A Structure Based Pattern Recognition on Antibodies

Dario CORRADA [1], Giulia MORRA [1], Giorgio COLOMBO [1]

[1] Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, via privata Mario Bianco, 9 - 20131 Milan (Italy)

dario.corrada@icrm.cnr.it

Istituto di Chimica del Riconoscimento Molecolare

Motivation

Protein-protein interactions are deeply involved in the antibody::antigen recognition process. Crystallographic data offer evidence of conformational changes between *apo* and *holo* forms of antibodies complexed with antigens. Nevertheless, the dynamical aspects of intermolecular relationships still remain a challenging issue. Extensive molecular dynamics (MD) simulations offer the suitable tool for generating statistical ensembles of conformations from which various energetic, structural and dynamic properties can be collected. The interaction energy correlations between all residue pairs can be investigated in order to find relevant regions involved in the fold stability; furthermore, a global overview of these sites can highlight preferential signaling pathways along the protein structures. **In the present work, we will describe those conformational change events which derive from the formation of antibody::antigen complex. In particular, we will define those pathways that start from the paratope region and propagate through the immunoglobulin domains.**

Methods

We have taken into account a dataset of 28 Fab::antigen complexes, whose structures have been deposited at the Protein Data Bank. For most of them, 50ns unrestrained MD simulations were computed. Seven structures targeting the same epitopes and sharing similar paratopes were submitted to 200ns MD simulations. Each case of the dataset was duplicated, considering two systems: the isolated Fab structure (*apo* form) and the complex (*holo* form). The GROMACS 4.0.7 program with the *Gromos96 ffG43a1* forcefield have been used for the simulations. The temperatures of proteins and solvent were separately coupled, through a Berendsen thermostat, to a bath with temperature 300K and time constant 0.2ps. Isotropic pressure coupling of the systems was based on the Berendsen weak coupling algorithm, with time constant 1ps and compressibility 4.6 X 10^{-5} bar^{-1}. **The structures sampled from the trajectories were clustered with the purpose of defining the conformational space explored.** Long simulations better scan the conformational space; indeed, the 200ns simulations share a higher number of clusters in respect of 50ns ones (Fig. 1B). Nevertheless, the sample size of the most populated clusters seems not to be affected (Fig. 1A).

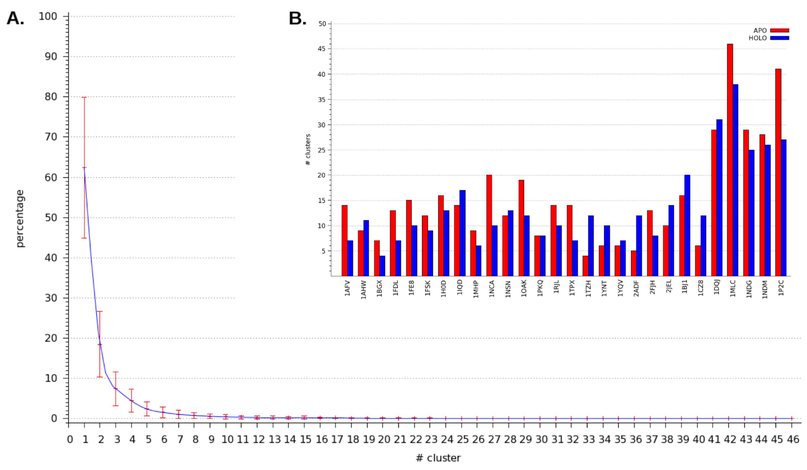

Figure 1: structural clustering along MD simulations. *A. Portion of those structures which belong to the same cluster; clusters are numbered according to their sample sizes. B. Amount of the clusters found from each MD simulation.*

From the most representative structures (medoids) of the most populated clusters energy interaction matrices were calculated, considering all the non-bonded interactions between each residue pair. Principal component analyses were performed over such matrices (Energy Decomposition Analysis)[1]. Since Fabs are composed of two chains of multidomain proteins, we have considered the group of those eigenvectors which bring together sufficient information to describe the stabilization energy of the whole system, according to [2]. Every component of the selected eigenvectors renders the relative contribution of each residue to the overall stabilization energy, therefore **we selected those components whose values are higher than a threshold value which depends on the number of the residues in the protein** (see also header figure at *Motivation*)[3]. All the residues were mapped to a common reference, obtained from joining the Chothia numbering schema [4] with structural multiple alignment based schema [5], for variable and constant domains respectively. Finally, **we mapped the energetic relevant residues of each case to the numbering schema** described above, in order to summarize a common residue pattern referred to a generic Fab structure. We then collected the occurrencies for each position, and we termed them as *Interaction Energy Recurrent Positions* (IERPs).

