SUPPORTING INFORMATION

Supporting Methods

MD simulation setup

The crystallographic structure of the $PDZ2^{APO}$ (PDB ID: 3LNX) and $PDZ2^{PEP}$ (PDB ID:3LNY) were used as starting structures for unrestrained, classical MD simulations. The input structures were prepared through the Prepwizard module of the Maestro molecular modeling suite (Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012). Optimization of hydrogen-bonding network as well as predictions of the protonation state of titratable groups at pH 7.4 were performed through the PROPKA software.[1] The N- and C-terminal of the PDZ2 domain were neutralized by adding Acetyl- and N-methyl-capping groups, respectively. The N-terminal of the RA-GEF2 peptide was acetylated as well.

The GROMACS v4.5.5 simulation package[2] was employed together with the AMBER99SB-ILDN all atoms force field[3] and the TIP3P water model to describe the solvent. Depending on the dimensions of the systems, a variable number of Na+ and Cl- ions placed at optimum electrostatic positions were added in order to neutralize the system. In detail, the systems included: 6318 water molecules, 12 Na^+ and 12 Cl^- ions for $PDZ2^{APO}$; 6490 water molecules, 15 Na^+ and 13 Cl^- ions for $PDZ2^{PEP}$. Periodic Boundary Conditions (PBC) were applied using an octahedric box as a unit cell and imposing a minimum distance of 12 Å between the solute and the box boundaries.

All the input crystallographic structures were subjected to energy minimization keeping restricted the positions of main chain atoms and all the Cβ atoms. The systems were then equilibrated at 300 K for 4 ns of backbone-restricted MD simulations. The Particle Mesh Ewald (PME) method was employed to compute the electrostatic interactions. Short range repulsive and attractive interactions were computed using a Lennard-Jones potential with a cutoff of 10 Å. The LINCS algorithm[4] was used to constrain all bond lengths except those in water molecules, allowing for an integration time step of 2 fs by the leap-frog algorithm. The v-rescale thermostat[5] was employed to keep the system at a constant temperature of 300 K,

by using a coupling constant (τ_t) of 0.1 ps. The pressure of the system was kept fixed at 1 atm, using the Berendsen weak coupling algorithm[6] with a coupling constant (τ_p) of 1 ps. The pre-equilibrated systems were then released for 5 ns prior to 20 ns of unrestrained isothermal-isobaric (T = 300K, P = 1 atm) MD simulations. For each system, three independent replicas were carried out by randomizing initial velocities.

References

1. Olsson, M. H., Protein electrostatics and pKa blind predictions; contribution from empirical predictions of internal ionizable residues. *Proteins* 79, 3333-3345.
2. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory. Comput.* **2008,** 4, 435-447.
3. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** 78, 1950-1958.
4. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** 18, 1463-1472.
5. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** 126, 014101.
6. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** 81, 3684-3690.

