability and folding: Comparative analysis of beta-lactoglobulins and liver basic fatty acid binding protein. *Proteins* **2005**, *61*, 366–376.
- (17) Chen, Y.; Wiesmann, C.; Fuh, G.; Li, B.; Christinger, H. W.; McKay, P.; De Vos, A. M.; Lowman, H. B. Selection and analysis of an optimized anti-VEGF antibody: Crystal structure of an affinity-matured Fab in complex with antigen. *J. Mol. Biol.* **1999**, *293*, 865–881.
- (18) Muller, Y. A.; Chen, Y.; Christinger, H. W.; Li, B.; Cunningham, B. C.; Lowman, H. B.; De Vos, A. M. VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 Å resolution and mutational analysis of the interface. *Structure* **1998**, *6*, 1153–1167.
- (19) Ferrara, N.; Hillan, K. J.; Gerber, H.-P.; Novotny, W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat. Rev. Drug Discovery* **2004**, *3*, 391–400.
- (20) Ferrara, N.; Henzel, W. J. Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. *Biochem. Biophys. Res. Commun.* **1989**, *161*, 851–858.
- (21) Ferrara, N. Vascular endothelial growth factor: Basic science and clinical progress. *Endocr. Rev.* **2004**, *25*, 581–611.
- (22) Ranieri, G.; Patruno, R.; Ruggieri, E.; Montemurro, S.; Valerio, P.; Ribatti, D. Vascular endothelial growth factor (VEGF) as a target of bevacizumab in cancer: From the biology to the clinic. *Curr. Med. Chem.* **2006**, *13*, 1845–1857.
- (23) Wiesmann, C.; Fuh, G.; Christinger, H. W.; Eigenbrot, C.; Wells, J. A.; De Vos, A. M. Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. *Cell* **1997**, *91*, 695–704.
- (24) Fiser, A.; Do, R. K.; Sali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773.
- (25) *The PyMOL Molecular Graphics System*, version 1.3r1; Schrödinger LLC: New York, 2010.
- (26) Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (27) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krieger, P.; Van Gunsteren, W. F. The GROMOS Biomolecular Simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- (28) Krause, E. F. *Taxicab Geometry*; Dover: New York, 1987.
- (29) Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65.
- (30) Kaufman, L.; Rousseeuw, P. *Finding Groups in Data*; John Wiley: New York, 1990.
- (31) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
- (32) Presta, L. G.; Chen, H.; O'Connor, S. J.; Chisholm, V.; Meng, Y. G.; Krummen, L.; Winkler, M.; Ferrara, N. Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res.* **1997**, *57*, 4593–4599.
- (33) Tiana, G.; Simona, F.; De Mori, G. M. S.; Broglia, R. A.; Colombo, G. Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Sci.* **2004**, *13*, 113–124.
- (34) Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *J. Chem. Phys. Lett.* **1993**, *215*, 617–621.
- (35) Gerber, H.-P.; Wu, X.; Yu, L.; Wiesmann, C.; Liang, X. H.; Lee, C. V.; Fuh, G.; Olsson, C.; Damico, L.; Xie, D.; Meng, Y. G.; Gutierrez, J.; Corpuz, R.; Li, B.; Hall, L.; Rangell, L.; Ferrando, R.; Lowman, H.; Peale, F.; Ferrara, N. Mice expressing a humanized form of VEGF-A may provide insights into the safety and efficacy of anti-VEGF antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 3478–3483.
- (36) Torres, M.; Fernández-Fuentes, N.; Fiser, A.; Casadevall, A. The immunoglobulin heavy chain constant region affects kinetic and thermodynamic parameters of antibody variable region interactions with antigen. *J. Biol. Chem.* **2007**, *282*, 13917–13927.
- (37) Pritsch, O.; Hudry-Clergeon, G.; Buckle, M.; Petillot, Y.; Bouvet, J. P.; Gagnon, J.; Dighiero, G. Can immunoglobulin C(H)1 constant region domain modulate antigen binding affinity of antibodies? *J. Clin. Invest.* **1996**, *98*, 2235–2243.
