PROTEINS WILEY

Computational modeling reveals key molecular properties and dynamic behavior of disruptor of telomeric silencing 1-like (*DOT1L*) and partnering complexes involved in leukemogenesis

Timothy J. Stodola[1,2] | Young-In Chi[1,2] | Thiago M. De Assuncao[1,3] |
Elise N. Leverence[1] | Swarnendu Tripathi[1,2] | Nikita R. Dsouza[1,2] |
Angela J. Mathison[1,2,3] | Brian F. Volkman[4] | Brian C. Smith[4] |
Gwen Lomberk[1,3,5] | Michael T. Zimmermann[1,2,4,6] | Raul Urrutia[1,2,3,4]

[1]Genomic Sciences and Precision Medicine Center (GSPMC), Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[2]Bioinformatics Research and Development Laboratory, and Precision Medicine Simulation Unit, GSPMC, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[3]Division of Research, Department of Surgery, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[4]Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[5]Department of Pharmacology and Toxicology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[6]Clinical and Translational Sciences Institute, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

Correspondence
Raul Urrutia, Genomic Sciences and Precision Medicine Center (GSPMC), Medical College of Wisconsin, Milwaukee, WI, USA.
Email: rurrutia@mcw.com

Abstract

Disruptor of telomeric silencing 1-like (*DOT1L*) is the only non-SET domain histone lysine methyltransferase (KMT) and writer of H3K79 methylation on nucleosomes marked by H2B ubiquitination. DOT1L has elicited significant attention because of its interaction or fusion with members of the AF protein family in blood cell biology and leukemogenic transformation. Here, our goal was to extend previous structural information by performing a robust molecular dynamic study of DOT1L and its leukemogenic partners combined with mutational analysis. We show that statically and dynamically, D161, G163, E186, and F223 make frequent time-dependent interactions with SAM, while additional residues T139, K187, and N241 interact with SAM only under dynamics. Dynamics models reveal DOT1L, SAM, and H4 moving as one and show that more than twice the number of DOT1L residues interacts with these partners, relative to the static structure. Mutational analyses indicate that six of these residues are intolerant to substitution. We describe the dynamic behavior of DOT1L interacting with AF10 and AF9. Studies on the dynamics of a heterotrimeric complex of DOT1L1-AF10 illuminated describe coordinated motions that impact the relative position of the DOT1L HMT domain to the nucleosome. The molecular motions of the DOT1L–AF9 complex are less extensive and highly dynamic, resembling a swivel-like mechanics. Through molecular dynamics and mutational analysis, we extend the knowledge previous provided by static measurements. These results are important to consider when describing the biochemical properties of DOT1L, under normal and in disease conditions, as well as for the development of novel therapeutic agents.

KEYWORDS
AF proteins, DOT1L, methyltransferase, molecular dynamics, protein complex, SAM, structural analysis

1 | INTRODUCTION

Epigenetics bestows regulatory mechanisms that underlie the faithful transmission of the instructions inherited through the genome.[1] Not surprisingly, epigenomic abnormalities are found in most diseases, and in particular, cancer.[2,3] These epigenomic dysfunctions occur at many levels, most often at the level of enzymatic function of key epigenomic regulators that act across the genome.[4] For their role in shaping cell fate, there is great interest in epigenetic regulatory enzymes as drug targets that could sensitize or re-sensitize tumors to treatments. Mutations in genes that encode epigenetic regulators (epi-mutations) have recently emerged as a cause of several cancers, both in pediatric and adult populations,[5–7] and little investigation has occurred for epimutational dependence on therapies. Thus, studies on the molecular mechanism underlying the functions of these genetic alterations, to better understand the contributions of each amino acid to enzyme structure and function, is an area of need for clinical genomics and precision oncology.

DOT1L (disruptor of telomeric silencing 1-like) is a conserved epigenomic regulator. Its enzymatic methyltransferase region adds methyl groups to H3 lysine 79, forming H3K79me1, H3K79me2, and H3K79me3 using S-adenosyl methionine (SAM) as the methyl donor.[8] The H3K79 methylation marks have been shown correlate with open chromatin and active transcription.[9,10] The H3K79me1/me2/me3 mark is found in gene bodies and is associated with open chromatin, cell-cycle regulation, and the DNA damage response.[11,12] Additionally, H3K79me2/me3mark is found at a subset of putative enhancer elements, where this mark is necessary to maintain open chromatin and promote histone acetylation in MLL-AF4 fusion cells.[13]

DOT1L is the only human methyltransferase (MTase) that does not contain a SET domain. Instead, DOT1L belongs to an ancient class of proteins known as seven-β-strand (7BS) MTases, known for their conserved structure that contains a Rossmann fold.[14] This fold is present in myriad proteins across all kingdoms of life where they modify not only proteins but also lipids and small molecules.[15–17] Recent studies have led to the discovery of novel 7BS lysine MTases (KMTs), like those that belong to the yeast MTase family 16 (MTF16) which methylate lysine residues in non-histone proteins.[18,19] Therefore, studying the biophysical and biochemical features of DOT1L is important to understand the mechanistic functions of DOT1L and these related KMTs at a fundamental level, in addition to the role of DOT1L in leukemogenesis.[12]

One of the longest known examples of genomic alterations affecting epigenomic regulators are oncogenic fusion proteins.[20] Oncogenic fusion proteins are found in a variety of cancers spanning across the three germ layers.[21] In particular, oncogenic fusion of mixed lineage leukemia (MLL) proteins with different AF (acute lymphoid leukemia [ALL1]-fused gene) proteins induces leukemogenesis in humans.[12,22–25] Notably, all these fusion proteins cause leukemia by using the AF region of the fusion protein to recruit DOT1L to MLL-targeted genes, leading to inappropriate activation of genes in these cells.[12,26]

Here, our goal was to extend previous structural information by performing a robust molecular dynamic study of DOT1L in complex with cofactor SAM, key nucleosome interaction partners histone H4 and ubiquitin, and leukemogenic partners AF10 and AF9. We interrogated the structural and dynamic features of DOT1L within these complexes that are important for DOT1L's function and localization. The results reported here provide the first characterization of the time-dependent behavior of a key functional domain of DOT1L, coordinated movement and stability of interactions after mutation. This is important as many experimental structures are available for DOT1L and yet, insight into how the structure of DOT1L produces motions of functional consequence remains to be fully understood. Therefore, since DOT1L is a well-known epigenomic regulator and is targeted in MLL-AF fusion leukemias, this new knowledge bears both biochemical and biomedical relevance.[12] We envision these studies may aid the development of small drugs to inhibit key disease-relevant functions of these complexes.

2 | METHODS

The modeling for DOT1L alone and in complexes was done using previously solved experimental structures downloaded from RCSB PDB (rcsb.org)[27] and plotted in Figure 1A. The structure of the human DOT1L enzymatic region with substrate SAM has been solved with the natural S-adenosyl-L-methionine (SAM) cofactor and by two groups with nearly identical results (1NW3[28] and 3QOW[29]). Moreover, DOT1L bound to an H2B-ubiquitinated nucleosome has been resolved using cryo-electron microscopy[30–34] and includes a model of DOT1L H3 methylation in the active state (PDB ID 6NQA[30]). Two additional structures have been resolved, one through X-ray diffraction of an octamer of heterodimers of DOT1L and AF10 (PDB ID 6JN2[35]), and the other through NMR for DOT1L heterodimer complexes with AF9 (PDB ID 2MV7[36]).

