THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

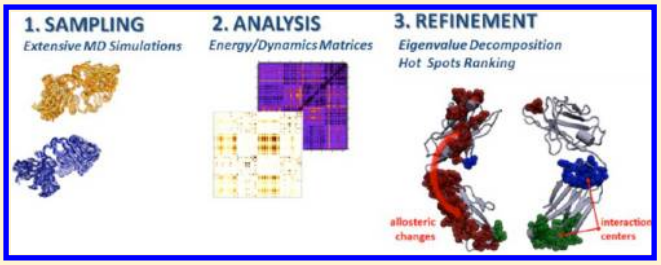
Investigating Allostery in Molecular Recognition: Insights from a Computational Study of Multiple Antibody−Antigen Complexes

Dario Corrada, Giulia Morra, 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−antigen recognition plays a key role in the immune response against pathogens. Here, we have investigated various aspects of this problem by analyzing a large and diverse set of antibodies and their respective complexes with protein antigens through atomistic simulations. Common features of antibody response to the presence of antigens are elucidated by the analysis of the proteins' internal dynamics and coordination in different ligand states, combined with the analysis of the interaction networks implicated in the stabilization of functional structures. The use of a common structural reference reveals preferential changes in the dynamic coordination and intramolecular interaction networks induced by antigen binding and shared by all antibodies. Such changes propagate from the binding region through the whole immunoglobulin domains. Overall, complexed antibodies show more diffuse networks of nonbonded interactions and a general higher internal dynamic coordination, which preferentially involve the immunoglobulin (Ig) domains of the heavy chain. The combined results provide atomistic insights into the correlations between the modulation of conformational dynamics, structural stability, and allosteric signal transduction. In particular, the results suggest that specific networks of residues, shared among all the analyzed proteins, define the molecular pathways by which antibody structures respond to antigen binding. Our studies may have implications in practical use, such as the rational design of antibodies with specifically modulated antigen-binding affinities.



INTRODUCTION

Proteins carry out their functions and participate in biochemical interaction pathways by switching among a limited number of substates, which favor the adaptation of the proteins to different partners. Modulation of conformational dynamics, while maintaining the native 3D structure, allows biomolecules to fine-tune their functions in response to varying conditions. This guarantees the specific and tightly regulated molecular recognition that underlies all aspects of cell life, from the regulation of cell development and signaling to the neutralization of pathogens by the immune system.[1−7] Unraveling the role of protein structure and dynamics in determining the fundamental mechanisms of molecular recognition is therefore central to understanding biology at the molecular level.

In this context, the process of complex formation between antibodies (Ab) and antigens (Ag) is one of the most relevant examples, in which highly specific and dynamic adaptation between the two binding partners is required to trigger a fundamental biological process, namely, the immune response to invading foreign molecules. Understanding the molecular-level determinants of how an Ab recognizes its Ag is thus an important issue for both fundamental[1,3] and practical reasons.[2,5,6] From the fundamental point of view, such knowledge could help in furthering our understanding of the relationships between protein sequence, structure, and functional evolution as well as shedding light on the molecular determinants of immune response. On this basis, from the practical point of view, we will be able to begin to rationally engineer antibodies or select novel antigens with characteristics suitable for a particular application, e.g., in diagnostics or vaccine development.

Antibodies (which belong to the immunoglobulin superfamily) are multichain proteins that consist of two identical light chains and two identical heavy chains. Each chain is composed of multiple domains in tandem (two for the light chain and usually four for the heavy chain), each consisting of seven to nine antiparallel β strands folded into a Greek key β sandwich (immunoglobulin fold). Antibody molecules are Y-shaped structures composed of three regions connected by flexible protease-sensitive linkers: two antigen binding regions, each referred to as Fab, and a constant region known as Fc. Each Fab fragment can be in turn subdivided into four distinct immunoglobulin (Ig) domains: the variable domains of the light (V_L) and heavy (V_H) chains and the constant domains of the light (C_L) and heavy (C_{H1}) chains (see also Figure 1A for further details).

Received: October 31, 2012
Revised: December 14, 2012
Published: December 17, 2012

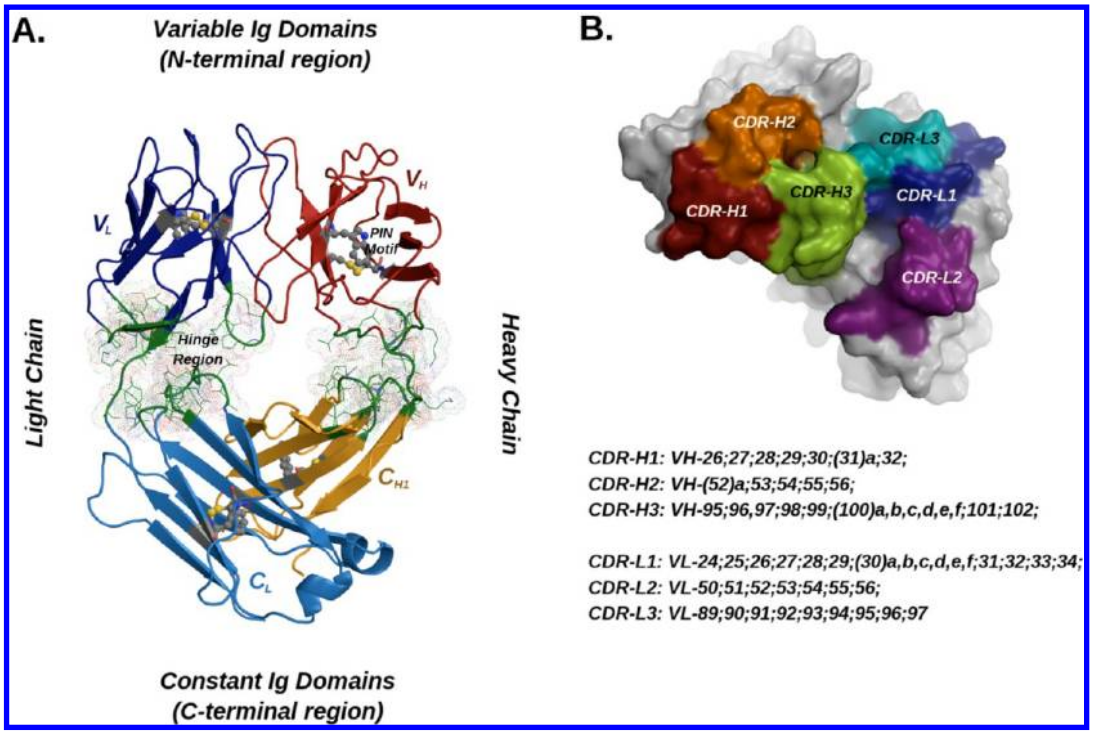


Figure 1. Schematic representation of a Fab fragment. The PDB structure 1BGX is used as a reference. (A) Structural elements of a Fab fragment. (B) A close-up view of the antigen binding site. Complementarity determining regions are highlighted in colors together with the list of the composing residues according to the nomenclature scheme adopted herein (derived from the Chothia numbering scheme).

Antibodies can recognize and bind with high specificity and high affinity a wide range of diverse antigens that span from small molecules to large proteins.[8] Six accessible loops are involved in antigen binding, and they are defined as complementarity determining regions (CDRs, Figure 1B). These regions show high variability in sequence and spatial arrangement, which determine the conformational flexibility and the physicochemical properties necessary to adapt to the shape and composition of the antigen. Optimization of such properties occurs during the process of affinity maturation of antibodies.[9−12] The regions of the variable domains outside these loops are called framework regions and are highly conserved in both sequence and structures. In some cases, the portion of the framework regions proximal to CDRs is involved in antigen binding.[13−16]

One of the major open issues is the comprehension of the molecular mechanisms by which antigen binding is translated into molecular signals directed to activate other sites, modulate interactions with other molecules (e.g., the complement or Ab aggregation), or switch on/off specific communication networks.[17,18] Indeed, interactions of the antibody with the ligand entail a precisely defined location, but their effects are clearly propagated to distal regions, determining the functional response.

In this context, differences in specificity and affinity have been observed among isotypes which share the same identical variable domains but different amino acid composition in the C_{H1} domain.[19−22] Structural evidence indicated allosteric antigen-induced conformational changes in the constant region.[23,24] However, it was not possible to discern whether such effects were due exclusively to antigen binding.[25] Additional experimental evidence has shown that the perturbation of the constant domains could impact the thermodynamic stability of the variable domains.[26−29]

Despite the high level of sophistication, there is still no experimental technique that can provide insights at an atomic level into the Ab::Ag recognition process and their possible allosteric consequences. Theoretical approaches offer a precious tool to understand such phenomena at atomistic resolution.

Recent results based on the development and application of novel computational approaches have started to shed light on the dynamic coordination between distal sites[5,30−38] in proteins. Structural bioinformatics strategies have been used to compare different antibodies and identify the sites and modes of interaction between the heavy and light chain variable domains that affect the relative position of the antigen-binding loops and the overall conformation of the binding site.[9,39]

To make further progress in understanding the relationships between structure, molecular recognition, and allosteric events in antibodies, it is necessary to compare the modulation of internal dynamics and energetics of a diverse set of isolated antibodies and their respective antigen-bound complexes. To this end, herein we have applied recently developed methods to analyze the dynamic coordination and energetic properties of biological systems starting from multiple all atom molecular dynamics (MD) simulations.[36,40] In particular, we have considered 28 different antigen binding fragments, or Fab molecules, in isolation (*apo* system) or in complex with a protein antigen (*holo* system, Fab::Ag), resulting in a total of 56 simulated systems.

The comparative analysis of specific patterns of inter-amino acid fluctuations allowed us to identify possible common mechanisms of coordinated conformational changes determined by antigen binding at the CDR site and diffused throughout the whole Ab structure. Moreover, we analyzed the energy networks that connect the antigen binding site to other antibody residues that are distant from it, in sequence and structure. In this context, we identified common modulation

patterns of such networks that are consistently observed in the presence of the antigen.

The combined results provide atomistic insights into the correlations between the modulation of conformational dynamics, structural stability, and allosteric signal transduction. In particular, the results suggest that specific networks of residues, shared among all the analyzed proteins, define the molecular pathways by which antibody structures respond to antigen binding.

In summary, in this article, we provide both a dynamic and an energetic view of the process of molecular recognition between antibodies and antigens and we merge these two facets into an integrated model that, together with those already proposed, will possibly aid in the design of new antibodies that are able to adapt to specific antigens in diagnostic and immunological applications.

METHODS

Structure Data Set. For all calculations, we have considered 28 Fab::Ag complexes, whose X-ray crystal structures are available form the Protein Data Bank (Table 1), expanding the data set used in a previous work.[41] Some structure was not complete (PDB ID: 1AHW, 1BGX, 1BJ1, 1CZ8, 1FE8, 1IQD, 1OAK, 1PKQ, 1TZH, and 2FJH), due to nonresolved atomic coordinates for short loops and isolated residues. In all of these cases, the missing residues are distal from the binding sites of the Fab. Missing short loops were modeled using the *Modeler 9v8* software. The structures for subsequent MD analyses were selected on the basis of the best DOPE score.[42]

Molecular Dynamics Simulations. For most of the structures, 50 ns MD simulations were computed. Seven complexes (PDB ID: 1BJ1, 1CZ8, 1MLC, 1NDG, 1NDM, and 1P2C) exhibit similar antigen binding sites, termed paratopes, and the literature related to these complexes shows structural evidence that the antibodies involved are expected to recognize almost identical patches, i.e., epitopes, on the antigen molecules.[43−45] The main difference consists in the Ab affinitities for Ag's, due to affinity maturation processes and/or isotype switching events. The simulation time for these structures was extended to 200 ns. For each member of the data set, two simulations were run: one for the isolated Fab structure, called the *apo* system, and one for the Fab::Ag complex, called the *holo* system. The *apo* systems were obtained from the original coordinate files, after removing the Ag molecule. Every system was submitted to unrestrained MD simulation in isothermal−isobaric (NPT) conditions; the time step was set to 2 fs. The *GROMACS 4.0.7* software,[46] with the *Gromos96 ffG43a1* forcefield,[47] was used for the simulations and for the subsequent trajectory analyses. The energy of each system was initially minimized with 2000 steps of the steepest descent method, to remove bad contacts and nonoptimal lengths and angles. The MD trajectories were calculated in periodic orthorhombic boxes, and the simple point charge (SPC) water model[48] was chosen for explicit solvation of the system. The box dimensions were defined in order to allow at least 1.2 nm between the protein and the boundaries. The box dimensions, the number of water molecules, and the counterions added for neutralization are listed in Table S1 (Supporting Information). All of the bonds were constrained by the LINear Constraint Solver (LINCS) algorithm,[49] the atom velocities for start up runs were obtained according to a Maxwell distribution at 300 K. The temperatures of proteins and solvent were separately coupled, through a Berendsen thermostat,[50] to a bath with temperature 300 K and time constant 0.2 ps. Isotropic pressure coupling of the systems was based on the Berendsen weak coupling algorithm, with time constant 1 ps and compressibility 4.6×10^{-5} bar^{-1}. The Particle-Mesh Ewald method (PME)[51] was adopted for electrostatics and a cutoff distance of 0.9 nm was used for van der Waals interactions. All analyses were carried out on the part of the trajectory obtained after discarding the first 10 ns to avoid equilibration artifacts.

