Protein Structure and Function

A structure based pattern recognition on antibodies

D. Corrada✉, G. Morra, G. Colombo

Institute of Chemistry of Molecular Recognition, National Research Council, Milano, Italy

Motivations

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 24 Fab::antigen complexes, whose structures have been deposited at the Protein Data Bank. For most of them, 50ns unrestrained MD simulations were computed, with a 2fs time-step. Three structures targeting the same epitope 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). An amount of 3,300ns of MD simulations was performed. The time-curves of RMSD on backbone atoms were used to monitor system equilibration. For every simulation the first 10ns were discarded for the further analyses. The structures sampled from the trajectories were clustered with the purpose of defining the conformational space explored. From the most representative structures (medoids) of the most populated clusters energy interaction matrices were calculated, consider-ing all the non-bonded interactions between each residue pairs. Principal component analyses were performed over such matrices (Energy Decomposition Analysis), and only the relevant eigenvectors were considered. Since every component renders the relative contribution of each residue to the overall stabilization energy, we selected those components whose values are higher than a threshold value which depends on the number of the residues in the protein. All the residues were mapped to a common reference, obtained from joining the Chothia numbering schema with structural multiple alignment based schema, for variable and constant domains respectively. Finally, we mapped the energetic relevant residues of each case to the numbering schema depicted 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 at identifying 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 analyses of global distance fluctuation matrices show that the motions of residues belonging to the same domain appear more correlated with each other. 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. 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. The distribution of these subsets along the polypeptide chains of the Fab structures indicates that heavy chains contain most of the shared IERPs; in contrast, it seem that apo and holo forms show

Protein Structure and Function

differerential IERPs profiles along the sequence of light chains. Then, we compared the distribution of IERPs, typical of apo and holo forms, along the four Ig domains VH, VL, CH and CL. In holo forms CL domains show lesser amount of IERPs, while CH domains appear enriched; variable domains (VH and VL) do not show significant differences between apo and holo. In holo form, CL domains still mantain few sparse IERPs that could be considered as local stabilizing hotspots. 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 starts from the framework regions of variable domains and propagate downstream, preferentially along the CH domain. Finally, the signal arrive at the base of costant domains, focusing around the C-terminal regions. On the opposite side unbound Fabs (apo forms) do not show a preferential pattern.