

Combined Monte Carlo / Torsion-Angle Molecular Dynamics for Ensemble Modeling of Proteins, Nucleic Acids and Carbohydrates

Weihong Zhang^a, Steven C. Howell^b, David W. Wright^c, Andrew Heindel^{d,1},
Xiangyun Qiu^e, Jianhan Chen^{a,*}, Joseph E. Curtis^{b,**}

^a*Kansas State University, Manhattan, KS, USA*

^b*NIST Center for Neutron Research, 100 Bureau Drive, Gaithersburg, MD, USA*

^c*Centre for Computational Science, Department of Chemistry, University College London,
10 Gordon St., London, UK*

^d*James Madison University, Department of Chemistry & Biochemistry, Harrisonburg, VA,
USA*

^e*Department of Physics, The George Washington University, Washington, DC, USA*

Abstract

We describe a general method to use Monte Carlo simulation followed by torsion-angle molecular dynamics simulations to create ensembles of structures to model a wide variety of soft-matter biological systems. Our particular emphasis is focused on modeling low-resolution small-angle scattering and reflectivity structural data. We provide examples of this method applied to HIV-1 Gag protein and derived fragment proteins, TraI protein, linear B-DNA, a nucleosome core particle, and a glycosylated monoclonal antibody. This procedure will enable a large community of researchers to model low-resolution experimental data with greater accuracy by using robust physics based simulation and sampling methods which are a significant improvement over traditional methods used to interpret such data.

Keywords: Monte Carlo, Torsion Angle Molecular Dynamics,
Structural Biology, Small-angle scattering, SANS, SAXS

*Corresponding author

**Corresponding author

Email addresses: jianhanc@ksu.edu (Jianhan Chen), joseph.curtis@nist.gov (Joseph E. Curtis)

¹current address: Drug Product Formulation Technologies, Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320, United States

1. Introduction

Molecular simulation encompasses a powerful and comprehensive set of methods to provide atomistic insight into a diverse set of material, chemical, and biochemical systems. The type of problem that we aim to address is the modeling of low-resolution experimental structural biology data from small-angle scattering (SAS) using either neutrons (SANS) or X-rays (SAXS). Many structural biology problems involve intrinsically disordered proteins and nucleic acids and flexible multi-domain complexes that often adopt a large range of conformations in solution. The use of molecular dynamics simulation to study such systems is often a challenging if not intractable task as many systems sample broad time and length scales that are beyond the current capabilities of algorithms and hardware. Thus a large number of structural biology problems exist where low-resolution experimental data are often modeled using analytical [1, 2] and dummy-ball models [3, 4]. While these methods are fast, robust, widely used and are commensurate with the idea that a low-resolution model adequately reflects low-resolution data, they are fundamentally limited by not using atomistic information.

Chemical bonding, topology and interactions are key factors to accurately model experimental data. In recent years, several groups have applied existing or developed new methods to model SAS experimental data of biological systems [5, 6, 7, 8, 9]. To different degrees, these methods incorporate atomistic aspects into the modeling process. We have developed a set of algorithms to perform Monte Carlo (MC) sampling of backbone-dihedral angles that can quickly generate ensembles of proteins and/or single-stranded nucleic acids to model low-resolution scattering and reflectivity data [10, 11]. Recently we have implemented a MC algorithm to simulate B-DNA that utilizes an intermediate worm-like chain coarse grain representation that is mapped back to all-atom coordinates [12]. These all-atom models are created and defined by mature force

fields, are structurally robust and directly available for further advanced simulation methods such as accelerated molecular dynamics [13], replica-exchange [14] and reverse MC methods [15] as needed. While powerful, advanced MD simulation methods can have limited utility based on the size of the system and/or available computational resources.

MC sampling is fast and numerically inexpensive since the number of degrees of freedom are dramatically reduced. Typical systems are well sampled in minutes using commodity single CPU hardware. Such algorithms inherently sample a rugged energy landscape and thus there are limitations to these methods. Additionally, sequential moves can be highly correlated, and the basic move-set is not designed for internal or concerted moves, such as internal loops between two rigid domains. Nonetheless this approach has been successfully used to model a large number of important structural biology problems [16, 17, 18, 19, 20, 21]. The combination of MC sampling with advanced simulation methods is largely not used, thus there remains a need to develop additional sampling procedures to improve the quality and robustness of such simulations. These procedures' will help realize the full-potential of fast atomistic modeling of scattering data.

In this report we discuss the combination of MC sampling with torsion-angle molecular dynamics (TAMD) simulations [22] that overcome some of the limitations of the rugged MC sampling methods. TAMD samples molecular configurations in torsion space and allows convenient specification of rigid domains and flexible degrees of freedoms consistent with sampling of MC trajectories. Using TAMD allows large time steps when propagating the equations of motion [22]. In TAMD, the molecule is represented by a branched tree structure consists of clusters of rigid-body atoms linked by hinges. Importantly, forces arising from the traditional Cartesian force field are projected along the internal coordinates in TAMD [23], thus allowing one to benefit from the significant improvement and accuracy of physics-based force fields. Using TAMD greatly increases sampling efficiency over traditional MD simulations and is widely used in NMR structure refinement [24, 25].

60 Details of subsampling of MC ensembles and the performance of the improved algorithm are explored with several representative examples covering systems containing not only protein but also nucleic acid and carbohydrate components. In the development of the algorithm several implicit solvent models and constrained simulation variables greatly affected the efficacy of the resulting
65 algorithm. The combination of MC and TAMD will enable the generation of robust models to interpret SAS, neutron and X-ray reflectivity, NMR, electron microscopy and other experimental data.

2. Methods

Seven systems are used in this study to illustrate the utility of the combined
70 MC/TAMD modeling protocol. Four protein only systems of differing complexity are studied: full-length human immunodeficiency virus type 1 (HIV-1) Gag protein, two truncated Gag protein constructs (A-B and A-tail) and a TraI protein fragment. Two example nucleic acid systems are highlighted, a linear 60 bp B-DNA molecule and nucleosome core particle (NCP), the latter of which
75 combines protein and DNA elements. The application of the methodology to carbohydrate containing systems is represented using a glycosylated and truncated human monomer IgA1 antibody molecule (PTerm455). Systems were prepared for MC simulations using CHARMM27 [26, 27, 28, 29] and CHARMM36 force fields [30, 31, 32, 33]. A summary of model systems is shown in Table 1.

80 *Model proteins.* Three of the protein systems used in this study were derived from the full-length human immunodeficiency virus type 1 (HIV-1) Gag protein. Full-length HIV-1 Gag protein can be divided into five globular domains, specifically, MA domain (residues 1-122), N-terminal domain of CA (residues 144-276), C-terminal domain of CA (residues 282-353), p2 “spacer” (residues
85 374-377), and NC domain (residues 390-431). The construction of structural model of the full-length HIV-1 Gag protein has been previously described [16]. An A-B system (first two domains of full-length HIV-1 Gag protein, residues 1-276) and an A-tail system (first domain with a flexible linker of full-length

HIV-1 Gag protein, residues 1-143) were built from the full-length model. A
90 fourth protein model taken from a fragment of TraI protein consisting of residues
381-858 was constructed as previously described [34]. This fragment contains
three globular domains spanning residues 381-573, residues 577-789 and residues
804-849, respectively connected by regions of flexible amino-acids. Initial mod-
els were energy minimized for 2000 steps following by 10 ps of vacuum MD
95 simulation using NAMD [35] and the CHARMM27 force field prior to use in the
MC simulations.

Model Linear B-DNA and NCP. A 60 bp linear DNA model was generated using
psfgen, a plug-in of VMD [36], and based on a model of a random sequence
generated by the 3D-DART DNA structure modeling server [37]. The initial
100 NCP model was generated using psfgen and based on the PDB-ID 1KX5 X-ray
structure of the NCP solved at 1.9 Å resolution [38]. Using NAMD with the
CHARMM36 force field, the DNA models were prepared for MC simulations
by independently performing 2000 energy minimization steps followed by 200
vacuum MD steps (0.2 ps) then another 2000 energy minimization steps. Note
105 that the CHARMM36 force field more accurately represents the experimentally
measured distribution of BI to BII DNA, a behavior not modeled well by the
CHARMM27 force field.

Model protein - carbohydrate system. The initial model structure was built using
coordinates from a Fc domain crystal structure (PDB: 1OW0) [39] and glycan
110 coordinates from a previous solution modelling study [40] as described in [41].
The composition of the biantennary N-glycan incorporated in the model was
(NeuAc)2(Gal)2(GlcNAc)2(Fuc)1(Man)3(GlcNAc)2. The PTerm455 structure
was prepared for simulation using the glycan reader component of CHARMM-
GUI [42, 43] using the CHARMM36 forcefield. The PTerm455 construct
115 contains the full antibody structure truncated at residue 455 in the heavy chain
(removing the flexible tailpiece region and a pair of N-linked glycans). The
remaining structure contains two N-linked glycans and the Fc domain. Cur-
rently there is no MC move-set available to sample carbohydrate degrees of

Table 1: Model systems, flexible residues and experimental radius of gyration (RGYR) values. Errors are reported as ± 1 standard deviation.

Name	Flexible Residues	RGYR (\AA)
HIV-1 Gag	123-143, 277-281, 354-373, 378-389, 408-412	34 ± 1 [16]
A-B (HIV-1 Gag 1-276)	123-143	29 ± 1 [16]
A-tail (HIV-1 Gag 1-143)	123-143	no data
TraI (381-858)	574-576, 790-803	36.9 ± 0.6 [34]
Linear B-DNA	chain A & B: 1-60	no data
NCP	chain A & B: 1-31, 117-147	no data
Pterm455	N-linked glycans	no data

freedom therefore MC simulations were not carried out and an energy mini-
120 minimized PTerm455 was used directly as input for TAMD simulations.

Monte Carlo simulations. MC trajectories were generated for each model pro-
tein using the ‘‘Monomer Monte Carlo’’ module of the SASSIE program [10].
Flexible residues were subjected to backbone dihedral angle sampling at 300 K.
The maximum single step rotation angle was set to 30 degrees and structures
125 with heavy atom overlap were discarded. For the full-length HIV-1 Gag, A-B,
A-tail and TraI systems, a total number of 150000 trial attempts were performed
and generated 126196, 136281, 142311 and 131014 structures, respectively. Ac-
cepted structures from MC simulations were energy minimized. MC trajectories
containing 71479 and 8166 structures were respectively generated for the linear
130 B-DNA and NCP using the MC simulation algorithm for B-DNA described by
Howell et al. [12]. For MC simulation of linear B-DNA and NCP flexible re-
gions are listed in Table 1, the last 30 base pairs on each end of the wrapped
DNA were designated as flexible. For both DNA containing simulations, the

maximum single rotation angle was set to 10 degrees and structures with heavy
135 atom overlap were discarded.

