CoSIMS: An Optimized Trajectory Based Collision Simulator for Ion Mobility Spectrometry
SUPPORTING INFORMATION

Christopher A. Myers,[a] Rebecca J. D'Esposito,[b] Daniele Fabris,[bcd]
Srivathsan V. Ranganathan,[c] and Alan A. Chen*[bc]

[a] *Department of Physics, University at Albany (SUNY), New York, USA.*
[b] *Department of Chemistry, University at Albany (SUNY), Albany, New York, USA.*
[c] *The RNA Institute, University at Albany (SUNY), New York, USA.*
[d] *Department of Biological Sciences, University at Albany (SUNY), New York, USA.*

S1 Introduction

In the following sections, details of numerical algorithms implemented into CoSIMS will be presented. First, a more detailed derivation of these methods will be discussed followed by some parameterization results and CPU benchmarks. The final section of this document contains tabulated data for many of the plots included either here or in the main text.

S2 Determining the atomic clusters

Before either the multipole or dispersion cut-off approximations can be applied, all of the atoms in the ion must be separated into clusters. Clustering allows CoSIMS to effeciently search for atoms that meet our approximation criteria by looking at groups of atoms at a time while also providing a means for formulating the multipole expansion moments. The algorithm used for the clustering implements a principle component analysis (PCA) in the following manor.

1. First, find the Covariance matrix σ_{ij} of the entire N atom molecule, defined as

$$\sigma_{ij} = \sum_{n}^{N} (x_{ni} - \mu_i)(x_{nj} - \mu_j) \tag{S1}$$

   where x_{ni} is the ith position vector component of the nth atom in the molecule and μ_j is the mean vector component of all N atoms.

2. Compute the eigenvalues and eigenvectors of σ_{ij}. Since this is only a 3×3 matrix, CoSIMS uses an exact solution for the digitalization.

3. The eigenvector corresponding to the largest eigenvalue will be the principle component vector that we use to split the cluster of atoms. This vector will define a plane such that any atoms lying above this plane go into one cluster and those below the plane go into another.

4. Compute the center of each new cluster by taking the arithmetic mean of all atoms. The size of each cluster is then defined by the maximum vector magnitude of all atoms from this center.

5. Go to step 2 and repeat for each cluster who's size is smaller than some desired tolerance, which we will call R_c.

6. Compute the total charge, dipole, and quadrupole of each cluster.

S3 Multipole Expansion

As mentioned in section 3.3 of the main article, the multipole expansion is performed by partitioning the electrostatic potential into an exact term $V^{(e)}$ and an approximate term $V^{(m)}$ such that $V = V^{(e)} + V^{(m)}$. The exact potential uses all atoms that belong to clusters that are either within or intersect the multipole cut-off sphere region $\mathcal{R}$ and is computed using

$$V^{(e)} = \sum_{i \in \mathcal{R}} \frac{q_i}{|\mathbf{R} - \mathbf{r}_i|}. \tag{S2}$$

Here, the summation runs over all atoms in $\mathcal{R}$. The approximate multipole term thus sums over all potentials computed from all other clusters of atoms. For an nth order approximation, the multipole potential $V^{(n)}$ will have the form

$$V^{(n)} = \frac{1}{n! d^{2n+1}} d_{\alpha_1} d_{\alpha_2} \cdots d_{\alpha_n} Q^{(n)}_{\alpha_1 \alpha_2 \cdots \alpha_n} \tag{S3}$$

where $\mathbf{d}_i = \mathbf{R} - \mathbf{a}_i$ is the distance from buffer gas particles position $\mathbf{R}$ to the vector in which the nth multipole moment $Q^{(n)}$ is expanded about. Throughout this article, we will invoke an implicit summation over repeated indices's of tensors. The multipole representation of the electrostatic potential is thus

$$V = V^{(e)} + \sum_n V^{(n)} \tag{S4}$$

The electric field E_ν evaluated at $\mathbf{R}$ is given by the negative gradient of the potential $-\partial_\nu V$ and is also partitioned into an exact and approximated field $E_\nu = E^{(e)}_\nu + E^{(m)}_\nu$. For small localized charges assigned to each atom, the induced dipole in the buffer gas atom p_ν is proportional to the electric field so that the potential energy U is

$$U = -\frac{1}{2} p_\nu E_\nu = -\frac{\alpha}{2} E_\nu E_\nu = -\frac{1}{2}\alpha \left(E^{(m)}_\nu + E^{(e)}_\nu\right)\left(E^{(m)}_\nu + E^{(e)}_\nu\right) \tag{S5}$$

with polarization constant α. Further taking gradients of the above energy term gives the force f_μ exerted on the buffer gas

$$f_\mu = -\partial_\mu U = \alpha \left(F^{(m)}_{\mu\nu} + F^{(e)}_{\mu\nu}\right)\left(E^{(m)}_\nu + E^{(e)}_\nu\right), \tag{S6}$$

where $F_{\mu\nu} = \partial_\mu E_\nu$ is denoted as the vector derivative of the electric field. From differentiating equation S3, we can write following general forms for the nth multipole field and vector derivatives used in CoSIMS:

$$E^{(n)}_\nu = \frac{1}{n! b^{2n+3}} \left((2n+1) b_\nu Q_{\alpha_1 \alpha_2 \cdots \alpha_n} \prod_{i=1}^{n} b_{\alpha_i} - n b^2 Q_{\alpha_1 \alpha_2 \cdots \alpha_{n-1} \nu} \prod_{i=1}^{n-1} b_{\alpha_i} \right) \tag{S7}$$

$$\begin{aligned} F^{(n)}_{\mu\nu} = \frac{2n+1}{n! b^{2n+5}} \Bigg[& \left(b^2 \delta_{\mu\nu} - (2n+3) b_\mu b_\nu\right) Q_{\alpha_1 \cdots \alpha_n} \prod_{i=1}^{n} b_{\alpha_i} \\ & + n b^2 \left(Q_{\alpha_1 \cdots \alpha_{n-1} \nu} b_\mu + Q_{\alpha_1 \cdots \alpha_{n-1} \mu} b_\nu\right) \prod_{i=1}^{n-1} b_{\alpha_i} \\ & - \frac{n b^4 (n-1)}{2n+1} Q_{\alpha_1 \cdots \alpha_{n-2} \mu\nu} \prod_{i=1}^{n-2} b_{\alpha_i} \Bigg] \end{aligned} \tag{S8}$$

CoSIMS uses up to the quadrupole potential, that is $n = 3$, although the framework presented here has the potential to incorporate higher order approximations in future versions of the software.

To test the accuracy of the multiple expansion, we applied a uniformly distributed $+20e$ charge to a handful of the DNA strands referenced in the main article. For the short strands, we recognize that this is an unrealistic charge distribution, however, the purpose is to use a charge that is large enough where ion induced dipole interactions become significant relative to the strength of the Lennard-Jones interactions. The results of this test are presented in Figure S1, where the relative percent difference σ_p is defined as $\sigma_p = 100 \times |A - B|/A$ where A is the CCS of the molecule using the exact potential and B is the CCS using the multipole approximation. Although neither the dipole or quadrupole approximation is consistently better than the other, all of the tests give a percent difference less than 0.02% which is well below the numerical accuracy of the tests.

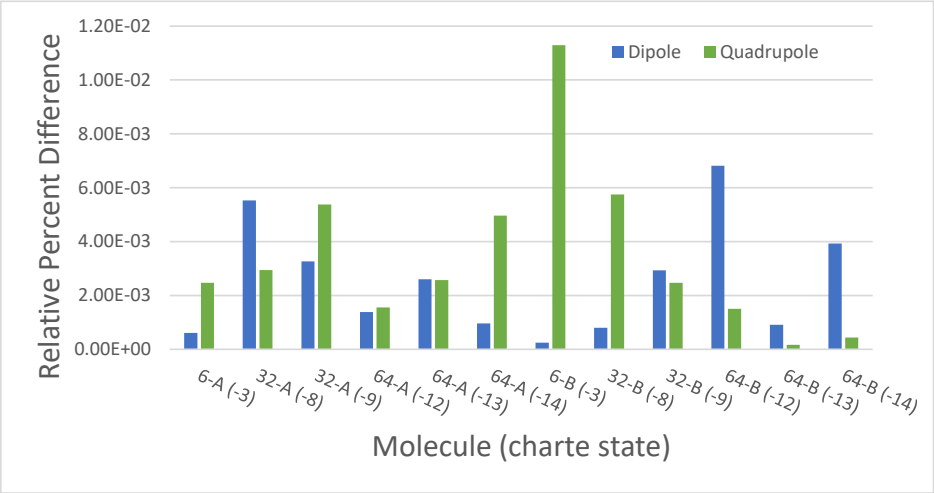


Figure S1: Relative percent difference between exactly calculated strands with a $+20e$ charge and the dipole or quadrupole approximation. A total of 2.5×10^5 trajectories were used at a simulation temperature of 298 Degrees Kelvin. A cut-off radius of 35Å and a cluster size of 4Å were used.

