Computational and Structural Biotechnology Journal 20 (2022) 4885–4891

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Concentration-dependent thermodynamic analysis of the partition process of small ligands into proteins

Leonardo Cirqueira, Letícia Stock, Werner Treptow *

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

ARTICLE INFO

Article history:
Received 2 June 2022
Received in revised form 19 August 2022
Accepted 20 August 2022
Available online 1 September 2022

Keywords:
Non-specific low-affinity interactions
Partition coefficient
Free energy stability
Flooding MD
Cosolvent MD

ABSTRACT

In the category of functional low-affinity interactions, small ligands may interact with multiple protein sites in a highly degenerate manner. Better conceived as a partition phenomenon at the molecular interface of proteins, such low-affinity interactions appear to be hidden to our current experimental resolution making their structural and functional characterization difficult in the low concentration regime of physiological processes. Characterization of the partition phenomenon under higher chemical forces could be a relevant strategy to tackle the problem provided the results can be scaled back to the low concentration range. Far from being trivial, such scaling demands a concentration-dependent understanding of self-interactions of the ligands, structural perturbations of the protein, among other molecular effects. Accordingly, we elaborate a novel and detailed concentration-dependent thermodynamic analysis of the partition process of small ligands aiming at characterizing the stability and structure of the dilute phenomenon from high concentrations. In analogy to an "aggregate" binding constant of a small molecule over multiple sites of a protein receptor, the model defines the stability of the process as a macroscopic equilibrium constant for the partition number of ligands that can be used to analyze biochemical and functional data of two-component systems driven by low-affinity interactions. Acquisition of such modeling-based structural information is expected to be highly welcome by revealing more traceable protein-binding spots for non-specific ligands.

As we expand our scientific knowledge and gain insight about previously unknown phenomena, it becomes evident that low-affinity non-specific interactions may play functional roles in the realm of molecular biology. Recent examples include reactions of chemicals, lipids and disordered proteins with a variety of macromolecular targets ranging from proteins [1], chromatin [2,3] to membraneless compartments [4]. With implications for drug repurposing and side-effects [5], enhancement of proteasome based therapies [6] and pathologies [7], low-affinity non-specific interactions have just started to be uncovered across a number of emergent research fields.

In the category of functional low-affinity interactions, small ligands may interact with multiple sites of a protein target in a highly degenerate manner to a degree that it is better conceived as a partition phenomenon at the molecular interface of proteins [8]. Such low-affinity interactions are structurally and functionally challenging to characterize and thus might be hidden in the most relevant physiological, low-concentration regime [9]. Investigation of the partition phenomenon under higher chemical thermodynamic forces or potentials could be one strategy to tackle the problem provided that the translation of results back to the low-concentration regime is made possible. The latter is a non-trivial task that requires among other molecular effects ligand self-interactions and protein structural perturbations to be well understood in a concentration-dependent manner.

To asses the feasibility of said strategy, we elaborate a novel and detailed concentration-dependent thermodynamic analysis of the partition process of small ligands to proteins, aiming at characterizing the free-energy stability and structure of the dilute phenomenon from high concentration sampling. According to the model, acquisition of dilute structural information derives directly from the spatial distribution of the ligand at high concentrations. The high-concentration distribution helps to define the free-energy stability of the dilute partition process as a macroscopic equilibrium constant that in analogy to an "aggregate" binding constant of a small molecule over multiple microscopic sites of a protein receptor is expected to be useful for analysis of biochemical and functional data of two-component systems governed by low-affinity non-specific interactions. In more technical terms,

* Corresponding author.
E-mail address: treptow@unb.br (W. Treptow).

https://doi.org/10.1016/j.csbj.2022.08.049

the thermodynamic analysis establishes energetic conditions in which concentration effects are ineffective and the dilute partition coefficient of the ligand into the protein site becomes a concentration invariant quantity of the molecular distribution across two well-defined phases of the system. The outcome is a simple formulation that allows the partition number, free-energy stability and structure of the dilute partition process to be predicted from the unitary probability density of the ligand at high concentrations $\bar{\rho}'(\boldsymbol{R})$. Illustration and discussion of the approach is shown in the context of flooding or cosolvent Molecular Dynamics (MD) which is a widespread method that can be efficiently applied to resolve $\bar{\rho}'(\boldsymbol{R})$ for a variety of systems under concentration effects and/or mixed-solvent compositions [10]. Here, we focus on the general anesthetic sevoflurane and the neuronal membrane protein Kv1.2, a mammalian ion channel for which structural data is available [11]. The fact that one or more anesthetic molecules do interact with the channel over multiple degenerate sites makes the system an important benchmark for numerical validation of our formulation [12].

1. Theory and methods

We consider a macromolecular system comprised of M chemical species, including a single protein P embedded in a large solvent volume V that contains N indistinguishable ligands L. The volume v within a certain cut-off distance of the protein is assumed to be a continuous phase occupied by a certain number of molecules

$$\langle n\rangle = N\int d\boldsymbol{R}\bar{\rho}(\boldsymbol{R}) \tag{1}$$

dictated by the spatial unitary density of the ligand across the system $\bar{\rho}(\boldsymbol{R})$.

Next, we rely on that definition to devise a thermodynamic model for the partition process of non-specific low-affinity ligands. Because such ligands occupy the protein site in a highly degenerate manner, the model assumes by construction that volume v is a nearly-homogeneous phase characterized by position-independent desolvation free-energy of the ligand.