Subsampling of MC ensembles. Spatially representative configurations of each
model system were selected from the structural ensembles generated by MC
simulation module for subsequent TAMD simulations. The selection of the
representative configurations involved subsampling in regards to the relative
140 positions of different globular-like domains by clustering similar configurations
on a geometric grid. The spatial coordinate system was divided into voxels,
where the size was proportional to the size of the globular-like domains to ensure
a proper fraction of occupancy in the voxels. Structures with the centers of
mass of individual domains occupying the same voxels were clustered, and one
145 representative structure in each cluster was extracted and added to the sub-
ensemble. This resulted in a total of 200, 175, 107, and 155 representative
structures sampled from the initial MC ensembles for full-length HIV-1 Gag
protein, A-B system, A-tail system and TraI protein respectively. For linear
B-DNA and NCP, a total of 110 and 178 structures were subsampled from the
150 MC ensembles, respectively.

Torsion angle molecular dynamics. TAMD simulations were carried out for the
subsampled ensembles for each model protein using the TAMD module [22] im-
plemented in CHARMM [44, 45]. For each protein, the globular domains were
clustered as rigid bodies. Rigid bodies were connected by flexible regions that
155 sample all torsional degrees of freedom in each flexible region. 2500 steps of
steepest descent energy minimization followed by 2500 steps of adopted basis
Newton-Raphson energy minimization method were performed before TAMD
simulations. TAMD simulations were then carried out for all representative
structures with a length of 500 ps and time step of 2 fs. Distance dependent
160 dielectric (RDIE) was used to describe electrostatic interactions. Structures
were saved every 1 ps. For comparison among different implicit solvent models,
additional TAMD simulations of four representative structures of A-B system
and TraI protein were carried out in the generalized Born with smooth switch-

ing (GBSW) [46, 47] and solvent accessible surface area (SASA) [48] models.
165 GBSW in particular is one of the latest generation of implicit solvent mod-
els that can provide a better description of solvent dependent molecular in-
teractions and allow more realistic descriptions of biomolecular conformational
equilibria [49]. It was observed that DNA molecules are unstable in the implicit
solvent used in the TAMD simulations. To maintain structural integrity of the
170 DNA molecules in TAMD simulations, a harmonic restraint with a force con-
stant of $0.1 \text{ kcal/mol/\AA}^2$ on the backbone atoms of the DNA molecules was used.
For each DNA containing subsampled structure, a 10 ps TAMD simulation was
performed to mainly relax the structure and sample local fluctuations. TAMD
simulations of PTerm455 were carried out with the protein segment remained
175 fixed whilst the carbohydrate was free to move.

Data analysis. Density plots to visually evaluate conformational coverage from
MC simulations, subsampling, and TAMD trajectories were calculated using
“Density Plot” module in SASSIE [10] with grid spacing of 6.0 \AA . The analysis
of Ramachandran plots, RMSD, and radius of gyration were carried out using
180 CHARMM and additional analysis scripts.

Implementation. Scripts to enable the automatic generation of CHARMM in-
put files for proteins, nucleic acids, carbohydrates and their complexes have been
incorporated into the SASSIE program [10] as a separate “Torsion Angle Molec-
ular Dynamics” module. Thus the capability to perform sequential MC and
185 TAMD simulations is accessible to a wide variety of structural biology problems
in a single software suite. Note that a valid license to use CHARMM is required
to run TAMD (http://charmm.chemistry.harvard.edu/charmm_lite.php)
and an automated license verification scheme is incorporated with SASSIE-web
(<https://sassie-web.chem.utk.edu/sassie2>).

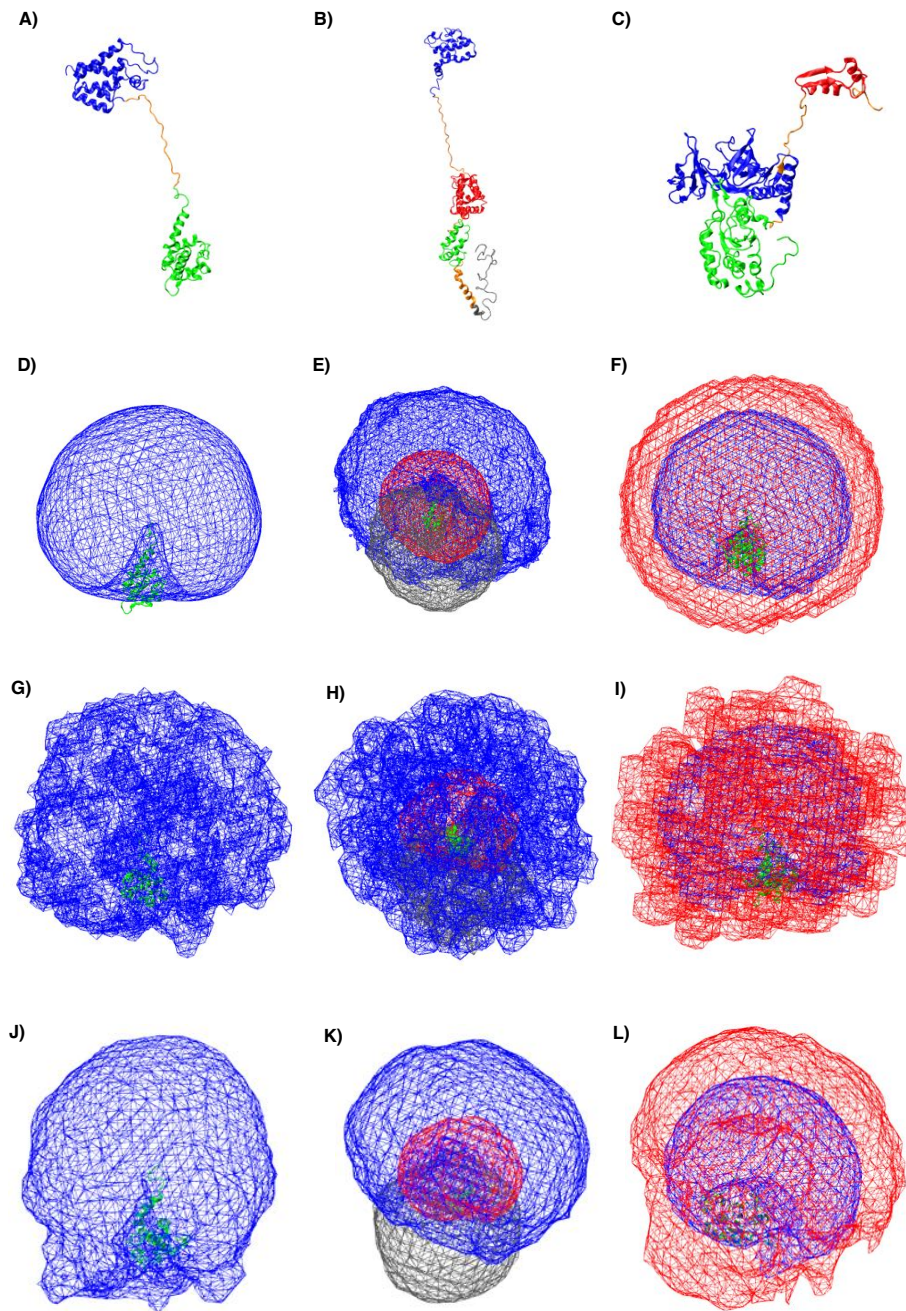


Figure 1: Cartoon representation of A-B system, HIV-1 Gag protein and TraI protein (A-C), and their density plots of conformational coverage from MC simulation (D-F), from subsampled representative structures (G-I) and after TAMD simulations (J-K). A), D), G) and J) A-B system, MA domain is shown in green, NTD domain is shown in blue; B), E), H) and K) the full-length HIV-1 Gag protein the MA⁹ domain is shown in blue, the NTD domain is shown in red, the CTD domain is shown in green and the NC domain is shown in grey. B), E) and H) full-length HIV-1 Gag protein is aligned with CTD domain. C), F), I) and L) TraI protein, the first globular domain (residues 381 to 573) is shown in green, the second globular domain (residues 577 to 789) is shown in blue and the third globular domain (residues 804 to 849) is shown in red.

190 **3. Results**

3.1. *Simulations of proteins*

This section is divided up into several subsections to explore the conditions required to apply MC simulation with TAMD to adequately sample configuration space for multi-domain protein systems. This involves how one can sub-sample the MC trajectories to carry out a limited yet representative set of TAMD simulations and the role of implicit solvent models on how the TAMD simulations compared to experimental data.

Monte Carlo and subsampling. For intrinsically disordered and flexible protein systems MC sampling using backbone torsion degrees of freedom is an efficient way to generate a wide range of structures that cover the large conformational space of these molecules. MC sampling alone is robust way to evaluate structural models to compare to low resolution SAS experimental data. The example systems, A-B, full-length HIV-1 Gag and TraI colored by domains Figure 1 (A-C) covered a wide-range of conformational space in their MC simulations as depicted in the density plots shown in Figure 1 (D-F). Since performing advanced MD simulations on the entire MC ensemble is generally not possible and likely unnecessary, a simple clustering algorithm was devised and implemented to extract a tractable and representative subsample from each MC trajectory. As an example of the characteristics of the the details of subsampling the A-B system is shown in Figure 2 and for A-B, HIV-1 Gag and TraI as density plots in Figure 1 (G-I). For the A-B system the average of the Cartesian dimensions of the two globular domains were 43 and 39 Å respectively, thus using voxel size \sim one-third of the globular domain dimensions yielded 175 structures in the sub-sample.

