UC Riverside

UC Riverside Previously Published Works

Title

GeomBD3: Brownian Dynamics Simulation Software for Biological and Engineered Systems

Permalink

https://escholarship.org/uc/item/6sf8g826

Journal

Journal of Chemical Information and Modeling, 62(10)

ISSN

1549-9596

Authors

Cholko, Timothy

Kaushik, Shivansh

Wu, Kingsley Y

et al.

Publication Date

2022-05-23

DOI

10.1021/acs.jcim.1c01387

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2022 July 24.

Published in final edited form as:

J Chem Inf Model. 2022 May 23; 62(10): 2257–2263. doi:10.1021/acs.jcim.1c01387.

GeomBD3: Brownian Dynamics Simulation Software for Biological and Engineered Systems

Timothy Cholko,
Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States;

Shivansh Kaushik,
Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States;

Kingsley Y. Wu,
Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States;

Ruben Montes,
Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States;

Chia-en A. Chang
Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States;

Abstract

GeomBD3 is a robust Brownian dynamics simulation package designed to easily handle natural or engineered systems in diverse environments and arrangements. The software package described herein allows users to design, execute, and analyze BD simulations. The simulations use all-atom, rigid molecular models that diffuse according to overdamped Langevin dynamics and interact through electrostatic, Lennard-Jones, and ligand desolvation potentials. The program automatically calculates molecular association rates, surface residence times, and association statistics for any number of user-defined association criteria. Users can also extract molecular association pathways, diffusion coefficients, intermolecular interaction energies, intermolecular

Corresponding Author: Chia-en A. Chang – Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States; chiaenc@ucr.edu.

Author Contributions

The manuscript was written through contributions from T.C., S.K., and K.Y.W. T.C. and SK contributed to the source code. All authors applied and tested the software. All authors have given approval to the final version of the manuscript.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c01387.
Additional simulation condition, sample parametrized (.pqr) file of acetate ion and table for hybridization time, hybridizing probe site, fraction of 2D hybridization and DNA-SAM interaction energies for case 2 simulation condition, illustration figure of simulation setups for crowding and diluted systems and table for average association time, direct binding probability and surface dwell time for case 3 simulation condition, and figure for probability density of AchE-ach system for case 4 (PDF)

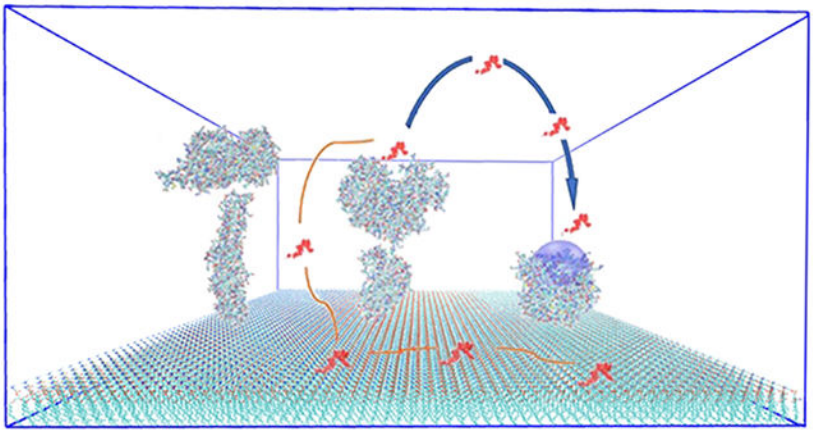
Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c01387

The authors declare no competing financial interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

contact probability maps, and more using the provided supplementary analysis scripts. We detail the use of the package from start to finish and apply it to a protein–ligand system and a large nucleic acid biosensor. GeomBD3 provides a versatile tool for researchers from various disciplines that can aid in rational design of engineered systems or play an explanatory role as a complement to experiments. GeomBD version 3 is available on our website at http://chemcha-gpu0.ucr.edu/geombd3/ and KBbox at https://kbbox.h-its.org/toolbox/methods/molecular-simulation/geombd/.

Graphical Abstract



INTRODUCTION

Biological and chemical processes *in vivo* and *in vitro* typically start out with diffusional processes by which two or more molecules encounter each other. These diffusional processes can take place on temporal and spatial scales up to milliseconds and micrometers,[1–3] rendering them beyond the reaches of methods such as molecular dynamics simulation. However, such processes must be studied to elucidate the kinetics and mechanisms of important reactions. Moreover, environmental, structural, and orientational modifications to complex systems can dramatically affect reactions taking place within, so the ability to rapidly assess these variables is extremely valuable. Understanding driving forces and mechanisms for binding assists the design of nanoscale molecular systems. In addition, most experimental techniques lack the ability to resolve subtle differences in device functionality that result from nanoscale modifications. Molecular simulation tools can bridge the gap and guide rational design of novel biotechnologies.[4,5]

Brownian dynamics (BD) has been used to investigate long time scale molecular processes for several decades.[6,7] Many software packages are available for studying a variety of problems related to bimolecular association.[8] MacroDox, UHBD, BrownDye, Simulation of Diffusional Association (SDA), BD_BOX, BROMOC, ReaDDy, Smoldyn, SEEKR, and BDpack are used to probe ligand–protein associations with a flexible or rigid biomolecular model.[9–18] Molecular interactions can be affected by their environments, and several BD programs also consider some environmental factors, such as molecular crowders. However, not all important system features can be modeled by existing programs. For example, nanostructures involving multienzyme cascades, systems with important surface properties,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and systems with ligands under diffusive flux in a particular direction cannot be simulated with existing BD packages.

