Electronic Supporting information

Predicting long term cooperativity and specific modulators of receptor interactions in human transferrin from dynamics within a Single Microstate

Haleh Abdizadeh and Canan Atilgan*
Faculty of Engineering and Natural Sciences, Sabanci University, Turkey

*Corresponding Author
Email: canan@sabanciuniv.edu
**Supplementary text**

**Text S1. Protein structures.** As mentioned in the main text, we utilize the holo and apo X-ray crystal structures of hTf. In the holo form, highly conserved iron liganding residues in N- (C-) lobes are comprised of two tyrosines (Y95/188 and Y426/517), an aspartic acid (D63/D392), and a histidine (H249/H585).\(^1\) Two oxygen atoms from synergistic anion, carbonate, complete the octahedral coordination of iron (Fig. 1a). For holo hTf, we select the chain B of the protein data bank (PDB) code 3V83\(^2\) which has the maximum number of residues (1-677). There are 68 missing atoms on chain B which we add using the “Build and Edit” feature of the Discovery Studio Visualizer 4.0 software.\(^3\) We call this structure **H** in the manuscript (Fig. 1b). The closed cleft of holo form buries iron 10 Å under the protein surface.\(^4\) The open form has 2HAV code\(^5\) and the first three residues are not present in the crystal structure; we refer to this structure as **A** throughout. RMSD based on C\(\alpha\) atoms between the two structures is 8.2 Å.\(^6\) The conformational difference between **H** and **A** involves the opening at the two hinges between the N\(_1\)/N\(_2\)- and C\(_1\)/C\(_2\)-subdomains, and a twist between the two lobes facilitated by the linker.\(^5, 7\)

**Text S2. MD simulations.** The NAMD package is used to model the dynamics of the protein-water systems.\(^8\) Solvation is achieved via the VMD 1.9.1 program solvate plug-in version 1.2.\(^9\) The protein is soaked in a cubic solvent box such that there is at least a 10 Å layer of solvent in each direction from any atom of the protein to the edge of the box. Ionic strength in all the simulations is 150 mM. Charge neutrality is obtained by adding 70, 63, 63 and 77 of Na\(^+\) and 63, 69, 69 and 71 of Cl\(^-\) ions to **H**\(_\text{p}\), **H**\(_\text{e}\), **A** and **A**\(_\text{C}\) systems, respectively. There are 78214, 77947, 86426 and 86422 in the respective systems. The CharmM22 all-atom force field is used to parameterize the topology of the protein atoms.\(^10\) The force field parameters for synergistic carbonate ion and Fe\(^{3+}\) reported in the literature are adopted.\(^11, 12\) TIP3P model is used to describe water molecules.\(^13\) During the simulations, periodic boundary conditions are imposed on the simulation boxes that have 88×107×90 and 97×99×96 Å\(^3\) dimensions for the holo and apo systems, respectively. Long range electrostatic interactions are
calculated by the particle mesh Ewald method\textsuperscript{14}, with a cutoff distance of 12 Å and a switching function at 10 Å. RATTLE algorithm\textsuperscript{15} is applied and a time step of 2 fs in the Verlet algorithm is used. Temperature control is carried out by Langevin dynamics with a dampening coefficient of 5/ps. Pressure control is attained by a Langevin piston. All systems were first subjected to 10000 steps of energy minimization with the conjugate gradients algorithm. 10 ns MD runs in the NVT ensemble at 310 K were carried out on the resulting systems. The resulting structures are then run in the NPT ensemble at 1 atm and 310 K until volumetric fluctuations stabilize and the desired average pressure is maintained.

For the selection of protonation sites to mimic environments at pH 7.4 and 5.6, we assign the most probable states of ionizable groups according to their pK\textsubscript{a} calculated by both PROPKA and H++ servers.\textsuperscript{16, 17} A detailed list and description of how they were selected is described in ref.\textsuperscript{18}. As a result, for the A system, all residues are assigned their standard protonation states. One 200 ns long, and another 50 ns long simulation for the A system displays the same average structure, RMSD profiles, and dominant modes of motion. For H\textsuperscript{a}, several runs whereby the tyrosine residues coordinating the irons (Fig. 1a) are assigned different charge states are carried out. When one or both of these tyrosines on each lobe are charged, the system remains close to the x-ray structure, maintaining 2 Å RMSD in two runs, each of 60 ns length. When both tyrosines on both lobes are neutral, a well-defined domain splitting on both lobes is observed in two separate runs of 500 and 150 ns length. The results of the latter have been discussed in detail in ref.\textsuperscript{18} and the observed intermediate has been related to the partially open crystal structure of holo hTf. In H\textsuperscript{b}, histidines 14, 25, 242, 289, 349, 578, 606, and 642 are protonated (one run of 300 and another of 200 ns length). We note that in reference \textsuperscript{18} we have verified that these choices of protonation states reflect the experimentally observed dynamics of hTf under different environmental conditions.