215 The voxel size one should use for other systems to account for asymmetry of domain sizes and orientations that may have to be taken into account in specific situations to adequately extract representative subsamples is beyond the scope of this report. Subsampling using 15 Å voxels was adequate for the illustrative purpose for the examples in this study. One quantitative measure of

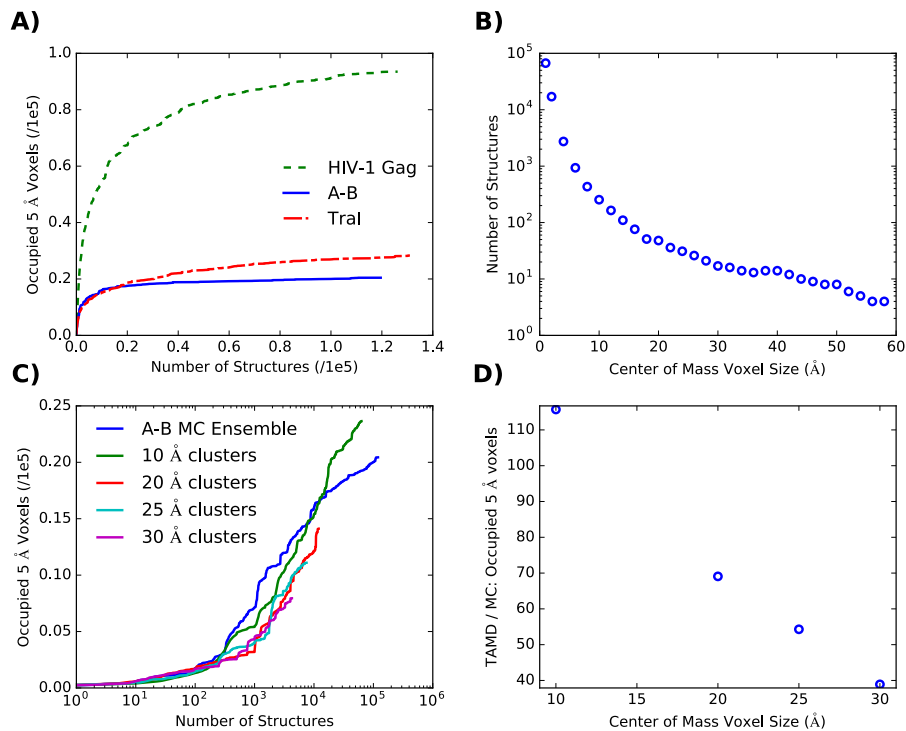


Figure 2: Representative characteristics of subsampling MC ensembles. A) The total number of unique 5 Å voxels occupied by an alpha carbon as a function of the number of structures in each MC ensemble. B) The number of center of mass voxels occupied by structures in the A-B MC Ensemble as a function of voxel size. C) Comparison between the total number of unique 5 Å voxels occupied by an alpha carbon for the A-B MC Ensemble and the ensembles created using TAMD simulations on various sizes center of mass voxels. D) TAMD/MC ratio of the total number of unique 5 Å voxels occupied by an alpha carbon as a function of the center of mass voxel size.

220 conformational space coverage is to enumerate the number of occupied voxels in
three-dimensional space as shown in Figure 2A. Additionally, for low-resolution
methods such as SAS which one can calculate theoretical data from atomic co-
ordinates [50], convergence of theoretical SAS profiles can be used in concert
with spatial convergence to obtain a objective evaluation of the extent of confor-
225 mational space coverage [12]. Although not used in this study, an advantage of
using SAS convergence is that structural configurations that are not discernible
by SAS are deemed redundant thus reducing the number of structures one needs
to consider.

TAMD enhances local sampling. TAMD propagates the equations of motion
230 directly in torsion space. It does not sample bond or angle degrees of freedom
removing the limitation imposed by the short timescale of bond vibrations and
allowing the use of much larger time steps. The process of subsampling the
original MC generated sampling of A-B, HIV-1 Gag and TraI results in rougher
spatial coverage for all systems (as can be seen by comparing 1(J-L) to 1(D-F)).
235 The domain sampling is much smoother following local enhanced sampling by
TAMD (Figure 1(G-I)), thus much of the spatial coverage in the original MC
trajectory is recovered in this process. This is highlighted in quantitative detail
for the A-B system in Figure 2. The spatial coverage of individual domains is
much smoother than that observed for the subsampled ensembles (Figure 1(G-
240 I)) after local enhanced sampling by TAMD thus recovering much of the spatial
coverage in the original MC trajectory (Figure 1(D-F)). Backbone torsion angle
MC samples configurations on a rugged energy surface whereby local intra-
residue bond and angle degrees of freedom are not relaxed completely even
following an energy minimization step. A short MC trajectory was carried out
245 where only residue 124 of the flexible region of the A-B system was sampled
and energy minimized. 136281 structures from MC were subsampled into 175
representative structures and simulated using TAMD for 1 ns. In Figure 3 the
results indicate that MC on its own generally samples distinct highly populated
states. A total of 17500 structures from the TAMD simulations were used

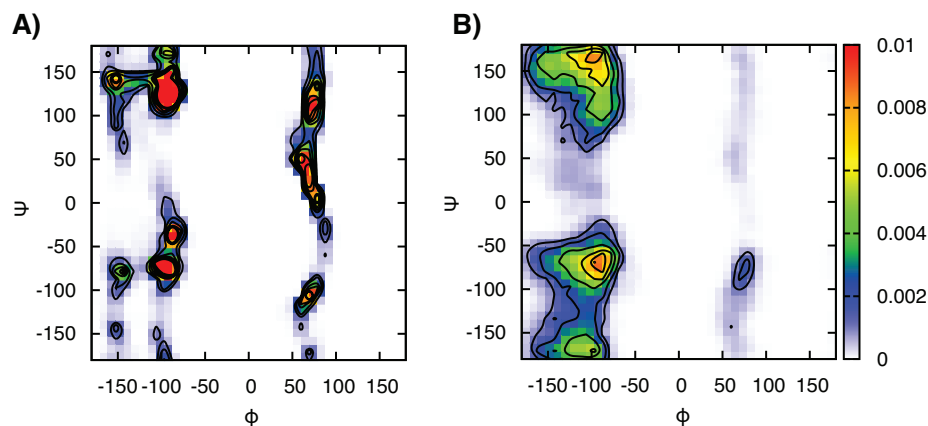


Figure 3: Ramachandran plot for residue 124 in A-B system. A) Dihedral space sampled by 17500 MC configurations generated using only residue 124 as the flexible torsion. B) Dihedral space sampled by MC combined with TAMD simulations. 175 representative configurations were subsampled from 136281 MC configurations and used to carry out independent 1 ns of TAMD simulations. A total of 17500 structures were used in calculating Ramachandran plot shown in B.

250 to calculate the Ramachandran plot shown in Figure 3 B. TAMD smoothly sampled the residue configurations in torsion angle space. The torsion angle populations sampled by TAMD are more physically realistic, with the two most populated states found at the alpha helical and extended torsion angle regions as expected. Note that at this length scale there are no discernible differences in the SAS profiles between the MC and TAMD configurations (data not shown).
 255

One way to evaluate the range of conformational space sampled in TAMD simulations is to calculate the RMSD with respect to the starting structures. Sampling the flexible region in the A-B system for 500 ps on average resulted in structures that with RMSD of 17 Å from the initial structure as shown in
 260 Figure 4. For the TraI protein fragment the average RMSD of the final structure after 500 ps simulation was approximately 10 Å. This suggests that even short TAMD simulations efficiently sample a large conformational space.

For more complex systems, such as full-length HIV-1 Gag protein, TAMD simulation is capable of sampling large domain rearrangements. The RMSD

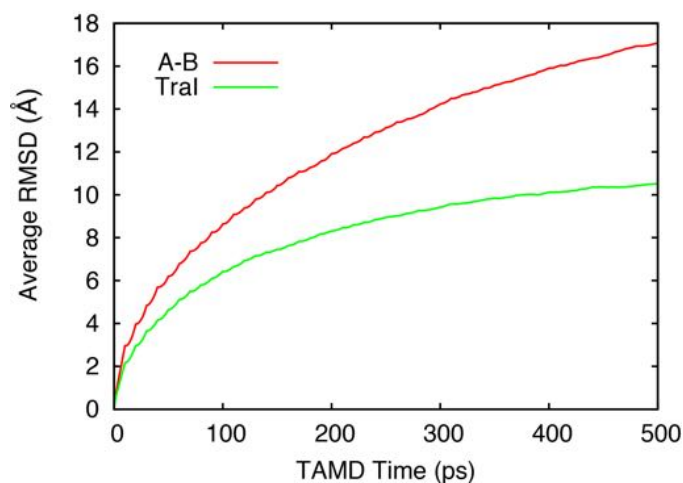


Figure 4: Red: Averaged backbone RMSD calculated from 175 independent TAMD simulations of A-B system. RMSD are calculated by aligning domain A. Green: Averaged backbone RMSD calculated from 155 independent TAMD simulations of TraI protein. RMSD are calculated by aligning the first domain (residue 1-192).

265 of the MA, NTD and NC domains calculated by aligning the CTD domain ranged from 20 Å to 80 Å (Figure S1). Importantly, TAMD simulation of the subensembles were able to refine structural features in several ways. Figure S2 demonstrates that the two domains in the A-B system remain relatively stable during the TAMD simulation while loop regions assume a more compact structure and reorient at the hinge ends. With two domains interacting with each other, TAMD is able to more rigorously sample the rearrangement of the flexible loop regions. Sampling of this cooperative movement of the loop region is limited in the backbone torsion MC method. Compared with dummy ball models, TAMD simulations are driven by a physics based force field and are thus capable of more accurately describe the interactions between domains. TAMD simulation also removed some of the steric clashes introduced by MC sampling, along with the optimization of the relative orientations of domains and the interaction interface as shown in Figure S3. Even though the algorithm mainly employs short TAMD simulations and focuses on local sampling, some large scale domain movement was observed (Figure S4).