Modeling and molecular dynamics (MD) analysis for each structure were performed as previously described.[37] Briefly, each RCSB PDB structure 1NW3, 3QOW, 6NQA, 6JN2, and 2MV7 was used as a template and refined with two cycles of 1000 steps of steepest descent and one cycle with 1000 steps of conjugate gradient. MD simulations were performed using the CHARMM22 force field[38] for the initial 2 ns and extended 10 ns with NAMD 2.13,[39] with each replicate using unique random number seeds, NPT ensemble, and implicit solvent. Visualizations were developed with Discovery Studio (Biovia).[40]

Root-mean-square fluctuation (RMSF) values were calculated on the residue level as average of atomic RMSF across trajectories aligned to the initial conformation. For each conformation, the radius of gyration and root-mean-square deviation (RMSD) values were calculated, the latter after trajectories aligned and relative to the initial conformation. Interacting residues were assessed for effects of mutations. For each interacting residue, the effect of mutations on the binding affinity with SAM, H4, ubiquitin, AF10, or AF9 ($\Delta\Delta G_{binding}$, kcal/mol) was calculated as the difference in the binding free energy
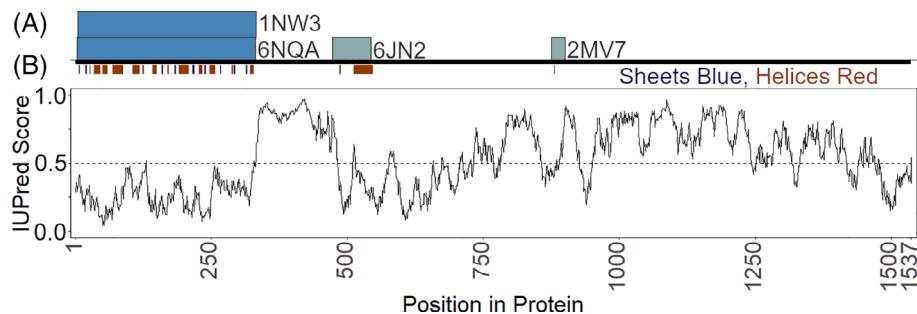
FIGURE 1 Overview of DOT1L protein and available 3D structures. (A) DOT1L consists of an enzymatic histone methyltransferase region (PDB 1NW3 and 6NQA, near *N*-terminus of the protein) and interprotein interactions with AF10 (PDB 6JN2) and AF9 (PDB 2MV7). Locations of sheets and helices are shown below the structures. (B) The largest structured region of the protein aligns with the enzymatic region, which is followed by a disordered region. The remainder of the protein alternates between ordered and disordered regions

in the mutated versus the wild type protein at pH 7.4. Mutation energies were reported as follows: stabilizing (blue) mutation energy is less than -0.5 kcal/mol, destabilizing (red) mutation energy is greater than 0.5 kcal/mol, and neutral (white) is in between.

3 | RESULTS

3.1 | Molecular dynamics predict the static structure under-estimates DOT1L interaction with SAM

3.1.1 | Static structure and interactions between DOT1L and SAM

To understand interactions with SAM, we first characterized the interactions present in the static experimental structure. The enzymatic region of DOT1L is at the *N*-terminal of the protein (Figure 1A). The structure can be topologically divided into three distinct domains, namely: (1) a DOT1L-specific histone-binding domain, (2) a KMT domain containing the SAM-binding site, and (3) the ubiquitin-based activation domain (Figure 2A). The SAM-binding site adopts a funnel-like configuration with a wider hydrophobic entrance, which accommodates the adenine group and a narrower, negatively charged region that fits the methionyl moiety. Residues in DOT1L contributing to these interactions included P133, E138, T139, D161, G163, Q168, E186, K187, G221, D222, and F223, which together created 19 hydrogen bonds and 4 π interactions to support binding (Table 1, Figure 2B,C). Residues from both the histone-binding region and KMT domain Rossmann folds were interacting with SAM. Notably, despite several charged amino acids from DOT1L within the binding site, no salt bridge participated in binding. SAM contributed 286 $Å^2$ of contact surface area, composed of a 161 $Å^2$ of polar surface and 125 $Å^2$ nonpolar surface. Complementarily, DOT1L displayed a binding pocket area of 309 $Å^2$, with a polar surface of 180 $Å^2$ and 121 $Å^2$ nonpolar surface. Thus, the SAM cofactor makes numerous interactions that are likely important for DOT1L function.

3.1.2 | General dynamics

To study the coordinated pattern of the enzyme's molecular movements, we analyzed simulations of multiple time lengths, which can be quantitatively described using different measurements. RMSF showed that both the *N*-terminal histone binding and the *C*-terminal ubiquitin binding domains were the most mobile regions of DOT1L (Figure 2D). The behavior of within these regions can be characterized as highly mobile loops alternating with less dynamic helical regions, except for the *C*-terminal helix. This terminal helix, however, would likely behave differently in the context of the full-length protein, since it connects the nucleosome-binding enzymatic region (PDB ID: 6NQA[30]) to a disordered region followed by coiled-coiled interactions with AF10 proteins (PDB ID: 6JN2[35]) or an alpha-beta fold interaction with AF9 (2MV7,[36] Figure 1A). Mobile regions flank a domain defined by low RMSF values, which corresponds to the conserved KMT core surrounding the cofactor SAM. An additional high-RMSF region corresponds to a loop that aligns the DOT1L enzymatic region on the acid patch of H2A/H2B.[30] Therefore, these results represent the coordinated motions that define the dynamics of the enzymatic region of DOT1L bound to SAM, showing that the DOT1L–SAM complex moves as a unit.

Principal component analysis was used as a data reduction method to isolate the predominant molecular motions observed over the combined set of 10 ns MD trajectories across replicates. These motions were summarized by RMSF and the Radius of Gyration of each of the first three PCs (Supplemental Figure S1A,B). The largest contributions to these motions came from the ubiquitin-interacting *C*-terminal helix, as well as the loops and small helix in the region spanning the H4 and SAM interactions (Figure S1C). Thus, the DOT1L structural flexibility is at regions that facilitate key interactions that lead to aligning DOT1L on the nucleosome through H4 and ubiquitin.

