Conformational Dynamics of the Hepatitis B Virus Pre-genomic RNA on Multiple Time Scales: Implications for Viral Replication

Lukasz T. Olenginski [1], Wojciech K. Kasprzak [2], Christina Bergonzo [3], Bruce A. Shapiro [4] and Theodore K. Dayie [1]*

*1 - **Center for Biomolecular Structure and Organization,** Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA*
*2 - **Basic Science Program,** Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA*
*3 - **Institute for Bioscience and Biotechnology Research,** National Institute of Standards and Technology and University of Maryland, Rockville, MD 20850, USA*
*4 - **RNA Biology Laboratory,** National Cancer Institute, Frederick, MD 21702, USA*

***Correspondence to Theodore K. Dayie:** dayie@umd.edu (T.K. Dayie)*
https://doi.org/10.1016/j.jmb.2022.167633
Edited by Philip C Bevilacqua

Abstract

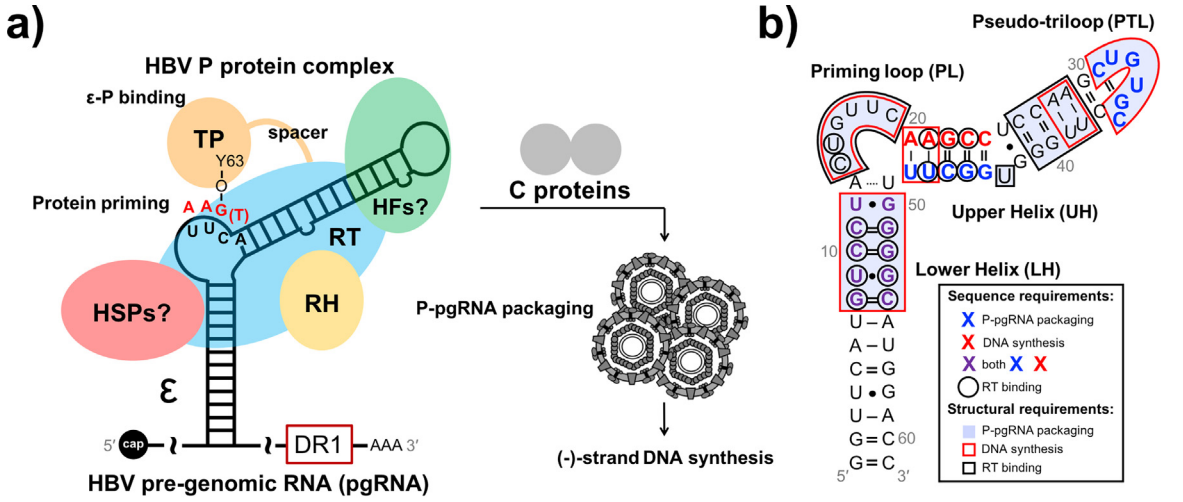
Human hepatitis B virus (HBV) replication is initiated by the binding of the viral polymerase (P) to epsilon (ε), an ≈85-nucleotide (nt) *cis*-acting regulatory stem-loop RNA located at the 5′-end of the pre-genomic RNA (pgRNA). This interaction triggers P and pgRNA packaging and protein-primed reverse transcription and is therefore an attractive therapeutic target. Our recent nuclear magnetic resonance (NMR) structure of ε provides a useful starting point toward a detailed understanding of HBV replication, and hints at the functional importance of ε dynamics. Here, we present a detailed description of ε motions on the ps to ns and μs to ms time scales by NMR spin relaxation and relaxation dispersion, respectively. We also carried out molecular dynamics simulations to provide additional insight into ε conformational dynamics. These data outline a series of complex motions on multiple time scales within ε. Moreover, these motions occur in mostly conserved nucleotides from structural regions (i.e., priming loop, pseudo-triloop, and U43 bulge) that biochemical and mutational studies have shown to be essential for P binding, P-pgRNA packaging, protein-priming, and DNA synthesis. Taken together, our work implicates RNA dynamics as an integral feature that governs HBV replication.

Introduction

Human hepatitis B virus (HBV) chronically infects an estimated 270 million people worldwide, leading to liver cancer and cirrhosis that cause ≈800,000 deaths each year.[1–3] Currently, there are no cures to wholly eliminate the virus without side-effects or indefinite treatment.[4–6] The global burden imposed by HBV therefore necessitates the study of its replication to enable the design of more effective antiviral drugs. HBV is a small DNA hepadnavirus that replicates its genome by reverse transcription of a pre-genomic RNA (pgRNA) intermediate.[7] Following transcription of the viral DNA by host RNA polymerase II, transcripts are transported to the cytoplasm where pgRNA provides the template for reverse transcription and the mRNA translated into HBV core (C) and polymerase (P) proteins.[7–8] Viral P comprises a reverse transcriptase (RT) domain, a middle spacer, a RNase H (RH) domain, and a terminal protein (TP) domain, which acts as a protein primer for reverse transcription.[9–14] HBV replication is initiated by the binding of P to epsilon (ε), an ≈85-nucleotide (nt) *cis*-acting regulatory RNA located at

the 5′-end of the pgRNA.[15–19] A second ε motif is also located at the 3′-end of the pgRNA but this copy does not bind P. This interaction at the 5′-end triggers packaging of both P and the pgRNA into subviral core particles by C protein dimers[20–21] and the initiation of protein-primed reverse transcription[22–25] (Figure 1(a)). The product of the priming reaction is a 3-4-nt DNA, whose 5′-end is covalently attached to a tyrosine residue (Y63) in the TP domain (Figure 1(a)), and which is synthesized using, as a template, the 6-nt priming loop (PL) bulge within ε.[12–14,22–25] The complex then translocates to a 3′-proximal element in the pgRNA, designated direct repeat 1 (DR1), where full-length (-)-strand DNA synthesis commences from the 3-4-nt DNA primer.[26–27]

With biochemical and mutational analyses, researchers have defined the basic requirements for the ε-P interaction,[28–29] P-pgRNA packaging,[17–18,24,30] and DNA synthesis.[22,24,29–30] In terms of proteins, the TP and RT domains of P are required for ε-P binding.[28,31] Cellular chaperones (e.g., heat shock proteins, HSPs) and host factor (s) (HFs) are further needed to assist the ε interaction and later P-pgRNA packaging.[28–34] As regards ε, mutational studies[17–18,22,24,28–30] show that different regions contribute to these functions in both a sequence- and structure-specific manner. As outlined in Figure 1(b, c)), the upper portion of the lower helix (LH) has primary sequence requirements for P binding[28] and P-pgRNA packaging.[17,24] Intriguingly, the lower portion of the upper helix (UH) has sequence requirements for P binding[28] and DNA synthesis on its 5′-side[24] and P-pgRNA packaging on its 3′-side,[17,24] whereas the upper portion of the UH solely has a structural role.[17] The PL bulge



c)

HBV replication function	Sequence requirements: Nucleotides	References	Structure requirements: Regions	References
P-pgRNA packaging	a. G8-U12 and G50-C54 b. C31-C36 c. G44-U48	a. 24, 28 b. 17, 28 c. 17, 24, 28	a. upper part of LH b. PL c. upper part of UH d. PTL e. U43 bulge	a. 17, 24, 28 b. 17, 18, 22, 24, 28, 29 c. 17, 28 d. 18, 28 e. 17, 28
DNA synthesis	a. A20-C24 b. G8-U12 and G50-C54	a. 24 b. 24	a. PL b. AU bps in UH c. PTL	a. 22, 24, 29 b. 30 c. 29
RT binding	a. G8-C11 and G51-C54 b. C14 and U15 c. A21-C23 and G45-U47	a. 28 b. 28 c. 28	a. PL b. upper part of UH c. U43 bulge	a. 28, 29 b. 28 c. 28

Figure 1. Key roles and requirements of the ε-P interaction in HBV replication. (a) Model of ε-P binding. While the ε motif is located at both the 3′- and 5′-ends of the pgRNA, the specific interaction of the TP and RT domains of P with the 5′-end ε initiates HBV replication. The specific location of HSPs and HFs and their sites of interactions are not completely known. The P-pgRNA complex is then packaged into viral core particles by C protein dimers, which forms the priming-competent complex that initiates TP primed (-)-strand DNA synthesis. Protein-primed reverse transcription initiates from the 6-nt priming loop bulge of ε to produce a 3-4-nt DNA (shown in red), whose 5′-end is covalently attached via a phosphotyrosyl linkage to Y63 in the TP domain. The P-DNA complex then translocates to the 3′-proximal DR1 element where full-length (-)-strand DNA synthesis is finalized. (b) RNA sequence and secondary structure requirements for ε-P binding, P-pgRNA packaging, and DNA synthesis are shown, as summarized from previous mutational studies.[17–18,22,24,28–30] ε structural regions are labeled with abbreviations that will be used throughout the text. (c) Same summary as in (b) but shown as a table, including relevant references to key work.

structure is essential for P binding,[28] P-pgRNA packaging,[17–18,22,28–29] and DNA synthesis,[17–18,24,29] whereas its 5′- and 3′-ends function separately in P binding[28] and protein priming,[24] respectively. The pseudo-triloop (PTL)[17,28] and U43 bulge[18,28] are essential for P-pgRNA packaging but only the former structure is dispensable for P binding.[28] Finally, the AU base pairs after the PL and preceding the PTL are required for protein-priming and (-)-strand DNA extension, respectively.[30] While this information provides a useful starting point, a detailed mechanistic understanding of the ε-P interaction and its subsequent functions require high-resolution structural studies. These data are completely lacking for P, making structural analysis of ε a necessary alternative.

In 2021, our group solved the structure of a 61-nt ε using solution nuclear magnetic resonance (NMR) spectroscopy (Figure S1(a)).[35] This ε construct contains the entire stem-loop region and will therefore be referred to as full-length ε throughout the text. Our NMR-derived model shows good agreement with the previously determined truncated 27-nt apical loop (AL) (i.e., nucleotides from the UH and PTL) ε NMR structure from Wijmenga and coworkers (Figure S1(a))[35–37] and provides insight into how the 6-nt PL bulge might engage the RT machinery. Moreover, our ^{13}C cross-correlated (η_{xy}) relaxation measurements hint at ps to ns motions within ε,[35] in agreement with previous studies on the AL ε[38] (Figure S1(b)). However, several open questions remain about the dynamics of full-length ε: what is the extent and time scales of motion it experiences? Does it adopt multiple conformations, and how might these modulate ε-P binding interactions and downstream functions? Work from the Nassal group using Systematic Evolution of Ligands by EXponential Enrichment (SELEX) indicates diverse ε sequence mutations are compatible with P binding, suggesting structural flexibility of ε.[39]

As a first step toward assessing the relevance of RNA dynamics for the ε-P interaction, we present a detailed description of ε motions. We used ^{13}C NMR spin relaxation and ^{1}H relaxation dispersion measurements to probe motions on the ps to ns and μs to ms time scales, respectively. We also carried out molecular dynamics (MD) simulations to provide additional insight into ε conformational dynamics. These data outline a series of complex motions on multiple time scales within the PL, PTL, and U43 bulge. Importantly, these motions occur in mostly conserved nucleotides from structural regions that biochemical and mutational studies have shown to be essential for P binding, P-pgRNA packaging, and DNA synthesis. Taken together, our work implicates RNA dynamics as an integral part of HBV replication and suggests avenues for targeting by anti-HBV therapeutics.

Results

ε undergoes fast motions on the ps to ns time scale

Previous ^{13}C spin relaxation measurements on AL ε reveal that nucleobase and ribose moieties in the PTL (U32-U34 and C36) and U43 bulge experience fast ps to ns motions (Figure S1(b)).[38] Whether or not full-length ε undergoes similar or additional motions has remained an open question, until now. Our recent ^{13}C η_{xy} measurements indicate that nucleotides in the PL (C14-C19 and adjacent U49), PTL (U32-U34 and C36), and U43 bulge show mobility in their nucleobases (Figure S1(b)).[35] These data[35] agree with what has been previously reported[38] and provide new information on unexplored ε regions (i.e., the LH and PL).