The BD simulations carried out in our software, GeomBD3 (GeomBD version 3), are specifically designed for this purpose.[19] GeomBD3 simulates the molecular recognition process of two molecules over long distances and time scales in a vast array of different systems and system sizes. Systems may include many separate molecules and may utilize a range of various boundary conditions and initial ligand positions. The package simulates multiple replicas of a ligand in parallel and generates multiple trajectories of the same ligand. The software is easily capable of handling molecular models up to hundreds of nanometers in scale, thus allowing simulation of simple protein–ligand systems or very large nanoscale technologies, such as biosensors or multi-enzyme arrays. For example, the software has been used to study enzyme bioconjugates, understand DNA hybridization in biosensors,[20] and replicate SPR experiments under diffusive flux.[21] Users may also arrange multiple enzymes in space or on a membrane to study ligand diffusion in crowded environments. In many biosensing and catalyst applications, the device surface properties are of high importance, as they can control substrate adsorption and diffusion and thus overall efficiency.[22,23] GeomBD3 provides the ability to examine different experimental settings *in silico* to rapidly evaluate multiple design variables.

Here, we present a detailed description of GeomBD3 along with two applications of the software and basic analysis of the results. Users are guided through the general simulation workflow, from setup to execution to final analysis. We then present applications of the program to a simple protein–ligand system and a complex nucleic acid biosensor system.

METHODS

Implementation.

GeomBD3 source code is written entirely in C++, whereas several pre- and post-simulation scripts are available in both C++ and Python3 to aid in simulation setup or analysis. Simulation setup, execution, and analysis are carried out using standard PDB, PQR, and DCD file types. The source code can be divided into two categories: pre-simulation code (*Parameterize.py* and *Gridder*) for parametrization of the molecular models and calculation of potential energy grids and simulation code (*GBD*) for running, recording, and monitoring the Brownian dynamics.[19] These codes can be found on our group GitHub at http://github.com/chang-group.

Simulation Algorithm.

The basic workflow of the software package (Figure 1) is as follows. Initially, users should obtain PDB format molecular models of the system. GeomBD3 uses all-atom, rigid molecular models with no internal degrees of freedom. Generally, systems consist of a ligand and one or more other molecules that may or may not be receptors of the ligand. The ligand is the molecule which diffuses during the simulation, and the receptor(s) is(are) its intended binding partner(s). The ligand may range in size from a protein to a small drug molecule, but it should be chosen as the smaller of the binding partners since simulation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

time scales linearly with ligand size. The package also allows one to include multiple conformations of a ligand. Next, the PDBs are converted into PQR type files by using the *Parameterize.py* script. These PQRs are then fed into the *Gridder* programs that precompute as much of the potential energy calculations as possible to save time. Users are not required to generate all three available potential grids and should choose which are needed for the study.[19]

The remaining steps involve calculating the Brownian dynamics trajectories and monitoring output. An input configuration file should be supplied containing the system PQR files, desired grid files, and simulation parameters. Upon running the *GBD* code, first, several independent ligand replicates are positioned according to the desired starting conditions. If ligand replica lie on the potential energy grids, the intermolecular forces are calculated by the finite difference method. As an example, forces in the x dimension F_x are calculated as

$$F_x = \frac{U(gx+1, gy, gz) - U(gx-1, gy, gz)}{2d} \tag{1}$$

where the potential U is a function of an atom's location on the grid specified by *gx, gy,* and *gz*, and δ is the grid spacing. During simulation, the potentials at the coordinates of each ligand atom are approximated by trilinear interpolation between grid points. The linear interpolation of the force may lead to some inaccuracy for highly charged systems. However, the accuracy gained by using nonlinear interpolation is insignificant compared with the increased computational time, especially for large systems. A comparison between linear and cubic spline interpolation shows a 40% increase in the CPU time.[24] A random force R with the properties

$$\langle R(t) \rangle = 0 \tag{2}$$

$$\langle R(t) R(t') \rangle = \delta(t - t') \tag{3}$$

generated from a Gaussian process with a zero mean is also applied at this time, and ligands are moved to their next positions according to the overdamped Langevin equation

$$r_{t+\Delta t} = r_t - \frac{D}{k_B T} F \Delta t + \sqrt{2 D \Delta t} R \tag{4}$$

where r is the ligand position, D the appropriate diffusion coefficient, $k_B T$ is Boltzmann's constant times the absolute temperature, F the three-dimensional force calculated from its x, y, and z components such as eq 1, and Δt the linearly scaled time step. The translational and rotational diffusion coefficients are calculated based on the hydrodynamic radius r_h of either the ligand only or the ligand and receptor. The former case is only used when the receptor is attached to a surface i.e., not diffusing.

$$D_{\text{trans}} = \frac{k_B T}{6\pi\gamma} \left(\frac{1}{r_{h,\text{lig}}} + \frac{1}{r_{h,\text{rec}}} \right) \tag{5}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$D_{\text{rot}} = \frac{k_B T}{8\pi\gamma}\left(\frac{1}{r_{h,\text{lig}}{}^3} + \frac{1}{r_{h,\text{rec}}{}^3}\right) \quad (6)$$

Hydrodynamic interactions (HI) are not included. It has been observed by Northrup et al. that the effect of HI on k_{ON} in a similar modeling condition is ~6%.[25] At each new position, each binding criterion is checked, and if satisfied, a binding event is counted and stored. The bound replica of the ligand is then reset according to the starting conditions to begin a new trajectory. In this way, the number of ligand trajectories being simulated is constant. These steps are repeated until certain simulation parameters converge below a defined threshold, or the user manually terminates the program.