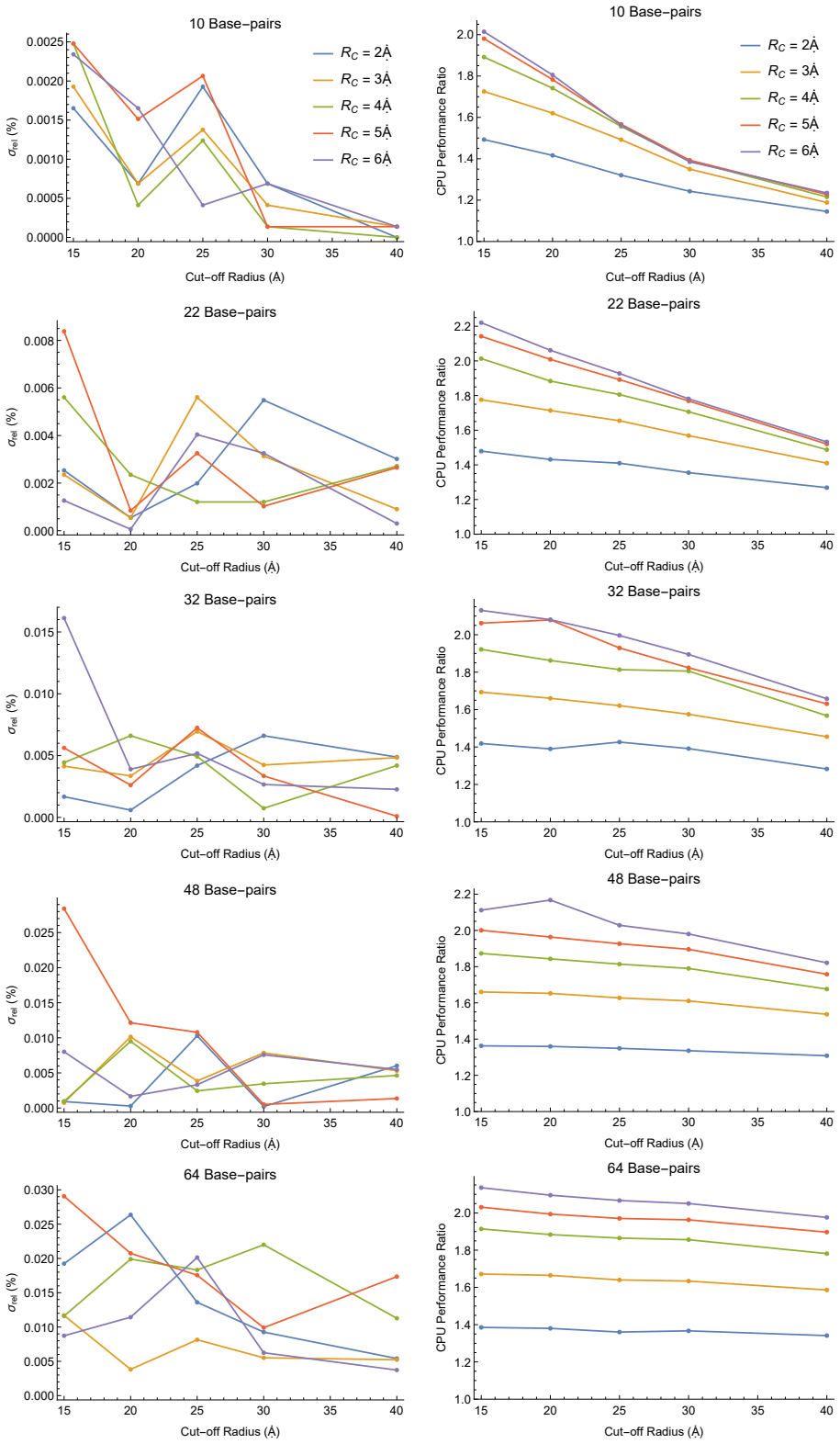


Figure S2: CCS dependence on cut-off radius and cluster size for the multipole approximation (dipole). All DNA strands used here were the same A-form strands referred to in the main text.

A further analysis of the change in accuracy and CPU performance was conducted and show in Figure S2. A single strand form the ensemble of strands referenced to in the main text were used in each of the tests. The relative percent error is the same definition used in the main article, and the CPU performance ratio is the ratio of the total CPU time used used to calculate the CCS using an exact potential to the total CPU time used to calculate the CCS with the dipole approximation.

S4 Dispersion Cut-off Approximation

For atoms distant form the buffer gas particle, the Lennard-Jones interactions will have a negligible effect on the particles force. To approximate the distance at which this occurs, consider a single particle centered at the origin that is surrounded by a uniformly dense solid that interacts with the particle through a Lennard-Jones interaction. We will assume that the particle is centered within a hollow, spherical cavity of radius σ. The total potential energy U of this particle will be

$$U = \int \mathrm{d}^3\mathbf{r}\ 4\rho \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] = 4\pi \int_{\sigma}^{\infty} \mathrm{d}r r^2\ 4\rho \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] = -\frac{32}{9}\pi\rho\sigma^3 \tag{S9}$$

with energy per unit volume ρ. By assigning a cut-off radius, we require that ratio of the total energy of this fictitious particle using the cutoff sphere to the total energy computed above is some number α with $0 < \alpha < 1$. The total energy $\tilde{U}$ within this cut-off sphere of radius a is

$$\begin{aligned} \tilde{U} &= 4\pi \int_{\sigma}^{a} \mathrm{d}r r^2\ 4\rho \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \\ &= \frac{16}{9}\pi\rho\sigma^3 \left[3\frac{\sigma^3}{a^3} - 2 - \frac{\sigma^9}{a^9} \right] \end{aligned} \tag{S10}$$

If the cut-off radius a is sufficiently much larger than σ, then their ratio will be small and the (σ^9/a^9) term can be dropped. Applying this approximation and taking the ratio of $\tilde{U}/U$ gives us

$$\tilde{U}/U = 1 - \frac{3}{2}\left(\frac{\sigma}{a}\right)^3 = \alpha \quad \rightarrow \quad a = \sigma \left(\frac{3}{2-2\alpha}\right)^{1/3} \tag{S11}$$

CoSIMS chooses the largest possible σ parameter available in the forcefield, which is $3.5\mathring{A}$, and a ratio $\alpha = 0.995$. For these conditions, $a \approx 23.43\mathring{A}$.

Similar to the dipole accuracy tests presented in Figure S2, we performed the same tests for dispersion cut-off approximation on the same strands and with the same simulation parameters, except that no charge was applied to the molecule. The results are shown in Figure S3.

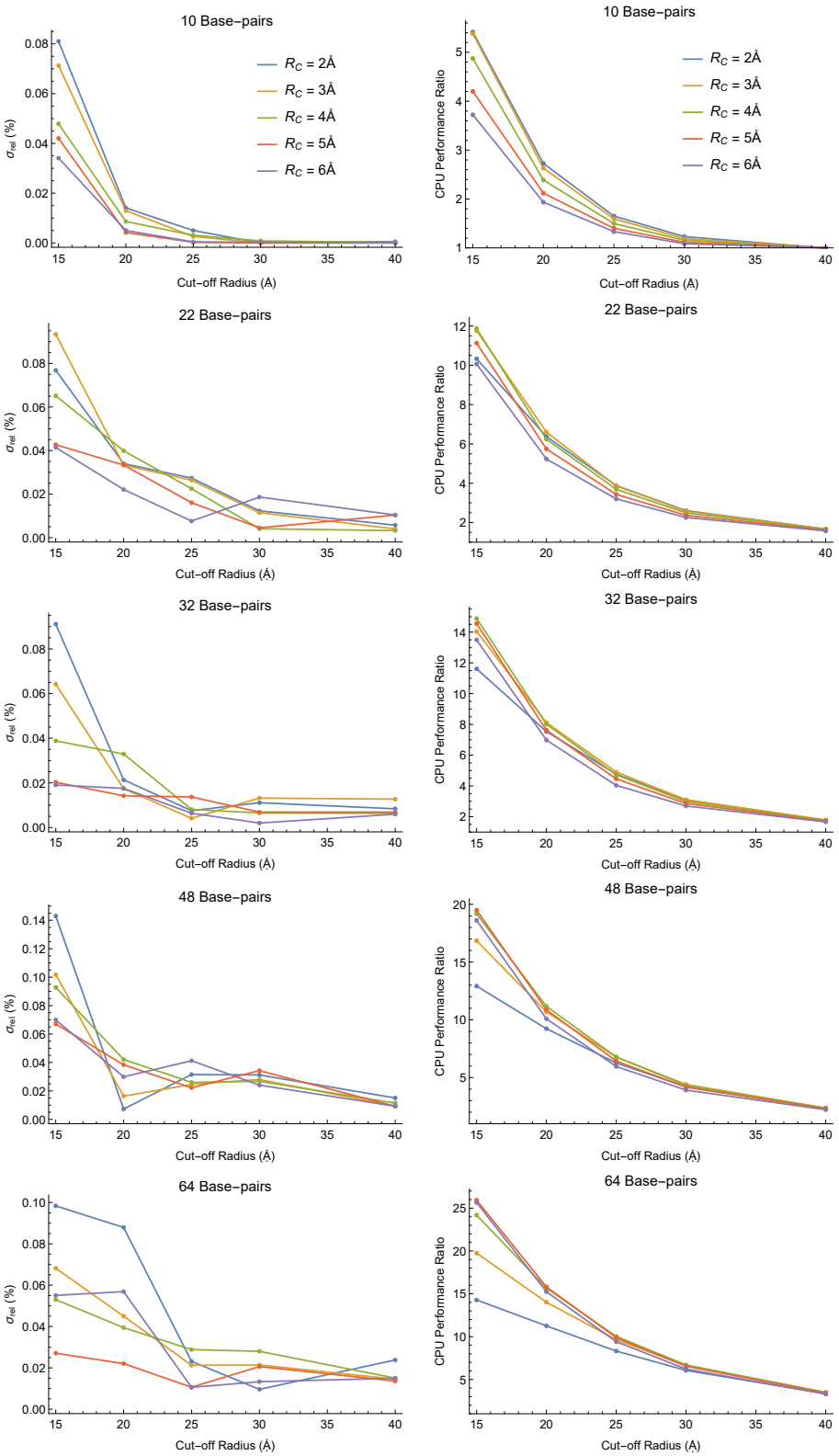
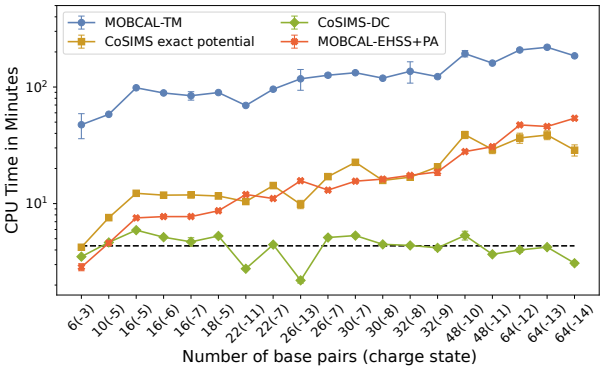


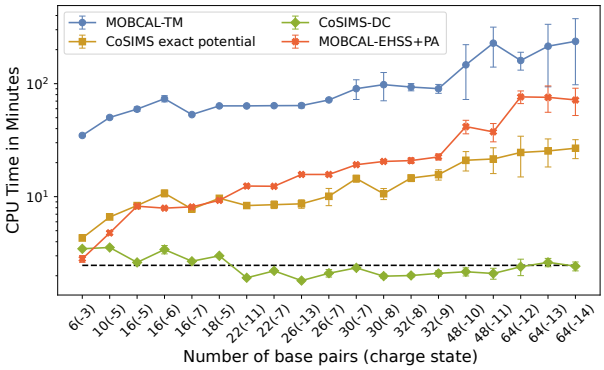
Figure S3: CCS dependence on cut-off radius and cluster size for the dispersion cut-off approximation. All DNA strands used here were the same A-form strands referred to in the main text.