To confirm and extend these preliminary findings to full-length ε, we employed ^{13}C spin relaxation experiments to probe motions on the ps to ns time scale in both the nucleobase and ribose moieties. In all ^{13}C spin relaxation experiments, atom-specifically (i.e., purine ^{13}C8, pyrimidine ^{13}C6, and ribose ^{13}C1′)[40–41] labeled ε samples (Figure 2) were used to remove complications from ^{13}C–^{13}C dipolar couplings.[40–43] Measurements were carried out at two magnetic field strengths (600 MHz and 800 MHz) for nucleobase sites and a single magnetic field strength (800 MHz) for ribose sites. Longitudinal (R_1) and transverse (R_2) relaxation rates and steady-state {^{1}H}-^{13}C heteronuclear Overhauser effect (hNOE) values were determined for all well-resolved purine C8, pyrimidine C6, and ribose C1′ nuclei in ε (Figures S2-4). Monoexponential decay was observed in all relaxation experiments, indicating the absence of ^{13}C–^{13}C cross-relaxation and Hartman-Hahn transfer (Figure S5). Slight variations in data are observed for purine C8 and pyrimidine C6 nuclei within the same secondary structural elements, reflecting their different chemical shift anisotropies (CSAs) (Figure 3(a)). Given that the C1′ CSA is small and less dependent on secondary structure,[44] these data provide a view of dynamics orthogonal to that from nucleobase C6 and C8 sites (Figure 3(b)). All measurements show that most helical regions of ε are rigid whereas nucleotides in the PL (C14-C19 and adjacent U48 and U49), PTL (U32-C36), and U43 bulge show elevated R_1 and hNOE values (>1 standard deviation (SD) above the mean) and attenuated R_2 values (<1 SD below the mean), indicative of increased flexibility (Figure 3). All raw relaxation rates can be found in Tables S1-S3.

We then used the ratio ρ[45] to determine which regions of ε undergo motion on the ps to ns time scale (see Methods). This analysis is approximately independent of the site-specific variations in dipolar coupling and CSA and is therefore more robust than a simpler R_2/R_1 analysis. These data demonstrate

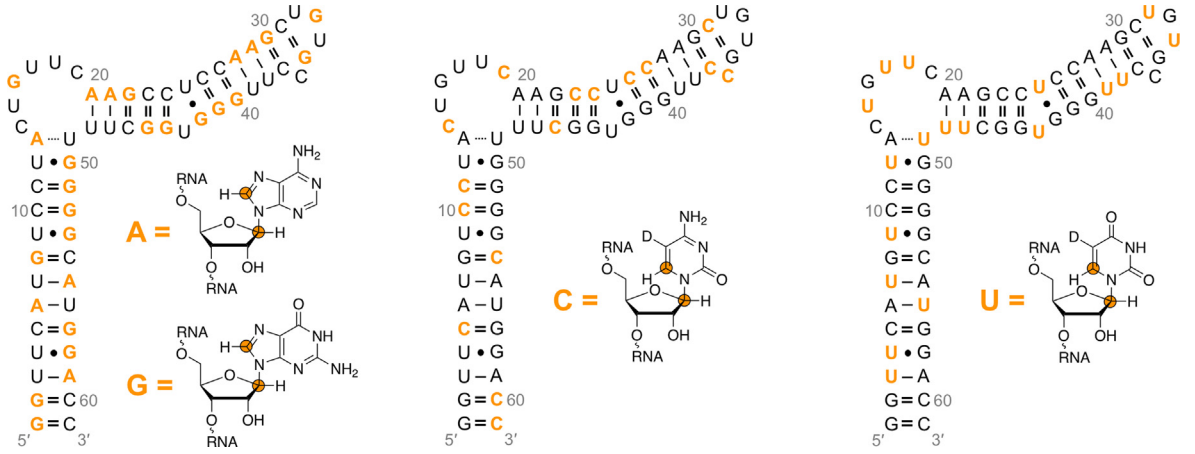


Figure 2. NMR samples used to probe ε dynamics. (a) For all NMR experiments, three atom-specifically labeled[40–41] ε samples were prepared by T7 RNA polymerase-based *in vitro* transcription: (1) [1′,8-$^{13}C_2$]-ATP and [1′,8-$^{13}C_2$]-GTP labeled, (2) [1′,6-$^{13}C_2$, 5-^{2}H]-CTP labeled, and (3) [1′,6-$^{13}C_2$, 5-^{2}H]-UTP labeled. Nucleotides harboring isotope labels are shown in orange and bolded. Orange circles and D represent ^{13}C and ^{2}H nuclei, respectively. These labeling topologies approximate isolated spin-pairs in both the nucleobase (i.e., purine H8-C8 and pyrimidine H6-C6) and ribose (i.e., H1′-C1′) moieties.

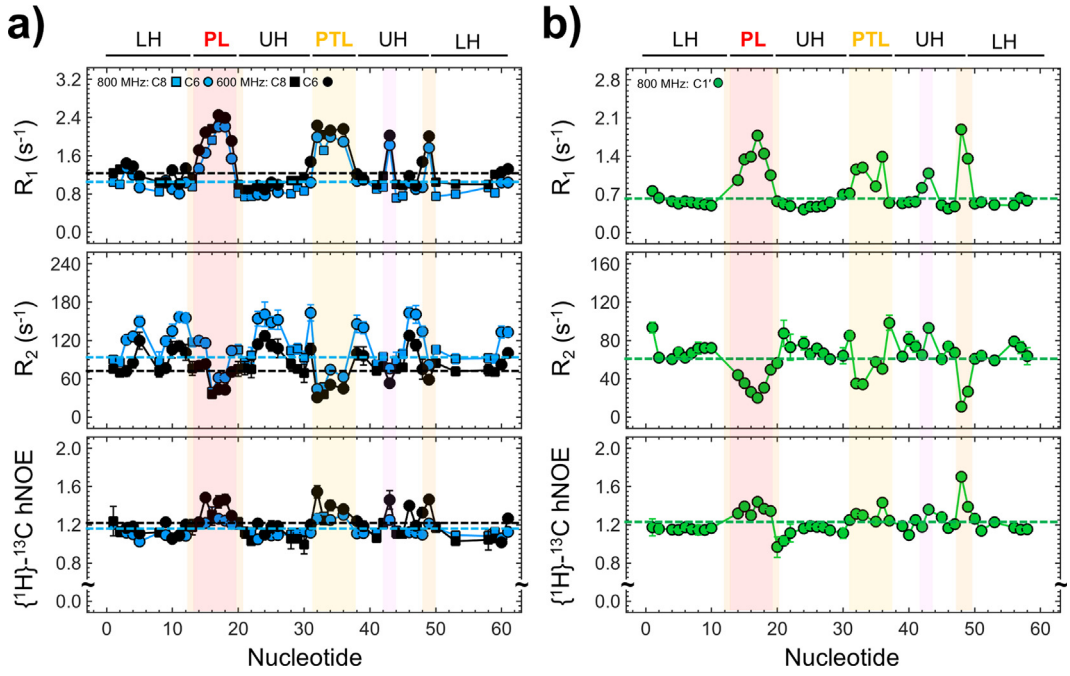


Figure 3. ^{13}C relaxation experiments indicate fast motions in ε. Measurements of ^{13}C R_1, R_2, and hNOE values are shown for all well-resolved (a) nucleobase (i.e., purine C8 and pyrimidine C6) and (b) ribose (i.e., C1′) nuclei with ε structural regions abbreviated as in Figure 1(b). NMR data were collected at 298 K on both 600 MHz and 800 MHz spectrometers. Data from nuclei in the PL, PL-adjacent (i.e., A13, A20, U48, and U49), PTL, and U43 bulge are highlighted with red, orange, yellow, and purple shaded boxes, respectively. Error bars represent ± SD and dashed lines represent 1 SD above (R_1 and hNOE) or below (R_2) the mean (calculated from helical nuclei in G1-A13, A20-C30, C37-G42, G44-U47, and G50-C61).

that, compared to helical regions, nucleobase sites in the PL (C14-C19 and U49), PTL (U32-U34 and C36), U43 bulge (and adjacent G41), and 5′- and 3′-ends (G1-U3 and G58) show attenuated (<1 SD below the mean) ρ values, suggestive of high-frequency motions (Figure 4(a)). Fast dynamics were also observed for ribose nuclei in the PL (C14-C19 and adjacent A20, U48, and U49), PTL

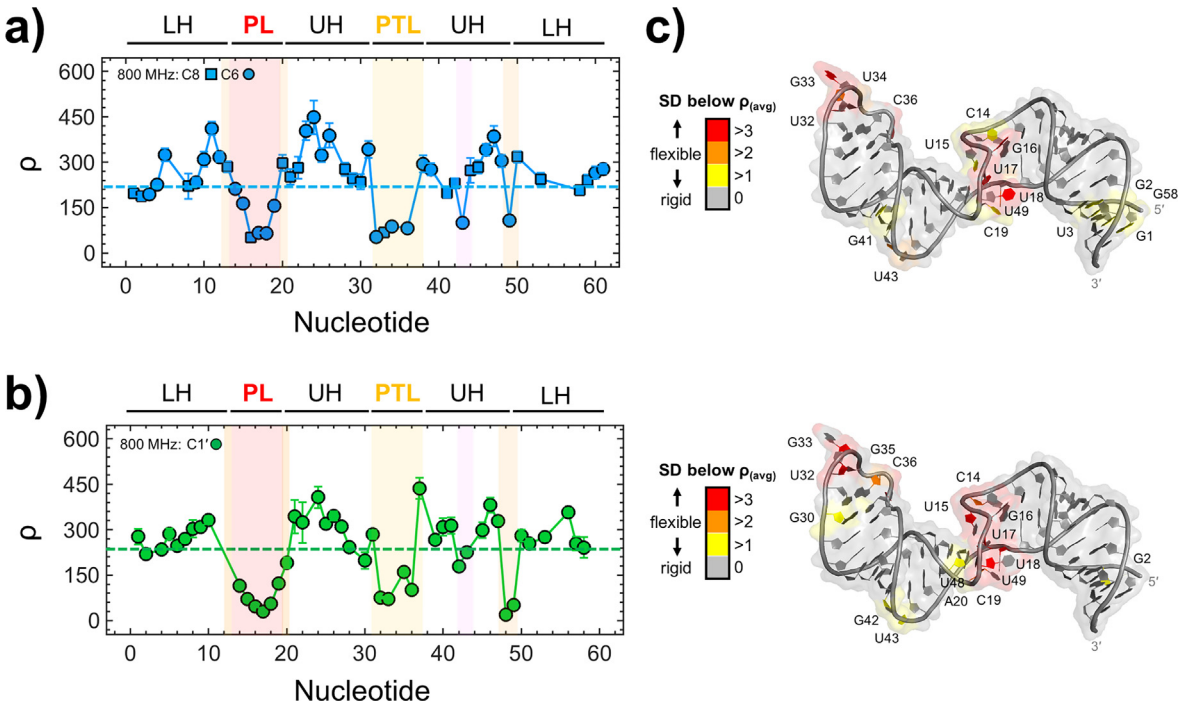


Figure 4. Mapping dynamic "hot spots" in ε. ρ values (see Methods) derived from the data shown in Figure 3 for all well-resolved (a) nucleobase (i.e., purine C8 and pyrimidine C6) and (b) ribose (i.e., C1′) nuclei, with ε structural regions abbreviated as in Figure 1(b). NMR data were collected at 298 K on both 600 MHz and 800 MHz spectrometers but only 800 MHz data are shown here. Data from nuclei in certain ε structural regions are shaded as in Figure 3. Error bars represent ± SD and dashed lines represent 1 SD below the mean (calculated from helical nuclei as in Figure 3). (c) Summary of data in (a) and (b) mapped onto the top-ranked ε NMR conformer (Protein Data Bank (PDB) code 6VAR[35]) to indicate dynamic "hot spots" that undergo fast ps to ns motions.

(U32, G33, G35, C36, and adjacent G30), U43 bulge (and adjacent G42), and 5′-end (G2) (Figure 4(b)). Taken together, our ^{13}C relaxation data (Figures 3 and 4) indicate that the PL, PTL, and U43 bulge are dynamic "hot spots" that undergo fast ps to ns motions (Figure 4(c)).

