Supporting Information

Role of Base Sequence Context in Conformational Equilibria and Nucleotide Excision Repair of Benzo[*a*]pyrene Diol Epoxide Adenine Adducts

Shixiang Yan,[†] Min Wu,[†] Tonko Buterin,[‡] Hanspeter Naegeli,[‡] Nicholas E. Geacintov,[*,†] and Suse Broyde[*,§]

[†] *Department of Chemistry, New York University, New York, NY 10003, USA*

[‡] *Institute of Pharmacology and Toxicology, University of Zürich-Tierspital, CH-8008, Zürich, Switzerland*

[§] *Department of Biology, New York University, New York, NY 10003, USA*

Running Title: Base Sequence Effects in BP-dA Adducts

Abbreviations Used: (+)-*anti*-BPDE, (+)-(7*R*,8*S*,9*S*,10*R*)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene; (−)-*anti*-BPDE, (−)-(7*S*,8*R*,9*R*,10*S*)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene; (+)-*syn*-BPDE, (+)-(7*S*,8*R*,9*S*,10*R*)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene; BP, benzo[*a*]pyrene; BPDE, benzo[*a*]pyrene diol epoxide; DNA, deoxyribonucleic acid; MD, molecular dynamics; MM-PBSA, molecular mechanics Poisson–Boltzmann surface area; NER, nucleotide excision repair; NMR, nuclear magnetic resonance; PAH, polycyclic aromatic hydrocarbon; RESP, restrained electrostatic potential fitting; RMSD, root-mean-square deviation; SASA, solvent-accessible surface area

This research is supported by NIH Grant CA-28038 to S.B., NIH Grant CA-76660 to N.E.G., and Swiss National Science Foundation Grant 31-61494.00 to H.N.

[*] To whom correspondence should be addressed. (N.E.G.) Telephone: 212-998-8407. Fax: 212-998-8421. Email: `ng1@nyu.edu`; (S.B.) Telephone: 212-998-8231. Fax: 212-995-4015. Email: `broyde@nyu.edu`.

Table S1: Distortion Free Energy Analysis of the $10S$ (+) and $10R$ (−) Adducts in the CA*C Sequence Context[a]

	$10S$ (+) (*anti*)	$10S$ (+) (*syn*)	$10R$ (−)
ΔE^d_{int}	6.1	−1.1	−3.7
ΔE^d_{vdW}	29.4	30.7	30.3
$\Delta E^d_{\text{electrostatic}}$	−193.4	−100.0	−123.8
ΔG^d_{PB}	178.4	94.7	110.0
$\Delta G^d_{\text{nonpolar}}$	1.2	1.0	1.2
$\Delta E^d_{\text{electrostatic}} + \Delta G^d_{\text{PB}}$	−15.0	−5.3	−13.9
$\Delta G^{\text{distort}}$	21.8	25.4	13.9

[a] All energies are in kcal/mol.

Table S2: Comparision of Structural Parameters for CA*C and CA*A Sequence Contexts Near Lesion Site

	CA*C		CA*A	
	$10S$ (+) (*anti*)[a]	$10R$ (−)	$10S$ (+)	$10R$ (−)
Rise (Å)	8.0	8.3	7.6	8.6
Unwinding (°)	28	20	41[b]	29
Quality of Watson–Crick hydrogen bonding[c]	560	316	599	363
Roll (°)	30	4	13	9
Trajectory average bend angle (°)	60[d]	22	25	23

[a] *syn* conformers are not amenable to reliable helical parameter calculations which depend on Watson–Crick base pairing (*73*, *74*).

[b] 26° at intercalation pocket, 15° at adjacent site stemming from steric effect of BP hydroxyl groups (*36*).

[c] Computed as detailed in previous work (*36*). A value of 0 represents ideal Watson–Crick base pairs. In *syn* conformers this index is in the range of 5000 or higher.

[d] 25° in *syn* conformer.