Table 1. Data Set of Crystallographic Structures

complex (PDB ID)	resolution (Å)	source	isotype	description (mAb::antigen)
1AFV	3.70	mouse	IgG	25.3::HIV-1 capsid protein (P24) complex
1AHW	3.00	mouse	IgG1	5G9::extracellular domain of tissue factor
1BGX	2.30	mouse	*na*	TP7::TAQ polymerase
1BJ1[a]	2.40	mouse	*na*	mAb::vascular endothelial growth factor
1CZ8[a]	2.40	mouse	IgG	mAb::vascular endothelial growth factor
1DQJ[b]	2.00	mouse	IgG2a	HyHEL-63 2::hen egg white lysozyme
1FDL	2.50	mouse	IgG1	D1.3::hen egg white lysozyme
1FE8	2.03	mouse	IgG2a	RU5::von Willebrand factor A3 domain
1FSK	2.90	mouse	IgG1	mAb::brich pollen Bet v 1(major allergen)
1H0D	2.00	human	IgG1	26-2F::human angiogenin
1IQD	2.00	human	IgG4	BO2C11::human factor VIII C2 domain
1MHP	2.80	mouse	IgG1	mAb:: alpha1 integrin I-domain
1MLC[c]	2.50	mouse	IgG1	D44.1::hen egg white lysozyme
1NCA	2.50	mouse	IgG2a	NC41::influenza virus N9 neuraminidase
1NDG[b]	1.90	mouse	IgG1	HyHEL-8::hen egg white lysozyme
1NDM[b]	2.10	mouse	IgG1	HyHEL-26::hen egg white lysozyme
1NSN	2.80	mouse	IgG1	N10::staphylococcal nuclease
1OAK	2.20	mouse	IgG1	NMC-4::von Willebrand factor A1 domain
1P2C[c]	2.00	mouse	IgG1	mAb::hen egg white lysozyme
1PKQ	3.00	mouse	IgG1	8-18C5::myelin oligodendrocyte glycoprotein
1RJL	2.60	mouse	IgG2a	H6831::OSPB-CT
1TPX	2.56	mouse	*na*	VRQ14::ovine recombinant PrP
1TZH	2.60	mouse	IgG	YADS1::vascular endothelial growth factor
1YNT	3.10	mouse	IgG1	mAb::*Toxoplasma gondii* surface antigen 1
1YQV	1.70	mouse	IgG1	HyHEL5::hen egg white lysozyme
2ADF	1.90	human	IgG	82D6A3::von Willebrand factor A3 domain
2FJH	3.10	human	IgG1	B20-4::vascular endothelial growth factor
2JEL	2.50	mouse	IgG1	JEL42::HPR

[a,b,c]These complexes share similar epitopes; then, 200 ns MD simulations are performed for them.

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

Equilibration of Molecular Dynamics Simulations. To examine the equilibration of the systems and if the Fab::Ag complexes were stable in the simulation conditions used, the atomic positional Root Mean Square Deviation (RMSD) of the backbone atoms with respect to the starting minimized structure was calculated. The *apo* and *holo* systems appear to be stable, as depicted in Figure S1 (Supporting Information). The distribution of RMSD values is comparable among all the systems (see Figure S2, Supporting Information) with central values (ie medians) ranging from 1.8 to 2.9 Å for *holo* systems and from 1.8 to 3.2 Å for *apo* systems, and no major structural rearrangements. Considering the time-evolution of RMSD for all systems, we decided to discard the initial 10 ns of each MD simulation from the analysis.

Primary Sequence Numbering Scheme. For the variable domains, V_H and V_L, we have adopted the improved Chothia numbering scheme proposed by Abhinandan and Martin in:[52] this version of the Chothia scheme was shown to be structurally correct throughout the CDRs and framework regions (see ref [52] for a detailed discussion of the errors in previous numbering schemes and the strategies used to correct them). Such corrections were obtained by exploiting the knowledge of annotated sites of mutations and sequence insertions/deletions (*indels*). Since the variable domain residues are well-defined in sequence based (Kabat) and structure based (Chothia) terms, the identifier of residues involved in *indel* events are typically characterized by a lower-case, one-letter suffix, e.g. CDR-H1−31a; CDR-H1−31b; etc.

In the case of the constant domains, to the best of our knowledge, no certain rules are provided for mapping the *indels*. Therefore, to alleviate this problem, we have introduced a numbering scheme obtained from the multiple structural alignment over the individual constant domains, C_{H1} and C_L. The alignment − and the subsequent superposition of domains − was performed by *MUSTANG* software.[53] Figure S3 (Supporting Information) shows the resulting structural alignment, with the worst aligned regions highlighted in colors. Such regions correspond to short external loops or stretches of poorly conserved sequences. The structural alignment proposed shows a high degree of similarity, therefore the constant domains could be considered homologous.

Finally, the nomenclature followed herein for the assignment of the secondary structure elements was described in details by Sun and Boyington in,[54] where constant Ig domains have been assigned to a C1 type fold while variable Ig domains have been assigned to a V type fold.

The residues of Fab structures were thus labeled according to the following scheme: *[VH|VL|CH|CL]-nnnl*. The capital-letter prefix identifies the specific Ig domain (*Variable* or *Constant*) associated to the specific peptidic chain (*Heavy* or *Light*). The subsequent sequence id is a number *n* plus an optional one-letter suffix *l*, characterizing the *indels* of variable domains only. For example, VH-26:VH-33 is the range of residues which span the CDR-H1 region. A visual representation of the assignment of different regions to a certain nomenclature descriptor for Fab's is shown in Figure 1. Table S2 (Supporting Information) lists the correspondences between the nomenclature used herein and the crystallographic numbering provided by the original PDB files.

Analysis of the Networks of Coordinated Distance Fluctuations. To characterize the dynamic response of the antibodies to the presence of the antigens, we analyzed the fluctuations of the distances of all of the amino acids pairs during simulations. Such analysis provides valuable information on the existence of quasi-rigid regions in the molecule as well as significant mechanical and dynamical correlations in the motion of protein regions that may be far away from the Ag binding site and may be separated in space and/or along the primary sequence. The matrix of distance fluctuations is defined as follows, for i not equal to j:

$$D_{ij} = \frac{\langle d_{ij} \rangle}{\sqrt{\langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle}} \quad (1)$$

where d_{ij} is the pairwise distance between the Cα atoms of amino acids i and j and the brackets indicate the time average over the trajectory. Compared to the definition presented in,[36] the value of D_{ij} used here defines an index of dynamic coordination which allows for larger fluctuation for residue pairs that are separated by larger distance, as can be inferred from eq 1.[55] Given their spatial proximity, amino acids belonging to the same domain should be associated to smaller distance fluctuations, resulting in larger D_{ij} values. High D_{ij} values for amino acids in different domains are expected to indicate a significant degree of mechanical coordination, as shown by previous studies carried out by our group.[31,36]

Analysis of the Networks of Interacting Residues. The networks of highly interacting residues in the *apo* and *holo* states of the Fab's were analyzed by the Energy Decomposition method, which allows detecting the energetic residue couplings important for the stabilization of a structure.[56] The method provides a simplified view of residue−residue pair interactions, extracting the major contributions to the energetic stability of the native structure from the results of all-atom MD simulations. For a protein of N residues, the NxN matrix (M) of nonbonded interactions between pairs of residues can be built by averaging over the structures visited during an MD trajectory.[40,57−59] The rather noisy energy matrix is then simplified through eigenvalue decomposition. Analysis of the N components of the eigenvector associated with the lowest eigenvalue was shown to identify residues behaving as strong interaction centers, characterized by components with an intensity higher than the threshold value corresponding to a "flat" normalized vector whose residues would all provide the same contribution. We verified that applying this analysis to the representative conformation of the most populated structural cluster from the simulation yields the same results as the averaging over the equilibrated part of the trajectory.[60,61] This approach requires the most populated structural cluster to be significantly more populated than the others. Analysis of the cluster number and population at different MD length scales supports this assumption (see also Figure S4, Supporting Information). The representative structures of the most populated cluster from each simulation were minimized. Nonbonded energies for each pair of amino acid i and j were computed using the Molecular Mechanics Poisson−Boltzmann Surface Area (MM-PBSA) method.[62] The complete network of interactions defines the global stabilization energy of the protein, it is described by a matrix M_{ij}:

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

Nonbonded energies were measured by summing electrostatic $E_{elect,ij}$ and van der Waals $E_{vdW,ij}$ contributions. In addition, solvation effects $G_{solv,ij}$ has been taken into account.

Eigenvalues Decomposition of Distance and Energy Matrices. Eigenvalues decomposition was carried out on both

matrices of both matrices D_{ij} and M_{ij}. The aim of this step is to identify a subset of highly coordinated (from D_{ij}) and/or strongly interacting residue pairs (from M_{ij}) that recapitulate the principal distance fluctuations and coupled contacts, yielding a global and compact description of the conformational and energetic response of Fab to Ag binding. In Figure S5 (Supporting Information), a workflow of the decomposition analysis protocol is provided. A generic matrix A_{ij} describing any quantitative relationship between residues i and j is diagonalized,[56,58] and re-expressed in the form:

$$A_{ij} = \sum_{k=1}^{N} \lambda_k w_i^k w_j^k \tag{3}$$

where N is the number of amino acids; λ_k is the k^{th} eigenvalue with k ranging from *1* to N; $w^k{}_i$ and $w^k{}_j$ are the i^{th} and j^{th} components of the associated normalized eigenvector. Then, the list of eigenvectors is sorted according to their respective eigenvalues. Based on the definition of the two measures for the dynamic and energetic couplings, the eigenvalues obtained from the distance fluctuation matrix D_{ij} were sorted in decreasing order, so λ_1 is associated with the most positive value. The ones obtained from interaction nonbonded energy matrix M_{ij} are sorted in increasing order, so λ_1 is associated with the most negative value, associated to the principal contribution to stabilization. Such decomposition method was broadly documented in,[40,41,56,58,59,63] with applications mainly oriented to nonbonded energy network analysis. In the case of the energy decomposition methods, it was observed that small or single-domain proteins can be accurately described considering only the eigenvector associated to the first eigenvalue λ_1.[58] This approximation is not suitable for more complicated cases, such as multidomain structures like the Abs. Indeed, it has been observed that more eigenvectors are generally needed to completely represent all the interactions that are fundamental for the protein stability. This issue was extensively investigated by Genoni and co-workers in[64] and, in agreement with the method proposed by the authors, we have considered a set of essential eigenvectors which bring together a sufficient information content. Afterward, these eigenvectors are combined (see *STEP 2* in Figure S5, Supporting Information). Finally, residues were defined as important hot spots for determining the dynamics or interaction networks if their component in the respective matrix was higher than a threshold value t, which depends only on the number N of residues:

$$t = \sqrt{\frac{1}{N}} \tag{4}$$

This threshold value t was calculated as the value corresponding to a normalized vector whose components provide the same contribution for each site, ie a flat eigenvector.[58]

Ranking of Hot Spots. The rank products method has been adopted[65] to classify recurrent positions (CMRP and IERP) and identify hot spots. The theoretical approach of positional methods can be applied to combine ranked lists in various biological problems, where general feature selection is required.[66,67] Every system i of the data set is considered as an independent replicate for scoring a recurrent position. The rank products score R_g is calculated as follows:

$$R_g = \prod_{i=1}^{K} \frac{r_{i,g}}{N_i} \tag{5}$$

For each residue g found in K replicates we calculate the ranking position $r_{i,g}$ of residue g among the N_i residues of the i^{th} replicate, sorted by their component values extracted from decomposition analyses. The R_g value is related to the combined probability of observing residue g ranked so high just by chance (e.g.: "the third residue with the higher component value or better"). The condition under control here is the presence of antigen molecules during MD simulations, determining two states to be compared: the unbound state of the Fab (in *apo* systems) versus the complexed state (in *holo* systems). Equation 5 was applied for ranking residues in both systems and a combined score was calculated through the two-sample rank products variant:[68]

$$\log(RP_g) = \log(R_g^{apo}) - \log(R_g^{holo}) \tag{6}$$

The sites with the smallest RP_g values are those residues that are differentially relevant in *holo* systems; viceversa, sites with the biggest RP_g values are those residues which better describe the *apo* systems. Statistical significance of $R_g{}^{apo}$ and $R_g{}^{holo}$ have been assessed for multiple testing through 10,000 permutations. Only sites that hold an $E_{value} \leq 0.05$ have been taken into account. The detailed list of rank products scores related to CMRPs and IERPs are presented in Tables S3 and S4 (Supporting Information), respectively.

