

Solvation Thermodynamic Mapping of Molecular Surfaces in AmberTools: GIST

Steven Ramsey,^[a,f] Crystal Nguyen,^[c] Romelia Salomon-Ferrer,^[d] Ross C. Walker,^[d,e] Michael K. Gilson,^[c] and Tom Kurtzman^[a,b,f]

The expulsion of water from surfaces upon molecular recognition and nonspecific association makes a major contribution to the free energy changes of these processes. In order to facilitate the characterization of water structure and thermodynamics on surfaces, we have incorporated Grid Inhomogeneous Solvation Theory (GIST) into the CPPTRAJ toolset of AmberTools. GIST is a grid-based implementation of Inhomogeneous Fluid Solvation Theory, which analyzes the output from molecular dynamics simulations to map out solvation thermodynamic and structural properties on a high-resolution,

three-dimensional grid. The CPPTRAJ implementation, called GIST-cpptraj, has a simple, easy-to-use command line interface, and is open source and freely distributed. We have also developed a set of open-source tools, called GISTPP, which facilitate the analysis of GIST output grids. Tutorials for both GIST-cpptraj and GISTPP can be found at ambermd.org. © 2016 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24417

Introduction

The work of displacing surface water to bulk, when a macromolecule binds another macromolecule or drug-like small molecule, can contribute strongly to the overall free energy of binding.^[1–13] As a consequence, the properties of water at molecular surfaces are of central interest in both molecular recognition and the nonspecific association of molecules in aqueous solution. Water can donate and accept hydrogen bonds with polar groups, while also making favorable van der Waals contacts with both polar and apolar parts of the surface. Such interactions tend to make it harder for a drug to displace water. On the other hand, the hydrophobic effect makes hydration of apolar surfaces net unfavorable, due, potentially, to a combination of entropic and enthalpic costs, and thus facilitates the displacement of water. It has also been argued that a map of water–protein interactions in a binding pocket creates a three-dimensional template, which strongly binding drugs replicate.^[10] Thus, the pattern of hydration of a binding pocket can provide useful information regarding key features that a drug molecule could mimic when binding its target.

Despite the intricacy of the protein–water interactions in typical binding pockets most structure-based computational drug–design methodologies have used simplified treatments of water, such as continuum models, which do not account for the granularity and geometric restrictions imposed by the finite size and specific hydrogen-bonding geometry of the water molecule.

In recent years the need for more detailed descriptions of water has been more widely acknowledged and this change is reflected in the growing set of computational tools that provide more detailed molecular descriptions of solvation. For example, amongst commercial software companies,

Schrodinger Inc. released WaterMap,^[14] Chemical Computing Group implemented 3D-RISM,^[15] Molecular Discovery implemented WaterFLAP,^[16,17] and OpenEye developed Szmap.^[18] Similarly, major pharmaceutical companies have integrated water structure and thermodynamic information into their drug discovery workflows, either by licensing commercial applications like the ones listed above, or by developing their own methodologies. Thus, Novartis developed WatMD,^[19,20] and GlaxoSmithKline developed SPAM.^[21] Further developments have emerged from academic laboratories, including the early WATsite method,^[22] as well as STOW,^[23] JAWS,^[24] WRAPPA,^[25] and WatClust.^[26]

[a] S. Ramsey, T. Kurtzman
Department of Chemistry, Lehman College, 205 W Bedford Pk Blvd, Bronx, New York 10468

[b] T. Kurtzman
Ph. D. Program in Chemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York, New York 10016

[c] C. Nguyen, M. K. Gilson
Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0736
E-mail: mgilson@ucsd.edu

[d] R. Salomon-Ferrer, R. C. Walker
San Diego Supercomputer Center, University of California San Diego, 9500 Gilman Drive MC0505, La Jolla, California 92093-0505

[e] R. C. Walker
Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0505
E-mail: ross@rosswalker.co.uk

[f] S. Ramsey, T. Kurtzman
Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York, New York 10016
E-mail: thomas.kurtzman@lehman.cuny.edu

Contract grant sponsor: National Institutes of Health; Contract grant number: GM100946

© 2016 Wiley Periodicals, Inc.

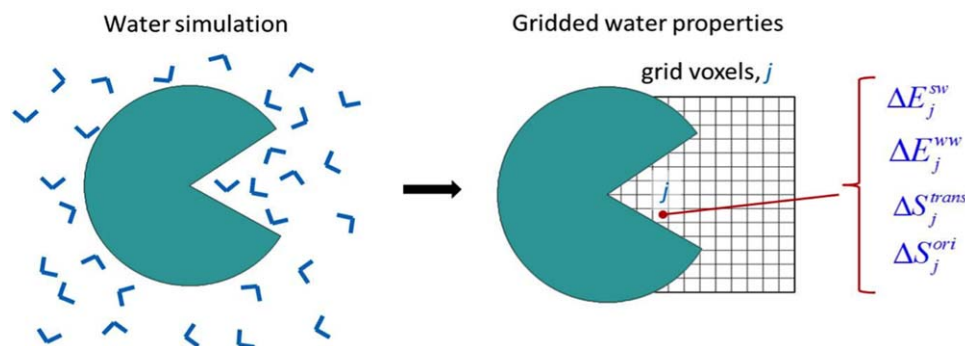


Figure 1. Diagram of a GIST calculation of water properties on a grid (right panel) in and near the binding pocket of a solute (green), based on a molecular dynamics simulation (left panel). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

While many of these important approaches have shown utility in the drug design setting they have significant limitations. Of particular concern for methods based on Inhomogeneous Fluid Solvation Theory (IST), such as WatClust, STOW, WATSite, and WaterMap, is that they are restricted to analysis of high-occupancy hydration sites and hence do not account for the significant fraction of binding-site water found in lower density regions. Among other limitations this restriction prevents a rigorous thermodynamic end-states analysis for comparison to free energy methods such as free energy perturbation (FEP) and thermodynamic integration (TI). In addition, many existing software packages for water analysis are closed-source and/or difficult to access and use.