270

275

280

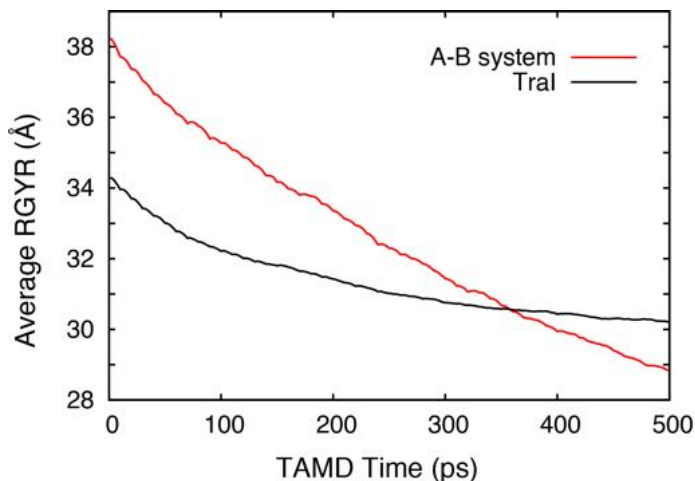


Figure 5: Averaged RGYR calculated from TAMD simulations of A-B system (red) and TraI protein (black).

Implicit solvent models. The quality of conformations sampled during TAMD simulations are determined by the force field employed. Efficiency considerations require deployment of efficient implicit solvent-based force fields, where the crucial solvent effects on biomolecule structures are captured by direct estimation of so-called solvation free energy. Only the molecule of interest need
 285 to be represented at the atomistic level with implicit solvent, thus reducing the system size dramatically. Distance dependent dielectric (RDIE) is a simple model to describe the solvent dependence of electrostatic interactions [51], yet it has been shown to be quite effective in docking and ranking of protein-ligand
 290 interactions [52]. In comparison, solvent accessible surface area (SASA) and generalized Born with smooth switching (GBSW) are two classes of more advanced implicit solvent models that can provide more realistic estimation of conformational dependence of solvation free energies [49]. Both models have demonstrated reasonable successes in folding simulations of small proteins, whereas
 295 they are more limited when applied to larger systems [49].

Using RDIE during TAMD simulations, both the A-B and TraI proteins tend to collapse during TAMD simulations, particularly with longer simulation times.

As shown in Figure 5, the radius of gyration (RGYR) gradually reduced from 38 to 29 and 34 to 30 for A-B system and TraI protein, respectively, during 500 ps TAMD runs. The value of RGYR for TraI is much lower than experimentally determined value of 36.9 Å as shown in Table 1.

Unfortunately, it does not appear that the tendency for the proteins to compact over the duration of the simulations could be effectively suppressed by the more advanced SASA and GBSW implicit solvent models. The structures of the A-B system in all three implicit solvent force fields collapsed, albeit to varying extents (Figure S5). The RGYRs generally remain at the same level within the first 100 ps, and quickly reduced to a plateau around 400 ps. TAMD simulations of TraI in implicit solvents showed less severe collapse, with an RGYR decrease of only about 5Å with values typically reaching a plateau after 400 ps (Figure S6). It should be noted that the TAMD simulation of TraI using all the implicit solvent models led to RGYR values less than that observed experimentally [34]. These observations reflect a common limitation of the current implicit solvent models, which energetically favor and over-stabilize compact state of proteins [53]. Several factors contribute to this bias, including the underlying protein force field, inadequate description of solvent screening of dispersion interactions and others [53]. We also note that the limitations associated with GB-class of implicit solvent models are shared by the corresponding Poisson-Boltzmann approaches. This will reduce the accuracy of solvent accessible surface area calculations thus affecting the estimation of the hydrophobic effect in domain-domain interactions. As such, we conclude that TAMD coupled with implicit solvent should be mainly used for refinement of local structural features and not for sampling large-scale conformational re-arrangements. Thus large scale rearrangements as noted in Figure S4 may entirely be due to deficiencies in the implicit solvent model and not a natural interaction predicted by TAMD. Specifically, we chose to perform short (100 ps) TAMD simulations to sample substantial conformational space without artificial collapse of the initial structures.

TAMD with RGYR restraints. One way to reduce the tendency for structures to collapse is to apply RGYR restraints on the initial structures in the TAMD simulations. The restraint force stabilizes the central moment of the selected atoms around their center of geometry. Figure 10 shows the RGYR of A-B system from three representative TAMD simulations both without and with RGYR restraints using a force constant of 20 kcal/mol/Å. The restrained TAMD simulations were extended up to 1 ns duration and the RGYR restraints effectively suppressed the artifactual protein collapse introduced by the implicit models. Although applying RGYR restraints prevents TAMD from sampling large scale conformational transitions, the sampling of local structural features and the refinement of the MC structures could be efficiently achieved.

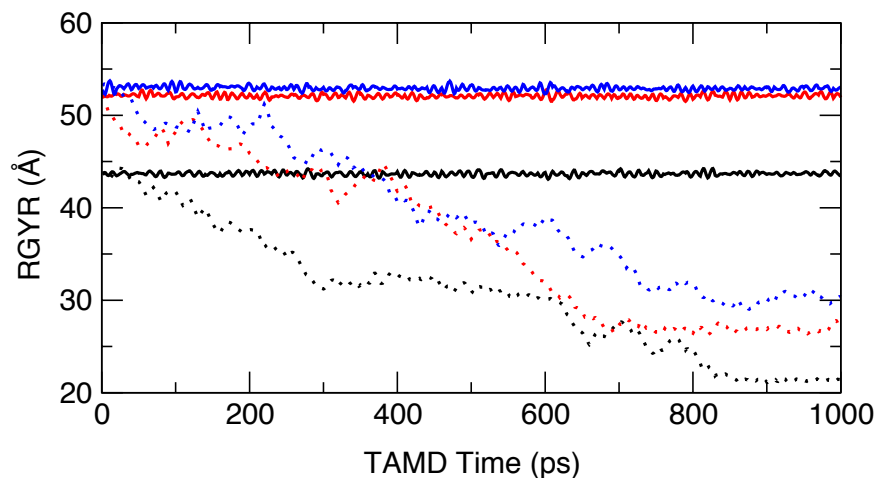


Figure 6: Radii of gyration of A-B system from three representative TAMD simulations with (solid lines) or without (dashed lines) RGYR restraints.

3.2. Simulations of Polypeptide Chains

While we have explored MC simulation followed by TAMD to sample large protein systems connected by flexible residues, the use of simple backbone torsion-angle MC sampling to model disordered chains in isolation is largely unexplored and there are many advanced and mature simulation and sampling

methods available that inherently capture the relevant states with greater accuracy [54, 55]. The simulation of chains attached at the termini of globular macromolecules may still benefit by the rugged MC simulation approach to provide rapid comparison of models to experimental scattering data. Many intrinsically disordered proteins contain regions at the terminus of globular segments that can be considered as disordered tails or polypeptide chains [56, 57].

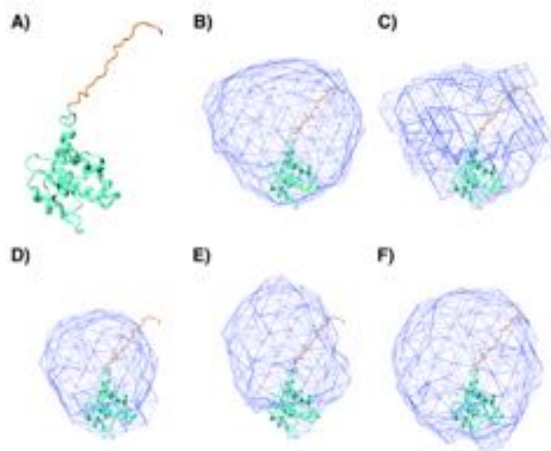


Figure 7: A-tail system and the density plot of different ensembles. A. Cartoon rendering of the A-tail system, extracted from the HIV-GAG protein (residue 1 to 143), tail region is flexible in the simulations and is shown in orange. B. Density plot of the Monte Carlo ensemble. C. Density plot of the sub-sampled ensemble. D. Density plot of the TAMD ensemble without constraints. E. Density plot of the TAMD ensemble with RGYR constraints imposed on the entire A-tail system. F. Density plot of the TAMD ensemble with RGYR constraints imposed on the tail only.

While the contribution to scattering profiles of disordered regions often have little influence relative to the orientation of large globular domains that they connect, the sampling of configurations of disordered chains at the end of globular domains needs to be accounted for with greater care as successive structures in a MC simulation can be highly correlated and backbone torsion sampling is limiting. As a model system of a disordered polypeptide chain we considered residues 1-143 of HIV1 Gag which is referred to as A-tail, where residues 123-

143 were considered flexible as shown in Figure 7A. Initial MC simulation and subsampling and unconstrained TAMD density plots are shown in Figure 7B-D. In addition, two TAMD simulations were run using the initial RGYR value of each subsampled structure applied to all atoms in the A-tail model or only to the flexible residues of the A-tail model with density plots shown in Figure 7E-F respectively.

To evaluate the efficacy of TAMD to simulate polypeptide chains attached to globular protein regions it is instructive to evaluate a few representative TAMD trajectories in detail. In Figure 8A and B, RGYR is plotted as a function of simulation time for independent TAMD simulations (shown in red, green, and blue in A and red, green, and black in B). Each simulation was carried out with no RGYR constraints, RGYR constraints on the entire molecule, or RGYR constraints applied to the flexible residues only (tail). Without RGYR constraints on the entire molecule all three simulations resulted in compact structures with RGYR values plateauing at $\sim 17-18 \text{ \AA}$ and the tail RGYR plateauing at $\sim 8-10 \text{ \AA}$. In simulations where RGYR constraints were applied to the entire system the resulting RGYR of the entire molecule in some cases would collapse (blue lines with circles in Figure 8A) and in other cases remained stable (green and red lines with circles in Figure 8A). The RGYR of the tail was less effected when the RGYR was applied to the entire system where a broad range of RGYR values of the tail were observed (Figure 8B red, green and black lines with circles). In simulations where the RGYR constraints were only applied to the tail residues the resulting RGYR of the entire molecule (Figure 8A) collapsed to $\sim 17-18 \text{ \AA}$ while RGYR of the tail (Figure 8B) resulted in values $\sim 13-16 \text{ \AA}$ and thus less compacted. Representative structures from a selection of TAMD simulations are shown in Figure 8C-F.

The average RGYR and end-to-end distance distributions are shown in Figure 9. TAMD simulations without RGYR constraints led to a dominant population of collapsed molecules while constraining RGYR on all atoms leads to the broadest distribution of RGYR values using both metrics shown in Figure 9A and B. While we do not have access to an experimental RGYR value for

the A-tail system, an average experimental RGYR of residues 1-131 of A-Tail system was calculated from an ensemble of NMR structures [58] and was found
390 to be 17.4 ± 1 Å. Note that the NMR ensemble had 12 fewer residues than the A-tail system. Taken together, the results suggest that a proper way to provide broad conformational sampling requires applying RGYR constraints to the flexible regions of such disordered tails and not to the entire molecule.