Supporting Figures

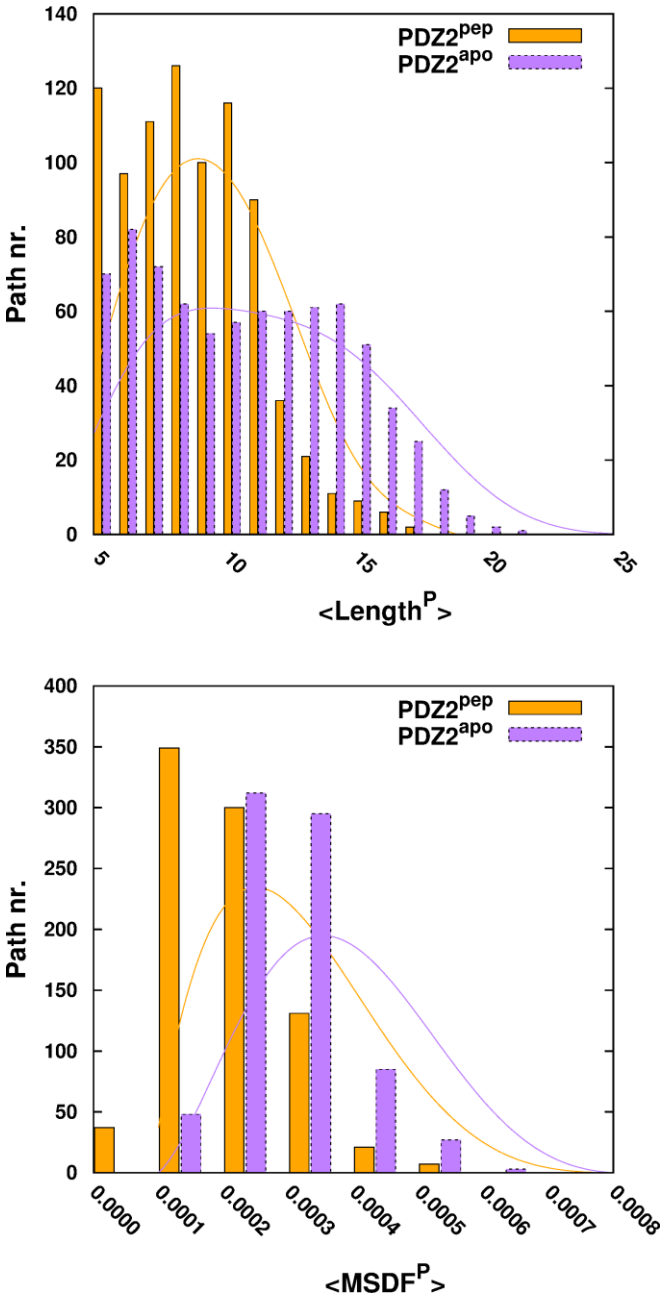

Figure S1. The distributions of the path length and path MSDF averaged over the total paths concerning $PDZ2^{APO}$ (violet) and $PDZ2^{PEP}$ (orange) are shown.