Based on this, a recurrent position g is regarded as an [*apo*|*holo*] hot spot if at least one fold change is detectable through the comparison of $R_g{}^{apo}$ and $R_g{}^{holo}$ scores. Recurrent positions which show $RP_g \approx 1$ are ranked in similar way in *apo* and *holo* systems and are defined as "*both*" hot spots.

RESULTS

This work is based on an all atom molecular dynamics (MD) study of a data set of 28 Fab's in isolation and in complex with their respective Ag (56 systems in total). The structures are listed in Table 1. The data set was defined to be nonredundant and represents different structural motifs for the antibody binding epitopes on the antigens, including random coils as well as ordered secondary structures. Structural diversity at the binding interface is relevant in our analysis since it may (or may not) reflect in different long-range dynamic and energetic responses across the Fab structures.

Equilibration and convergence of the different simulations was evaluated by calculating the time evolution of the Root Mean Square Deviation (RMSD) of the atomic positions along MD simulations with respect to the corresponding starting structures. The RMSD curves depicted in Figure S1 (Supporting Information) show that the systems could be considered equilibrated after the initial 10 ns of simulation. Furthermore, the stability of the systems was evaluated by measuring the respective total internal protein energies, as reported in Table S5 (Supporting Information). No substantial fluctuation could be observed. The resulting coefficient of variation was about 7×10^{-4} around the time-averaged total energy values of each system.

In order to carry out a comparative analysis of the dynamics and energetics of multiple proteins, sharing homology at the sequence and structure levels, we have used a standardized numbering scheme that allows the direct comparison of properties without alignment. All Fab's were thus referred to

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

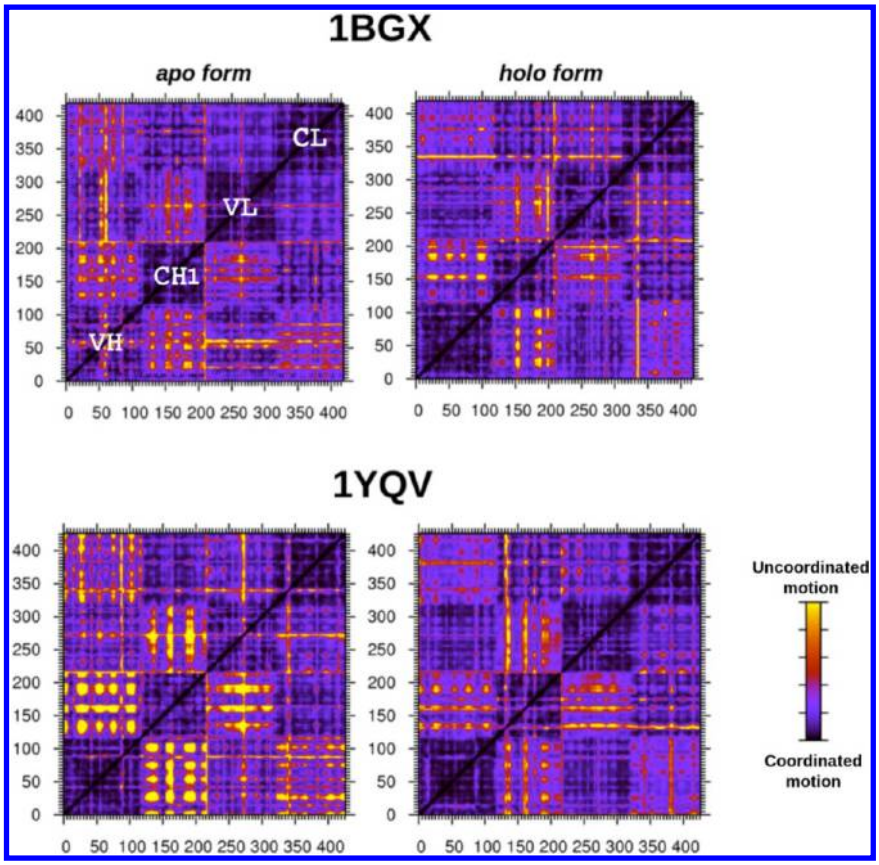


Figure 2. Representative coordination matrices. The matrices for two structures of the data set are shown. The D_{ij} values and the corresponding color ramp scale have been normalized; the lower the distance fluctuation value, the more uncorrelated motion (see also eq 1). The most correlated motions are clustered around the Ig domains of the Fab protein.

a common structural reference and the topologically equivalent regions were labeled using the same nomenclature for each structure. The details of the numbering scheme are reported in Methods.

To analyze the short- and long-ranged antigen-dependent dynamic changes we first computed the fluctuations of pairwise amino acid distances in the MD trajectories of the *apo* and *holo* complexes (see Methods). These quantities are used to define matrix D_{ij}, in which each entry defines an index of dynamic coordination between residues i and j. For the sake of simplicity, high D_{ij} values indicate high coordination (rigidity) between i and j. Equation 1 in Methods shows that this parameter accounts for larger fluctuations for residue pairs that are separated by larger distances. Given their spatial proximity, amino acids belonging to the same domain are expected to be associated to smaller distance fluctuations, resulting in larger D_{ij} values. High D_{ij} values for amino acids in different domains are hence expected to indicate a significant degree of mechanical coordination.[31,36] By means of this description, a residue based map of the dynamical coordination within each antibody in the *apo* and *holo* forms is obtained, which can highlight modulation signals induced by the antigen (Figure S6, Supporting Information). Each matrix is next simplified by means of eigenvector decomposition. The higher components of the main eigenvector obtained from all the *apo* and *holo* systems identify the positions (hot spots) that contribute the most to intramolecular coordination.

It is important to emphasize that the definition of *hot spot* adopted in this work is quite different from the one used in other studies of Fab::Ag or protein−protein complexes. In the latter cases, the *hot spot* definition refers to those residues characterized by peculiar binding properties at the interface. In our context, *hot spot* refers to those residues that are important in defining the conformational responses on passing from the *apo* to the *holo* state of the Fab.

Next, to shed light on the networks of energetic interactions between aminoacids that are crucial for the energetic stability of functional conformational states, we applied the energy decomposition method (EDM). This method, as a first step, computes the matrix of nonbonded interaction energies M_{ij} (namely, van der Waals and electrostatic interactions) between pairs of residues. This matrix is afterward diagonalized, and from the analysis of the eigenvectors associated with the lowest eigenvalues, it is possible to determine which residues behave as the most strongly interacting and/or stabilizing centers.

Definition of Positions with Recurrent Dynamic and/or Energetic Coupling Properties in the Data Set. The topologically equivalent sites that consistently show high dynamic coordination in different representatives of the data set, identified from the analysis of the D_{ij} matrices, are labeled coordinated motion recurrent positions (CMRPs). By analogy, the sites that consistently behave as strong interaction centers, identified from the analysis of the energy matrix M_{ij}, are named interaction energy recurrent positions (IERPs). A recurrent position is determined by evaluating the number of occurrences, over the data set, in which a topologically equivalent site is found to exceed a threshold value t for coordination or energetic interaction (see eq 4 in Methods). As

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

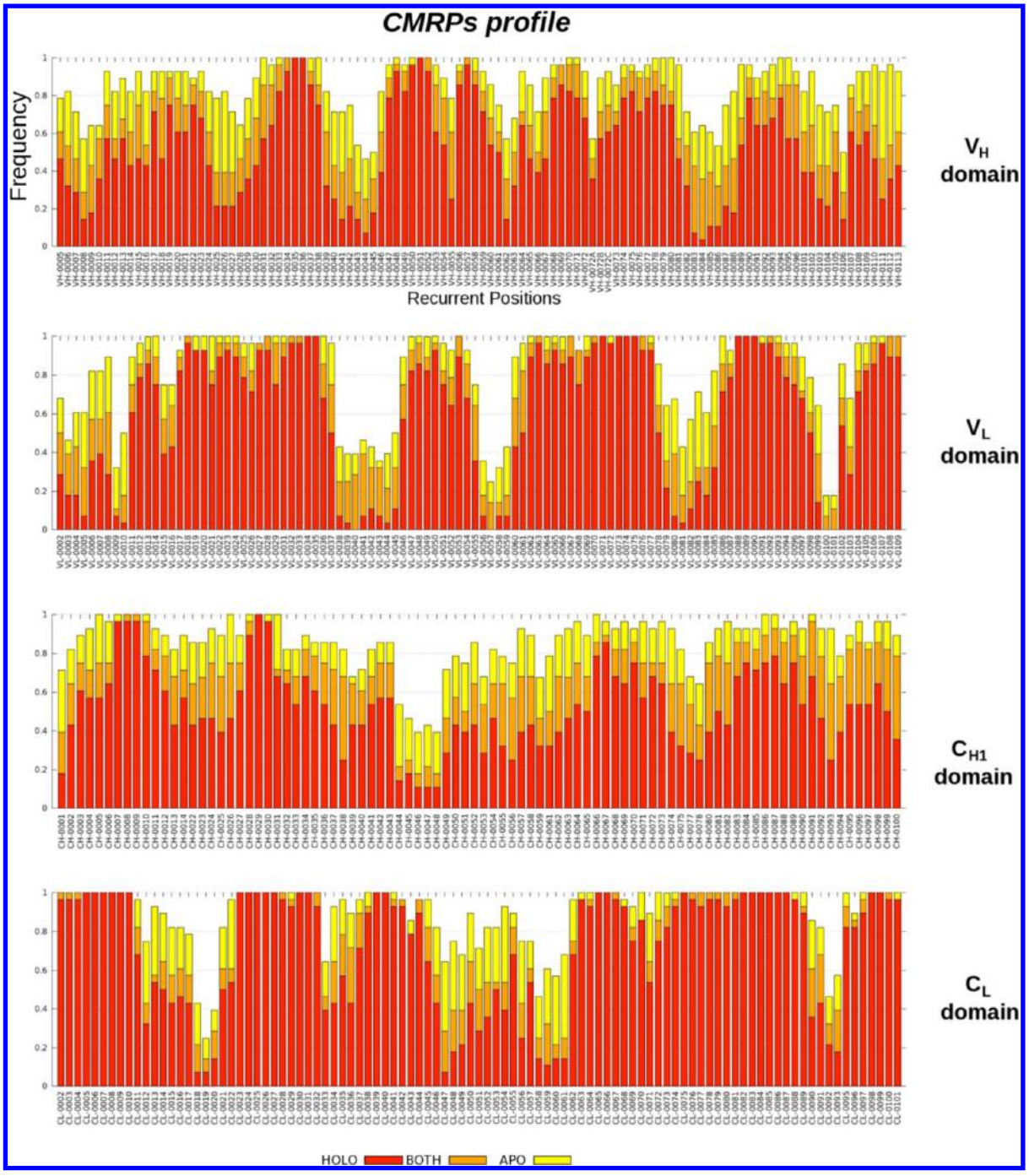


Figure 3. Recurrent position profiles along individual Ig domains, according to the coordination analysis. The stacked histograms show the frequency by which every recurrent position is found in the *apo* and/or *holo* forms of each structure. The frequency scale is normalized to the amount of the entries of the data set. Those recurrent positions that exclusively appear in the *apo* or *holo* forms account for all yellow or all red bins, respectively.

an example, let us consider position CH-36, in the context of the coordination analysis. From our data set, the component value related to this site is found to be above the threshold t in 53.6% of the *holo* systems and in 10.7% of the *apo* systems. In 21.4% of cases, this site is above the threshold t in both. In this context, site CH-36 is defined as a relevant site for regulating the internal coordination of *holo* systems. It is thus defined as a *holo* coordinated motion recurrent position (CMRP). In an analogous way, we define the *apo* CMRPs. If the component value is found to be above the threshold in most of the simulations in both the *apo* and *holo* states, the recurrent position is defined as a *both* CMRP. The same reasoning and definitions are applied to IERPs.

The recurrent positions are then mapped onto the common reference Fab structure to obtain a systematic and general description of possible common patterns of effects determined by antigen binding.

Coordination Analysis. A preliminary visual inspection of the coordination matrices D_{ij} immediately shows common features among the different structures of the data set. The matrices have a well-defined block-character, which reflects the multidomain organization of the Fab's. Indeed, high internal coordination can be observed within each domain. In Figure 2,

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

reporting on two representative instances, the four distinct dark areas along the main diagonal delineate the boundaries of the Ig domains C_{H1}, C_L, V_H, and V_L.

In general, the constant domain of the heavy chain C_{H1} shows lower coordination (higher flexibility) with respect to the remainder of the Fab, in particular toward both the variable domains V_L and V_H. However, this preliminary analysis does not permit a clear discrimination between the *apo* and *holo* systems (see Figure S6 (Supporting Information) for a complete view of all of the D_{ij} matrices).

Consequently, coordinated motions were investigated in more detail, first through eigenvector decomposition of the coordination matrices and then by focusing on the coordination of the subset of residues defining the Fab binding pockets with the rest of the protein.