Relevant data pertaining to association rates, binding probabilities, site/surface-specific binding dwell times, ligand–receptor interactions, and simulation convergence are automatically output to a log file. Simulation trajectories are output as a DCD format file for viewing and analysis.

Simulation Setup.

The initial step in the simulation workflow is to obtain PQR format input files containing the molecular structures of the ligand and receptor. These PQR files can be generated by the script *Parameterize.py*, which takes in PDB files and converts them to the PQR format.[19]

Any file in a standard PDB format is acceptable, such as those downloadable from the Protein Data Bank.[26] GeomBD3 also handles nonstandard systems, as long as PDB molecular models can be created by the user and the parameters for new molecules are added to the parameter files. The parameter files are simple to read and edit, making addition of new parameters very straightforward. Instructions on preparing a simulation of a novel system are provided in the user manual at http://chemcha-gpu0.ucr.edu/geombd3/. The PQR files contain the coordinates, partial charges, and vdW radii of all atoms, which are later used in energy calculations. The values of partial charges for amino acids and standard nucleotides are taken from AMBER ff14sb,[27] while those for other molecules can be obtained using any software that performs quantum or semiempirical charge calculations. For instance, partial charges for Ach (case 4, SI) were calculated using the AM1-BCC method. The vdW radii were taken from the GAFF[28] force fields. Input PDB files of SAM and gold surfaces were modeled by the AMBER GAFF force field using the xleap package and then converted to PQR files.

After obtaining parametrized PQR files, users should run the series of *Gridder* programs to precompute the potential energy at all points in grid-based representations of the receptor(s) and any other non-ligand molecules included in the simulation. The following *Gridder* programs are currently available.[19]

Gridder-EX.cc.—This creates a grid (the EX grid) defining the excluded volume of the simulation space due to the presence of the receptor(s). It defines the space occupied by the receptor and prevents unphysical interpenetration of atoms. The calculation maps each atom of the receptor to a grid point, then searches a distance equal to the atom's LJ radii in *x*, *y*,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and z dimensions. Every grid falling within this search range is considered to be occupied by an atom. The EX grid should always be generated before running a simulation.

Gridder-ES.cc.—This creates a grid (the ES grid) holding the electric field strength of the receptor molecule at each point calculated by the screened Coulomb method.[29] Methods such as the Adaptive Poisson–Boltzmann Solver (APBS) were deemed impractical due to very high computational cost and memory usage. During simulation, the partial charge, q_n, on each ligand atom is multiplied by the appropriate grid value, $\varphi_{\mathrm{SC,grid}}(g)$, to give the electrostatic potential between that atom and the receptor.

$$\varphi_{\mathrm{SC,grid}}(g) = \sum_m \frac{k_e q_m e^{-kr_{gm}}}{r_{gm} \in} \tag{7a}$$

$$E_{\mathrm{SC,grid}}(g, q_n) = q\varphi_{\mathrm{SC,grid}}(g) \tag{7b}$$

$$\kappa = \sqrt{\frac{\sum_i (c_i q_i^2)}{\in \ \in_0 k_{\mathrm{B}} T}} \tag{8}$$

where k_e is Coulomb's constant and g a grid coordinate, and q_m and q_n are the charges on the mth receptor atom and nth ligand atom, respectively. r_{gm} is the distance between a grid point g and mth receptor atom, ε the dielectric of water, and ε_0 vacuum permittivity. κ is the inverse Debye length, where c_i and q_i are the molar concentration and formal charge of ionic species i.

Gridder-LJ.cc.—This creates grids (the LJ grids) storing the 12–6 Lennard-Jones potential E_{LJ} between the receptor and each atom type in the ligand.

$$E_{\mathrm{LJ}} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{9}$$

where ε is the potential well depth, σ the interatomic distance corresponding to zero potential, and r the interatomic distance.

Gridder-D.cc.—This creates a grid holding the desolvation potential of the ligand. This potential is based on the degree of desolvation of each atom and the atom's desolvation parameter, S. The function used in GeomBD3 borrows the functional form and parameters used in AutoDock4.2,[30] which is similar to the concept developed by Stouten et al.[31] Here, eq 10 is the energy calculation required to hydrate an atom using the volume-summing method, and eq 11 calculates atomic solvation parameter for each atom type required.

$$E_{\mathrm{desolv}} = \sum_{j=1}^{N} S_i V_j e^{-r_{ij}^2/2\sigma^2} \tag{10}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$S_i = \text{solpar} + 0.01097 q_i \tag{11}$$

where S_i is the desolvation parameter of ligand atom i, V_j the fragmental volume of receptor atom j, and r_{ij} the distance between them. Sigma is 3.5 Å, and N is the number of atoms in the receptor molecule. In S_i, solpar is an empirically derived part of the desolvation parameter specific to each atom type, and q_i is the partial charge on the atom.

Each grid has user-definable *padding* and *spacing* values. The *padding* variable is the distance the grid will extend in each dimension beyond the minimum/maximum atomic coordinates which can be configured by the user. The default *padding* values are 40, 12, and 12 Å for the ES, LJ, and D grids, respectively. The *spacing* variable is the distance between adjacent grid points and has a default 0.5 Å for all grids in the discussed cases. After observing highly similar results if the grid spacing is ≤ 0.5 Å, we suggest 0.375 to 0.5 Å for good performance in most systems.

Configuring and Running the Simulation.

Once all the desired precomputed grids have been made, the simulation configuration file should be set. The file consists of keywords followed by either a number or string. A user manual with a detailed description of each keyword and guidance on selecting appropriate values is provided on our website at http://chemcha-gpu0.ucr.edu/geombd3/. Here, we discuss only a brief overview of the file.

