A Generalized Fundamental Measure Theory for Atomistic Modeling of Macromolecular Crowding
Sanbo Qin and Huan-Xiang Zhou
Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, USA

Supplementary Information

1. Further details on the crowder-exclusion surface
   We point out that the volume $v_p$ and area $s_p$ as described in the main text are those of the crowder-exclusion surface, but the points used for calculating $r_g$ are located on a surface that is slightly inflated. The crowder-exclusion surface and the inflated surface are shown in Fig. S1.

Fig. S1  Comparison of (a) the crowder-exclusion surface for calculating $v_p$ and $s_p$ and (b) the inflated surface for calculating $r_g$. 
The difference between the two surfaces can be illustrated by a simple test particle: a spherocylinder (see Fig. S2). For this convex particle, the crowder-exclusion surface is identical to the surface of the particle. The inflated surface for calculating $r_g$ consists of the proximal intersection points between rays emanating from the center of the test particle and crowders around the test particle; the crowders are all in contact with the test particles and the rays end at the center positions of the crowders. Note that, when the crowder radius (denoted as $R_c$ hereafter) is zero, the inflated surface coincides with the crowder-exclusion surface.

![Fig. S2](image)

Fig. S2 The inflated surface (in blue) illustrated on a spherocylindrical test particle (in red). A crowder in contact with the test particle is shown as a circle in dash. The arrow shows a ray pointing from the center of the test particle to the center of the crowder. A black dot indicates the proximal intersection of the ray and the crowder; the collection of such intersection points as the crowder rolls around the test particle defines the inflated surface.

Let the spherocylinder, with radius $R_{sc}$, be oriented such that its center is at the origin and its central axis be along the $z$ axis of a Cartesian coordinate system. For a crowder located at $(x_c, 0, z_c)$, the proximal intersection is located at $(x_i, 0, z_i)$, where

$$x_i = x_c \left[ 1 - 2 \frac{R_c}{(x_c^2 + z_c^2)^{1/2}} \right]$$

$$z_i = z_c \left[ 1 - \frac{R_c}{(x_c^2 + z_c^2)^{1/2}} \right]$$

(1a) (1b)

A crowder in contact with the cylindrical portion of the test particle has $x_c = R_{sc} + R_c$ and $-L_{sc} < z_c < L_{sc}$, where $L_{sc}$ the half length of the cylindrical portion. A crowder in contact with the upper hemisphere of the test particle has $x_c = (R_{sc} + R_c)\sin\theta$ and $z_c = (R_{sc} + R_c)\cos\theta + L_{sc}$, where $0 < \theta < \pi/2$.

For a test protein represented at the atomic level, we found that $r_g$ calculated on the inflated surface is $\sim 10\%$ higher than $r_g$ calculated on the crowder-exclusion surface. We also found that the higher $r_g$ leads to better agreement with the $\Delta\mu$ results obtained by the insertion procedure, explaining why we use the inflated surface for $r_g$. Note that both the crowder-exclusion surface and the inflated surface depend on the crowder radius.
2. Comparison of $l_p$ and $r_g$ for a spherocylindrical test protein

According to the fundamental measure theory [1], the linear size $l_p$ of a spherocylindrical test protein is

$$l_p = R_{sc} + L_{sc}/2. \tag{2}$$

The radius of gyration can be calculated from

$$r_g^2 = \frac{\oint r_i^2 \, ds}{\oint ds}, \tag{3}$$

where $r_i = (x_i^2 + z_i^2)^{1/2}$ is the distance from the center of the spherocylinder to a point on the inflated surface; the integration is over the entire inflated surface, and $ds$ is the surface area element. For comparison, we also calculate the integration on the crowder-exclusion surface (which in the present case is the same as the surface of the test particle), and denote the corresponding result by $r_{g0}$. When $R_c = 0$, $r_g$ reduces to $r_{g0}$. For $r_{g0}$, the integration can be calculated analytically. The result is

$$r_{g0}^2 = \frac{L_{sc}/R_{sc}}{L_{sc}/R_{sc} + 1} \left( R_{sc}^2 + L_{sc}^2 / 3 \right) + \frac{1}{L_{sc}/R_{sc} + 1} \left( R_{sc}^2 + L_{sc}^2 + R_{sc} L_{sc} \right). \tag{4}$$

We have not been able to calculate the integration analytically for $r_g$; therefore we calculate $r_g$ by numerical integration.

