Supporting Information

Allosteric Modulation of Membrane Proteins by Small Low-Affinity Ligands

Werner Treptow

Laboratório de Biologia Teórica e Computacional (LBTC), Universidade de Brasília, Brasília, DF, CEP 70904-970, Brasil

Corresponding author and lead contact: E-mail treptow@unb.br

SI MATERIALS AND METHODS

Detailed solution of the thermodynamic cycle shown in the main-text Figure 1 is presented as supporting information for publication. Although a detailed derivation of the two state model can be found in classic mechanical statistical text books,[1] our second derivation of the model based on free-energy perturbation is particularly useful in the context of MD simulations for proper identification and computation of the energy contributions arising from the interaction of protein, ligand and external forces.

Conformational equilibrium of the protein free of ligands. By relying on the canonical free energy of the extended system

$$e^{-\beta F_Y(\boldsymbol{R}^n,\chi)} \propto \int d\boldsymbol{r}^P \delta[\boldsymbol{Y}^{'}(\boldsymbol{r}^P)-\boldsymbol{Y}]\int d\boldsymbol{r}_1 \delta[\boldsymbol{R}^{'}(\boldsymbol{r}_1)-\boldsymbol{R}_1]...\int d\boldsymbol{r}_n \delta[\boldsymbol{R}^{'}(\boldsymbol{r}_n)-\boldsymbol{R}_n]\int_V d\boldsymbol{r}^{N-n}\int d\boldsymbol{r}^S\int d\bar{\chi} e^{-\beta[U(\boldsymbol{r}^{P+N+S})-\chi\bar{\chi}]} ,$$ [S1]

an equilibrium constant for reaction $A_0 \underset{\leftarrow}{\overset{\chi}{\rightleftharpoons}} B_0$ may be properly defined

$$K(\chi) \equiv \frac{P_{B_0}(\chi)}{P_{A_0}(\chi)} = \frac{\int_V d\boldsymbol{R}^n e^{-\beta F_B(\boldsymbol{R}^n,\chi)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_A(\boldsymbol{R}^n,\chi)}}$$ [S2]

following integration of eq. [S1] over microscopic configurations in which all ligands are in the bath volume V. Evaluation of the equilibrium constant is attained in the context of free-energy perturbation (FEP) by detailing the reaction process along a reversible thermodynamic cycle in which: (*i*) the applied stimulus χ is first turned off at the protein state A_0 ; (*ii*) protein changes its conformation to B_0 in the absence of stimulus; and (*iii*) the applied stimulus is turned on at the protein state B_0. From that construction, the ratio of configuration integrals in eq. [S2] rewrites exactly

$$\frac{\int_V d\boldsymbol{R}^n e^{-\beta F_B(\boldsymbol{R}^n,\chi)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_A(\boldsymbol{R}^n,\chi)}} = \overbrace{\frac{\int_V d\boldsymbol{R}^n e^{-\beta F_A(\boldsymbol{R}^n,0)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_A(\boldsymbol{R}^n,\chi)}}}^{\equiv e^{+\beta\Delta F(A_0\rightarrow\chi)}} \times \overbrace{\frac{\int_V d\boldsymbol{R}^n e^{-\beta F_B(\boldsymbol{R}^n,0)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_A(\boldsymbol{R}^n,0)}}}^{\equiv e^{-\beta\Delta F(A_0\rightarrow B_0,0)}} \times \overbrace{\frac{\int_V d\boldsymbol{R}^n e^{-\beta F_B(\boldsymbol{R}^n,\chi)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_B(\boldsymbol{R}^n,0)}}}^{\equiv e^{-\beta\Delta F(B_0,0\rightarrow\chi)}}$$ [S3]

and the equilibrium constant resumes

$$K(\chi)=K(0)\times e^{-\beta\Delta\Delta F(A_0\rightarrow B_0,0\rightarrow\chi)}, \quad K(0)\equiv e^{-\beta\Delta F(A_0\rightarrow B_0,0)}$$ [S4]

as a function of the intrinsic free energy of the system $\Delta F(A_0\rightarrow B_0,0)\equiv F(B_0,0)-F(A_0,0)$ and the *excess* free energy due to the applied stimulus $\Delta\Delta F(A_0\rightarrow B_0,0\rightarrow\chi)\equiv\Delta F(B_0,0\rightarrow\chi)-\Delta F(A_0,0\rightarrow\chi)$.