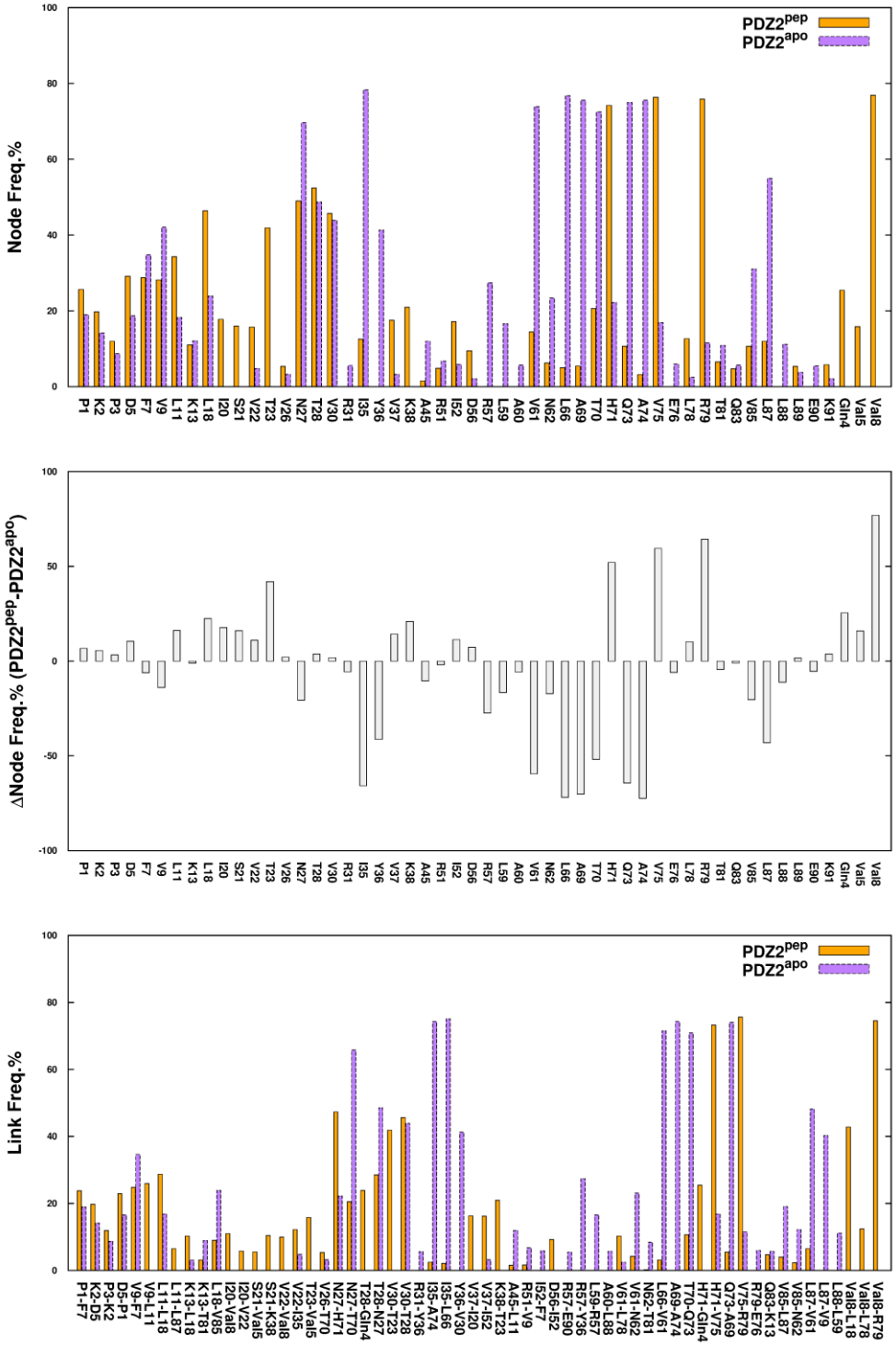

Figure S2. The distributions of nodes (top and middle) and links (bottom) over the total number of paths concerning $PDZ2^{APO}$ (violet) and $PDZ2^{PEP}$ (orange) and achieved by PSN-ENM are shown. Only nodes present in ≥5% of paths (i.e. frequent nodes) and links satisfying both conditions of being present in ≥5% of the paths and of connecting "frequent nodes" are shown. The middle panel shows the node-frequency differences between unbound and bound states. To distinguish between protein and ligand amino acids, the three-letter code is exclusively used for the latter.