When $L_{sc}/R_{sc} \rightarrow 0$, the spherocylinder becomes a sphere, we have $r_g = r_{g0} = R_{sc}$, which is the correct result for $l_p$. In the opposite limit $L_{sc}/R_{sc} \rightarrow \infty$, the spherocylinder becomes a long needle; then $r_g \rightarrow r_{g0} \rightarrow L_{sc}/3^{1/2}$, which slightly overestimates the result of Eq. (2) for this case, $l_p = L_{sc}/2$. Figure S3 shows that $r_g$ is numerically close to $l_p$ over a wide range of $L_{sc}/R_{sc}$.

![Figure S3](image_url)

Fig. S3 Comparison of $l_p$ and $r_g$ for a spherocylindrical test protein. The crowder radius, scaled by $R_{sc}$, is either 0 or 1.
3. Comparison of methods for calculating $s_p$ and $v_p$

The $s_p$ and $v_p$ results obtained by our method are very close to those obtained by a method recently developed by Voss et al. [2]. Figure S4 shows a comparison on the results of the two methods for 8 test systems when the crowder radius is 15 Å. The relative differences in $s_p$ are < 0.5%; those in $v_p$ are < 1.3%.

![Figure S4](image)

Fig. S4 Comparison of (a) $s_p$ and (b) $v_p$ results between our method and that of Voss et al. Areas and volumes are in Å$^2$ and Å$^3$, respectively. Diagonal lines indicate exact match of the two methods. $s_p$ values from small to large correspond to barstar, θ subunit of E. coli DNA polymerase III (Pol III), barnase, folded cytochrome $b_{562}$, unfolded cytochrome $b_{562}$, ε subunit of Pol III, barnase-barstar complex, and Pol III ε-θ complex. The orders between barstar and Pol III θ and between barnase and folded cytochrome $b_{562}$ are reversed according to $v_p$ values. Averages over protein conformations (see below) are shown, with error bars representing variations among the conformations.

Our method is significantly faster than that of Voss et al. The CPU times of our method are independent of the crowder radius, and are ~10 s per conformation for the barnase-barstar complex. In contrast, the CPU times of the method of Voss et al. show a significant increase with increasing crowder radii, going from 54 s at a crowder radius of 15 Å to 1030 s at a crowder radius of 30 Å for each conformation of the barnase-barstar complex.

4. Information on molecular dynamics simulations

In all direct simulations of crowding reported to date [3-7], coarse-grained representations of the test protein have been used. We have developed an alternative approach [8-10], referred to as postprocessing. In the postprocessing approach, the motions of the test protein and those of the crowders are followed in two separate simulations. The effects of crowding are then modeled by calculating $\Delta \mu$, the change in the chemical potential of the test protein. The calculation entailed fictitiously placing one by one the protein conformations randomly inside snapshots of the crowder trajectory, much like implementing Widom’s insertion theorem [11]. Like in direct simulations [3-7], the interactions between...
the atoms of the test protein and the crowders were assumed to be hard-core repulsion. In that case, $\Delta\mu$ is related to the average probability of successfully placing protein conformations into the snapshots of crowder configurations. A successful placement is one that is free of any protein-crowder overlap.

In essence, the generalized fundamental measure theory (GFMT) presented here is a new method for calculating $\Delta\mu$. To directly test this method, we study the same 8 test systems (listed in Fig. S4) as those in our previous work developing and applying the postprocessing approach [8, 9], and use the same protein conformations generated in the absence of crowders. The numbers of conformations used were 1000 each for folded and unfolded cytochrome $b_{562}$; 548 each for barnase, barstar, and their complex; and 700 each for Pol III $\varepsilon$ and $\theta$ subunits and their complex. Briefly, these conformations were sampled from molecular dynamics simulations of proteins solvated by TIP3P, which were run in the AMBER force field. For 7 of the 8 test systems, the simulations were at room temperature. For unfolded cytochrome $b_{562}$, a high temperature (500 K) was used to denature the protein. Once the conformations are selected, temperature enters the GFMT only together with $k_B$ to serve as the unit of $\Delta\mu$ [see Eq. (1) of the main text].