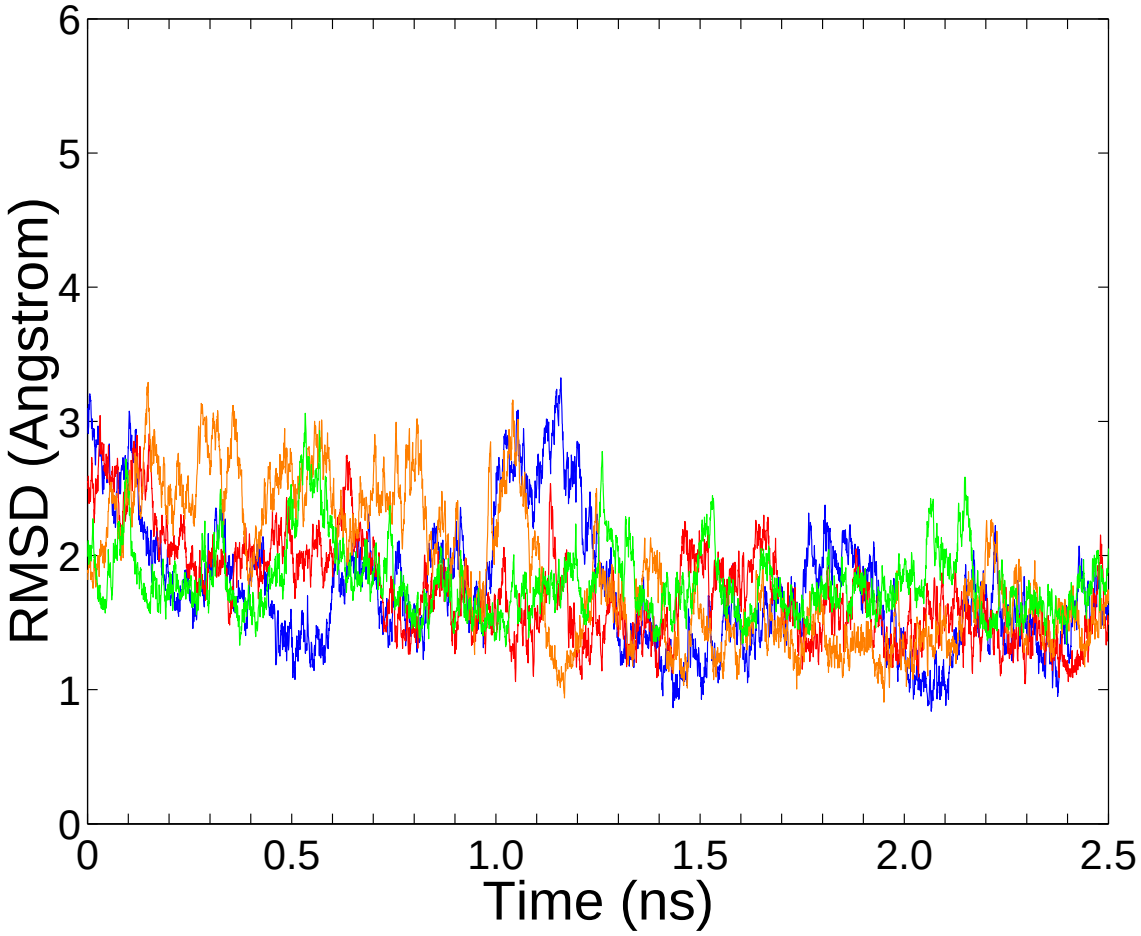


Figure S1: Root mean sequare deviations (RMSDs) for the 10*S* (+) adduct *anti* conformer (red), 10*S* (+) adduct *syn* conformer (orange), 10*R* (−) adduct (blue), and the unmodified control structure (green) in the CA*C sequence context over the 2.5-ns production MD simulation. The RMSDs were computed relative to the respective average structures over 1–2.5 ns.