Grid inhomogeneous solvation theory (GIST)^[27,28] discretizes the integrals of IST onto a three-dimensional grid that fills the binding pocket region, covering both high- and low-occupancy regions (Fig. 1). Thus, unlike other IST methods, it provides a smooth map of water structure and thermodynamics in the entire region of interest, rather than being restricted

to high-occupancy water sites. In order to facilitate broad application and further development of this technology, we have incorporated GIST into the freely distributable, open-source, CPPTRAJ package of AmberTools.^[27,29–31]

In this communication, we first discuss the algorithms utilized within GIST-cpptraj. Next we outline how to set up and run a GIST-cpptraj analysis of a molecular dynamics trajectory. Then we illustrate how to visualize the results, and how to manipulate the output files with the GISTPP postprocessing tool suite. Finally, we analyze the convergence of thermodynamic quantities calculated by GIST.

Methods

Overview of GIST-cpptraj and GISTPP

GIST-cpptraj is a straightforward, command-line tool, included within CPPTRAJ,^[29] which maps out the structural and thermodynamic quantities of water in a user-defined region of interest. This region is partitioned into a high-resolution, three-dimensional (3D) grid (Fig. 2), each grid box of which may be viewed as a 3D pixel, or voxel. The grid's desired size, position, and resolution (i.e., the size of the voxels) are specified via the command line, as detailed in Table 1. Thermodynamic quantities are calculated and mapped to each voxel via the grid based implementation of IST.^[27,28] In addition, a variety of structural measures are calculated for the water in each voxel. The thermodynamic and structural properties of water are listed in Table 2.

By default GIST outputs density weighted thermodynamic quantities from Table 2 in Data Explorer (.dx) file format,

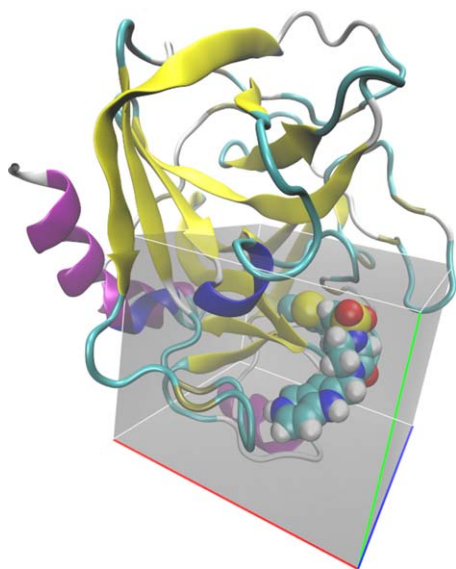


Figure 2. Co-crystal structure of Coagulation Factor Xa (cartoon) with ligand RRR (CPK) from pdb structure 1NFW.^[34] The region of interest corresponding to the GIST grid is shaded in gray. Individual voxels are not shown. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. GIST command options that define the region of interest.

Option	Description	Default value
gridcnr	Grid centroid position ^[a] <x> <y> <z>	0.0 0.0 0.0
griddim	Grid side length ^[b] <x> <y> <z>	40 40 40
gridspanc	Voxel side length ^[a]	0.5
refdens	Reference density ^[c]	0.0334

[a] Quantities in units of angstroms (Å). [b] Quantity in units of voxel count. [c] Density in molecules per Å³; a list of reference densities for common solvent models can be found in the Supporting Information.

Table 2. Quantities mapped by GIST on to a user-defined volume.

Quantity	Description	Units
TS_{trans}	Translational entropy density	[kcal/(mol Å ³)]
TS_{orient}	Oriental entropy density	[kcal/(mol Å ³)]
TS_{six}	Total entropy density	[kcal/(mol Å ³)]
E_{ww}	Water-water energy density	[kcal/(mol Å ³)]
E_{sw}	Solute-water energy density	[kcal/(mol Å ³)]
$E_{ij}^{\text{[a]}}$	Voxel pair interaction energy density	kcal/mol
Dipole ^[b]	Mean water dipole density	Debye/Å ³
Neighbor count	Mean number of water neighbors ^[c]	molecules
Order ^[a]	Average tetrahedral order parameter	None

[a] Optional outputs. [b] Dipole moments are reported as time-averaged x, y, and z, components, along with the mean magnitude. [c] Neighbor defined by O–O distance of less than 3.5 Å.

enabling visualization with standard graphics packages, such as VMD and PyMol.^[32,33] Thermodynamic quantities in per water units are not output in the dx file, but are accessible through the space-delimited text output that includes all calculated quantities. A complete list of quantities calculated by GIST can be found in the supplemental information. The GIST postprocessing suite of tools, GISTPP, further aids in the analysis of GIST outputs by supporting an array of mathematical and logical operations on dx files.

GIST algorithms

Calculation of hydration entropy terms. The first-order entropy calculated by GIST for a voxel k is:

$$S_{\text{sw}} = \frac{R\rho^\circ}{8\pi^2} \int_k g(\mathbf{r}, \boldsymbol{\omega}) \ln g(\mathbf{r}, \boldsymbol{\omega}) d\boldsymbol{\omega} d\mathbf{r} \quad (1)$$

where $g(\mathbf{r}, \boldsymbol{\omega})$ is the pair correlation function at position \mathbf{r} and water orientation $\boldsymbol{\omega}$, ρ° is the number density of bulk water, and R is the gas constant. For a voxel k , the spatial integral is over the volume of the voxel as indicated by the subscript on the integral symbol. In GIST-cpptraj, S_{sw} is evaluated using two different methods as described below.

Method 1. The first method relies on the approximation that the orientation is independent of the position within each voxel.^[27,35]

$$g_{\text{vox}}(\mathbf{r}, \boldsymbol{\omega}) \approx g_{\text{vox}}(\mathbf{r})g_{\text{vox}}(\boldsymbol{\omega}) \quad (2)$$

This allows the orientational and translational entropy to be evaluated separately within each voxel:

$$S_{\text{k}}^{\text{trans}} = -R\rho^\circ \int_k g_{\text{sw}}(\mathbf{r}) \ln g_{\text{sw}}(\mathbf{r}) d\mathbf{r} \quad (3)$$

$$S_{\text{k}}^{\text{orient}} = \frac{-R\rho^\circ}{8\pi^2} \int_k g_{\text{sw}}(\boldsymbol{\omega}) \ln g_{\text{sw}}(\boldsymbol{\omega}) d\boldsymbol{\omega} \quad (4)$$