S5 CPU Benchmarks

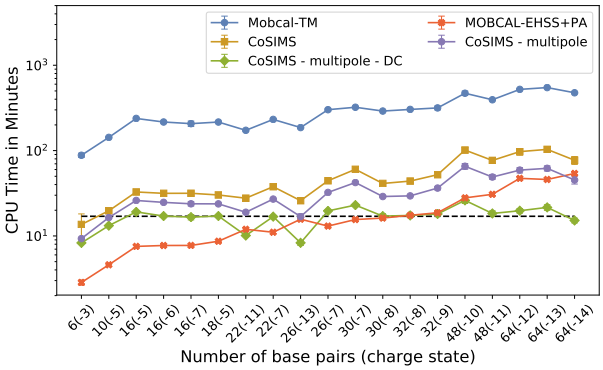
In Figure 7 of the main article, we showed the CPU performance for neutral DNA strands for MOBCAL vs. CoSIMS using the exactly calculated potential and the dispersion cut-off approximation for the B-form DNA strands. Shown in Figure S4 are the remaining figures for the A-form DNA strands and for using a uniformly distributed charge.



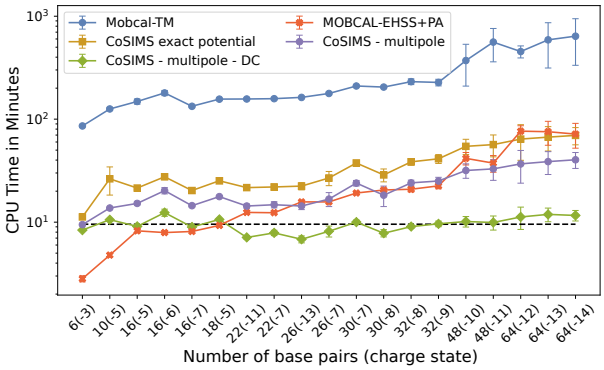
(a) Neutral A-form DNA strands.
Dashed line average: 4.33 ± 14.70 minutes.



(b) Charged A-form DNA strands.
Dashed line average: 2.47 ± 0.52 minutes.



(c) Neutral A-form DNA strands.
Dashed line average: 16.99 ± 15.76 minutes.



(d) Charged A-form DNA strands.
Dashed line average: 9.54 ± 9.25 minutes.

Figure S4: CPU runtimes for neutral and charged DNA strands. The dashed line represents an average CPU time using the DC approximation.

CoSIMS also utilizes the OpenMP multi-threading library to distribute the calculation of the programs many trajectories over multiple processors. Figure S5 demonstrates the efficiency of the OpenMP implementation in CoSIMS for the 64 base pair B-form DNA strands. The Ratio of the total CPU time used to calculate the CCS using 1 CPU thread t_0 to the total time to calculate with multiple threads t is shown on the left axis while the total time is shown on the right axis. This test was performed using the same hardware configuration describe in the main text.

S6 Stability and RMSD Tests for Long DNA Strands

As mentioned in the main article, MOBCAL gives inaccurate CCSs for the large B-form strands of DNA in our data set. An example of such occurrences and how the RMSD difference between two consecutive MD simulation frames does not explain this error was presented for the 64 base pair, B-form, -14 charge state

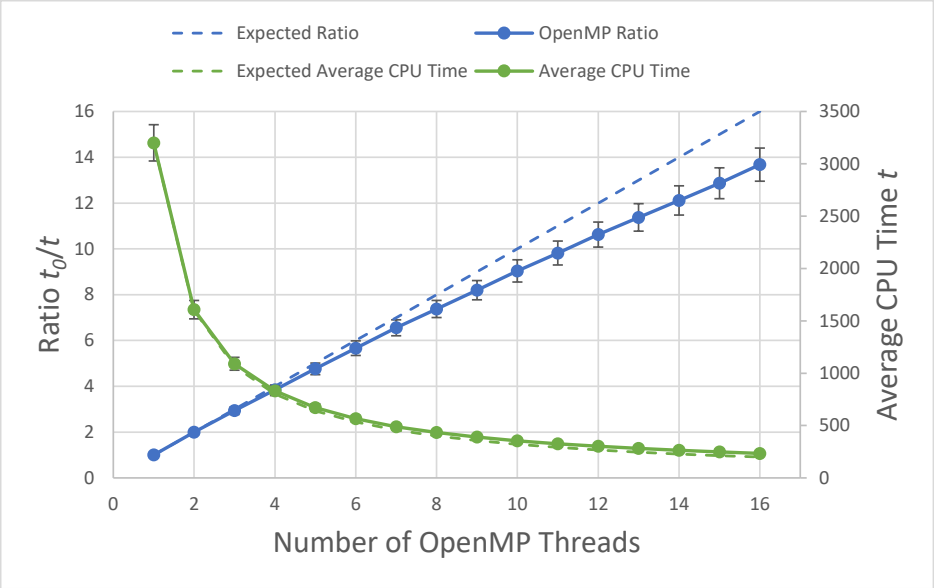
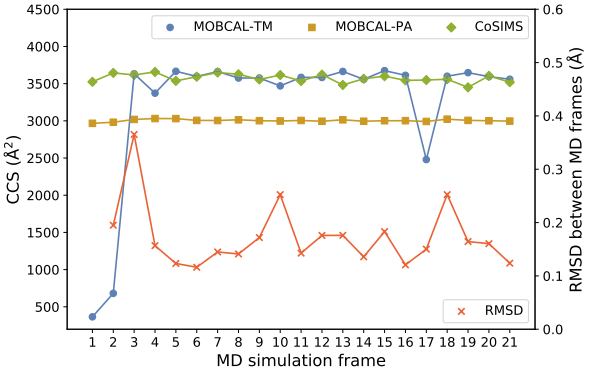
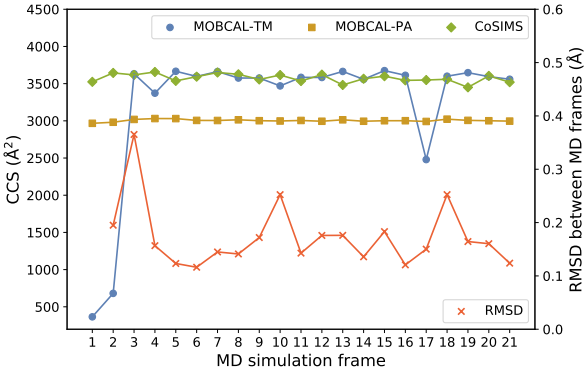


Figure S5: CPU Scaling for OpenMP threads for the 64-B(-12) DNA strand. Data points represented as a average CPU wall time for all 21 MD snapshots and Error bars represent the standard deviation of the respected averages.

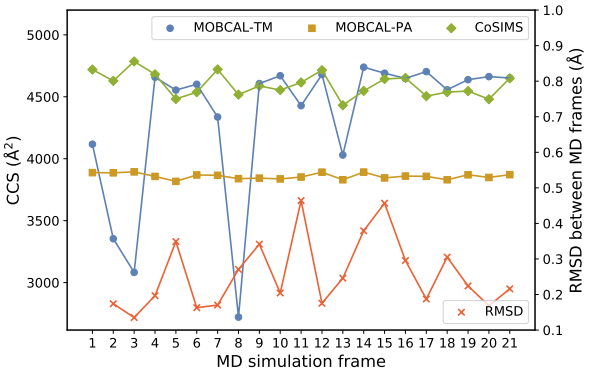
set of structures. For remaining 64 base pair, B-form strands, including the one used in the main article, are shown in Figure S6.



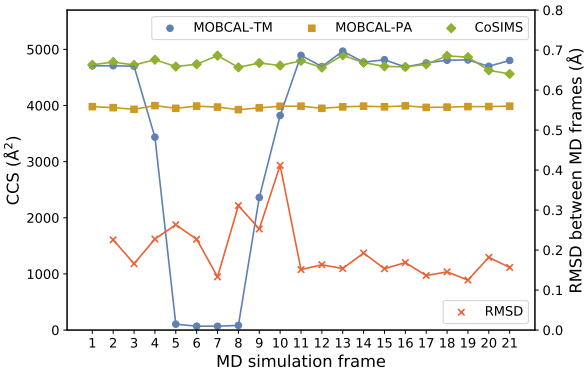
(a) Neutral, B-form, 48 base pair, -10 charge state



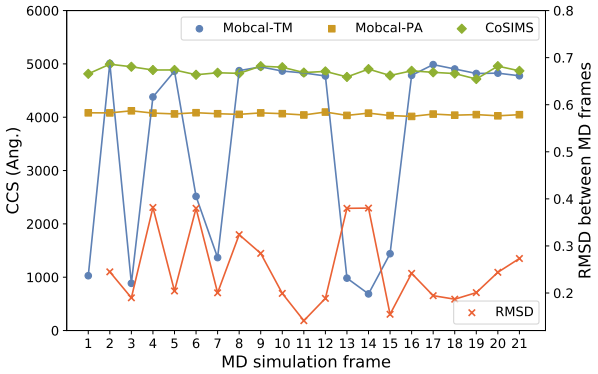
(b) Neutral, B-form, 48 base pair, -11 charge state



(c) Neutral, B-form, 64 base pair, -12 charge state



(d) Neutral, B-form, 64 base pair, -13 charge state



(e) Neutral, B-form, 64 base pair, -14 charge state

Figure S6: Stability of CoSIMS vs. MOBCAL and their relation to the RMSD difference between MD simulation frames.