Comparison of the initial structure against the time-dependent motions simulations revealed modifications in the molecular volume, shape, and secondary structure. Chief among them was a more compact general structure along with changes in secondary structures within the *N*-terminal domain. In fact, the molecular volume of DOT1L decreased from 53 749 $Å^3$ before simulation to 50 237 ± 94 $Å^3$ at the
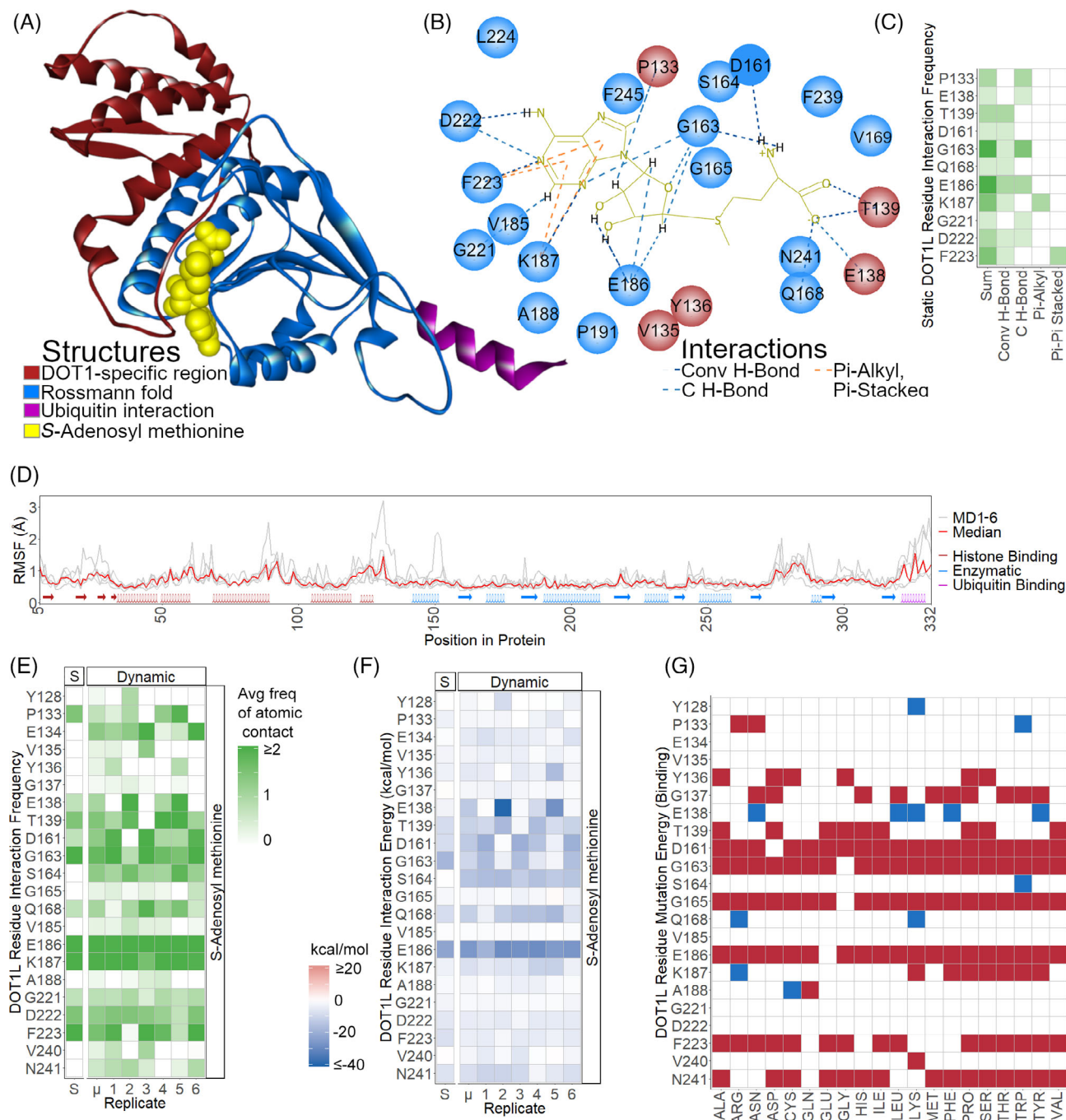
FIGURE 2 DOT1L residues D161, G163, G165, E186, F223, and N241 facilitate important interactions with SAM and are intolerant to mutation. (A) The DOT1L enzymatic region can be divided into three regions, the first is specific to DOT1L (red), the second is a conserved histone methyltransferase domain from a series of Rossmann folds (blue), and a region that interacts with ubiquitin (purple; PDB 1NW3). The substrate SAM (yellow) fits in a pocket formed by the Rossmann folds. (B) Two-dimensional projection showing DOT1L residues interacting with SAM colored as in (A). (C) The frequency (increasing green) and type of noncovalent interactions (hydrogen bond, electrostatic, and hydrophobic) in the static structure. (D) RMSF shows larger fluctuations in DOT1L-specific histone interacting region and less mobility in enzymatic region. Individual replicates are shown in gray, group median is shown in red. Arrows highlight residues with intermolecular interactions. (E) Average frequency of noncovalent atomic contact (increasing green) in dynamics illustrates E134 and S164 generate dynamic interactions absent in the static state. (F) Residue interaction energy shows T139, D161, G163, and E186 with the most negative energy (van der Waals and electrostatic) contributions in the static and dynamic conditions, while E138, S164, and Q168 are lowest in the dynamics only. Blue represents negative, white zero and red positive energy values. (G) Mutational energy (binding) between a DOT1L residue and SAM, thresholded as destabilizing (>0.5 kcal/mol, red), or stabilizing (<0.5 kcal/mol, blue). Interactions with D161, G163, E186, F223, and N241 are destabilized by the majority of possible variants

TABLE 1 Noncovalent interactions in each structure

Structure	Distance (Å)	Bond type	From	To
1NW3	1.767	Conventional hydrogen bond	A:THR139:HN	A:SAM500:OCT1
1NW3	1.858	Conventional hydrogen bond	A:THR139:HG1	A:SAM500:OCT2
1NW3	2.075	Conventional hydrogen bond	A:GLN168:HE21	A:SAM500:OCT1
1NW3	2.594	Conventional hydrogen bond	A:LYS187:HN	A:SAM500:N3
1NW3	2.040	Conventional hydrogen bond	A:PHE223:HN	A:SAM500:N1
1NW3	2.225	Conventional hydrogen bond	A:SAM500:H2	A:ASP161:OD1
1NW3	1.781	Conventional hydrogen bond	A:SAM500:H3	A:GLY163:O
1NW3	1.845	Conventional hydrogen bond	A:SAM500:HO3'	A:GLU186:OE1
1NW3	1.796	Conventional hydrogen bond	A:SAM500:HO2'	A:GLU186:OE2
1NW3	2.107	Conventional hydrogen bond	A:SAM500:H61	A:ASP222:OD1
1NW3	2.203	Carbon hydrogen bond	A:GLU138:HA	A:SAM500:OCT1
1NW3	2.585	Carbon hydrogen bond	A:GLY163:HA1	A:SAM500:O4'
1NW3	2.744	Carbon hydrogen bond	A:GLY163:HA2	A:SAM500:N3
1NW3	2.448	Carbon hydrogen bond	A:ASP222:HA	A:SAM500:N1
1NW3	2.488	Carbon hydrogen bond	A:SAM500:H4'	A:GLY163:O
1NW3	2.840	Carbon hydrogen bond	A:SAM500:H4'	A:GLU186:OE1
1NW3	2.704	Carbon hydrogen bond	A:SAM500:H2'	A:PRO133:O
1NW3	2.495	Carbon hydrogen bond	A:SAM500:H1'	A:GLU186:OE1
1NW3	2.193	Carbon hydrogen bond	A:SAM500:H8	A:PRO133:O
1NW3	2.507	Carbon hydrogen bond	A:SAM500:H4	A:GLY221:O
1NW3	3.782	Pi–Pi stacked	A:PHE223	A:SAM500
1NW3	4.499	Pi–Pi stacked	A:SAM500	A:PHE223
1NW3	4.838	Pi–Alkyl	A:SAM500	A:LYS187
1NW3	3.966	Pi–Alkyl	A:SAM500	A:LYS187
6NQA	2.310	Salt bridge; attractive charge	F:LYS20:HZ3	K:ASP28:OD2
6NQA	2.873	Conventional hydrogen bond	F:ALA15:HT3	K:ASN116:O
6NQA	1.928	Conventional hydrogen bond	F:ARG19:HH12	E:NLE79:O
6NQA	2.237	Conventional hydrogen bond	F:ARG19:HH22	E:GLN76:O
6NQA	1.823	Conventional hydrogen bond	K:ASN116:HD21	F:ALA15:O
6NQA	2.822	Carbon hydrogen bond	F:ARG17:HD2	K:SER304:OG
6NQA	2.732	Carbon hydrogen bond	F:LYS20:HE2	K:ASP28:OD2
6NQA	2.381	Pi–cation; Pi–donor H bond	F:LYS16:HZ2	K:HIS31
6NQA	5.594	Pi–Pi stacked	F:HIS18	K:TYR136
6NQA	4.082	Alkyl	F:ALA15	K:PRO122
6NQA	5.080	Pi–Alkyl	K:HIS31	F:LYS16
6NQA	2.502	Conventional hydrogen bond	K:ARG319:HE	L:LEU71:O
6NQA	3.038	Carbon hydrogen bond	K:ARG319:HD2	L:LEU71:O
6NQA	5.115	Alkyl	K:ALA258	L:LEU73
6NQA	5.182	Alkyl	K:ILE290	L:LEU73
6NQA	5.027	Alkyl	K:LEU322	L:LEU71
6NQA	5.217	Alkyl	L:ARG74	K:LEU284
6NQA	4.788	Pi–Alkyl	K:PHE326	L:ILE36
6NQA	5.359	Pi–Alkyl	K:PHE326	L:PRO37
6JN2 Intradimer	2.705	Salt bridge; attractive charge	B:LYS508:HZ3	A:ASP739:OD2
6JN2 Intradimer	4.587	Attractive charge	B:LYS521:NZ	A:GLU758:OE2
6JN2 Intradimer	2.929	Conventional hydrogen bond	A:GLN751:HN	B:GLU518:OE2

TABLE 1 (Continued)