The generated PQR files should be specified after the *ligand* and *receptor* keywords. This is followed by the specifications of the grid files to be loaded. The *grid* keyword followed by a specific string denoting the grid type, for example, es for electrostatic, should be given first. Then, the path to the appropriate grid file should follow.[19]

The next set of keywords specifies the simulation parameters. These include *ligandStart*, *boundary*, *boundWall*, and *fixedReceptor*. For example, with *ligandStart* and *boundary*, users can select periodic or reflective boundary conditions, and the bounding cell may be either cuboid, hexagonal prismatic, or spherical. In addition, ligands may be started from an *xy*-aligned plane, a random point, or a specified point. Alternatively, simulations can be run under NAM (after Northrup, Allison, and McCammon) conditions,[25] in which ligands start from a spherical plane and exit at a larger spherical plane. The value of the inner spherical plane must be large enough to satisfy these conditions: the interaction between ligand and receptor can be considered centrosymmetric, and the ligand has no angle dependence with isotropic reactivity. These options are made available so that simulations can be made to match experimental and natural conditions or so that more appropriate comparisons can be made between simulated and theoretical results. Association rate constants in most simulations are calculated as

$$k_{\text{on}} = \frac{1}{ct_{\text{avg}}} \tag{12}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where t_{avg} is the average association time, and c is the ligand concentration estimated by the simulation volume. The concentration can be varied by using a different simulation volume.[7] Under NAM conditions, the association rate constant is calculated as

$$k_{on} = \frac{4\pi r D\beta}{1-(1-\beta)\frac{k_d(b)}{k_d(q)}} \quad (13)$$

where r is the ligand–receptor distance required for binding, D the relative diffusion coefficient of ligand and receptor, and β the fraction of ligands that bound before exiting, and b and q are the radii of the starting and exiting spherical planes, respectively.[25] Collison probability, β, is determined on the basis of simulations which yield similar results, and the starting position of the ligand must be angle independent. The *biasForce* keyword allows users to apply a biasing force along one dimension to mimic flow of the ligands inside the system. For example, such conditions were used to simulate surface plasmon resonance experimental conditions to investigate the effect of flow on ligand association to HIV protease.[21] Due to the external flow effect in the system, the calculation of the rate constant using eq 12 or 13 is not possible. The ligand diffusive flux resulting from this force can be calculated as

$$\mathcal{T} = \frac{cF_{bias}}{f} \quad (14)$$

where c is the ligand concentration, F_{bias} the biasing force, and f the friction force on the diffusing ligand.

The *bindgroups* keyword defines the criteria that must be met to achieve binding between the ligand and receptor. This should be carefully set based on visualization of the molecular models and knowledge of the system/process being simulated. Each binding criterion consists of an XYZ coordinate, an index number of a ligand atom, and a distance. Taken altogether, this represents the distance between the ligand atom and specified coordinate below which the two molecules are considered bound. For example, the input file line *bindgroups 5.0 10.0 15.0 6 12.0* specifies a condition in which ligand atom number 6 must be less than 12.0 Å from the Cartesian point 5.0 10.0 15.0 to record a binding event. The user can also include multiple binding sites with different combinations using “AND” and “OR” terms in *bindgroups*, which may further affect the binding pathways and rate. Notably, the bindgroups and distance between a ligand and bidgroup are system dependent. If an experimental k_{ON} is available, that value may be used to benchmark appropriate choice of binding conditions.

Simulations can be set to automatically terminate under specific conditions, or they can be manually terminated by the user at any time. Most often, simulations are set to terminate when some value of interest has converged below a certain threshold. For example, if mean binding times t_{avg} are desired, GeomBD3 can be set to terminate when the SEM (standard error of mean) of binding times is below a threshold set by the *convergence* input keyword. In such a case, setting *convergence* to 0.01 would cause the program to stop when the SEM/t_{avg} was less than 0.01. The calculation of t_{avg} gives insight into diffusion-limited rate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

constants, trends in reaction rates, and diffusion rates in chemically diverse environments. Because the program simulates a constant number of ligand trajectories, when many replicates are being run (e.g. 5000), the initial data in the log file are necessarily biased toward fast-binding trajectories. Generally, one needs to sample several times *Nreplicates* binding events before results are reasonably converged.

Simulation Output.

Five files are automatically output by GeomBD3. The log file (.log extension) displays updated information about the number of association events, the average association time, the SD (standard deviation) and SEM of association times, convergence of the mean association time, ligand exit events, dwell time, and ligand–receptor interaction energies. The ligand's time experiencing non-zero LJ potential is considered as dwell time, since it is diffusing under strong short-range forces from the receptor in such a case. All information about binding events is displayed separately for each specified binding criterion.[19]

The output PQR and DCD files are needed for viewing the trajectory and can be loaded together into any program which supports these file formats, for example, VMD[32] and Pymol.[33] The PQR contains the initial system configuration, that is, the starting positions of all ligand replicates; it does not contain the receptor coordinates. The DCD file containing the simulation frames should be loaded into this PQR file. Users will see trajectories of all ligand replicates displayed at once. Individual trajectories can be viewed by selecting the appropriate representation in the chosen software. The receptor PQR or PDB used as the simulation input should be loaded separately to view ligand–receptor interactions. Trajectory visualization tips and analysis are given in the user manual at http://chemcha-gpu0.ucr.edu/geombd3/.