ε experiences conformational exchange on the μs to ms time scale

Given that previous ^{13}C rotating-frame ($R_{1\rho}$)-based measurements hinted at μs to ms motions within AL ε,[38] we sought to unambiguously probe for such conformational exchange in full-length ε. We therefore employed ^{1}H Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion measurements.[46–48] In all ^{1}H CPMG experiments, atom-specifically (i.e., purine $^{13}C8$, pyrimidine $^{13}C6$, and ribose $^{13}C1'$)[40–41] labeled ε samples (Figure 2) were used to remove complications from ^{13}C–^{13}C scalar couplings.[46,49] Measurements were carried out at 800 MHz and CPMG curves were obtained for all well-resolved purine H8, pyrimidine H6, and ribose H1′ nuclei in ε. In these experiments, the effective R_2 rates ($R_{2,eff}$) were measured during a constant-time period that contains a variable number of equally spaced 180° refocusing pulses. The dependence of $R_{2,eff}$ on the frequency at which the refocusing pulses were applied (ν_{CPMG}) is indicative of conformational exchange.

This analysis indicated that several nucleotides in the PL (C14, U15, U17-C19, U48, and U49), PTL (U32, U34, and C36), and U43 bulge experience conformational exchange whereas the majority of helical nucleotides do not (Figure 5). Interestingly, as with our ^{13}C spin relaxation data, some nucleotides exhibit different motions in their nucleobase and ribose moieties. For example, exchange was observed in both nucleobase and ribose nuclei for U17, U18, U32, and U43 (Figure 5). Conversely, C14, C19, and U34 displayed exchange only in nucleobase sites (Figure 5), whereas U15, C36, U48, and U49 exhibited exchange only in ribose nuclei (Figure 5).

Fitting the CPMG dispersion curves to a two-site exchange model provides insight on the kinetics (i.e., exchange rate, k_{ex}) and thermodynamics (i.e., populations of major and minor states, p_a and p_b, respectively) of the interconverting species, depending on the exchange regime.[46–48] In fast exchange, it is not possible to fit p_a, p_b, or the chemical shift difference ($\delta\omega$) between them. Instead, only R_2 auto-relaxation ($R_{2,0}$), the chemical exchange contribution (R_{ex}) to R_2, k_{ex}, and δ_{ppm}^{min} can

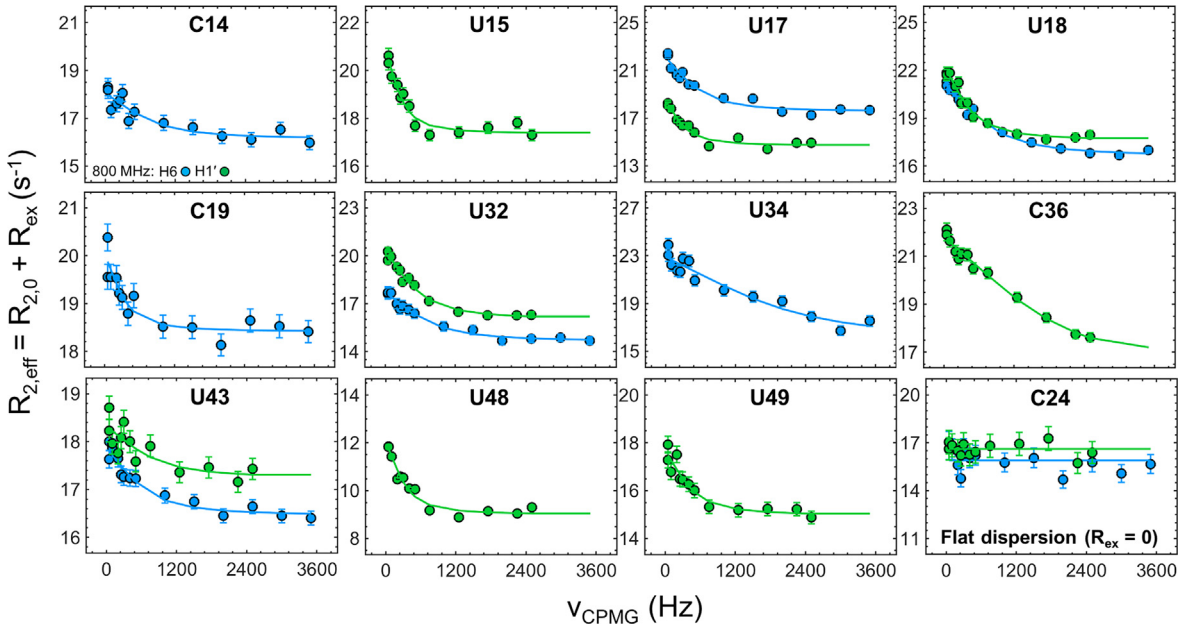


Figure 5. ^{1}H CPMG experiments reveal conformational exchange in ε. ^{1}H CPMG relaxation dispersion profiles are shown for nucleobase (i.e., pyrimidine H6) and ribose (i.e., H1′) nuclei showing evidence of exchange (i.e., a non-flat dispersion where $R_{ex} > 0$). NMR data were collected at 298 K on an 800 MHz spectrometer. All CPMG curves are shown with best fits to the fast exchange model (i.e., Luz-Meiboom[50]) using RING NMR Dynamics.[51] Representative nucleobase (C24-H6) and ribose (C24-H1′) profiles that do not experience conformational exchange (i.e., $R_{ex} = 0$) are shown as a reference. Error bars represent ± SD.

be obtained, where the latter parameter represents the δω that would be present if the populations were equal.[50–52] The presence of chemical exchange and the appropriate exchange regime (i.e., fast, slow, or no exchange) was identified and selected based on the Akaike Information Criterion (AIC)[51,53] (see Methods and Tables S4 and S5). Moreover, all CPMG data were analyzed with three different fitting programs to ascertain the robustness of the fits (see Methods and Table S6). These analyses revealed that all ε nuclei with R_{ex} exhibit exchange that is fast on the NMR time scale (Tables S4-S6). We are therefore unable to obtain additional structural information (i.e., populations and chemical shifts) of the low populated, transient ε states.

Interestingly, conformational exchange processes in ε occur with a variety of k_{ex} values (Table S4 and Figure 6(a)). Considering such events in both nucleobase and ribose sites, PL nuclei have a k_{ex} ranging from ≈1,700 s^{-1} to 4,700 s^{-1}, those in the PTL have k_{ex} between ≈3,400 s^{-1} to 11,700 s^{-1}, and U43 has a k_{ex} of ≈3,800 s^{-1} to 4,500 s^{-1} (Table S4 and Figure 6(a)). These k_{ex} values correspond to lifetimes (τ_{ex}) of ≈90 μs to 600 μs (Table S4 and Figure 6(a)). Global fitting of the relaxation dispersion data to a two-site fast exchange model suggests an overall exchange event with a k_{ex} value of ≈2,700 s^{-1} and a lifetime of ≈370 μs (Table S4). It is important to note that the global fit parameters do not significantly change with and without the inclusion of the very fast exchange events in U34-H6 and C36-H1′ nuclei (Table S4).

Taken together, our ^{1}H CPMG data reveal vast conformational exchange within ε. Interestingly, these μs to ms motions localize to the same structural regions (i.e., PL, PTL, and U43 bulge) (Figure 6(b)) as those on the ps to ns time scale (Figure 4(c)), implying that relatively slower motions are superimposed onto faster ones.

MD simulations show vast global conformational sampling of ε

MD simulations were employed to provide some structural insight and report on motions that occur on time scales between that of NMR spin relaxation and relaxation dispersion (i.e., ns to μs motions). Three independent 1 μs simulations were carried out on ε NMR conformer 3, designated R3. This model was chosen due to its unique PL orientation that was previously shown to dock small molecule ligands.[35] A large ensemble of states is evident from overlays of ε conformations sampled throughout an MD run (Figure 7(a)). When aligning the backbone phosphorus atom of all nucleotides, the ensemble centers around an average fold, similar to that of the deposited NMR model.[35] However, when the alignment only includes nucleotides in the LH, this no longer holds true. Instead, a vast conformational ensemble is

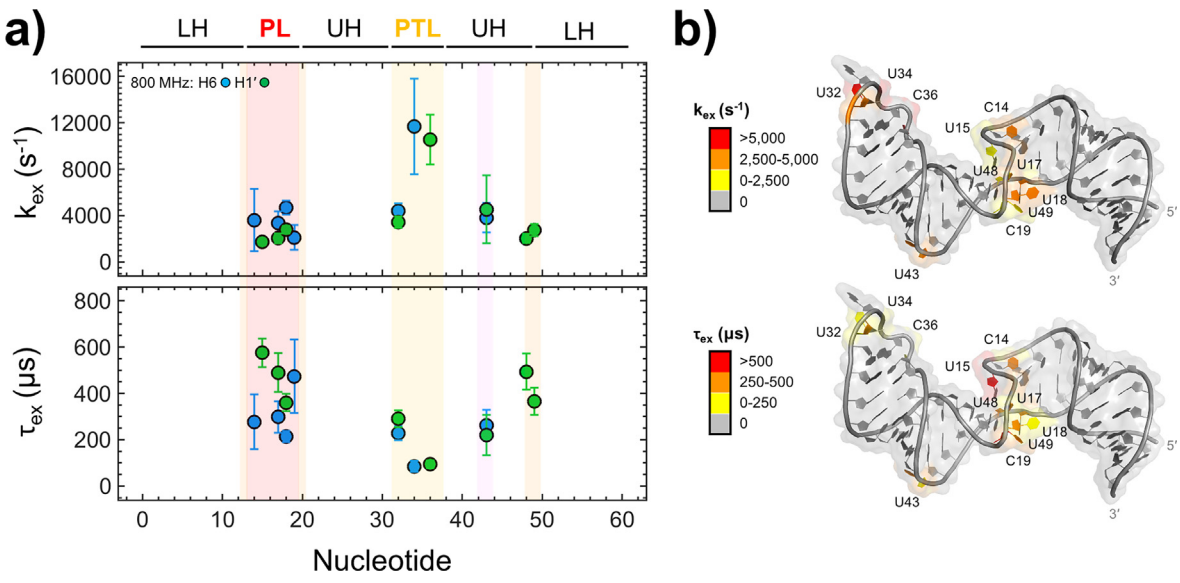


Figure 6. Quantifying the rate and lifetime of conformational exchange processes in ε. (a) All k_{ex} and τ_{ex}, values, derived from the ^{1}H CPMG data presented in Figure 5 and Table S4, are shown with ε structural regions abbreviated and shaded as in Figure 1(b) and Figure 3, respectively. The k_{ex} values are the direct output from fast exchange model (i.e., Luz-Meiboom[50]) fitting using RING NMR Dynamics[51] and $\tau_{ex} = 1/k_{ex}$. Error bars represent ± SD. (b) Summary of data in (a) mapped onto the top-ranked ε NMR conformer (PDB code 6VAR[35]).

observed with large deviations within the remaining portion of the RNA (Figure 7(a)). Taken together, this suggests that global ε flexibility correlates with its PL, PTL, and U43 bulge (Figure 7(a)). In agreement with this idea, MD simulations show elevated root-mean-square-deviation (RMSD) values (>1 SD above the mean) for nucleotides in the PL (A13-A21 and U48-G50), PTL (C31-C37), and U43 bulge (G42-G44), indicative of increased flexibility far above those in helical regions (Figure 7(b) and Supplementary Script S1). While variations exist between the three MD runs, the overall trends are unambiguous (Figure S6). Moreover, MD data (Figure 7(b)) show agreement with ^{13}C spin relaxation data (Figures 3 and 4), and we therefore conclude that ε exhibits increased sampling on the ns to μs time scale within the same structural regions that undergo ps to ns motions (i.e., PL, PTL, and U43 bulge).

To provide further structural insight into the conformational sampling of ε, PDB snapshots were taken throughout the MD trajectories. These analyses reveal that nucleotides in the PL undergo a series of nucleobase flipping events and backbone reorganizations that occur on the ns to μs time scale. For example, the nucleobases of A13 and C14 move toward one another into a partially stacked conformation (Figure 7(c)). The nucleobase of G16, on the other hand, flips away from the PL central cavity to partially stack with U15 (Figure 7(c)). Finally, the nucleobases of U17, U18, and C19 flip back-and-forth between the PL central cavity, facilitated by kinking of their backbone dihedrals (Figure 7(c)). Similar types of motions are also observed in the PTL and U43 bulge, though to a lesser extent (Figure S7).