3.3. Simulations of Nucleic Acids

395 Nucleic acid simulations are a particularly difficult problem due to the greater number of charges than proteins and the difficulty in capturing multi-body electrostatic affects. While simulation of double-stranded DNA, including NCP and higher order DNA complexes and structured RNA containing systems has been extensively studied [59, 60, 61, 62, 63, 64, 65, 66, 12, 19, 67] there is a continuing
400 need to improve conformational sampling for these challenging and important systems. As an example of the use of combined MC / TAMD on a disordered nucleic acid we used MC trajectories of the 60 bp B-Form DNA molecule and the NCP (which combines protein and DNA elements) using a MC algorithm derived for B-DNA [12].

405 Figure 10 demonstrates the spatial coverage of the structures from the MC ensemble (blue), subsampled ensemble (red) as well as the total ensemble from TAMD simulation (green). The density plot shows that the subsampled ensemble covers most of the MC ensemble in conformational space. We have found that DNA molecules are unstable using the implicit solvent models in
410 the TAMD simulations. To maintain a stable double stranded structure, and prevent DNA molecules from deviating from the initial subsampled structures in the TAMD simulation, a small harmonic restraint was applied on each DNA backbone atom with a force constant of 0.1 kcal/mol/Å^2 . Despite the harmonic restraints in the short TAMD simulations, the structure ensemble generated by
415 TAMD simulation adequately sampled configuration space.

To ensure the DNA molecules did not unwind during the TAMD simulations, we calculated the change of RGYR during the entire course of TAMD simula-

tions. As shown in Figure [11](#), the RGYR of the 60 BP DNA and the NCP molecule is stable (~ 0.3 Å for 60 bp DNA and ~ 0.15 Å for the NCP molecule) during the TAMD simulations. TAMD mainly samples the local fluctuations and serves as a tool to quickly relax DNA structures.

3.4. Simulations of Carbohydrates

Many mammalian proteins undergo post-translational modifications which incorporate carbohydrates into their structure with important consequences for their biological activity. Our approach to incorporate carbohydrate flexibility is to perform MC simulations on the protein component with rigid carbohydrate followed by subsampling and TAMD on both protein and carbohydrate degrees of freedom. [Ensemble modeling using molecular dynamics of glycosylated proteins to analyze SAXS data](#) and examples of using MC simulations of monoclonal antibodies have been described in the literature [\[68, 18, 41\]](#). In this example we focus on the TAMD simulation of the carbohydrate moiety in detail using a single configuration of the antibody protein fragment for brevity. Specifically, we carried out TAMD simulations the PTerm455 construct derived from the human monomer IgA1 antibody [\[41\]](#) containing two N-linked glycans. Based on the solution structure (PDB: 1IGA) by Boehm *et al.* [\[40\]](#) these glycans were modelled pointing away from the protein and into the solvent, in a splayed conformation (with the terminal sialic acids spread by 27.8 Å) as shown in Figure [12A](#). More recently, a crystal structure of the IgA1 Fc domain bound to a receptor (PDB: 1OW0) has been solved [\[39\]](#) in which one of the two arms of each of the biantennary N-linked glycans are resolved. This branch of the glycan runs along the Fc surface with the closest protein residue to the carbohydrate terminus being arginine 392, which is located towards the C-terminus of the protein (see Figure [12A](#)). Ten independent 500,000 step TAMD simulations were performed of the PTerm455 construct. In all simulations the protein segment remained fixed whilst the carbohydrate was free to move. During the simulations the glycans explore a wide range of conformations, both protruding into the solvent and lying along the protein surface as shown in Figure [12B&C](#).

However, all simulations predominantly sampled structures with one branch ap-
proaching arginine 392 as observed in the crystal structure with the terminal
450 sialic acids of the two arms within 10 Å of one another. These results are in
agreement with more recent solution studies where the best fitting structures
featured glycan conformations based upon the crystal structure pose [41].

4. Conclusion

The combination of MC and TAMD to rapidly generate physically repre-
455 sentative structural ensembles to compare to experimental data is a natural
evolution to overcome sampling and solvation limitations in molecular dynam-
ics simulations. While MC alone has the ability to sample a large configuration
space the limited degrees of freedom can lead to configurations that are not
adequately relaxed. It was found by evaluating several implicit solvent mod-
460 els that in order to prevent the prediction of unrealistic compact states, that
a simple RGYR constraint corresponding to the starting RGYR for each MC
configuration be applied. Thus a simple procedure has been described whereby
MC ensembles are spatially subsampled followed by RGYR constrained TAMD.
Due to the current state of current implicit solvent models the predictive ability
465 of the combined MC / TAMD method is limited. Regardless, the generation
of physically realistic and robust ensembles to be compared to experimental
data is a valuable advance that will improve as more accurate implicit sol-
vent models are developed. Programs to carry out the MC and TAMD for
proteins, nucleic acids, and carbohydrates are freely available ([https://sassie-
web.chem.utk.edu/sassie2](https://sassie-
470 web.chem.utk.edu/sassie2)).

5. Acknowledgements

This work benefitted from CCP-SAS software developed through a joint
National Science Foundation [CHE-1265821] and Engineering and Physical Sci-
ences Research Council [EP/K039121/1] grant. WZ and JC were supported
475 by the National Science Foundation [NSF CHE 1265850]. SCH acknowledges

support from the National Research Council Research Associateship program. AH was supported by a Summer Undergraduate Research Fellowship by the National Science Foundation [DMR-1508249]. XQ was supported by the National Science Foundation [MCB-1616337]. The authors thank M. Karplus for helping
480 setup the automatic CHARMM license verification implemented in SASSIE-
web.

References

- [1] S. R. Kline, Reduction and analysis of SANS and USANS data using IGOR Pro, Journal of Applied Crystallography 39 (6) (2006) 895–900. doi:10.1107/S0021889806035059.
485 URL <http://dx.doi.org/10.1107/S0021889806035059>
- [2] F. Spinozzi, C. Ferrero, M. G. Ortore, A. D. M. Antolinos, P. Mariani, Genfit: software for the analysis of small-angle x-ray and neutron scattering data of macromolecules in solution, Journal of Applied Crystallography
490 47 (3) (2014) 1132–1139.
- [3] P. Chacon, F. Moran, J. F. Diaz, E. Pantos, J. M. Andreu, Low-resolution structures of proteins in solution retrieved from x-ray scattering with a genetic algorithm, Biophysical Journal 74 (6) (1998) 2760–2775.
- [4] D. Franke, D. I. Svergun, DAMMIF, a program for rapid *ab-initio* shape
495 determination in small-angle scattering, Journal of Applied Crystallography 42 (2) (2009) 342–346. doi:10.1107/S0021889809000338.
URL <http://dx.doi.org/10.1107/S0021889809000338>
- [5] P. Bernado, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, Structural characterization of flexible proteins using small-angle x-ray scattering, J. Am. Chem. Soc. 129 (129) (2007) 5656–5664.
500
- [6] S. J. Perkins, A. I. Okemefuna, A. N. Fernando, A. Bonner, H. E. Gilbert, P. B. Furtado, X-ray and neutron scattering data and their constrained

- molecular modeling, in: J. J. Correia, I. H. W. Detrich (Eds.), Biophysical tools for biologists, volume one: in vitro techniques, Vol. 84 of Methods in Cell Biology, Academic Press, 2008, Ch. 13, pp. 375–423.
- 505
- [7] F. Spinozzi, M. Beltramini, Quafit: A novel method for the quaternary structure determination from small-angle scattering data, Biophysical Journal 103 (3) (2012) 511–521.
- [8] M. Pelikan, G. L. Hura, M. Hammel, Structure and flexibility within proteins as identified through small angle x-ray scattering, Gen. Physiol. Biophys. 28 (2) (2009) 174–189.
- 510
- [9] P. chia Chen, J. Hub, Interpretation of solution x-ray scattering by explicit-solvent molecular dynamics, Biophysical Journal 108 (10) (2015) 2573 – 2584. doi:http://dx.doi.org/10.1016/j.bpj.2015.03.062.
- 515 URL <http://www.sciencedirect.com/science/article/pii/S0006349515004038>
- [10] J. E. Curtis, S. Raghunandan, H. Nanda, S. Krueger, Sassi: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints, Computer Physics Communications 183 (2012) 382389.
- 520
- [11] J. E. Curtis, H. Zhang, H. Nanda, Sldmol: A tool for the structural characterization of thermally disordered membrane proteins, Computer Physics Communications 185 (11) (2014) 3010 – 3015. doi:http://dx.doi.org/10.1016/j.cpc.2014.07.006.
- 525 URL <http://www.sciencedirect.com/science/article/pii/S0010465514002483>
- [12] S. C. Howell, X. Qiu, J. E. Curtis, Monte carlo simulation algorithm for b-dna, Journal of Computational Chemistry 37 (29) (2016) 2553–2563. doi:10.1002/jcc.24474.
- 530 URL <http://dx.doi.org/10.1002/jcc.24474>

- [13] D. Hamelberg, J. Mongan, J. A. McCammon, Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules, The Journal of Chemical Physics 120 (24) (2004) 11919–11929. doi:http://dx.doi.org/10.1063/1.1755656.
535 URL <http://scitation.aip.org/content/aip/journal/jcp/120/24/10.1063/1.1755656>
- [14] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chemical Physics Letters 314 (1-2) (1999) 141 – 151. doi:http://dx.doi.org/10.1016/S0009-2614(99)01123-9.
540 URL <http://www.sciencedirect.com/science/article/pii/S0009261499011239>
- [15] B. Aoun, Fullrnc, a rigid body reverse monte carlo modeling package enabled with machine learning and artificial intelligence, Journal of Computational Chemistry 37 (12) (2016) 1102–1111. doi:10.1002/jcc.24304.
545 URL <http://dx.doi.org/10.1002/jcc.24304>
- [16] S. A. Datta, J. E. Curtis, W. Ratcliff, P. K. Clark, R. M. Crist, J. Lebowitz, S. Krueger, A. Rein, Conformation of the hiv-1 gag protein in solution, J Mol Biol 365 (2007) 812–24.
- [17] S. Krueger, J.-H. Shin, S. Raghunandan, J. Curtis, Z. Kelman, Atomistic ensemble modeling and small-angle neutron scattering of intrinsically disordered protein complexes: Applied to minichromosome maintenance protein, Biophysical Journal 101 (12) (2011) 2999 – 3007. doi:http://dx.doi.org/10.1016/j.bpj.2011.11.006.
550 URL <http://www.sciencedirect.com/science/article/pii/S0006349511013191>
555
- [18] N. J. Clark, H. Zhang, S. Krueger, H. J. Lee, R. R. Ketchem, B. Kerwin, S. R. Kanapuram, M. J. Treuheit, A. McAuley, J. E. Curtis, Small-angle neutron scattering study of a monoclonal antibody using free-energy