Eigenvalue decomposition was applied to the D_{ij} matrices. The comparison of the Fab eigenvector profiles in the *holo* vs *apo* state in each of the 28 pairs studied was used to label the coordinated motion recurrent positions (CMRPs) as either *holo* or *apo*, according to the procedure described above. The results were then reported in the form of stacked histograms (Figure 3). The histograms focus only on the sites for which the respective main eigenvector component is above the threshold in at least one case (*apo* or *holo*). The *x*-axes report thus only such positions (not the whole sequence). The *y*-axes report the frequency by which each position is assigned a *holo* (red bins), *apo* (yellow bins), or *both* (orange bins) label. The *holo* and *apo* systems show differential profiles along the entire Ig primary sequence (Kolmogorov–Smirnov test, p value ≤ 0.05). Strikingly, several sites are classified as CMRP in all of the Fab::Ag complexes analyzed. The presence of a number of residues with a higher eigenvector component in the bound cases (*holo* forms) indicates a general increase in internal coordination of the Fab's upon complex formation.

We next investigated the coordination between the antigen binding site and the rest of the protein. To this end, we defined $D_{i,\text{paratope}}$ as the subset of the $D_{i,j}$ values that describe the coordination between any residue i in the antibody and any residues j of the binding site, directly involved in binding. In this context, the *paratope* group of residues entails the amino acids involved in hydrogen bonds and/or salt bridges with the Ag molecule. The *apo* systems globally show slightly lower and less spread values (mean: 40.35 ± 8.96) compared to the *holo* systems (mean: 42.28 ± 10.38); the two distributions are statistically different (F test, p value ≤ 0.05). Globally, the *holo* systems appear to be more coordinated to the antigen-binding site.

It is interesting to observe that the impact of antigen binding on the values of $D_{i,\text{paratope}}$ within each of the Fab subdomains depends on the identity of the domain. The average $D_{i,\text{paratope}}$ values for each domain, in the *apo* vs *holo* systems, are reported in Table 2. The data indicates that the Fab::Ag complex formation leads to a diffuse variation of the internal coordination properties of the Fab's. All domains in the complexed antibodies significantly increase their average value of coordination with the binding site, except for C_{H1}. C_{H1} flexibility remains higher (lower coordination) in both *apo* and *holo* systems. Moreover, $D_{i,\text{paratope}}$ for the C_{H1} domain spans a large ensemble of values, suggesting higher intrinsic internal mobility for this domain in both states (for further details, a scatter plot of $D_{i,\text{paratope}}$ values is provided in Figure S7 (Supporting Information)).

Next, we mapped the $D_{i,\text{paratope}}$ values onto the reference Fab structure and color-coded different positions according to their coordination with the binding site (Figure 4) to highlight whether specific regions of the antibodies directly respond to the presence of the antigen.

In the case of the variable domains, the distribution of the most coordinated sites (colored in orange/red) with the binding site upon Ag binding depends on the specific Ag present and there is no conserved pattern along V_L or V_H domains. On the other hand, the modulation of the coordination with the binding site of constant domains upon Ag binding exhibits common features. In the case of the C_L domains, changes in coordination values entail mainly the strands of the β-sandwich motif and the small helices at the bottom of the Fab. The C_{H1} shows high coordination patterns that entail solvent exposed regions (namely, the strands F and G).

Interaction Networks in the *apo* and *holo* Systems. The functional structures of proteins are stabilized by sparse networks of residues, which underlie the basic properties of a certain fold or conformation. These networks can be spatially organized such that different functional sites (binding sites, conformational hinges, surface interaction patches, etc.) are connected to one another throughout the structure. To identify such networks in the Fab structures and their modulation in response to the presence or absence of an antigen ligand, we applied the energy decomposition method. The method has been introduced to extract the major contributions to the energetic stability of a protein conformation, and its results have been validated against a diverse set of experimental data.[40,41,56−59,61−64] The analysis of the components of the eigenvectors associated with the principal eigenvalues of the average interaction energy matrix has proven to be able to single out those residues behaving as strongly interacting centers, hence determining the stability hot spots of a given protein fold or conformational state.

Here, we applied the energy decomposition approach both to all the Fab::Ag complexes and to the isolated Fab structures in the data set. The results of the eigenvalue decomposition analyses of the energy matrices were used to define the energy-based recurrent positions (see Methods), namely, interaction energy recurrent positions (IERPs).

The stacked histograms in Figure 5 represent the occurrences of IERPs in the *apo* or *holo* systems. The histograms are constructed following the same rules as described for the coordination analysis.

Overall, the distributions indicate that the *apo* and *holo* systems show differential profiles along the entire Ig primary sequence (Kolmogorov–Smirnov test, p value ≤ 0.05), but a remarkable portion of IERPs along the profile of V_H domain is shared by both the *apo* and *holo* systems (orange bins). The C_{H1} domain, which in the $D_{i,j}$ analysis showed the lowest degree of coordination upon Ag binding, is characterized by a limited and sparse number of stabilizing energy couplings.

Table 2. $D_{i,\text{paratope}}$ Average Values in *apo* and *holo* Systems

domain	*apo* system	*holo* system	F test p value
V_H	40.75 ± 8.52	44.24 ± 9.60	≤0.05
V_L	42.77 ± 8.23	45.37 ± 9.41	≤0.05
C_{H1}	34.19 ± 8.08	34.26 ± 8.77	≤0.05
C_L	43.32 ± 8.05	44.64 ± 9.61	≤0.05

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

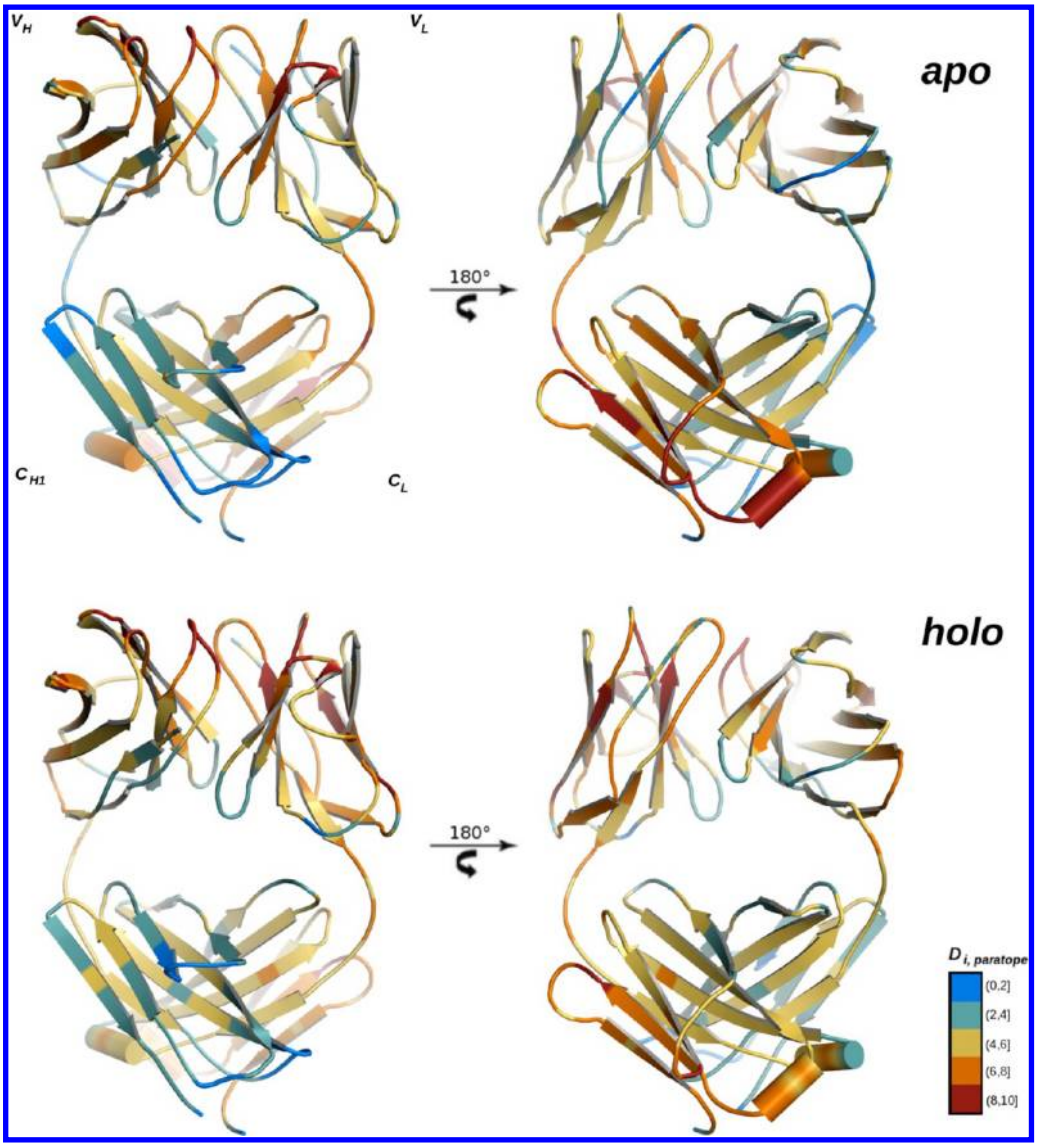


Figure 4. Projection of $D_{i,\text{paratope}}$ values along the Fab structure in the *apo* and *holo* systems. Numerical data has been normalized and clustered in five intervals.

Globally, the *apo* systems show a lower amount of recurrent positions than the *holo* systems. The IERP profiles also show that Ag binding determines an increase in the number of residues participating in energy networks, which span large portions of the Fab structures. Interestingly, such energy networks include residues that are not directly involved in Fab::Ag contacts. The higher number and structurally diffuse character of energetic couplings may help stabilize the bound conformation of the Fab, favoring the formation of the complex with the Ag.

Identification of the Most Relevant Residues for the Dynamic and Energetic Response to Antigen Binding. The CMRP and IERP profiles shed light on the overall propensity of a recurrent position to be involved in dynamic coordination and/or in energetically coupled networks in the different states of a Fab. In order to gain better insights into the role of specific substructures in the selection of the dynamic states of Fab's, we set out to quantitatively identify and rank the sites characterized by a significantly different behavior in the *apo* and *holo* systems. The residues corresponding to recurrent positions (CMRP and IERP) that fall into this category are defined as dynamic or energetic hot spots. For the sake of clarity, let us consider the *holo* CMRPs. If the component values associated with one (or a subset) of these recurrent positions is consistently higher in the *holo* compared to the *apo* systems, this recurrent position (or subset of residues) is classified as a *holo* hot spot. A similar reasoning is used to define the *apo* hot spots. To put this definition on a quantitative ground, the recurrent positions were ranked through the adoption of a classical positional technique, assigning to each site a score according to its positional order along a list. See Methods and Tables S3 and S4 (Supporting Information) for details on the ranking procedure and lists of rank product scores related to CMRP and IERP, respectively.

Consistently with the CMRP and IERP analyses, hot spots are defined as *apo* or *holo* hot spots. Attention should be devoted to the residues that are hot spots in both systems (defined as a "*both*" hot spots). They can be considered as core elements that participate in the general stabilization of the structures and in the definition of the native, functionally oriented internal dynamics of the folded states of the Fab's. Such traits are not influenced by the perturbations induced by