For a sufficiently low stimulus, the *excess* free energy can be formulated

$$\begin{aligned}\Delta F(\boldsymbol{Y}_0,0\rightarrow\chi)&=F(\boldsymbol{Y}_o,\chi)-F(\boldsymbol{Y}_o,0)\\&=\int_0^\chi d\chi'\frac{\partial}{\partial\chi'}\Delta F(\boldsymbol{Y}_0,0\rightarrow\chi')\\&=-\int_0^\chi d\chi'\langle\bar{\chi}\rangle_{\boldsymbol{Y}_0,\chi'}\\&=-[\chi\langle\bar{\chi}\rangle_{\boldsymbol{Y}_0,0}+\frac{1}{2}c\chi^2]\end{aligned}$$ [S5]

in terms of linear response considerations $\langle\bar{\chi}\rangle_{\boldsymbol{Y}_0,\chi'}\approx\langle\bar{\chi}\rangle_{\boldsymbol{Y}_0,0}+c\chi$ involving equilibrium fluctuations of the system *i.e.*, $c=\frac{\partial}{\partial\chi}\langle\bar{\chi}\rangle_{\boldsymbol{Y}_o,\chi}\big|_{\chi=0}$.[2] As previously shown for displacement charges Q under the influence of a membrane potential V ,[3] equilibrium fluctuations or variance of the system in the absence of any stimulus are expected to little depend on the protein conformations. In the limit in which c is comparable between protein's states, eq. [S5] then conveniently establishes the *excess* free-energy difference

$$\Delta\Delta F(A_0\rightarrow B_0,0\rightarrow\chi)\approx-\chi[\langle\bar{\chi}\rangle_{B_0,0}-\langle\bar{\chi}\rangle_{A_0,0}]\equiv-\chi\Delta\langle\bar{\chi}\rangle .$$ [S6]

as the variation of the stimulus energy resulting from the net displacement of $\langle\bar{\chi}\rangle_{\boldsymbol{Y}_0,0}$ at $\chi=0$. The consequence for the energetics of the extended system along reaction $A_0 \underset{\leftarrow}{\overset{\chi}{\rightleftharpoons}} B_0$ is clear and the equilibrium constant rewrites accordingly

$$K(\chi)=K(0)\times e^{+\beta\chi\Delta\langle\bar{\chi}\rangle}, \quad K(0)=e^{-\beta\chi_m\Delta\langle\bar{\chi}\rangle} \tag{S7}$$

by taking into consideration the midpoint equilibrium χ_m in which both states of the system are equally likely *i.e.*, $\Delta F(A_0\rightarrow B_0,0)=\chi_m\Delta\langle\bar{\chi}\rangle$. As an extensive change in the external device, $\Delta\langle\bar{\chi}\rangle$ can be directly related to an opposite equivalent variation in the protein subsystem

$$\Delta\langle\bar{\chi}\rangle=-\Delta\langle\bar{\chi}(\boldsymbol{r}^P)\rangle \tag{S8}$$

and consequently, eq. [S7] becomes useful to describe the conformational equilibrium of the protein in terms of the external variables $\{\chi_m,\Delta\langle\bar{\chi}\rangle\}$ that can be learned as adjustment of the model's parameters over measurements. From eq. [S7], the probability of the specific state B_0 then writes as the familiar two-state Boltzmann equation

$$P_{B_0}(\chi)\equiv\left[1+P_{A_0}(\chi)/P_{B_0}(\chi)\right]^{-1}=\left[1+K(\chi)^{-1}\right]^{-1}=\left[1+e^{+\beta(\chi_m-\chi)\Delta\langle\bar{\chi}\rangle}\right]^{-1}. \tag{S9}$$