We then performed K-means clustering analyses on all the conformations sampled in each of the three 1 μs MD trajectories. Clustering was carried out using either the global RMSD of the backbone phosphorus atom of nucleotides G2-C61 or backbone and chi dihedral angles from PL nucleotides A13-A20. We used five clusters and identified centroid structures (designated c0, c1, c2, c3, and c4) that are MD-sampled conformations that have the lowest cumulative distance to every other point in a given cluster. Centroid structures are therefore representative conformations of the five clusters, where c0 represents the most populated centroid and c4 represents the least. Interestingly, individual simulations largely sampled individual clusters, with minimal sampling across clusters (Tables S7 and S8). However, when the ensemble was considered (i.e., including conformations from all three ε R3-based MD runs) (Supplementary Scripts S2 and S3), there was a high degree of correlation between centroid structures identified by clustering on global RMSD and centroid structures identified by clustering on the PL backbone dihedral angles. Figure 8(a) shows a two-dimensional (2D)-RMSD plot comparing the backbone phosphorus atom RMSD of centroid structures from global RMSD (x-axis) and PL dihedral (y-axis) clustering. Comparatively low backbone RMSD along the diagonal suggests that

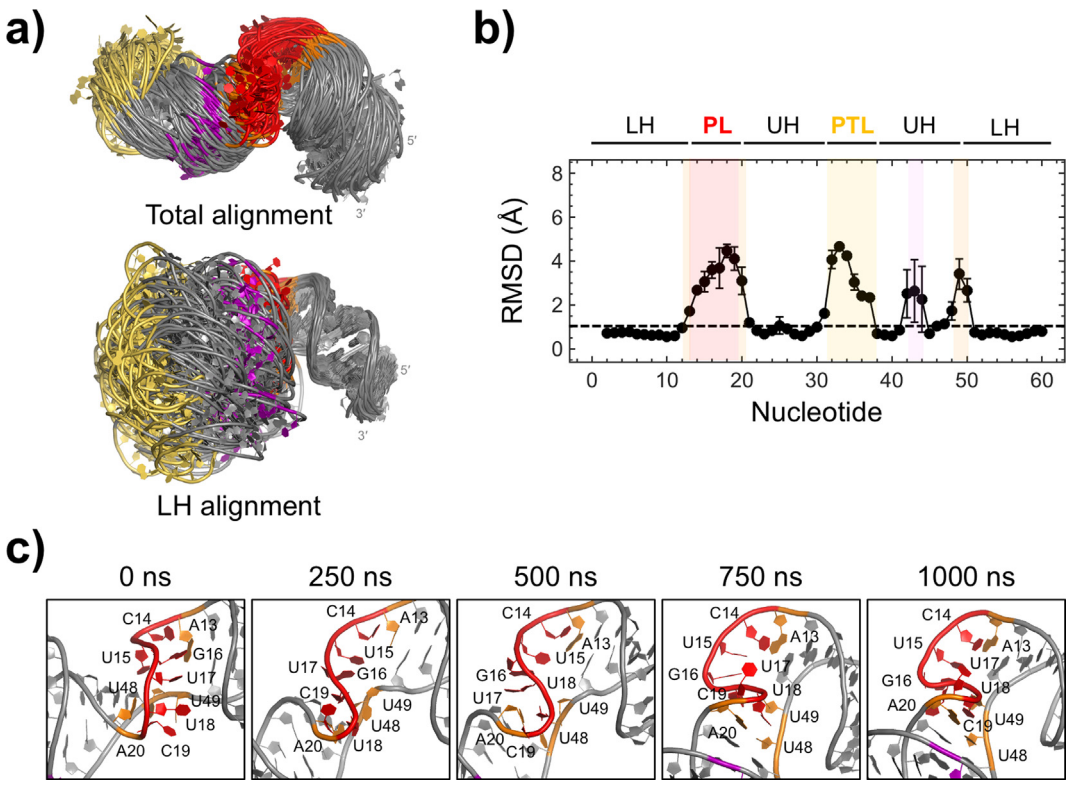


Figure 7. MD simulations reveal high-frequency motions in ε. (a) Representative structural overlay of PDB snapshots taken every 10 ns of the first (of three) 1 μs MD trajectory, with backbone phosphorus atom alignments shown using all (i.e., G2-C61) (top) or LH (i.e., G1-U12, G50-C60) (bottom) nucleotides. (b) The average whole-atom RMSD (calculated against ε R3, PDB code 6VAR[35]) of the three MD runs for each nucleotide window (see Methods), averaged over the entire 1 μs trajectory. Error bars represent ± SD and the dashed line represents 1 SD above the mean (calculated from helical nucleotides). (c) Representative PDB snapshots are shown at equally spaced time points along the first (of three) 1 μs MD trajectory and are aligned at the backbone phosphorus atom of PL nucleotides (i.e., A13-A20, U49 and G50). ε snapshots are shown in cartoon, and all structural regions are abbreviated and colored or shaded as in Figure 1(b) and Figure 3, respectively.

these centroid structures (e.g., c0 from RMSD clustering and c0 from PL dihedral clustering) have the same overall conformations (Figure 8(a, b)). Importantly, this correlation is absent when clustering is performed using PTL or LH nucleotide dihedral angles (Figure 8(c)) instead of the PL nucleotide dihedral angles. Taken together, these data suggest that global ε motions uniquely correlate with PL conformations (Figure 8).

We then analyzed the link between PL flexibility and ε conformational sampling by calculating the dynamic cross-correlation matrix from our three ε R3 1 μs MD trajectories, as previously described[54] (Supplementary Script S4). This analysis can reveal correlated and anti-correlated motion with nucleotide resolution. As expected, strong correlations (≈1.0, red regions on the plot) were observed along the diagonal (Figure 9(a)), corresponding to self and base pair interactions. Interestingly, there was a significant (<−0.5, blue regions on the plot) anti-correlation between nucleotides in the PL (A13-C19) and the PTL (A28-G35) and the 3′-end (A55-G58) (Figure 9(a)). These are clearly long-range events that do not involve direct interactions between nucleotides (Figure 9(b)). Nevertheless, these anti-correlated motions between the PL and PTL and 3′-end likely explain the vast spread in global ε orientations that are observed in the centroid structures (Figure 8(b)).

Finally, we carried out additional 1 μs MD trajectories on all ε NMR conformers, designated R1-R10.[35] As one means of quantifying the global sampling of ε, we measured the overall global bend between the helices. To this end, a simple angle (Θ) convention was used, where the center of mass of the first turn of the LH (i.e., G2-A6 and U56-C60), the PL and its flanking nucleotides (i.e., A13-A20, U48, and U49), and the second turn of the UH (i.e., U25-G30 and C37-G42) defined the three points (Figure 10(a) and Supplementary Script S5). Using data from all MD trajectories (i.e., three of ε R3 and new simulations of ε R1-R10), the vast sampling of Θ angles from 65° to 180° suggests a large conformational space with many global rearrangements (Figure 10(b)). We then employed principal component analysis (PCA) of the covariance

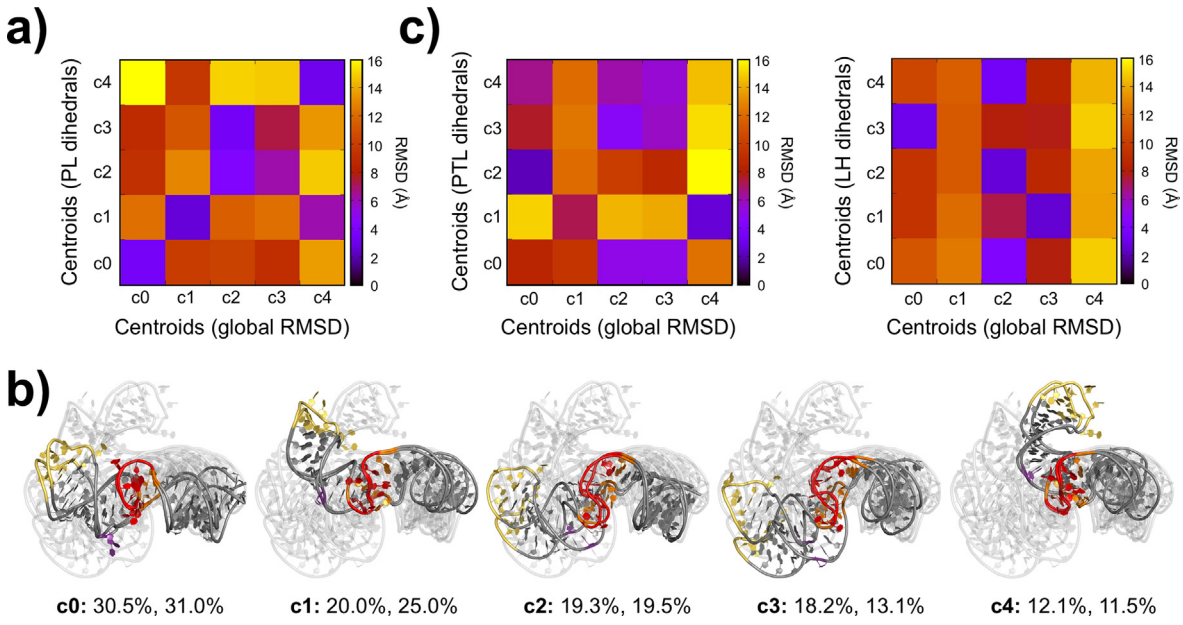


Figure 8. Global ε motions correlate with priming loop conformations. (a) 2D-RMSD plot comparing centroid structures from K-means clustering analysis of the set of three ε R3-based 1 μs MD runs using global RMSD of the backbone phosphorus atom of nucleotides G2-C61 (x-axis) and backbone and chi dihedrals (i.e., α, β, γ, δ, ε, ζ, and χ) of the PL nucleotides A13-A20 (y-axis). (b) Representative structural overlays with percentage of ensemble from global RMSD (first percentage) and PL dihedral (second percentage) clustering. Centroids of interest are shown in cartoon with structural regions colored as in Figure 1(b). All remaining centroids are shown in transparent cartoon, all aligned at the backbone phosphorus atom of the PL (i.e., A13-A20, U49, and G50). (c) Same 2D-RMSD plot as in (a) except clustering by dihedrals (y-axis) of the PTL nucleotides C31-C36 (left) or LH nucleotides G2-A13 and U49-C61 (right).

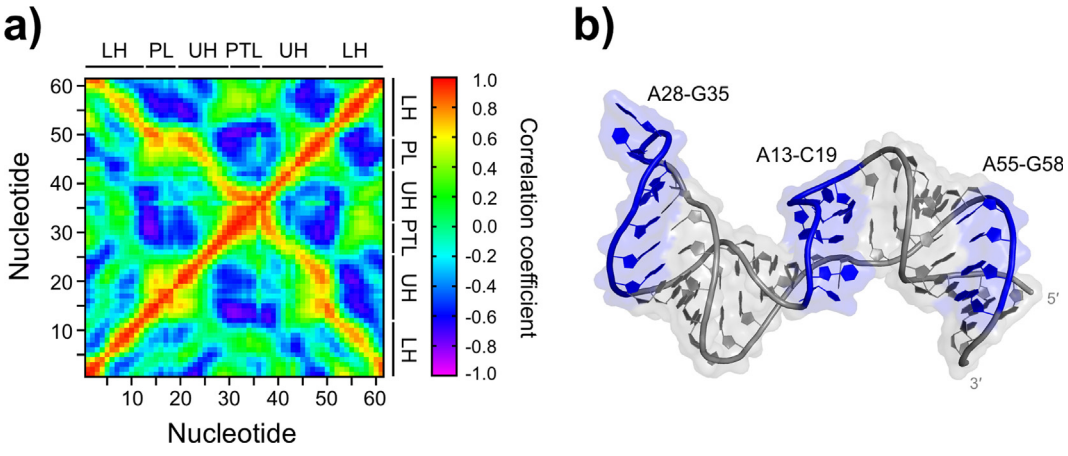


Figure 9. The ε priming loop shows anti-correlated motion with the pseudo-triloop and 3′-end. (a) Dynamic cross-correlation matrix is shown with ε structural regions abbreviated as in Figure 1(b). The matrix was calculated for all non-hydrogen atoms per nucleotide from the conformational space sampled in the three ε R3 1 μs MD trajectories, as previously described.[54] Correlation coefficients signify which regions display correlated (>0.5, red regions on the plot) and anti-correlated (<0.5, blue regions on the plot) motions. (b) Regions of anti-correlation mapped onto ε R3 (PDB code 6VAR[35]).