Structure	Distance (Å)	Bond type	From	To
6JN2 Intradimer	2.782	Conventional hydrogen bond	A:GLN751:HE22	B:GLU518:OE1
6JN2 Intradimer	2.108	Conventional hydrogen bond	A:ASN754:HD22	B:ASN522:OD1
6JN2 Intradimer	1.988	Conventional hydrogen bond	B:LYS503:HZ1	A:GLN734:O
6JN2 Intradimer	2.306	Conventional hydrogen bond	B:ASN522:HD21	A:GLU753:OE2
6JN2 Intradimer	2.768	Carbon hydrogen bond	A:LEU747:HA	B:GLU518:OE2
6JN2 Intradimer	2.420	Carbon hydrogen bond	A:GLN772:HA	B:GLN539:OE1
6JN2 Intradimer	2.565	Carbon hydrogen bond	B:LYS508:HE1	A:ASP739:OD2
6JN2 Intradimer	2.192	Carbon hydrogen bond	B:LYS521:HE1	A:ASN754:OD1
6JN2 Intradimer	2.239	Carbon hydrogen bond	B:ASN522:HA	A:ASN754:OD1
6JN2 Intradimer	4.811	Alkyl	A:MET743	B:LEU511
6JN2 Intradimer	4.825	Alkyl	A:LEU747	B:LEU514
6JN2 Intradimer	4.519	Alkyl	A:LEU747	B:LEU515
6JN2 Intradimer	4.946	Alkyl	A:LEU757	B:LEU525
6JN2 Intradimer	5.252	Alkyl	A:LEU764	B:LEU532
6JN2 Intradimer	5.028	Alkyl	A:LEU764	B:LEU533
6JN2 Intradimer	5.330	Alkyl	A:LEU778	B:LEU546
6JN2 Intradimer	5.366	Alkyl	B:LYS519	A:LEU750
6JN2 Intradimer	4.439	Alkyl	B:ALA528	A:ILE761
6JN2 Intradimer	4.791	Alkyl	B:ALA529	A:LEU757
6JN2 Intradimer	4.750	Alkyl	B:ALA529	A:ILE761
6JN2 Intradimer	5.045	Alkyl	B:LYS540	A:LEU771
6JN2 Intradimer	5.166	Pi–Alkyl	B:TYR507	A:MET743
6JN2 Interdimer	5.441	Attractive charge	A:ARG770:NH1	E:PRO781:OCT1
6JN2 Interdimer	2.306	Carbon hydrogen bond	A:TRP724:HD1	G:LEU720:O
6JN2 Interdimer	2.774	Carbon hydrogen bond	A:PRO781:HA	C:GLU769:OE1
6JN2 Interdimer	2.774	Carbon hydrogen bond	C:PRO781:HA	A:GLU769:OE1
6JN2 Interdimer	2.305	Carbon hydrogen bond	G:TRP724:HD1	A:LEU720:O
6JN2 Interdimer	5.477	Alkyl	A:ILE716	G:LEU731
6JN2 Interdimer	5.439	Alkyl	A:LEU720	G:LEU731
6JN2 Interdimer	5.477	Alkyl	A:LEU731	G:ILE716
6JN2 Interdimer	5.440	Alkyl	A:LEU731	G:LEU720
6JN2 Interdimer	5.363	Alkyl	A:LEU774	E:LEU778
6JN2 Interdimer	5.363	Alkyl	A:LEU778	E:LEU774
6JN2 Interdimer	4.228	Alkyl	A:VAL780	C:LEU773
6JN2 Interdimer	5.238	Alkyl	B:LEU488	H:LEU488
6JN2 Interdimer	4.875	Alkyl	B:LYS492	H:LEU484
6JN2 Interdimer	4.277	Alkyl	B:LYS492	HLEU488
6JN2 Interdimer	4.228	Alkyl	C:VAL780	A:LEU773
6JN2 Interdimer	4.875	Alkyl	h:LYS492	B:LEU484
6JN2 Interdimer	4.276	Alkyl	h:LYS492	B:LEU488
6JN2 Interdimer	4.336	Pi–Alkyl	A:TRP724	G:LEU720
6JN2 Interdimer	5.360	Pi–Alkyl	A:TRP724	G:LEU720
6JN2 Interdimer	4.877	Pi–Alkyl	A:PHE730	h:LEU487
6JN2 Interdimer	5.390	Pi–Alkyl	B:TYR495	h:LEU484
6JN2 Interdimer	4.765	Pi–Alkyl	B:PHE498	G:LEU719
6JN2 Interdimer	5.365	Pi–Alkyl	B:PHE547	F:LYS540

(Continues)

TABLE 1 (Continued)

Structure	Distance (Å)	Bond type	From	To
6JN2 Interdimer	5.374	Pi–Alkyl	B:PHE547	F:ILE543
6JN2 Interdimer	5.365	Pi–Alkyl	F:PHE547	B:LYS540
6JN2 Interdimer	5.374	Pi–Alkyl	F:PHE547	B:ILE543
6JN2 Interdimer	4.335	Pi–Alkyl	G:TRP724	A:LEU720
6JN2 Interdimer	5.360	Pi–Alkyl	G:TRP724	A:LEU720
6JN2 Interdimer	4.877	Pi–Alkyl	G:PHE730	B:LEU487
6JN2 Interdimer	5.389	Pi–Alkyl	H:TYR495	B:LEU484
6JN2 Interdimer	4.765	Pi–Alkyl	H:PHE498	A:LEU719
2MV7	2.005	Conventional hydrogen bond	A:GLN524:HE22	B:VAL889:O
2MV7	2.678	Conventional hydrogen bond	A:ASN528:HD22	B:VAL889:O
2MV7	2.610	Conventional hydrogen bond	A:PHE543:HN	B:ILE883:O
2MV7	2.702	Conventional hydrogen bond	A:PHE545:HN	B:VAL881:O
2MV7	2.560	Conventional hydrogen bond	A:LEU547:HN	B:LEU879:O
2MV7	2.159	Conventional hydrogen bond	A:CYS548:HG	B:ASN877:O
2MV7	2.175	Conventional hydrogen bond	B:VAL881:HN	A:PHE545:O
2MV7	1.920	Conventional hydrogen bond	B:ILE883:HN	A:PHE543:O
2MV7	2.398	Conventional hydrogen bond	B:VAL889:HN	A:GLN524:OE1
2MV7	2.307	Carbon hydrogen bond	A:HIS511:HE1	B:PRO880:O
2MV7	2.317	Carbon hydrogen bond	A:ASP544:HA	B:VAL881:O
2MV7	2.334	Carbon hydrogen bond	A:ASP546:HA	B:LEU879:O
2MV7	2.550	Carbon hydrogen bond	B:PRO880:HA	A:PHE545:O
2MV7	2.472	Carbon hydrogen bond	B:PRO880:HA	A:ASP546:OD1
2MV7	2.357	Carbon hydrogen bond	B:SER882:HA	A:PHE543:O
2MV7	2.551	Carbon hydrogen bond	B:VAL888:HA	A:GLN524:OE1
2MV7	2.869	Pi–Sigma	B:VAL881:HB	A:PHE545
2MV7	4.594	Alkyl	A:LEU523	B:ILE883
2MV7	4.689	Alkyl	A:VAL527	B:VAL888
2MV7	5.270	Alkyl	A:VAL527	B:LEU890
2MV7	5.299	Alkyl	A:ILE538	B:LEU890
2MV7	5.043	Alkyl	A:LEU547	B:LEU879
2MV7	5.169	Alkyl	B:VAL881	A:LEU514
2MV7	4.099	Pi–Alkyl	A:HIS511	B:LEU879
2MV7	5.037	Pi–Alkyl	A:HIS511	B:VAL881
2MV7	5.236	Pi–Alkyl	A:PHE543	B:LEU890

end of simulations. Similarly, the surface area decreased from 12 381 Å^2 to 10 036 ± 165 Å^2. The changes in DOT1L upon the comparison of the structure before the simulation to after dynamics simulations are depicted in Figure S2A,B. Interestingly, some conspicuous changes occurred at the *N*-terminal helical domains in a manner that resemble similar regions found in the yeast dot1 structure (Figure S2, white, PDB ID: 1U2Z[41]). These changes were characterized by remodeling of the loop that joins the DOT1L-specific region to the KMT domains, where the loop becomes helical (Figure S2, yellow). These transient helical regions appear to transmit downward forces that coordinate movements between these domains, thereby determining the dynamic spatial location of both the SAM and H4 tail binding. This mechanism is important since the spatial location of the SAM pocket in

relationship to its substrate, H3K79, is required for H3K79 methylation.[30–32] These dynamic-induced changes that remodel regions of the human protein to mimic the yeast ortholog suggest that the molecular architecture for these enzymes, when considering possible dynamic variations, is likely more similar across species than previously inferred from static crystallography-derived models.

