Sup data 1:

Molecular Dynamics simulations

Molecular Dynamics (MD) simulations were performed using GROMACS 4.0.7 software [1] with the OPLS-AA force-field [2]. L33, P33 and V33 forms of β3 were soaked in a rhombic dodecahedral simulation box with 60,622 TIP3P water molecules and 28 Na^+ ions. The distance between any atom of the protein and the box edges was set to at least 10 Å. The total energy of the system was minimized twice (before and after the addition of the ions) with a steepest descent algorithm. MD simulations were run under the NPT thermodynamic ensemble and periodic boundary conditions were applied in all directions. We used the weak coupling algorithm of Berendsen [3] to maintain the system at a constant physiological temperature of 310 K using a coupling constant of 0.1 ps (protein and water ions separately). Pressure was held constant using the Berendsen algorithm [3] at 1 atm with a coupling constant of 1 ps. Water molecules were kept rigid using the SETTLE algorithm [4]. All other bond lengths were constrained with the LINCS algorithm [5], allowing a 2 fs time step. We used a short-range coulombic and van der Waals cut-off of 10 Å and calculated the long-range electrostatic interactions using the smooth particle mesh Ewald (PME) algorithm [6,7] with a grid spacing of 1.2 Å and an interpolation order of 4. The neighbor list was updated every 10 steps. After a 1 ns equilibration (with position restraints on the protein), each system was simulated for 50 ns. For the three systems, five 50 ns simulations were performed (using different initial velocities) and one was extended until 100 ns for a total simulation time of 300 ns. Molecular conformations were saved every 100 ps for further analysis.

Trajectory analyses

The first 5 ns of each MD simulation were discarded and trajectory analyses were conducted on a set of 2,749 MD snapshots for each system and were performed with the

GROMACS software and in-house Python and R scripts. Root mean square deviations (RMSD) and root mean square fluctuations (RMSF) were calculated on Cα atoms only. To analyze the number of contacts, two residues were defined as contacting each other when the distance between their Cα atoms was less or equal to 8 Å [8]. The absolute accessible surface area (ASA) was computed using GROMACS software. The relative accessible surface area (rASA) was obtained by normalizing the absolute ASA to the ASA calculated on an extended peptide Ala-X-Ala (where X stands for the amino acid in question), in this case 183 $Å^2$, 142 $Å^2$ and 154 $Å^2$ for Leu, Pro and Val respectively. Distances between residue 33 and the I-EGF-1 and I-EGF-2 domains were computed as the distance between the center of mass of the atoms in residue 33, and the I-EGF-1 and I-EGF-2 domains.

Protein Blocks analysis

Protein Blocks (PBs) [9] are a structural alphabet composed of 16 local prototypes. Each specific PB is characterized by the φ, ψ dihedral angles of five consecutive residues. Obtained through an unsupervised training approach performed on a representative non-redundant databank, PBs give a reasonable approximation of all local protein 3D structures [10]. PBs are very efficient in modeling tasks such as protein superimpositions [11] and MD analyses [12].

PB assignments are done for every residue of the β3 protein and for every snapshot extracted from the MD simulations. The equivalent number of PBs [10] (N_{eq}) is a statistical measurement similar to entropy and represents the average number of PBs a given residue takes. N_{eq} is calculated as follows:

$$N_{eq} = exp\left(-\sum_{x=1}^{1} f_x \ln f_x\right)$$

Where f_x is the probability of PB *x*. A N_{eq} value of 1 indicates that only one type of PB is observed, while a value of 16 is equivalent to a random distribution. To underline the main differences between the two proteins, ΔN_{eq} values corresponding to the absolute difference between N_{eq} from L33 and P33 simulations were computed. Differences were considered significant when at least one of the N_{eq} was less than 2.5 and ΔN_{eq} was greater than 1.5.

Reference List

1. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4: 435-447.

2. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS All-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118: 11225-11236.

3. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81: 3684-3691.

4. Miyamoto S, Kollman PA (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13: 952-962.

5. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18: 1463-1472.

6. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N . log(N) method for Ewald sums in large systems. J Chem Phys 98: 10089-10092.

7. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L (1995) A smooth particle mesh Ewald method. J Chem Phys 103: 8577-8593.

8. Faure G, Bornot A, de Brevern AG (2008) Protein contacts, inter-residue interactions and side-chain modelling. Biochimie 90: 626-639. S0300-9084(07)00316-1 [pii];10.1016/j.biochi.2007.11.007 [doi].