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

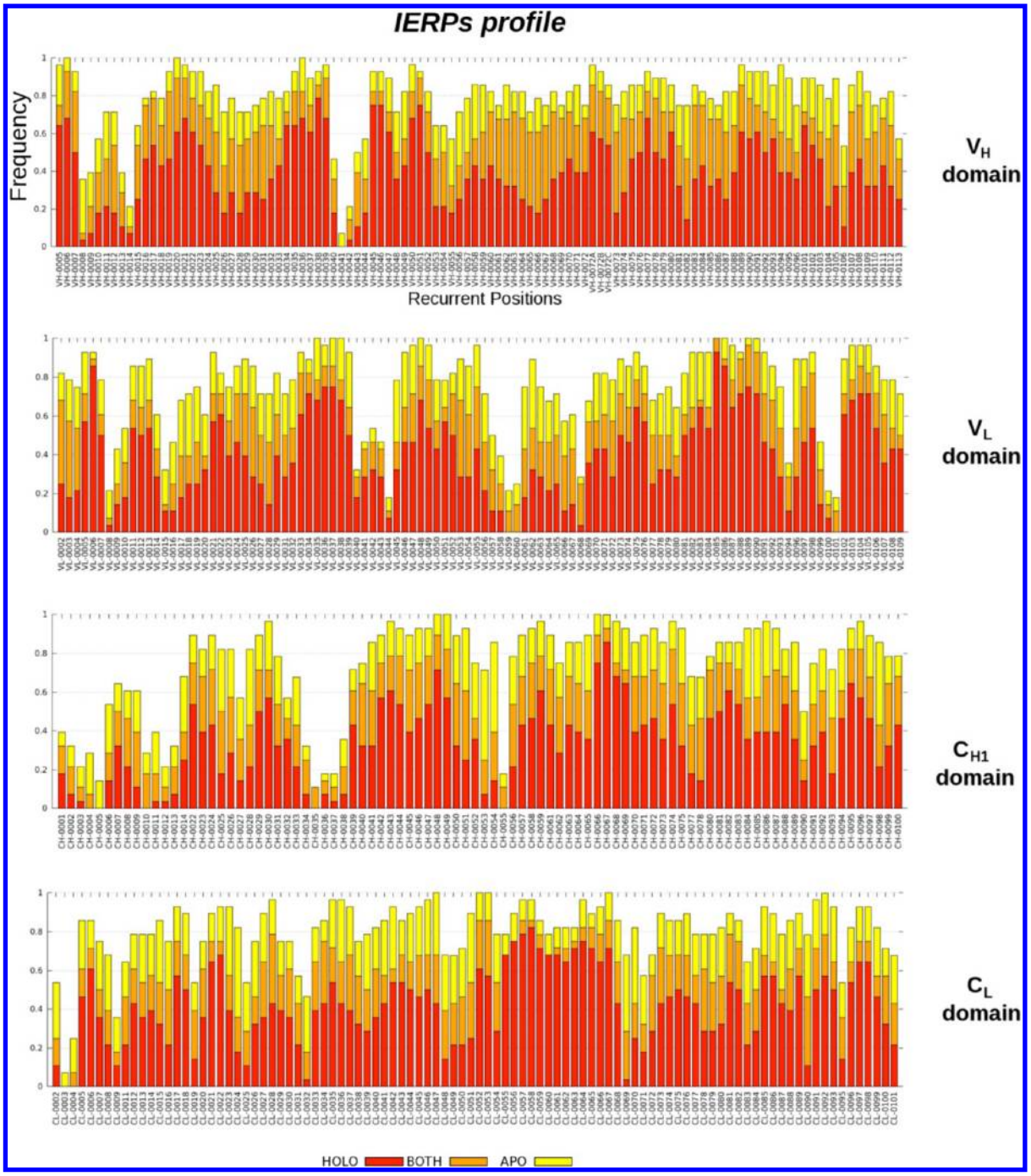


Figure 5. Recurrent position profiles along individual Ig domains, according to the energetic analysis. The stacked histograms are constructed similarly to what is shown in Figure 3. Those recurrent positions that exclusively appear in the *apo* or *holo* forms account for all yellow or all red bins, respectively.

the presence of the Ag molecule. The complete list of hot spots is listed in Table S6 (Supporting Information).

Dynamic Hot Spots. The light and heavy chain domains show a differential distribution of the CMRP hot spots that depends on the binding state of the Fab. Their spatial location is shown in Figure 6. The dynamic hot spots (shown in CPK representation) are distributed asymmetrically on the Fab 3D structure, indicating that the distinct Ig domains are affected in different ways, depending on the presence/absence of the Ag.

The variable region does not show any *apo* hot spot (blue spheres), while *holo* hot spots (red spheres) are localized essentially on the V_H domain. These results suggest that coordinated motions are not markedly relevant for the variable region when the Fab is in the free conformation. In contrast, coordination in the internal dynamics emerges during the simulations of complexed Fab's and only through the V_H domain, independently of the specific location of the antigen binding site, which in some cases may even span the V_L domain. This cluster of hot spots also influences a few residues of the V_L domain, which are in direct contact with the V_H domain (VL-92 which belong to the CDR-L3 region).

The constant region is characterized by a differential behavior for C_{H1} and C_L. In the case of C_{H1}, a continuous, solvent exposed stretch of CMRP hot spots appears only in the *holo*

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

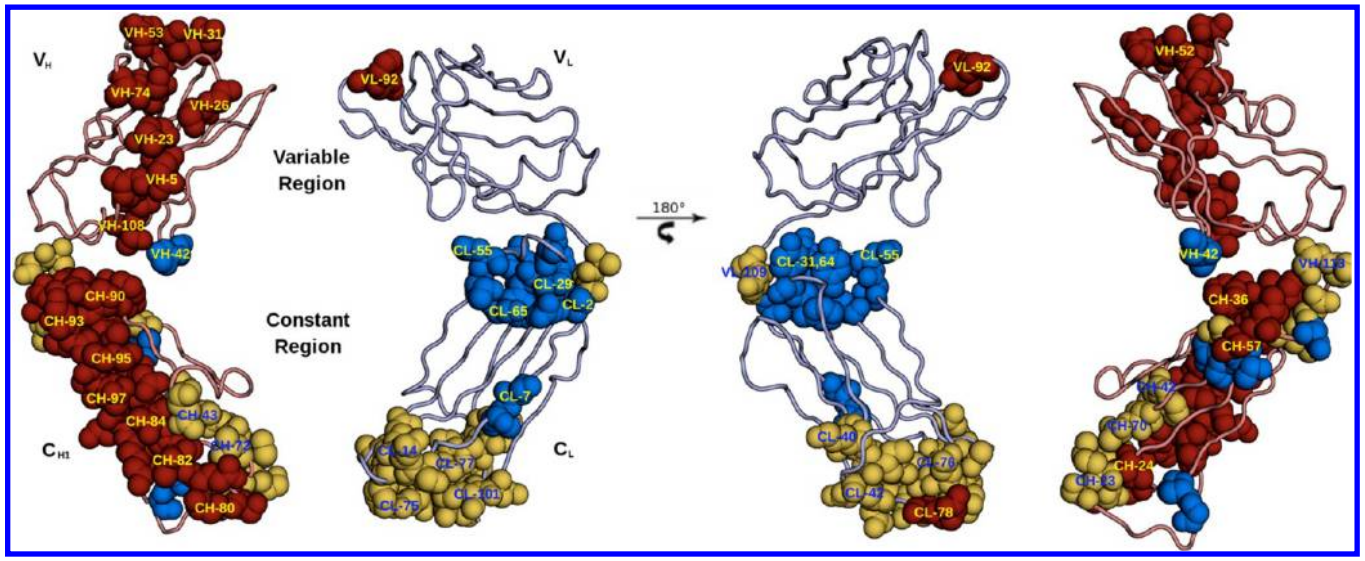


Figure 6. Spatial distribution of CMRP hot spots. The hot spots are plotted and numbered according to the common nomenclature scheme described in Methods. The dynamic hot spots labeled as *apo* or *holo* are shown in blue and red spacefill representations, respectively. Relevant hot spots for both systems are highlighted in yellow.

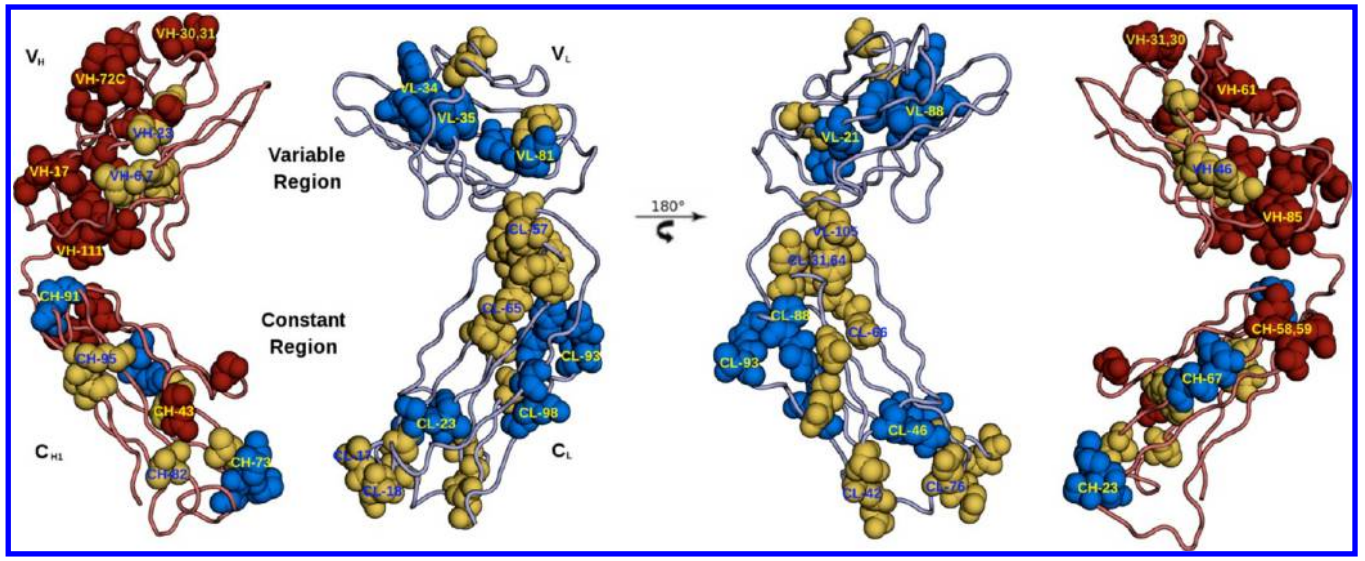


Figure 7. Spatial distribution of IERP hot spots. The hot spots are plotted and numbered according to the common nomenclature scheme described in Methods. The energy hot spots labeled as *apo* or *holo* are shown in blue and red spacefill representations, respectively. Relevant hot spots for both systems are highlighted in yellow.

state (*holo* dynamic hot spots), and connects the upper loops (CH-34..37; CH-90..92) to the β-strands F (CH-83..89) and G (CH-91..100).

The upper portion of the C_L domain is characterized by a compact group of *apo* hot spots (CL-2; CL-29..31; CL-55; CL-62..65), highlighting a rigid core of amino acids that is peculiar of the free state of the Fab (see also Figure 4). These hot spots may define the core element that determines the rigidity of the C_L domain in the *apo* states. Furthermore, an amino acid of the V_H domain facing the C_L domain appears sensitive to such kind of coordination (VH-42).

The C_L domain also shows another group of hot spots, labeled as *both* (yellow spheres). These hot spots define two small helices at the bottom portion of the Ig domain (CL-13..16; CL-73..79), and they are always present in the bound/unbound state of the Fab. The constitutive aspect of these hot spots suggests that this region is essential for the maintenance of the spatial arrangement of the structural elements of the Fab.

Energetic Hot Spots. The comprehensive analysis of the hot spots derived from IERP recurrent positions was aimed at identifying the clusters of residues that may be relevant in stabilizing functional Fab structures. A global overview of such hot spots is depicted in Figure 7. For the sake of clarity, we have maintained the same representation (CPK) and color code adopted for Figure 6 (residues are colored in blue, red, and yellow for *apo*, *holo*, and *both* hot spots, respectively).

The variable domain V_H shows hot spots labeled *holo* or *both*, indicating the formation of a network of interactions that characterizes the Fab::Ag complexes. Energetic hot spots span the full Ig structure and define a pathway of strongly coupled residues that connects the apex of the domain, interacting with the Ag molecule, with the bottom of the domain, interacting with C_{H1} domain. Specifically, this cluster of hot spots partially