S7 Tabulated Data

In this section, the data for the various plots used in the main article are presented. A description of the data and what figure it is used in is given in the caption of each table.

Molecule	CoSIMS (min)	CoSIMS - DC (min)	MOBCAL (min)	MOBCAL/ CoSIMS	MOBCAL/ CoSIMS-DC
6-A (-3)	4.22 ± 2.0%	3.50 ± 2.11%	50.4 ± 23.0%	11.92	14.37
10-A (-5)	7.58 ± 1.5%	4.64 ± 2.10%	62.9 ± 1.4%	8.30	13.56
16-A (-5)	12.24 ± 1.6%	5.91 ± 1.33%	106.1 ± 1.2%	8.66	17.94
16-A (-6)	11.81 ± 1.8%	5.15 ± 1.58%	96.7 ± 1.5%	8.19	18.78
16-A (-7)	11.86 ± 5.3%	4.70 ± 8.33%	91.9 ± 7.4%	7.75	19.55
18-A (-5)	11.63 ± 1.7%	5.27 ± 1.63%	98.3 ± 1.4%	8.45	18.64
22-A (-11)	10.42 ± 3.0%	2.76 ± 1.76%	81.5 ± 0.6%	7.82	29.51
22-A (-7)	14.27 ± 2.5%	4.45 ± 1.38%	106.8 ± 2.4%	7.49	24.03
26-A (-13)	9.85 ± 8.3%	2.20 ± 5.41%	133.3 ± 18.0%	13.53	60.72
26-A (-7)	17.04 ± 2.8%	5.12 ± 2.27%	139.4 ± 3.0%	8.18	27.23
30-A (-7)	22.61 ± 3.1%	5.31 ± 2.71%	148.2 ± 1.8%	6.56	27.89
30-A (-8)	15.79 ± 4.8%	4.47 ± 2.03%	135.2 ± 0.9%	8.56	30.26
32-A (-8)	16.87 ± 2.5%	4.38 ± 1.49%	153.8 ± 18.4%	9.12	35.14
32-A (-9)	20.66 ± 5.4%	4.17 ± 4.15%	141.4 ± 1.9%	6.84	33.90
48-A (-10)	38.91 ± 7.2%	5.33 ± 8.73%	221.3 ± 5.3%	5.69	41.51
48-A (-11)	29.11 ± 7.9%	3.67 ± 1.96%	191.2 ± 1.0%	6.57	52.04
64-A (-12)	36.53 ± 9.9%	4.00 ± 4.73%	255.1 ± 2.0%	6.98	63.72
64-A (-13)	38.72 ± 8.7%	4.23 ± 3.27%	265.2 ± 1.7%	6.85	62.68
64-A (-14)	28.74 ± 11.0%	3.08 ± 3.53%	239.2 ± 1.2%	8.32	77.65
6-B (-3)	4.31 ± 1.80%	3.46 ± 1.64%	37.6 ± 2.0%	8.73	10.88
10-B (-5)	6.62 ± 2.06%	3.55 ± 2.34%	55.0 ± 2.5%	8.31	15.50
16-B (-5)	8.35 ± 3.55%	2.62 ± 2.86%	67.7 ± 3.9%	8.11	25.82
16-B (-6)	10.71 ± 7.03%	3.41 ± 8.42%	81.5 ± 5.6%	7.61	23.88
16-B (-7)	7.78 ± 2.43%	2.68 ± 1.96%	61.4 ± 1.9%	7.89	22.86
18-B (-5)	9.68 ± 1.84%	3.00 ± 2.24%	72.8 ± 1.9%	7.52	24.27
22-B (-11)	8.36 ± 4.84%	1.92 ± 2.95%	75.9 ± 2.2%	9.09	39.59
22-B (-7)	8.51 ± 7.74%	2.21 ± 3.55%	76.2 ± 1.7%	8.96	34.49
26-B (-13)	8.65 ± 8.55%	1.81 ± 2.55%	79.7 ± 3.6%	9.22	44.00
26-B (-7)	10.09 ± 17.30%	2.11 ± 7.54%	87.5 ± 2.0%	8.67	41.53
30-B (-7)	14.46 ± 7.04%	2.35 ± 2.73%	109.5 ± 16.3%	7.57	46.62
30-B (-8)	10.62 ± 11.14%	1.98 ± 4.17%	118.5 ± 23.2%	11.16	59.72
32-B (-8)	14.62 ± 6.89%	2.01 ± 3.79%	114.1 ± 5.1%	7.80	56.77
32-B (-9)	15.66 ± 10.50%	2.10 ± 5.63%	112.7 ± 6.5%	7.20	53.76
48-B (-10)	20.98 ± 19.56%	2.17 ± 8.23%	188.4 ± 37.1%	8.98	86.86
48-B (-11)	21.55 ± 25.61%	2.09 ± 11.10%	265.4 ± 31.8%	12.32	126.7
64-B (-12)	24.61 ± 39.25%	2.40 ± 16.62%	236.8 ± 9.3%	9.62	98.60
64-B (-13)	25.38 ± 27.88%	2.63 ± 8.49%	289.9 ± 37.1%	11.42	110.4
64-B (-14)	26.82 ± 19.13%	2.43 ± 9.07%	308.3 ± 39.2%	11.49	127.1

Table S1: CPU benchmarks for the neutrally charged MD-frames. This data is used in Figures S4a and S4b and in Figure 7

Molecule		No. Atoms	Minimum RMSD (Å)	Maximum RMSD (Å)	CoSIMS (Å^2)			MOBCAL (Å^2)			MOBCAL/ CoSIMS
6-A	(-3)	383	0.509	1.125	483.0	±	1.07%	489.5	±	0.99%	1.01
10-A	(-5)	641	0.522	0.787	720.9	±	0.83%	731.9	±	1.12%	1.02
16-A	(-5)	1039	0.799	1.330	992.0	±	0.88%	1006.7	±	0.89%	1.01
16-A	(-6)	1038	0.809	1.630	1068.1	±	1.45%	1078.7	±	1.36%	1.01
16-A	(-7)	1037	1.092	3.718	1165.5	±	2.25%	1183.2	±	2.07%	1.02
18-A	(-5)	1161	0.558	1.212	1120.6	±	0.97%	1134.4	±	1.02%	1.01
22-A	(-11)	1415	0.656	1.065	1644.5	±	0.83%	1672.2	±	0.85%	1.02
22-A	(-7)	1419	0.607	1.035	1383.1	±	0.99%	1402.1	±	0.78%	1.01
26-A	(-13)	1673	2.626	4.528	2149.5	±	1.25%	2181.0	±	1.43%	1.01
26-A	(-7)	1679	0.702	2.450	1597.3	±	1.07%	1618.1	±	0.94%	1.01
30-A	(-7)	1958	0.768	1.886	1697.1	±	1.06%	1712.3	±	1.09%	1.01
30-A	(-8)	1957	1.046	1.669	1786.9	±	0.93%	1804.9	±	0.91%	1.01
32-A	(-8)	2087	1.044	1.337	1928.5	±	0.96%	1954.3	±	0.64%	1.01
32-A	(-9)	2086	1.103	3.212	1982.7	±	1.07%	2019.6	±	1.08%	1.02
48-A	(-10)	3134	1.274	5.749	2693.3	±	1.19%	2728.0	±	1.41%	1.01
48-A	(-11)	3133	0.937	2.322	2922.4	±	0.97%	2962.4	±	1.27%	1.01
64-A	(-12)	4182	1.097	2.774	3617.2	±	1.20%	3668.4	±	1.06%	1.01
64-A	(-13)	4181	1.198	2.981	3548.2	±	1.19%	3598.6	±	0.87%	1.01
64-A	(-14)	4180	1.219	2.530	3852.2	±	0.83%	3905.6	±	0.95%	1.01
6-B	(-3)	383	0.655	1.098	496.3	±	1.62%	502.3	±	1.61%	1.01
10-B	(-5)	641	0.825	1.447	820.4	±	1.44%	831.2	±	1.16%	1.01
16-B	(-5)	1039	0.898	2.040	1258.8	±	1.68%	1280.0	±	1.32%	1.02
16-B	(-6)	1038	1.339	4.159	1245.3	±	1.79%	1259.5	±	1.89%	1.01
16-B	(-7)	1037	0.762	1.481	1248.0	±	1.02%	1267.2	±	1.09%	1.02
18-B	(-5)	1161	0.849	1.464	1368.1	±	0.85%	1388.2	±	1.04%	1.01
22-B	(-11)	1415	1.391	2.573	1720.7	±	1.39%	1728.4	±	0.78%	1.00
22-B	(-7)	1419	0.711	1.704	1685.6	±	1.23%	1713.7	±	0.97%	1.02
26-B	(-13)	1673	0.721	1.495	1987.0	±	0.97%	2008.1	±	1.24%	1.01
26-B	(-7)	1679	1.556	5.807	2011.6	±	1.36%	2032.5	±	0.99%	1.01
30-B	(-7)	1958	1.319	2.123	2264.5	±	1.44%	2300.8	±	0.94%	1.02
30-B	(-8)	1957	1.127	2.665	2379.4	±	1.57%	2393.1	±	1.07%	1.01
32-B	(-8)	2087	0.891	3.021	2409.8	±	1.35%	2422.8	±	2.99%	1.01
32-B	(-9)	2086	1.325	2.464	2399.6	±	1.20%	2406.8	±	3.70%	1.00
48-B	(-10)	3134	1.951	6.268	3574.0	±	1.52%	3245.5	±	28.29%	0.91
48-B	(-11)	3133	1.880	5.107	3623.6	±	1.69%	2115.2	±	58.32%	0.58
64-B	(-12)	4182	1.667	6.182	4597.1	±	2.01%	4339.5	±	13.00%	0.94
64-B	(-13)	4181	1.948	4.278	4747.3	±	1.79%	3655.4	±	50.10%	0.77
64-B	(-14)	4180	2.450	4.930	4861.2	±	1.44%	3645.0	±	47.00%	0.75

Table S2: CCS calculations for the neutrally charged MD-frames used in Figures 5a and 5b. Minimum and maximum RMSD values are calculated over all frames in the ensemble with respect to the first frame.