-
- 560 [constraints](#), The Journal of Physical Chemistry B 117 (45) (2013) 14029–
14038. [arXiv:http://dx.doi.org/10.1021/jp408710r](#), [doi:10.1021/
jp408710r](#).
URL [http://dx.doi.org/10.1021/jp408710r](#)
- [19] Y. Peng, J. E. Curtis, X. Fang, S. A. Woodson, [Structural model
of an mrna in complex with the bacterial chaperone hfq](#), Proceed-
565 ings of the National Academy of Sciences 111 (48) (2014) 17134–
17139. [arXiv:http://www.pnas.org/content/111/48/17134.full.pdf](#),
[doi:10.1073/pnas.1410114111](#).
URL [http://www.pnas.org/content/111/48/17134.abstract](#)
- [20] C. A. Castaneda, A. Chaturvedi, C. M. Camara, J. E. Curtis, S. Krueger,
570 D. Fushman, [Linkage-specific conformational ensembles of non-canonical
polyubiquitin chains](#), Phys. Chem. Chem. Phys. 18 (2016) 5771–5788. [doi:
10.1039/C5CP04601G](#).
URL [http://dx.doi.org/10.1039/C5CP04601G](#)
- [21] M. Green, L. Hatter, E. Brookes, P. Soultanas, D. J. Scott, [Defining
575 the intrinsically disordered c-terminal domain of {SSB} reveals dna-
mediated compaction](#), Journal of Molecular Biology 428 (2, Part A)
(2016) 357 – 364, study of biomolecules and biological systems: Proteins.
[doi:http://dx.doi.org/10.1016/j.jmb.2015.12.007](#).
URL [http://www.sciencedirect.com/science/article/pii/
580 S0022283615006981](#)
- [22] J. Chen, W. Im, r. Brooks, C. L., Application of torsion angle molecular
dynamics for efficient sampling of protein conformations, J Comput Chem
26 (2005) 1565–78.
- [23] S. Jian, K. H. Chen, Y. Choi, A coupled implicit method for chemical
585 non-equilibrium flows at all speeds, Journal of Computational Physics 106
(1993) 306–318.

- [24] E. G. Stein, L. M. Rice, A. T. Brünger, [Torsion-angle molecular dynamics as a new efficient tool for nmr structure calculation](#), *Journal of Magnetic Resonance* 124 (1) (1997) 154 – 164.
590 [doi:http://dx.doi.org/10.1006/jmre.1996.1027](#).
URL <http://www.sciencedirect.com/science/article/pii/S1090780796910277>
- [25] P. Gntert, C. Mumenthaler, K. Wüthrich, [Torsion angle dynamics for nmr structure calculation with the new program dyanal](#) edited by
595 [p. e. wright](#), *Journal of Molecular Biology* 273 (1) (1997) 283 – 298.
[doi:http://dx.doi.org/10.1006/jmbi.1997.1284](#).
URL <http://www.sciencedirect.com/science/article/pii/S0022283697912845>
- [26] A. D. Mackerell, M. Feig, C. L. Brooks, [Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations](#), *Journal of Computational Chemistry* 25 (11) (2004) 1400–1415. [doi:10.1002/jcc.20065](#).
600 URL <http://dx.doi.org/10.1002/jcc.20065>
- [27] J. A. D. MacKerell, D. Bashford, M. Bellott, R. L. D. Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wierkiewicz-Kuczera,
610 D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins., *Journal of Physical Chemistry B*, 102 (1998) 3586–3616.
- [28] N. Foloppe, A. D. MacKerell, Jr., [All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data](#), *Journal of Computational Chemistry*
615

- 21 (2) (2000) 86–104. [doi:10.1002/\(SICI\)1096-987X\(20000130\)21:2<86::AID-JCC2>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G).
URL [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2<86::AID-JCC2>3.0.CO;2-G](http://dx.doi.org/10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G)
- 620 [29] A. D. MacKerell, N. K. Banavali, [All-atom empirical force field for nucleic acids: Ii. application to molecular dynamics simulations of dna and rna in solution](#), *Journal of Computational Chemistry* 21 (2) (2000) 105–120. [doi:10.1002/\(SICI\)1096-987X\(20000130\)21:2<105::AID-JCC3>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P).
625 URL [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2<105::AID-JCC3>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P)
- [30] O. Guvench, E. R. Hatcher, R. M. Venable, R. W. Pastor, A. D. Mack-
erell, [Charmm additive all-atom force field for glycosidic linkages be-
tween hexopyranoses.](#), *J Chem Theory Comput* 5 (9) (2009) 2353–2370.
630 [doi:10.1021/ct900242e](https://doi.org/10.1021/ct900242e).
URL <http://dx.doi.org/10.1021/ct900242e>
- [31] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck,
M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy,
L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo,
635 D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C.
Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin,
M. Karplus, [All-atom empirical potential for molecular modeling and dy-
namics studies of proteins.](#), *J Phys Chem B* 102 (18) (1998) 3586–3616.
[doi:10.1021/jp973084f](https://doi.org/10.1021/jp973084f).
640 URL <http://dx.doi.org/10.1021/jp973084f>
- [32] A. D. Mackerell, Jr, M. Feig, C. L. Brooks, 3rd, [Extending the treat-
ment of backbone energetics in protein force fields: limitations of gas-phase
quantum mechanics in reproducing protein conformational distributions in
molecular dynamics simulations.](#), *J Comput Chem* 25 (11) (2004) 1400–

1415. [doi:10.1002/jcc.20065](https://doi.org/10.1002/jcc.20065),
URL <http://dx.doi.org/10.1002/jcc.20065>
- [33] E. P. Raman, O. Guvench, A. D. MacKerell, Jr, [Charmm additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses.](#), J Phys Chem B 114 (40) (2010) 12981–12994. [doi:10.1021/jp105758h](https://doi.org/10.1021/jp105758h),
URL <http://dx.doi.org/10.1021/jp105758h>
- [34] N. J. Clark, M. Raththagala, N. T. Wright, E. A. Buenger, J. F. Schildbach, S. Krueger, J. E. Curtis, Structures of traI in solution, J Mol Model 20 (2014) 2308.
- [35] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with namd, Journal of Computational Chemistry 26 (2005) 1781–1802.
- [36] W. Humphrey, A. Dalke, K. Schulten, VMD – Visual Molecular Dynamics, Journal of Molecular Graphics 14 (1996) 33–38.
- [37] M. van Dijk, A. M. J. J. Bonvin, [3d-dart: a dna structure modelling server](#), Nucleic Acids Research 37 (suppl 2) (2009) W235–W239. [arXiv:http://nar.oxfordjournals.org/content/37/suppl_2/W235.full.pdf+html](https://arxiv.org/http://nar.oxfordjournals.org/content/37/suppl_2/W235.full.pdf+html),
[doi:10.1093/nar/gkp287](https://doi.org/10.1093/nar/gkp287).
URL http://nar.oxfordjournals.org/content/37/suppl_2/W235.abstract
- [38] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, [Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 resolution](#), Journal of Molecular Biology 319 (5) (2002) 1097 – 1113. [doi:http://dx.doi.org/10.1016/S0022-2836\(02\)00386-8](https://doi.org/http://dx.doi.org/10.1016/S0022-2836(02)00386-8).
URL <http://www.sciencedirect.com/science/article/pii/S0022283602003868>
- [39] A. B. Herr, E. R. Ballister, P. J. Bjorkman, [Insights into iga-mediated immune responses from the crystal structures of human fcα1 and its](#)

-
- complex with iga1-fc., Nature 423 (6940) (2003) 614–620. doi:10.1038/nature01685.
- 675 URL <http://dx.doi.org/10.1038/nature01685>
- [40] M. K. Boehm, J. M. Woof, M. A. Kerr, S. J. Perkins, The fab and fc fragments of iga1 exhibit a different arrangement from that in igg: a study by x-ray and neutron solution scattering and homology modelling., J Mol Biol 286 (5) (1999) 1421–1447. doi:10.1006/jmbi.1998.2556.
- 680 URL <http://dx.doi.org/10.1006/jmbi.1998.2556>
- [41] G. K. Hui, D. W. Wright, O. L. Vennard, L. E. Rayner, M. Pang, S. C. Yeo, J. Gor, K. Molyneux, J. Barratt, S. J. Perkins, The solution structures of native and patient monomeric human iga1 reveal asymmetric extended structures: implications for function and igan disease, Biochemical Journal 471 (2) (2015) 167–185. arXiv:<http://www.biochemj.org/content/471/2/167.full.pdf>, doi:10.1042/BJ20150612.
- 685 URL <http://www.biochemj.org/content/471/2/167>
- [42] S. Jo, T. Kim, V. G. Iyer, W. Im, Charmm-gui: a web-based graphical user interface for charmm., J Comput Chem 29 (11) (2008) 1859–1865.
- 690 doi:10.1002/jcc.20945.
- URL <http://dx.doi.org/10.1002/jcc.20945>
- [43] S. Jo, K. C. Song, H. Desaire, A. D. MacKerell, Jr, W. Im, Glycan reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins., J Comput Chem 32 (14) (2011) 3135–3141.
- 695 doi:10.1002/jcc.21886.
- URL <http://dx.doi.org/10.1002/jcc.21886>
- [44] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, Charmm - a program for macromolecular energy, minimization, and dynamics calculations, J Comput Chem 4 (1983) 187–217.
- 700 [45] B. R. Brooks, r. Brooks, C. L., J. Mackerell, A. D., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch,

- L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, Charmm: The biomolecular simulation program, *J Comput Chem* 30 (2009) 1545–614.
- [46] W. Im, M. S. Lee, r. Brooks, C. L., Generalized born model with a simple smoothing function, *J Comput Chem* 24 (2003) 1691–702.
- [47] J. Chen, Effective approximation of molecular volume using atom-centered dielectric functions in generalized born models, *J Chem Theory Comput* 6 (2010) 2790–2803.
- [48] P. Ferrara, J. Apostolakis, A. Caffisch, Evaluation of a fast implicit solvent model for molecular dynamics simulations, *Proteins* 46 (2002) 24–33.
- [49] J. Chen, C. L. Brooks, 3rd, J. Khandogin, Recent advances in implicit solvent based methods for biomolecular simulations, *Curr. Opin. Struct. Biol.* 18 (2008) 140–148.
- [50] M. C. Watson, J. E. Curtis, Rapid and accurate calculation of small-angle scattering profiles using the golden ratio, *Journal of Applied Crystallography* 46 (4) (2013) 1171–1177. doi:10.1107/S002188981301666X. URL <http://dx.doi.org/10.1107/S002188981301666X>
- [51] D. J. Tobias, Electrostatics calculations: recent methodological advances and applications to membranes, *Curr. Opin. Struct. Biol.* 11 (2001) 253–261.
- [52] P. Ferrara, H. Gohlke, D. Price, G. Klebe, C. L. Brooks, 3rd, Assessing scoring functions for protein-ligand interactions, *Journal of Medicinal Chemistry* 47 (2004) 3032–3047.
- [53] J. Chen, C. L. Brooks, 3rd, Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions, *Phys. Chem. Chem. Phys.* 10 (2008) 471–481.