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

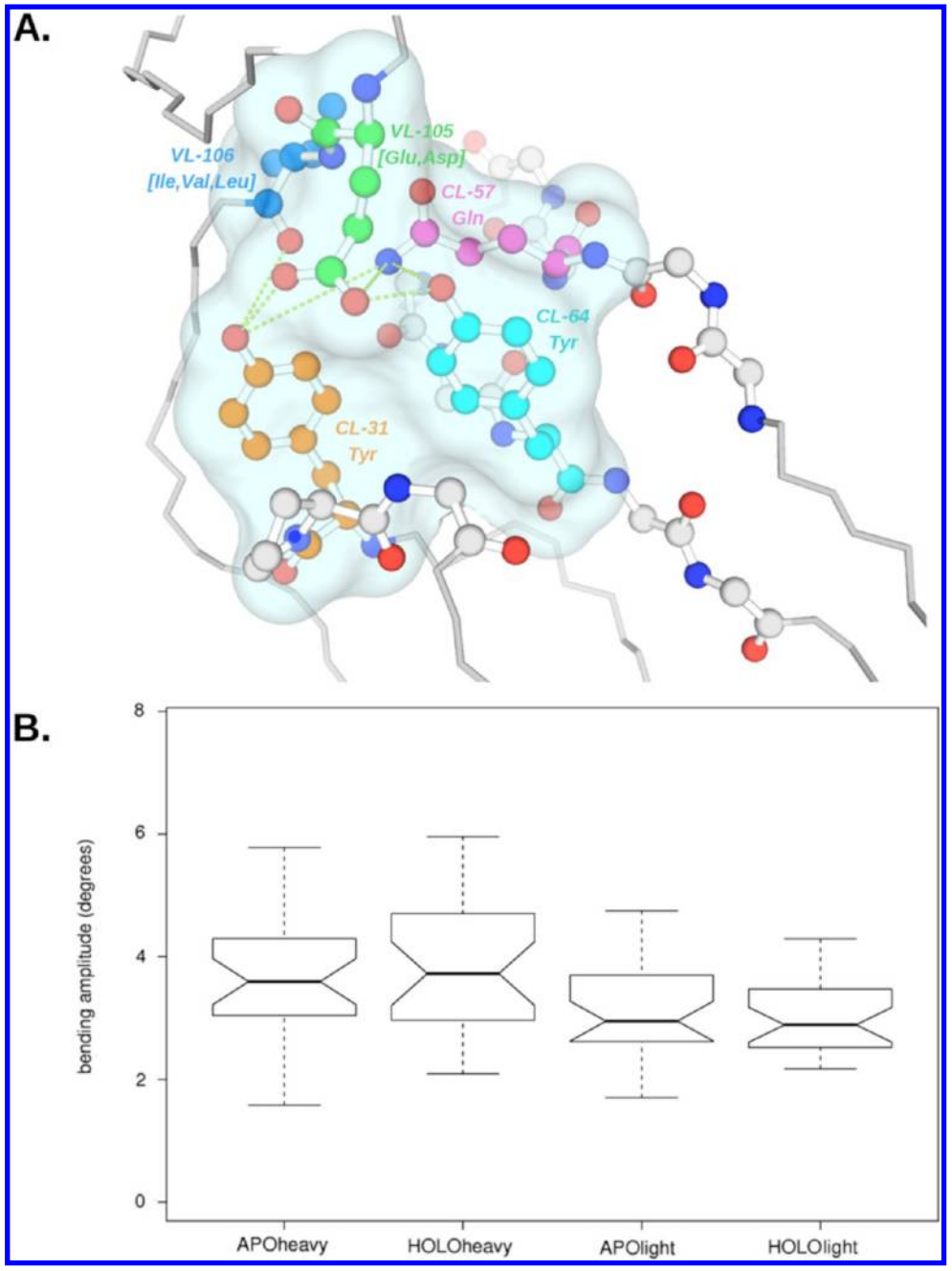


Figure 8. IERP hot spots at the interface between C_L and V_L domains. (A) Snapshot taken from the 1FE8 structure. The hot spot residues are shown in ball and sticks; residues with Cα atoms colored in light gray belong to the CMRP hot spots located nearby. (B) Boxplot of the standard deviation of bending angles measured for the heavy and light chains in apo and holo systems.

overlaps the CDR regions (VH-30,31) and propagates through the framework regions, especially HFR1 (VH-6,7; VH-16..20; VH-23,24) and HFR3 (VH-68; VH-72..72C; VH-85,86). Hot spots are also present at the loop connecting the C_{H1} domain (VH-111).

In contrast, the variable domain V_L shows only *apo* hot spots, concentrated around the PIN motif (VL-35; VL-88). The PIN motif is a core structural element, typical of Ig and Ig-like folds, in which a tryptophan packs against a disulfide bridge. The sequence pattern is *C-m-[LFYW]-n-C*, and the sequence distances *m* and *n* between the conserved amino acids are variable according to the specific type of Ig fold.[69] Ag binding affects the stability of the PIN motif, whose residues are not strongly coupled in the *holo* system. Such perturbation extends in the CDR regions CDR-L1 (VL-34) and CDR-L3 (VL-89) and in the framework regions LFR1 (VL-21) and LFR2 (VL-36).

Another relevant group of hot spots is located at the interface between the variable and constant domains of the light chain. A complex and dynamic network of hydrogen bonds is observed in all of the MD simulations of our data set. The interacting partners are characterized by conserved (CL-31; CL-57; CL-64) or semiconserved (VL-105) amino acids (see a detailed view in Figure 8A). Such residues are found to be IERP hot spots, and they define a strong interaction center that could be observed irrespective of the binding state of the Fab. Other IERP hot spots are found close to the interface (CL-58,59; CL-65,66; CL-88..93), indicating that the neighboring regions are

dx.doi.org/10.1021/jp310753z | J. Phys. Chem. B 2013, 117, 535−552

part of this energetic network. Since this network is located at the interface between distinct domains, it may play a role in determining the overall flexibility of the immunoglobulins. We have thus measured the bending angle between the centers of mass of the constant domain, the variable domain, and the connecting loop, for the heavy and light chains. In Table S7 (Supporting Information), we have reported the bending angles and the standard deviations, averaged on each MD simulation. The heavy chain shows a more closed conformation (89.90°) than the light chain (108.42°), probably due to the packing interactions determined by tyrosynes CL-31 and CL-64. If we consider the bending fluctuation, the light chain shows a smaller amplitude, as shown in Figure 8B. We have also performed a two-way ANOVA (analysis of variance) in order to estimate the statistical significance of this difference and if this difference also depends on the specific conformation of the Fab protein. The results, illustrated in Table 3, show a significant difference between the two chains, even if no significant difference can be found between *apo* and *holo* systems. In summary, a network of interactions is well-defined in the interface region between the domains of the light chain, but no analogous interactions have been observed in the interface region of the heavy chain. In the latter case, a conserved phenylalanine (CH-34) occupies the position of one of the two tyrosines present in the light chain case (CL-31), preventing the formation of stable hydrogen bonds.

Table 3. Two-Way ANOVA Performed over Bending Angle Variation

factors	levels	*F* value	*p* value
system	*apo*|*holo*	0.014	9.07×10^{-1}
chain	heavy|light	12.859	5.06×10^{-4} ***[a]
system:chain		0.332	5.66×10^{-1}

[a]The asterisks denote the statistical significance of results when testing hypotheses. When the likelihood that a result occurred by chance alone is below a certain level, one or more asterisks are displayed. Usually, significance levels are 0.05 (*), 0.01 (**), and 0.001 (***).

In the C_L domains, energetic hot spots are located in the two specific α-helices at the lower portion of the domain (CL-17,18; CL-75,76). They have been considered as critical folding units that render C_L domains capable of folding autonomously.[70]

Finally, sparse clusters of energy hot spots could be found in the C_{H1} domain. A group of IERP *holo* hot spots are located in the vicinity of the solvent exposed β sheet (CH-43; CH-82; CH-95,96), while the *apo* hot spots map to the internal β sheet (CH-66..68) and to the lower portion of the C_{H1} domain (CH-22,23; CH-72..74). Moreover, another group of *holo* hot spots appears in the upper portion of strand D (CH-58..61). Such results show that the C_{H1} domain could be described by distinct and nonconnected groups of coupled residues that do not merge in determining a well-defined stabilization (folding) nucleus. In this domain, it is possible to separate the nonbonded energy networks relevant for the *apo* system from those of *holo* systems.

DISCUSSION

Long-range coordination and interactions are crucial to define protein function and evolution. In this paper, we have focused on Fab::Ag recognition, using MD simulations to investigate the molecular mechanisms underlying the antibody response to the presence of protein antigens. We have elucidated common features characterizing the interplay between structure, dynamics, and interactions in different complexes involved in the same fundamental molecular process. We have tackled specific aspects of this general problem by considering how the long-range inter-residue coordination, internal dynamics, and energetic interaction networks are affected by the bound protein antigen in a diverse set of 28 Fab's. To this end, we have considered the immunoglobulins in the *apo* and *holo* states and used recently introduced MD analysis methods aimed at identifying corresponding dynamical and energetic traits in the data set.[36,40]

Overall, the integrated results allow pinpointing common substructures in all the members of the data set that share the same dynamical and energetic response to the binding of the antigen. Importantly, many such substructures are distal from the antigen-binding site. Moreover, the use of a common reference frame has allowed us to obtain a unified view of the relevant internal motions and stabilizing interaction networks that link antigen recognition to the modulation of the conformational properties of different motifs in the 3D organization of the Ab.

Overall, our analysis identifies the relevant structural stabilization elements of Fab proteins in connection to the internal dynamic changes that underlie allosteric responses to antigen binding. Considering the *holo* systems, Figures 6 and 7 clearly show a continuous network of amino acids (red spheres) that connects the antigen binding site to the bottom of the Fab, encompassing the internal framework residues of the variable regions and the solvent exposed residues of the constant regions. The hot spots characterizing the Fab::Ag complexes suggest that allosteric conformational signals span the entire antibody structure. Strikingly, the two peptidic chains composing the Fab show differential behaviors upon antigen binding, despite the high structure similarity and sequence composition. Indeed, from the dynamic coordination analysis, only the heavy chain seems sensitive to antigen binding, even if the binding sites also comprise CDR regions belonging to the light chain. On the other hand, energetic analyses show that the light chain holds important interaction centers for the constitutive stabilization of the immunoglobulin fold. With this perspective, the structures of Fab's can be decomposed into different sectors in terms of functional roles. While the heavy chain domains are involved in dynamically adapting and responding to the antigen (effector function),[20,25,71] the light chain domains act as scaffolds for the stabilization of the quaternary structure of Fab fragments.[70,72]

Dynamic and Energetic Modulation of Fab Structures upon Ag Binding. Our analysis started by considering the fluctuations of the distances of all pairs of amino acids, as reporters of internal dynamic coordination between residues at variable distances.[31,36] Globally, the *holo* systems show more coordinated motions with respect to the *apo* ones. Several Fab residues show high coordination with the active site, even if they are physically separated at distances of more than 6 Å from the CDR regions (in this case, 6 Å was chosen as the cutoff distance—minimal distance—between two residues to define the long-range coordination). This observation points to the existence of networks of long-range interactions between remote regions of the protein. Moreover, the comparison of *apo* and *holo* systems exhibits different patterns of internal coordination, in the absence of any major conformational rearrangement of the Fab structure.

Focusing on the internal dynamics of the constant domains, the results show that the residues belonging to the same Ig domain show highly internal coordinated motions. The domain boundaries in each Fab can be identified as regions of lower coordination, immediately evident from the visual inspection of the fluctuation matrices (Figure 2). In the *apo* systems, C_L domains appear more rigid and coordinated than the respective counterpart of the heavy chain, the C_{H1} domains (Figure 4). In contrast, in the *holo* systems, the C_L domain becomes more flexible. At the same time, the C_{H1} domains only show a slight increase of internal coordination in the bound state, while conserving higher motional flexibility with respect to the other domains.

Indeed, the dynamic behavior of C_{H1} shows interesting features. A systematic inspection of the coordination matrices (see Figure S6, Supporting Information) shows highly uncoordinated motions between the residues of C_{H1} and the residues of both V_H and V_L. The higher mobility of the C_{H1} domain with respect to the others could be related to its functional role as a flexible adaptor that allows the switch of the conformational states of the antibodies required to optimally accommodate (multiple) epitopes located on antigen molecules.

Indeed, the C-terminal portion of the C_{H1} domain is part of the upper hinge (UH) region, which ends in the first Cys forming an interheavy chain disulfide bond[73] in the major hinge in a complete antibody structure. UH regions have been shown to play a key role in determining the structural plasticity required to adapt to multivalent Ag's.[74] The multivalent interaction between Ag and Ab would clearly depend on the spatial distribution of epitopes and the possibility for Fab arms to conformationally adapt to such distribution. Sizeable distortions from the canonical Y shape of the Ab have been observed by X-ray crystallography[75,76] as well as from immunoelectron microscopy.[74,77] Cryo-electron tomography (cryo-ET) studies suggest that Ab molecules possess sufficient flexibility to host antigens with a size of 50−90 Å.[78,79] Moreover, the flexibility of C_{H1} has recently been pinpointed as an essential requirement for the neutralizing activity of Ab's against HIV-1.[22]

The modulation of the flexibility and coordination patterns observed for the C_{H1} domains may be reflected in changes in the elbow angle, which defines the relative orientation between the constant region and the variable region of a Fab. Some reports, based on the analysis of crystal structures or short MD simulations, have hypothesized that variations in this angle can be linked to the Ag binding and as a consequence favor the transmission of conformational signals throughout the Ab structure.[80−87]

In this work, we have considered the individual bending angles of the heavy and light chains. A significant difference in angle variation could be detected between chains. In contrast, no significant difference could be observed between the *apo* and *holo* systems. The reduced bending mobility of the light chain may be attributed to extensive interfacial hydrogen bond networks between the Ig domains. Indeed, the interaction surface between V_L and C_L is much larger than that between V_H and C_{H1}, as shown in an extensive analysis of X-ray structures.[26] The smaller number of contacts between V_H and C_{H1} is not sufficient to restrain the respective motions of the heavy chain domains. As a result, the bending angle variation is higher than that in the light chain.