1.1. Low-concentration partition coefficient, free-energy stability and structure

From Eq. (1), a partition coefficient may be defined relative to the number of ligand molecules $N-\langle n\rangle$ in the bulk $V-v$

$$\wp = \left(\frac{N-\langle n\rangle}{V-v}\right)^{-1}\frac{\langle n\rangle}{v} \tag{2}$$

In the present form, Eq. (2) does not clarify any potential dependence that $\wp$ may have with concentration. To analyze that aspect more carefully, we reinterpret in Fig. 1 the partition coefficient along an idealized thermodynamic cycle in which a single ligand molecule is reversibly decoupled from the bulk (*) and the protein interface (**) into the gas phase with the use of auxiliary external potentials [13]. In case of small ligands, the molecule typically does not adopt any special configuration along the partition process and the external potentials $\{u^*, u^{**}\}$ are purely translational flat wells applied to confine the ligands within three-dimensional volumes. According to that construction, the partition coefficient may then be directly linked to the desolvation free-energy of the ligand in each phase of the system $\{W^*, W^{**}\}$

$$\wp = e^{-\beta[W^*-W^{**}]} \tag{3}$$

at a fixed temperature $\beta=(k_BT)^{-1}$. As detailed in Supporting Information, desolvation free energies are computed along decoupling transformations in FEP simulations and W^*-W^{**} reports the resulting free-energy difference to transfer the ligand from the bulk into the layer around the protein. Because each free-energy corresponds to reversible external work against non-bonded molecular forces imposed by the local environment including other ligand molecules, $\{W^*, W^{**}\}$ are concentration-dependent quantities that according to a linear version of the partition model for binary mixtures [14]

$$\begin{cases} W^* = \bar{\mu} - h^*\frac{N-\langle n\rangle}{V-v}c \\ W^{**} = \bar{w} - h^{**}\frac{\langle n\rangle}{v}c \end{cases} \tag{4}$$

might deviate from their dilute reference values $\{\bar{\mu},\bar{w}\}$ at high molar concentrations ($c=1,660\text{Å}^3$), in consequence of self interactions of the ligand in the bulk and in the protein site $\{h^*, h^{**}\}$. Self-

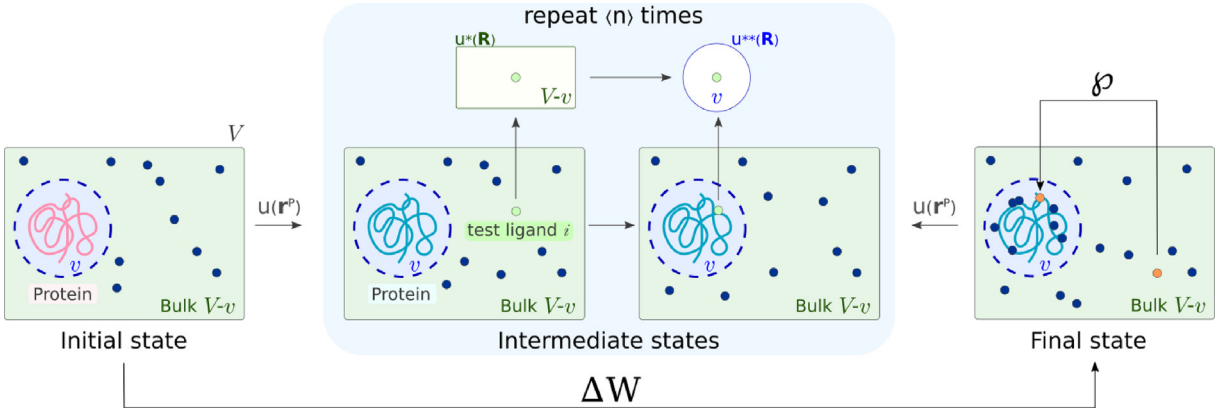

Fig. 1. Partition process. In a typical isothermal flooding MD simulation, one simulates a molecular system comprised of a protein embedded in a ligand-rich environment. As the simulation proceeds, ligands dissolve across the system to reversibly partition into the protein under equilibrium conditions, regardless of any assumptions about the underlying phenomenon. The partition process can be described along an idealized thermodynamic cycle in which (*i*) the protein P is first restrained with energy u to its final equilibrium structure; (*ii*) $\langle n\rangle$ ligands are decoupled from the bulk and the protein interface into the gas phase under applied restraints $\{u^*, u^{**}\}$; and (*iii*) restraints are released in the final state. Implicit in step (*ii*), and demonstrated in the results, is the consideration that in case of small ligands, the molecule typically does not adopt any special configuration in the final state. External potentials are then purely translational flat wells applied to confine the ligands within their equilibrium three-dimensional volumes in the bulk $V-v$ and in the protein interface v. As discussed in the text, the overall free-energy stability of the process ΔW can be described in terms of the partition coefficient of the ligand $\wp$.

interaction energies of the ligand, $\{h^*, h^{**}\}$, can be either attractive (<0) or repulsive (>0). In the approximated limit of comparable self-interaction energies in both phases

$$\delta \equiv \left[h^* \frac{N-\langle n\rangle}{V-v}c - h^{**}\frac{\langle n\rangle}{v}c\right] \approx 0, \tag{5}$$

such concentration effects are expected to cancel out over the free-energy difference $W^* - W^{**}$ and the partition coefficient to be conveniently re-expressed as a concentration invariant quantity of the desolvation free energies of the ligand under dilution *i.e.*, $\wp \approx e^{-\beta[\bar{\mu}-\bar{w}]}$ (Fig. 2). Note that invariance of $\wp$ may hold true only at non-saturating concentrations in which the protein site is not maximally occupied by the ligand. At saturating conditions, the partition coefficient is expected to become concentration dependent and to decrease monotonically with increasing concentrations of the ligand in the bulk – a conclusion that agrees with the recent work by Carlson and coworkers dealing with binding of a single probe molecule in the microscopic cavity of a protein receptor under saturation conditions [15].