- (38) Feige, M. J.; Hendershot, L. M.; Buchner, J. How antibodies fold. *Trends Biochem. Sci.* **2010**, *35*, 189–198.
- (39) Fenley, A. T.; Muddana, H. S.; Gilson, M. K. Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20006–20011.
- (40) Kang, J.; Warren, A. S. Enthalpy–entropy compensation in the transition of a monospecific antibody towards antigen-binding promiscuity. *Mol. Immunol.* **2007**, *44*, 3623–3624.
- (41) Adhikary, R.; Yu, W.; Oda, M.; Zimmermann, J.; Romesberg, F. E. Protein dynamics and the diversity of an antibody response. *J. Biol. Chem.* **2012**, *287*, 27139–27147.
- (42) Feige, M. J.; Groscurth, S.; Marcinowski, M.; Shimizu, Y.; Kessler, H.; Hendershot, L. M.; Buchner, J. An unfolded CH1 domain controls the assembly and secretion of IgG antibodies. *Mol. Cell* **2009**, *34*, 569–579.
- (43) Pieckarska, B.; Drozd, A.; Konieczny, L.; Król, M.; Jurkowski, W.; Roterman, I.; Spólnik, P.; Stopa, B.; Rybarska, J. The indirect generation of long-distance structural changes in antibodies upon their binding to antigen. *Chem. Biol. Drug Des.* **2006**, *68*, 276–283.
- (44) Król, M.; Roterman, I.; Pieckarska, B.; Konieczny, L.; Rybarska, J.; Stopa, B.; Spólnik, P. Analysis of correlated domain motions in IgG light chain reveals possible mechanisms of immunological signal transduction. *Proteins* **2005**, *59*, 545–554.
- (45) Thielges, M. C.; Zimmermann, J.; Yu, W.; Oda, M.; Romesberg, F. E. Exploring the energy landscape of antibody–antigen complexes: protein dynamics, flexibility, and molecular recognition. *Biochemistry* **2008**, *47*, 7237–7247.
- (46) Thorpe, I. F.; Brooks, C. L. Molecular evolution of affinity and flexibility in the immune system. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8821–8826.
- (47) Takahashi, H.; Tamura, H.; Shimba, N.; Shimada, I.; Arata, Y. Role of the domain-domain interaction in the construction of the antigen combining site. A comparative study by 1H-15N shift

- correlation NMR spectroscopy of the Fv and Fab fragments of anti-dansyl mouse monoclonal antibody. *J. Mol. Biol.* **1994**, *243*, 494–503.
- (48) Williams, D. C.; Rule, G. S.; Poljak, R. J.; Benjamin, D. C. Reduction in the amide hydrogen exchange rates of an anti-lysozyme Fv fragment due to formation of the Fv-Lysozyme complex. *J. Mol. Biol.* **1997**, *270*, 751–762.
- (49) Rajewsky, K. Clonal selection and learning in the antibody system. *Nature* **1996**, *381*, 751–758.
- (50) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
- (51) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6358–6363.
- (52) Zimmerman, J.; Oakman, E. L.; Thorpe, I. F.; Shi, X.; Abbyad, P.; Brooks, C. L.; Boxer, S. G.; Romesberg, F. E. Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13722–13727.
- (53) Jeschke, G. DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* **2012**, *63*, 419–446.
- (54) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11937–11942.
- (55) Lippow, S. M.; Wittrup, K. D.; Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **2007**, *25*, 1171–1176.
- (56) Tharakaraman, K.; Robinson, L. N.; Hatas, A.; Chen, Y.-L.; Siyue, L.; Raguram, S.; Sasisekharan, V.; Wogan, G. N.; Sasisekharan, R. Redesign of a cross-reactive antibody to dengue virus with broad-spectrum activity and increased in vivo potency. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E1555–E1564.
- (57) Abhinandan, K. R.; Martin, A. C. R. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. *Mol. Immunol.* **2008**, *45*, 3832–3839.