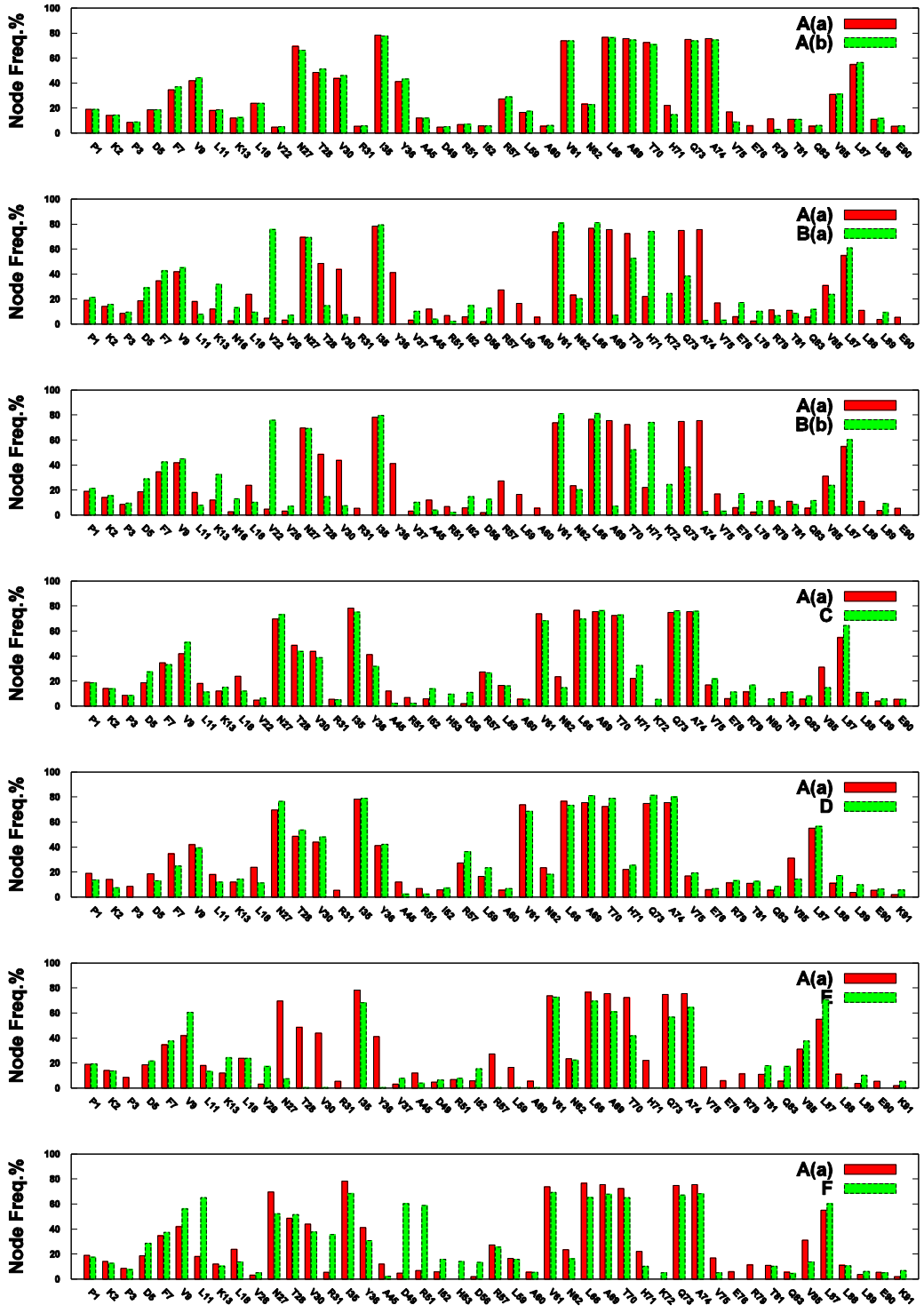

Figure S3. The distributions of those nodes contributing to the global meta paths computed on the alternative chains (i.e. **A**-**F**) or side chains (i.e. only for chains **A** and **B** and indicated as lower-case "**a**" and "**b**") in the 3LNX structure of PDZ2^{APO} are shown. Red concerns the structure the results shown in this study refer to, whereas green concerns the alternative structures found in 3LNX.

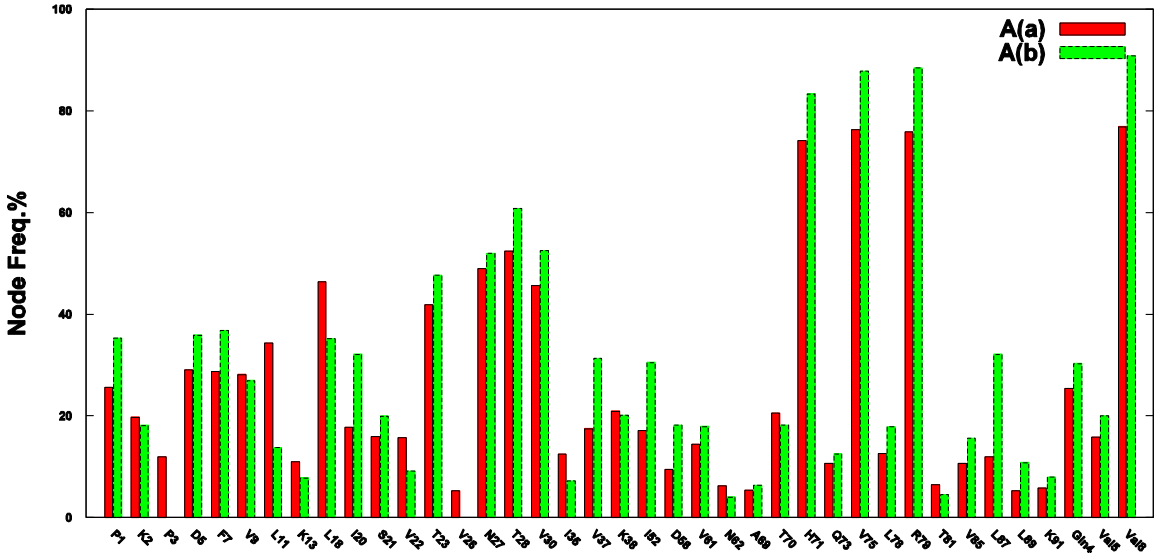

Figure S4. The distributions of those nodes contributing to the global meta paths computed on the alternative side chains (i.e. indicated as lower-case "**a**" and "**b**") in the 3LNY structure of PDZ2^{PEP} are shown. Red concerns the structure the results shown in this study refer to, whereas green concern the alternative structure found in 3LNY.