The integrals are estimated using a nearest neighbors approach.^[36]

$$S_{\text{k}}^{\text{trans}} \approx \frac{-R}{N_{\text{k}}} \left(\sum_{i=1}^{N_{\text{k}}} \ln g_{\text{NN}}(\mathbf{r}) + \gamma \right) \quad (5)$$

$$S_{\text{k}}^{\text{orient}} \approx \frac{-R}{N_{\text{k}}} \left(\sum_{i=1}^{N_{\text{k}}} \ln g_{\text{NN}}(\boldsymbol{\omega}) + \gamma \right) \quad (6)$$

where N_{k} is the number of water molecules found in voxel k and γ is Euler's constant. The g_{NN} quantities are the nearest neighbor estimates for the respective g quantities:

$$g_{\text{NN},i}(\mathbf{r}) = \frac{1}{\rho^\circ V_{\text{NN},i}^{\text{trans}}} \quad (7)$$

and

$$g_{\text{NN},i}(\boldsymbol{\omega}) = \frac{8\pi^2}{V_{\text{NN},i}^{\text{orient}}} \quad (8)$$

Here, $V_{\text{NN},i}$ is the volume of a sphere whose radius is the distance from the current water molecule, i , to its nearest neighbor.

$$V_{\text{NN},i}^{\text{trans}} = \frac{4\pi d_{\text{trans}}^3}{3} \quad (9)$$

$$V_{\text{NN},i}^{\text{orient}} = \frac{4\pi d_{\text{orient}}^3}{3} \quad (10)$$

The distance in translational space is defined as the Euclidean distance between water oxygens, and the rotational distance is a quaternion distance.^[37]

$$d_{\text{trans}} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (11)$$

$$d_{\text{orient}} = 2 \arccos(q_i \cdot q_j) \quad (12)$$

The NN distances for Methods 1 and 2, below, for waters are calculated from a superset of configurations that includes water coordinates from all stored frames of the simulation. For rotational distances only waters found in the same voxel are used and for translational distances waters within the same and neighboring voxels are used. A voxel is considered to be a neighboring voxel if it shares any of the current voxels vertices, therefore each voxel has 26 neighbors and a voxel combined with all its neighbors form a $3 \times 3 \times 3$ cube of voxels.

Method 2. The second method uses an NN approach to directly estimate the six-dimensional integral in eq. (1).^[38]

$$S_{\text{k}} \approx \frac{-R}{N_{\text{k}}} \left(\sum_{i=1}^{N_{\text{k}}} \ln g_{\text{NN},i}(\mathbf{r}, \boldsymbol{\omega}) + \gamma \right) \quad (13)$$

where the nearest neighbors density, $g_{\text{NN}}(\mathbf{r}, \boldsymbol{\omega})$, is:

$$g_{\text{NN},i}(\mathbf{r}, \boldsymbol{\omega}) = \frac{8\pi^2}{\rho^\circ V_{\text{NN},i}^{\text{combined}}} \quad (14)$$

Here, $V_{\text{NN},i}^{\text{combined}}$ is the nearest neighbor volume for water i , which is the volume of a six dimensional sphere whose radius

is the distance between water i and its nearest neighbor, in the combined translational and orientational space.

$$V_{NN,i}^{\text{combined}} = \frac{\pi^3 d_{\text{combined}}^6}{6} \quad (15)$$

$$d_{\text{combined}} = \sqrt{(d_{\text{orient}}^2 + d_{\text{trans}}^2)} \quad (16)$$

A water's combined nearest neighbor distance is obtained by considering all waters found within the same and neighboring voxels. Limiting the nearest neighbor search space to this region is computationally efficient and avoids the possibility that position and orientation unit magnitudes can skew the resulting nearest neighbor distance to favor rotations or translations as opposed to their combination.^[39]

Molecular dynamics simulations

GIST analysis was applied to data from molecular dynamic simulations of cucurbit[7]uril (CB7),^[40,41] in a box of 1096 TIP3P water molecules, and to coagulation Factor Xa^[34] in a box of 27381 TIP3P water molecules. The simulations were performed with Amber14^[30,31] GAFF/AM1-BCC^[42,43] parameters for cucurbit[7]uril (CB7), Amber14SB^[44] parameters for Factor Xa, and the TIP3P water model.^[45] GIST mapping was performed with a prerelease version of AmberTools16,^[30] and all visualizations were conducted using Visual Molecular Dynamics (VMD) 1.91.^[32] The systems were prepared in tleap, minimized with 20,000 steps of the steepest descent algorithm, and then heated to 300 K over 240 ps in the isochoric–isothermal (NVT) ensemble. Each system was then equilibrated for 20 ns in isobaric–isothermal (NPT) conditions at 300 K and 1 atm pressure. The production simulation of CB7 was performed for 1 μ s and the production simulation of Factor Xa was performed for 150 ns. Both production runs were in NVT conditions maintaining a temperature of 300 K. The heavy atoms of Factor Xa and CB7 were restrained to the lab frame using harmonic restraints with a force constant of 2.5 kcal/Å². Temperature was controlled using a Langevin thermostat^[46] with a collision frequency of $\gamma = 1.0 \text{ ps}^{-1}$. The Berendsen barostat^[47] with isotropic volume scaling and a coupling constant of 1.0 ps was utilized for pressure regulation. The SHAKE^[48] algorithm was utilized to constrain the length of bonds to hydrogen at their equilibrium bond length permitting a 2 fs integration time step to be used. Production simulation trajectory coordinate frames were recorded at 1 ps intervals.

Results

Using CPTRAJ-GIST

GIST requires as input a CPTRAJ-readable molecular dynamics trajectory file and a corresponding topology file. It is recommended that the simulation be carried out with the atoms of the solute of interest (e.g., a protein) restrained in the lab frame, and that the simulations be run for a minimum of 30 ns,^[27] with snapshots saved every 3–5 ps for optimal use of nearest neighbor algorithms.^[38]

A GIST Study can be Broken into Three Steps:

Step 1. Identify and define the region of interest. The region of interest is a rectangular prism with center (x, y, z) and dimensions (L_x, L_y, L_z) input by the user. Visualization software, such as AutoDockTools,^[49] can be utilized to determine the grid center and dimensions.

Step 2. Run GIST-cpptraj. The required inputs for running GIST are the MD trajectory and topology files, the coordinates specifying both the dimensions and center of the region of interest and the name of the output file.