9. Joseph AP, Agarwal G, Mahajan S, Gelly JC, Swapna LS, Offmann B, Cadet F, Bornot A, Tyagi M, Valadie H, Schneider B, Etchebest C, Srinivasan N, de Brevern AG (2010) A short survey on protein blocks. Biophys Rev 2: 137-147. 10.1007/s12551-010-0036-1 [doi].

10. de Brevern AG, Etchebest C, Hazout S (2000) Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins 41: 271-287. 10.1002/1097-0134(20001115)41:3<271::AID-PROT10>3.0.CO;2-Z [pii].

11. Joseph AP, Srinivasan N, de Brevern AG (2012) Progressive structure-based alignment of homologous proteins: Adopting sequence comparison strategies. Biochimie . S0300-9084(12)00216-7 [pii];10.1016/j.biochi.2012.05.028 [doi].

12. de Brevern AG, Wong H, Tournamille C, Colin Y, Le Van KC, Etchebest C (2005) A structural model of a seven-transmembrane helix receptor: the Duffy antigen/receptor for chemokine (DARC). Biochim Biophys Acta 1724: 288-306. S0304-4165(05)00159-5 [pii];10.1016/j.bbagen.2005.05.016 [doi].

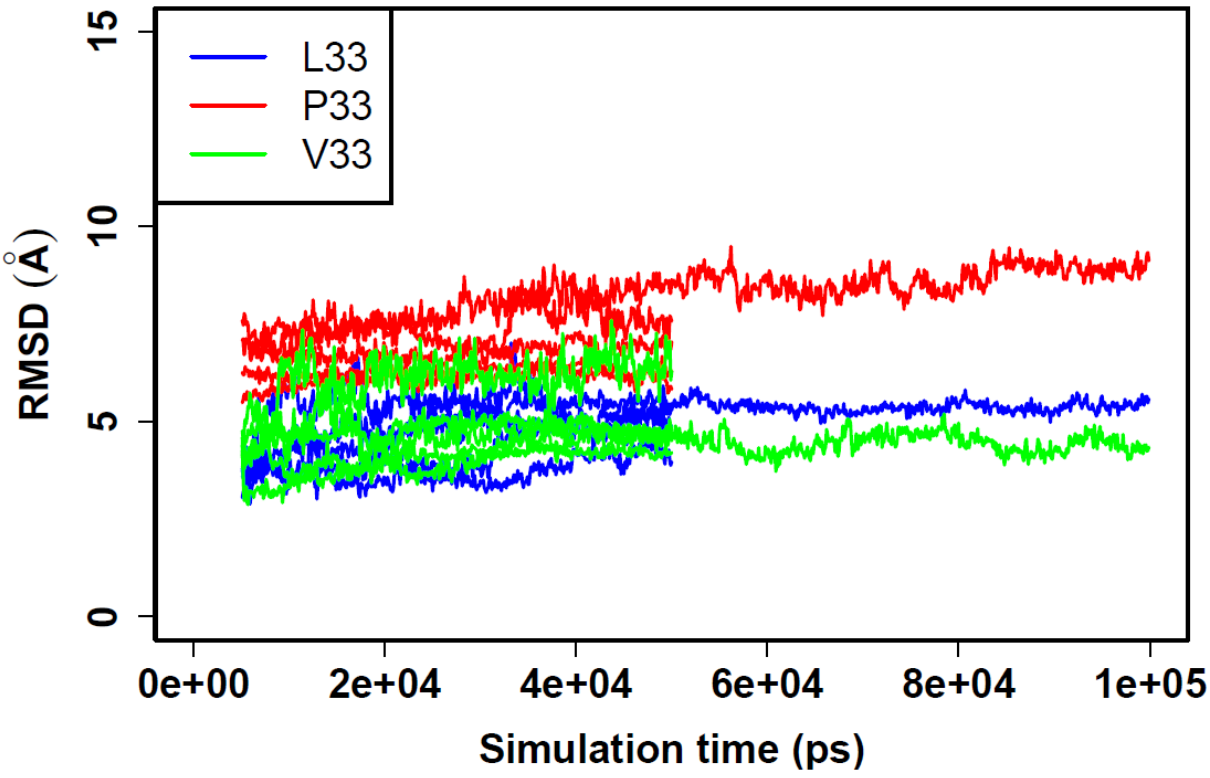

Figure S1. RMSD of the PSI, I-EGF-1 and I-EGF-2 domains. Root-mean-square deviations (RMSD) calculated for Cα atoms are individually presented for the four MD simulations of 50 ns and the fifth of 100 ns for the L33-β3, P33-β3 and V33-β3 (blue, red and green lines, respectively). The first 5 ns of each simulation, i.e. the time to reach stability, were omitted. RMSD values were stable throughout the MD simulations.