Simulation algorithms are parallelized with the Intel compiler's Cilk+ extension. GeomBD3 performance depends on the system, ligand, and number of ligand replicates. An example of the HIVp-xk263 system is included to show the efficiency of the program. The system is a box having HIVp (3128 atoms) placed at the bottom and 100 replicas of xk263 (84 atoms each replica) starting as a plane from the top. Simulations were run on 24 CPU cores of Intel(R) Xeon(R) CPU E5-2620 v2 with a step rate of 2308.3 steps/s. Considering each step is of 50 fs, a 7 μs simulation of this system takes around 17 h.

Restarting GeomBD3 Simulations.

There are two additional output files with the extensions .t and .crd. The former contains the restart times and binding information, while the later contains the coordinates for all ligand replicates. These files can be used to restart a simulation by restoring the coordinates and trajectory lifetimes of all ligand replicates as well as the statistics for all binding criteria. This provides the option of restarting a simulation at any point.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

EXAMPLE APPLICATIONS

Case 1: Ligand Association to a Surface-Bound Protein under Diffusive Flux.

In one application, GeomBD3 was used to study the effect of ligand flow and ambient surface properties. The simulation consisted of surface-bound HIV protease (HIVp) as the receptor and xk263 as the ligand. We analyzed the association time and probability while varying the flow rate of xk263, the surface properties, and HIVp's orientation with respect to the surface. We also examined the effect of HIVp's accessible surface for xk263 by allowing it to bind nonspecifically to the protein interface.

The simulation cell was a rectangular prism containing HIVp bound to a self-assembled monolayer (SAM) surface (open conformation) and xk263 (PDB ID: 1HVR). The HIVp used in the model is an open conformer derived from a MD simulation of a closed conformer (PDB ID: 1HHP). The SAM surface with CH_3 terminal groups provides a hydrophobic surface environment. The ligand replicas started to diffuse from an *yz*-aligned plane at the $-x$ direction boundary with a drift in the $+x$ direction (Figure 2). The flux effect was introduced by boosting the force value with the biasForce in the x component of the overdamped Langevin equation only. Boundaries in the y direction were periodic, and a termination boundary was used $+x$ direction. We considered two orientations for HIVp and seven flux rates for xk263 leading to 56 different system configurations. We analyzed diffusion coefficients, xk263 interactions with the SAM surface, and the distribution of xk263 with respect to the flow.

We found that the weak vdW attractions of the surface with xk263 increases the binding time in the absence of diffusion flux. With the increase in flux, the association time decreases even in the presence of a SAM surface. The simulated average association time of xk263 with HIVp in the presence of SAM but no flow is 1.491 μs. The remaining association time data for each system according to flux can be found in the original research publication.[21] With the increase in ligand flux, the concentration in the middle region of the system increases and leads to lessened ligand–surface interactions. We have also found that placing the HIVp 20 Å above the surface made the bottom of HIVp accessible for xk263 interaction. This exposed an attractive nonspecific association site, increasing association time by 10%–20% in the static environment.

Case 2: DNA Hybridization in a Biosensor.

GeomBD3 is also useful for simulations of large or complex engineered systems that span hundreds of nanometers. As a second GeomBD3 application, we present previously published data[19,20] from a simulation of DNA hybridization (binding) taking place in a biosensor (Figure 3). The sensor consisted of ssDNA probes inserted into a self-assembled monolayer (SAM) of functionalized alkanethiol molecules. With GeomBD3, it was possible to rapidly study the sensor under different probe surface densities and equipped with different functional groups on the self-assembled monolayer (SAM). These simulations used one or more ssDNA probes inserted into a slab of SAM as the receptor and a ssDNA molecule complementary to the probe strands as the ligand. The simulation space was bounded by a cuboid box with the biosensor placed at its bottom. Periodic boundary

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conditions were used in the x and y dimensions and a hard wall at the top. The x and y dimensions of the box were varied to encompass different amounts of SAM, yielding three different probe surface densities. Additionally, hybridization with clusters of five tightly packed probes was also studied on the $-CH_3$ and $-COO^-$ SAMs. The probe density was varied while also varying SAM functional groups between −OH, $-COO^-$, and $-CH_3$, yielding 15 different system permutations in total.

Varying the surface properties by changing SAM functional groups can dramatically alter DNA–surface interactions and hybridization rates. In some cases, DNA adsorbs and diffuses along the SAM two dimensionally, which may slow down or speed up hybridization depending on the strength of the interaction. This surface-mediated hybrization is defined as two dimensional, whereas diffusion directly to a probe through the solvent is defined as three-dimensional hybridization. Using the included script *Pathfinder.cc*, we analyzed the DCD trajectories to find the fraction of hybridization taking place two dimensionally through a surface-mediated mechanism, which we label as γ. Our results indicated that when surface interactions are strong, two-dimensional hybridization was significantly slower than the three-dimensional hybridization.

In the biosensor modeled with a cluster of probes (Figure 3), each probe (labeled 1–5; 5 at center) represented a possible binding site. GeomBD3 accepts multiple binding criteria and can independently track binding data for each criterion. We quantified the hybridization time and fraction of hybridization taking place at each probe to understand how tight probe packing may influence accessibility. The results are presented in Table S1, and they indicate that the center probes are nearly inaccessible when the spacing between neighbors is less than or equal to the target DNA length. The reversal of hybridization efficiency at high densities observed experimentally agrees with our simulation results.[34]

Cases 3 and 4.

Two more application examples can be found in the Supporting Information. Case 3 models multiple enzymes from a biocatalytic cascade in a dilute or crowded environment, and case 4 examines the rate of protein–ligand bindings using an acetylcholinesterase–acetylcholine (AchE-ach) system.