In order to gain more insight into the collective protein motions that characterize the Fab conformational changes upon antigen binding, we have carried out a normal mode analysis (NMA) of all the structures and complexes in our study, as described in ref 88, and characterized the lowest-frequency modes. The reference structures are the representative structures of the most populated clusters (see Methods). The analysis of the extreme projections along the top five normal modes of each system clearly shows that the motions of the Fab proteins can be approximated by the motions of two quasi-rigid blocks, defined by domains V_H:V_L and C_{H1}:C_L, respectively. The large-amplitude motions described by the normal modes involve bending, torsional, and stretching movements between the constant region and the variable region. The bending motion reproduces the variation observed above for the elbow angle.[85,86] The torsional and stretching motions are oriented along the axis connecting the Ag binding site to the C-terminal region of the Fab. While the normal modes related to bending and torsional motions collectively involve the whole protein, the ones related to stretching motion show a collectivity degree of about 60% (collectivity refers to "the fraction of residues that are significantly affected by a given [normal] mode"[88]). This value suggests that the differences remarked in Table 3 may be due to a different mutual coordination of the heavy and light chain. Indeed, the distance fluctuation matrices related to such normal modes (Figure S7, Supporting Information) highlight high flexibility between the domains C_{H1} and V_H (upper left quarter of the matrix), while the light chain domains appear more constrained (lower right quarter), with sporadic relevant signals.

The Dynamic and Energetic Properties of Fab's in Comparison with Experimental Findings. The residues defining the binding sites are mobile and show a low degree of coordination with the rest of the Ab in the absence of the antigen. Upon ligand binding, the dynamics of complementary determining regions show a complex pattern of increased coordination with the remaining portion of the Fab. Experimental evidence shows that modulation of rigidity/flexibility, for which coordination based on fluctuation analysis is a valid reporter, is not restricted to the binding pocket but may entail distant regions[87,89,90] that are unevenly distributed throughout the Ab structure. Their identity, conformational properties, and functional relevance have been discussed in refs 23, 24, and 91.

In this context, our integrated analysis of the dynamic and energetic responses to antigen binding identified hot spots that play a key role in connecting distal substructures. A persistent network of energetically coupled residues (IERP hot spots) connects the binding pocket with the framework region and the hinge region in the V_H domain, as shown in Figure 7. In particular, positions VH-6 and VH-7 are defined as IERP hot spots of the "both" type (yellow spheres in Figure 7). This points to their critical role in the structural core stabilization of the Ab, consistent with observations by Honegger and Plückthun in ref 92. VH-6 and VH-7 establish a stable network of hydrogen bonds in order to connect the opposite β-bulges—internal to the strands A and G—of the β-sandwich motif of V_H. It is also worth noting that substitutions at the hot spots we identify result in a clear decrease of Ab thermodynamic stability and consequent lower production yields.[93,94] For its structural relevance, Honegger and Plückthun have classified the region spanning VH-5..10 residues in four different topological subtypes.[92] Apart from the specific classification, we can

consider this region as important for the dynamic coordination in *holo* systems. Indeed, a continuous stretch of CMRP hot spots is observed for this region (red spheres in Figure 6).

From the experimental point of view, the dynamic behavior of different Fab domains has been investigated through NMR-based amide hydrogen exchange (HX, protection factor) analysis. Williams and co-workers measured variations in the protection factor of the variable fragment (Fv) of a Fab in the *apo* and *holo* forms.[95] The highest protection factors are observed for the *holo* form, indicating a general dynamic coordination of the Fab structure upon binding, combined with a strong energy coupling of the solvent exposed amino acids with the more buried residues (VH-23..26; VH-46; VH-52,53; VH-72..74; VH-85,86). The residues with the highest protection factor are preferentially located at the VH domain, reflecting the asimmetric distribution of *holo* hot spots both from a dynamic (CMRPs, Figure 6) and energetic (IERPs, Figure 7) perspective.

Other HX NMR studies showed that the higher differences in hydrogen exchange measures map to the framework regions.[72] These experimental results are consistent with our observation of *apo* hot spots around the PIN motif of the V_L domain (VL-35; VL-88 in Figure 7). Ag binding perturbs the stability of the PIN motif, whose residues are not strongly coupled in the *holo* system. Such perturbation extends in the CDR regions CDR-L1 (VL-34) and CDR-L3 (VL-89) and in the framework regions LFR1 (VL-21) and LFR2 (VL-36).

Experimental data shows that the constant domains C_L and C_{H1} are coupled to each other in terms of stability. Fluorescence spectroscopy measurements have shown that the C_L::C_{H1} heterodimer undergoes highly cooperative unfolding. The large, well-packed hydrophobic interdomain interface is responsible for the mutual stabilization of the domains.[26] The interface between C_L and C_{H1} is highly coordinated from a dynamic point of view. The internal β-sheets of the constant domains show similar values of $D_{i,j}$ (yellow in Figure 4) in the *apo* as well as *holo* systems. This region thus constitutes an interaction framework by which the dynamic alterations at one chain are efficiently transferred to the other.

In the folded structure, the hydrophobic core of the C_{H1} domain defined by the hot spots CH-23, CH-42, CH-70, and CH-72 (Figure 6) moves as a coordinated unit, as confirmed by experimental approaches in ref 96. The conformational dynamics and structural stability of the C_{H1} domain was shown to be relevant for the selection of the Ab conformation required for productive antigen binding. Surface plasmon resonance (SPR) experiments performed on similar Fab's from different isotypes, in which the primary sequence composition is identical for all the domains except for C_{H1},[20,71] showed variations in fine specificity and affinity for Ag. The structural determinants responsible for the differences were mapped to the solvent exposed β-sheet and the upper portion of the C_{H1} domain.[20] Our analysis consistently identifies a continuous stretch of dynamic hot spots (CH-34..37; CH-80..100) that are present only in *holo* systems (red spheres, Figure 6).

SPR has also been used to investigate the conformational changes induced in C_{H1} by Ag binding. The reporter system was represented by the Staphylococcal protein G (SpG), which binds antibodies through establishing contacts with C_{H1}. Antigen binding to Fab was shown to cause an allosteric change at the C_{H1} that induced the dissociation of Fab::Ag complex from SpG.[25] Importantly, the SpG binding site indentified by the authors of the experimental study overlaps with the C_{H1} region defined by dynamic *holo* hot spots emphasized in Figure 6. Antigen binding thus activates an allosteric perturbation/rigidification of the C_{H1} domain, which propagates through strands F (CH-83..89) and G (CH-91..100).

The structural relevance of the upper loops of C_L as an interaction center typical of the *apo* system (CL-2; CL-29..31; CL-55; CL-62..65; highlighted by blue spheres in Figure 6) was predicted by Kabat and co-workers through sequence analysis of conserved residues,[97] and subsequently hypothesized by normal-mode analysis (NMA) on a crystallographic structure of a Fab::Ag complex.[98] These studies support our findings of high coordinated motion in this region.

Finally, a persistent group of highly coupled and strongly interacting amino acids highlights two peculiar α-helices at the lower portion of the C_L domain (CL-13..18; CL-73..79 in Figures 6 and 7). They have been considered as critical folding units that render C_L domains capable of folding autonomously.[70] Indeed, such helices are not present in C_{H1}, which requires an interdomain interaction with C_L to properly fold into its Ig structure.[99] The C_L terminal helices may play a scaffolding role in maintaining the correct spacing and orientation of the C_L β-strands.[100] As a consequence, these regions are defined as hot spots of the both type (yellow spheres), consistently with their role in the constitutive stabilization of Fab's.

CONCLUSIONS

The results of this study indicate that the response of antibodies to the presence of antigens may be linked to changes in the internal dynamics, coordination patterns, and interaction networks of specific substructures that determine the adaptation of Fab's to the binding partner. The use of multiple Fab's in complex with a diverse set of protein antigens combined with a common reference frame for antibody structure has allowed us to generate an integrated view of the residues that are most important in defining the structural stability and modulation of the conformational dynamics and flexibility that are required for productive Ag binding.[24,26,101,102]

We would like to mention that, to the best of our knowledge, this represents the first study where antibody conformational dynamics is extensively studied on a large data set at atomistic resolution. Such information may be relevant for antibody design. As an applicative example, the surface residues of an antibody can be mutagenized to avoid aggregation, while keeping the dynamic and energetic hot spots, which are functionally indispensable, unchanged. In this context, we envisage that our MD based findings can be integrated with, e.g., the SAP technology[103,104] for the identification of aggregation-prone regions to rationally optimize Ab sequences. A second applicative perspective might entail the application of directed evolution strategies to the regions identified as strongly coupled (energetically or dynamically) with the CDRs, in order to modulate the affinity for the antigen for the design of biomarkers and bioprobes. The analysis presented here cannot provide a detailed, time-resolved picture of the cause−effect relationships of how the binding site is transmitted to remote regions. Nonetheless, the identification of dynamic and energetic *holo* hot spots, outside the region of contact with antigen, clearly shows a network that starts from the Ag binding sites and has the C_{H1} domain as a final target. Such information can be used for a third experimental application, in which the

identified residues are targeted by site directed mutagenesis, and their effect on antibody binding affinity[20,71] is evaluated.

In this work, we have identified relevant differences in flexibility between the heavy and light chains. This may offer suggestions on future directions of theoretical analyses: How the modulation of such flexibility affects the energetic and mechanical coordination of the Fab molecules can be investigated, for example, by restraining or accelerating dynamics at the hinge regions. Moreover, a natural extension of our work would entail the investigation of whole antibody structures in order to completely elucidate the transmission networks from the bidnign site to the effector region, which triggers complement activation and other immune responses, and is located in the Fc fragment. In this context, Zimmermann and co-workers have analyzed the motions of a full IgG model and compared them to the use of Fab fragments.[105]

■ ASSOCIATED CONTENT

Supporting Information

The PDF file (jp310753z_si_001.pdf) contains the following: Figure S1: RMSD time-curves of the systems. Figure S2: Distributions of RMSD values for *apo* and *holo* systems, respectively. Figure S3: Multiple structural alignment on constant domains. Figure S4: Structural clustering along MD simulations. Figure S5: Decomposition analysis workflow. Figure S6: $D_{i,j}$ coordination matrices of all of the structures of the data set. Figure S7: Scatter plot of $D_{i,\text{paratope}}$ values versus the Cα:Cα distances between each residue of the Fab protein and the subset of those including the paratope region. Figure S8: Distance fluctuation matrices related to the normal mode analysis. Figure S9: The main clusters of hot spots. Complete references for refs 16, 21, and 37. Table S1 (jp310753z_si_002.xls): Solvation box settings. Table S2 (jp310753z_si_003.xls): Cross-references of individual residue Ds; the one-letter prefix defines the chain name. Table S3 (jp310753z_si_004.xls): Rank product values calculated over coordinated motion recurrent positions. Table S4 (jp310753z_si_005.xls): Rank product values calculated over interaction energy recurrent positions. Table S5 (jp310753z_si_006.xls): Total energies of the MD simulations. Table S6 (jp310753z_si_007.xls): List of hot spots extracted from recurrent position profiles of IERP and CMRP. Table S7 (jp310753z_si_008.xls): Average bending angles between variable and constant domains in heavy and light chains, respectively. 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.

■ ABBREVIATIONS

Ab, antibody; Ag, antigen; CDR, complementarity determining region; Fab, antigen binding fragment; Fc, crystallizable fragment; Fv, variable domains fragment; IERP, interaction energy recurrent position; Ig, immunoglobulin; MD, molecular dynamics; RMSD, root-mean-square deviation; CMRP, coordinated motion recurrent position; VL, VH, CL, CH1, variable and constant domains of the light and heavy chain

■ REFERENCES

(1) Bahadur, R. P.; Zacharias, M. *Cell. Mol. Life Sci.* **2008**, *65*, 1059–1072.

(2) Chen, S.-W. W.; Van Regenmortel, M. H. V; Pellequer, J.-L. *Curr. Med. Chem.* **2009**, *16*, 953–964.

(3) De Vries, S. J.; Bonvin, A. M. J. J. *Curr. Protein Pept. Sci.* **2008**, *9*, 394–406.

(4) Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644–2676.

(5) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. *Chem. Rev.* **2008**, *108*, 1225–1244.

(6) Reichmann, D.; Rahat, O.; Cohen, M.; Neuvirth, H.; Schreiber, G. *Curr. Opin. Struct. Biol.* **2007**, *17*, 67–76.

(7) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596–6616.

(8) Oda, M.; Uchiyama, S.; Robinson, C. V; Fukui, K.; Kobayashi, Y.; Azuma, T. *FEBS J.* **2006**, *273*, 1476–1487.

(9) Thorpe, I. F.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8821–8826.

(10) Zimmermann, J.; Oakman, E. L.; Thorpe, I. F.; Shi, X.; Abbyad, P.; Brooks, C. L.; Boxer, S. G.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13722–13727.

(11) Jimenez, R.; Salazar, G.; Baldridge, K. K.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 92–97.

(12) Jimenez, R.; Salazar, G.; Yin, J.; Joo, T.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3803–3808.

(13) Murali, R.; Sharkey, D. J.; Daiss, J. L.; Murthy, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12562–12567.

(14) Li, Y.; Li, H.; Smith-Gill, S. J.; Mariuzza, R. A. *Biochemistry* **2000**, *39*, 6296–6309.

(15) Braden, B. C.; Souchon, H.; Eiselé, J. L.; Bentley, G. A.; Bhat, T. N.; Navaza, J.; Poljak, R. J. *J. Mol. Biol.* **1994**, *243*, 767–781.