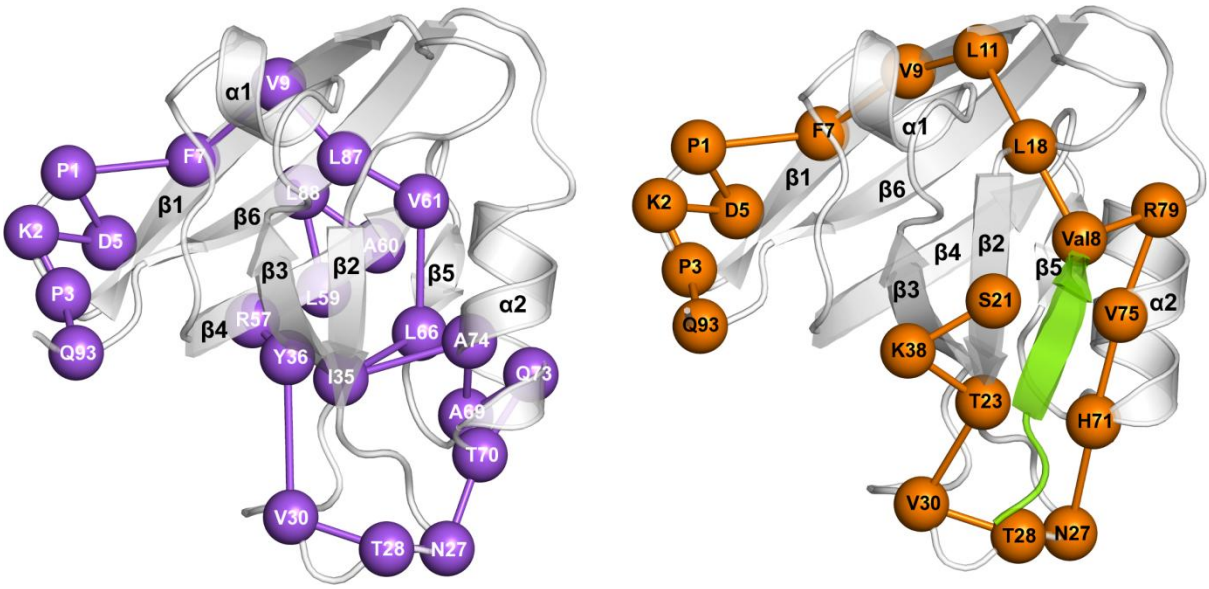

Figure S5. Centers of the first populated cluster obtained for PDZ2^{APO} (left) and PDZ2^{PEP} (right) by clusterization method 1. Violet and orange spheres indicate the unbound and bound states, respectively.