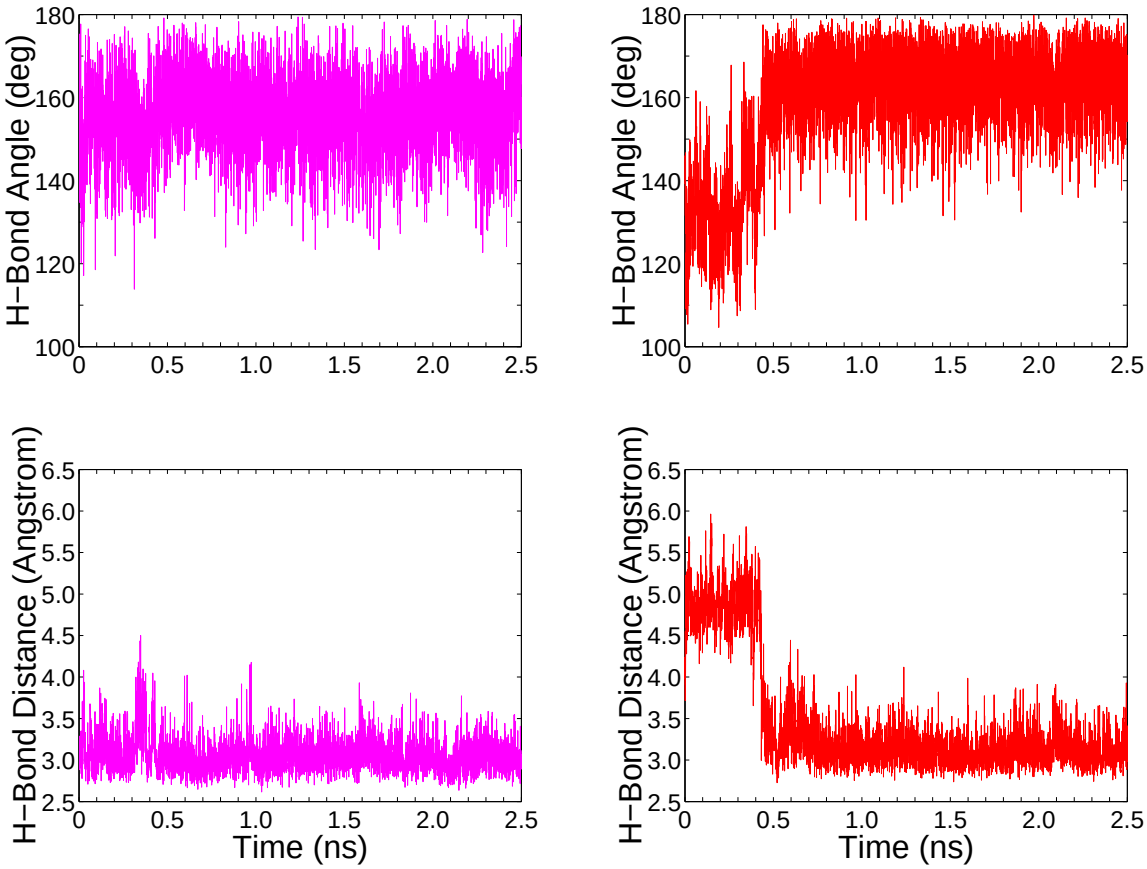


Figure S2: Watson–Crick hydrogen bond angles and distances (heavy atom to heavy atom) for the A*6–T17 base pair of the 10*S* (+) adduct *anti* conformer in the CA*C sequence context over the 2.5-ns production MD simulation. The angles and distances for the N^6–H6 (A*6)$\cdots$O4 (T17) hydrogen bond are in magenta; the angles and distances for the N3–H3 (T17)$\cdots$N1 (A*6) hydrogen bond are in red.