3.1.3 | Dynamics interactions between DOT1L and SAM

We assessed changes to the DOT1L–SAM interaction patterns observed after MD. Here, residues P133, E134, E138, T139, D161,

G163, S164, G165, Q168, V185, E186, K187, G221, D222, and F223 most consistently interacted with SAM across MD replicates (Figure 2E) and that T139, D161, G163, Q168, E186, D222, and F223 contributed the lowest interaction energy (van der Waals and electrostatic energy) in the conformations sampled during the dynamic trajectories (Figure 2F). Four residues, E134, S164, G165, and V185, interacted with SAM in dynamics but were not identified as interacting with SAM in the static structure, increasing the interacting residues by 36% over the static structure.

We subsequently performed scanning mutagenesis using all 20 natural amino acids for each position identified to interact with SAM in either the static or dynamic models (Figure 2G). Using this approach, we found that the majority of substitutions of E134, E138, S164, Q168, G221, and D222 were tolerated, while most substitutions at D161, G163, and E186 were destabilizing. Lastly, changes at P133 and A188 were either stabilizing or destabilizing depending on the amino acid replacement. Therefore, this method reveals that D161, G163, G165, E186, F223, and N241 facilitate important interactions with SAM and are intolerant to mutation.

3.2 | Dynamics analysis identifies increased DOT1L–H4 tail and DOT1L–ubiquitin interactions with DOT1L loop stabilizations

3.2.1 | Static structure and interactions between DOT1L, histone H4, and ubiquitin

Next, we studied a complex of DOT1L with the histone H4 tail and ubiquitin, derived from the structure of nucleosome-bound form of this enzyme (Figure 3A). Distance-dependent cutoffs showed that DOT1L residues D28, H30, H31, N116, P122, Y136, and S304 form a pocket that accommodates the flexible H4 tail, contacting amino acids A15, H16, and H20 (Figure 3B,C, Table 1). These residues stabilized the complex by forming four hydrogen bonds, a salt bridge, and four hydrophobic interactions. Interface surface analyses indicated that the H4 tails provide a contact area of 157 Å^2 with 90 Å^2 polar and 67 Å^2 non-polar surfaces. The DOT1L contact surface area for this binding encompassed 156 Å^2, with 96 Å^2 polar and 60 Å^2 non-polar. Hence, in comparison, the binding surface for histone is half the size of the surface used for SAM binding. Using similar approaches as above, we also studied the interaction between DOT1L and ubiquitin attached to H2BK120, which is necessary for DOT1L docking to the nucleosome and activation of its enzymatic activity. Analyses of the DOT1L–ubiquitin complex, before MD simulations, indicated that the binding between these molecules was stabilized by two hydrogen and six hydrophobic interactions (Figure 3C,D, Table 1). DOT1L residues A258, L284, I290, R319, L322, and F326 were involved in this intermolecular interaction (Figure 3C), forming a total contact area of 400 Å^2, with 196 Å^2 of polar and 203 Å^2 nonpolar surface contributions. Thus, the DOT1L–ubiquitin interaction interface had a larger surface than either SAM or H4 and contained a higher proportion of non-polar contacts. The complementary ubiquitin surface, which is formed by T7, I36, P37, L71, and L73, presented a contact area of 386 Å^2, with 176 Å^2 and 210 Å^2 polar and nonpolar contributions, respectively. In this regard, the DOT1L-binding surfaces for both SAM and ubiquitin were extensive, attesting to their key roles as substrate and cofactor, respectively.

3.2.2 | General dynamics

Adding histone and ubiquitin to our MD simulations did not significantly alter the overall pattern of DOT1L mobilities relative to the models with DOT1L and SAM alone (Figure 3d). The largest RMSF occurred at the N-terminal sheets. The structures also exhibited increased RMSF from the beginning of DOT1L helix 2 to the end of the loop between helix 3 and helix 4. This indicates that the molecular motions after histone binding are concordant with the motions of DOT1L without the histone, and therefore future work on the monomer is of value. Relative to DOT1L–SAM alone, there was lower RMSF from the end of helix 4 through the loop after helix 5, which corresponds to the region that interacts with histone H4. Finally, there was reduced RMSF at DOT1L helix 12 whose residues interact with ubiquitin. This is important as ubiquitin serves as a signal for DOT1L, which recognizes Ub-K120 H2B containing nucleosomes but not as an allosteric activator.[42] We found no changes for SAM binding in the DOT1L–ubiquitin–H4 tail complex. This is consistent with the fact that processing of SAM for H3K79 methylation does not involve changes in DOT1L shape but rather a significant rotation along the surface of the nucleosomes to contact H3K79.[30-32] Therefore, the dynamic results presented here are congruent with the function attributed to all members of the complex and show that the binding of DOT1L, SAM, and H4 stabilizes the complex resulting in less dynamic range of motion.