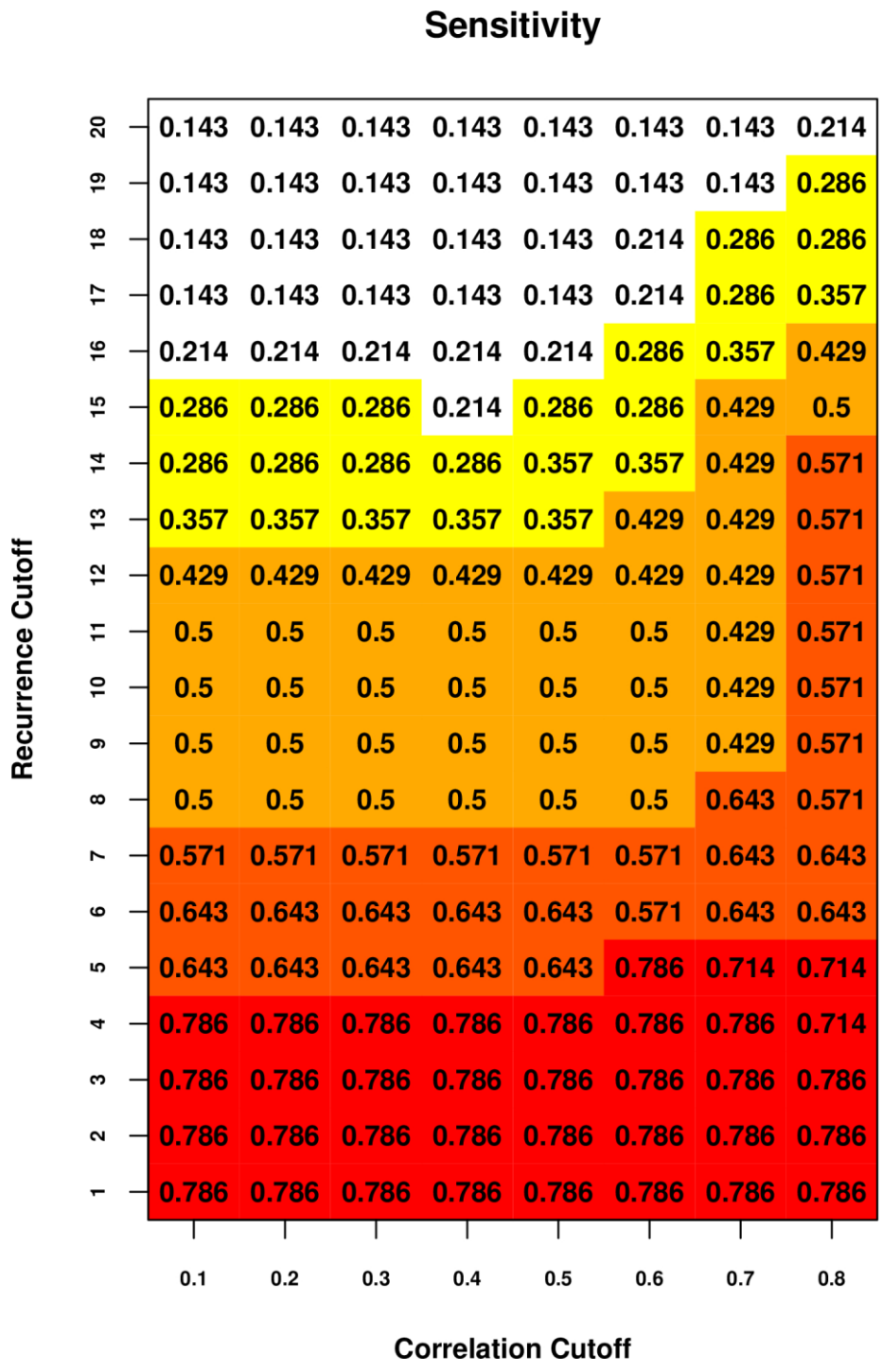

Figure S6. Variation of sensitivity as a function of two cutoffs: the motion correlation coefficient and the recurrence of nodes in the meta paths of the bound form. Sensitivity of results is referred to NMR spin relaxation determinations (Table 2; Fuentes at al., *J Mol Biol* **2004**, *335*, 1105-15).