CONCLUSIONS

GeomBD3 provides researchers with a robust Brownian dynamics software package for modeling molecular association under a wide range of macromolecular arrangements and environments. The program uses rigid all-atom molecular models of a diffusing ligand and stationary grid-based representation of one or more receptors or crowder molecules. Molecular recognition is simulated by diffusing a molecule according to overdamped Langevin dynamics and evaluating user-defined binding criteria to track binding events. The program is capable of quickly simulating binding kinetics for drug–protein systems as well as identifying binding pathways and nonspecific adsorption sites. GeomBD3 can also elucidate the effects of specific system properties on molecular recognition to reveal phenomena unseen in experiments or to help interpret results. We demonstrate this in our application of the program to a nucleic acid biosensor with varied probe surface density

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and surface properties. This also demonstrates the applicability of the program to extremely large, nonstandard engineered systems for which the program's parameter files can be easily customized. An online interface for molecule parametrization and grid construction is currently being planned. Thus, the GeomBD3 package can efficiently meet the needs of researchers in a wide range of disciplines as both an explanatory and predictive tool.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This study was supported by the U.S. National Science Foundation (MCB-1932984), U.S. National Institutes of Health (GM-109045), and NSF national supercomputer centers (TG-CHE130009).

DATA AND SOFTWARE AVAILABILITY

All software is open source and is made available on GitHub, which can be accessed from our group website at http://chemcha-gpu0.ucr.edu/software/ and a method toolbox KBbox at https://kbbox.h-its.org/toolbox/methods/molecular-simulation/geombd/. The GeomBD3 page of the site also contains a helpful user manual, installation tips, example input files, and bash scripts to streamline system setup at http://chemcha-gpu0.ucr.edu/geombd3/.

REFERENCES

(1). Berg OG; von Hippel PH Diffusion-controlled macromolecular interactions. Annual review of biophysics and biophysical chemistry 1985, 14 (1), 131–158.

(2). Mereghetti P; Kokh D; McCammon JA; Wade RC Diffusion and Association Processes in Biological Systems: Theory, Computation and Experiment; Springer, 2011.

(3). Schreiber G; Haran G; Zhou H-X Fundamental aspects of protein-protein association kinetics. Chem. Rev 2009, 109 (3), 839–860. [PubMed: 19196002]

(4). Gao Y; Roberts CC; Toop A; Chang C. e. A.; Wheeldon I Mechanisms of enhanced catalysis in enzyme-DNA nanostructures revealed through molecular simulations and experimental analysis. ChemBioChem. 2016, 17 (15), 1430–1436. [PubMed: 27173175]

(5). Bellucci L; Delfino F; Tozzini V In silico design, building and gas adsorption of nano-porous graphene scaffolds. Nanotechnology 2021, 32 (4), 045704. [PubMed: 33017808]

(6). Huber GA; McCammon JA Brownian Dynamics Simulations of Biological Molecules. Trends Chem. 2019, 1 (8), 727–738. [PubMed: 32309795]

(7). Roberts CC; Chang C.-e. A. Modeling of enhanced catalysis in multienzyme nanostructures: effect of molecular scaffolds, spatial organization, and concentration. J. Chem. Theory Comput 2015, 11 (1), 286–292. [PubMed: 26574226]

(8). Bruce NJ; Ganotra GK; Richter S; Wade RC KBbox: A toolbox of computational methods for studying the kinetics of molecular binding. J. Chem. Inf. Model 2019, 59 (9), 3630–3634. [PubMed: 31381336]

(9). Northrup S; Laughner T; Stevenson G MacroDox Macromolecular Simulation Program; Department of Chemistry, Tennessee Technological University, Cookeville, TN, 1999.

(10). Madura JD; Briggs JM; Wade RC; Davis ME; Luty BA; Ilin A; Antosiewicz J; Gilson MK; Bagheri B; Scott LR; McCammon JA Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comput. Phys. Commun 1995, 91 (1–3), 57–95.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(11). Huber GA; McCammon JA Browndye: a software package for Brownian dynamics. Comput. Phys. Commun 2010, 181 (11), 1896–1905. [PubMed: 21132109]

(12). Martinez M; Bruce NJ; Romanowska J; Kokh DB; Ozboyaci M; Yu X; Öztürk MA; Richter S; Wade RC SDA 7: A modular and parallel implementation of the simulation of diffusional association software. Journal of computational chemistry 2015, 36 (21), 1631–1645. [PubMed: 26123630]

(13). Długosz M; Zieliński P; Trylska J Brownian dynamics simulations on CPU and GPU with BD_BOX. Journal of computational chemistry 2011, 32 (12), 2734–2744. [PubMed: 21638295]

(14). De Biase PM; Markosyan S; Noskov S BROMOC suite: Monte Carlo/Brownian dynamics suite for studies of ion permeation and DNA transport in biological and artificial pores with effective potentials. Journal of computational chemistry 2015, 36 (4), 264–271. [PubMed: 25503688]

(15). Andrews SS Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. Bioinformatics 2017, 33 (5), 710–717. [PubMed: 28365760]

(16). Votapka LW; Jagger BR; Heyneman AL; Amaro RE SEEKR: simulation enabled estimation of kinetic rates, a computational tool to estimate molecular kinetics and its application to trypsin-benzamidine binding. J. Phys. Chem. B 2017, 121 (15), 3597–3606. [PubMed: 28191969]

(17). Saadat A; Khomami B Matrix-free Brownian dynamics simulation technique for semidilute polymeric solutions. Phys. Rev. E 2015, 92 (3), 033307.