matrix of all MD trajectories to extract the dominant modes of global RNA motion sampled in these simulations (Supplementary Script S6). PC1 and PC2 account for 43.6 % and 18.8 % of the eigenvalues in the covariance matrix, respectively, and therefore describe the majority (>61 %) of motional modes. While ε samples a vast amount of conformational space (Figures 10(c) and S8), no single simulation

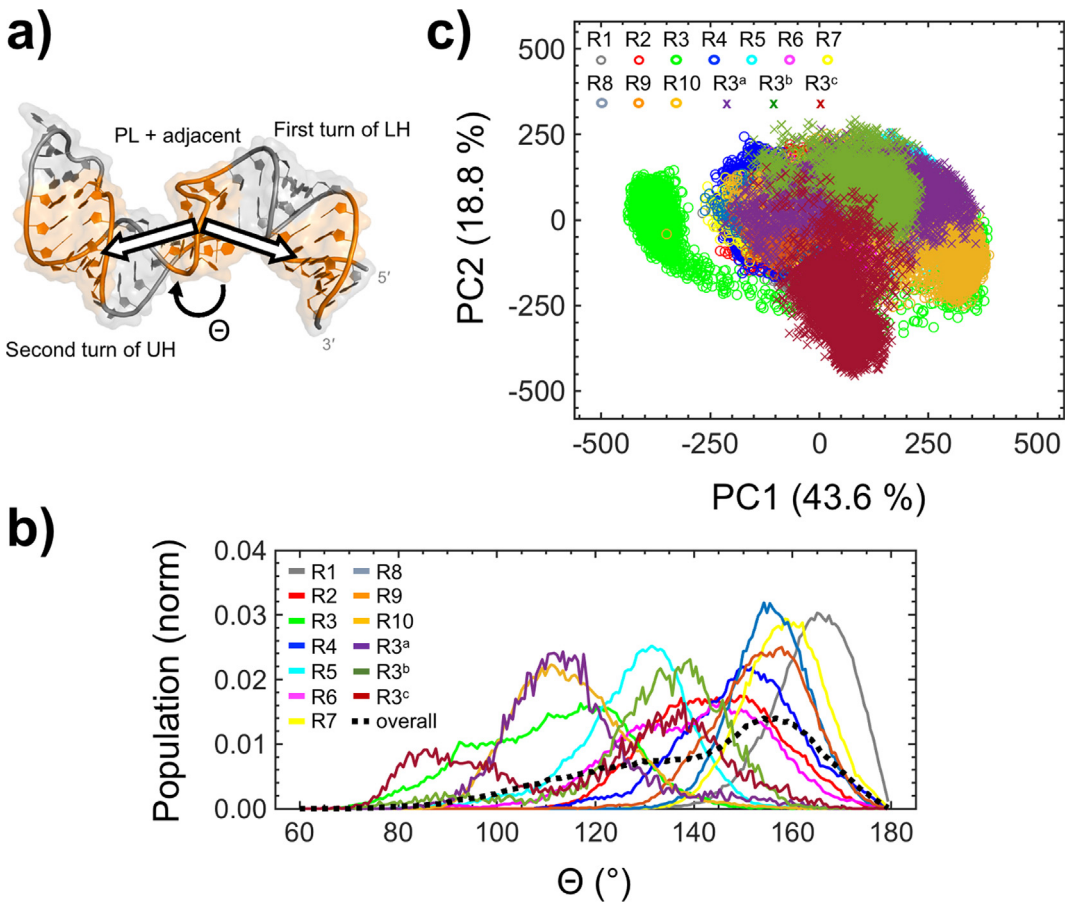


Figure 10. Vast conformational sampling of ε. (a) Schematic definition of the angle Θ mapped onto the top-ranked ε NMR conformer (PDB code 6VAR[35]). (b) Normalized population data of the distribution of Θ sampled for all 13 MD trajectories. (c) PCA of the modes of motion along PC1 and PC2 for all 13 MD trajectories. $R3^a$, $R3^b$, and $R3^c$ refer to the initial three independent 1 μs MD runs of ε R3.

visits all the possible conformations, suggesting that our simulations show no obvious convergence. Instead, our simulations investigate potential conformations and demonstrate that wide global rearrangements in ε are possible and likely occur on a time scale beyond the MD sampling (i.e., 1 μs).

Discussion

The *cis*-acting regulatory stem-loop ε plays a central role in the HBV life cycle.[15–19] Its interaction with P is critical for genome replication and therefore an attractive therapeutic target.[15–19] Unfortunately, lack of structural data on P limits our understanding of this interaction and its subsequent functions. Our recent NMR structure of full-length (61-nt) ε[35] provides an important starting point to an improved understanding of HBV replication. In addition to structural data, initial ^{13}C η_{xy} relaxation measurements suggest that nucleotides in the PL, PTL, and U43 bulge are flexible on the ps to ns time scale (Figure S1(b)).[35] Given that these region are required for the ε-P interaction[28–29] and subsequent P-pgRNA packaging[17–18,24,30] and DNA synthesis,[22,24,29–30] our data imply the functional importance of these motions. Whether or not ε by itself experiences conformational dynamics on additional time scales beyond ps to ns was completely unknown until this work, and we anticipate that such knowledge would add valuable insight to the ε-P binding mechanism.

ε undergoes motion on multiple time scales in conserved nucleotides critical for HBV replication

To begin to fill this critical knowledge gap, we combined NMR and MD simulations to provide a robust description of ε dynamics in the free state. ^{13}C spin relaxation measurements were used to assess fast motions on the ps to ns time scale. In agreement with our recent work[35] and previous NMR studies,[38] nucleotides in the PL (C14-C19 and adjacent A20, U48, and U49), PTL (U32-C36 and adjacent G30), and U43 bulge (and adjacent G41 and G42) undergo fast ps to ns motions (Figures 3 and 4). To unambiguously determine if ε experiences motion on the slower μs to ms time

scale, we employed ^{1}H CPMG relaxation dispersion measurements. Interestingly, nucleotides in the PL (C14, U15, U17-C19, U48, and U49), PTL (U32, U34, and C36), and U43 bulge all experience conformational exchange (Figures 5 and 6) that is fast on the NMR time scale with a global exchange rate and lifetime of ≈2,700 s^{-1} and ≈370 μs, respectively (Table S4). These observations suggest that slower motions are superimposed onto faster ones, with the notable exception of G16 and G33 in the PL and PTL, respectively (Figures 3–6). It is important to note that our NMR measurements only sample a subset of motional time scales. Additional studies are therefore needed to determine if ε undergoes dynamics on time scales outside the detectable range of spin relaxation and CPMG. Some of these experiments (e.g., chemical exchange saturation transfer (CEST) and off-resonance $R_{1\rho}$) would also provide the structural characterization (i.e., populations and chemical shifts) that our initial CPMG measurements are lacking.

To gain additional insight into the conformational dynamics of ε, MD simulations were carried out on ε R3. RMSD analysis reveals that nucleotides in the PL (A13-A21 and U48-G50), PTL (C31-C37), and U43 bulge (G42-G44) show increased sampling on the ns to μs time scale (Figures 7(a, b) and S6), consistent with our NMR data (Figures 3 and 4). In an attempt to provide some structural insight regarding local dynamics in ε, we took PDB snapshots along our ε R3 MD trajectories. These analyses reveal that nucleotides in the PL, PTL, and U43 bulge undergo a series of nucleobase flipping events and backbone reorganizations (Figures 7(c) and S7). While the MD runs do not extend to the time scale that our ^{1}H CPMG data report on, these data may nevertheless be useful in hinting at the structural rearrangements that take place on the longer μs to ms time scale.

To better understand the conformational variety in ε, we clustered together of all MD-sampled conformations of ε R3 (Figure 8). Interestingly, there was a correlation between the dominant PL orientations (cluster centroids from backbone and chi dihedral angle clustering) and the global motions of the RNA (cluster centroids from backbone phosphorus RMSD clustering) (Figure 8 (a, b)). Indeed, the representative centroid structures display of a wide variety of both PL conformations and global sampling (Figure 8(b)), suggesting that the high degree of PL flexibility prevents ε from adopting a single, stable conformation. Instead, PL flexibility enables global ε flexibility. Dynamic cross-correlation matrix analysis suggests that anti-correlated motions between the PL and PTL and 3′-end (Figure 9) likely explain the vast spread in global ε orientations. To gain a better understanding of the conformational space of ε, we carried out 1 μs MD simulations on all ε NMR conformers R1-R10. We observed a vast amount of sampling, as measured by the interhelical angle Θ (Figure 10 (b)) and PCA (Figure 10(c)). Given that there was no obvious convergence in our MD simulations, global rearrangements in ε likely occur on a time scale beyond the MD sampling (i.e., 1 μs). This is consistent with the overall interpretation of our ^{1}H CPMG data, which suggests that ε samples low populated, transient states on the μs to ms time scale (Figures 5 and 6). Interestingly, while the conformational variation in ε is dramatic, the vectors of amplitudes of motion of the PC1 projection demonstrate that these motions revolve around an average structure similar to our structural model (PDB code 6VAR[35]) (Figure S8).

Taken together, our combined NMR and MD data suggest that ε undergoes conformational dynamics on multiple time scales (Figure 11(a)). Moreover, these motions are localized to nucleotides that are highly conserved and functionally important (Figure 11(b)). That is, every dynamic nucleotide in the PL, PTL, and U43 bulge is completely conserved except C14, G16, U32, and G35, and all nucleotides but G42 vary in fewer than three HBV strains (Figure 11).[37] While nucleotides in rigid structural elements (e.g., helices) may be conserved to maintain their structure, the motions in the non-helical PL, PTL, and U43 bulge nucleotides are likely conserved for their biological function.

Conformational dynamics of ε: Implications for HBV replication

The diversity of RNA sequence- and structure-specific requirements for HBV replication (Figure 1 (b, c)) suggests that each step (i.e., P binding, P-pgRNA packaging, protein-priming, and DNA synthesis) (Figure 1(a)) may require distinct conformations of ε. Moreover, there are ε conformational states beyond unbound, P-bound, and priming-competent, as evidenced by the case where ε-P binding occurs without protein-priming.[28] Nucleotide motions may therefore enable binding to P and cellular HFs, and facilitate conformational changes required for P-pgRNA packaging, protein-priming, and DNA synthesis, as suggested for the related duck HBV.[27,55]

Using this framework, we hypothesize that dynamics may be linked to function as follows. Given that the PL and U43 bulge are required for P binding (Figure 1(b, c)),[28–29] nucleotides with fast ps to ns motions in these regions could facilitate this interaction. Although the PTL is dispensable for P binding (Figure 1(b, c)),[28] it is proposed to interact with cellular HFs that are needed for P-pgRNA packaging and protein-priming.[28] Again, ps to ns motions in the PTL may promote binding to HFs. Both scenarios could involve conformational capture mechanisms, whereby mobile nucleotides facilitate binding interactions. Superimposed on these fast motions, it is likely that a series of complex conformational changes on a slower time scale could