Ligand Binding to a fixed Protein Conformation. Reaction $Y_0+nL\overset{\chi}{\underset{}{\rightleftarrows}}Y_n$ describes a process in which n ligands unambiguously bind the protein conformation Y at a microscopic volume v *i.e.*,

$$n\equiv\int_v d\boldsymbol{R}\sum_{i=1}^N\delta[\boldsymbol{R}(\boldsymbol{r}_i)-\boldsymbol{R}] . \tag{S10}$$

According to that definition, an equilibrium constant for the binding process under stimulus χ may be properly defined

$$K_Y(n,\chi)\equiv\frac{P_Y(n,\chi)}{P_Y(0,\chi)}=\omega\frac{\int_v d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,\chi)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,\chi)}} \tag{S11}$$

at a certain number of degenerate states $\omega\approx N^n/n!$ in the thermodynamic limit $n\ll N$. Again, evaluation of $K_Y(n,\chi)$ is attained in the context of free-energy perturbation (FEP) by detailing the binding process along a reversible thermodynamic cycle with well-defined intermediate states constructed by the use of auxiliary external potentials.[4] As shown in Figure 1, (*i*) the applied stimulus χ is turned off at the protein state Y_0, (*ii*) n ligands are step-wisely decoupled from the bath (*) and re-coupled into the protein site (**) under applied restraints $\{u^*(\boldsymbol{R}^n),u^{**}(\boldsymbol{R}^n)\}$ and, (*iii*) stimulus is turned on at the bound state Y_n. For the case of small ligands, the binder typically does not adopt any special configuration in the bound state and external potentials are purely translational flat wells applied to confine the molecule within equilibrium three-dimensional volumes in the bath V and protein site v.[5] From that construction, the ratio of configuration integrals in eq. [S11] rewrites exactly

$$\frac{\int_v d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,\chi)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,\chi)}}=\overbrace{\frac{\int_V d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,0)}}{\int_V d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,\chi)}}}^{\equiv e^{+\beta\Delta F(Y_0,0\rightarrow\chi)}}\times\overbrace{\frac{\int_V d\boldsymbol{R}^n e^{-\beta[F_Y(\boldsymbol{R}^n,0)+u^*(\boldsymbol{R}^n)]}}{\int_V d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,0)}}}^{\equiv e^{-\beta W_n^*}}\times\overbrace{\frac{\int_v d\boldsymbol{R}^n e^{-\beta[F_Y(\boldsymbol{R}^n,0)+u^{**}(\boldsymbol{R}^n)]}}{\int_V d\boldsymbol{R}^n e^{-\beta[F_Y(\boldsymbol{R}^n,0)+u^*(\boldsymbol{R}^n)]}}}^{\equiv(v/V)^n}\times\overbrace{\frac{\int_v d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,0)}}{\int_v d\boldsymbol{R}^n e^{-\beta[F_Y(\boldsymbol{R}^n,0)+u^{**}(\boldsymbol{R}^n)]}}}^{\equiv e^{+\beta W_Y^{**}(n)}}\times\overbrace{\frac{\int_v d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,\chi)}}{\int_v d\boldsymbol{R}^n e^{-\beta F_Y(\boldsymbol{R}^n,0)}}}^{\equiv e^{-\beta\Delta F(Y_n,0\rightarrow\chi)}}$$

and the equilibrium constant resumes

$$K_Y(n,\chi)=\rho^n K_Y(n)\times e^{-\beta\ \Delta\Delta F(Y_0\rightarrow Y_n,0\rightarrow\chi)}, \qquad K_Y(n)\equiv\frac{v^n}{n!}e^{-\beta[W_n^*-W_Y^{**}(n)]} \tag{S12}$$

in terms of the molecular density ρ of the ligand, its desolvation energies $W_n^*-W_Y^{**}(n)$ and the associated *excess* free-energy difference $\Delta\Delta F(Y_0\rightarrow Y_n,0\rightarrow\chi)\equiv\Delta F(Y_n,0\rightarrow\chi)-\Delta F(Y_0,0\rightarrow\chi)$. To simplify notation, any potential dependence of the binding site volume with the protein conformation is not shown in eq. [S12] *i.e.*, $v\equiv v_Y$.