(18). Schöneberg J; Noé F ReaDDy-a software for particle-based reaction-diffusion dynamics in crowded cellular environments. PloS one 2013, 8 (9), No. e74261. [PubMed: 24040218]

(19). Cholko T, Simulating Chemical Processes from Brownian Diffusion to Binding Thermodynamics, in Chemistry; UC Riverside, 2021; ProQuest ID: Cholko_ucr_0032D_14455; Merritt ID: ark:/13030/m5f82kqc. https://escholarship.org/uc/item/4rh1c4cn (accessed May 2022).

(20). Cholko T; Chang C.-e. A. Modeling Effects of Surface Properties and Probe Density for Nanoscale Biosensor Design: A Case Study of DNA Hybridization near Surfaces. J. Phys. Chem. B 2021, 125 (7), 1746–1754. [PubMed: 33591751]

(21). Kaushik S; Chang C-EA Molecular Mechanics Study of Flow and Surface Influence in Ligand-Protein Association. Front. Mol. Biosci 2021, 8, 8.

(22). Cholko T; Kaushik S; Chang C.-e.A. Dynamics and molecular interactions of single-stranded DNA in nucleic acid biosensors with varied surface properties. Phys. Chem. Chem. Phys 2019, 21 (29), 16367–16380. [PubMed: 31309941]

(23). Gong P; Levicky R DNA surface hybridization regimes. Proc. Natl. Acad. Sci. U. S. A 2008, 105 (14), 5301–5306. [PubMed: 18381819]

(24). English NJ; MacElroy J Atomistic simulations of liquid water using Lekner electrostatics. Mol. Phys 2002, 100 (23), 3753–3769.

(25). Northrup SH; Allison SA; McCammon JA Brownian dynamics simulation of diffusion-influenced bimolecular reactions. J. Chem. Phys 1984, 80 (4), 1517–1524.

(26). Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE The protein data bank. Nucleic acids research 2000, 28 (1), 235–242. [PubMed: 10592235]

(27). Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput 2015, 11 (8), 3696–3713. [PubMed: 26574453]

(28). Wang J; Wolf RM; Caldwell JW; Kollman PA; Case DA Development and testing of a general amber force field. Journal of computational chemistry 2004, 25 (9), 1157–1174. [PubMed: 15116359]

(29). Hassan SA; Mehler EL A critical analysis of continuum electrostatics: the screened Coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins. Proteins: Struct., Funct., Bioinf 2002, 47 (1), 45–61.

(30). Huey R; Morris GM; Olson AJ; Goodsell DS A semiempirical free energy force field with charge-based desolvation. Journal of computational chemistry 2007, 28 (6), 1145–1152. [PubMed: 17274016]

(31). Stouten PF; Frömmel C; Nakamura H; Sander C An effective solvation term based on atomic occupancies for use in protein simulations. Mol. Simul 1993, 10 (2–6), 97–120.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(32). Humphrey W; Dalke A; Schulten K VMD: visual molecular dynamics. J. Mol. Graphics 1996, 14 (1), 33–38.

(33). Schrodinger L The PyMOL Molecular Graphics System, Version, 2.0; Schrödinger, LLC, 2010. https://pymol.org/2/support.html (accessed May 2022).

(34). Peterson AW; Heaton RJ; Georgiadis RM The effect of surface probe density on DNA hybridization. Nucleic acids research 2001, 29 (24), 5163–5168. [PubMed: 11812850]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