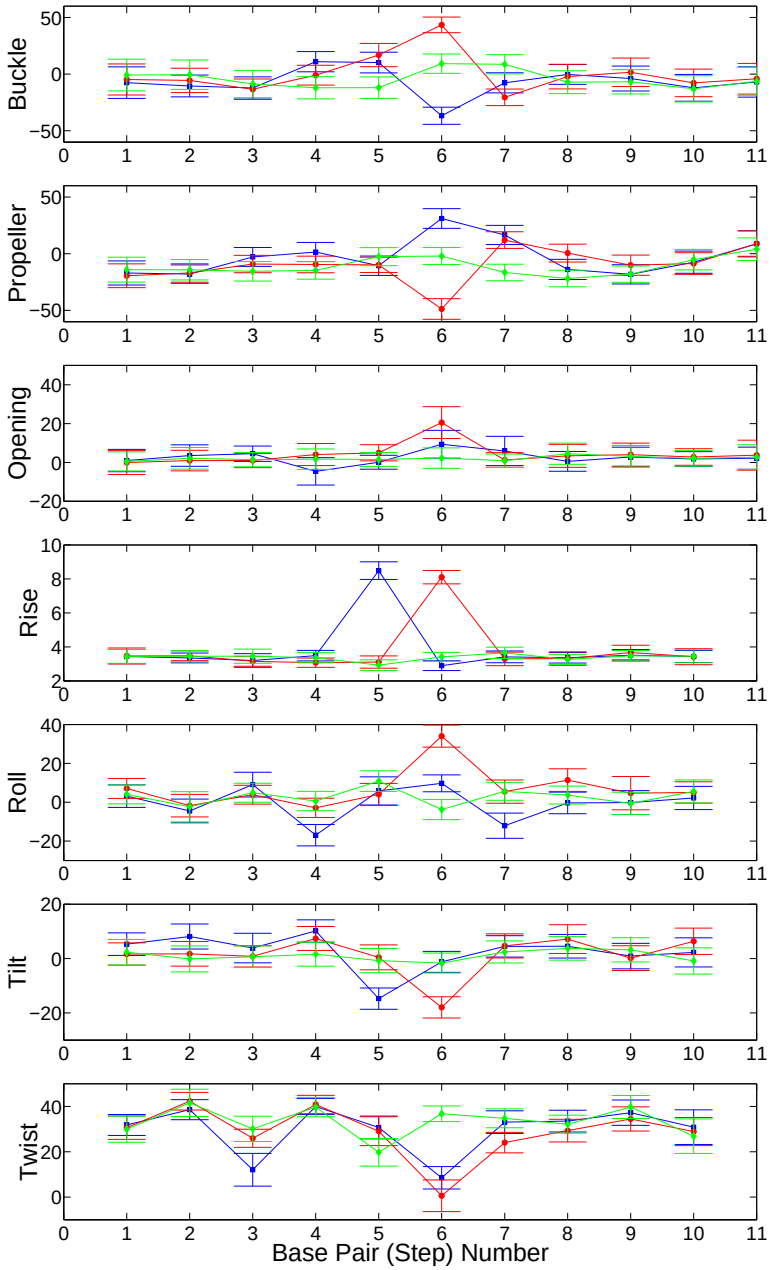


Figure S3: Average helicoidal parameters for the structures of the 10*S* (+) adduct *anti* conformer (red circles), 10*R* (−) adduct (blue squares), and the unmodified control duplex (green diamonds) over 1–2.5 ns (3000 structures). The standard deviations are shown as error bars. The numbering scheme for the nucleotide base pair steps is that the C1–G22 to G2–C21 is step 1, the T2–A21 to C3–G20 is step 2, ..., and so on.