As made explicit in eq. [S12], ligand interaction may affect the energy coupling of the protein with the external device and therefore, it may impact the conformational free-energy of the protein in a stimulus dependent manner. With similar arguments as from eq. [S6], the *excess* free-energy difference between bound and unbound states is clear and mostly derives from the net displacement of $\langle\bar{\chi}\rangle_{Y_n,0}$ induced by the ligands

$$\Delta\Delta F(Y_0\rightarrow Y_n,0\rightarrow\chi)\approx-\chi[\langle\bar{\chi}\rangle_{Y_n,0}-\langle\bar{\chi}\rangle_{Y_0,0}]\equiv-\chi\Delta\langle\bar{\chi}\rangle_{n,Y} . \tag{S13}$$

On the other hand, $K_X(n)$ is the stimulus-independent binding constant of the ligand to the protein site volume v. The relevance of $K_Y(n)$ is clear as it ensures the probability of any occupancy state of the protein to be known accordingly

$$P_Y(n)=Z_Y^{-1}\rho^n K_Y(n) \tag{S14}$$

with, normalization given by the grand canonical partition function $Z_Y\equiv\sum_n\rho^{n'}K_Y(n')$. Implicit in eq. [S14] is the fact that $P_Y(n)$ depends on the molecular density of the ligand thus implying a concentration-dependent average occupancy of the protein by $\langle n\rangle_Y$ molecules where, $\langle n\rangle_Y=\sum_{n'}n'P_Y(n')$.

Conformational Equilibrium of the Protein in Presence of Ligands. From eq. [S7, S12, S13 and S14], system's probabilities in reaction $A_n\overset{\chi}{\underset{}{\rightleftarrows}}B_n$ can be expanded across every occupancy state of the protein conformations at a fixed ligand concentration

$$\begin{cases}P_{A_n}(\chi)=\sum_{n'}P_A(n',\chi)=P_A(0,\chi)Z_A e^{+\beta\chi\Delta\langle\bar{\chi}\rangle_{n,A}} \\ P_{B_n}(\chi)=\sum_{n'}P_B(n',\chi)=P_A(0,\chi)K(\chi)Z_B e^{+\beta\chi\Delta\langle\bar{\chi}\rangle_{n,B}}\end{cases}.$$

The resulting probability of state B_n then rewrites in the form of the two-state Boltzmann equation

$$P_{B_n}(\chi)\equiv\left[1+P_{A_n}(\chi)/P_{B_n}(\chi)\right]^{-1}=\left[1+\alpha K(\chi)^{-1}\right]^{-1}=\left[1+\alpha e^{+\beta(\chi_m-\chi)\Delta\langle\bar{\chi}\rangle}\right]^{-1} \tag{S15}$$

2

now embodying the grand-canonical partition functions of the protein subsystem and the subjacent free-energy contributions derived from ligand binding $\alpha \equiv \dfrac{Z_A e^{+\beta\chi\Delta\langle\chi\rangle_{n,A}}}{Z_B e^{+\beta\chi\Delta\langle\chi\rangle_{n,B}}}$.

For small low-affinity ligands, α derives essentially from the grand-canonical partition function ratio Z_A/Z_B (see main-text for more details). For a scenario in which two or more distinct sites of the protein $v=v_1+...+v_s$ are respectively occupied by $n=n_1+...+n_s$ small low-affinity molecules, the partition ratio at a fixed concentration may be more conveniently expressed in terms of a generalized binding constant $K_Y(n_1,...,n_s)$ by detailing eq. [S12] over every possible multi-occupancy state of the protein subsystem.[6] Because computation of $W_Y^{**}(n_1,...,n_s)$ increases significantly with the number of inter-correlated sites along the protein structure, the generalized constant cannot be applied to solve in practice multiple correlated binding events unless for the condition of independent sites in which $K_Y(n_1,...,n_s) \approx K_Y(n_1) \times ... \times K_Y(n_s)$

$$K_Y(n_1) = \frac{v_1^{n_1}}{n_1!} e^{-\beta[n_1\Pi - W_Y^{**}(n_1)]}$$
$$...$$ [S16]
$$K_Y(n_s) = \frac{v_s^{n_s}}{n_s!} e^{-\beta[n_s\Pi - W_Y^{**}(n_s)]}$$