3.2.3 | Dynamics interactions between DOT1L, H4, and ubiquitin

The time-dependent behavior of the DOT1L–H4 interactions in MD showed that 22 DOT1L residues were interacting with H4, and D28, Y115, N116, E138, K300, S304, W305, and T306 have the largest quantity of interactions consistently across replicates (Figure 3E), whereas D28, E138, and K300 have with the most consistent energy lowering contributions (Figure 3F). Of these only D28, N116, and S304 were captured in the static structure analysis. Between DOT1L and ubiquitin, 16 DOT1L residues were contributing interactions, and L284, R319, E323, F326, and K330 had the largest number of interactions in every replicate. DOT1L residues E323 and K330 exhibited the largest stabilizing interaction energy contributions with ubiquitin. In summary, the dynamics analysis increased the number of DOT1L residues interacting with H4 3.6-fold and with ubiquitin 2.6-fold. The static model missed four of eight most frequency interacting residues with H4 and missed two of the three residues with the most energetically stabilizing interactions with H4. It also missed two of the five
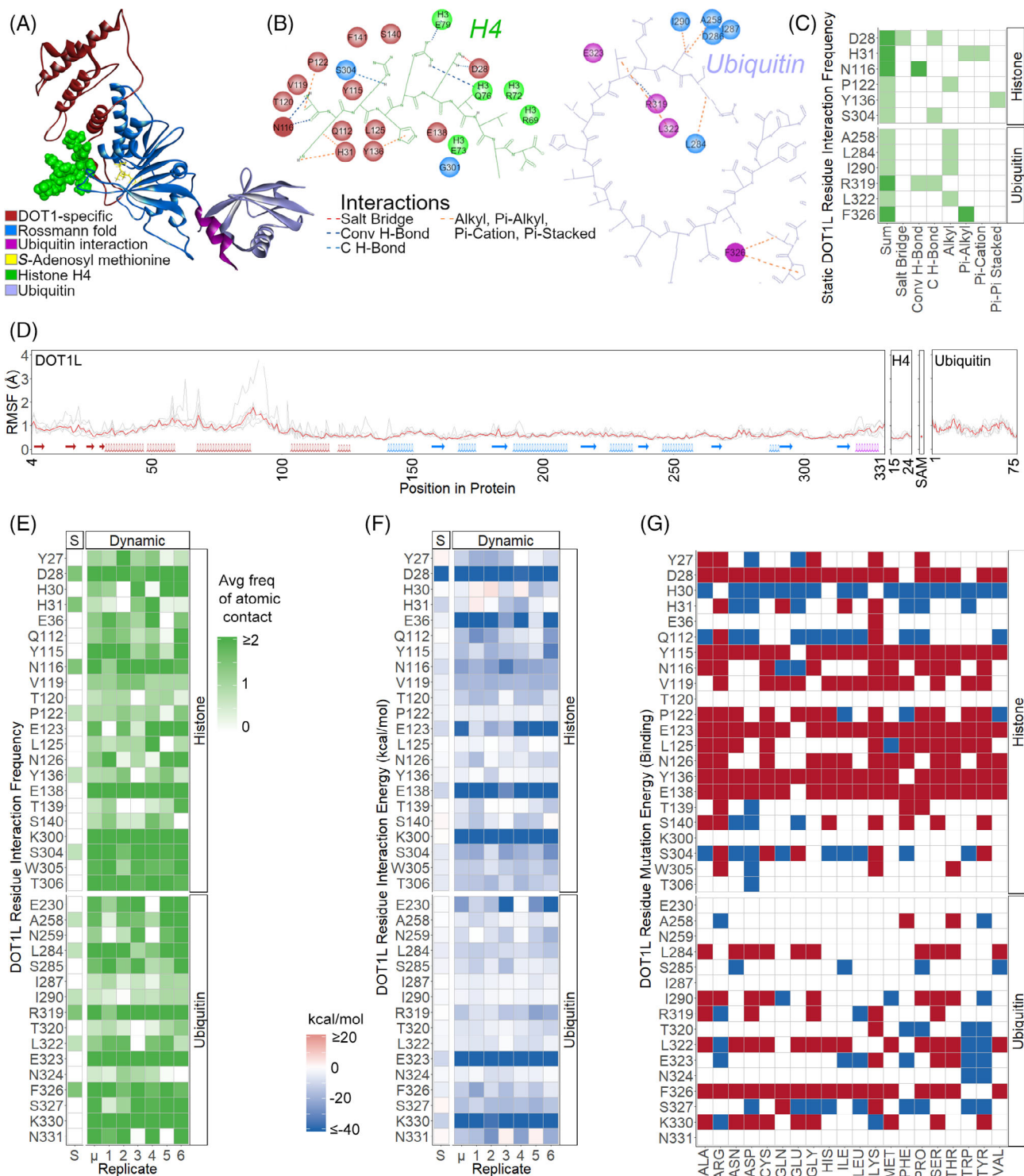
FIGURE 3 Interface and coordinated motions of the DOT1L enzymatic domain, histone H4 and ubiquitin. (A) DOT1 enzymatic region, colored as in Figure 2A, with histone H4 (green) and ubiquitin (violet) (from PDB 6NQA). (B) Two-dimensional projection of DOT1L residues interacting with histone H4 and ubiquitin. (C) Six DOT1L residues have non-covalent bonds with residues of histone H4 and an additional six DOT1L residues have non-covalent bonds with ubiquitin in the static structure. (D) RMSF analysis shows a similar pattern to Figure 2D, with high flexibility in the DOT1L-specific histone H4 interacting region and the ubiquitin binding region. Individual replicates are shown in gray, group median is shown in red. (E) Thirty-six DOT1L residues consistently contribute non-covalent interactions under dynamics. (F) Interaction energy shows D28 is important in the static structure, and with dynamics E138, K300, E323, and K330 in addition to D28 have the largest stabilizing energetic contributions. (G) Mutational energy calculation shows mutations that are destabilizing (red) or stabilizing (blue). Interactions with D28, Y115, E123, Y138, E138, L322, and F326 are destabilized by most mutations

most frequently interacting residues with ubiquitin and the two most energetically stabilizing interactions with ubiquitin. These residues represent the most important targets for disrupting the alignment of DOT1L on a nucleosome.

We also performed scanning mutagenesis, as described above for SAM, to calculate differences in the free energy of binding between the wild type and mutants interacting with ubiquitin and the H4 tail (Figure 3G). Scanning mutagenesis revealed most substitutions in H30 to be stabilizing for the DOT1L–histone H4 interaction, whereas substitutions in D28, Y115, E123, L125, N126, Y136, and E138 were destabilizing. Substitutions in L322 and F326 were also primarily destabilizing to the DOT1L–ubiquitin interaction. In conclusion, the interactions including residues in DOT1L helices 4 and 5 critical for the DOT1L–H4 interface are damaged by most variants.

3.3 | Dynamic analyses of DOT1L–AF10 reveals increased coil–coil interactions in heterodimers and displacement of N-terminal coils with anchoring C-terminal coils in heterotetramer complexes

Efforts to define the role of DOT1L in leukemogenesis have yielded the structures of DOT1L–AF10 and DOT1L–AF9. These complexes form in regions of DOT1L downstream of the N-terminal enzymatic region. DOT1L is disordered between the enzymatic region and AF10 and AF9 binding regions and is predicted to alternate between ordered and disordered regions from the AF-protein interacting regions to the C-terminus of the protein (Figure 1B). Through MLL-AF fusion proteins, DOT1L is recruited to de novo target regions of the genome thereby leading to neoplastic transformation. Notably, however, these interactions are not exclusive to leukemia, as AF proteins, including AF10 and AF9, also bind DOT1L in normal cells.[9,12] Thus, the activity of fusion proteins cannot be turned off by disassembly of the complex in tumors without also disrupting normal DOT1L function. Both DOT1L–AF structures will be analyzed in the following sections.

3.3.1 | Static structure and interactions between DOT1L and AF10

Figure 4A shows the composition of the heterodimer of DOT1L with AF10. As shown in Table 1, bonds established by the dimer were primarily hydrophobic (13) and hydrogen bonds (10), with minimal contribution of salt bridges (2) (Figure 4B,C). Using distance cutoffs, the dimer formed an "alkyl" zipper, which involves a set of aliphatic amino acids including several leucine residues. Eighteen DOT1L residues participated in AF10 interactions. Therefore, the interactions between DOT1L and AF10 form a stable interface of electrostatic and hydrophobic interactions. Interface surface analyses showed that DOT1L contributed a contact surface area of 486 Å^2, split into 192 Å^2 polar and 294 Å^2 nonpolar. A surface of similar area (479 Å^2) divided into 164 Å^2 polar and 315 Å^2 nonpolar was found for AF10. Thus, the

dimerization occurred in the form of a zipper, primarily by creating large nonpolar complementary surface contacts.

3.3.2 | General dynamics

MD simulations indicated that each member of the dimer can form more extensive bonding. At the N-terminus, the chains came together and moved against each other. The combined movement of the dimer resembled a flexible polymer rope that folds and extends around a kink located immediately upstream of the C-terminal leucine zipper. These patterns of motion were characterized by low RMSF at key force redirecting points of the secondary structure (Figure 4D), such as the region surrounding the kink at the C-terminus of the structure. Of note, areas outside of the kink had high RMSF with the highest values obtained at the edges of the molecules. This result is characteristic for dimers of this type, which are simulated in isolation from the main structure and/or other surrounding coiled-coiled partners (e.g., microfilament bundles) and therefore tend to bend in the middle. The RMSF plots demonstrated for both, DOT1L and AF10, that the ends of the structures had the highest fluctuations, while the loops connecting the two helices and the middle of the structures had the least fluctuations.

3.3.3 | Dynamic interactions between DOT1L and AF10

In dynamics, 28 of the 82 DOT1L residues in this fragment interacted with AF10 (Figure 4E), a 1.5-fold increase. The most stabilizing interaction energies were from residues K492, K503, K508, K519, K521, and R545 (Figure 4F). The leucine residues in this region were the least tolerant toward mutation (Figure 4G). Thus, we conclude that, like other polymeric structures of similar design, the behavior of this DOT1L–AF10 structure displays movements characteristic of isolated coiled-coiled and leucine zipper domains.