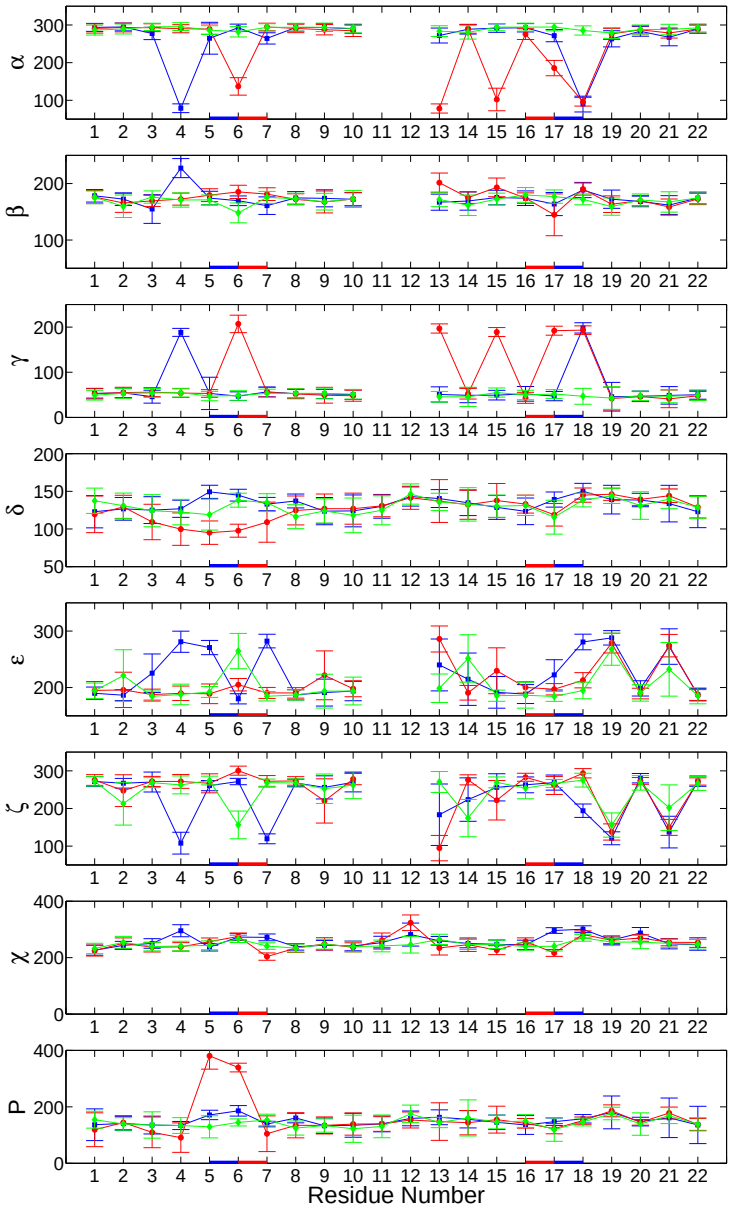


Figure S4: Average backbone torsional parameters for the structures of the 10*S* (+) adduct *anti* conformer (red circles), 10*R* (−) adduct (blue squares), and the unmodified control duplex (green diamonds) over 1–2.5 ns (3000 structures). The standard deviations are shown as error bars. The red and blue bars on the axes indicate the intercalation pocket of the 10*S* (+) adduct (A*6–T17 and C7–G16) and the intercalation pocket of the 10*R* (−) adduct (C5–G18 and A*6–T17), respectively. It should be noted that the residue numbers in Dials and Windows (*56*) differ from the IUPAC convention as follows: For α, β and γ, residue numbers 1–10 should be shifted +1, and for ϵ and ζ, residues 13–22 should be shifted −1 to accord with the IUPAC convention.