Step 3. Analyze GIST results. Visualize and manipulate output.

These steps are detailed below.

Step 1. Identify and define the region of Interest. The location, dimensions and spacing of the 3D grid are defined on the cpptraj command line:

```
gist gridcntr <xval> <yval> <zval> griddim
<xval> <yval> <zval> gridspacn <val> out
<filename>
```

Here the `gridcntr` x , y , and z values define the center of the grid in units of Angstroms (Å). For a protein-binding site a good choice for the center of the grid is at one of the central atoms of a cocrystallized ligand (Fig. 2). Note, however, that a common usage is the analysis of water in a binding site in the absence of a ligand. In such simulations the GIST analysis provides information on the structure and thermodynamics of water that is displaced upon ligand binding. In these cases the coordinates of an arbitrary atom in the active site may be convenient, or one may use any preferred method to identify a central position in the binding site. In choosing the grid placement and dimension several docking GUIs may be useful to visualize the region of interest, such as AutoDock Tools.^[49]

The grid is divided into cubic voxels with sides of length `gridspacn` (Å), and hence volume $V_{\text{vox}} = (\text{gridspacn})^3$, and the number of voxels along each dimension of the box is given by `griddim`. Smaller voxels provide finer spatial resolution in the GIST maps but require longer simulations to converge. Based on earlier studies the default grid spacing of 0.5 Å provides a reasonable balance of speed, resolution, and convergence of voxel-by-voxel thermodynamic quantities.

Step 2. Run GIST-cpptraj. Since GIST is incorporated into cpptraj, the topology, trajectory files, and GIST command line can all be specified in a simple cpptraj input file. A sample input file, `cpptraj.gist.in`, follows, with filenames rendered in *italics*.

```
parm topology.prmtop
trajin trajectoryfile.mdcrd 1 30000
gist gridcntr 25.0 31.0 30.0 griddim 40 40 40
gridspacn 0.50 out gist.out
go
```

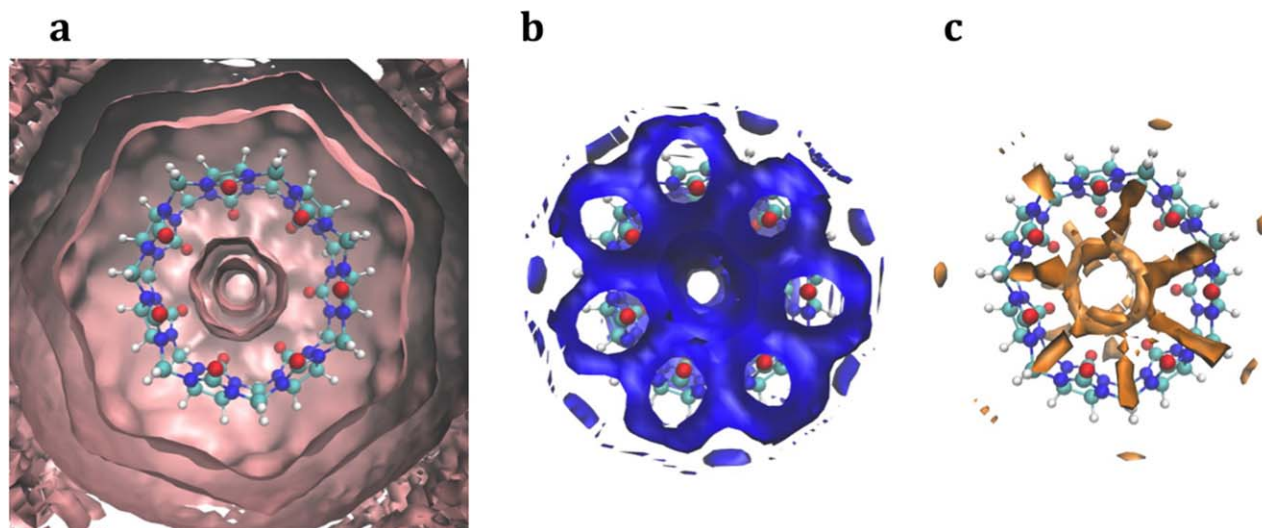


Figure 3. Sample visualizations of GIST output for host molecule CB7. a) Solvation shells around CB7 produced by visualizing *gist-gO.dx* with cutoff 0.98. b) Favorable solute–water interactions found near the oxygens of CB7, produced by visualizing *gist-Esw-dens.dx* with cutoff -0.24 (kcal/Å³). c) Unfavorable combined entropy found between the oxygens of CB7, visualized from *gist-dTsix-dens.dx* with cutoff -0.18 (kcal/Å³). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The first line defines which topology file to use; the second defines the trajectory file and the range of frames to be analyzed, here the first 30,000; the third line defines the center of the grid region to be analyzed and specifies that the region is a cube of dimensions $40 \times 40 \times 40$ voxel lengths; the fourth line gives the voxel length of 0.50 Å and defines the name of the output file; and the fifth line tells *cpptraj* to start the analysis.

The syntax for running *cpptraj* with input file *cpptraj.gist.in* is:

```
cpptraj -i cpptraj.gist.in
```

Step 3. Analyze GIST Results. GIST-*cpptraj* outputs gridded maps of solvent thermodynamic and structural quantities in dx file format, which can be visualized in most model viewing software, such as VMD and PyMol.^[32,33] Quantities not printed in dx files by default can be found in the text output and can be converted into dx format, if desired, by GISTPP. When visualized, these maps highlight key features of surface solvation such as high-density regions, or regions where the solvent interacts favorably with the surface or has unfavorable entropy, as illustrated for CB7 in Figure 3. Though in this study we demonstrate the utility of GIST on a host-guest system an in depth application of GIST on a protein-ligand system has been demonstrated previously and is also shown in the AmberTools GIST tutorial.^[28,35,50]

Postprocessing results with GISTPP

To aid in GIST analysis we have developed a suite of postprocessing tools, GISTPP, which facilitates the manipulation of GIST output files. The GISTPP package provides a comprehensive set of operations that allow end-users to easily manipulate the dx files generated by GIST. Each command specifies an operation, one or two input dx files, an output file, and any options related to the operation:

```
gistpp -i infile -i2 infile2 -op operation -  
opt options -o outfile
```

A list of the key operations and brief explanations can be found in Table 3. A user manual can be found at ambermd.org.^[50]

For example, two dx files can be easily summed together from the command line:

```
gistpp -i gist-Eww-dens.dx -i2 gist-Esw-  
dens.dx -i op add -o Etot.dx
```

Here, the solute–water energy map and the water–water energy map are summed to obtain a dx map, *Etot.dx*, of the total water–system energy.