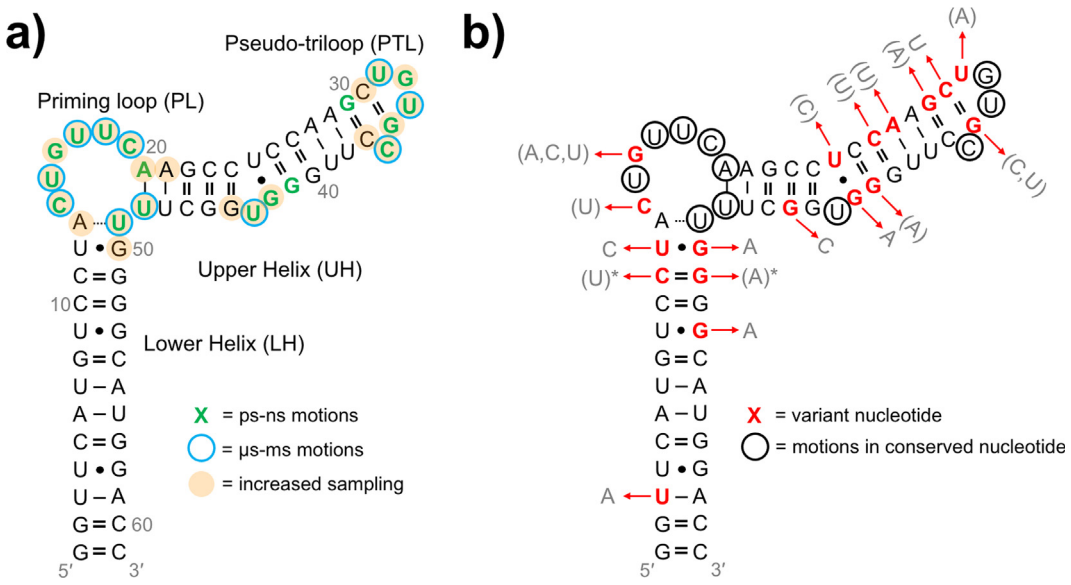


Figure 11. Summary of ε dynamics data as they relate to nucleotide conservation. (a) ^{13}C spin relaxation, ^{1}H CPMG relaxation dispersion, and MD whole-atom RMSD data mapped onto the secondary structure of ε. Even though each nucleotide is denoted as either experiencing ps to ns and/or μs to ms motions there are instances where this is true only for the nucleobase (i.e., purine-H8/C8 or pyrimidine-H6/C6) or ribose (i.e., C1′ or H1′) moieties (Figures 3-6). Dynamic nucleotides from the 5′- and 3′-ends have been excluded given that they likely experience motions due to base pair fraying. (b) Nucleotide mutations and conserved mobile nucleotides mapped onto the secondary structure of ε. Nucleotides with at least one mutation described among 1,025 HBV strains from the literature, and/or observed by Wijmenga and co-workers[37] in 205 strains, are shown in gray. Mutations shown in parentheses have been seen in fewer than three strains. Asterisks show a coincidental double base-pairing mutation in one strain. Conserved mobile nucleotides were shown to be dynamic by both NMR (^{13}C spin relaxation or ^{1}H CPMG) and MD.

alter contact networks between ε, P, and HFs to encourage P-pgRNA packaging, protein-priming, and DNA synthesis. Fast exchange events in the PTL (e.g., U34-H6 and C36-H1′) (Table S4) could initiate the transition of the ε-P-HF complex to its priming-competent state, which likely involves cellular HFs.[28] Then, other exchange processes may position ε within the P complex and cellular HFs to prepare the priming complex and present the 3′-end of the PL to the TP domain to prime and initiate reverse transcription. In the absence of additional experimental data, this must remain purely speculative.

In conclusion, our combined NMR and MD data indicate a series of complex motions on multiple time scales within ε (Figure 11(a)). This work provides a useful follow-up to our recent ε NMR structure[35] and underscores an important reality: rather than a single, stable structure, ε exhibits a dynamic conformational ensemble with motions that occur in conserved structural regions (i.e., PL, PTL, and U43 bulge) that are critical for function (Figures 1(b, c) and 11). Taken together, our work implicates RNA dynamics as an integral part of HBV replication. RNA dynamics must therefore be taken into consideration in future therapeutic efforts using ε as a target. Explicit confirmation of our proposed dynamic model would require NMR measurements of ε in the presence of P, or at least its RT domain, which at present is not feasible. Our dynamic description of unbound ε therefore provides a necessary starting point for a detailed understanding how RNA dynamics regulates HBV replication, which must consider important metal ion requirements (e.g., Mg^{2+}), structural dynamics of protein cofactors, and their interplay within the cellular context.

Methods

NMR sample preparation

ε RNA samples were prepared by T7 RNA polymerase-based *in vitro* transcription following well established protocols,[56] and as recently described.[42] After transcription, RNA was extracted with acid-phenol:chloroform, ethanol precipitated, purified by preparative denaturing polyacrylamide gel electrophoresis, and electroeluted. The samples were subsequently dialyzed five times against UltraPure ddH_2O, folded in NMR buffer (10 mmol/L Na_3PO_4 pH 6.5, 0.1 mmol/L EDTA), lyophilized, and resuspended in D_2O (99.9%). This NMR buffer, which lacks Mg^{2+}, was used in order to remain consistent with previous ε NMR studies.[35–38] NMR sample concentrations were between 0.5–

0.7 mmol/L in 0.3 mL (calculated using a molar extinction coefficient of 768.3 $(mmol/L)^{-1}cm^{-1}$). For all NMR experiments, three atom-specifically[40–41] labeled ε samples were used: (1) [1′,8-$^{13}C_2$]-ATP and [1′,8-$^{13}C_2$]-GTP labeled, (2) [1′,6-$^{13}C_2$, 5-^{2}H]-CTP labeled, and (3) [1′,6-$^{13}C_2$, 5-^{2}H]-UTP labeled (Figure 2). These labeling patterns approximate isolated heteronuclear spin pairs in both the nucleobase (i.e., purine H8-C8 and pyrimidine H6-C6-) and ribose (i.e., H1′-C1′) moieties to reduce spectral crowding and permit NMR measurements free of complications from ^{13}C–^{13}C scalar and dipolar couplings.[40–43]

^{13}C NMR spin relaxation experiments

Transverse relaxation optimized spectroscopy (TROSY)-detected ^{13}C R_1, on-resonance $R_{1\rho}$, and steady-state {^{1}H}-^{13}C hNOE experiments were adapted from previous pulse sequences[57–58] and collected at 298 K on an Avance III Bruker Ultrashield Plus 600 MHz spectrometer with a room temperature triple resonance probe and an 800 MHz Avance III Bruker spectrometer equipped with a triple resonance cryogenic probe. R_1 and $R_{1\rho}$ experiments were each collected in an interleaved manner with relaxation delays specified in Table S9. In these experiments, recycle delays of 1.5 s and 2.5 s were used at 600 MHz and 800 MHz, respectively. R_2 rates were obtained from R_1 and $R_{1\rho}$ according to the following relations:[59]

$$R_{1\rho} = R_1(\cos^2\theta) + R_2(\sin^2\theta) \text{ and } \theta = \tan^{-1}\left(\frac{\omega_1}{\Omega}\right)$$

where ω_1 is the strength of the spin-lock field and Ω is the offset from the spin-lock carrier frequency. In the $R_{1\rho}$ experiments, ω_1 and Ω were 1.9 kHz and 6.0 Hz, respectively. In the saturation hNOE experiments, recycle delays of 1.5 s were used and saturation delays of 5.0 s and 7.0 s were used at 600 MHz and 800 MHz, respectively. Proton saturation was achieved using a hard 180° pulse. In the no saturation hNOE experiments, delays of 6.5 s and 8.5 s were used at 600 and 800 MHz, respectively, to match delays in the saturation experiments. hNOE values were obtained according to the following relations:[60–62]

$$\text{hNOE} = 1 + \left(\frac{I_{sat} - I_{eq}}{I_{eq}}\right)$$

where I_{sat} and I_{eq} are intensities of the ^{13}C resonances when the ^{1}H resonances are saturated or not. Additional experimental parameters are provided in Table S10. ^{13}C relaxation measurements were carried out at 600 MHz and 800 MHz for nucleobase sites and only 800 MHz for ribose sites. The latter was due to worse signal-to-noise and spectral overlap in the ribose region at lower magnetic field.

NMR spectra were processed and analyzed with TopSpin 4.0 and NMRFx processor.[63] R_1 and $R_{1\rho}$ relaxation rates were determined by fitting peak intensities to an exponential decay function using RELAXFIT.[64] Errors in hNOE and R_1 values were estimated by propagating the error in peak intensities from duplicated experiments whereas errors in $R_{1\rho}$ values were estimated from spectral signal-to-noise using RELAXFIT.[64] We did not conduct a Model-Free[65] analysis of the ^{13}C relaxation data given its many assumptions (e.g., molecular rotational anisotropy, ^{13}C–^{13}C dipolar couplings, and uncertainty in nucleobase CSA values, especially in different structural regions). We therefore opted for a qualitative analysis, as previously described.[66] That is, local motions were determined from individual relaxation parameters, whereby elevated R_1 and hNOE values (>1 SD above the mean) and attenuated R_2 values (<1 SD below the mean) indicate fast motions on the ps to ns time scale. For a more robust analysis, we calculated ρ according to the following relation:[45]

$$\rho = \frac{4J(0)}{3J(\omega_C)} \approx \frac{2R_2 - R_1 - HF_2}{R_1 - HF_1}$$

where J(0) and J(ω) are spectral density components at ω = 0 and ω = ω_C and HF_1 and HF_2 are high frequency components of the spectral density which are defined in the original work.[45] ρ calculations were made with ROTDIF.[45] In this analysis, attenuated ρ values (<1 SD below the mean) indicate fast motions on the ps to ns time scale.

^{1}H CPMG experiments

TROSY-detected ^{1}H Carr-Purcell-Meiboom-Gill (CPMG) experiments were adapted from previous pulse sequences[46–47] and collected at 298 K on an 800 MHz Avance III Bruker spectrometer equipped with a triple resonance cryogenic probe. CPMG modules were used to obtain the effective relaxation rate ($R_{2,eff}$) as a function of the frequency at which 180° pulses are applied (ν_{CPMG}) during a constant time relaxation period (T_{relax}).[67] The CPMG module is a set of N repeated blocks of delay (τ)-180° pulse (π)-delay (τ) (τCPMG-π-τCPMG) such that $2 \times N \times \tau CPMG = T_{relax}$. In the CPMG experiment, T_{relax} was set to 40 ms, recycle delays of 1.5 s were used, and the ν_{CPMG} was varied as specified in Table S11. Additional experimental parameters are provided in Table S12. NMR spectra were processed and analyzed with TopSpin 4.0 and NMRFx processor.[63] $R_{2,eff}$ rates were determined according to the following relation:

$$R_{2,eff} = -\frac{1}{T_{relax}} \ln\left(\frac{I(\nu_{CPMG})}{I_0}\right)$$

where I_0 and $I(\nu_{CPMG})$ are peak intensities measured without and with the constant-time relaxation period, respectively. Errors in $R_{2,eff}$ were estimated from spectral signal-to-noise using RELAXFIT.[64]

For R_2 relaxation dispersion, three exchange models exist: (1) no exchange ($R_{2,eff} = R_{2,0}$), (2) fast exchange ($k_{ex} \gg \delta\omega$), and (3) slow exchange (i.e., outside of the fast regime). Fast exchange can be modeled using the Luz-Meiboom equations:[50,68]

$$R_{2,eff} = R_{2,0} + R_{ex}\left[1 - \frac{4\nu_{CPMG}}{k_{ex}}\tanh\left(\frac{k_{ex}}{4\nu_{CPMG}}\right)\right]$$

where $R_{2,0}$ is $R_{2,eff}$ at infinite ν_{CPMG} or no exchange (R_{ex} = 0) and k_{ex} is the exchange rate. In fast exchange, it is not possible to fit p_a, p_b, or $\delta\omega$. Instead, only $R_{2,0}$, R_{ex}, k_{ex}, and δ^{min}_{ppm} can be obtained.[50–52] For exchange outside the fast regime, data can be fit to a model using the Carver Richards equations[52] to extract $R_{2,0}$, k_{ex}, p_a, p_b, and $\delta\omega$.

The presence of chemical exchange and the appropriate exchange regime (i.e., fast, slow, or no exchange) was identified and selected based on the AIC[51,53] (see Methods and Tables S4 and S5). To increase the robustness of our results, the CPMG data were fit to both ShereKhan[69] and *rdnmr* (version 1.5)[70] and the outputs were compared across the three programs. All exchange processes were determined to be in the fast exchange regime, and the extracted parameters from each analysis showed strong agreement (Table S6).

MD simulations

The Amber20 software package[71] was used to perform MD simulations with the ff99LJbb[72] force field (source file leaprc.RNA.LJbb), which combines the OL3[73] parameter set, the Steinbrecher and Case phosphate oxygen van der Waals radii,[74] and the OPC water model[75–76] to model RNA. The receptor PDB files were input into the Amber LEaP module which combined them with OPC waters, Joung-Cheatham[77] monovalent ions (Na^+/Cl^-), and the RNA-specific force field parameters mentioned above to generate the topology and coordinate files. Explicit solvent molecular particle mesh Ewald (PME) dynamics simulations were utilized.[78] RNA was placed in a cuboid solvent box with OPC waters and the minimum distance between the solute and solvent box boundary was set at 12 Å. The net solute charge was neutralized with Na^+ ions, and additional Na^+/Cl^- ion pairs were added to simulate the net 150 mmol/L salt concentration for the entire system. Simulations were run with 2 fs time steps, employing the SHAKE algorithm to constrain all hydrogen bonds in the system. The Berendsen thermostat[79] and algorithm were used to maintain the simulation temperature of 300 K and to maintain the pressure at 1.0 Pa in NPT simulations used in all phases of MD. A cut-off of 9 Å for the non-bonded interactions was used and explicit solvent periodic boundary conditions were employed.