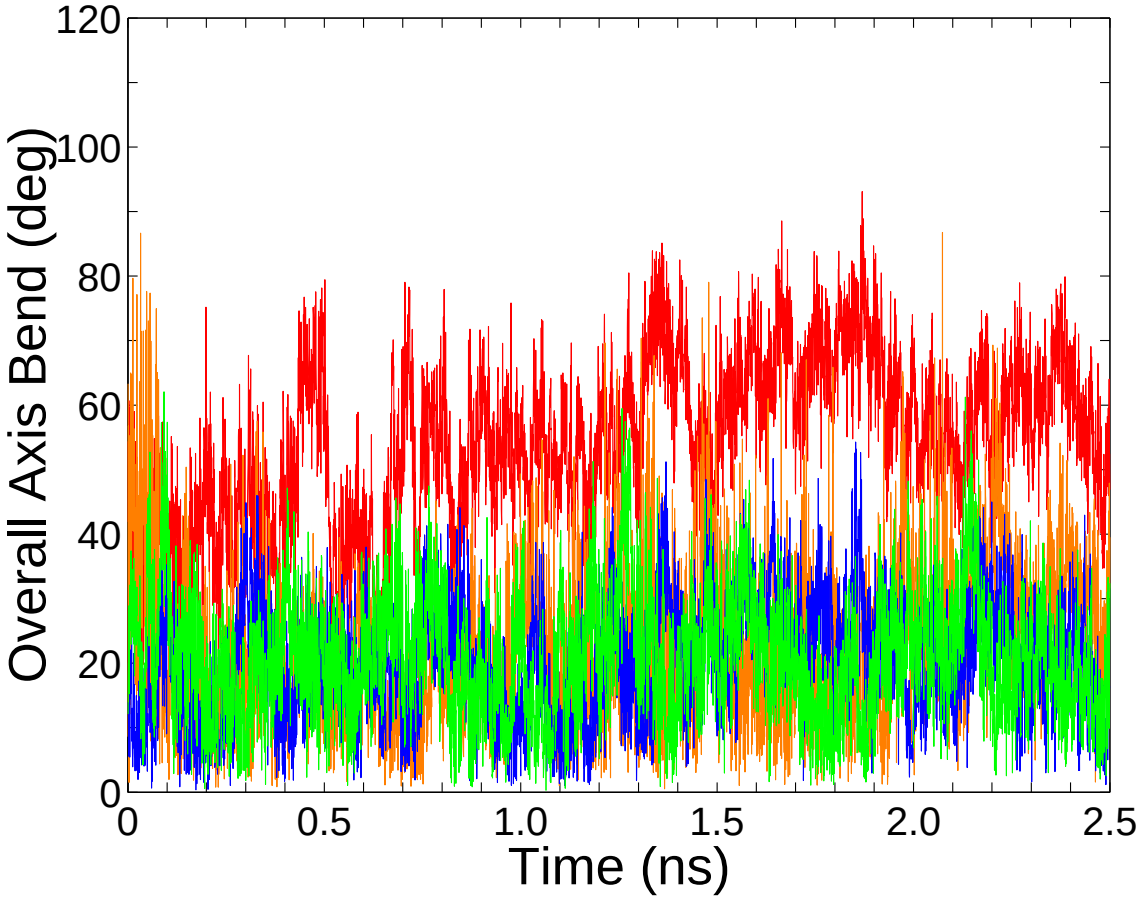
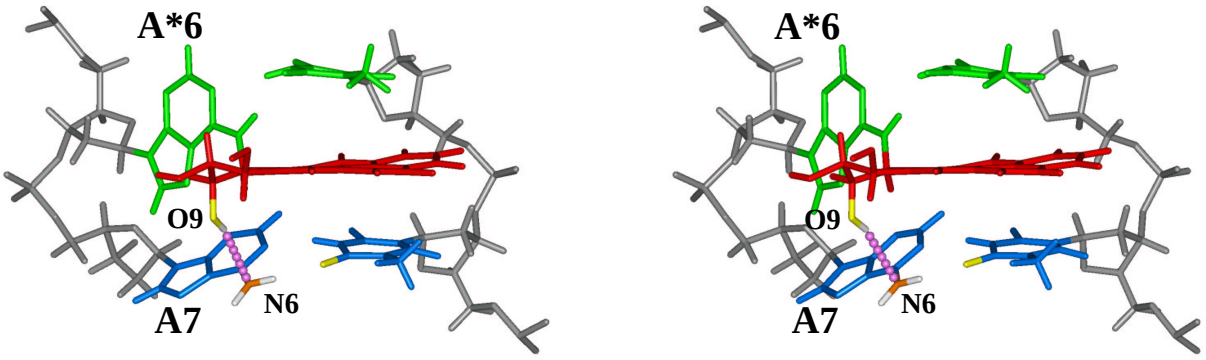


Figure S5: Time dependence of the overall axis bend of the 10*S* (+) adduct *anti* conformer (red), 10*S* (+) *syn* conformer (orange), 10*R* (−) adduct (blue), and the unmodified control structure (green) in the CA*C sequence context over the 2.5-ns production MD simulation.



10*S* (+) Adduct (CA*A, *syn* domain)

Figure S6: Stereo view of representative *syn* glycosidic conformation [$\chi = 47.1°$, see also Figure 8 of Yan *et al.* (*36*)] of the 10*S* (+) adduct [d(A*A)·d(TT)] in the CA*A sequence context (*36*). BP is in red, A*6–T17 in green, and A7–T16 in blue. The backbone and sugar atoms are in grey. The O9 on BP and O4 on T16 are in yellow, and the N^6 on A7 is in orange; the hydrogen atoms of the O9–HO9 hydroxyl group on BP and the N^6 amino group on A7 are in white. The weak hydrogen bond/electrostatic interaction between N^6 (A7) and O9–HO9 (BP) is shown as solid pink dots. The N^6 (A7) to O9 (BP) distance is 3.67 Å.