- 730 [54] A. Liwo, C. Czaplewski, S. Oldziej, H. A. Scheraga, [Computational techniques for efficient conformational sampling of proteins](#), *Current Opinion in Structural Biology* 18 (2) (2008) 134 – 139, theory and simulation / *Macromolecular assemblages*.
[doi:http://dx.doi.org/10.1016/j.sbi.2007.12.001](http://dx.doi.org/10.1016/j.sbi.2007.12.001),
735 URL <http://www.sciencedirect.com/science/article/pii/S0959440X07002011>
- [55] D. M. Zuckerman, [Equilibrium sampling in biomolecular simulations](#), *Annual Review of Biophysics* 40 (1) (2011) 41–62, pMID: 21370970. [arXiv:http://dx.doi.org/10.1146/annurev-biophys-042910-155255](#), [doi:10.1146/annurev-biophys-042910-155255](#),
740 URL <http://dx.doi.org/10.1146/annurev-biophys-042910-155255>
- [56] V. N. Uversky, C. J. Oldfield, A. K. Dunker, [Intrinsically disordered proteins in human diseases: Introducing the d2 concept](#), *Annual Review of Biophysics* 37 (1) (2008) 215–246, pMID: 18573080. [arXiv:http://dx.doi.org/10.1146/annurev.biophys.37.032807.125924](#),
745 [doi:10.1146/annurev.biophys.37.032807.125924](http://dx.doi.org/10.1146/annurev.biophys.37.032807.125924).
URL <http://dx.doi.org/10.1146/annurev.biophys.37.032807.125924>
- [57] D. D. Boehr, R. Nussinov, P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, *Nature Chemical Biology* 5 (11) (2009) 789–796. [arXiv:http://doi.org/10.1038/nchembio.232](#).
750
- [58] J. S. Saad, J. Miller, J. Tai, A. Kim, R. H. Ghanam, M. F. Summers, [Structural basis for targeting hiv-1 gag proteins to the plasma membrane for virus assembly](#), *Proceedings of the National Academy of Sciences* 103 (30) (2006) 11364–11369. [arXiv:http://www.pnas.org/content/103/30/11364.full.pdf](#), [doi:10.1073/pnas.0602818103](#),
755 URL <http://www.pnas.org/content/103/30/11364.abstract>
- [59] A. Savelyev, A. D. MacKerell, All-atom polarizable force field for DNA

- based on the classical drude oscillator model, *Journal of Computational Chemistry* 35 (16) (2014) 1219–1239. [doi:10.1002/jcc.23611](https://doi.org/10.1002/jcc.23611).
- [60] D. Norouzi, V. B. Zhurkin, Topological Polymorphism of the Two-Start Chromatin Fiber, *Biophysical Journal* 108 (10) (2015) 2591–2600. [doi:10.1016/j.bpj.2015.04.015](https://doi.org/10.1016/j.bpj.2015.04.015).
- [61] Y. Chen, J. M. Tokuda, T. Topping, J. L. Sutton, S. P. Meisburger, S. A. Pabit, L. M. Gloss, L. Pollack, Revealing transient structures of nucleosomes as DNA unwinds, *Nucleic Acids Research* 42 (13) (2014) 8767–8776. [doi:10.1093/nar/gku562](https://doi.org/10.1093/nar/gku562).
- [62] K. Andresen, I. Jimenez-Useche, S. C. Howell, C. Yuan, X. Qiu, Solution Scattering and FRET Studies on Nucleosomes Reveal DNA Unwrapping Effects of H3 and H4 Tail Removal, *PLoS ONE* 8 (11) (2013) e78587. [doi:10.1371/journal.pone.0078587](https://doi.org/10.1371/journal.pone.0078587).
- [63] A. Garai, S. Saurabh, Y. Lansac, P. K. Maiti, DNA Elasticity from Short DNA to Nucleosomal DNA, *The Journal of Physical Chemistry B* 119 (34) (2015) 11146–11156. [doi:10.1021/acs.jpccb.5b03006](https://doi.org/10.1021/acs.jpccb.5b03006).
- [64] N. Alegret, E. Santos, A. Rodríguez-Fortea, F. X. Rius, J. M. Poblet, Disruption of small double stranded DNA molecules on carbon nanotubes: A molecular dynamics study, *Chemical Physics Letters* 525–526 (2012) 120–124. [doi:10.1016/j.cplett.2011.12.077](https://doi.org/10.1016/j.cplett.2011.12.077).
- [65] S. Bowerman, J. Wereszczynski, Effects of MacroH2A and H2A.Z on Nucleosome Dynamics as Elucidated by Molecular Dynamics Simulations, *Biophysical Journal* 110 (2) (2016) 327–337. [doi:10.1016/j.bpj.2015.12.015](https://doi.org/10.1016/j.bpj.2015.12.015).
- [66] H. Zhang, S. Khodadadi, S. L. Fiedler, J. E. Curtis, [Role of water and ions on the dynamical transition of rna](https://doi.org/10.1021/acs.jpclett.4c01911), *The Journal of Physical Chemistry Letters* 4 (19) (2013) 3325–3329. [arXiv:http://dx.doi.org/10.1021/](https://arxiv.org/http://dx.doi.org/10.1021/)

[jz401406c](#), [doi:10.1021/jz401406c](#)

URL <http://dx.doi.org/10.1021/jz401406c>

[67] P. Auffinger, E. Westhof, [Simulations of the molecular dynamics of nucleic acids](#), *Current Opinion in Structural Biology* 8 (2) (1998) 227 – 236.

790 [doi:http://dx.doi.org/10.1016/S0959-440X\(98\)80044-4](#).

URL <http://www.sciencedirect.com/science/article/pii/S0959440X98800444>

[68] M. Guttman, P. Weinkam, A. Sali, K. Lee, [All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosy-](#)

795 [lated proteins](#), *Structure* 21 (3) (2013) 321 – 331. [doi:http://dx.doi.org/10.1016/j.str.2013.02.004](#).

URL <http://www.sciencedirect.com/science/article/pii/S0969212613000440>

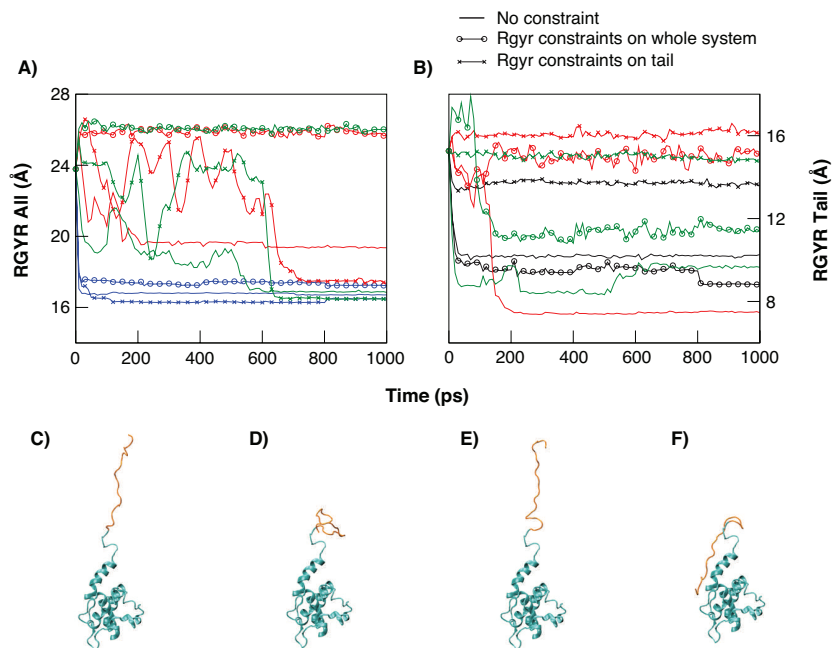
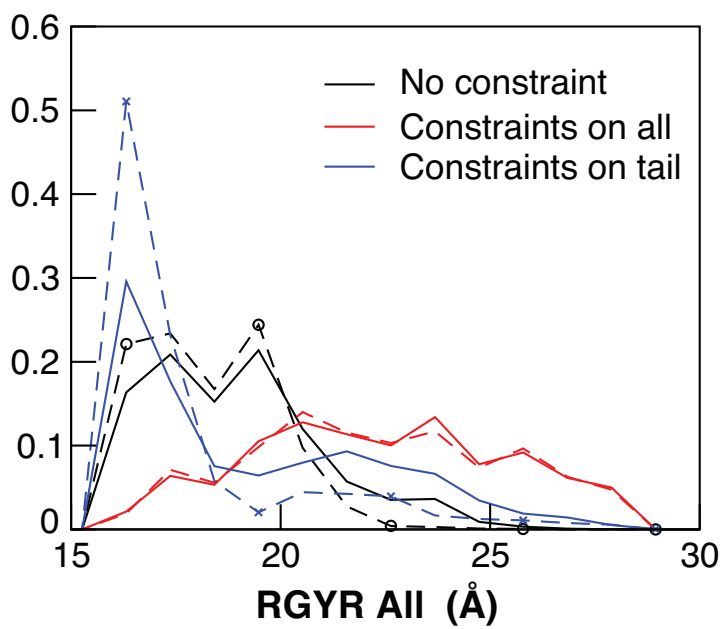


Figure 8: Affect of RGYR constraints on A-tail TAMD simulations. A). RGYR plot of the entire A-tail system (residues 1-143) from three representative TAMD simulations. B). RGYR plot of tail region (residues 123-143) from three representative TAMD simulations. For both panels A and B straight lines without symbols indicate the TAMD simulations without constraints; lines with circles indicate the RGYR from the TAMD simulations with RGYR constraints on the entire A-tail system; lines with crosses indicate RGYR from TAMD simulations with RGYR constraints on the tail region only. Different colors indicate simulations starting from different initial structures. C). Cartoon rendering of the initial structure of one representative TAMD simulation (red in A and B). D). Final snapshot of the structure in TAMD simulation without constraints. E) Final snapshot of the structure in TAMD simulation with RGYR constraints on the entire A-tail system. F) Final snapshot of the structure in TAMD simulation with RGYR constraints on the tail region only.