In AutoDock, the center of mass of the docked molecule (carbonate) was selected from the coordinates of the C\textsubscript{\alpha} atoms of the contacting residues listed above. The carbonate was enclosed in a box with number of grid points 58×60×70 and 60×92×60 in x-, y- and z-directions, respectively for the N-lobe and C-lobe carbonate, with a grid spacing of 0.375 Å. Lamarckian genetic algorithms, as implemented
in AutoDock, were employed to perform docking calculations. All other parameters were set to their default values. For each of the docking cases, the lowest energy docked conformation, according to the Autodock scoring function, was selected as the binding mode. We have produced two sets of MD runs, 100 ns each, of the carbonate docked (Ac) system.

**Text S3. PRS method.** PRS relies on scanning the protein C\textsubscript{\textalpha} atoms using eqn (1) and applying the force $\Delta F$ on each residue in different directions. For the perturbation of a single residue $i$, \[ (AF)^T = \{000...\Delta F_i^x \Delta F_i^y \Delta F_i^z ...000\}_{j=3,N} \]. In this work, the C matrix generated from all atom MD simulations describes the interactions with detailed potentials that directly take into account their distance dependence for isolated hTf. An MD trajectory is coarse grained by tracing the coordinates of its C\textsubscript{\textalpha} atoms. The deviation of residue $j$ from the average structure obtained over the time window of observations (discretely counted by $w$ recorded steps) is $\Delta R_j(t) = R_j(t) - \langle R_j(t) \rangle$. For all residues, the deviations may be recorded in the $3N \times w$ trajectory matrix $\Delta R$. The covariance matrix $C$ is obtained from the matrix product $(\Delta R \Delta R^T)/w$. For the hTf-receptor complexes where MD simulations are prohibitively expensive, $C$ has been constructed from ANM with 10 Å cutoff distance.\textsuperscript{19} The slowest modes of motion contributing to the dynamics are obtained by a singular value decomposition of the $C$ matrix and projection the eigenvectors corresponding to the dominant eigenvalues onto the three dimensional protein structure.

Cross-correlations between residue pairs $i$ and $j$ are calculated by treating the $C$ matrix as an $N \times N$ supermatrix, whose $ij$th element is the $3 \times 3$ second moment matrix of correlations between the $x$-, $y$-, and $z$-components of the fluctuations $\Delta R$, and $\Delta R_i$ of residues $i$ and $j$:\textsuperscript{20}

\[
C_{ij} = \begin{bmatrix}
\langle \Delta X_i \Delta X_j \rangle & \langle \Delta X_i \Delta Y_j \rangle & \langle \Delta X_i \Delta Z_j \rangle \\
\langle \Delta Y_i \Delta X_j \rangle & \langle \Delta Y_i \Delta Y_j \rangle & \langle \Delta Y_i \Delta Z_j \rangle \\
\langle \Delta Z_i \Delta X_j \rangle & \langle \Delta Z_i \Delta Y_j \rangle & \langle \Delta Z_i \Delta Z_j \rangle 
\end{bmatrix}
\] (S1)

Finally, the cross-correlations between residues $i$ and $j$ is given by:
\[
\langle \Delta R_i \Delta R_j \rangle = \text{tr}(C_{ij})
\]  

(S2)

In PRS there is no a priori assumption on how a force might be generated at a particular point. Conversely, after finding the force/residue pair that best leads to the conformational change of interest, we relate this finding to possible causes. PRS is applied by scanning each residue in 200 random directions. Due to the linear response nature of the methodology, the calculated displacements will be directly proportional to the imposed force. Since we are interested in the relative displacements (see below), a force of unit magnitude is applied throughout.

The quality of the predicted displacements is assessed via eqn (2). On the other hand, overlaps between the compared vectors instead of correlations have been used in our previous work.\textsuperscript{21, 22} However, for large scale transitions, the overlap, calculated as the dot product \((\Delta \mathbf{R}_i \cdot \Delta \mathbf{S})/(|\Delta \mathbf{R}_i| \, |\Delta \mathbf{S}|)\), is not expected to be close to 1 since the direction of the initial motion may be very different from the overall displacements characterizing the end points of the conformational change, although the motion described may be very similar. In the multi-domain proteins we study in this work, the motions are always large scale. While this problem has been discussed and alternative approaches to the problem have been offered in the literature\textsuperscript{23, 24}, there is no best motion-type-independent solution to the problem. We have found that the correlation between the relative displacements of the residues is adequate in classifying the induced conformational change due to perturbation.\textsuperscript{12} We therefore apply eqn (2) throughout this work.
Figure S1. The maximum (out of 200 random directions) displacements due to perturbations for the 674 residues are compared with those of the crystal structures and correlations between the predicted and observed displacements (eqn (2)) are sorted from largest to smallest. For the A → H, a sharp drop from 0.75 to 0.70 after the first 16 residues is observed while for H → A, there is smooth decrease in correlation values, the largest being 0.52.

References:


