22 May 2026

Flexi-GIST: Solvation Thermodynamic Decomposition and Mapping for Flexible Molecules

Helmut Carter, Joseph Cruz, Steven Ramsey, Tom Kurtzman

Abstract

Solvation mapping methods provide spatially resolved estimates of solvent thermodynamic quantities but are formally restricted to rigid representations of molecular systems. This restriction limits their applicability to biologically relevant molecules, for which conformational flexibility can significantly influence both solvation free energies and the spatial distribution of solvent thermodynamics. Although the underlying statistical mechanical formalism has been defined for flexible systems, its direct evaluation requires an ensemble average over solute configurations and is computationally intractable for all but the simplest systems. As a result, solvation mapping is typically applied to only a single rigid solute conformation. To address this limitation, we present Flexi-GIST, an extension of Grid Inhomogeneous Solvation Theory (GIST) that enables spatially decomposed solvation thermodynamics estimates for flexible molecules. Flexi-GIST approximates the formally exact ensemble average over solute configurations by clustering them and evaluating GIST on representative conformations for each cluster. The representative solvation thermodynamic estimates are then combined to approximate the thermodynamics of solvating the flexible molecule. Solute configurational entropy contributions are accounted for using a nearest-neighbor entropy estimator with a truncated mutual information expansion. We validated the Flexi-GIST methodology on a set of flexible small molecules in explicit solvent. Solvation free energies obtained by spatial integration of solvation maps are compared to values computed using thermodynamic integration of flexible solutes, demonstrating agreement between the two methods. This framework extends solvation mapping to flexible molecular systems and provides a basis for estimating spatially resolved solvation thermodynamics in cases where rigid representations are insufficient.

Posted on 22 May 2026 — CC-BY 4.0 — This is a preprint and has not been peer reviewed. Data may be preliminary. — https://doi.org/10.26434/chemrxiv.15003769/v1

Flexi-GIST: Solvation Thermodynamic Decomposition and Mapping for Flexible Molecules

Helmut Carter, Joseph Cruz, Steven Ramsey, Tom Kurtzman

May 20, 2026

Abstract

Solvation mapping methods provide spatially resolved estimates of solvent thermodynamic quantities but are formally restricted to rigid representations of molecular systems. This restriction limits their applicability to biologically relevant molecules, for which conformational flexibility can significantly influence both solvation free energies and the spatial distribution of solvent thermodynamics. Although the underlying statistical mechanical formalism has been defined for flexible systems, its direct evaluation requires an ensemble average over solute configurations and is computationally intractable for all but the simplest systems. As a result, solvation mapping is typically applied to only a single rigid solute conformation. To address this limitation, we present Flexi-GIST, an extension of Grid Inhomogeneous Solvation Theory (GIST) that enables spatially decomposed solvation thermodynamics estimates for flexible molecules. Flexi-GIST approximates the formally exact ensemble average over solute configurations by clustering them and evaluating GIST on representative conformations for each cluster. The representative solvation thermodynamic estimates are then combined to approximate the thermodynamics of solvating the flexible molecule. Solute configurational entropy contributions are accounted for using a nearest-neighbor entropy estimator with a truncated mutual information expansion. We validated the Flexi-GIST methodology on a set of flexible small molecules in explicit solvent. Solvation free energies obtained by spatial integration of solvation maps are compared to values computed using thermodynamic integration of flexible solutes, demonstrating agreement between the two methods. This framework extends solvation mapping to flexible molecular systems and provides a basis for estimating spatially resolved solvation thermodynamics in cases where rigid representations are insufficient.

1 Introduction

Water plays a central role in molecular recognition, and substantial effort has been devoted to quantifying solvent contributions to protein–ligand binding. Solvation mapping methods characterize the spatial distribution of solvent thermodynamic properties around proteins and ligands, providing mechanistic insight into molecular recognition. A wide range of

tools have been developed for this purpose, including Grid Inhomogeneous Solvation Theory (GIST) and related approaches[1, 2], which map solvent energetic and entropic properties on a three-dimensional grid.

Despite their success[3, 4, 5], current solvation mapping methods are formally restricted to rigid representations of molecular systems. In practice, solvation thermodynamics are typically evaluated for a single protein conformation, often an experimentally-determined structure. This limitation is significant, as biologically relevant molecules like proteins are inherently flexible, and solvent structure and thermodynamics can be highly sensitive to even modest conformational fluctuations of binding sites [6]. As a result, single-conformation solvation maps may fail to capture conformation-dependent solvation thermodynamics, potentially leading to biased or inaccurate free energy estimates and overlooking thermodynamically relevant alternative binding-site conformations.

The theoretical foundations of inhomogeneous solvation theory (IST), on which GIST is based, are exact for both rigid and flexible solutes. However, applying IST-based methods to flexible systems is computationally infeasible, as it requires averaging solvent thermodynamic quantities over an ensemble of solute configurations. In practice, existing solvation mapping methods analyze only a single solute conformation. Recent work from our group has shown that this can lead to substantial overestimation of the solvent contribution to protein–ligand binding free energies, with errors exceeding 10 kcal/mol in some cases [6]. Furthermore, incomplete treatment of protein flexibility in structure-based drug design reduces the chemical space that is complementary to the binding site, limiting exploration in drug discovery. These findings highlight the need for solvation mapping approaches that explicitly account for solute flexibility. Several studies have suggested mitigating this limitation by performing solvation analyses on a small number of representative conformations, but no statistical mechanical foundation or systemic validation for this strategy has yet been published [2, 7].

Here, we introduce Flexi-GIST, an extension of GIST that enables spatially resolved solvation thermodynamic analysis of flexible molecular systems. Flexi-GIST approximates the formally exact ensemble average over solute configurations by applying GIST to representative conformations chosen by clustering, which are subsequently combined using statistically consistent weighting. By validating Flexi-GIST on solvation free energy estimates of small molecules, we offer a framework for analyzing how solute flexibility influences solvent structure and thermodynamics, with direct implications for binding affinity prediction and the identification of thermodynamically relevant binding site conformations.

We first present the theoretical background for extending solvation mapping to flexible solutes, including the decomposition of solvation free energy into solvent, solute, and conformational contributions. We then describe the computational workflow used to approximate the flexible-solute ensemble average, including MD sampling, conformational clustering, GIST calculations on representative conformations, and population-weighted recombination of thermodynamic quantities. The approximate method is validated in two stages: first, by comparing GIST-derived end-state solvation free energies for rigid conformations against thermodynamic integration calculations, and second, by comparing Flexi-GIST estimates for flexible solutes against fully flexible TI reference free energies. Finally, we discuss the implications of incorporating solute flexibility into solvation mapping, current limitations of the approach, and future applications to protein binding sites in structure-based drug design.

2 Conditional Probability Decomposition of the Entropy of a Solute in Solvent

The entropy S of a solute and solvent system can be fully represented by:

$$S_{sys} = -k_B \int P(\mathbf{q}, \mathbf{r}^N) \ln P(\mathbf{q}, \mathbf{r}^N) d\mathbf{q} d\mathbf{r}^N \tag{1}$$

Where k_B is Boltzmann's constant, P is the probability of sampling a given configuration, $\mathbf{q}$ fully specifies the position, orientation, and configuration of a solute molecule and $\mathbf{r}^N$ fully describes the positions and orientations of N solvent molecules.

Conditioning the solvent distribution on the solute configuration yields

$$S_{solvent} = -k_B \int P(\mathbf{q}) P(\mathbf{r}^N|\mathbf{q}) \ln P(\mathbf{q}) P(\mathbf{r}^N|\mathbf{q}) d\mathbf{q} d\mathbf{r}^N \tag{2}$$

Which separates into two exact contributions:

$$S_{solute} = -k_B \int P(\mathbf{q}) \ln P(\mathbf{q}) \, d\mathbf{q} \tag{3}$$

$$S_{solvent} = -k_B \int P(\mathbf{q}) \, P(\mathbf{r}^N|\mathbf{q}) \ln P(\mathbf{r}^N|\mathbf{q}) \, d\mathbf{q} \, d\mathbf{r}^N \tag{4}$$

so that $S_{sys} = S_{solute} + S_{solvent}$.

Evaluating $S_{solvent}$ in (4) exactly requires sampling the solvent distribution for every solute configuration $\mathbf{q}$ sampled during the simulation. In practice, this is computationally intractable: for a single rigid conformation of a small molecule, GIST requires approximately 8 hours of GPU time for MD simulation and an additional 5 hours of CPU time for GIST entropy analysis, and for proteins these simulations often take two to three times as long. Repeating this procedure for the thousands of solute configurations sampled in a typical MD simulation would therefore be infeasible.

Solvation mapping methods therefore adopt the Percus source particle approximation, in which the solute is treated as a rigid body fixed at a single conformation $\mathbf{Q}$ [8]. This is the standard mode of operation for IST-based methods such as STOW, GIST, WaterMap, and SSTMap, as well as other solvation methods including SZmap, 3D-RISM, and MM-GBSA [9, 10, 2, 11, 12, 13].

Under this approximation, the solute distribution collapses to a delta function:

$$P(\mathbf{q}) = \delta(\mathbf{q} - \mathbf{Q}), \tag{5}$$

and the joint distribution of solute and solvent configurations factorizes as

$$P(\mathbf{r}^N, \mathbf{q}) = P(\mathbf{r}^N \mid \mathbf{Q}) \, \delta(\mathbf{q} - \mathbf{Q}). \tag{6}$$

Since the solute has only a single conformation, the solute conformational entropy vanishes,

$$S_{solute} = -k_B \int P(\mathbf{q}) \ln P(\mathbf{q}) \, d\mathbf{q} = 0, \tag{7}$$

and the total solvation entropy reduces entirely to the solvent contribution, which depends only on the conditional solvent distribution at the fixed conformation $\mathbf{Q}$:

$$S_{solvent}(\mathbf{Q}) = -k_B \int P(\mathbf{r}^N \mid \mathbf{Q}) \ln P(\mathbf{r}^N \mid \mathbf{Q})\, d\mathbf{r}^N. \tag{8}$$

Thus, within the rigid-solute approximation, the solvation entropy calculation requires sampling only the solvent degrees of freedom $\mathbf{r}^N$ in the presence of the fixed solute conformation $\mathbf{Q}$, which is precisely the quantity evaluated by GIST and related methods.

3 Cluster Decomposition of the Solvent Entropy

Here we introduce a clustering method to approximately evaluate the solvent entropy of a flexible solute. We first divide the the integral over solute configurations in (4) into k non-overlapping subvolumes

$$S_{solvent} = -k_B \sum_{i=1}^{k} \int_i P(\mathbf{q})\, P(\mathbf{r}^N \mid \mathbf{q})\, \ln P(\mathbf{r}^N \mid \mathbf{q})\, d\mathbf{q}\, d\mathbf{r}^N. \tag{9}$$

Defining the probability that the solute configuration belongs to cluster i as

$$P_i = \int_i P(\mathbf{q})\, d\mathbf{q}, \tag{10}$$

we write $P(\mathbf{q}) = P_i \left(\dfrac{P(\mathbf{q})}{P_i} \right)$ within each cluster, so that (9) becomes

$$S_{solvent} = -k_B \sum_{i=1}^{k} P_i \int_i \frac{P(\mathbf{q})}{P_i}\, P(\mathbf{r}^N \mid \mathbf{q})\, \ln P(\mathbf{r}^N \mid \mathbf{q})\, d\mathbf{q}\, d\mathbf{r}^N. \tag{11}$$

Note that $P(\mathbf{q})/P_i$ is a properly normalized probability distribution over solute configurations within cluster i. We can therefore identify the exact, probability-weighted solvent entropy for each cluster:

$$S_{solvent} = \sum_{i=1}^{k} P_i\, S^{exact}_{i,solvent} \tag{12}$$

where

$$S^{exact}_{i,solvent} = -k_B \int_i \frac{P(\mathbf{q})}{P_i}\, P(\mathbf{r}^N \mid \mathbf{q})\, \ln P(\mathbf{r}^N \mid \mathbf{q})\, d\mathbf{q}\, d\mathbf{r}^N. \tag{13}$$

Equation (12) is exact. However, evaluating $S^{exact}_{i,solvent}$ still requires sampling the solvent distribution for every solute configuration $\mathbf{q}$ within each cluster, which is computationally intractable for the same reasons discussed above.

To address this, we make the approximation that within each cluster i, the solvent distribution is independent of the solute configuration $\mathbf{q}$:

$$P(\mathbf{q}, \mathbf{r}^N) = P(\mathbf{q}) P(\mathbf{r}^N|\mathbf{q}) \approx P(\mathbf{q}) P_i(\mathbf{r}^N) \tag{14}$$

and $P(\mathbf{r}^N|\mathbf{q}) = P_i(\mathbf{r}^N)$ within each cluster i where the subscript i in $P_i(\mathbf{r}^N)$ represents that it is the solvent distribution within cluster i

$$P_i \int_i \frac{P(\mathbf{q})}{P_i} P(\mathbf{r}^N|\mathbf{q}) \ln P(\mathbf{r}^N|\mathbf{q}) d\mathbf{q} d\mathbf{r}^N = P_i \int_i \frac{P(\mathbf{q})}{P_i} P_i(\mathbf{r}^N) \ln P_i(\mathbf{r}^N) d\mathbf{q} d\mathbf{r}^N \tag{15}$$

With this approximation, $P_i(\mathbf{r}^N)$ is independent of $\mathbf{q}$ within each cluster i and the integrals over solute and solvent coordinates in the RHS of (15) factorize.The integral over the solute configurations $\mathbf{q}$ gives P_i as $P_i = \int_i P(\mathbf{q}) d\mathbf{q}$ therefore equation (15) becomes:

$$P_i \int_i \frac{P(\mathbf{q})}{P_i} P(\mathbf{r}^N|\mathbf{q}) \ln P(\mathbf{r}^N|\mathbf{q}) d\mathbf{q} d\mathbf{r}^N = P_i \int P_i(\mathbf{r}^N) \ln P_i(\mathbf{r}^N) d\mathbf{r}^N \tag{16}$$

Note that the subscript i is dropped from the integral here because it is an integral over all solvent configurations $\mathbf{r}^N$ but the probability distribution of the solvent is from only the solute configurations within cluster i.

$$S_{i,solvent} = \int P_i(\mathbf{r}^N) \ln P_i(\mathbf{r}^N) d\mathbf{r}^N \tag{17}$$

Here, $S_{i,solvent}$ is the entropy evaluated from the solvent distribution $P_i(\mathbf{r}^N)$ for cluster i and the quantity estimated by GIST.

3.1 Full Entropy Decomposition

Combining the solute entropy (3) with the cluster-decomposed solvent entropy, the total entropy of a flexible solvated system is

$$\boxed{S = S_{solute} + \sum_{i=1}^{k} P_i \, S_{i,solvent}} \tag{18}$$

where S_{solute} is the full solute configurational entropy and each $S_{i,solvent}$ is the solvent positional and orientational entropy obtained by performing solvation mapping (e.g. GIST) for a representative solvent distribution around cluster i.

The solute entropy can itself be decomposed into cluster contributions using the identity

$$S_{solute} = -k_B \sum_{i=1}^{k} P_i \ln P_i + \sum_{i=1}^{k} P_i \, S_{i,vib}, \tag{19}$$

where $-k_B \sum_i^k P_i \ln P_i$ is the inter-cluster (conformational) entropy and

$$S_{i,vib} = -k_B \int_i \frac{P(\mathbf{q})}{P_i} \ln \frac{P(\mathbf{q})}{P_i} \, d\mathbf{q} \tag{20}$$

is the intra-cluster vibrational entropy. Substituting (19) into (18) recovers the full decomposition:

$$S_{sys} \approx -k_B \sum_{i=1}^{k} P_i \ln P_i + \sum_{i=1}^{k} P_i \, S_{i,vib} + \sum_{i=1}^{k} P_i \, S_{i,solvent}. \tag{21}$$

Equation (21) naturally defines a per-cluster solute entropy

$$S_{i,solute} = -k_B \ln P_i + S_{i,vib}, \tag{22}$$

which combines the inter-cluster conformational contribution $-k_B \ln P_i$ with the intra-cluster vibrational entropy $S_{i,vib}$. The total entropy for cluster i is then

$$S_i = P_i \, S_{i,solute} + P_i \, S_{i,solvent}, \tag{23}$$

so that

$$S_{sys} \approx \sum_{i=1}^{k} S_i \tag{24}$$

The two terms in eq. (23) represent, respectively, the solute entropy contribution and the solvation entropy contribution of cluster i, each weighted by the cluster's normalized population P_i. This cluster-level assignment provides a natural bookkeeping framework that we now extend to the internal energy of the system.

In practice, this approximate system entropy can be evaluated for a finite number of conformations, making the computational calculations tractable. The implementation requires that solute configurations be clustered such that their structural fluctuations are small enough that the solvent distribution, and therefore the entropy estimated from it, is insensitive to the solute fluctuations within each cluster.

3.2 Exact Cluster Decomposition of the Internal Energy

In contrast to the entropy, the decomposition of the system's internal energy into cluster contributions is exact and requires no approximation analogous to the one introduced in Section 3. The total internal energy of the solute–solvent system is

$$E = \int P(\mathbf{q}, \mathbf{r}^N) \, U(\mathbf{q}, \mathbf{r}^N) \, d\mathbf{q} \, d\mathbf{r}^N, \tag{25}$$

where $U(\mathbf{q}, \mathbf{r}^N)$ is the total potential energy of the configuration $(\mathbf{q}, \mathbf{r}^N)$. The potential energy separates exactly into three physically distinct contributions,

$$U(\mathbf{q}, \mathbf{r}^N) = U_s(\mathbf{q}) + U_{sw}(\mathbf{q}, \mathbf{r}^N) + U_{ww}(\mathbf{r}^N), \tag{26}$$

where $U_s(\mathbf{q})$ is the solute intramolecular potential energy, $U_{sw}(\mathbf{q}, \mathbf{r}^N)$ is the solute–water interaction energy, and $U_{ww}(\mathbf{r}^N)$ is the water–water interaction energy. Substituting (26) into (25) and using the linearity of integration, the total energy decomposes exactly as

$$E = E_s + E_{sw} + E_{ww}, \tag{27}$$

where each term is an exact ensemble average:

$$E_s = \int P(\mathbf{q}, \mathbf{r}^N) \, U_s(\mathbf{q}) \, d\mathbf{q} \, d\mathbf{r}^N, \tag{28}$$

$$E_{sw} = \int P(\mathbf{q}, \mathbf{r}^N) \, U_{sw}(\mathbf{q}, \mathbf{r}^N) \, d\mathbf{q} \, d\mathbf{r}^N, \tag{29}$$

$$E_{ww} = \int P(\mathbf{q}, \mathbf{r}^N) \, U_{ww}(\mathbf{r}^N) \, d\mathbf{q} \, d\mathbf{r}^N. \tag{30}$$

We note that GIST estimates these quantities directly from sampled molecular dynamics trajectories.

Partitioning the configuration space into the same k clusters used for the entropy decomposition, the total energy in (25) can be written without approximation as

$$E = \sum_{i=1}^{k} \int_i P(\mathbf{q}, \mathbf{r}^N)\, U(\mathbf{q}, \mathbf{r}^N)\; d\mathbf{q}\, d\mathbf{r}^N. \tag{31}$$

Defining the probability-weighted mean energy of cluster i as

$$\langle E \rangle_i = \frac{1}{P_i} \int_i P(\mathbf{q}, \mathbf{r}^N)\, U(\mathbf{q}, \mathbf{r}^N)\; d\mathbf{q}\, d\mathbf{r}^N, \tag{32}$$

equation (31) becomes the exact relation

$$\boxed{E = \sum_{i=1}^{k} P_i \, \langle E \rangle_i.} \tag{33}$$

Applying the same energy decomposition (26) within each cluster, each $\langle E \rangle_i$ separates exactly into

$$\langle E \rangle_i = \langle E_s \rangle_i + \langle E_{sw} \rangle_i + \langle E_{ww} \rangle_i, \tag{34}$$

where $\langle E_s \rangle_i$, $\langle E_{sw} \rangle_i$, and $\langle E_{ww} \rangle_i$ are the cluster-averaged solute, solute–water, and water–water energies, respectively. Crucially, these three quantities can be evaluated directly from simulation trajectories restricted to cluster i without invoking any approximation: they are simply conditional averages of interaction energies over the subset of frames belonging to cluster i. The sum over clusters is therefore exact,

$$E_s = \sum_{i=1}^{k} P_i \, \langle E_s \rangle_i, \qquad E_{sw} = \sum_{i=1}^{k} P_i \, \langle E_{sw} \rangle_i, \qquad E_{ww} = \sum_{i=1}^{k} P_i \, \langle E_{ww} \rangle_i, \tag{35}$$

and the total system energy is recovered without error:

$$E = \sum_{i=1}^{k} P_i \, (\langle E_s \rangle_i + \langle E_{sw} \rangle_i + \langle E_{ww} \rangle_i) \tag{36}$$

Combining with the entropy result (21), the total Helmholtz free energy of the solvated system is therefore approximated as

$$\boxed{A \;\approx\; \sum_{i=1}^{k} P_i \, [\langle E_{solute} \rangle_i + \langle E_{sw} \rangle_i + \langle E_{ww} \rangle_i - T S_{i,solute} - T S_{i,solvent}]} \tag{37}$$

where the only approximation entering (37) is the cluster independence of the solvent distribution assumed in (14); the energy terms are exact.

We define the Helmholtz free energy contribution of cluster i as

$$A_i = \langle E_{solute} \rangle_i + \langle E_{sw} \rangle_i + \langle E_{ww} \rangle_i - TS_{i,solute} - TS_{i,solvent}, \tag{38}$$

so that (37) becomes simply

$$A \approx \sum_{i=1}^{k} P_i \, A_i \tag{39}$$

The five terms in (38) are the solute intramolecular energy, the solute–water interaction energy, the water–water interaction energy, the solute entropic contribution, and the solvation entropic contribution of cluster i, respectively. The first three terms are evaluated exactly as conditional averages over frames belonging to cluster i, while $S_{i,solute}$ is estimated from the intra-cluster coordinate distribution and $S_{i,solvent}$ is obtained from GIST analysis of a simulation restrained to the centroid of cluster i.

3.3 Thermodynamic Decomposition and Mapping of Flexible Solutes with GIST

It is important to emphasize that solvation thermodynamic mapping is only rigorously connected to the solvation entropy when performed for a rigid solute. If the solute, whether a protein or a small-molecule ligand, is allowed to move during the simulation, the resulting solvent density distributions become smoothed over the ensemble of solute configurations, leading to artificially broadened density distributions and systematic overestimates of the entropy. In the limiting case, if the solute is free to translate and rotate, the solvent single-body density distributions $\rho(\mathbf{R}, \boldsymbol{\omega})$ will approach uniform translational and orientational distributions, yielding bulk-like entropy values that are devoid of information about solute–solvent ordering. Solvation thermodynamic mapping is therefore only rigorously connected to the solvent entropy $S_{i,solvent}$ for a fixed solute conformation.

The cluster decomposition introduced above resolves this limitation. By restraining the solute to the centroid of each cluster i during the corresponding GIST simulation, the solvent density distributions are sampled in the presence of an effectively rigid solute conformation, preserving the rigor of the solvation entropy estimate. This procedure yields a spatially resolved solvation map, i.e. a grid of solvent thermodynamic properties as produced by GIST, that is uniquely associated with each cluster i. The full solvation thermodynamics of the flexible solute is then recovered by combining these per-cluster maps with population weights P_i as in (37), providing both a free energy estimate and a conformation-dependent spatial decomposition of solvation thermodynamics.

4 Methods

4.1 Validation Dataset

A 20-molecule validation set was selected from a previously reported test set of 504 compounds [14]. Molecules were retained only if the range of solvation free energies across their

conformations exceeded 2.0 kcal/mol, ensuring a meaningful conformational dependence of the solvation free energy for each molecule investigated. The remaining molecules were stratified by reported flexible solvation free energy [14] using quantile binning into 20 bins. From each bin, one representative molecule was selected to ensure chemical diversity. Diversity was assessed using Morgan fingerprints (radius 2, 2048 bits, chirality included) and Tanimoto distance [15, 16]. Selection was performed using a MaxMin procedure to maximize dissimilarity among the chosen molecules [17]. The selected molecules and their experimental solvation free energies are listed in Table X. After clustering in solution (described below), all solutes with more than 4 clusters were excluded from further analysis, resulting in a final validation set of 15 molecules.

4.2 MD Simulations

The three-dimensional structures of the 20 small molecules were obtained from the previously published dataset [14]. Atomic partial charges were assigned using the AM1-BCC model [18] as implemented in Antechamber [19]. Solute forcefield parameters were assigned using GAFF2 [20], and systems were constructed using tLEaP [21]. Each solute was solvated in a rectangular box of TIP3P water molecules [22] such that the nearest periodic boundary is at least 15 Å away. All systems were subjected to energy minimization and equilibration following a previously described ten-step protocol [23]. Briefly, solvent molecules were minimized using steepest descent, after which solute atoms were gradually relaxed and equilibrated in stages. For rigid simulations, solute atoms were held in place with harmonic positional restraints of 100 kcal $\cdot$ mol^{-1}Å^{-2}. In all simulations, covalent bonds involving hydrogen were constrained using the SHAKE algorithm. Equations of motion were integrated using Langevin dynamics, with a collision frequency (γ) of 2 ps^{-1}. A Berendsen barostat with isotropic position scaling was used with a target pressure of 1 bar and a pressure relaxation time (τ) of 2 ps. All final equilibration and production simulations were performed under NPT conditions at 300 K using a 2 fs integration timestep, with system configurations saved every 1000 time steps. All simulations were equilibrated for 250 ps. Unrestrained production simulations were carried out for 100 ns and used as the basis for solute conformational clustering. Additional simulations used for thermodynamic integration and GIST calculations are described in the corresponding sections. Detailed molecular dynamics input files are provided in the Supplemental Information.

4.3 Thermodynamic Integration

Thermodynamic integration (TI) was used to calculate solvation free energies by alchemically decoupling solute–solvent interactions. Simulations were performed at 15 intermediate λ states (0.0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 1.0) as in our previous study[9], where λ scales the strength of the decoupling of solute–solvent interactions. Both Lennard–Jones and electrostatic interactions between the solute and solvent were decreased simultaneously as a function of λ. A soft-core potential was employed to avoid endpoint singularities, using the default AMBER parameters $\alpha_{LJ} = 0.5$ and $\beta_C = 12$. For solvation free energy calculations for flexible solutes, TI equilibration and production simulations were performed without restraints. To estimate solvation free energies of specific solute

conformations, TI simulations were performed with harmonic positional restraints applied to all solute atoms. Restraints with a force constant of 100 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2} were applied relative to the Cartesian coordinates of each cluster's centroid, using a Euclidean distance metric. The centroid was defined as the configuration with the smallest distance to all other configurations within a cluster. At intermediate λ states, simulations consisted of a 250 ps equilibration phase followed by a 5 ns production phase. At end states (λ = 0.0 and 1.0), simulations were run for 250 ps of equilibration and 100 ns of production, in order to facilitate GIST analysis on the same trajectories as TI. Free energy differences were computed by numerical integration of $\langle \partial H/\partial \lambda \rangle$ over λ using the trapezoid rule as implemented in the `alchemlyb` toolkit [24]. Uncertainties were estimated by propagating the standard errors of the λ-window mean $\langle \partial H/\partial \lambda \rangle$ values through the trapezoid formula.

4.4 GIST Calculations

GIST calculations were performed using the `gist` command implemented in `cpptraj` [25]. Analyses were carried out on 100 ns production trajectories, with atomic coordinates saved every 2 ps, yielding 50,000 frames per trajectory. A grid spacing of 0.5 Å was used, corresponding to a voxel volume of 0.125 Å^3. The grid was centered on the solute center of mass and extended 100 voxels, or 50 Å, in each Cartesian direction. This grid size and voxel number was sufficient to contain the entire periodic box for all systems.

Electrostatic energies were computed using the Particle Mesh Ewald method [26], and a long-range Lennard-Jones correction was applied. A cutoff of 9 Å was used for both short-range electrostatic and Lennard-Jones interactions. For each solute conformation analyzed, GIST calculations were performed on three independent replicate MD simulations, and reported values represent averages over these replicates. A reference bulk water number density of ρ_{ref} = 0.03288 mol/Å^3 was used for all density-based calculations.

4.4.1 GIST Integration

The output of GIST analysis is `.dx` files for each of the thermodynamic quantities analyzed. The total value of any quantity for the system can be calculated by summing the values for all the voxels. The value for the quantity in a subset of voxels can be calculated by summing the desired voxels, for example those within 10 Å of the solute's center of mass. Integration of the GIST output was performed using the command line utility `gistpp` [27]. Because inconsistent voxel occupancy at the edge of the solvent box can introduce local inaccuracies in the nearest-neighbor–based solvent entropy estimator, coordinates within 1 Å of the solvent box edges were excluded from analysis. Solvent properties in the excluded regions are effectively bulk-like, and their true contribution to the solvation entropy is therefore negligible.

The entropy for a neat fluid should be zero, but systemic bias in the nearest neighbor algorithm yields a nonzero value. To account for this, we run GIST on the decoupled (bulk, $\lambda = 1.0$) simulations to calculate a per-water entropy correction term S_{ref}:

$$S_{ref} = \sum_{voxels} \frac{S_{dens} \cdot V_{voxel}}{N_{wat}}$$

(40)

where V_{voxel} is the voxel volume, S_{dens} is the solvation entropy density estimated by GIST, and N_{wat} is the total number of water molecules on the grid. Similarly, all thermodynamic quantities calculated by GIST can be integrated in this way, as previously [9]:

$$E = \sum_{voxels} \frac{E_{dens} \cdot V_{voxel}}{N_{wat}} \tag{41}$$

Alternatively, the reference values can be formulated as the initial (non-interacting) state of the solvation process:

$$A_{initial} = E_{total} + U_{solute} - TS_{solvent} - TS_{solute}^{config} \tag{42}$$

where E_{total} is the sum of the water-water (E_{ww}) and solute-water interaction energy (E_{sw}) calculated by GIST, U_{solute} is the internal energy of the solute calculated from the trajectory using `cpptraj`, T is temperature, $S_{solvent}$ is the entropy of water as calculated by GIST, and S_{solute}^{config} is the solute's configurational entropy, discussed in 4.6. In the "gas" (non-interacting) phase, E_{sw} is zero. For the solute in water:

$$A_{final} = E_{total} + U_{solute}^{bonded} - TS_{solvent} - TS_{solute}^{config} \tag{43}$$

U_{solute}^{bonded} is used here because unlike in the gas-phase calculation, the solute's non-bonded interactions are included in E_{sw}. Thus we can calculate the solvation free energy change:

$$\Delta A_{solv} = A_{final} - A_{initial} \tag{44}$$

where A_{final} is calculated at $\lambda = 0.0$, analogous to the liquid phase, and $A_{initial}$ is calculated at $\lambda = 1.0$, equivalent to the gas phase. By utilizing lambda scaling of the solute-solvent interactions for GIST end-states calculations, MD trajectories can be shared between GIST and alchemical free energy analysis for validation, removing sources of variance for the sake of validation. This also obviates the need to run a separate *in vacuo* simulation of the solute to obtain the solute's gas-phase energy and entropy.

4.5 Clustering

Unrestrained molecular dynamics simulations were performed for each solute to sample configurational space, as described above. Solute configurations were clustered using k-means with N-Ary Natural Initialization (NANI) [28], based on aligned Cartesian coordinates. For each molecule, the number of clusters (k) was selected automatically to balance inter-cluster separation and intra-cluster compactness using the Davies–Bouldin criterion [29]. Distances were calculated using a Euclidean metric.

4.6 Solute Entropy Estimation

We define the solute entropy contributions as follows:

$$S_{config} = S_{conform} + S_{vib} \tag{45}$$

where S_{config} is the solute's configurational entropy, $S_{conform}$ is the solute's conformational entropy, and S_{vib} is the solute's vibrational entropy.

We estimated solute vibrational entropy using a nearest-neighbor method [30, 31] applied to internal coordinates. Correlations between coordinates were accounted for using a mutual information expansion truncated after the second-order term [32]. For each trajectory frame, a vector of solute internal coordinates x, consisting of bond lengths, bond angles, and torsion angles, was constructed. Periodicity of torsional coordinates was not explicitly accounted for; distances were computed using a Euclidean metric in the internal coordinate space. Torsional distributions remained localized within conformational basins, limiting artifacts associated with angular discontinuities.

4.6.1 Nearest-Neighbor Entropy Estimator

For a set of N samples $\{x_i\}_{i=1}^{N}$ in a given coordinate space, the entropy estimator used in this work is given by

$$S = k_B \left(\frac{1}{N} \sum_{i=1}^{N} \ln r_i + C \right), \tag{46}$$

where r_i is the nearest-neighbor distance for sample i, and C is an additive constant that depends on the dimensionality of the space and the sample size N.

For one-dimensional entropy estimates, the constant is

$$C = \ln(2N) + \gamma, \tag{47}$$

where γ is Euler's constant. For two-dimensional entropy estimates, used in calculating the second-order mutual information term, the constant is

$$C = \ln(N\pi) + \gamma. \tag{48}$$

4.6.2 Mutual Information Expansion

To account for correlations between internal coordinates, vibrational entropy was approximated using a truncated mutual information expansion,

$$S_{\text{vib}} \approx \sum_i S_i - \sum_{i<j} I_{ij}, \tag{49}$$

where S_i is the marginal entropy of coordinate i, and I_{ij} is the pairwise mutual information between coordinates i and j.

Pairwise mutual information terms were computed as

$$I_{ij} = S_i + S_j - S_{ij}, \tag{50}$$

where S_{ij} is the joint entropy of the coordinate pair (i, j), estimated using the nearest-neighbor estimator in the corresponding joint coordinate space.

All marginal and joint entropy terms were computed using the nearest-neighbor estimator in the corresponding one- and two-dimensional coordinate spaces. The expansion was truncated after the pairwise (second-order) term; higher-order correlations were not included.

4.6.3 Conformational Entropy

Conformational entropy of each solute was estimated from the populations of conformational clusters obtained from trajectory analysis. The conformational entropy was computed as

$$S_{\text{conform}} = -k_B \sum_{i=1}^{k} P(i) \ln P(i), \tag{51}$$

where $P(i)$ is the normalized population of cluster i (eq. 10), and k is the solute's total number of clusters.

$S_{conform}$ was calculated using the percentage of configurations assigned to each conformation by clustering a fully flexible 100 ns MD simulation (as described above), either at $\lambda = 0.0$ for the solvated molecule or $\lambda = 1.0$ for the gas phase molecule, and $\Delta S_{conform}$ was calculated as the difference of the two.

4.7 Flexi-GIST Free Energy Calculation

Solvation free energies for flexible solutes were computed by combining cluster-resolved free energy contributions from representative conformations. Representative conformations were defined as the centroids obtained from clustering an unrestrained MD trajectory. For each cluster i, solvation thermodynamic quantities were evaluated for the centroid using GIST and combined using population-weighted averaging with weights $P(i)$. All energy and entropy terms were defined relative to the bulk solvent reference state as described above.

The Flexi-GIST solvation free energy was computed as

$$\Delta A^{flex}_{solv} \approx \sum_{i=1}^{k} P(q_i) \Delta E^{i}_{solv} - T\Delta S^{config}_{solute} - T \sum_{i=1}^{k} P(q_i) \Delta S^{i}_{solvent} \tag{52}$$

where k is the number of clusters, $P(q_i)$ is the normalized population of cluster i, ΔE^{i}_{solv} is the solvation energy of cluster i, $\Delta S^{config}_{solute}$ is the change in solute configurational entropy upon solvation, and $\Delta S^{i}_{solvent}$ is the solvation entropy of cluster i. The energy decomposition is exact (see Section 3.2), while the solvent entropy requires the cluster-independence approximation.

$$\Delta A^{flex}_{solv} \approx \Delta E_{solv} - T\Delta S^{config}_{solute} - T \sum_{i=1}^{k} P(q_i) \Delta S^{i}_{solvent} \tag{53}$$

where k is the number of clusters and $P(i)$ is the normalized population of cluster i.

In the limit of a single conformational cluster ($k = 1$), this expression reduces to the existing rigid approximation for GIST solvation free energy [9], as in eq. 8. In the limit of

complete sampling over conformational space ($\lim_{k\to\infty}$) the expression reduces to the exact flexible solvation free energy [33], as in eq. 1.

5 Results

We validate Flexi-GIST through a staged series of comparisons designed to isolate and assess each approximation introduced by the method. We first test whether the solvation free energy of a flexible molecule can be recovered as a probability-weighted sum over the solvation free energies of its rigid conformational states, using thermodynamic integration as a reference for both the rigid and flexible calculations. This isolates the clustering and population-weighting approximation from any errors associated with GIST itself. We then verify that GIST accurately reproduces TI solvation free energies for individual rigid conformations, establishing that the end-state free energy estimator introduces no significant additional error. Finally, we combine both approximations and evaluate the full Flexi-GIST procedure, in which cluster-resolved GIST thermodynamic quantities are used in place of TI, against fully flexible TI reference values. This progression allows any discrepancies observed in the final comparison to be attributed to specific stages of the methodology.

5.1 Flexible Solvation Free Energies as Probability-Weighted Sums over Rigid Conformations

To assess whether solvation free energies of flexible molecules can be reconstructed from rigid conformational states, we approximate the free energy of the flexible system as a weighted sum over the free energy of solvation of each rigid cluster centroid:

$$\Delta A_{\text{solv}}^{\text{flex}} \approx \sum_{i=1}^{k} P(q_i)\, \Delta A_{\text{solv}}^{i} \tag{54}$$

where k is the number of clusters, $P(q_i)$ is the probability of the solute occupying cluster i, and $\Delta A_{\text{solv}}^{i}$ is the solvation free energy of the rigid centroid of cluster i obtained by thermodynamic integration. This expression approximates the flexible solvation free energy as a probability-weighted average of the solvation free energies computed independently for each rigid cluster centroid.

The results are shown in (Fig.1). For each system, an unrestrained MD trajectory was clustered to obtain representative conformations, and ΔA_{solv} was estimated for each cluster's centroid using three independent TI replicates; fully flexible ΔA_{solv} values were computed using the same replicate protocol.

Across the test set, centroid-specific solvation free energies often exhibited variation within a given solute, although some showed little conformational dependence. These results indicate that even for relatively small, drug-like compounds, ΔA_{solv} can exhibit conformational dependence. In most cases, the fully flexible TI value was not reproduced by any single centroid alone. This behavior is consistent with the expectation that a molecule's free energy reflects an ensemble average over conformational states rather than a property

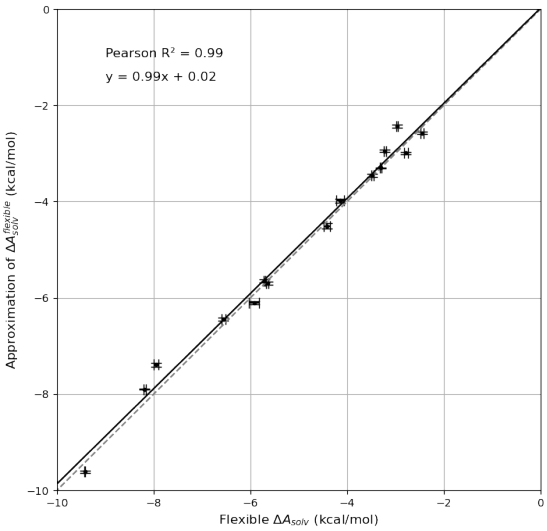

Figure 1: Approximate treatment of conformation-dependent solvation thermodynamics validated using thermodynamic integration (TI). In order to validate the approximate treatment of solute flexibility underlying Flexi-GIST, we first applied the approximation to ΔA_{solv} values estimated by TI for rigid solute conformations. Conformation-specific solvation free energies were weighted and combined using eq. 54 and compared to flexible free energies, also estimated by TI. Data points represent the mean of three independent replicates, and error bars are ± the standard error of the mean (SEM). Mean absolute error (MAE) = 0.19 kcal/mol.

of any individual structure. Even when individual centroids yield similar total ΔA_{solv} values, the underlying thermodynamic contributions and solvent structure can differ between conformations.

A population-weighted average over centroid TI values provided a reasonable approximation to the fully flexible result and, in many cases, improved upon single-conformation estimates. These results confirm that conformational populations contribute to solvation thermodynamics and support an ensemble-based treatment.

5.2 GIST Matches TI for Rigid Solutes

To evaluate the accuracy of GIST for rigid solute conformations, we compared solvation free energies computed using GIST to those estimated with TI for the same centroid structures (Fig. 2). For each representative conformation, GIST-derived ΔA_{solv} values were obtained from end-state analysis and compared directly to the corresponding TI estimates.

Across the full set of centroid structures, GIST and TI showed strong agreement, with a mean absolute error (MAE) of 0.35 kcal/mol and a Pearson R-squared correlation coefficient of $R^2 = 0.92$. Deviations between the two methods were comparable to the expected uncertainty from finite sampling and numerical integration. No systematic bias was observed

across the range of solvation free energies considered.

These results demonstrate that GIST accurately reproduces solvation free energies for rigid solute conformations, consistent with prior work [9]. Therefore, any discrepancies observed for flexible molecules are unlikely to arise from inaccuracies in the underlying GIST calculation, but instead reflect the approximate treatment of conformational heterogeneity. In order to validate the approximate treatment of solute flexibility underlying Flexi-GIST,

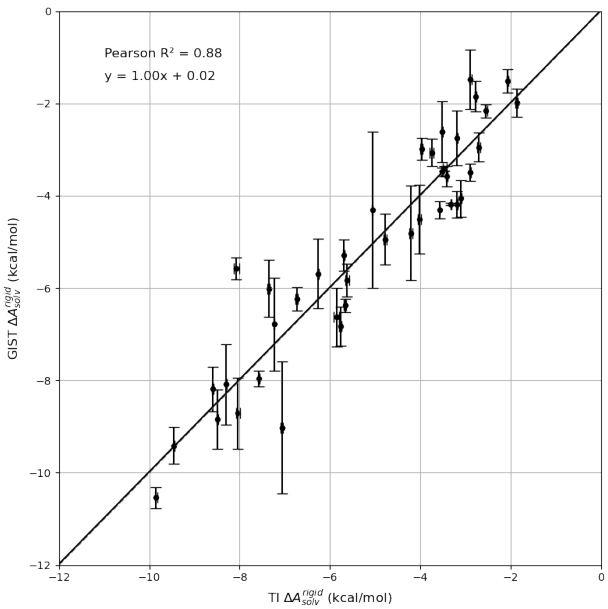

Figure 2: GIST solvation free energies agree with TI for rigid conformations. In order to validate GIST for solvation free energy calculations, we applied GIST and TI to rigid conformations from the test set. Conformations per solute ranged from 2-4, and were defined as the centroid of each cluster. Eq. 44 was used to sum all the free energy contributions implicitly included by TI to ensure agreement. The solid black line is the line of best fit using linear regression, the points are the means of three independent replicates, and the error bars are ±SEM. MAE = 0.66 kcal/mol

we first applied the approximation to ΔA_{solv} values estimated by TI for rigid solute conformations. Conformation-specific solvation free energies were weighted and combined using eq. 54 and compared to flexible free energies, also estimated by TI. Data points represent the mean of three independent replicates, and error bars are ± the standard error of the mean (SEM). Mean absolute error (MAE) = 0.19 kcal/mol.

5.3 Flexi-GIST Approximates Flexible Solvation Free Energy

We next evaluated how well the Flexi-GIST approach approximates the solvation free energies of flexible molecules. For each solute, cluster-resolved GIST thermodynamic quantities were computed for representative conformations and combined using the population-weighted procedure described in eq. 53.

Flexi-GIST ΔA_{solv} values showed good agreement with fully flexible TI estimates across the test set (Fig.3), with a mean absolute error of 0.76 kcal/mol. This represents an increase relative to the error observed for rigid solute comparisons (0.39 kcal/mol), reflecting the additional approximations associated with clustering and configurational entropy estimation. Nonetheless, the overall level of agreement remains within ∼1 kcal/mol, indicating that the cluster-based approach captures the dominant contributions of conformational heterogeneity to solvation thermodynamics.

Relative to single-conformation approximations, Flexi-GIST provides a more accurate estimate of the flexible-molecule free energy by incorporating ΔA_{solv} contributions from alternative conformations of the solute. Residual deviations likely arise from finite sampling, clustering approximations, and limitations of the entropy estimation procedure. These results demonstrate that Flexi-GIST provides a practical approach for approximating flexible solvation free energies while preserving the conformation-dependent spatial decomposition of solvent thermodynamics.

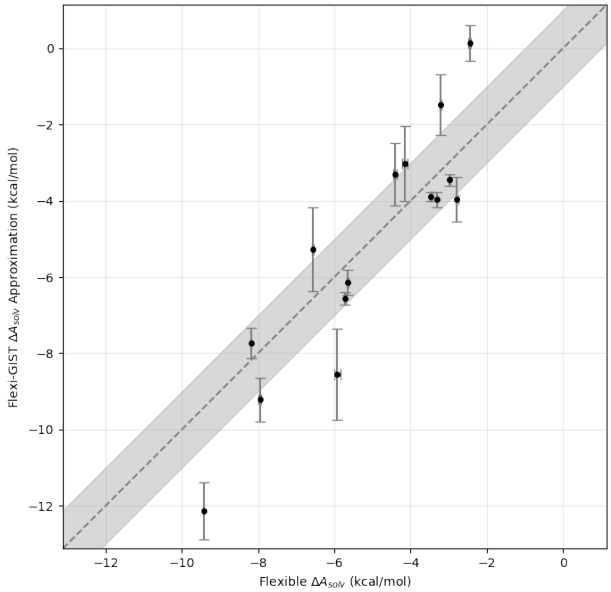

Figure 3: Flexi-GIST captures conformation-dependent solvation effects. Solvation free energy estimates for rigid solute conformations combined by Flexi-GIST (eq. 53) show agreement with TI estimates for flexible molecules. Dashed line represents exact agreement between Flexi-GIST and flexible TI, shaded area represents ± 1 kcal/mol within exact agreement. Data points represent the mean of three independent replicates, and error bars are ±SEM. MAE = 1.266 kcal/mol.

6 Discussion

In this work, we introduced Flexi-GIST, a cluster-based extension of grid inhomogeneous solvation theory that enables an approximate treatment of solvation mapping for flexible

molecules. By combining GIST calculations performed on representative conformations with a population-weighted ensemble framework and an explicit treatment of solute configurational entropy, Flexi-GIST accurately reproduces solvation thermodynamics of flexible molecules.