More complex manipulations can be achieved with sequential GISTPP operations. For example, in a previous study, we

Table 3. Selected GISTPP operations with descriptions.

Operation	Description
Mult	Multiplies two dx maps together
Add	Adds two dx maps together
Sub	Subtracts dx map 2 from dx map 1
Div	Divides dx map 1 by dx map 2
filter1	Create a binary dx ^[a] map of voxels that are above or below a threshold
sum	Integrates all voxel quantities in a given dx map, printing the sum
defbp	Create a binary map around a pdb structure by flagging all voxels found within a user defined distance of all heavy atoms
SAS	Creates a dx map of the real solvent accessible surface, ^[19,20] which comprises voxels where the water density is greater than 0.1 of bulk density ($g(O) > 0.1$) and at least one neighbor voxel has water density less than 0.1 of bulk ($g(O) < 0.1$)

[a] A binary map contains a 1 in each voxel that meets a criterion and a 0 in each voxel that does not.

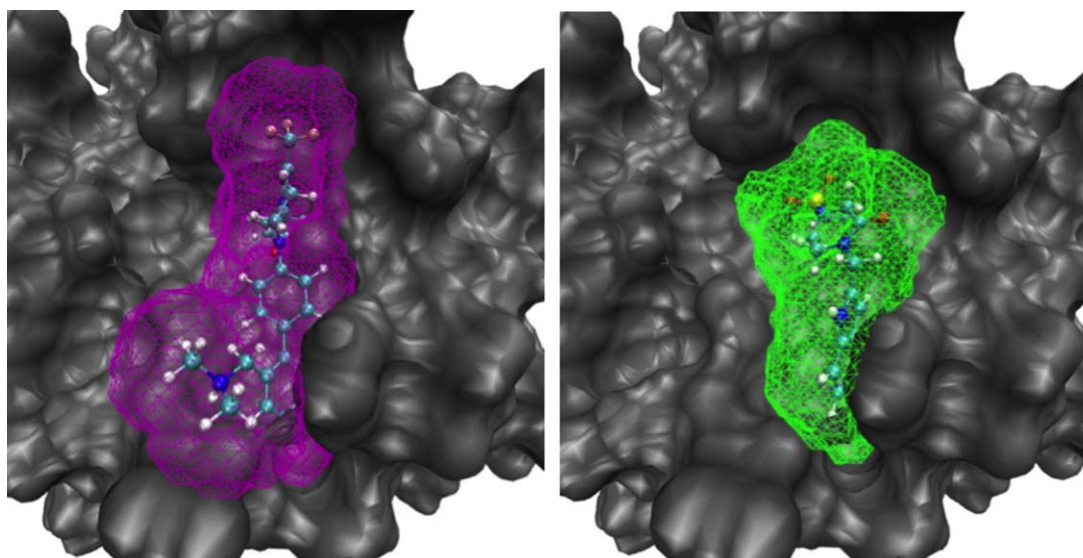


Figure 4. Scoring regions of two ligands from a congeneric pair study of Factor Xa, ligand RRR (left, purple) and 4QC (right, green).^[10] [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

found that the displacement of high energy and high density water from the Factor Xa binding site was highly correlated to protein–ligand binding affinity.^[35] To map out such regions, we would first add the water–water energies and the solute–water energies to obtain a total energy map (as shown above). One may then create binary maps, which define voxels with high density and high energy, respectively:

```
gistpp -i gist-gO.dx -op filter1 -opt gt1 -opt
cutoff1 2 -o Highdens.dx
gistpp -i Etot.dx -op filter1 -opt gt1 -opt
cutoff1 -9.53 -o HighEnergy.dx
```

Multiplying the binary maps of high density and high energy creates a binary scoring map, which marks only those voxels at both high density and high energy:

```
gistpp -i Highdens.dx -i2 HighEnergy.dx -op
mult -o ScoringRegions.dx
```

The scoring regions displaced by a given ligand can then be mapped out by creating a mask of water that is displaced by a given ligand, defined by a PDB file, here RRR.pdb from a congeneric pair study.^[10]

```
gistpp -i ScoringRegions.dx -i2 RRR.pdb -op
defbp -o DisplacedRRR.dx
gistpp -i DisplacedRRR.dx -i2 ScoringRe-
gions.dx -op mult -o ScoringRegionsRRR.dx
```

As in our previous study the solvent displacement score for each ligand can then be calculated by finding the number of scoring voxels displaced (Fig. 4):

```
gistpp -i ScoringRegionsRRR.dx -op sum
```

In our previous study,^[35] we found that the differences in these scores for congeneric pairs of ligands were highly correlated ($R^2 = 0.92$) with relatively binding affinities from experiment.

Convergence of entropy

The convergence of the first-order hydration entropy was examined by applying both Method 1 (separated) and Method 2 (combined) to a region within 3 Å of cocrystallized ligand RRR in the Factor Xa binding site, for MD snapshots saved at 1 ps intervals. As shown in Figure 5, the translational entropy converges to within 0.002 [kcal/(mol water)] of the final value after 2 ns. The orientational entropy converges to within 0.2 [kcal/(mol water)] of the final value after 2 ns; 0.04 [kcal/(mol water)] of the final value after 10 ns; and 0.01 [kcal/(mol water)] after 100 ns. Convergence of the summed entropy (Fig. 5) is thus limited by the orientational component. Interestingly, the combined