Extension of the thermodynamic cycle shown in Fig. 1 over additional intermediate states constructed by the use of external potentials [13] allows the overall stability of the partition process ΔW to be properly evaluated relative to an initial state of the system in which all ligands are in the bulk. More specifically, ΔW can be evaluated along a reversible path in which (i) the protein P is first restrained with energy u to its final equilibrium structure, (ii) $\langle n\rangle$ ligands are decoupled from the bulk and the protein interface under restraints $\{u^*, u^{**}\}$ and (iii) restraints are released in the final state. According to that construction, the overall stability of the partition process with ω degenerate states then writes

$$e^{-\beta\Delta W} = \omega\left(\frac{v}{V-v}\right)^{\langle n\rangle} e^{-\beta\left[\Delta W_P + W^*_{\langle n\rangle} - W^{**}_{\langle n\rangle}\right]} \tag{6}$$

in terms of free-energy variations of the protein and ligands between initial and final states. For better readership of the work, rigorous derivation of Eq. (6) is presented on SI Eq. S5 through S13 in which each step of the thermodynamic cycle in Fig. 1 is properly defined as a reversible external work under restraints (see SI section-II for details). Briefly, ΔW_P is the net amount of reversible work coupled to the transformation of the protein internal structure between initial and final states and as such, it is a concentration dependent quantity that results from structural perturbations imposed by multi-ligand interactions. On the other hand,

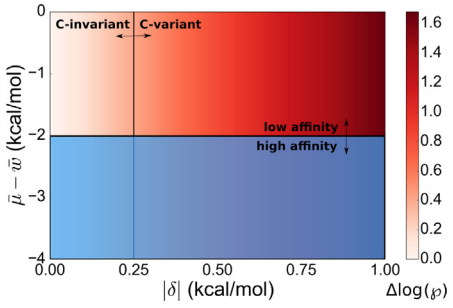

Fig. 2. Energetic conditions for the concentration-dependence of the partition coefficient of a small low-affinity molecule at non-saturating conditions. For typical low-affinity interactions $\bar{\mu} - \bar{w} \geq -2.0$ kcal/mol, the partition coefficient $\log(\wp) \leq 3.5$ is expected to be concentration independent $\Delta\log(\wp) \leq 0.5$ in case of minor concentration effects in the range of thermal energy $|\delta| \leq 0.25$ kcal/mol. Above that threshold, concentration effects are important and the partition coefficient depends on concentration. Color bar shows the variation of the partition coefficient relative to the case in which concentration effects are absent *i.e.*, $\Delta log(\wp) \equiv log(\wp(|\delta|)) - log(\wp(0))$.

$W^*_{\langle n\rangle} - W^{**}_{\langle n\rangle}$ is the net reversible work to decouple ligands from the system and as a state function, it can be expressed from Eq. (4)

$$W^*_{\langle n\rangle} - W^{**}_{\langle n\rangle} = \sum_{0\leqslant i\leqslant\langle n\rangle} W^*\left(\frac{N-i}{V-v}c\right) - W^{**}\left(\frac{i}{v}c\right) \tag{7}$$

along a step-wise equilibrium process in which a single molecule i is decoupled from the bulk and re-coupled into the protein interface occupied by $i-1$ molecules [16]. In the low-concentration regime, structural perturbations of the protein are expected to vanish $\Delta W_P \to 0$ and desolvation free-energies of the ligand to converge $W^*_{\langle n\rangle} - W^{**}_{\langle n\rangle} \to \langle n\rangle(\bar{\mu} - \bar{w})$ implying that,

$$e^{-\beta\Delta W} = \omega\left(\frac{v\wp}{V-v}\right)^{\langle n\rangle} = \left(\frac{N}{V-v}\right)^{\langle n\rangle} K \tag{8}$$

is the overall stability of the partition process of the ligands under dilution. In analogy to an aggregate binding constant of a small molecule over multiple microscopic binding sites of a protein receptor [8], Eq. (8) defines a macroscopic equilibrium constant for the average partition number that can be useful to analyze biochemical and functional data of the ligand. By taking into consideration the number of degenerate states $\omega \approx N^{\langle n\rangle}/\langle n\rangle!$ in the thermodynamic limit $\langle n\rangle \ll N$, note that proper regularization of Eq. (8) by concentration $(N/V-v)^{\langle n\rangle}$ provides us with the familiar binding constant equation *i.e.*, $K \equiv \frac{v^{\langle n\rangle}}{\langle n\rangle!} exp[-\beta\langle n\rangle(\bar{\mu} - \bar{w})]$. The analogy is particularly important as knowledge of the binding constant ensures the occupancy probability of the protein to be known and therefore, quantification of any ensemble average thermodynamic property of the system with biochemical and functional implications [8].

For completeness, note further that under moderate structural perturbations of the protein and surroundings, the unitary density of the ligand in the final partition state is expected to be a concentration-independent three-dimensional map that satisfies $\int d\boldsymbol{R}\bar{\rho}(\boldsymbol{R}) = 1$. The spatial distribution of the ligands in the low-concentration regime may be then conveniently approximated from Eq. (1) as a weighted function of the unitary density nearby the protein *i.e.*, $\rho(\boldsymbol{R}) = \langle n\rangle\bar{\rho}(\boldsymbol{R})\ \forall \boldsymbol{R} \in v$. For large protein receptors embedded in their native environment, structural perturbations are expected to be moderate within typical rmsd equilibrium values of the reference state free of ligands ≤5.0 Å.

1.2. Reconstruction of partition process from high concentrations

Taken together, the present analysis offers a self-consistent formulation for investigation of the dilute partition phenomenon of small low-affinity ligands from high concentrations. In practice, if the unitary-three-dimensional density of the ligand $\bar{\rho}'(\boldsymbol{R})$ and its derived partition coefficient $\wp'$ are known from high concentrations then a number of ligands is expected to partition into the protein when the total number N or their concentration is decreased in the system N/V

$$\langle n\rangle \approx \left(1 + \frac{V-v}{v\wp'}\right)^{-1} N \tag{9}$$

with an approximate free-energy stability and spatial density respectively given by

$$\Delta W \approx -\beta^{-1} ln(\omega) - \beta^{-1}\langle n\rangle ln\left(\frac{v\wp'}{V-v}\right) \tag{10}$$

and

$$\rho(\boldsymbol{R}) \approx \langle n\rangle\bar{\rho}'(\boldsymbol{R}) \tag{11}$$

Table 1
Equilibrium properties of flooding MD simulations.