Protein	CoSIMS (Å^2) Exact Potential	CoSIMS (Å^2) $R_d = 30$ Å $R_c = 3$ Å	CoSIMS (Å^2) $R_d = 30$ Å $R_c = 6$ Å	CoSIMS (Å^2) $R_d = 25$ Å $R_c = 6$ Å	MOBCAL (Å^2) Exact Potential
2m1k	2669.1 ± 0.21%	2671.0 ± 0.23%	2671.8 ± 0.21%	2671.7 ± 0.22%	2708.9 ± 0.71%
2mwg	5394.8 ± 0.41%	5387.9 ± 0.40%	5396.4 ± 0.41%	5390.5 ± 0.40%	5443.0 ± 0.35%
2mz6	846.8 ± 0.52%	846.8 ± 0.52%	846.8 ± 0.51%	846.8 ± 0.52%	863.8 ± 0.55%
4p3v	1137.9 ± 0.24%	1137.9 ± 0.23%	1138.2 ± 0.24%	1137.5 ± 0.24%	1156.9 ± 0.38%
4r06	4313.9 ± 0.41%	4314.0 ± 0.39%	4313.5 ± 0.40%	4315.0 ± 0.40%	4339.8 ± 0.65%
4r8z	3739.6 ± 0.20%	3741.7 ± 0.20%	3742.4 ± 0.20%	3742.9 ± 0.20%	3805.4 ± 0.41%
4rna	3251.5 ± 0.30%	3250.6 ± 0.30%	3250.6 ± 0.29%	3251.3 ± 0.30%	3315.6 ± 0.55%
4xmq	4570.1 ± 0.21%	4570.1 ± 0.23%	4569.1 ± 0.22%	4567.6 ± 0.25%	4633.9 ± 0.29%
4zn8	2314.8 ± 0.18%	2312.9 ± 0.18%	2313.9 ± 0.18%	2313.5 ± 0.19%	2343.0 ± 0.44%
5arm	1563.4 ± 0.56%	1563.8 ± 0.55%	1564.0 ± 0.55%	1563.8 ± 0.55%	1607.0 ± 0.46%
5b6o	4835.0 ± 0.24%	4833.6 ± 0.24%	4833.9 ± 0.24%	4830.5 ± 0.26%	4864.3 ± 0.42%
5fmn	2202.6 ± 0.49%	2201.9 ± 0.50%	2201.9 ± 0.49%	2200.6 ± 0.50%	2229.4 ± 0.52%
5h6w	4130.1 ± 0.06%	4130.0 ± 0.06%	4128.2 ± 0.06%	4128.7 ± 0.05%	4186.5 ± 0.44%
5ieu	2390.4 ± 0.26%	2391.2 ± 0.26%	2391.5 ± 0.25%	2390.3 ± 0.25%	2418.5 ± 0.53%
5iew	1178.9 ± 0.53%	1178.8 ± 0.53%	1178.7 ± 0.53%	1178.8 ± 0.53%	1195.8 ± 0.35%
5iir	1184.4 ± 0.65%	1184.3 ± 0.65%	1184.4 ± 0.65%	1184.2 ± 0.64%	1187.5 ± 0.41%
5ilg	3994.0 ± 0.27%	3991.7 ± 0.27%	3992.5 ± 0.26%	3993.1 ± 0.26%	4046.7 ± 0.28%
5jip	5125.2 ± 0.33%	5120.4 ± 0.33%	5118.5 ± 0.32%	5122.4 ± 0.34%	5156.7 ± 0.34%
5lvz	2536.3 ± 0.35%	2535.8 ± 0.35%	2535.6 ± 0.35%	2537.2 ± 0.34%	2571.3 ± 0.55%
5lx4	3427.3 ± 0.47%	3426.7 ± 0.45%	3424.8 ± 0.45%	3425.7 ± 0.45%	3446.8 ± 0.60%
5mus	5229.3 ± 0.37%	5229.1 ± 0.40%	5231.4 ± 0.39%	5230.9 ± 0.40%	5317.7 ± 0.30%
5my9	2629.8 ± 0.39%	2629.3 ± 0.39%	2629.2 ± 0.39%	2630.0 ± 0.38%	2688.0 ± 0.27%
5t95	4218.1 ± 0.60%	4217.6 ± 0.60%	4215.1 ± 0.60%	4216.2 ± 0.60%	4227.0 ± 0.35%
5ujl	1981.0 ± 0.28%	1980.5 ± 0.27%	1980.5 ± 0.26%	1980.8 ± 0.25%	2018.2 ± 0.60%
5x29	3416.0 ± 0.19%	3419.3 ± 0.21%	3410.4 ± 0.21%	3411.2 ± 0.19%	3431.6 ± 0.45%

Table S3: CCS calculations for symmetric proteins with and without the dispersion cutoff approximation for various radii parameters. Here, R_d is the cut-off radius and R_c is the cluster radius. The first column using the exact potential and the last column for MOBCAL contains the data used to generate Figure 3 in the main article. The relative percent difference between the CCS calculated using exact potential and the approximated potential are shown in Figure S5

Protein	CoSIMS (Å^2) Exact Potential	CoSIMS (Å^2) $R_d = 30$ Å $R_c = 3$ Å	CoSIMS (Å^2) $R_d = 30$ Å $R_c = 6$ Å	CoSIMS (Å^2) $R_d = 25$ Å $R_c = 6$ Å	MOBCAL (Å^2) Exact Potential
2mr8	1348.3 ± 0.36%	1348.2 ± 0.36%	1347.8 ± 0.35%	1347.9 ± 0.36%	1347.7 ± 0.44%
2n3u	2973.8 ± 0.29%	2972.6 ± 0.28%	2974.6 ± 0.26%	2972.9 ± 0.26%	3003.0 ± 0.29%
2n44	3452.4 ± 0.69%	3452.1 ± 0.71%	3453.0 ± 0.70%	3453.9 ± 0.70%	3496.9 ± 0.47%
2n9z	830.5 ± 0.26%	830.5 ± 0.26%	830.6 ± 0.26%	830.5 ± 0.25%	747.5 ± 0.45%
4p2b	4494.9 ± 0.53%	4491.2 ± 0.51%	4491.6 ± 0.52%	4493.5 ± 0.49%	4517.1 ± 0.45%
4rlo	4620.8 ± 0.41%	4622.1 ± 0.42%	4624.7 ± 0.42%	4625.6 ± 0.43%	4682.5 ± 0.51%
4ub0	3716.1 ± 0.37%	3717.0 ± 0.35%	3717.0 ± 0.35%	3716.8 ± 0.35%	3726.4 ± 0.55%
4w6f	3956.4 ± 0.38%	3954.7 ± 0.37%	3957.9 ± 0.37%	3954.5 ± 0.36%	4023.6 ± 0.42%
4xol	2004.5 ± 0.36%	2004.9 ± 0.36%	2005.4 ± 0.36%	2004.5 ± 0.37%	385.8 ± 0.91%
4y4q	3350.3 ± 0.31%	3349.4 ± 0.31%	3350.8 ± 0.32%	3349.7 ± 0.31%	3352.6 ± 0.45%
4zpf	3167.7 ± 0.37%	3165.4 ± 0.37%	3165.6 ± 0.36%	3166.6 ± 0.37%	3202.3 ± 0.43%
5bz0	2878.7 ± 0.16%	2877.4 ± 0.17%	2878.5 ± 0.16%	2877.9 ± 0.17%	2897.3 ± 0.39%
5cxv	4346.2 ± 0.27%	4347.6 ± 0.28%	4347.4 ± 0.27%	4348.7 ± 0.28%	4380.0 ± 0.40%
5e21	1344.3 ± 0.36%	1344.5 ± 0.36%	1344.5 ± 0.36%	1344.2 ± 0.36%	1362.5 ± 0.70%
5h7b	5354.6 ± 0.57%	5351.3 ± 0.59%	5351.1 ± 0.57%	5350.0 ± 0.57%	5376.2 ± 0.52%
5imk	2547.5 ± 0.23%	2548.1 ± 0.22%	2548.6 ± 0.23%	2548.7 ± 0.22%	2593.0 ± 0.60%
5jqf	648.2 ± 0.47%	648.2 ± 0.47%	648.2 ± 0.47%	648.3 ± 0.47%	662.4 ± 0.25%
5lr1	2170.9 ± 0.41%	2171.6 ± 0.41%	2171.1 ± 0.42%	2171.7 ± 0.41%	2228.1 ± 0.52%
5lt8	2536.3 ± 0.26%	2536.3 ± 0.27%	2536.3 ± 0.27%	2534.4 ± 0.24%	2549.8 ± 0.43%
5pdo	1596.1 ± 0.31%	1596.1 ± 0.30%	1596.2 ± 0.31%	1595.5 ± 0.30%	1347.9 ± 0.21%
5uk5	5570.4 ± 0.25%	5565.6 ± 0.22%	5565.6 ± 0.22%	5569.5 ± 0.23%	5619.9 ± 0.95%
5viz	749.0 ± 0.44%	749.0 ± 0.44%	749.0 ± 0.44%	748.9 ± 0.44%	756.4 ± 0.48%
5wxg	1086.5 ± 0.39%	1086.4 ± 0.39%	1086.5 ± 0.39%	1086.5 ± 0.39%	1109.7 ± 0.51%
5x3l	473.2 ± 0.36%	473.2 ± 0.36%	473.2 ± 0.36%	473.2 ± 0.36%	479.1 ± 0.45%
5yfg	2290.6 ± 0.50%	2290.3 ± 0.52%	2291.1 ± 0.51%	2289.6 ± 0.51%	2288.1 ± 0.53%