(16) Graille, M.; Stura, E. A.; Bossus, M.; Muller, B. H.; Letourneur, O.; Battail-Poirot, N.; Sibaï, G.; Gauthier, M.; Rolland, D.; Le Du, M. H.; et al. *J. Mol. Biol.* **2005**, *354*, 447–458.

(17) Sinha, N.; Smith-Gill, S. J. *Curr. Protein Pept. Sci.* **2002**, *3*, 601–614.

(18) Jefferis, R.; Lund, J.; Pound, J. D. *Immunol. Rev.* **1998**, *163*, 59–76.

(19) Torres, M.; Casadevall, A. *Trends Immunol.* **2008**, *29*, 91–97.

(20) Torres, M.; Fernández-Fuentes, N.; Fiser, A.; Casadevall, A. *J. Biol. Chem.* **2007**, *282*, 13917–13927.

(21) Torosantucci, A.; Chiani, P.; Bromuro, C.; De Bernardis, F.; Palma, A. S.; Liu, Y.; Mignogna, G.; Maras, B.; Colone, M.; Stringaro, A.; et al. *PLoS One* **2009**, *4*, No. e5392.

(22) Tudor, D.; Yu, H.; Maupetit, J.; Drillet, A.-S.; Bouceba, T.; Schwartz-Cornil, I.; Lopalco, L.; Tuffery, P.; Bomsel, M. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12680–12685.

(23) Król, M.; Roterman, I.; Piekarska, B.; Konieczny, L.; Rybarska, J.; Stopa, B.; Spólnik, P. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 545–554.

(24) Piekarska, B.; Drozd, A.; Konieczny, L.; Król, M.; Jurkowski, W.; Roterman, I.; Spólnik, P.; Stopa, B.; Rybarska, J. *Chem. Biol. Drug Des.* **2006**, *68*, 276–283.

(25) Oda, M.; Kozono, H.; Morii, H.; Azuma, T. *Int. Immunol.* **2003**, *15*, 417–426.

(26) Röthlisberger, D.; Honegger, A.; Plückthun, A. *J. Mol. Biol.* **2005**, *347*, 773–789.

(27) Cooper, L. J.; Shikhman, A. R.; Glass, D. D.; Kangisser, D.; Cunningham, M. W.; Greenspan, N. S. *J. Immunol.* **1993**, *150*, 2231−2242.
(28) Kato, K.; Matsunaga, C.; Odaka, A.; Yamato, S.; Takaha, W.; Shimada, I.; Arata, Y. *Biochemistry* **1991**, *30*, 6604−6610.
(29) Pritsch, O.; Magnac, C.; Dumas, G.; Bouvet, J. P.; Alzari, P.; Dighiero, G. *Eur. J. Immunol.* **2000**, *30*, 3387−3395.
(30) Colombo, G.; Meli, M.; Morra, G.; Gabizon, R.; Gasset, M. *PLoS One* **2009**, *4*, No. e4296.
(31) Morra, G.; Verkhivker, G.; Colombo, G. *PLoS Comput. Biol.* **2009**, *5*, No. e1000323.
(32) Torella, R.; Moroni, E.; Caselle, M.; Morra, G.; Colombo, G. *BMC Struct. Biol.* **2010**, *10*, 42.
(33) Meli, M.; Gasset, M.; Colombo, G. *PLoS One* **2011**, *6*, No. e19093.
(34) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. *Cell* **2011**, *147*, 1564−1575.
(35) Kurkcuoglu, Z.; Bakan, A.; Kocaman, D.; Bahar, I.; Doruker, P. *PLoS Comput. Biol.* **2012**, *8*, 11.
(36) Morra, G.; Potestio, R.; Micheletti, C.; Colombo, G. *PLoS Comput. Biol.* **2012**, *8*, No. e1002433.
(37) Pagano, K.; Torella, R.; Foglieni, C.; Bugatti, A.; Tomaselli, S.; Zetta, L.; Presta, M.; Rusnati, M.; Taraboletti, G.; Colombo, G.; et al. *PLoS One* **2012**, *7*, No. e36990.
(38) Kamberaj, H.; Van der Vaart, A. *Biophys. J.* **2009**, *97*, 1747−1755.
(39) Chailyan, A.; Marcatili, P.; Tramontano, A. *FEBS J.* **2011**, *278*, 2858−2866.
(40) Morra, G.; Colombo, G. *Proteins: Struct., Funct., Bioinf.* **2008**, *72*, 660−672.
(41) Scarabelli, G.; Morra, G.; Colombo, G. *Biophys. J.* **2010**, *98*, 1966−1975.
(42) Fiser, A.; Do, R. K.; Sali, A. *Protein Sci.* **2000**, *9*, 1753−1773.
(43) Li, Y.; Li, H.; Yang, F.; Smith-Gill, S. J.; Mariuzza, R. A. *Nat. Struct. Mol. Biol.* **2003**, *10*, 482−488.
(44) Chen, Y.; Wiesmann, C.; Fuh, G.; Li, B.; Christinger, H. W.; McKay, P.; De Vos, A. M.; Lowman, H. B. *J. Mol. Biol.* **1999**, *293*, 865−881.
(45) Cauerhff, A.; Goldbaum, F. A.; Braden, B. C. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3539−3544.
(46) Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(47) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; Van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(48) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(49) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(50) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(51) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(52) Abhinandan, K. R.; Martin, A. C. R. *Mol. Immunol.* **2008**, *45*, 3832−3839.
(53) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. *Proteins: Struct., Funct., Bioinf.* **2006**, *64*, 559−574.
(54) Sun, P. D.; Boyington, J. C. *Curr. Protoc. Immunol.* **2001**, *Appendix 1*, Appendix 1N.
(55) Mollica, L.; Morra, G.; Colombo, G.; Musco, G. *Chem. Asian J.* **2011**, *6*, 1171−1180.
(56) Tiana, G.; Simona, F.; De Mori, G. M. S.; Broglia, R. A.; Colombo, G. *Protein Sci.* **2004**, *13*, 113−124.
(57) Colacino, S.; Tiana, G.; Broglia, R. A.; Colombo, G. *Proteins: Struct., Funct., Bioinf.* **2006**, *62*, 698−707.
(58) Colacino, S.; Tiana, G.; Colombo, G. *BMC Struct. Biol.* **2006**, *6*, 17.
(59) Ragona, L.; Colombo, G.; Catalano, M.; Molinari, H. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 366−376.
(60) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(61) Morra, G.; Baragli, C.; Colombo, G. *Biophys. Chem.* **2010**, *146*, 76−84.
(62) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Luo, R.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 3986−3994.
(63) Genoni, A.; Morra, G.; Merz, K. M.; Colombo, G. *Biochemistry* **2010**, *49*, 4283−4295.
(64) Genoni, A.; Morra, G.; Colombo, G. *J. Phys. Chem. B* **2012**, *116*, 3331−3343.
(65) Breitling, R.; Armengaud, P.; Amtmann, A.; Herzyk, P. *FEBS Lett.* **2004**, *573*, 83−92.
(66) Jurman, G.; Merler, S.; Barla, A.; Paoli, S.; Galea, A.; Furlanello, C. *Bioinformatics* **2008**, *24*, 258−264.
(67) Monwar, M. M.; Gavrilova, M. L. *IEEE Trans. Syst. Man Cybern. B. Cybern.* **2009**, *39*, 867−878.
(68) Koziol, J. A. *FEBS Lett.* **2010**, *584*, 4481−4484.
(69) Smith, D. K.; Xue, H. *J. Mol. Biol.* **1997**, *274*, 530−545.
(70) Feige, M. J.; Hendershot, L. M.; Buchner, J. *Trends Biochem. Sci.* **2010**, *35*, 189−198.
(71) Pritsch, O.; Hudry-Clergeon, G.; Buckle, M.; Petillot, Y.; Bouvet, J. P.; Gagnon, J.; Dighiero, G. *J. Clin. Invest.* **1996**, *98*, 2235−2243.
(72) Takahashi, H.; Tamura, H.; Shimba, N.; Shimada, I.; Arata, Y. *J. Mol. Biol.* **1994**, *243*, 494−503.
(73) Burton, D. R. *Mol. Immunol.* **1985**, *22*, 161−206.
(74) Løset, G. A.; Roux, K. H.; Zhu, P.; Michaelsen, T. E.; Sandlie, I. *J. Immunol.* **2004**, *172*, 2925−2934.
(75) Harris, L. J.; Larson, S. B.; Skaletsky, E.; McPherson, A. *Immunol. Rev.* **1998**, *163*, 35−43.
(76) Harris, L. J.; Skaletsky, E.; McPherson, A. *J. Mol. Biol.* **1998**, *275*, 861−872.
(77) Roux, K. H.; Strelets, L.; Michaelsen, T. E. *J. Immunol.* **1997**, *159*, 3372−3382.
(78) Sandin, S.; Ofverstedt, L.-G.; Wikström, A.-C.; Wrange, O.; Skoglund, U. *Structure* **2004**, *12*, 409−415.
(79) Bongini, L.; Fanelli, D.; Piazza, F.; De Los Rios, P.; Sandin, S.; Skoglund, U. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6466−6471.
(80) Wilson, I. A.; Stanfield, R. L. *Curr. Opin. Struct. Biol.* **1994**, *4*, 857−867.
(81) Padlan, E. A. *Adv. Protein Chem.* **1996**, *49*, 57−133.
(82) Prasad, L.; Vandonselaar, M.; Lee, J. S.; Delbaere, L. T. *J. Biol. Chem.* **1988**, *263*, 2571−2574.
(83) Rini, J. M.; Stanfield, R. L.; Stura, E. A.; Salinas, P. A.; Profy, A. T.; Wilson, I. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6325−6329.
(84) Harris, L. J.; Larson, S. B.; Hasel, K. W.; McPherson, A. *Biochemistry* **1997**, *36*, 1581−1597.
(85) Sotriffer, C. A.; Rode, B. M.; Varga, J. M.; Liedl, K. R. *Biophys. J.* **2000**, *79*, 614−628.
(86) Sotriffer, C. A.; Liedl, K. R.; Linthicum, D. S.; Rode, B. M.; Varga, J. M. *J. Mol. Biol.* **1998**, *278*, 301−306.
(87) Lim, K.; Herron, J. N. *Biochemistry* **1995**, *34*, 6962−6974.
(88) Suhre, K.; Sanejouand, Y.-H. *Nucleic Acids Res.* **2004**, *32*, W610-4.
(89) Kroon, G. J. A.; Mo, H.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **2003**, *12*, 1386−1394.
(90) Lim, K.; Jameson, D. M.; Gentry, C. A.; Herron, J. N. *Biochemistry* **1995**, *34*, 6975−6984.
(91) Lange, O. F.; Grubmüller, H. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1294−1312.
(92) Honegger, A.; Plückthun, A. *J. Mol. Biol.* **2001**, *309*, 687−699.
(93) Kipriyanov, S. M.; Moldenhauer, G.; Martin, A. C.; Kupriyanova, O. A.; Little, M. *Protein Eng.* **1997**, *10*, 445−453.
(94) De Haard, H. J.; Kazemier, B.; Van der Bent, A.; Oudshoorn, P.; Boender, P.; Van Gemen, B.; Arends, J. W.; Hoogenboom, H. R. *Protein Eng.* **1998**, *11*, 1267−1276.
(95) Williams, D. C.; Rule, G. S.; Poljak, R. J.; Benjamin, D. C. *J. Mol. Biol.* **1997**, *270*, 751−762.

(96) Feige, M. J.; Groscurth, S.; Marcinowski, M.; Shimizu, Y.; Kessler, H.; Hendershot, L. M.; Buchner, J. *Mol. Cell* **2009**, *34*, 569−579.
(97) Kabat, E. A.; Padlan, E. A.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2785−2788.
(98) Adachi, M.; Kurihara, Y.; Nojima, H.; Takeda-Shitaka, M.; Kamiya, K.; Umeyama, H. *Protein Sci.* **2003**, *12*, 2125−2131.
(99) Lee, Y. K.; Brewer, J. W.; Hellman, R.; Hendershot, L. M. *Mol. Biol. Cell* **1999**, *10*, 2209−2219.
(100) Feige, M. J.; Groscurth, S.; Marcinowski, M.; Yew, Z. T.; Truffault, V.; Paci, E.; Kessler, H.; Buchner, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13373−13378.
(101) Ewert, S.; Huber, T.; Honegger, A.; Plückthun, A. *J. Mol. Biol.* **2003**, *325*, 531−553.
(102) Vargas-Madrazo, E.; Paz-García, E. *J. Mol. Recognit.* **2003**, *16*, 113−120.
(103) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. *J. Phys. Chem. B* **2010**, *114*, 6614−6624.
(104) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11937−11942.
(105) Zimmermann, M. T.; Skliros, A.; Kloczkowski, A.; Jernigan, R. L. *Immunome Res.* **2011**, *7*, 5.