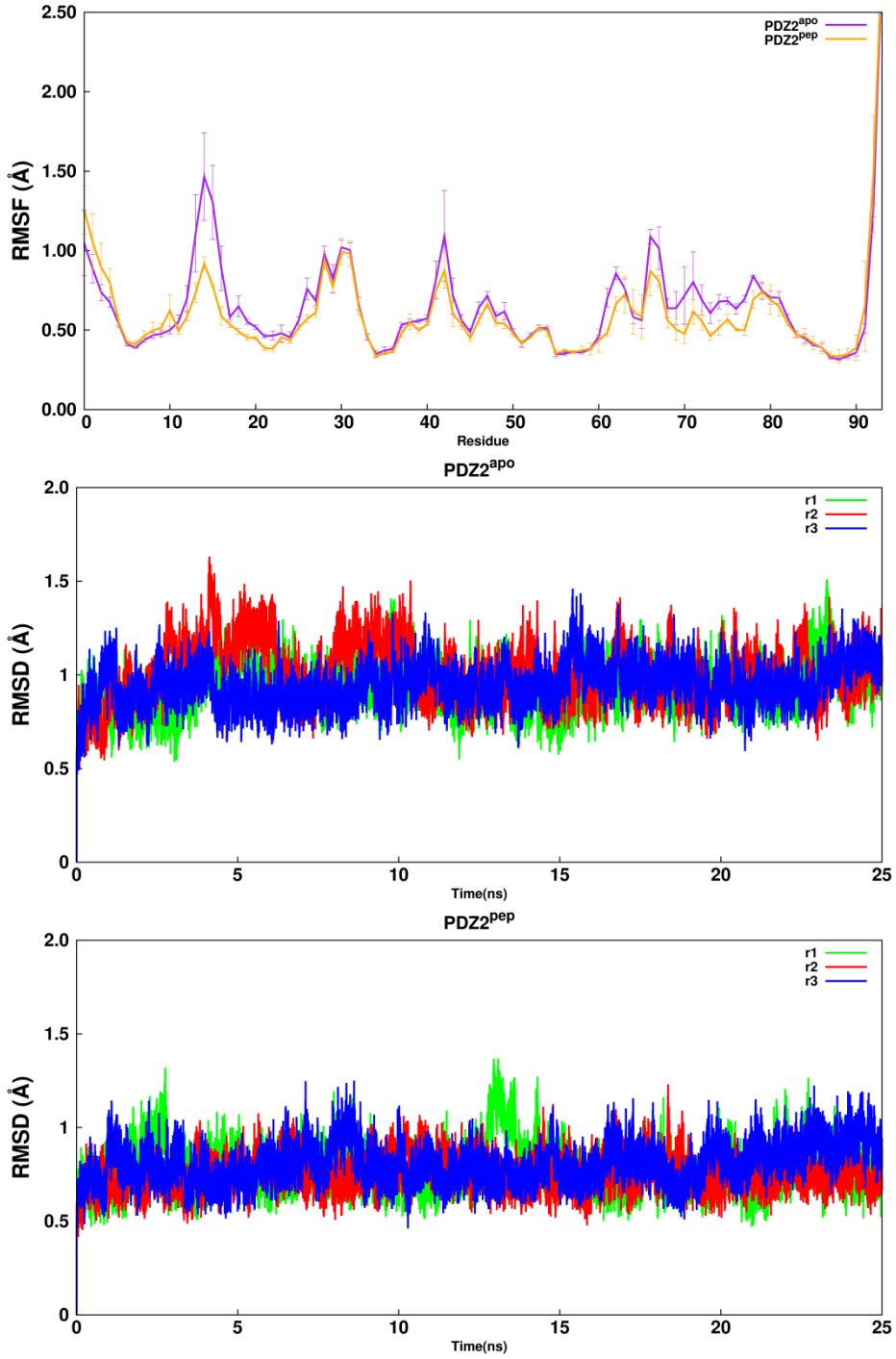

Figure S7. The Cα-RMSF and Cα-RMSD concerning the three MD simulation replicas (r1, r2, and r3) are shown for $PDZ2^{APO}$ and $PDZ2^{PEP}$.

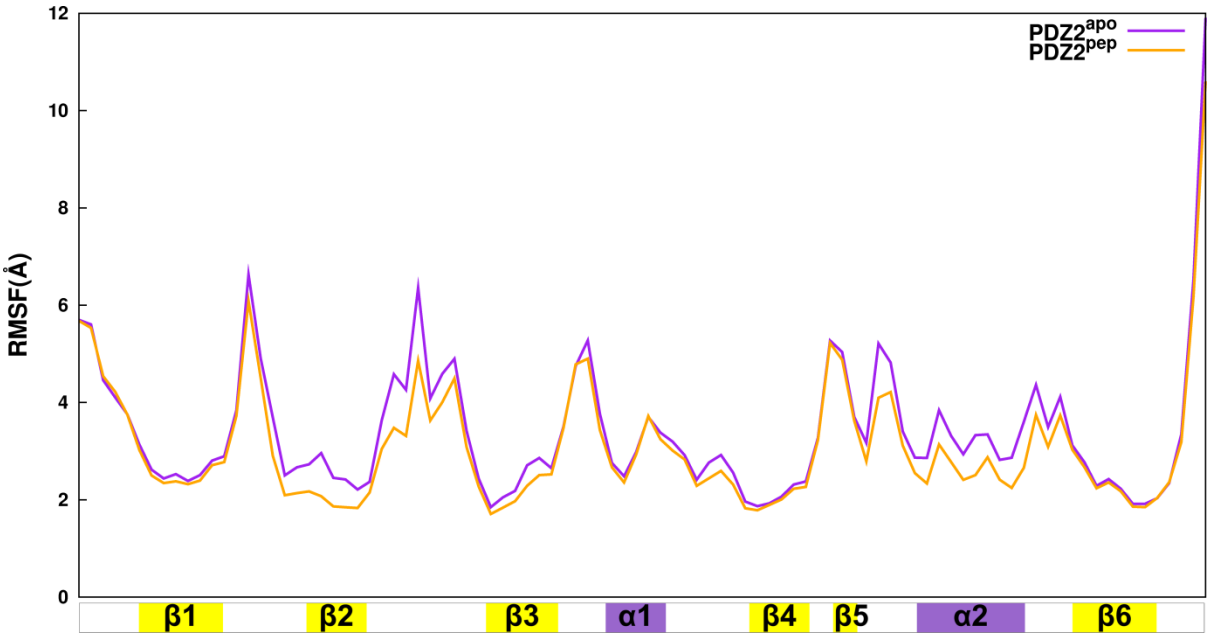

Figure S8. The Cα-RMSF profiles concerning $PDZ2^{APO}$ and $PDZ2^{PEP}$ and inferred from ENM-NMA are shown.