Table S4: CCS calculations for asymmetric proteins with and without the dispersion cutoff approximation for various radii parameters. Here, R_d is the cut-off radius and R_c is the cluster radius. The first column using the exact potential and the last column for MOBCAL contains the data used to generate Figure 3 in the main article. The relative percent difference between the CCS calculated using exact potential and the approximated potential are shown in Figure S6

Protein	Number of Atoms	$\sigma_{rel}(\%)$ $R_d = 30$ Å $R_c = 3$ Å	$\sigma_{rel}(\%)$ $R_d = 30$ Å $R_c = 6$ Å	$\sigma_{rel}(\%)$ $R_d = 25$ Å $R_c = 6$ Å
2m1k	3402	0.073	0.103	0.098
2mwg	8272	0.129	0.030	0.079
2mz6	566	0.000	0.002	0.000
4p3v	1025	0.000	0.025	0.034
4r06	4075	0.002	0.010	0.025
4r8z	3385	0.056	0.075	0.088
4rna	2504	0.027	0.026	0.006
4xmq	4200	0.000	0.022	0.054
4zn8	1735	0.084	0.042	0.058
5arm	965	0.023	0.037	0.025
5b6o	4733	0.029	0.023	0.093
5fmn	1398	0.032	0.032	0.088
5h6w	3858	0.002	0.045	0.034
5ieu	1549	0.034	0.045	0.005
5iew	1096	0.009	0.010	0.003
5iir	1096	0.004	0.002	0.013
5ilg	3941	0.057	0.037	0.022
5jip	9986	0.094	0.131	0.055
5lvz	1869	0.021	0.030	0.033
5lx4	3040	0.015	0.073	0.044
5mus	5151	0.004	0.040	0.030
5my9	1996	0.017	0.021	0.011
5t95	4084	0.010	0.071	0.044
5ujl	4688	0.025	0.027	0.012
5x29	4710	0.095	0.164	0.142

Table S5: CCS accuracy comparisons for various cut-off radius and cluster sizes for symmetric proteins with the dispersion cut-off approximation invoked. Here, $\sigma_{rel} = 100 \times |A - B|/B$ with B being the CCS generated using the exact potential and B is the CCS generated using the approximate potential.

Protein	Number of Atoms	$\sigma_{rel}(\%)$ $R_d = 30$ Å $R_c = 3$ Å	$\sigma_{rel}(\%)$ $R_d = 30$ Å $R_c = 6$ Å	$\sigma_{rel}(\%)$ $R_d = 25$ Å $R_c = 6$ Å
2mr8	4383	0.006	0.031	0.028
2n3u	5734	0.042	0.025	0.030
2n44	652	0.008	0.016	0.045
2n9z	3830	0.005	0.005	0.008
4p2b	8162	0.082	0.072	0.032
4rlo	3262	0.028	0.086	0.105
4ub0	3539	0.024	0.024	0.018
4w6f	1176	0.042	0.038	0.049
4xol	2805	0.020	0.042	0.000
4y4q	2877	0.027	0.015	0.017
4zpf	2545	0.073	0.066	0.035
5bz0	3500	0.045	0.009	0.029
5cxv	1434	0.032	0.026	0.058
5e21	3387	0.014	0.014	0.008
5h7b	3738	0.061	0.066	0.086
5imk	633	0.024	0.044	0.049
5jqf	1646	0.001	0.001	0.005
5lr1	1847	0.032	0.007	0.035
5lt8	943	0.002	0.002	0.073
5pdo	7740	0.001	0.007	0.039
5uk5	3681	0.086	0.086	0.016
5viz	400	0.002	0.001	0.005
5wxg	549	.013	0.007	0.006
5x3l	323	0.000	0.000	0.001
5yfg	2858	0.011	0.020	0.044

Table S6: CCS accuracy comparisons for various cut-off radius and cluster sizes for asymmetric proteins with the dispersion cut-off approximation invoked. Here, $\sigma_{rel} = 100 \times |A - B|/B$ with B being the CCS generated using the exact potential and B is the CCS generated using the approximate potential.

Protein	No. of Atoms	Radius of Gyration	CoSIMS (min) Exact	CoSIMS (min) $R_d = 30$Å $R_c = 3$Å	CoSIMS (min) 30Å 6Å	CoSIMS (min) 25Å 6Å	MOBCAL (min)	Ratio to MOBCAL			
			A	B	C	D	E	E/A	E/B	E/C	E/D
2mz6	566	12.2	7.6	5.0	5.8	4.2	41.5	5.4	8.2	7.2	9.8
4p3v	1025	14.0	13.6	8.2	9.5	7.0	95.7	7.0	11.6	10.1	13.7
5iir	1096	14.9	10.0	6.3	7.2	5.5	91.4	9.1	14.5	12.7	16.6
5iew	1096	14.9	9.5	6.1	6.9	5.3	84.6	8.9	13.8	12.3	15.9
5arm	965	15.8	10.8	5.5	6.4	4.5	83.5	7.8	15.2	13.0	18.5
4zn8	1735	17.8	27.3	10.9	12.4	8.3	201.2	7.4	18.5	16.2	24.4
5ieu	1549	18.4	22.5	7.2	8.3	5.3	116.6	5.2	16.1	14.0	21.8
5ujl	4688	19.1	53.7	22.0	25.8	17.9	471.8	8.8	21.4	18.3	26.4
5lvz	1869	19.3	29.3	8.8	10.2	6.5	147.0	5.0	16.7	14.5	22.6
5my9	1996	19.5	34.0	9.6	11.3	7.2	178.5	5.2	18.7	15.9	24.9
5fmn	1398	20.1	16.4	6.0	6.9	4.7	111.9	6.8	18.5	16.1	24.0
2m1k	3402	20.7	66.5	13.7	15.3	9.5	229.0	3.4	16.7	14.9	24.1
5x29	4710	21.5	98.6	29.9	33.4	22.2	641.8	6.5	21.5	19.2	28.9
5lx4	3040	22.4	52.3	9.0	9.4	5.9	202.7	3.9	22.5	21.5	34.4
4r8z	3385	23.5	47.9	12.6	13.8	8.8	348.3	7.3	27.5	25.2	39.4
5ilg	3941	23.8	56.5	14.3	15.5	10.0	416.7	7.4	29.0	26.8	41.7
4rna	2504	24.3	38.0	8.2	8.9	5.7	196.7	5.2	24.1	22.0	34.5
5h6w	3858	24.6	67.9	17.4	18.2	11.8	434.0	6.4	25.0	23.8	36.8
4r06	4075	24.8	69.6	15.6	16.4	10.4	435.0	6.3	28.0	26.6	41.9
5b6o	4733	25.8	105.0	20.0	20.3	13.0	557.0	5.3	27.8	27.4	42.9
4xmq	4200	26.1	61.6	14.2	14.9	9.7	419.2	6.8	29.6	28.1	43.3
5mus	5151	27.5	92.9	19.7	20.1	13.2	649.3	7.0	32.9	32.2	49.1
5jip	9986	27.5	168.4	31.7	34.5	21.5	1101.2	6.5	34.7	31.9	51.1
5t95	4084	29.1	48.6	10.5	11.5	7.4	377.0	7.8	35.9	32.7	50.8
2mwg	8272	32.8	113.6	21.8	23.1	15.2	832.2	7.3	38.2	36.1	54.7

Table S7: CPU benchmark comparisons for various cut-off radius and cluster sizes for symmetric proteins with the dispersion cut-off approximation invoked. These are instead sorted by increasing CPU ratio. The data used in Columns A, D, and E was used to generate Figure 4 of the main text.

Protein	No. of Atoms	Radius of Gyration	CoSIMS (min) Exact	$R_d = 30$Å $R_c = 3$Å	30Å 6Å	25Å 6Å	MOBCAL (min)	Ratio to MOBCAL			
			A	B	C	D	E	E/A	E/B	E/C	E/D
5jqf	1646	9.53	6.7	6.3	6.2	5.2	61.4	9.2	9.8	9.9	11.7
5viz	400	9.76	4.8	4.0	4.3	3.4	36.5	7.6	9.2	8.4	10.9
5x3l	323	9.88	2.6	2.2	2.3	2.1	22.9	8.9	10.3	9.9	11.1
2n9z	3830	11.54	7.3	5.3	5.9	4.4	59.4	8.2	11.1	10.1	13.6
5wxg	549	12.47	6.7	4.2	4.8	3.3	38.8	5.8	9.2	8.1	11.9
5e21	3387	13.40	20.7	11.5	11.5	7.0	104.5	5.1	9.1	9.1	15.0
2mr8	4383	13.61	16.5	8.8	10.1	6.4	103.0	6.2	11.7	10.2	16.1
5pdo	7740	15.88	14.0	4.7	5.6	3.2	123.5	8.8	26.1	22.2	38.3
5lr1	1847	16.86	27.3	10.6	12.1	7.2	180.3	6.6	17.0	14.9	25.1
5yfg	2858	17.47	45.4	16.8	19.5	11.2	262.7	5.8	15.6	13.4	23.5
5lt8	943	18.81	27.6	10.4	10.3	6.1	162.1	5.9	15.6	15.7	26.5
5bz0	3500	19.49	41.3	12.7	14.0	8.2	270.2	6.5	21.3	19.3	33.1
5imk	633	19.88	47.8	16.4	18.7	11.1	321.8	6.7	19.6	17.2	28.9
2n3u	5734	20.38	75.5	21.9	24.8	14.4	441.3	5.8	20.1	17.8	30.7
4zpf	2545	20.69	46.0	12.7	14.0	8.1	296.5	6.4	23.3	21.2	36.5
4y4q	2877	22.11	42.1	11.0	12.2	7.0	300.3	7.1	27.3	24.6	42.7
2n44	652	22.62	86.1	24.5	26.5	15.4	600.8	7.0	24.5	22.7	39.1
4ub0	3539	24.04	47.5	13.6	13.0	7.7	350.7	7.4	25.8	27.0	45.5
4w6f	1176	25.88	48.3	11.6	12.5	7.3	346.0	7.2	29.9	27.6	47.6
4xol	2805	27.26	10.7	3.5	3.9	2.5	101.6	9.5	29.1	25.9	40.2
5cxv	1434	28.48	49.4	10.0	10.5	6.1	334.8	6.8	33.4	31.9	54.9
4rlo	3262	28.86	122.2	22.1	23.9	13.6	789.5	6.5	35.7	33.0	58.0
4p2b	8162	30.50	57.2	9.5	9.9	5.8	345.3	6.0	36.4	35.0	60.0
5h7b	3738	38.44	69.5	8.8	8.3	5.3	237.7	3.4	26.9	28.6	44.8
5uk5	3681	45.51	35.7	4.7	4.6	2.9	218.7	6.1	46.6	47.8	76.0