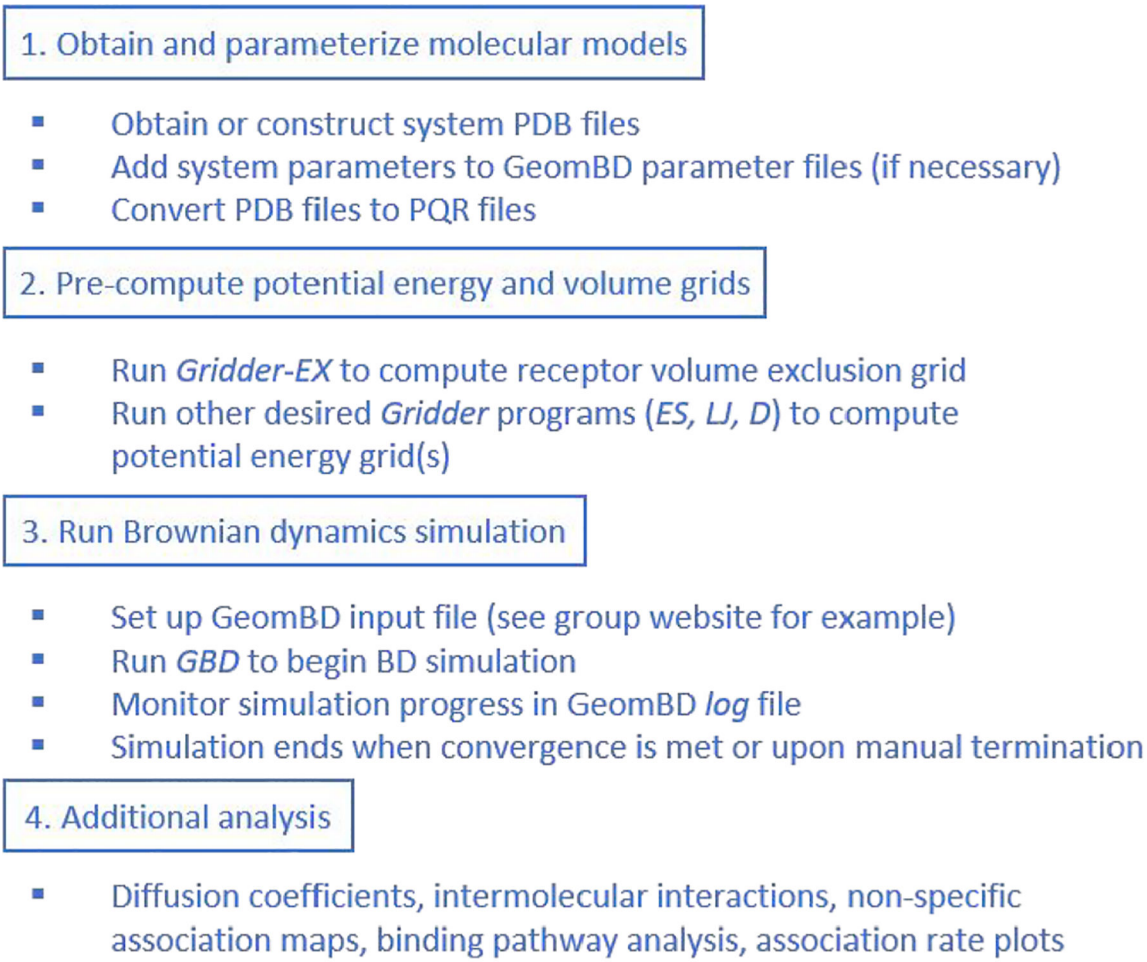


Figure 1.
Diagram of GeomBD3 workflow.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

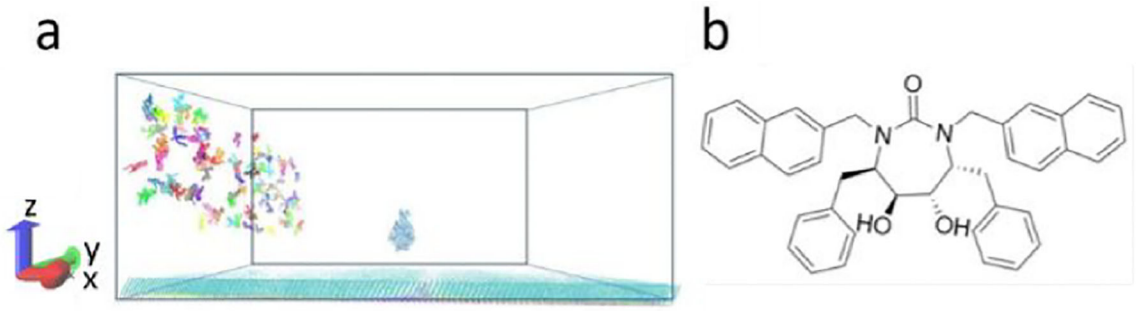


Figure 2.

(a) HIVp placed 20 Å above the SAM parallel to the *yz*-plane. (b) Structure of xk263 ligand.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

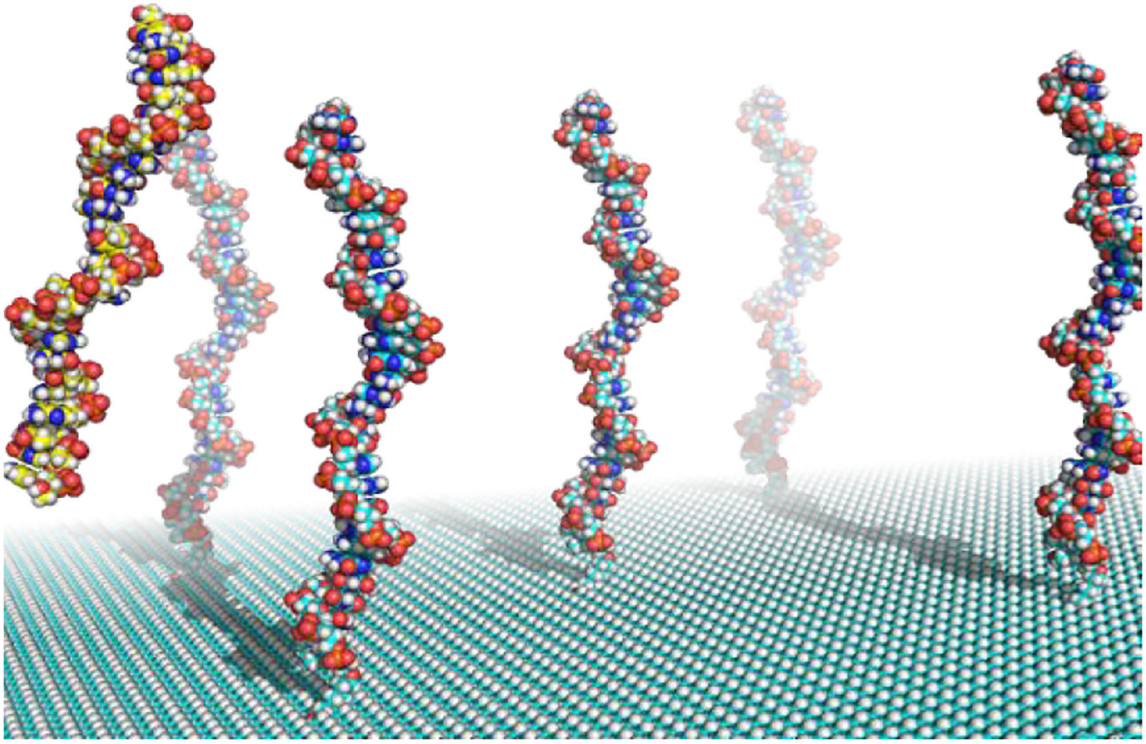


Figure 3.
GeomBD3 molecular model of a DNA biosensor with multiple probes (blue strands) and the target DNA (yellow strand) which binds to the probes.