A)



B)

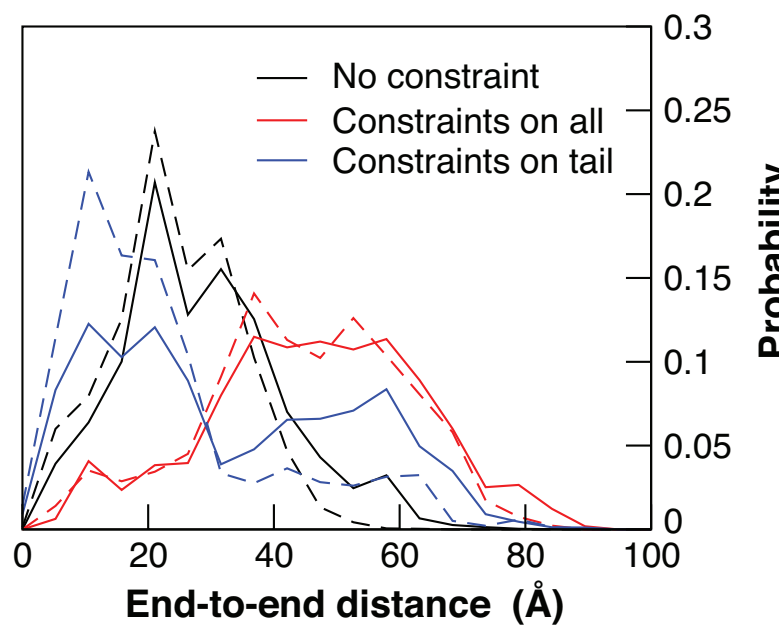


Figure 9: A). RGYR distribution of the A-tail system from TAMD simulations. Solid lines indicate the distribution from the first 500ps simulations and dashed lines indicate the distribution from the last 500ps simulations.

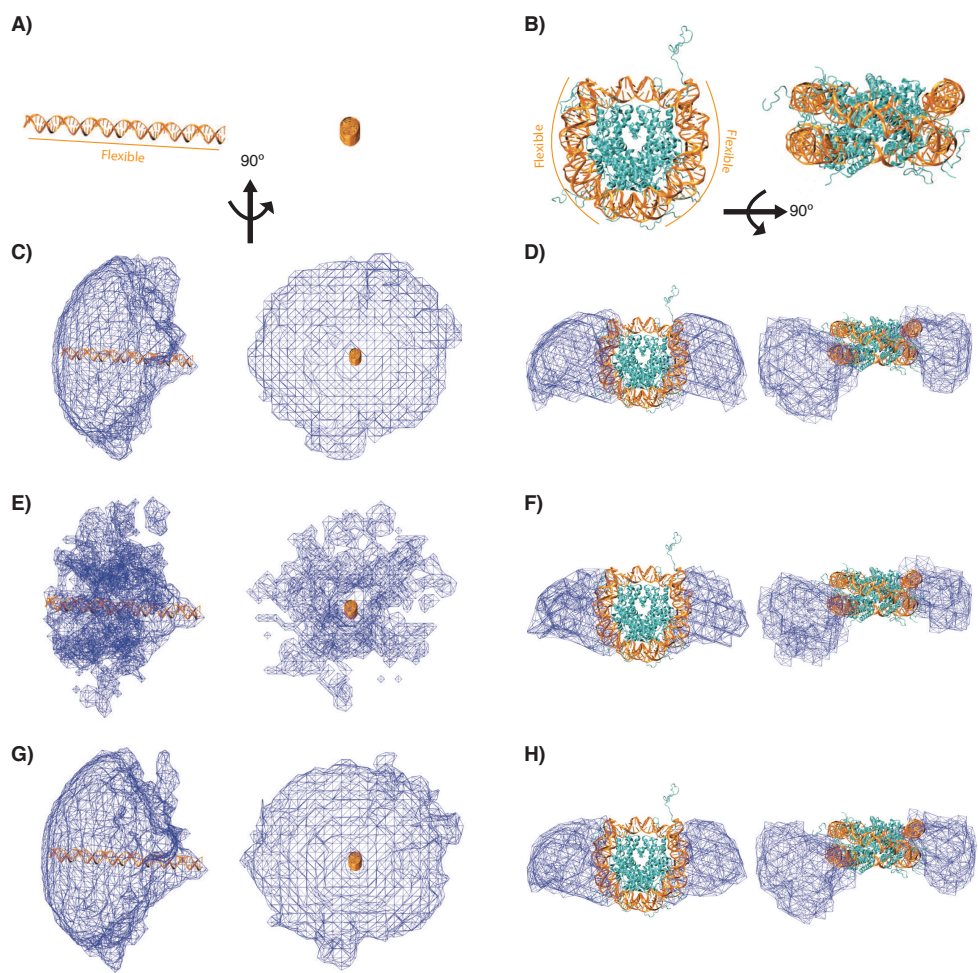


Figure 10: Cartoon representation of A) 60 bp B-DNA, and B) NCP molecule. Protein molecules are colored in cyan and DNA molecules are colored in orange. The flexible regions sampled in MC simulations are highlighted by lines. C) and D) density plot of spatial coverage of MC ensemble. E and F) density plot of the spatial coverage of subsampled ensemble. G) and H) density plot of the spatial coverage of TAMD ensemble. 90 degree rotation about the z-axis is shown for each case.

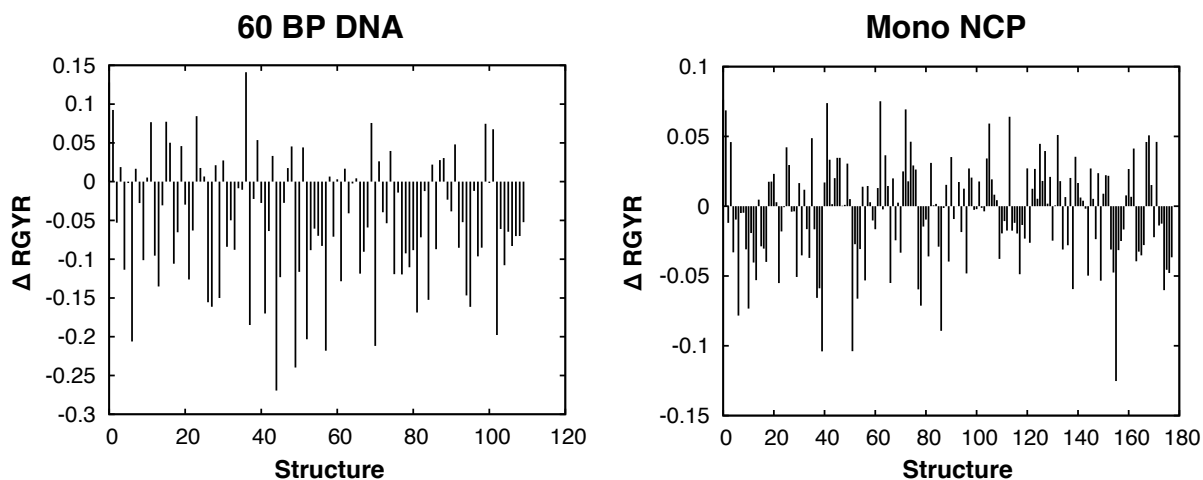


Figure 11: The change of RGyr of the DNA molecules during the TAMd simulations.

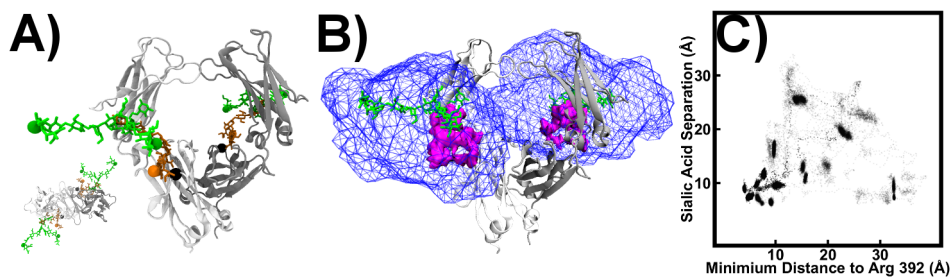


Figure 12: TAMd simulation of the N-glycan in the PTerm455 construct based on IgA1. (A) The location of the N-linked glycans in the Fc fragment of the construct for both the crystal structure (orange) and the initial structure used for simulation which is based on the model of Boehm *et al.* [40] (green). The C1 atoms of the terminal sialic acid monosaccharides are shown as spheres. The carbon alpha of arginine 392 is shown as a black sphere. The inset figure shows the structure viewed along the symmetry axis. (B) Occupancy plot showing the space explored by the N-glycans during simulations. The blue wireframe volume represents all regions of space sampled during the simulation and the red shaded region that sampled by more than 50 % of the simulation frames. The crystal structure glycan conformation (orange) is shown for comparison. (C) The variation in glycan conformation observed in the simulation characterized by the minimum distance between a C1 atom of one of the branch terminal sialic acids to arginine 392 and the separation between the two branches.