- here, it is understood that each binding constant refers to the equilibrium reaction between the unbound and bound state of the protein at the level of a single binding site. In the condition of independent binding sites, the partition function factorizes as a product of individual contributions

$$\frac{Z_A}{Z_B} = \frac{\sum_{n_1',...,n_s'} \bar{\rho}^{\,n_1'+...+n_s'} K_A(n_1',...,n_s')}{\sum_{n_1',...,n_s'} \bar{\rho}^{\,n_1'+...+n_s'} K_B(n_1',...,n_s')} \approx \frac{Z_{A,1} \times ... \times Z_{A,s}}{Z_{B,1} \times ... \times Z_{B,s}} \ .$$ [S17]

By describing ligand binding over multiple independent sites, eq. [S15, and S17] are generalizations of the classical Monod-Wyman-Changeux (MWC) model devised for allosteric modulation of proteins[7] and commonly used in the study of transcription factors in presence of ligand molecules.[8]

Extrapolation of the analysis to the case of highly degenerate interactions allows eq. [S17] to be expressed in terms of a partition phenomenon at the molecular interface of proteins. Under low-affinity interactions, the energetics of the protein subsystem rewrites as a function of the partition process at a fixed concentration

$$K_Y(n_Y) \equiv \frac{v^{n_Y}}{n_Y!} \wp_Y^{n_Y}, \qquad \wp_Y = e^{-\beta[W_1^* - W_Y^{**}(1)]}$$ [S18]

in which a discrete average number of ligands $n_Y = \lfloor\langle n\rangle_Y\rfloor$ migrates from the bath (*) into the protein (**) according to a conformation-dependent partition coefficient $\wp_Y$.[9] In analogy to an "aggregate" binding constant of the ligand over multiple microscopic sites of the protein $K_Y(n_1,...,n_s)$, eq. [S18] then defines an equilibrium constant for the partition process at the macroscopic volume of the protein interface v such that,

$$\frac{Z_A}{Z_B} = \frac{1+\rho^{n_A} K_A(n_A)}{1+\rho^{n_B} K_B(n_B)} \ .$$ [S19]

While eq. [S18] is only valid under dilution, the partition coefficient might deviate from its dilute value in consequence of self interactions of the ligand. In the case of non-specific low-affinity binders, self interaction energies were demonstrated to cancel out over the free-energy difference $W_1^* - W_Y^{**}(1)$ and the partition coefficient to be an effective concentration-independent function of the desolvation energies under dilution. As such, for the case of non-specific low-affinity binders, eq. [S18] can be resolved under dilution by taking into consideration high-concentration estimates of $\wp_Y'$ that can be efficiently learned from flooding-MD simulations of the protein embedded in a non-saturating environment rich of ligands.[9]

References

(1)   Hill, T. L. *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*; Springer Series in Molecular and Cell Biology; Springer-Verlag: New York, 1985. https://doi.org/10.1007/978-1-4612-5082-1.

(2)   McQuarrie, D. A. *Statistical Mechanics*, 1st edition.; University Science Books, 2000.

(3)   Roux, B. Influence of the Membrane Potential on the Free Energy of an Intrinsic Protein. *Biophys. J.* **1997**, *73* (6), 2980–2989.

(4)   Dubois, J.-M.; Ouanounou, G.; Rouzaire-Dubois, B. The Boltzmann Equation in Molecular Biology. *Prog. Biophys. Mol. Biol.* **2009**, *99* (2), 87–93. https://doi.org/10.1016/j.pbiomolbio.2009.07.001.

(5)   Woo, H.-J.; Roux, B. Calculation of Absolute Protein–Ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6825–6830. https://doi.org/10.1073/pnas.0409005102.