The dynamic behavior of the DOT1L–AF10 leucine zipper was found to produce a very stable scaffold, which, for the most part, kept its overall shape, as described in Figure 4. Next, we determined how the tetramer of DOT1L–AF10 heterodimers (tetramer) described by Song et al.[35] acted as a whole, considering each DOT1L–AF10 heterodimer as a unit (Figure 5A). In the static structure, residues from a single DOT1L–AF10 heterodimer interacted with residues from five of the other six chains in the tetramer (Figure 5B,C). DOT1L residues L484, L487, L488, K492, Y495, K540, I543, and F547 contributed hydrophobic interactions in forming the tetramer. Under dynamic conditions, additional noncovalent interactions occurred between DOT1L residues A483, K486, E489, Q494, Q548 and Q549 and residues of other heterodimers in the majority of sampled trajectories (Figure 5F), while the lowest interaction energies were from DOT1L residues K486, K540, and Q549 (Figure 5G). Scanning mutagenesis revealed that most mutations in L484, L487, L488, Y495, and F547 were
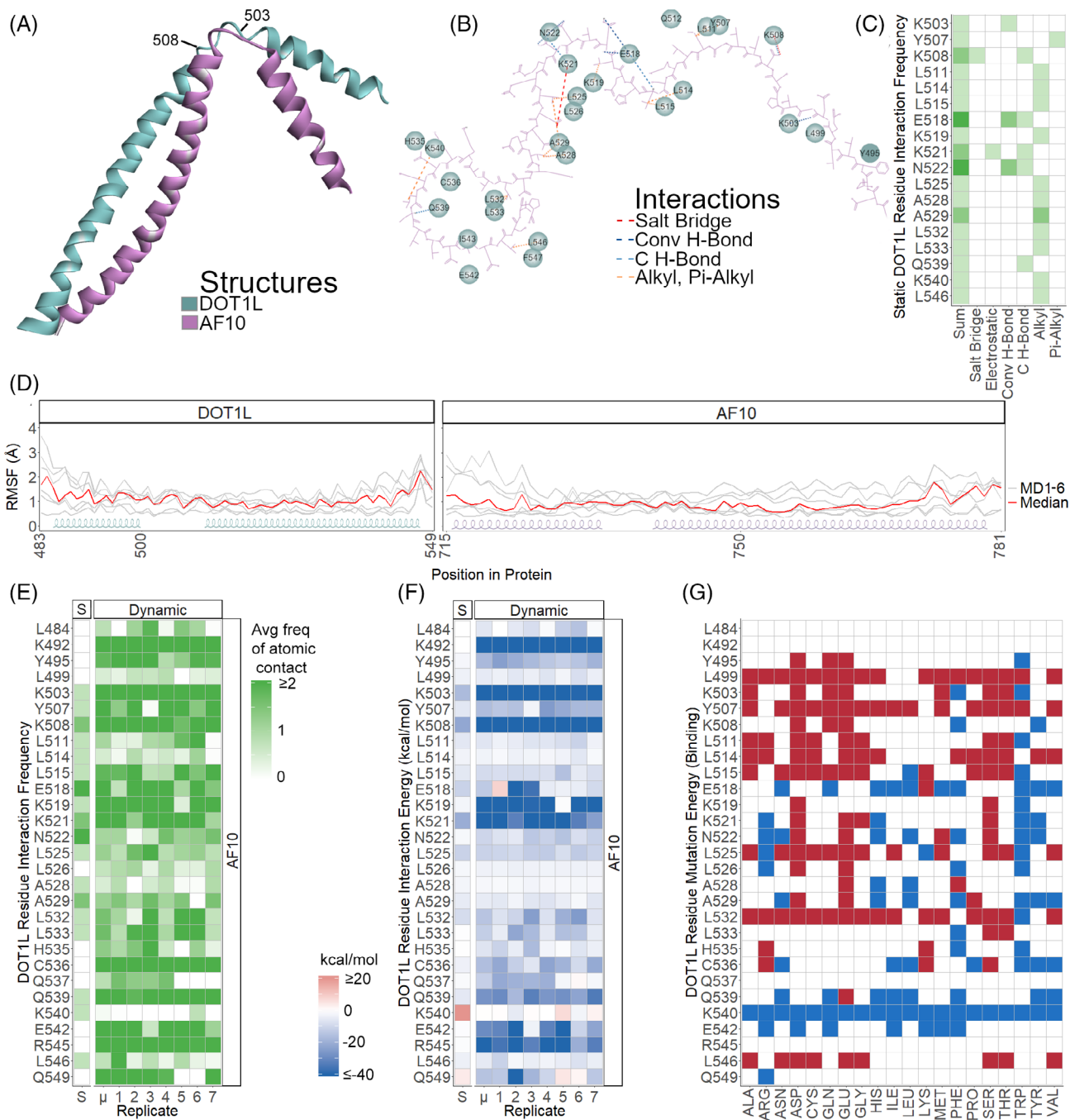
FIGURE 4 The stable scaffold of the DOT1L–AF10 heterodimer leucine zipper. (A) DOT1L (light blue) interaction with AF10 (pink) (PDB 6JN2). (B) Two-dimensional projection of DOT1L/AF10 interaction, DOT1L residues light blue circles, AF10 pink chain. (C) Eighteen DOT1L residues have noncovalent interactions in the static structure. (D) RMSF shows DOT1L and AF10 similar patterns of movement, with the both helix 1 more flexible relative to the hinge-loop, and with low flexibility in both helix 2 until close to the *C*-terminus. Individual replicates are shown in gray, group median is shown in red. (E) Twenty-five DOT1L residues participate in noncovalent interactions in more than half the replicates under dynamic conditions. (F) DOT1L residues K503, K508, and K521 have the lowest interaction energies (blue) in the static structure, while those plus K492, K519, and R545 are consistently the lowest (blue) during dynamics. Increases in interaction energy are shown in red. (G) Mutational energy (binding) between AF10 and specific DOT1L residues show interactions with L499, Y507, L514, L515, and L532 are destabilized (red) by most mutations, while K540 is stabilized (blue) by most mutations

destabilizing, whereas the majority in K492 were stabilizing (Figure 5H). Thus, dynamics revealed a 1.75-fold increase in DOT1L residues interacting with other heterodimers, especially the highly

stabilizing interactions at K486 and Q549 unique to dynamics, and that the highly interacting DOT1L residues Y495 and F547 are destabilized by most mutations.
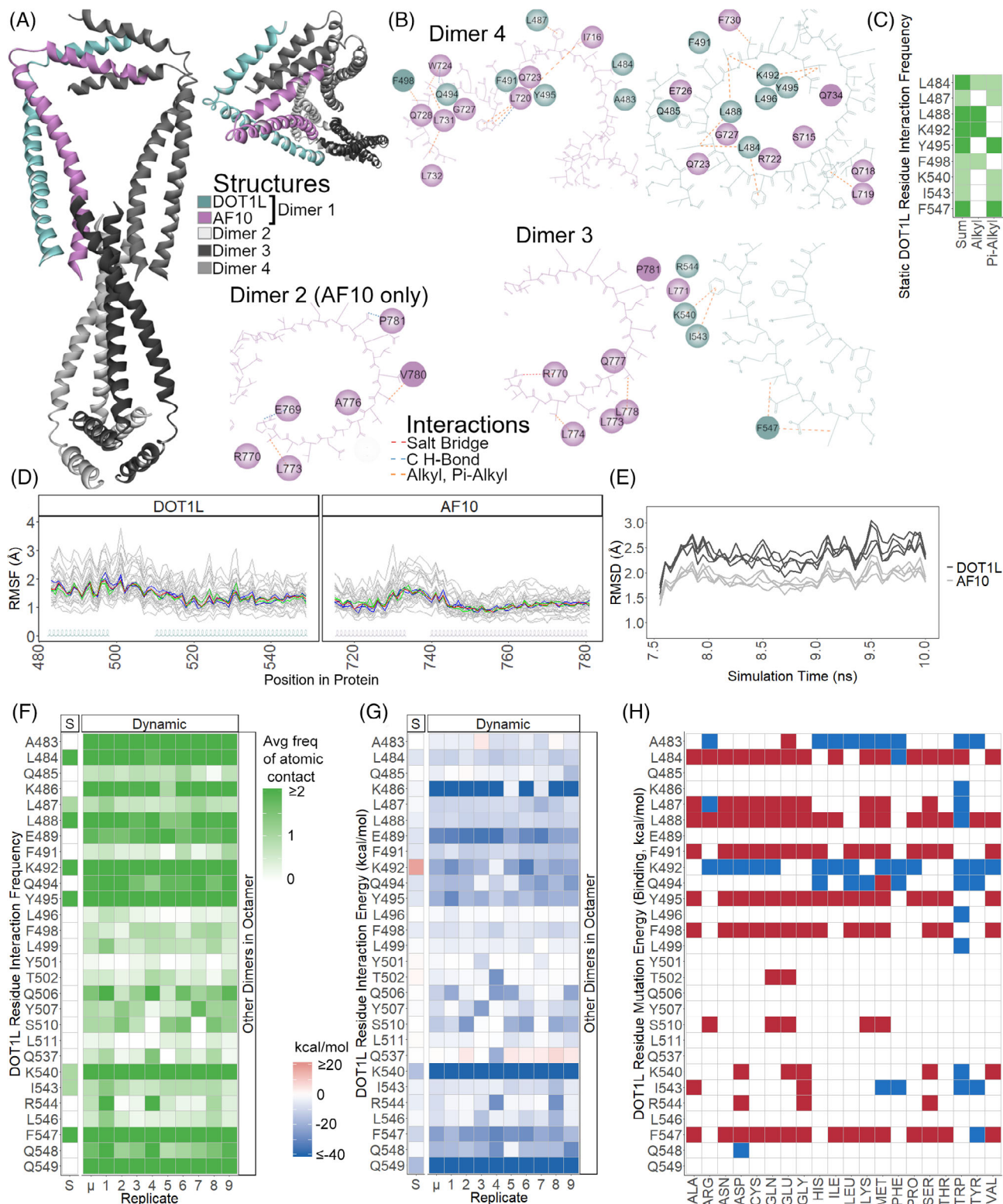
FIGURE 5 Legend on next page.