Simulation	N(#)	V(Å^3)	$\frac{N}{V}$(M)	$\langle n\rangle$(#)	v(Å^3)	$\frac{\langle n\rangle}{v}$(M)	$N-\langle n\rangle$(#)	$V-v$(Å^3)	$\frac{N-\langle n\rangle}{V-v}$(M)	log($\wp$)
1	2	1,729.326 × 10^3	0.002	0.82	159.713 × 10^3	0.008	1.18	1,569.613 × 10^3	0.001	1.93
2	29	1,729.326 × 10^3	0.028	13.67	159.713 × 10^3	0.142	14.33	1,569.613 × 10^3	0.015	2.17
3	58	1,729.326 × 10^3	0.056	22.87	159.713 × 10^3	0.238	35.13	1,569.613 × 10^3	0.037	1.86
4	87	1,729.326 × 10^3	0.083	36.30	159.713 × 10^3	0.377	50.70	1,569.613 × 10^3	0.054	1.95
5	116	1,729.326 × 10^3	0.111	45.57	159.713 × 10^3	0.474	70.43	1,569.613 × 10^3	0.074	1.85
6	174	1,729.326 × 10^3	0.167	66.10	159.713 × 10^3	0.687	107.90	1,569.613 × 10^3	0.114	1.80

1.3. Computational methods

We rely on Eqs. (9), (10) and (11) to investigate the molecular partition of sevoflurane into Kv1.2 in the context of flooding or cosolvent MD simulations at high concentrations. Details of the calculations are provided as Supporting Information. Scripts for analysis and molecular configurations for flooding MD simulations of the sevoflurane/kv1.2 system can be downloaded from github ZENODO repository under DOI number (https://doi.org/10.5281/zenodo.6964766).

2. Results and discussion

The main goal here is to investigate the partition phenomenon of small ligands into proteins. The work is illustrated and discussed in the context of flooding or cosolvent MD simulations of the general anesthetic sevoflurane and the voltage-gated channel Kv1.2 [11].

2.1. Partition coefficient

As detailed in Table 1, Fig. 3 and Fig. S1, independent MD simulations were carried out with the channel embedded in a phospholipid bilayer and flooded with sevoflurane at concentrations ranging from 0.02 to 0.167 M. Consistent with the previously reported finding that sevoflurane may impact the ion channel function by interacting with one or more binding sites in the low-affinity mM concentration range [12], the simulated three-dimensional distribution of the ligand nearby the protein was found to be highly degenerate, occurring over multiple spots of comparable unitary probability density. Non-specific interactions of sevoflurane were then analyzed as a partition phenomenon at the molecular interface of the channel.

For each simulation system, the bulk volume $V-v$ was determined from the interface volume v; whereas v is defined to be within typical (≤5.0 Å) non-bonded distances of the protein [17]. Analysis of the simulations according to that volume decomposition reveals that ligand partition into the protein converges at a characteristic timescale of $t^* \geq 0.1\ \mu s$, with an average concentration-independent partition coefficient $log(\wp)$ of ∼1.93. That invariant property of $log(\wp)$ was independently investigated in terms of bulk (*) and protein (**) desolvation energies of the ligand at the same local concentrations of the flooding-MD simulations at equilibrium. Because the bulk of membrane proteins is made itself of aqueous (subscript 1) and lipid (subscript 2) regions:

$$e^{-\beta W^*} = \left(\frac{v_1}{V-v}\right)e^{-\beta W^*_1} + \left(\frac{v_2}{V-v}\right)e^{-\beta W^*_2}, \qquad (12)$$

desolvation free energies $\{W^*, W^{**}\}$ were estimated across each of the system's phases accordingly. Structural analysis of the ligand supports that no specific internal conformation or spatial orientation is adopted by the molecule across the system. Energy estimates were then obtained in the context of FEP calculations under applied translational flat wells (see SI section-I for further details; Fig. S2 and S3; Table-S1 and Table-S2).

The desolvation free-energy of sevoflurane in the bulk is concentration-dependent and largely dominated by favorable interactions with lipids, in agreement with past studies [20]. FEP calculations further support that W^{**} is another concentration-dependent quantity characterized by non-specific ligand interactions with distinct regions of the protein interface. Adjustment of the partition model in Eq. (4) to FEP estimates shows that $\{W^*, W^{**}\}$ respectively converges to their dilute values, $\bar{\mu}$ = 3.8 kcal.mol^{-1} and $\bar{w}$ = 5.0 kcal.mol^{-1}, in the low-concentration range (≤0.05 M) in which repulsive self interactions of the ligand are absent (h^* = 3.8 kcal.mol^{-1}; h^{**} = 0.90 kcal.mol^{-1}). Because self-