(6)   Salari, R.; Joseph, T.; Lohia, R.; Hénin, J.; Brannigan, G. A Streamlined, General Approach for Computing Ligand Binding Free Energies and Its Application to GPCR-Bound Cholesterol. *J. Chem. Theory Comput.* **2018**, *14* (12), 6560–6573. https://doi.org/10.1021/acs.jctc.8b00447.

(7)   Stock, L.; Hosoume, J.; Treptow, W. Concentration-Dependent Binding of Small Ligands to Multiple Saturable Sites in Membrane Proteins. *Sci. Rep.* **2017**, *7* (1), 5734. https://doi.org/10.1038/s41598-017-05896-8.

(8)   Monod, J.; Wyman, J.; Changeux, J.-P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

(9)   Razo-Mejia, M.; Barnes, S. L.; Belliveau, N. M.; Chure, G.; Einav, T.; Lewis, M.; Phillips, R. Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. *Cell Syst.* **2018**, *6* (4), 456-469.e10. https://doi.org/10.1016/j.cels.2018.02.004.

(10)   Cirqueira, L.; Stock, L.; Treptow, W. Concentration-Dependent Thermodynamic Analysis of the Partition Process of Small Ligands into Proteins. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 4885–4891. https://doi.org/10.1016/j.csbj.2022.08.049.
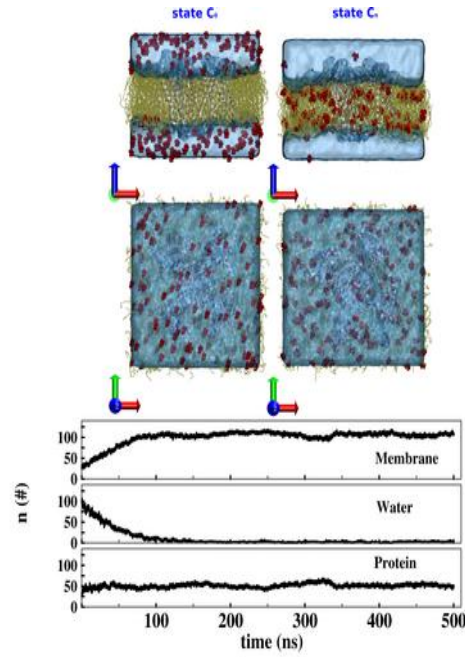
Figure S1. Simulation of the partition process. (Top) Representative initial and final equilibrium flooding-MD configurations of the channel (white) embedded in a fully-hydrated lipid bilayer (ochre) and in presence of a high concentration of the ligand (red). (Bottom) Instantaneous number of ligands in the water, membrane and within 5.0 Å of the protein. Steady-state was defined according to the simulation time in which sevoflurane fully partitions across the system $t^* > 100\,ns$.

Table-S1. Equilibrium properties of flooding MD simulations: Decomposition of the bath volume

Y	Simulation	Bath Waters				Bath Lipids			
		n_1 (#)	v_1 (Å³)	$\frac{n_1}{v_1}$ (M)	x_L	n_2 (#)	v_2 (Å³)	$\frac{n_2}{v_2}$ (M)	x_L
C	1	2.90±2.152	917.513x10³	0.005	8.08x10⁻⁵	106.13±4.450	567.148x10³	0.312	0.25
	2	1.55±1.709	917.513x10³	0.003	4.33x10⁻⁵	73.68±3.876	567.148x10³	0.216	0.17
	3	1.02±2.147	917.513x10³	0.002	2.83x10⁻⁵	53.38±2.983	567.148x10³	0.156	0.12
O	1	3.70±3.398	971.833x10³	0.006	9.58x10⁻⁵	104.17±5.092	598.743x10³	0.289	0.24
	2	3.06±2.246	971.833x10³	0.005	7.94x10⁻⁵	69.59±5.946	598.743x10³	0.193	0.16
	3	3.40±2.577	971.833x10³	0.006	8.80x10⁻⁵	48.06±3.666	598.743x10³	0.133	0.11

*Where, $V = v_1 + v_2$ and $N - n = n_1 + n_2$. Here, n_1 and n_2 were determined by averaging the number of ligands over the respective partition volumes of the bath. x_L is the mole fraction of the ligand.