Results

The results obtained herein are intended to identify those interactions that define the formation of antibody::antigen complexes. The correlated motions are also investigated, reducing the complex protein dynamics to its essential degrees of freedom. The analysis of global distance fluctuation matrices [6] shows that the motions of residues belonging to the same domain appear more correlated with each other, as depicted in Figure 2. **The constant domains of the heavy chain (CH) show a higher mobility, with respect to the remaining protein.** The CH domains are close to the boundary between Fab and Fc fragments; hence, CH domains may benefit from a greater degree of unrestrained motion.

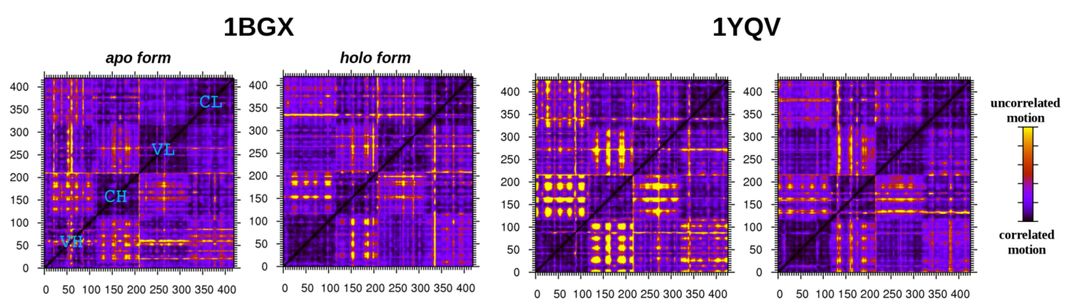

Figure 2: global distance fluctation matrices. *The color ramp is related to the values of distance fluctuation that is the standard deviation of the interatomic distances measured for each couple of* C_{alpha}*, along the entire MD simulation.*

From the interaction energy analyses, we collected three subset of IERPs: the first ones are extracted only from *apo* forms simulations, the second ones are typical of *holo* forms, the third ones appear as relevant in both forms (Figure 3). The distribution of these subsets, along the Fab polypeptide chains, indicates that *apo* and *holo* forms show differerential IERPs profiles. Then, we compared the distribution of IERPs along the four immunoglobulin domains VH, VL, CH and CL (Kolmogorov-Smirnov test, p-value ≤ 0.05). Globally, **the *apo* forms show a lesser amount of IERPS, indicating that a stronger network of non-bonded interaction arise upon the antigen binding.**

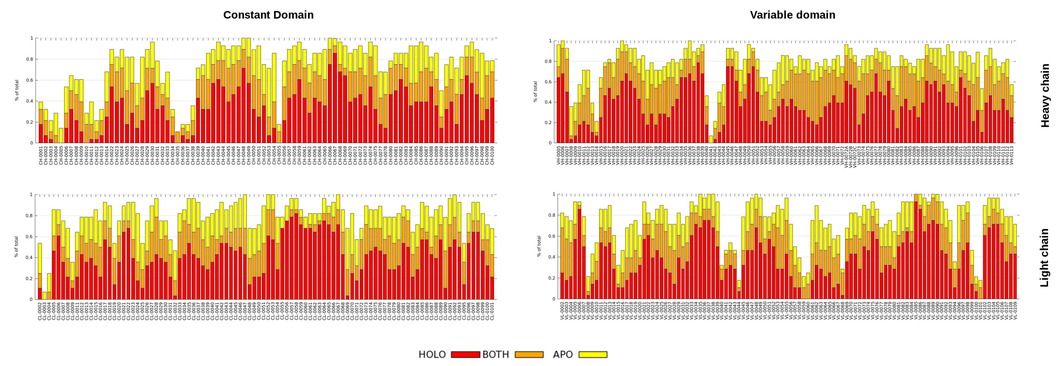

Figure 3: IERPs profile along Ig domains. *The stacked histograms show the occurrencies by which every IERP is found in the reference structure (see "Methods") from each MD simulation performed over the PDB dataset considered in this work.*

The visual inspection of the structures of the dataset illuminates the spatial distribution of IERPs. **In the antigen-bound Fabs (*holo* forms) the selected residues overlap with the framework regions of VH domains and propagate downstream to the CH domains** (Figure 4, red spheres). On the other hand, those IERPs that are more relevant for *apo* forms map preferentially towards the light chain (blue spheres). A consistent portion of IERPs are shared between *apo* and *holo* forms (green spheres), defining an ensemble of "housekeeping" hotspots, probably needed for the stabilization of the Fab structure itself. In particular, a subset of shared IERPs map to two small helices, located at the basement of the CL domains (Figure 4, at the bottom of *Light chain side* view). This region is known to be critical for the correct folding of Ig proteins [7].