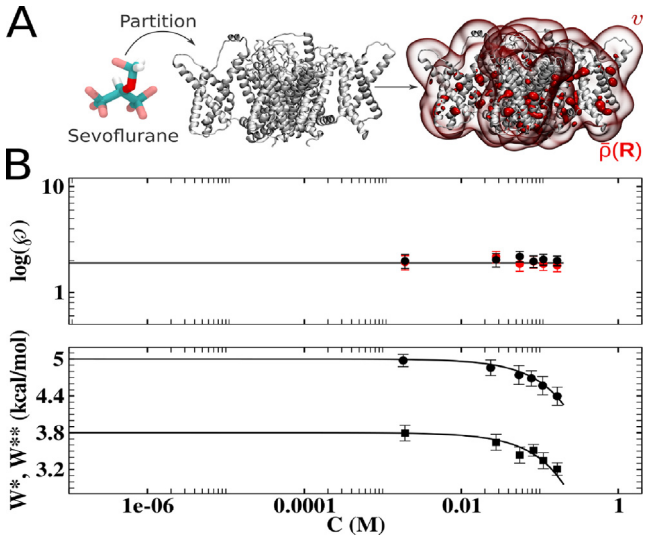

Fig. 3. Partition coefficient and desolvation free energies as a function of the total molar concentration C of the ligand. (A) Molecular structures of sevoflurane, Kv1.2 and isovalue representation of the unitary-three-dimensional distribution $\bar{\rho}(\boldsymbol{R})$ of the ligand within the interface volume v of the protein (red blobs). (B) Partition coefficient $log(\wp)$ and desolvation free energies$\{W^*, W^{**}\}$. Estimates of $log(\wp)$ (red circles) were computed from Eq. (2) by taking into consideration equilibrium MD properties reported in Table 1. Average value is ∼1.93 (continuous line). Estimates of $\{W^*, W^{**}\}$ and statistical errors were determined using the simple overlap sampling formula based on at least 10 independent FEP simulations *per* phase *per* concentration [18]. W^* was estimated by decoupling the ligand from bulk waters and lipids as devised in Eq. (12) (squares). W^{**} was estimated by decoupling the ligand from distinct regions of the channel including the S4 helix, S4S5 linker, S5S6 interface and central cavity (circles). According to the partition model (continuous line), the desolvation energy deviates from its dilute value as a result of unfavorable self interactions of the ligand at high concentration. The partition model was resolved as a linear fit of Eq. (4) to FEP estimates. Adjusted values with best regression coefficient (R > 0.99) are: ($\bar{\mu}$ = 3.8 kcal.mol^{-1}, h^*=3.8 kcal.mol^{-1}) and ($\bar{w}$ = 5.0 kcal.mol^{-1}, h^{**}= 0.90 kcal.mol^{-1}). Note that concentration effects cancel out over the free-energy difference $W^* - W^{**}$ and the derived partition coefficient (black circles) is a concentration-independent quantity that agrees with estimates obtained from flooding-MD. Statistical errors of $log(\wp)$ from flooding MD and FEP simulations were respectively obtained by bootstrap analysis of the partition number $\langle n\rangle$ and free-energy estimates $\{W^*, W^*\}$. Analysis involved 100 sets of 50 resampling values with replacement in each concentration. Analysis of the MD trajectories was performed in VMD [19]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

interaction energies are comparable between phases $\delta \approx 0$ (Table-S3), such concentration effects tend to cancel out over the free-energy difference $W^{*} - W^{**}$ and the derived partition coefficient $log(\wp)$ nicely agrees with the flooding-MD estimates in a concentration-independent manner. Note that, FEP and flooding-MD estimates are independent themselves thus strengthening the conclusion that $log(\wp)$ is invariant across the range of non-saturating concentrations under investigation.

2.2. Reconstruction of the partition process from high concentrations

From Fig. 3, the conclusion that $log(\wp)$ is concentration invariant allows the stability and structure of the dilute partition process to be known from high concentrations as devised in Eqs. (9), (10) and (11).

As a saturation curve, the number of protein-interacting ligands *versus* concentration $\langle n \rangle$ *x* $log(C)$ is a sigmoid function that reaches a plateau number of ligands at high concentrations. In Fig. 4A, the partition number resolved from flooding MD increases with concentration thus indicating that simulations were indeed carried out at non-saturating conditions. Eq. (9) provides the prediction of the number of ligands $\langle n \rangle$ that according to the high concentration partition coefficient $\wp'$ is expected to partition into the protein as a function of the total concentration of the system N/V. Based on the partition coefficient $log(\wp')$ of 1.80 known from flooding-MD at 0.167 M, Fig. 4A then shows the solution of Eq. (9) as a function of the number of molecules $0 \leq N \leq 200$ in the total solvent volume of the simulation box *i.e.*, V=1,729.326 $\times$ 10^3 Å^3. While agreeing with the flooding-MD estimates (regression coefficient R = 0.98), the model predicts that the number of partition events must vanish in the μM range as a consequence of the low affinity of the molecule to the protein – a conclusion that is highly consistent with experimental findings [21].

Given the small low-affinity nature of the ligand, the structural conditions underlying its partition across the system appear to fulfill the molecular premises in Eq. (6) thus allowing the stability of the process to be modeled accordingly, in terms of perturbations of the internal structure of the protein P, desolvation energies of $\langle n \rangle$ ligands L and entropic contribution S (see SI section-III for further details; Figs. S4, S5, S6 and S7). Despite the finite timescales of MD simulations, the protein structure shows a moderate yet clear tendency to deviate from its initial state with increasing ligand concentrations. Consistent with that structural perturbation induced by multi-ligand interactions, the free-energy change of the protein along the partition process ΔW_P was found to be a positive and concentration-dependent quantity that amount up to +10.0 kcal.mol^{-1} at high concentration and vanishes with depletion of the ligand. The total desolvation energy of the ligands along the partition process $\Delta W_L = W^{*}_{\langle n \rangle} - W^{**}_{\langle n \rangle}$ is in contrast a more significant, negative and concentration-dependent function that totals a minimum of -120.0 kcal.mol^{-1} at high concentration and vanishes with depletion of the ligand. Including entropic contributions ΔW_S, the total energetic outcome then points to a quite stable partition process $\Delta W = \sum \Delta W_i \leq 0$ primarily determined by favorable