Flexi-GIST predictions showed good agreement with fully flexible thermodynamic integration (TI) results, with a mean absolute error (MAE) of 0.76 kcal/mol. This represents a slight increase relative to the error observed for rigid solute comparisons (0.66 kcal/mol, fig. 2), reflecting the additional approximations introduced by clustering. These include finite sampling of conformational space and discretization of the ensemble into a limited number of representative states. Despite these sources of error, the overall level of agreement largely remains within $\sim$1 kcal/mol, indicating that the dominant contributions of conformational heterogeneity to solvation free energy are captured by Flexi-GIST.

Importantly, the objective of Flexi-GIST is not to outperform alchemical free energy methods in terms of absolute accuracy, but to apply solvation mapping to flexible systems in a manner that is consistent with the underlying statistical mechanical theory. In this context, Flexi-GIST provides a means of approximating the formally exact ensemble average over solute conformations while preserving the spatial decomposition of solvent thermodynamics that is central to GIST. This distinguishes it from alchemical methods, which yield accurate free energy differences but do not provide direct access to spatially resolved solvent contributions. A further strength of the approach is its ability to decompose the solvation free energy into physically meaningful component terms, including the solute entropy, solute energy, solute–water interaction energy, water–water interaction energy, and solvent entropy. When applied to molecular recognition, such a decomposition provides direct insight into which thermodynamic contributions favor or oppose binding affinity, thereby offering actionable guidance for the rational design of molecular modifications.

6.1 Limitations

Several limitations of the present approach should be noted. First, the accuracy of using representative rigid conformations chosen by clustering to approximate the flexible system depends on the quality of the underlying conformational ensemble and the clustering procedure used to define representative states. Highly disordered systems will resist meaningful clustering, and even for well-ordered systems the ability to exploit an alternative conformation is dependent both on sampling it and identifying it in clustering. Second, the configurational entropy term is approximated using a nearest-neighbor estimator with a truncated mutual information expansion, which may neglect higher-order correlations. Third, Flexi-GIST inherits the sampling requirements of both GIST and the nearest-neighbor entropy estimator, which can lead to noisy free energy values if not sufficiently converged. A summary of the error in each decomposed term can be found in supplemental information Table S1. Finally, the discretization of the conformational ensemble may introduce bias if important states are undersampled or omitted.

6.2 Applications

Flexi-GIST is expected to be most useful in applications where conformational flexibility plays a significant role in modulating solvation structure, particularly in structure-based drug design. In these contexts, the ability to relate changes in solvation thermodynamics to specific conformational states may provide insight that is not accessible from single-structure analyses or purely alchemical approaches. Future work will focus on applying Flexi-GIST to protein systems, including for bindability assessment, the incorporation of solvation thermodynamics into docking workflows, and lead optimization, where solvent-mediated effects are critical determinants of ligand binding.

An illustrative example of the value of decomposing the free energy into physically distinct contributions can be seen by comparison to the work of Fischer et al. [34], who showed that ensemble docking improves when each receptor conformation is assigned a Boltzmann-weighted energy penalty derived from its crystallographic occupancy. That penalty enters the docking score as a single term, effectively treating the conformational free energy as an undifferentiated quantity. Our decomposition reveals that the individual contributions to the conformational free energy have distinct influences upon ligand binding. The cavity solvation free energy is recovered when binding-site water is displaced into bulk, making conformations with high (unfavorable) cavity solvation more attractive targets because the energetic cost of maintaining that ordered water is returned as a favorable contribution to binding. The protein conformational entropy, by contrast, is lost: upon binding, the side chains in the binding site become restricted as they adopt positions complementary to the ligand, so conformations with low binding-site entropy suffer less entropy loss upon complexation. Similarly, the protein energetic strain is effectively locked into place as the protein rigidifies around the bound ligand, so conformations with low strain pay a smaller energetic penalty for maintaining the bound state. The overall free energy reflects the balance of these competing contributions. Conformations should therefore be prioritized not by total free energy alone but by the total free energy minus the binding-cavity solvation cost, further favoring low protein entropy and low protein strain in the binding site. Such a breakdownof solvation thermodynamic contributions across multiple conformations was not practical at the time of [34], but is enabled by the Flexi-GIST decomposition presented here.

References

[1] Tom Young, Robert Abel, Byungchan Kim, Bruce J. Berne, and Richard A. Friesner. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proceedings of the National Academy of Sciences*, 104(3):808–813, January 2007.

[2] Kamran Haider, Anthony Cruz, Steven Ramsey, Michael K. Gilson, and Tom Kurtzman. Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis of Water in Molecular Dynamics Trajectories. *Journal of Chemical Theory and Computation*, 14(1):418–425, 2018.

[3] Robert A. Pearlstein, Woody Sherman, and Robert Abel. Contributions of water transfer energy to protein-ligand association and dissociation barriers: Watermap analysis of a series of p38 MAP kinase inhibitors. *Proteins*, 81(9):1509–1526, September 2013.

[4] Daniel Cappel, Woody Sherman, and Thijs Beuming. Calculating Water Thermodynamics in the Binding Site of Proteins - Applications of WaterMap to Drug Discovery. *Current Topics in Medicinal Chemistry*, 17(23):2586–2598, 2017.

[5] Christopher Higgs, Thijs Beuming, and Woody Sherman. Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor. *ACS Medicinal Chemistry Letters*, 1(4):160–164, July 2010.

[6] Yeonji Ji, Vjay Molino, Steven Ramsey, and Tom Kurtzman. Solvation Thermodynamic Costs of Cognate Binding Site Formation. *ChemRxiv*, 2025.

[7] Valentin J. Egger-Hoerschinger, Franz Waibl, Vjay Molino, Helmut Carter, Monica Fernández-Quintero, Steven Ramsey, Daniel R. Roe, Klaus R. Liedl, Michael K. Gilson, and Tom Kurtzman. Quantifying Spatially Resolved Hydration Thermodynamics Using Grid Inhomogeneous Solvation Theory [v1.0]. *Living Journal of Computational Molecular Science*, 6(1):3059, August 2025.

[8] Jerome K. Percus and George J. Yevick. Analysis of Classical Statistical Mechanics by Means of Collective Coordinates. *Physical Review*, 110(1):1–13, 1958.

[9] Lieyang Chen, Anthony Cruz, Daniel R. Roe, Andrew C. Simmonett, Lauren Wickstrom, Nanjie Deng, and Tom Kurtzman. Thermodynamic Decomposition of Solvation Free Energies with Particle Mesh Ewald and Long-Range Lennard-Jones Interactions in Grid Inhomogeneous Solvation Theory. *Journal of Chemical Theory and Computation*, 17(5):2714–2724, 2021.

[10] Robert Abel, Tom Young, Ramy Farid, Bruce J. Berne, and Richard A. Friesner. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *Journal of the American Chemical Society*, 130(9):2817–2831, March 2008.

[11] Alexander S Bayden, Demetri T Moustakas, Diane Joseph-McCarthy, and Michelle L Lamb. Evaluating free energies of binding and conservation of crystallographic waters using SZMAP. *J. Chem. Inf. Model.*, 55(8):1552–1565, August 2015.

[12] Fumio Hirata. *Molecular Theory of Solvation*. Understanding Chemical Reactivity. Kluwer Academic, Tucson, AZ, December 2003.

[13] Tingjun Hou, Junmei Wang, Youyong Li, and Wei Wang. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.*, 51(1):69–82, January 2011.

[14] David L. Mobley, Ken A. Dill, and John D. Chodera. Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small Molecules. *The Journal of Physical Chemistry B*, 112(3):938–946, January 2008.

[15] David Rogers and Mathew Hahn. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, May 2010.

[16] Dávid Bajusz, Anita Rácz, and Károly Héberger. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*, 7(1):20, December 2015.