A 12-step equilibration protocol was used in all simulations that started with energy minimization of the solvent (while the RNA was restrained), followed by multiple short phases of heating to 300 K, dynamics at 300 K, and energy minimizations with gradually decreasing harmonic restraints applied to the solute. The last phase of the equilibration protocol was an unrestrained heating to 300 K, ramped up over 0.2 ns and kept at the steady target temperature for a total time of 2 ns. Following equilibration, three independent, unrestrained (production) MD simulations were performed for 1 μs for ε R3 (PDB code 6VAR[35]) (i.e., $R3^a$, $R3^b$, and $R3^c$).

Repeated simulations on all ε NMR conformers R1-R10 (PDB code 6VAR[35]) were solvated in an octahedral TIP4Pew[80] water box, with a 10 Å buffer between the solute and edge of the box. The ff99LJbb forcefield previously described was used, combined with Joung-Cheatham[77] monovalent ion parameters. The net solute charge was neutralized with Na^+ ions, and additional Na^+/Cl^- ion pairs were added to simulate the net 10 mmol/L salt concentration for the entire system. The protocol described by Roe and Brooks was used to prepare the system for dynamics.[81] MD were run in constant volume using the pmemd.cuda.MPI implementation in Amber20.[82] The Langevin thermostat[83] was used to maintain temperature at 300 K, with a collision frequency of 2 ps^{-1}. Random seeds were used for each restart to avoid synchronization artifacts.[84] Hydrogen mass repartitioning was used so a 4 fs timestep could be employed.[85] SHAKE and default PME settings were used as described above.

The Amber CPPTRAJ[86] module was used for analysis. RMSD calculations were performed using the MD trajectories (excluding the 2 ns equilibrations) with ε R3 used as the reference structure. The 1 μs trajectories were sampled in 0.1 ns steps to yield 10,000 data points and calculations were made considering all atoms within a 3-nt window (i.e., G1-U3,. . .,A59-C61). K-means clustering analysis was employed individually for the set of three ε R3-based 1 μs for MD runs or as an ensemble (Supplementary Scripts S2 and S3). Clustering was carried out using either the global RMSD of the backbone phosphorus atom of nucleotides G2-C61 or backbone and chi dihedral angles (i.e., α, β, γ, δ, ε, ζ, and χ) from PL (i.e., A13-A20), PTL (i.e., C31-C36), or LH (i.e., G2-A13 and U49-C61) nucleotides. From these data, five representative centroid structures can be selected that minimize the distance between all other points in the cluster. CPPTRAJ was also used for measuring the dynamic cross-correlation matrix,[54] interhelical angle Θ, and PCA (Supplementary Scripts, S4-S6).

Each set of simulations were carried out using different simulation conditions that approximate either the *in vivo* environment (i.e., 150 mmol/L NaCl as in $R3^a$, $R3^b$, and $R3^c$) or the ε NMR buffer (i.e., 10 mmol/L NaCl as in R1-R10). Moreover, the three independent simulations on R3 were run using OPC water model and the ff99LJbb forcefield. The follow-up MD runs on R1-R10 were carried out with the TIP4Pew water model and the ff99LJbb forcefield. It is important to note that the

same RNA forcefield was used and only the water model varied. Based on previous reports,[72] we do not expect our MD analysis or interpretations to be impacted by using either the TIP4P-Ew or OPC water model. This is evidenced in our interhelical angle Θ (Figure 10(a, b)), PCA (Figure 10(c)), and dynamic cross-correlation matrix (Figures 9 and S9) analyses, which show that data from R3^a, R3^b, and R3^c lie within the range of data from R1-R10. Moreover, previous NMR experiments suggest that ε experiences minimal chemical shift perturbations with different buffer conditions.[35]

CRediT authorship contribution statement

Lukasz T. Olenginski: Conceptualization, Investigation, Writing – original draft. **Wojciech K. Kasprzak:** Conceptualization, Investigation, Writing – original draft. **Christina Bergonzo:** Conceptualization, Investigation, Writing – original draft. **Bruce A. Shapiro:** Conceptualization, Supervision, Writing – original draft. **Theodore K. Dayie:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft.

Acknowledgements

We thank Bruce A. Johnson for technical help with RING NMR Dynamics and Regan M. Leblanc for NMR assistance and thoughtful discussion. This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research (B.A.S.) and with Federal funds from the Frederick National Laboratory for Cancer Research, NIH, under contract number [HHSN261200800001E to W.K. K.]. Funding for 800 MHz NMR instrumentation was supported by NSF [DBI1040158 to T.K.D.] and isotopic labeling work was funded by NSF [1808705 to T.K.D.].

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Certain commercial equipment, instruments, and materials are identified in this article in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2022.167633.

Received 1 February 2022;
Accepted 7 May 2022;
Available online 17 May 2022

Keywords:
hepatitis B virus;
viral replication;
dynamics;
NMR spectroscopy;
conformational selection

Abbreviations:
HBV, hepatitis B virus; pgRNA, pre-genomic RNA; C, HBV core protein; P, HBV polymerase protein; RT, reverse transcriptase; RH, RNase H; TP, terminal protein; ε, epsilon; nt, nucleotide; PL, priming loop; DR1, direct repeat 1; HSPs, heat shock proteins; HFs, host factors; LH, lower helix; UH, upper helix; PTL, pseudo-triloop; NMR, nuclear magnetic resonance; AL, apical loop; SELEX, Systematic Evolution of Ligands by Exponential Enrichment; MD, molecular dynamics; R_1, longitudinal relaxation rate; R_2, transverse relaxation rate; hNOE, $\{^1H\}$-^{13}C heteronuclear Overhauser effect; CSA, chemical shift anisotropy; SD, standard deviation; PDB, Protein Data Bank; $R_{1\rho}$, rotating-frame relaxation rate; R_{ex}, chemical exchange contribution to R_2; CPMG, Carr-Purcell-Meiboom-Gill; $R_{2,eff}$, effective R_2 rate; ν_{CPMG}, frequency of 180° refocusing pulse in CPMG experiment; k_{ex}, exchange rate; p_a, population of major state; p_b, population of minor state; $\delta\omega$, chemical shift difference between major and minor state; $R_{2,0}$, R_2 auto-relaxation; δ^{min}_{ppm}, $\delta\omega$ that would be present if the p_a and p_b were equal; AIC, Akaike Information Criterion; τ_{ex}, exchange lifetime; RMSD, root-mean-square-deviation; norm, normalized; 2D, two-dimensional; CEST, chemical exchange saturation transfer (CEST); TROSY, transverse relaxation optimized spectroscopy; J(0), spectral density component at $\omega = 0$; $J(\omega_C)$, spectral density component at $\omega = \omega_C$; HF_1, high frequency component 1 of the spectral density; HF_2, high frequency component 2 of the spectral density

References

1. MacLachlan, J.H., Cowie, B.C., (2015). Hepatitis B virus epidemiology. *Cold Spring Harb. Perspect. Med.* **5**, a021410
2. Stanaway, J.D., Flaxman, A.D., Naghavi, M., Fitzmaurice, C., Vos, T., Abubakar, I., Abu-Raddad, L.J., Assadi, R., Bhala, N., Cowie, B., Forouzanfour, M.H., Groeger, J., Hanafiah, K.M., Jacobsen, K.H., James, S.L., MacLachlan, J., Malekzadeh, R., Martin, N.K., Mokdad, A.A., Mokdad, A. H., Murray, C.J.L., Plass, D., Rana, S., Rein, D.B., Richardus, J.H., Sanabria, J., Saylan, M., Shahraz, S.,

So, S., Vlassov, V.V., Weiderpass, E., Wiersma, S.T., Younis, M., Yu, C., El Sayed Zaki, M., Cooke, G.S., (2016). The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. *Lancet* **388**, 1081–1088.
3. Ott, J.J., Stevens, G.A., Groeger, J., Wiersma, S.T., (2012). Global epidemiology of hepatitis B virus infection: New estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* **30**, 2212–2219.
4. Block, T.M., Rawat, S., Brosgart, C.L., (2015). Chronic hepatitis B: A wave of new therapies on the horizon. *Antiviral Res.* **121**, 69–81.
5. Zampino, R., Sagnelli, C., Boemio, A., Sagnelli, E., Coppola, N., (2016). Treatment of chronic HBV infection in developing countries. *Ann. Hepatol.* **15**, 816–823.
6. Bhattacharya, D., Thio, C.L., (2010). Review of hepatitis B therapeutics. *Clin. Infect. Dis.* **51**, 1201–1208.
7. Summers, J., Mason, W.S., (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* **29**, 403–415.
8. Rall, L.B., Standring, D.N., Laub, O., Rutter, W.J., (1983). Transcription of hepatitis B virus by RNA polymerase II. *Mol. Cell. Biol.* **3**, 1766–1773.
9. Hu, J., Seeger, C., (1996). Expression and characterization of hepadnavirus reverse transcriptases. *Methods Enzymol.* **275**, 195–208.
10. Radziwill, G., Tucker, W., Schaller, H., (1990). Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. *J. Virol.* **64**, 613–620.
11. Toh, H., Hayashida, H., Miyata, T., (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature* **305**, 827–829.
12. Zoulim, F., Seeger, C., (1994). Reverse transcription in hepatitis B viruses is primed by a tyrosine residue of the polymerase. *J. Virol.* **68**, 6–13.
13. Weber, M., Bronsema, V., Bartos, H., Bosserhoff, A., Bartenschlager, R., Schaller, H., (1994). Hepadnavirus P protein utilizes a tyrosine residue in the TP domain to prime reverse transcription. *J. Virol.* **68**, 2994–2999.
14. Wang, G.H., Seeger, C., (1992). The reverse transcriptase of hepatitis B virus acts as a protein primer for viral DNA synthesis. *Cell* **71**, 663–670.
15. Junker-Niepmann, M., Bartenschlager, R., Schaller, H., (1990). A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J.* **9**, 3389–3396.
16. Hirsch, R.C., Lavine, J.E., Chang, L.J., Varmus, H.E., Ganem, D., (1990). Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* **344**, 552–555.
17. Pollack, J.R., Ganem, D., (1993). An RNA stem-loop structure directs hepatitis B virus genomic RNA encapsidation. *J. Virol.* **67**, 3254–3263.
18. Knaus, T., Nassal, M., (1993). The encapsidation signal on the hepatitis B virus RNA pregenome forms a stem-loop structure that is critical for its function. *Nucleic Acids Res.* **21**, 3967–3975.
19. Wang, G.H., Zoulim, F., Leber, E.H., Kitson, J., Seeger, C., (1994). Role of RNA in enzymatic activity of the reverse transcriptase of hepatitis B viruses. *J. Virol.* **68**, 8437–8442.
20. Zhou, S., Standring, D.N., (1992). Hepatitis B virus capsid particles are assembled from core-protein dimer precursors. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10046–10050.
21. Bartenschlager, R., Schaller, H., (1992). Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* **1**, 3413–3420.
22. Nassal, M., Rieger, A., (1996). A bulged region of the hepatitis B virus RNA encapsidation signal contains the replication origin for discontinuous first-strand DNA synthesis. *J. Virol.* **70**, 2764–2773.
23. Lanford, R.E., Notvall, L., Beames, B., (1995). Nucleotide priming and reverse transcriptase activity of hepatitis B virus polymerase expressed in insect cells. *J. Virol.* **69**, 4431–4439.
24. Fallows, D.A., Goff, S.P., (1995). Mutations in the epsilon sequences of human hepatitis B virus affect both RNA encapsidation and reverse transcription. *J. Virol.* **69**, 3067–3073.
25. Rieger, A., Nassal, M., (1996). Specific hepatitis B virus minus-strand DNA synthesis requires only the 5′ encapsidation signal and the 3′-proximal direct repeat DR1. *J. Virol.* **70**, 585–589.
26. Tavis, J.E., Perri, S., Ganem, D., (1994). Hepadnavirus reverse transcription initiates within the stem-loop of the RNA packaging signal and employs a novel strand transfer. *J. Virol.* **68**, 3536–3543.
27. Beck, J., Nassal, M., (1998). Formation of a Functional Hepatitis B Virus Replication Initiation Complex Involves a Major Structural Alteration in the RNA Template. *Mol. Cell. Biol.* **18**, 6265–6272.
28. Hu, J., Boyer, M., (2006). Hepatitis B virus reverse transcriptase and ε RNA sequences required for specific interaction *in vitro*. *J. Virol.* **80**, 2141–2150.
29. Jones, S.A., Boregowda, R., Spratt, T.E., Hu, J., (2012). *In vitro* epsilon RNA-dependent protein priming activity of human hepatitis B virus polymerase. *J. Virol.* **86**, 5134–5150.
30. Feng, H., Chen, P., Zhao, F., Nassal, M., Hu, K., (2013). Evidence for multiple distinct interactions between hepatitis B virus P protein and its cognate RNA encapsidation signal during initiation of reverse transcription. *PLoS ONE* **8**, e72798
31. Hu, J., Flores, D., Toft, D., Wang, X., Nguyen, D., (2004). Requirement of heat shock protein 90 for human hepatitis B virus reverse transcriptase function. *J. Virol.* **78**, 13122–13131.
32. Hu, J., Toft, O.D., Seeger, C., (1997). Hepadnavirus assembly and reverse transcription require a multi-component chaperone complex which is incorporated into nucleocapsids. *EMBO J.* **16**, 59–68.
33. Hu, J., Toft, D., Anselmo, D., Wang, X., (2002). *In vitro* reconstitution of functional hepadnavirus reverse transcriptase with cellular chaperone proteins. *J. Virol.* **76**, 269–279.
34. Hu, J., Seeger, C., (1996). Hsp90 is required for the activity of a hepatitis B virus reverse transcriptase. *Proc. Natl. Acad. Sci.* **93**, 1060–1064.
35. LeBlanc, R.M., Kasprzak, W.K., Longhini, A.P., Olenginski, L.T., Abulwerdi, F., Ginocchio, S., Shields, B., Nyman, J., Svirydava, M., Del Vecchio, C., Ivanic, J., Schneekloth, J. S., Shapiro, B.A., Dayie, T.K., Le Grice, S.F.J., (2021). Structural insights of the conserved “priming loop” of hepatitis B virus pre-genomic RNA. *J. Biomol. Struct. Dyn.*, 1–13.
36. Flodell, S., Petersen, M., Girard, F., Zdunek, J., Kidd-Ljunggren, K., Schleucher, J., Wijmenga, S., (2006).

Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res.* **34**, 4449–4457.
37. Flodell, S., Schleucher, J., Cromsigt, J., Ippel, H., Kidd-Ljunggren, K., Wijmenga, S., (2002). The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.* **30**, 4803–4811.
38. Petzold, K., Duchardt, E., Flodell, S., Larsson, G., Kidd-Ljunggren, K., Wijmenga, S., Schleucher, J., (2007). Conserved nucleotides in an RNA essential for hepatitis B virus replication show distinct mobility patterns. *Nucleic Acids Res.* **35**, 6854–6861.
39. Feng, H., Beck, J., Nassal, M., Hu, K., (2011). A SELEX-screened aptamer of human hepatitis B virus RNA encapsidation signal suppresses viral replication. *PLoS ONE* **6**, e27862
40. Alvarado, L.J., Leblanc, R.M., Longhini, A.P., Keane, S.C., Jain, N., Yildiz, Z.F., Tolbert, B.S., D'Souza, V.M., Summers, M.F., Kreutz, C., Dayie, T.K., (2014). Regio-selective chemical-enzymatic synthesis of pyrimidine nucleotides facilitates RNA structure and dynamics studies. *ChemBioChem* **15**, 1573–1577.
41. Longhini, A.P., Leblanc, R.M., Becette, O., Salguero, C., Wunderlich, C.H., Johnson, B.A., D'Souza, V.M., Kreutz, C., Dayie, T.K., (2015). Chemo-enzymatic synthesis of site-specific isotopically labeled nucleotides for use in NMR resonance assignment, dynamics and structural characterizations. *Nucleic Acids Res.* **44**, e52
42. Olenginski, L.T., Dayie, T.K., (2021). Quantifying the effects of long-range ^{13}C–^{13}C dipolar coupling on measured relaxation rates in RNA. *J. Biomol. NMR* **75**, 203–211.
43. Nam, H., Becette, O., LeBlanc, R.M., Oh, D., Case, D.A., Dayie, T.K., (2020). Deleterious effects of carbon–carbon dipolar coupling on RNA NMR dynamics. *J. Biomol. NMR* **74**, 321–331.
44. Boisbouvier, J., Wu, Z., Ono, A., Kainosho, M., Bax, A., (2003). Rotational diffusion tensor of nucleic acids from ^{13}C NMR relaxation. *J. Biomol. NMR* **27**, 133–142.
45. Berlin, K., Longhini, A., Dayie, T.K., Fushman, D., (2013). Deriving quantitative dynamics information for proteins and RNAs using ROTDIF with a graphical user interface. *J. Biomol. NMR* **57** (2013), 333–352.
46. Leblanc, R.M., Longhini, A.P., Tugarinov, V., Dayie, T.K., (2018). NMR probing of invisible excited states using selectively labeled RNAs. *J. Biomol. NMR* **71**, 165–172.
47. Juen, M.A., Wunderlich, C.H., Nußbaumer, F., Tollinger, M., Kontaxis, G., Konrat, R., Hansen, D.F., Kreutz, C., (2016). Excited states of nucleic acids probed by proton relaxation dispersion NMR spectroscopy. *Angew. Chemie Int. Ed.* **128**, 12187–12191.
48. Ishima, R., Torchia, D.A., (2003). Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. *J. Biomol. NMR* **25**, 243–248.
49. Vallurupalli, P., Scott, L., Williamson, J.R., Kay, L.E., (2007). Strong coupling effects during X-pulse CPMG experiments recorded on heteronuclear ABX spin systems: artifacts and a simple solution. *J. Biomol. NMR* **38**, 41–46.
50. Luz, Z., Meiboom, S., (1963). Nuclear magnetic resonance study of the protolysis of trimethylammonium ion in aqueous solution-order of the reaction with respect to solvent. *J. Chem. Phys.* **39**, 366–370.
51. Beckwith, M.A., Erazo-Colon, T., Johnson, B.A., (2021). RING NMR dynamics: software for analysis of multiple NMR relaxation experiments. *J. Biomol. NMR* **75**, 9–23.
52. Carver, J.P., Richards, R.E., (1972). A general two-site solution for the chemical exchange produced dependence of T_2 upon the Carr-Purcell pulse separation. *J. Magn. Reson.* **6**, 89–105.
53. Cavanaugh, J.E., Neath, A.A., (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdiscip. Rev. Comput. Stat.* **11**, e1460
54. Karplus, M., Ichiye, T., (1995). Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* **252**, 492–503.
55. Beck, J., Nassal, M., (1997). Sequence- and structure-specific determinants in the interaction between the RNA encapsidation signal and reverse transcriptase of avian hepatitis B viruses. *J. Virol.* **71**, 4971–4980.
56. Milligan, J.F., Uhlenbeck, O.C., (1989). Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62.
57. Hansen, A.L., Al-Hashimi, H.M., (2007). Dynamics of large elongated RNA by NMR carbon relaxation. *J. Am. Chem. Soc.* **129**, 16072–16082.
58. Lakomek, N.A., Kaufman, J.D., Stahl, S.J., Louis, J.M., Grishaev, A., Wingfield, P.T., Bax, A., (2013). Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales. *Angew. Chemie Int. Ed.* **52**, 3911–3915.
59. Akke, M., Palmer, A.G., (1996). Monitoring macromolecular motions on microsecond to millisecond time scales by $R_{1\rho}$–R_1 constant relaxation time NMR spectroscopy. *J. Am. Chem. Soc.* **118**, 911–912.
60. Palmer, A.G., Rance, M., Wright, P.E., (1991). Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**, 4371–4380.
61. Eldho, N.V., Dayie, K.T., (2007). Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: Implications for catalysis. *J. Mol. Biol.* **365**, 930–944.
62. Kay, L.E., Torchia, D.A., Bax, A., (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to *Staphylococcal* nuclease. *Biochemistry* **28**, 8972–8979.
63. Norris, M., Fetler, B., Marchant, J., Johnson, B.A., (2016). NMRFx Processor: a cross-platform NMR data processing program. *J. Biomol. NMR* **65**, 205–216.
64. Fushman, D., Cahill, S., Cowburn, D., (1997). The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: Analysis of ^{15}N relaxation with monomer/dimer equilibration. *J. Mol. Biol.* **266**, 173–194.
65. Lipari, G., Szabo, A., (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559.
66. Shajani, Z., Drobny, G., Varani, G., (2007). Binding of U1A protein changes RNA dynamics as observed by ^{13}C NMR relaxation studies. *Biochemistry* **46**, 5875–5883.
67. Mulder, F.A.A., Skrynnikov, N.R., Hon, B., Dahlquist, F.W., Kay, L.E., (2001). Measurement of slow (μs-ms) time scale dynamics in protein side chains by ^{15}N relaxation

dispersion NMR spectroscopy: Application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.* **123**, 967–975.
68. Palmer, A.G., Koss, H., (2019). Chemical Exchange. *Methods Enzymol.* **615**, 177–236.
69. Mazur, A., Hammesfahr, B., Griesinger, C., Lee, D., Kollmar, M., (2013). ShereKhan—calculating exchange parameters in relaxation dispersion data from CPMG experiments. *Bioinformatics* **29**, 1819–1820.
70. Strebitzer, E., Nußbaumer, F., Kremser, J., Tollinger, M., Kreutz, C., (2018). Studying sparsely populated conformational states in RNA combining chemical synthesis and solution NMR spectroscopy. *Methods* **148**, 39–47.
71. Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J., (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688.
72. Bergonzo, C., Cheatham, T.E., (2015). Improved force field parameters lead to a better description of RNA structure. *J. Chem. Theory Comput.* **11**, 3969–3972.
73. Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham, T.E., Jurečka, P., (2011). Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902.
74. Steinbrecher, T., Latzer, J., Case, D.A., (2012). Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.* **8**, 4405–4412.
75. Izadi, S., Anandakrishnan, R., Onufriev, A.V., (2014). Building water models: A different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871.
76. Mukhopadhyay, A., Fenley, A.T., Tolokh, I.S., Onufriev, A. V., (2012). Charge hydration asymmetry: The basic principle and how to use it to test and improve water models. *J. Phys. Chem. B* **116**, 9776–9783.
77. Joung, I.S., Cheatham, T.E., (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041.
78. Cheatham, T.E., Miller, J.L., Fox, T., Darden, T.A., Kollman, P.A., (1995). Molecular dynamics simulations on solvated biomolecular systems: The particle mesh ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **117**, 4193–4194.
79. Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., Dinola, A., Haak, J.R., (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
80. Horn, H.W., Swope, W.C., Pitera, J.W., Madura, J.D., Dick, T.J., Hura, G.L., Head-Gordon, T., (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **120**, 9665.
81. Roe, D.R., Brooks, B.R., (2020). A protocol for preparing explicitly solvated systems for stable molecular dynamics simulations. *J. Chem. Phys.* **153**, 054123
82. Salomon-Ferrer, R., Götz, A.W., Poole, D., Le Grand, S., Walker, R.C., (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* **9**, 3878–3888.
83. Loncharich, R.J., Brooks, B.R., Pastor, R.W., (1992). Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **32**, 523–535.
84. Sindhikara, D.J., Kim, S., Voter, A.F., Roitberg, A.E., (2009). Bad seeds sprout perilous dynamics: Stochastic thermostat induced trajectory synchronization in biomolecules. *J. Chem. Theory Comput.* **5**, 1624–1631.
85. Hopkins, C.W., Le Grand, S., Walker, R.C., Roitberg, A.E., (2015). Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874.
86. Roe, D.R., Cheatham, T.E., (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095.