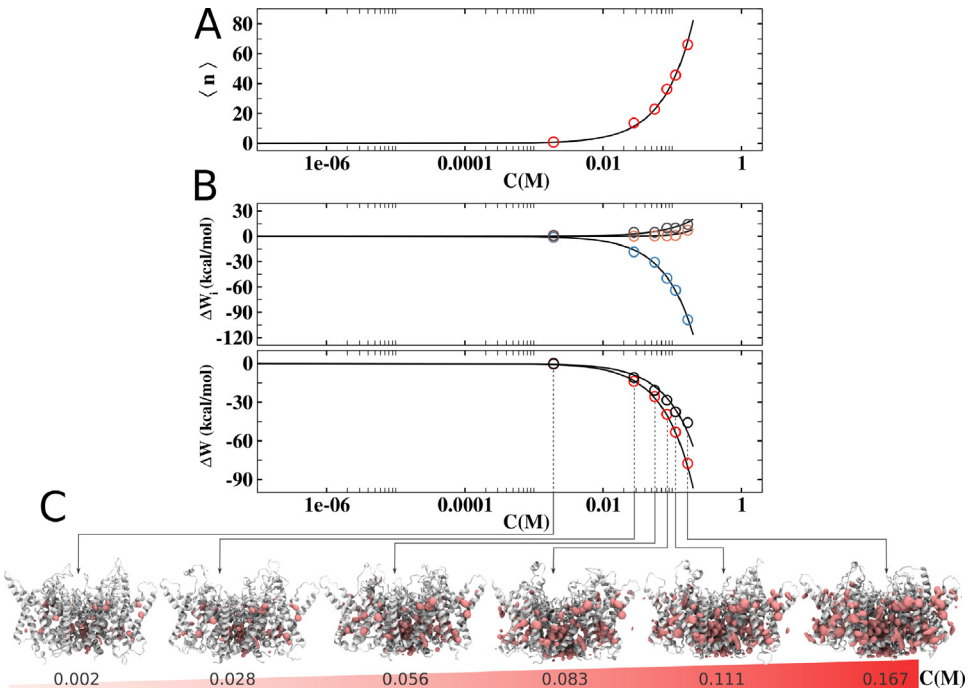

Fig. 4. Partition number, free-energy stability and structure as a function of the total molar concentration C of the ligand. (A) Partition number $\langle n \rangle$. According to Eq. (9), prediction of the partition number (line) agrees with flooding-MD estimates (circles). Eq. (9) was resolved for the partition coefficient $log(\wp') = 1.80$ as a function of the total concentration *i.e.*, number of molecules N in the total solvent volume of the simulation system V. (B) Partial ΔW_i and total ΔW free-energy variations. Estimates of ΔW_i stem from structural perturbations of the protein P (orange circles), desolvation energies of the ligand L (blue circles) and entropic S contributions (gray circles). Individual contributions were combined in the total free-energy variation of the partition process $\Delta W = \sum \Delta W_i$ (red circles) and subsequently compared to estimates obtained from Eq. (10) in which the total stability of the partition process is approximated from entropic contributions and the partition coefficient of the ligand *i.e.*, $\Delta W' \approx \Delta W_S - \beta^{-1} \langle n \rangle ln(\wp')$ (black circles). As respectively detailed in Figs. S4–S7, free-energy estimates $\Delta W_{i=\{P,L,S\}}$, ΔW and $\Delta W\prime$ were linearly plotted as a function of the total molar concentration of the system $C = (N/V)c$ and a linear or quadratic function (line) was adjusted to the data *i.e.*, $\Delta W_P = h_P C^2$, $\Delta W_L = h_L C$, $\Delta W_S = h_S C$, $\Delta W = hC$ and $\Delta W' = h'C$. Adjusted values with best regression coefficient (R > 0.93) are: (h_P=200.0 kcal.mol^{-1}, h_L=−580.0 kcal.mol^{-1}, h_S=100.0 kcal.mol^{-1}, h=−480.0 kcal.mol^{-1} and h'=−320.0 kcal.mol^{-1}). (C) Three-dimensional distribution $\rho(\boldsymbol{R})$ at isovalues of 1.0 $\times$ 10^{-4} Å^{-3}. Concentration-dependent spatial densities of the ligand (red) at the interface of the protein (white) were reconstructed at 0.002, 0.028, 0.056, 0.083, 0.111 and 0.167 M. Reconstruction followed Eq. (11), by respectively reweighing the unitary density of the molecule resolved from flooding MD at high concentration (0.167 M) by the partition number in each concentration. Because every reconstructed density derives from the same unitary density $\bar{\rho}'(\boldsymbol{R})$ at 0.167 M, differences between any two concentration-dependent distributions result from the respective reweighing partition numbers $\langle n \rangle$ considered in the procedure (Table 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

desolvation energies of the ligands that surmount minor destabilizing perturbations of the protein structure across the entire concentration range. Because concentration effects accounting for structural perturbations of the protein and self-interactions of the ligand disappear in the dilute regime, ΔW becomes a function of the partition coefficient of the molecule and its partition number, and as such, it can be well approximated at low concentrations by Eq. (10). According to that approximation, sevoflurane partition into Kv1.2 appears to be stabilized by less than 10.0 kcal.mol^{-1} across the range of experimental concentrations ($\leq$0.01 M) in which general anesthetics affect ion channels [22].

Because perturbations of the protein structure induced by ligand partition are moderate within typical root-mean-square deviations in solution $\leq$5.0 Å, the equilibrium spatial distribution of the ligand at the protein interface $\rho(\boldsymbol{R})$ was reconstructed from Eq. (11) by reweighing the unitary density of the molecule resolved from MD at high concentrations. Better conveying the partition phenomenon under investigation, $\rho(\boldsymbol{R})$ is uniformly spread over multiple concentration dependent sites that are preferentially localized at the protein interface with lipids as a consequence of favorable interactions of the ligand with that moiety of the bulk. Careful inspection of the distribution model supports that multiple interaction spots may equally contribute to drug modulation of the channel in the low-concentration regime ($\leq$0.01 M), including the gating-implicated S4S5 linker previously reported in photolabeling experiments [23,24]. Supporting the conclusion of a degenerate mode of action, the ability of general anesthetics to modulate Kv channels was indeed found not to depend on specific interactions of the molecule at established cavities of the protein [25]. Further implications for the action of general anesthetics on ion channels will be discussed elsewhere (manuscript in preparation).