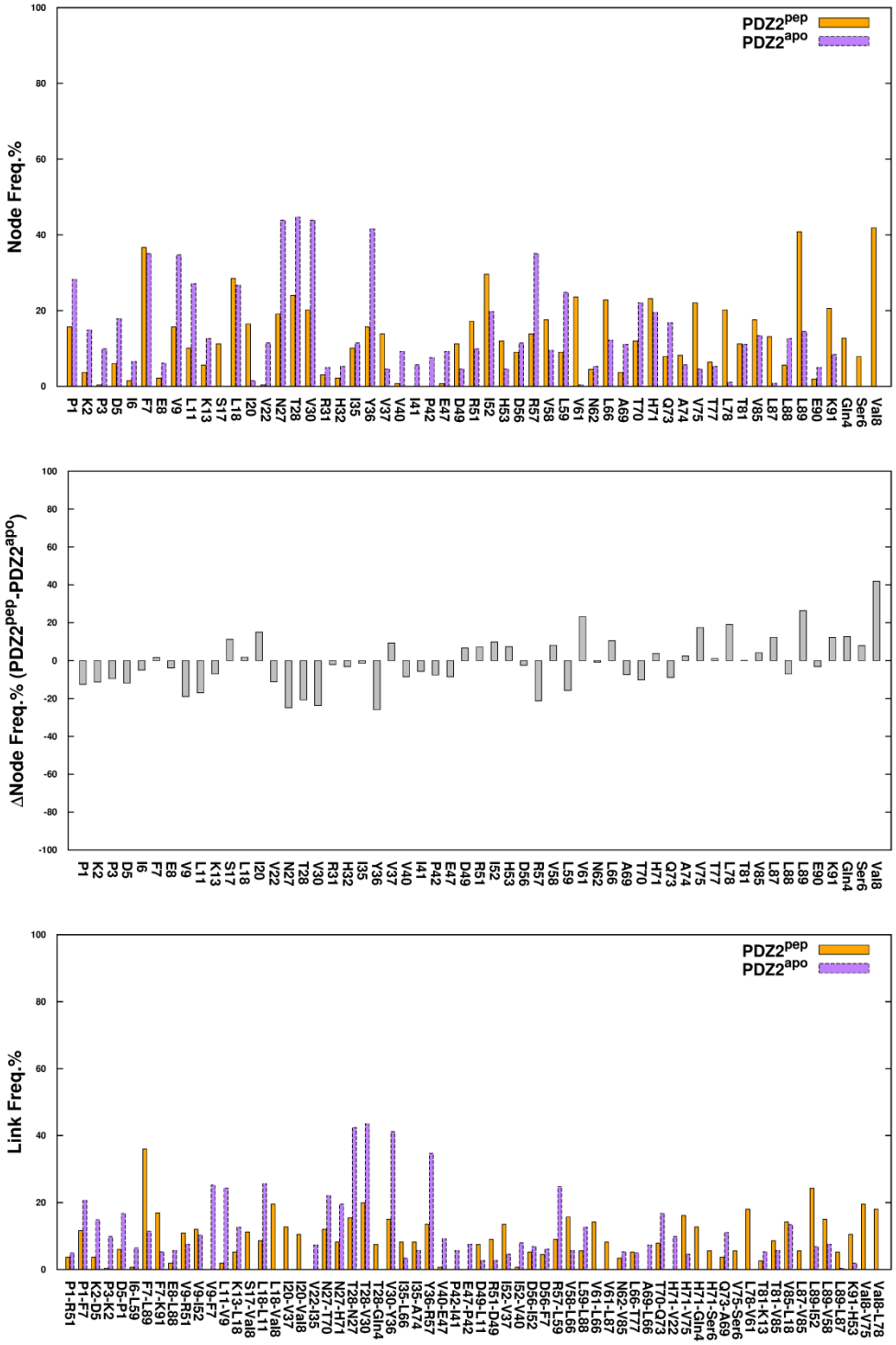

Figure S9. The distributions of nodes (top and middle) and links (bottom) over the total number of paths concerning PDZ2^{APO} (violet) and PDZ2^{PEP} (orange) and achieved by PSN-MD are shown. Only nodes present in ≥10% of paths (i.e. frequent nodes) and links satisfying both conditions of being present in ≥10% of the paths and of connecting "frequent nodes" are shown. The middle panel shows the node-frequency differences between unbound and bound states. To distinguish between protein and ligand amino acids, the three-letter code is exclusively used for the latter. The paths considered for this analysis derive from the merge of the three MD replica-specific path sets.