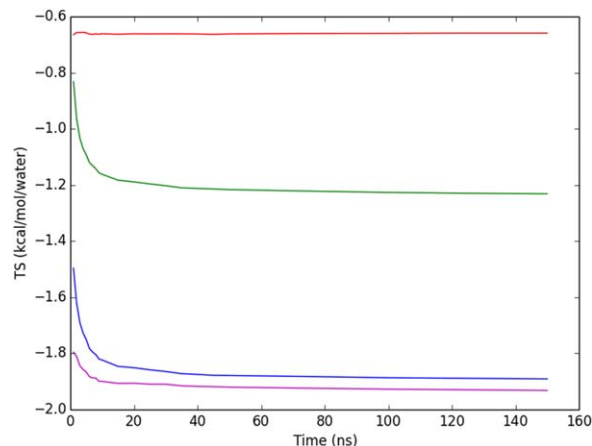


Figure 5. Convergence of per water entropy in a region within a 3Å around the heavy atoms of Factor Xa ligand RRR. Red: translational entropy; green: orientational; blue: sum of translational and orientational; purple: combined six dimensional total entropy (Method 2). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

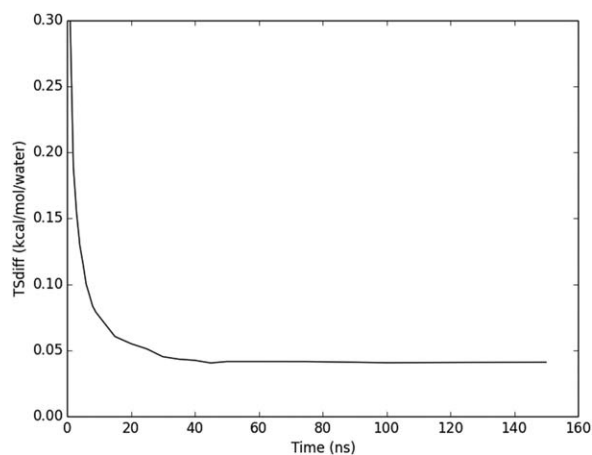


Figure 6. Difference in total first order per water entropy from Method 1 versus Method 2, for a 3 Å volume around the heavy atoms of Factor Xa ligand RRR.

method (Fig. 5), which directly provides the sum of the orientational and translational terms, converges more rapidly than the sum from the separated method, converging to within 0.1 [kcal/(mol water)] by 2 ns, within 0.04 [kcal/(mol water)] by 10 ns, and 0.004 [kcal/(mol water)] after 100 ns. The agreement between both methods is excellent, as is evident from Figure 5 and the graph of the difference between the two methods as a function of simulation time (Fig. 6). The slight difference between separated and combined nearest neighbor formulations represents the error from smoothing within voxels in the conditionally separated formulation.

Convergence of energy

GIST evaluates hydration energy as separate solute-water (E_{sw}) and water-water (E_{ww}) contributions. Here we examine the convergence of their sum ($E_{tot} = E_{sw} + E_{ww}$), again for the region within 3 Å of the heavy atoms of Factor Xa ligand RRR. As shown in Figure 7, the energy is converged to within 0.1 [kcal/(mol water)] of the final value within 2 ns, to within 0.02 [kcal/(mol water)] by 10 ns, and 0.004 [kcal/(mol water)] by 100 ns respectively.

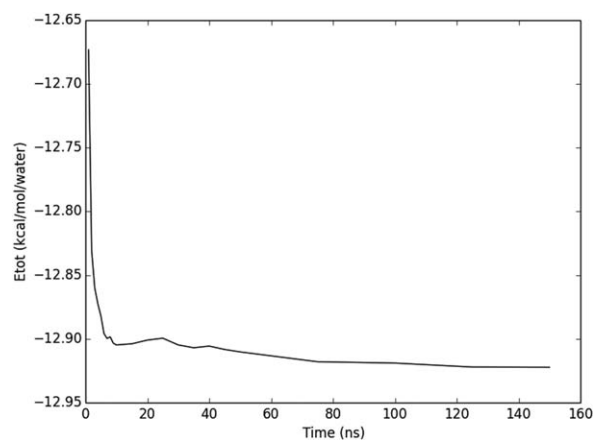


Figure 7. Convergence of E_{tot} , normalized per water molecule, for a region 3 Å around the heavy atoms of Factor Xa ligand RRR.

End states analysis

In previous studies we correlated Factor Xa binding affinities with water properties from GIST using a displaced solvent functional.^[10,35] This analysis relied upon the investigation of the displacement of water from the protein binding site alone and did not include contributions from the solvent reorganization upon binding or from the desolvation of the ligand.

Here we outline how GIST-cpptraj can be used to estimate ΔG_{solv} of a ligand-binding event that includes both the solvent reorganization around the complex and the desolvation of the ligand. In order to do this, three independent MD simulations would be run and analyzed by GIST: the solvated protein–ligand complex, the protein and the ligand. The first step of the analysis would be to identify a region in which the water is energetically and entropically perturbed from bulk values around the solute. This is illustrated in Figure 8.

The next step would be to calculate the perturbation of energy and entropy of the water from bulk in each of these three regions.

$$\Delta E_{solv} = \left(\sum_{i=1} E_i \right) - N E_{bulk}$$

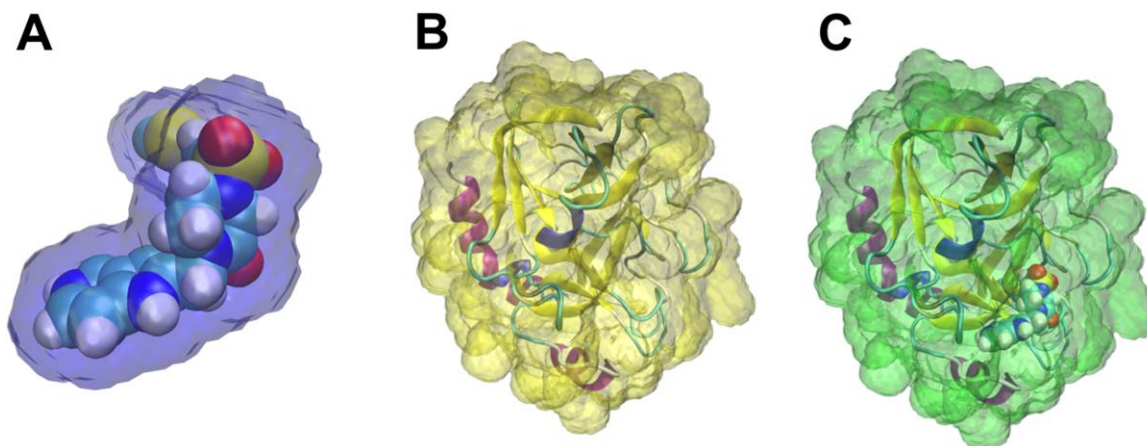


Figure 8. End states volumes defined around ligand RRR (a, blue), Factor Xa (b, yellow), and the complex formed by these two molecules (c, green). The free energy of solvation defined by these volumes can be used to estimate the free energy of solvation during the binding of ligand RRR. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$\Delta G_{\text{solv}} = \left(\sum_{i=1}^N S_i \right) - N S_{\text{bulk}}$$