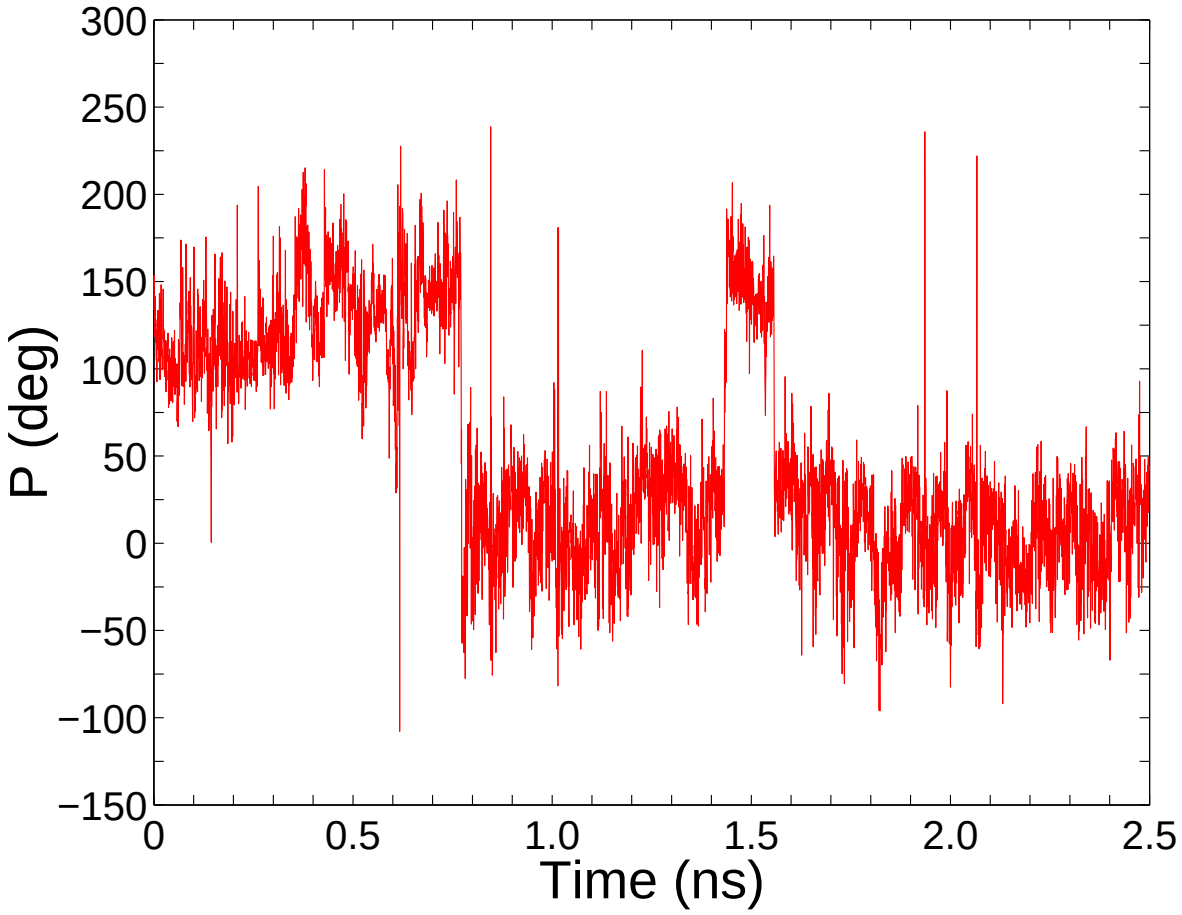


Figure S7: Time dependence of the pseudorotation phase angle P for C5 of the 10S (+) adduct *anti* conformer in the CA*C sequence context over the 2.5-ns production MD simulation.