3. Concluding remarks

Proteins are targets for a large family of ligands, including small low-affinity molecules featuring a wide spectrum of biological roles. How such ligands modulate protein function must build on understanding their highly degenerate atomic-level interactions under dilution, thus currently challenging the resolution of both theoretical and experimental routines. Looking for new developments in the field, we conceive the interaction mode of small low-affinity ligands as a partition phenomenon governed by non-specific interactions at the molecular surface of proteins. According to that molecular description, thermodynamic analysis of concentration-dependent self-interactions of the ligand and structural perturbations of the protein allows the partition coefficient, stability and structure of the dilute partition process to be reconstructed from the unitary equilibrium probability density of the ligand $\bar{\rho}'(\boldsymbol{R})$. The result stems essentially from the fact that under non-disruptive chemical forces on the protein structure and surroundings, the spatial distribution of the ligand is expected to be a concentration-independent three-dimensional map that can be efficiently learned at high concentrations.

Illustration and discussion of the thermodynamic analysis is made here in the context of flooding-MD simulations of the general anesthetic sevoflurane and the important neuronal membrane protein Kv1.2, illuminating its great utility for the structural biology field. Ideally, the dilute partition process of a small low-affinity molecule would be resolved from MD directly if such simulations did not require very long time scales to converge at low concentrations. By allowing the dilute partition process to be reconstructed directly from high-concentration conditions in which sampling is significantly enhanced over shorter time scales, our thermodynamic analysis is then expected to unlock the power of MD in the low concentration regime. Especially true for the more sampling-sensitive structural properties, the reconstructed spatial distribution of the ligand in the low-concentration regime provides us with a larger set of interaction spots nearby the protein than the same estimates directly resolved from MD in consequence of poor sampling of the partition process over the finite timescale of the simulation. Supporting that conclusion (Fig. S8), the low-concentration mismatch between the volume occupied by the ligand in the *observed* and *predicted* spatial probability densities is significantly decreased by enhancing sampling either via symmetrization across the fourfold symmetry of the channel or substantial extension of simulation time (over $\sim$5 μs at 5 mM of sevoflurane).

Considering the complete space of physical chemical characteristics of ligands, protein and environment, and the combinatorial arrangements thereof, a myriad of scenarios are possible. While it would be difficult to list and evaluate all of them, the present formulation, devised to study low-affinity ligands, has underlying assumptions that are required for it's proper applicability in the context of flooding MD simulations. First, we hereby define low-affinity ligand as a class of molecules that interact with proteins in the mM concentration range. Second, as a consequence of low affinity, the ligand's distribution nearby the protein should be degenerate, occurring over multiple spots of comparable unitary spatial probability density $\bar{\rho}'(\boldsymbol{R})$. If both conditions are true, than the ligand interaction is likely better described by a partition phenomenon, in which the molecule occupies the protein with a characteristic partition coefficient $log(\wp') \leq 3.5$ and predominant loss of translational freedom. Third, if the partition phenomenon is indeed concentration independent, $\log(\wp)$ calculated in a second, lower, concentration provides a $\Delta\log(\wp) \leq 0.5$ (*cf.* Fig. 2), indicating that the probability density of the ligand and the derived partition coefficient are concentration-invariant and can be efficiently learned from high concentrations.

Naturally, invariance of the unitary distribution and the derived partition coefficient only hold true at non-saturating concentrations in which the partition number $\langle n \rangle$ does not saturate and perturbations of the protein structure are moderate, within typical RMSD equilibrium values of the reference state free of ligands $\leq$5.0 Å. As formulated herein, it is important to clarify that the low-affinity description is not suitable for processes involving high affinity binders such as, charged molecules, lipids and peptides. In such high-affinity scenario, the binder interacts preferentially at one specific microscopic cavity of the protein receptor with a position-dependent desolvation energy $W^{**}(\boldsymbol{R})$ and loss of translational, rotational and conformational freedom. The implications of the latter are important for ligand interaction and its description must include additional gas-phase intermediates to define a proper binding constant for accurate description of the binding process and the derived equilibrium properties at the dilute low concentration range [13,14]. An aggregate constant may be defined accordingly in case of two or more high-affinity sites [8].

Abiding by the same requirements, independent flooding-MD calculations for the sedative hypnotic drug trichloroethanol [26] and the aqueous-soluble bovine serum albumin [27] further support that the thermodynamic model may be applicable to other two-component interacting systems guided by non-bonded low-affinity interactions (Table-S4, Fig. S9 and S10). Consistent with fluorescence quenching experiments showing that trichloroethanol interacts with albumin in the low-affinity mM range [28], the spatial distribution of the ligand resolved from simulation $\bar{\rho}'(\boldsymbol{R})$ is degenerate over multiple spots of comparable unitary probability density at the protein layer. Better conveying a partition phenomenon, the partition coefficient of the ligand $log(\wp')\approx3.21$ is shown to be clearly concentration-independent $\Delta log(\wp)\approx0.25$ across two non-saturating conditions 50 and 100 mM in which the partition number $\langle n \rangle$ does not saturate and structural

deviations of the protein are within equilibrium RMSD values of the reference state free of ligands $\leq$5.0 Å. Solution of the thermodynamic model then allows the partition coefficient, stability and structure of the dilute partition process to be reconstructed from the unitary density of the ligand at high concentration. Careful inspection of the distribution model supports that multiple interaction spots may equally contribute for drug interaction in the low-concentration regime, including tryptophan regions previously reported in fluorescence experiments [28].

Because proteins are primary targets for small low-affinity molecules that can be widely investigated by the scientific community in the context of a partition phenomenon at high concentrations, we thus believe the study is of broad interest and likely useful in producing new results in the field. With promising prospects, implementation and convergence of the model in the context of MD simulations appears easier than other more elaborate and time consuming approaches [29], making it especially attractive for a large number of applications. By revealing more traceable atomic-level spots at the molecular surface of proteins, we anticipate that modeling-based structural information might be especially useful to delineate and interpret novel high-throughput mass spectrometry experiments [9] aimed at characterizing functional interactions of non-specific ligands under physiological conditions. Reconstruction of the dilute spatial distribution of two distinct solvents from high concentrations might be of utility to potentiate identification of active and allosteric binding sites in protein receptors in the context of mixed-solvent simulations [30].