Table S8: CPU benchmark comparisons for various cut-off radius and cluster sizes for asymmetric proteins with the dispersion cut-off approximation invoked. These are instead sorted by increasing CPU ratio. The data used in Columns A, D, and E was used to generate Figure 4 of the main text.

S8 Comparison to Other TM Based Software

In this section, we compare CoSIMS to 3 other recently published trajectory method CCS software; IMoS[1], Collidoscope[2], and HPCCS[3]. Each program was run on the same CPU architecture as noted in the main text. Only A-form DNA strands were used, as MOBCAL provided the least amount of errors in terms of CCS accuracy (see the main text and Figure 5 for more details).

Each program was run with their settings kept at the default values except for the number of trajectories. An additional change was also made for Collidoscope (see below) in order to generate an error bar for the program's CCS. A total of 10 CCS integrals were computed with each software with a chosen number of total trajectories as to achieve a similar average percent error over all 19 different strands. Collidoscope did not have any direct way of choosing the number of trajectories. The total number of trajectories and their average percent error for each program is as follows:

Software	Number of Trajectories	Average Percent Error
MOBCAL	100,000	0.60 %
CoSIMS	250,000	0.56 %
IMoS	100,000	0.56 %
Collidoscope	280,000-560,000 depending on the molecule	0.98 %
HPCCS	1,000,000	0.89 %

The total CPU time required to run 10 CCS calculations on a single core and on 16 cores is presented in Figures S7 and S8 and the average CCS is compared to CoSIMS in Figures S9-S11. IMoS and HPCCS are Monte-Carlo based programs, so no changes to their settings needed to be made except for the number of trajectories. Collidoscope uses a Riemann sum for their integration and each program call produces the same integral evaluation for a given molecule. To measure the error between each integral evaluation, their spherical "vantage points" were chosen randomly instead of using a discrete spherical grid provided by the program. All other settings for Collidoscope were kept as default.

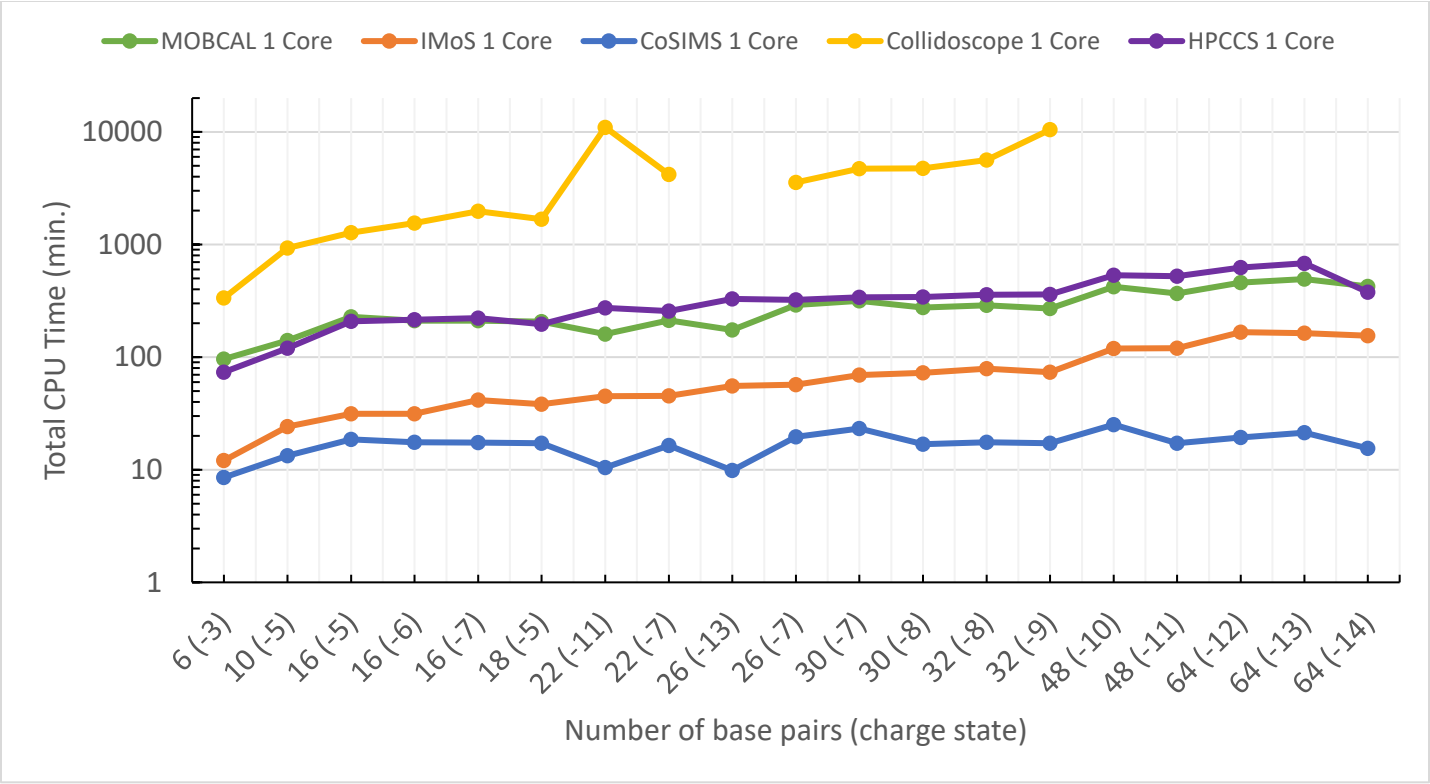


Figure S7: CPU runtimes for all five trajectory method programs tested run on a single CPU core. The gaps in the graph for Collidoscope are due to stopping the program after 24 hours of runtime.

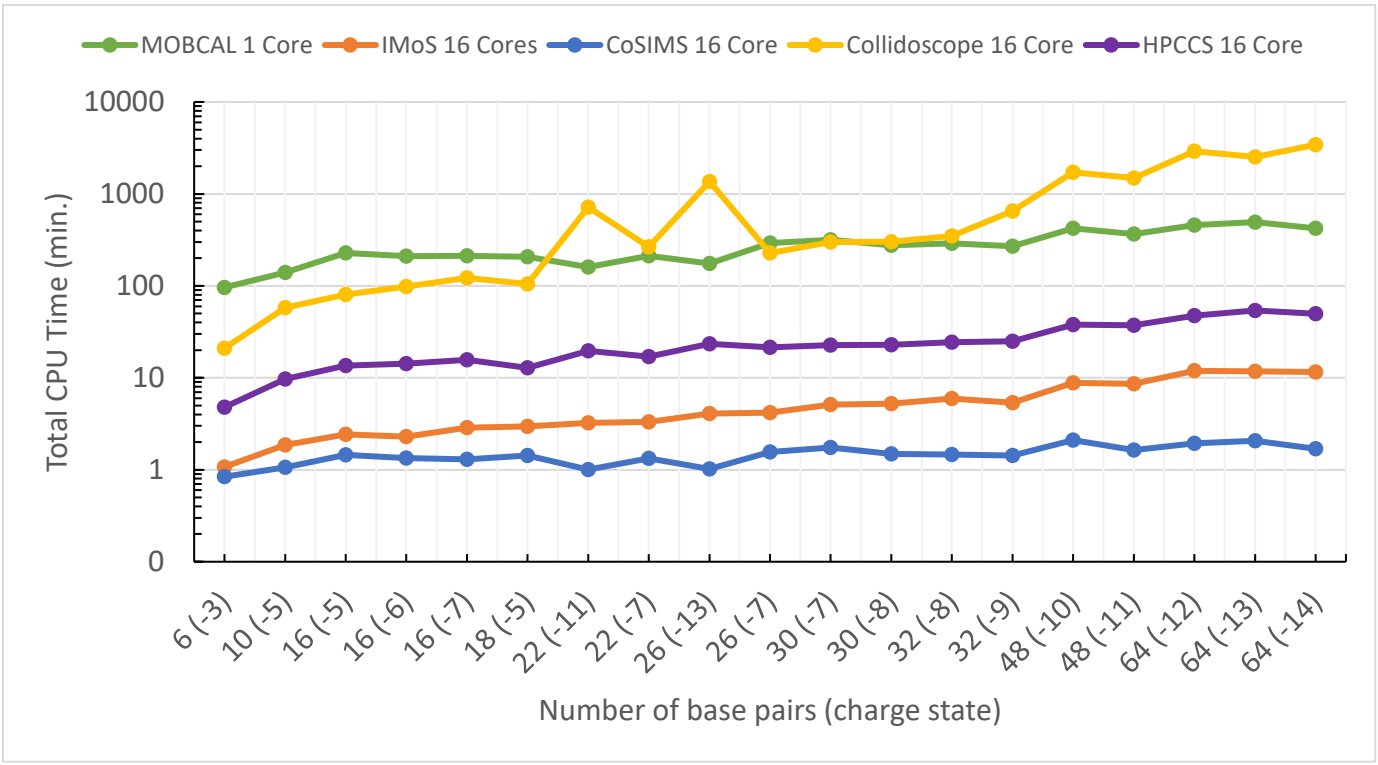


Figure S8: CPU runtimes for all five trajectory methods programs tested run on 16 CPU cores, with the exception of MOBCAL which can only run on one core.