[17] Mark Ashton, John Barnard, Florence Casset, Michael Charlton, Geoffrey Downs, Dominique Gorse, John Holliday, Roger Lahana, and Peter Willett. Identification of Diverse Database Subsets using Property-Based and Fragment-Based Molecular Descriptions. *Quantitative Structure-Activity Relationships*, 21(6):598–604, December 2002.

[18] Araz Jakalian, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*, 23(16):1623–1641, 2002. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.10128.

[19] Junmei Wang, Wei Wang, Peter Kollman, and David Case. ANTECHAMBER: an accessory software package for molecular mechanical calculations. *Journal of Chemical Information and Computer Sciences - JCISD*, 222, November 2000.

[20] David A. Case, Hasan Metin Aktulga, Kellon Belfon, David S. Cerutti, G. Andrés Cisneros, Vinícius Wilian D. Cruzeiro, Negin Forouzesh, Timothy J. Giese, Andreas W. Götz, Holger Gohlke, Saeed Izadi, Koushik Kasavajhala, Mehmet C. Kaymak, Edward King, Tom Kurtzman, Tai-Sung Lee, Pengfei Li, Jian Liu, Tyler Luchko, Ray Luo, Madushanka Manathunga, Matias R. Machado, Hai Minh Nguyen, Kurt A. O'Hearn, Alexey V. Onufriev, Feng Pan, Sergio Pantano, Ruxi Qi, Ali Rahnamoun, Ali Risheh, Stephan Schott-Verdugo, Akhil Shajan, Jason Swails, Junmei Wang, Haixin Wei, Xiongwu Wu, Yongxian Wu, Shi Zhang, Shiji Zhao, Qiang Zhu, Thomas E. Cheatham III, Daniel R. Roe, Adrian Roitberg, Carlos Simmerling, Darrin M. York, Maria C. Nagan, and Kenneth M. Merz Jr. AmberTools. *Journal of Chemical Information and Modeling*, 63(20):6183–6191, October 2023.

[21] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, J. T. Berryman, S. R. Brozell, F. S. Carvahol, D. S. Cerutti, T. E. III Cheatham, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, N. Forouzesh, M. Ghazimirsaeed, G. Giambaşu, T. Giese, M. K. Gilson, H. Gohlke, A. W. Goetz, J. Harris, Z. Huang, S. Izadi, S. A. Izmailov, K. Kasavajhala, M. C. Kaymak, I. Kolossváry, A. Kovalenko, T. Kurtzman, T. S. Lee, P. Li, Z. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, J. Wang, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D. M. York, C. Zhao, Q. Zhu, and P. A. Kollman. Amber 2025, 2025.

[22] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983.

[23] Daniel R. Roe and Bernard R. Brooks. A protocol for preparing explicitly solvated systems for stable molecular dynamics simulations. *The Journal of Chemical Physics*, 153(5):054123, August 2020. _eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0013849/15577893/054123_1_online.pdf.

[24] Zhiyi Wu, David L. Dotson, Irfan Alibay, Bryce K. Allen, Mohammad Soroush Barhaghi, Jérôme Hénin, Thomas T. Joseph, Ian M. Kenney, Hyungro Lee, Haoxi Li, Victoria Lim, Shuai Liu, Domenico Marson, Pascal T. Merz, Alexander Schlaich, David Mobley, Michael R. Shirts, and Oliver Beckstein. alchemlyb: the simple alchemistry library. *Journal of Open Source Software*, 9(101):6934, September 2024.

[25] Daniel R. Roe and Thomas E. Cheatham. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, July 2013.

[26] Romelia Salomon-Ferrer, Andreas W. Götz, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation*, 9(9):3878–3888, September 2013.

[27] Steven Ramsey, Crystal Nguyen, Romelia Salomon-Ferrer, Ross C. Walker, Michael K. Gilson, and Tom Kurtzman. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *Journal of Computational Chemistry*, 37(21):2029–2037, August 2016.

[28] Lexin Chen, Daniel R. Roe, Matthew Kochert, Carlos Simmerling, and Ramón Alain Miranda-Quintana. K-Means NANI: An Improved Clustering Algorithm for Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*, 20(13):5583–5597, 2024.

[29] David L. Davies and Donald W. Bouldin. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, 1979.

[30] Crystal N. Nguyen, Tom Kurtzman, and Michael K. Gilson. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]Uril. *Journal of Chemical Physics*, 137(4):044101, 2012.

[31] L. F. Kozachenko and N. N. Leonenko. Sample Estimate of the Entropy of a Random Vector. *Problemy Peredachi Informatsii*, 23(2):9–16, 1987.

[32] Benjamin J. Killian, Joslyn Y. Kravitz, and Michael K. Gilson. Extraction of Configurational Entropy from Molecular Simulations via an Expansion Approximation. *Journal of Chemical Physics*, 127(2):024107, 2007.

[33] Themis Lazaridis. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *The Journal of Physical Chemistry B*, 102(18):3531–3541, April 1998.

[34] Marcus Fischer, Ryan G. Coleman, James S. Fraser, and Brian K. Shoichet. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nature Chemistry*, 6(7):575–583, July 2014.

Supplemental Information

Table S1: Decomposed thermodynamic contributions to the solvation free energy and the standard error of the mean (SEM) for each term, for all small molecules in the validation set. All values are in kcal/mol and represent averages over three independent replicates.

Solute	ΔE_{ww}	SEM	ΔE_{sw}	SEM	$T\Delta S^{\text{solute}}_{\text{conform}}$	SEM	$T\Delta S^{\text{solute}}_{\text{vib}}$	SEM	ΔE_{solute}	SEM	$T\Delta S_{\text{solvent}}$	SEM
2-fluorophenol	−44.928	0.101	−12.060	0.032	−0.269	0.187	0.279	0.958	1.785	0.039	−6.362	0.008
2-methoxy-1,1,1-trimethoxyethane	−50.485	0.196	−18.034	0.062	0.135	0.120	2.782	1.271	1.684	0.199	−10.529	0.033
2-methylpyrazine	−47.309	0.019	−13.209	0.006	0.000	0.000	0.713	0.202	0.028	0.008	−7.122	0.013
2-nitroaniline	−49.020	0.192	−17.205	0.008	0.137	0.064	−0.019	0.389	1.226	0.015	−7.809	0.023
3-methoxyphenol	−47.155	0.082	−15.935	0.009	0.025	0.002	−0.689	1.107	0.454	0.020	−7.961	0.015
bis(2-chloroethyl) ether	−47.029	0.278	−12.087	0.014	0.108	0.134	−2.310	0.680	1.112	0.012	−7.226	0.007
diethyl succinate	−50.845	0.605	−20.474	0.012	−0.237	0.080	1.708	0.840	0.448	0.029	−10.861	0.021
dimethylamine	−40.583	0.097	−8.611	0.003	0.093	0.134	−0.249	0.211	0.165	0.006	−5.087	0.005
methyl cyclohexanecarboxylate	−48.995	0.321	−14.255	0.001	−0.013	0.001	−1.137	1.060	0.064	0.002	−8.604	0.009
methyl formate	−40.356	0.009	−8.950	0.041	0.092	0.182	0.133	0.121	0.120	0.063	−4.196	0.002
piperazine	−47.565	0.096	−15.653	0.006	0.344	0.052	0.146	0.807	−0.176	0.114	−8.701	0.014
4-methylpentan-2-ol	−46.464	0.205	−12.537	0.014	−0.098	0.108	0.980	0.548	0.249	0.074	−8.192	0.008
N-methylmorpholine	−46.787	0.103	−12.932	0.003	−0.021	0.021	0.168	0.176	−0.010	0.022	−8.246	0.016
N-methyl-N-(2,2,2-trifluoroethyl)aniline	−51.773	0.370	−15.716	0.004	0.000	0.005	−0.283	0.885	0.418	0.012	−9.336	0.028
prop-2-en-1-ol	−41.189	0.099	−9.736	0.003	−0.139	0.219	0.579	0.416	0.551	0.010	−5.232	0.011