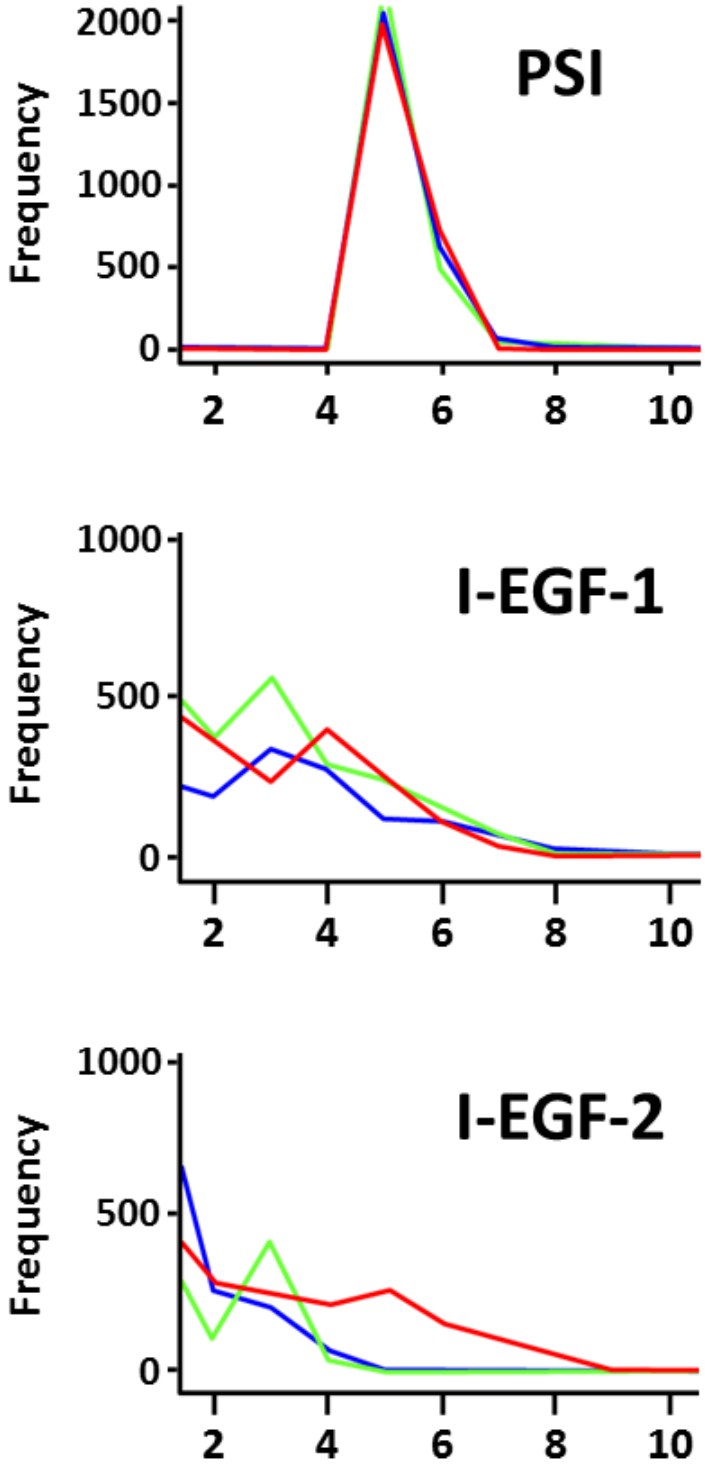

Figure S2. Contacts of residue 33 with β3 knee domains. Cα contact of L33, V33 and P33 (blue, green and red curves respectively) with atoms of the PSI, I-EGF-1 and I-EGF-2 domains. P33 has higher number of contact (up to ~~9~~8) with the I-EGF-2 domain not observed for the other variants (≤3).