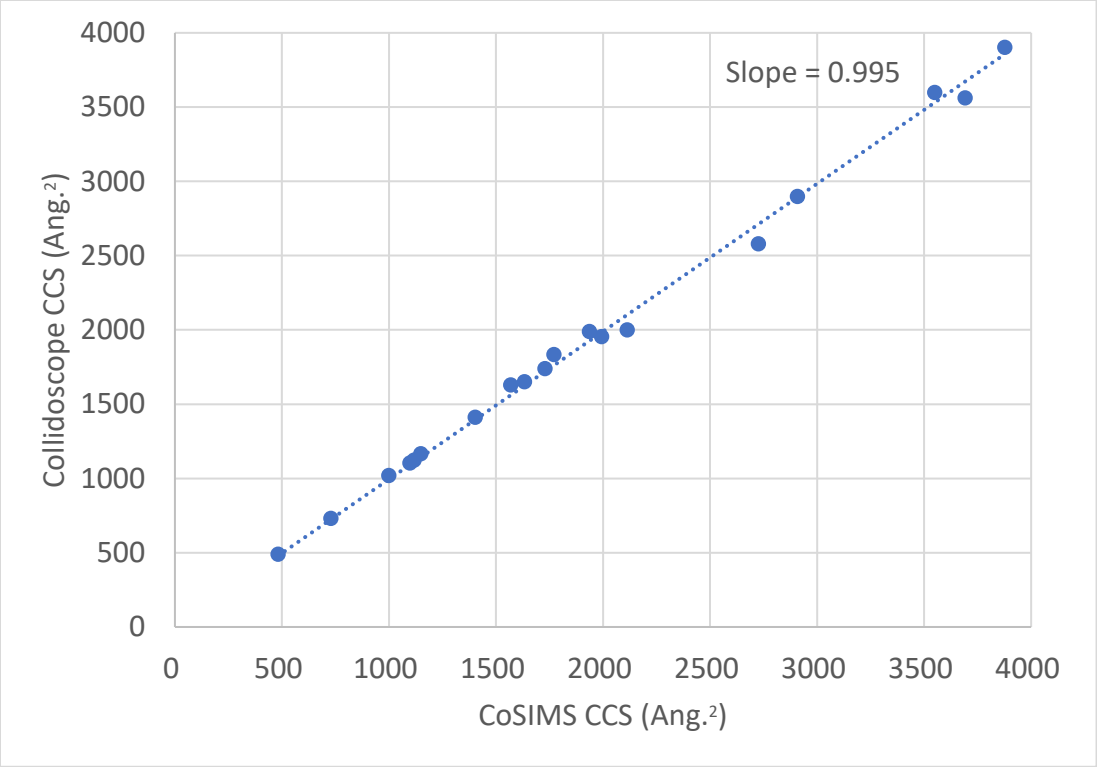


Figure S9: CoSIMS CCS vs Collidoscope CCS for A-form DNA strands.

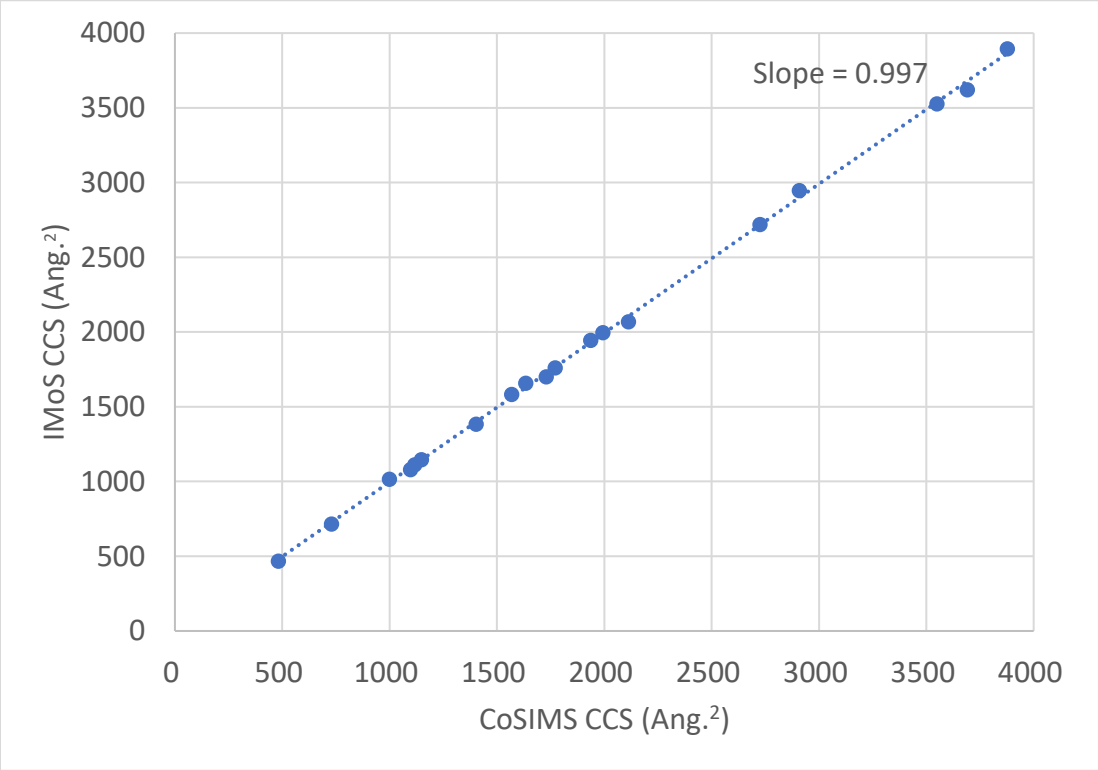


Figure S10: CoSIMS CCS vs IMoS CCS for A-form DNA strands.

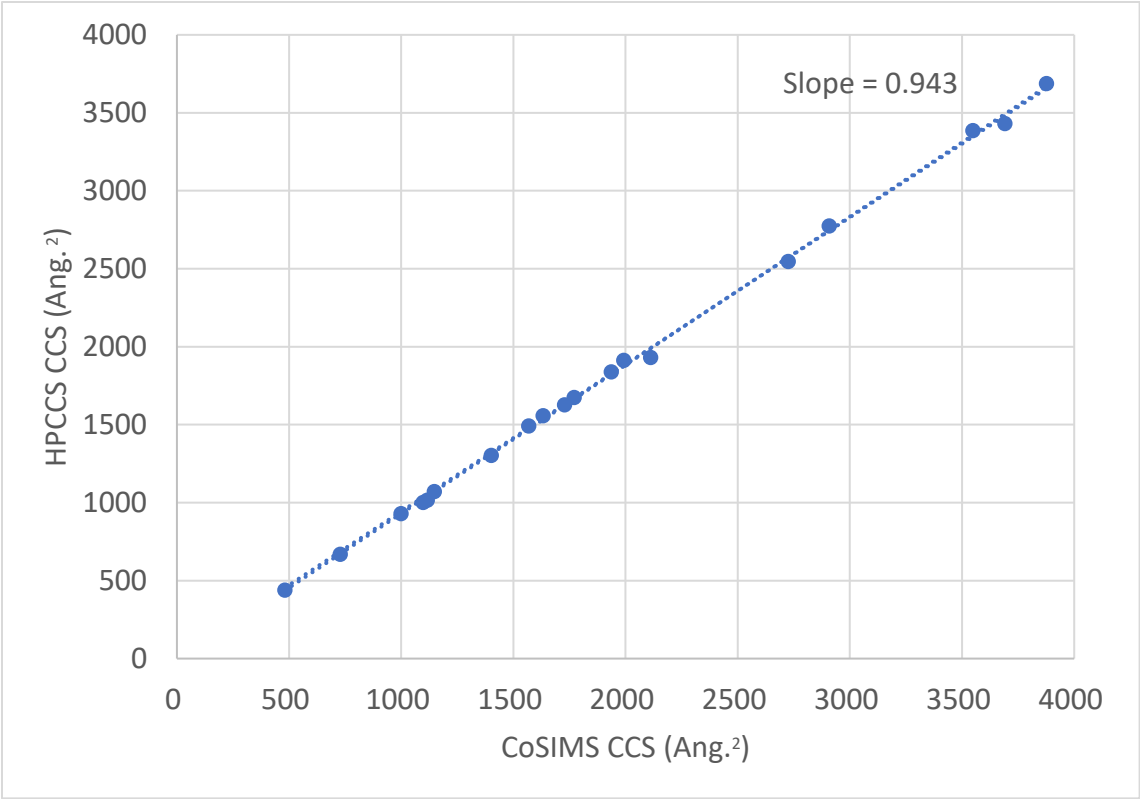


Figure S11: CoSIMS CCS vs HPCCS CCS for A-form DNA strands.

References

1) Larriba, C.; Hogan, C. J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *J. Comput. Phys.* 2013, *251*, 344 – 363.

2) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* 2017, *28*, 587-596.

3) Zanotto, L.; Heerdt, G.; Souza, P. C. T.; Araujo, G.; Skaf, M. S. High Performance Collision Cross Section Calculation-HPCCS. *J. Comput. Chem.* 2018, *39*, 1675-1681.