where E_i and S_i are, respectively, the energy and entropy of water found within voxels i the region of interest, N is the average number of waters in the perturbed volume, E_{bulk} and S_{bulk} are the energy and entropy of a single water molecule in the neat solution. The free energy of perturbation would then be:

$$\Delta G_{\text{solv}} = \Delta E_{\text{solv}} - T \Delta S_{\text{solv}}$$

The change in solvation free energy for the binding process, $\Delta G_{\text{solv}}^{\text{bind}}$, would be:

$$\Delta G_{\text{solv}}^{\text{bind}} = \Delta G_{\text{solv}}^{\text{complex}} - \left(\Delta G_{\text{solv}}^{\text{protein}} + \Delta G_{\text{solv}}^{\text{ligand}} \right)$$

Example GISTPP commands that accomplish this approach for Factor Xa and ligand RRR are provided in the Supporting Information.

Discussion and Conclusions

The incorporation of GIST calculations into the AMBER CPPTRAJ analysis suite affords a powerful open-source tool for computing and mapping the properties of water at molecular surfaces in high resolution. This is supplemented by the GISTPP postprocessing code, which facilitates a wide variety of analyses and visualizations.

Lower resolution implementations of IST using site based approaches have been successful in computer aided-drug design applications, such as describing scoring affinity differences between congeneric pairs of ligands,^[10,51] improving docking and virtual screening, and understanding specificity of binding between members of the same protein family.^[51–54] These techniques have been applied towards advancing our understanding of the hydrophobic effect^[55–60] and elucidating where implicit models deviate from explicit and how appropriate corrections to implicit water models can be made.^[61]

GIST's high-resolution approach has had similar success in correlating binding affinities in Factor Xa and we anticipate a higher resolution picture will afford greater insight in the above and other applications. In particular, the mapping of solvation structure and thermodynamics onto a grid is easily amenable to incorporation into commonly used grid based docking scoring functions. We have also illustrated how GIST can be used in an end states approach that will allow for the straightforward comparisons to other techniques aimed at predicting binding affinities such as free energy perturbation and thermodynamic integration.

Both GIST-cpptraj and GISTPP come with detailed manuals and tutorials. This new toolkit provides users with capabilities to consider surface solvation in basic and applied computational studies, and provides a basis for further method development.


Acknowledgments

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. M.K.G. has an

equity interest in and is a cofounder and scientific advisor of VeraChem LLC.

Keywords: water · solvation thermodynamics · inhomogeneous fluid solvation theory · drug design · molecular recognition · AMBER

How to cite this article: S. Ramsey, C. Nguyen, R. Salomon-Ferrer, R. C. Walker, M. K. Gilson, T. Kurtzman *J. Comput. Chem.* **2016**, DOI: 10.1002/jcc.24417

 Additional Supporting Information may be found in the online version of this article.

- [1] S. E. Wong, F. C. Lightstone, *Expert Opin Drug Discov* **2010**, *6*, 65.
- [2] R. Abel, L. Wang, R. A. Friesner, B. J. Berne, *J Chem Theory Comput* **2010**, *6*, 2924.
- [3] C. Bissantz, B. Kuhn, M. Stahl, *J Med Chem* **2010**, *53*, 5061.
- [4] S. Riniker, L. J. Barandun, F. Diederich, O. Kraemer, A. Steffan, W. F. van Gunsteren, *J Comput Aided Mol Des* **2012**, *26*, 1293.
- [5] C. S. Poornima, P. M. Dean, *J Comput Aided Mol Des* **1995**, *9*, 500.
- [6] A. T. Garcia-Sosa, S. Firth-Clark, R. L. Mancera, *J Chem Inform Model* **2005**, *45*, 624.
- [7] J. E. Ladbury, *Chem Biol* **1996**, *3*, 973.
- [8] G. Hummer, *Nat Chem* **2010**, *2*, 906.
- [9] R. L. Mancera, *Curr Opin Drug Discov Dev* **2007**, *10*, 275.
- [10] R. Abel, T. Young, R. Farid, B. J. Berne, R. A. Friesner, *J Am Chem Soc* **2008**, *130*, 2817.
- [11] Z. Li, T. Lazaridis, *J Am Chem Soc* **2003**, *125*, 6636.
- [12] Z. Li, T. Lazaridis, *J Phys Chem. B* **2006**, *110*, 1464.
- [13] R. Baron, P. Setny, J. A. McCammon, *J Am Chem Soc* **2010**, *132*, 12091.
- [14] T. Young, R. Abel, B. Kim, B. J. Berne, R. A. Friesner, *Natl Acad Sci* **2007**, *104*, 808.
- [15] T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata, *J Am Chem Soc* **2005**, *127*, 15334.
- [16] M. Baroni, G. Cruciani, S. Sciabola, F. Perruccio, J. S. Mason, *J Chem Inform Model* **2007**, *47*, 279.
- [17] High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks. Jonathan S. Mason, A. Bortolato, D. R. Weiss, F. Defforian, B. Tehan, and F. H. Marshall. In *Silico Pharmacology* **2013**, 1:23.
- [18] A. S. Bayden, D. T. Moustakas, D. Joseph-McCarthy, M. L. Lamb, *J Chem Inform Model* **2015**, *55*, 1552. doi: 10.1021/ci500746d
- [19] C. Velez-Vega, D. J. J. McKay, V. Aravamuthan, R. Pearlstein, J. S. Duca, *J Chem Inform Model* **2014**, *54*, 3344.
- [20] C. Velez-Vega, D. J. J. McKay, T. Kurtzman, V. Aravamuthan, R. A. Pearlstein, J. S. Duca, *J Chem Theory Comput* **2015**, *11*, 5090.
- [21] G. Cui, J. M. Swails, E. S. Manas, *J Chem Theory Comput* **2013**, *9*, 5539.
- [22] B. Hu, M. A. Lill, *J Comput Chem* **2014**, *35*, 1255.
- [23] Z. Li, T. Lazaridis, *Methods Mol Biol* **2012**, *819*, 393.
- [24] J. Michel, J. Tirado-Rives, W. L. Jorgensen, *J Phys Chem B* **2009**, *113*, 13337.
- [25] C. M. Fraser, A. Fernández, L. R. Scott, C. M. Fraser, *WRAPPA: A Screening Tool for Candidate Dehydron Identification*. University of Chicago (2011).
- [26] E. D. López, J. P. Arcon, D. F. Gauto, A. A. Petruk, C. P. Modenutti, V. G. Dumas, M. A. Marti, A. G. Turjanski, *Bioinformatics* **2015**. doi: 10.1093/bioinformatics/btv411.
- [27] C. N. Nguyen, T. K. Young, M. K. Gilson, *J Chem Phys* **2012**, *137*, 44101.
- [28] C. Nguyen, M. K. Gilson, T. Young, *arXiv:1108.4876* **2011**.
- [29] D. R. Roe, T. E. Cheatham, *J Chem Theory Comput* **2013**, *9*, 3084.
- [30] D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham, I. I. T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R.