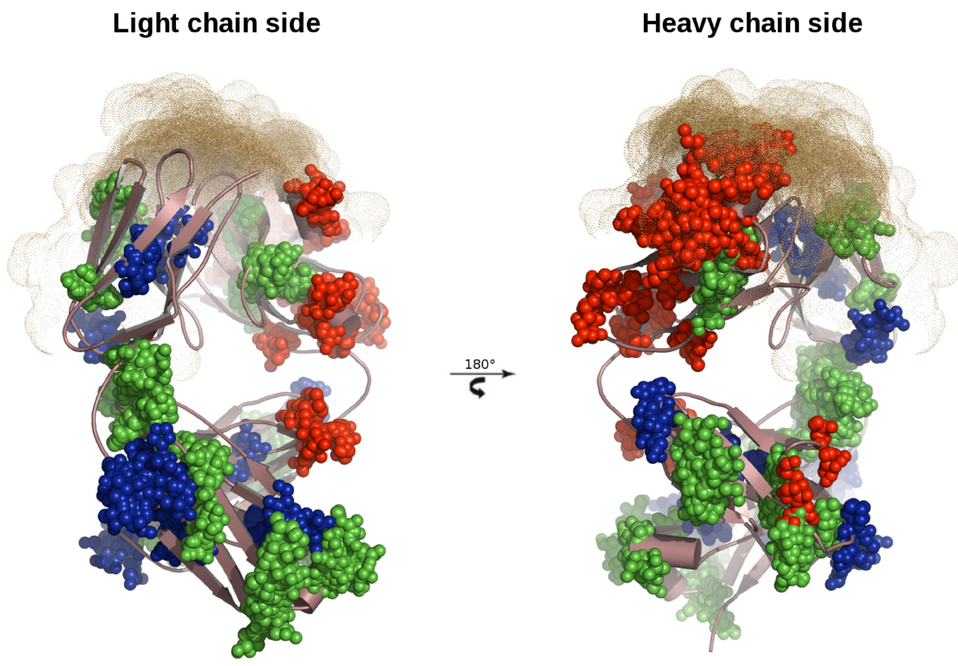

Figure 4: differential IERPs over Fab structures. *In spheres are showed those IERPs which are relevant in apo, holo or both forms (in blue, red and green respectively), mapped along the 28 PDBs of the dataset. For the sake of clarity a single global Fab structure is showed (in cartoon). Dotted surface highlights the contact region by which the antigens bind.*

References

[1] Tiana G, Simona F, De Mori GM, Broglia RA, Colombo G. Understanding the determinants of stability and folding of small globular proteins from their energetics. Protein Sci. 2004 Jan;13(1):113-24.

[2] Genoni A, Morra G, Colombo G. Identification of domains in protein structures from the analysis of intramolecular interactions. J Phys Chem B. 2012 Mar 15;116(10):3331-43.

[3] Colacino S, Tiana G, Colombo G. Similar folds with different stabilization mechanisms: the cases of Prion and Doppel proteins. BMC Struct Biol. 2006 Jul 21;6:17.

[4] Abhinandan KR, Martin AC. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. Mol Immunol. 2008 Aug;45(14):3832-9.

[5] Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: a multiple structural alignment algorithm. Proteins. 2006 Aug 15;64(3):559-74.

[6] 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 Mar;8(3):e1002433.

[7] Feige MJ, Hendershot LM, Buchner J. How antibodies fold. Trends Biochem Sci. 2010 Apr;35(4):189-98.

We gratefully acknowledge "Fondazione Cariplo: PROMOTING INNOVATIVE APPROACHES TO VACCINES DEVELOPMENT AGAINST INFECTIOUS DISEASES" for funding through a grant to the project: **GTA. From Genome to Antigen: a Multidisciplinary Approach towards the Development of an Effective Vaccine Against Burkholderia pseudomallei, the Etiological Agent of Melioidosis**