CRediT authorship contribution statement

Leonardo Cirqueira: Investigation, Software, Visualization, Validation, Writing – review & editing. **Letícia Stock:** Investigation, Software, Visualization, Validation, Writing – review & editing. **Werner Treptow:** Conceptualization, Methodology, Software, Visualization, Validation, Writing – original draft, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank B. Roux for helpful discussions. The work was supported by National Council for Scientific and Technological Development CNPq [grant number 302089/2019-5 and 200114/2020-4 (WT)], Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES [grant number 23038.010052/2013-95 (WT)], and Fundação de Apoio a Pesquisa do Distrito Federal FAPDF [grant number 193.001.202/2016 (WT)]. LS thanks CAPES for postdoctoral fellowship (grant number 88882.463151/2019-01).

Author contributions

LC and LS contributed equally to this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.08.049.

References

[1] Uversky VN. Intrinsic disorder-based protein interactions and their modulators. Curr Pharm Des 2013;19(23):4191–213.
[2] Silva ITG, Fernandes V, Souza C, Treptow W, Santos GM. Biophysical studies of cholesterol effects on chromatin. J Lipid Res 2017;58(5):934–40.
[3] Fernandes V, Teles K, Ribeiro C, Treptow W, Santos G. Fat nucleosome: Role of lipids on chromatin. Prog Lipid Res 2018;1(70):29–34.
[4] Musacchio A. On the role of phase separation in the biogenesis of membraneless compartments. EMBO J 2022:e109952.
[5] Chartier M, Morency LP, Zylber MI, Najmanovich RJ. Large-scale detection of drug off-targets: hypotheses for drug repurposing and understanding side-effects. BMC Pharmacol Toxicol 2017;18(1):18.
[6] George DE, Tepe JJ. Advances in Proteasome Enhancement by Small Molecules. Biomolecules 2021;11(12):1789.
[7] Uversky VN. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. Front Aging Neurosci [Internet]. 2015 [cited 2022 Feb 7]; Available from: https://www.frontiersin.org/article/10.3389/fnagi.2015.00018.
[8] 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.
[9] McLaren DG, Shah V, Wisniewski T, Ghislain L, Liu C, Zhang H, et al. High-throughput mass spectrometry for hit identification: current landscape and future perspectives. SLAS Discov Adv Sci Drug Discov 2021;26(2):168–91.
[10] Guvench Jr O, Adm.. Computational fragment-based binding site identification by ligand competitive saturation. PLOS Comput Biol 2009;5(7):e1000435.
[11] Long SB, Campbell EB, MacKinnon R. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. Science 2005;309:897–903.
[12] Stock L, Hosoume J, Cirqueira L, Treptow W. Binding of the general anesthetic sevoflurane to ion channels. PLOS Comput Biol 2018;14(11):e1006605.
[13] Woo HJ, Roux B. Calculation of absolute protein–ligand binding free energy from computer simulations. Proc Natl Acad Sci U S A 2005;102(19):6825–30.
[14] 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):6560–73.
[15] Ghanakota P, DasGupta D, Carlson HA. Free Energies and Entropies of Binding Sites Identified by MixMD cosolvent simulations. J Chem Inf Model 2019;59(5):2035–45.
[16] Roux B, Nina M, Pomès R, Smith JC. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. Biophys J 1996;71(2):670–81.
[17] Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 1996;38(3).
[18] Lu N, Kofke DA, Woolf TB. Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. J Comput Chem 2004;25(1):28–40.
[19] Humphrey W, Dalke A, Schulten K. VMD - visual molecular dynamics. J Molec Graph 1996;14(1):33–8.
[20] Pickholz M, Saiz L, Klein ML. Concentration effects of volatile anesthetics on the properties of model membranes: A coarse-grain approach. Biophys J 2005;88(3):1524–34.
[21] Barber AF, Liang Q, Covarrubias M. Novel activation of voltage-gated K+ channels by sevoflurane. J Biol Chem 2012;4(287):40425–32.
[22] Liang Q, Anderson WD, Jones ST, Souza CS, Hosoume JM, Treptow W, et al. Positive Allosteric Modulation of Kv Channels by Sevoflurane: Insights into the Structural Basis of Inhaled Anesthetic Action. PLoS ONE 2015 Nov 24;10(11):e0143363.
[23] Bu W, Liang Q, Zhi L, Maciunas L, Loll PJ, Eckenhoff RG, et al. Sites and functional consequence of alkylphenol anesthetic binding to Kv1.2 channels. Mol Neurobiol. 2017;1–11.
[24] Woll KA, Peng W, Liang Q, Zhi L, Jacobs JA, Maciunas L, et al. Photoaffinity ligand for the inhalational anesthetic sevoflurane allows mechanistic insight into potassium channel modulation. ACS Chem Biol 2017;12(5):1353–62.
[25] Bu W, Liang Q, Zhi L, Maciunas L, Loll PJ, Eckenhoff RG, et al. Sites and functional consequence of alkylphenol anesthetic binding to Kv1.2 channels. Mol Neurobiol 2018;55(2):1692–702.
[26] Kiametis AS, Stock L, Cirqueira L, Treptow W. Atomistic model for simulations of the sedative hypnotic drug 2,2,2-trichloroethanol. ACS Omega 2018;3(11):15916–23.
[27] Bujacz A. Structures of bovine, equine and leporine serum albumin. Acta Crystallogr D Biol Crystallogr 2012;68(Pt 10):1278–89.
[28] Solt K, Johansson JS. Binding of the active metabolite of chloral hydrate, 2,2,2-trichloroethanol, to serum albumin demonstrated using tryptophan fluorescence quenching. Pharmacology 2002;64(3):152–9.
[29] Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biol 2011;9(1):71.
[30] Ghanakota P, Carlson HA. Moving beyond active-site detection: MixMD applied to allosteric systems. J Phys Chem B 2016;120(33):8685–95.