- C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York, P. A. Kollman, Univ Calif San Francisco **2015**.
- [31] R. Salomon-Ferrer, D. A. Case, R. C. Walker, *Wiley Interdiscip Rev Comput Mol Sci* **2013**, 3, 198.
- [32] W. Humphrey, A. Dalke, K. Schulten, *J Mol Graph* **1996**, 14, 33.
- [33] Delano, W. The PyMOL Molecular Graphics System, DeLana Scientific: Palo Alto, CA, **2002**.
- [34] S. Maignan, J.-P. Guilloteau, Y. M. Choi-Sledeski, M. R. Becker, W. R. Ewing, H. W. Pauls, A. P. Spada, V. Mikol, *J Med Chem* **2003**, 46, 685.
- [35] C. N. Nguyen, A. Cruz, M. K. Gilson, T. Kurtzman, *J Chem Theory Comput* **2014**, 10, 2769.
- [36] V. H. Harshinder Singh, *Am J Math Manage Sci* **2003**, 23.
- [37] D. J. Huggins, *J Comput Chem* **2014**, 35, 377.
- [38] D. J. Huggins, *J Chem Theory Comput* **2014**, 10, 3617.
- [39] F. Fogolari, et al., *J Chem Theory Comput* **2016**, 12, 1.
- [40] J. Kim, I.-S. Jung, S.-Y. Kim, E. Lee, J.-K. Kang, S. Sakamoto, K. Yamaguchi, K. Kim, *J Am Chem Soc* **2000**, 122, 540.
- [41] W. S. Jeon, K. Moon, S. H. Park, H. Chun, Y. H. Ko, J. Y. Lee, E. S. Lee, S. Samal, N. Selvapalam, M. V. Rekharsky, V. Sindelar, D. Sobransingh, Y. Inoue, A. E. Kaifer, K. Kim, *J Am Chem Soc* **2005**, 127, 12984.
- [42] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J Comput Chem* **2004**, 25, 1157.
- [43] A. Jakalian, D. B. Jack, C. I. Bayly, *J Comput Chem* **2002**, 23, 1623.
- [44] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J Chem Theory Comput* **2015**. doi: 10.1021/acs.jctc.5b00255.
- [45] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J Chem Phys* **1983**, 79, 926.
- [46] R. W. Pastor, B. R. Brooks, A. Szabo, *Mol Phys* **1988**, 65, 1409.
- [47] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*; Oxford University Press: Oxford, **1989**.
- [48] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J Comput Phys* **1977**, 23, 327.
- [49] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. Olson, *J Comput Chem* **2009**, 30, 2785.
- [50] AMBER Advanced Tutorial 25—Grid Inhomogeneous Solvent Theory (GIST) Example: Introduction. <http://ambermd.org/tutorials/advanced/tutorial25/>; Accessed 4th January **2016**.
- [51] N. K. S. Robert Abel, *ChemMedChem* **2011**, 6, 1049.
- [52] F. Shah, J. Gut, J. Legac, D. Shivakumar, W. Sherman, P. J. Rosenthal, M. A. Avery, *J Chem Inform Model* **2012**, 52, 696.
- [53] R. A. Pearlstein, W. Sherman, R. Abel, *Proteins Struct Funct Bioinform* **2013**, 81, 1509.
- [54] T. Beuming, Y. Che, R. Abel, B. Kim, V. Shanmugasundaram, W. Sherman, *Proteins Struct Funct Bioinform* **2012**, 80, 871.
- [55] T. Lazaridis, *Acc Chem Res* **2001**, 34, 931.
- [56] T. Lazaridis, M. E. Paulaitis, *J Phys Chem* **1992**, 96, 3847.
- [57] J. M. Fox, K. Kang, W. Sherman, A. Heroux, G. M. Sastry, M. Baghbanzadeh, M. R. Lockett, G. M. Whitesides, *J Am Chem Soc* **2015**, 137, 3859.
- [58] P. W. Snyder, J. Mecinovic, D. T. Moustakas, S. W. ThomasIII, M. Harder, E. T. Mack, M.R. Lockett, A. Heroux, W. Sherman, G. M. Whitesides, *Proc Natl Acad Sci* **2011**, 108, 17889.
- [59] M. R. Lockett, H. Lange, B. Breiten, A. Heroux, W. Sherman, D. Rappoport, P. O. Yau, P. W. Snyder, G. M. Whitesides, *Angew Chem Int Ed* **2013**, 52, 7714.
- [60] B. Breiten, M. R. Lockett, W. Sherman, S. Fujita, M. Al-Sayah, H. Lange, C. M. Bowers, A. Heroux, G. Krilov, G. M. Whitesides, *J Am Chem Soc* **2013**, 135, 15579.
- [61] L. Wickstrom, N. Deng, P. He, A. Mentis, C. Nguyen, M. K. Gilson, T. Kurtzman, E. Gallicchio, R. M. Levy, *J Mol Recognit* **2016**.

Received: 27 January 2016
Revised: 19 April 2016
Accepted: 17 May 2016
Published online on 00 Month 2016