3.3.4 | General dynamics of the tetramer

The RMSF values showed similar and wide distributions across the entire structure, with peaks in the loops between the two helices of a heterodimer pair (Figure 5D). The hinge region of AF10 was the most

mobile, while the tail C-terminal regions of both DOT1L and AF10, now interacting with other tails, were the most stable. This pattern was present at multiple time scales, present in both 2 ns simulations, and continued through 10 ns after NAMD dynamics extensions (Figure S3A). Therefore, in the tetramer complex each DOT1L–AF10 heterodimer

pair was moving in concert, with displacement of the *N*-terminal helices relative to the *C*-terminal helices through the connecting hinge.

To further characterize these displacements, when the tetramer is viewed along the longest dimension as in Figure 5A, it resembles a bowtie with wide ends and a narrow middle. We detected movements that, in an alternating manner, decreased and increased the angles (94° to 141°) between the DOT1L helices within a single DOT1L–AF10 heterodimer pair (Figure S3B, i and ii). From the lateral view, the structure wiggled similar to a fish with the two ends of the molecule having a rhythmic lateral motion between 110 and 122 Å (Figure S3, iii). The diamond shape at the center of the tetramer, the *C*-terminal ends of each chain, moved in place and influenced the displacement of the DOT1 *N*-terminal ends, displayed in a CPK style (teal). These movements were reflected in a quantitative manner in 2D as RMSF values and in 3D after principal component analysis (PCA) (Figure 5D, Figure S4). Interestingly, the lateral displacement of the AF10 chains was smaller (Figure 5E) than those of DOT1L, which appeared to exert a slightly undulating movement. AF10 formed more of a column that, when intertwined with DOT1L, supports the lateral movements of both ends in a wider swinging motion. Each DOT1L *N*-terminal end in the complex connects through a disordered region to the DOT1L enzymatic domain, which contains a nucleosomal DNA binding domain. These molecular motions likely aid the DOT1L–AF10 heterotetramer to recognize nucleosomes marked by ubiquitination of H2B at K120. We also conjecture that these motions modulate DOT1L KMT function. Thus, dynamics studies of the tetrameric complex of DOT1L1-AF10 heterodimers illuminated displacement of the DOT1L *N*-terminal not observed in the heterodimer alone, pointing toward further studies that are needed to understand the role of this complex in leukemogenesis.

3.4 | Dynamic analyses of DOT1L–AF9 reveals an increased heterodimerization surface anchored by K878

### 3.4.1 | Static structure and interaction between DOT1L and AF9

Finally, we examined the structural, dynamic, and functional features of the DOT1L–AF9 complex (Figure 1B), which also plays a role in the genome-wide recruitment of this KMT and its mistargeted H3K79

methylation of nucleosomes during leukemogenesis due to MLL-AF9 fusions. One of the key structural features of AF9 is the domain involved in the recruitment and mistargeting of the DOT1L adopts an alpha-beta fold. In contrast to the leucine zipper of the DOT1L–AF10 complex, here, DOT1L forms a beta-strand that is parallel to the strand in AF9, with the AF9 helices providing a cavity for stabilizing the binding to the KMT (Figure 6A). DOT1L residues N877, L879, P880, V881, V888, V889, and L890 were involved in this interaction (Figure 6B,C). The corresponding amino acids in AF9 that formed the binding surface were represented by G524, N528, F543, F545, D546, L547, and C548. Hence, both members of the complex primarily bind through hydrophobic residues, although the interactions were almost equally distributed between hydrogen bonds and hydrophobic interactions. Note that these interactions are quite distinct from the L/K dominated coil-coil interactions between the helices of DOT1L and AF10. Interface surface analyses revealed that the DOT1L fragment provided a contact surface area of 377 Å^2 with 182 Å^2 polar and 195 Å^2 nonpolar contributions. Congruently, contact surface area for the AF9 domain was 381 Å^2, divided into 128 Å^2 polar and 253 Å^2 nonpolar. These measurements support the concept that most of the binding surface provided by AF9 is nonpolar, while the nonpolar *N*-terminal region of DOT1L is complementary and the *C*-terminal region remains outside of the major nonpolar binding region.

3.4.2 | General dynamics

MD revealed the DOT1L–AF9 heterodimer formed a stable complex and the coordinated movements between DOT1L and AF9. The only notable displacements occurred at the terminals. The 2D representation of these motions as RMSF values revealed the relative stability through the majority of the DOT1L–AF9 complex. The RMSF values also illustrated the *N*-terminal fragment of DOT1L and the *C*-terminal fragment of AF9 exploring the 3D space widely (Figure 6D). Notably, the *C*-terminal helix of AF9 (Figure 6D, AF9 helix 4) rotated back and forth in a manner that caused coordinated movement in the AF9 beta sheet 2, which is anchored to the beta sheet in DOT1L and thus transmitted the movement to the entire DOT1L fragment. Thus, these simulations provide insight into the coordinated motion through the DOT1L–AF9 complex, which may be useful to the design of complex disruptors for leukemias caused by an MLL-AF9 fusion.

FIGURE 5 Dynamics reveal DOT1L *N*-terminal helix displacements in tetramer of DOT1L–AF10 heterodimers. (A) Model of the DOT1L/AF10 heterodimers in tetramer structure (based on PDB 6JN2), with one tetramer pair colored (DOT1L light blue, AF10 pink), and the other pairs in different shades of gray. The view on the right is from the top-down. (B) Two-dimensional projections of interactions with the upper-left heterodimer pair (circles, DOT1L blue, AF10 pink) and the other chains of the tetramer are shown in the same orientation as in (A). Interactions with upper-right Dimer 4. Interactions with lower-left Dimer 2 (AF10 only), and interactions with lower-right Dimer 3. (C) There are eight DOT1L residues with noncovalent interactions in the static structure. (D) The RMSF plot of all replicates (gray) and median of each chain in tetramer (color) shows the flexibility seen near the *N*-terminal and *C*-terminal parts of the helices of the 6JN2 heterodimer is lost in the tetramer complex, which now have low RMSF, while the maximum RMSF in the structure is now observed in loops between the helices. (E) RMSD relative to initial conformation shows each DOT1L chain with greater displacement than AF10 chains. (F) Twenty DOT1L residues have consistent interaction with other heterodimers in the dynamic tetramer complex. (G) In the static structure, there is little interaction energy between DOT1L and members of other heterodimers, however, with dynamics consistently lowered interaction energies (blue) are identified with K486 and E489 in the smaller helix, and with K540 and Q549 at the *C*-terminus of the lower helix. (H) Mutational energy (binding) interaction energies between specific DOT1L residues in the upper-left heterodimer with residues in other heterodimers. Residues L484, L487, L488, F491, Y495, F498, and F547 are destabilized by most mutations (red), while K492 is stabilized (blue) by most mutations
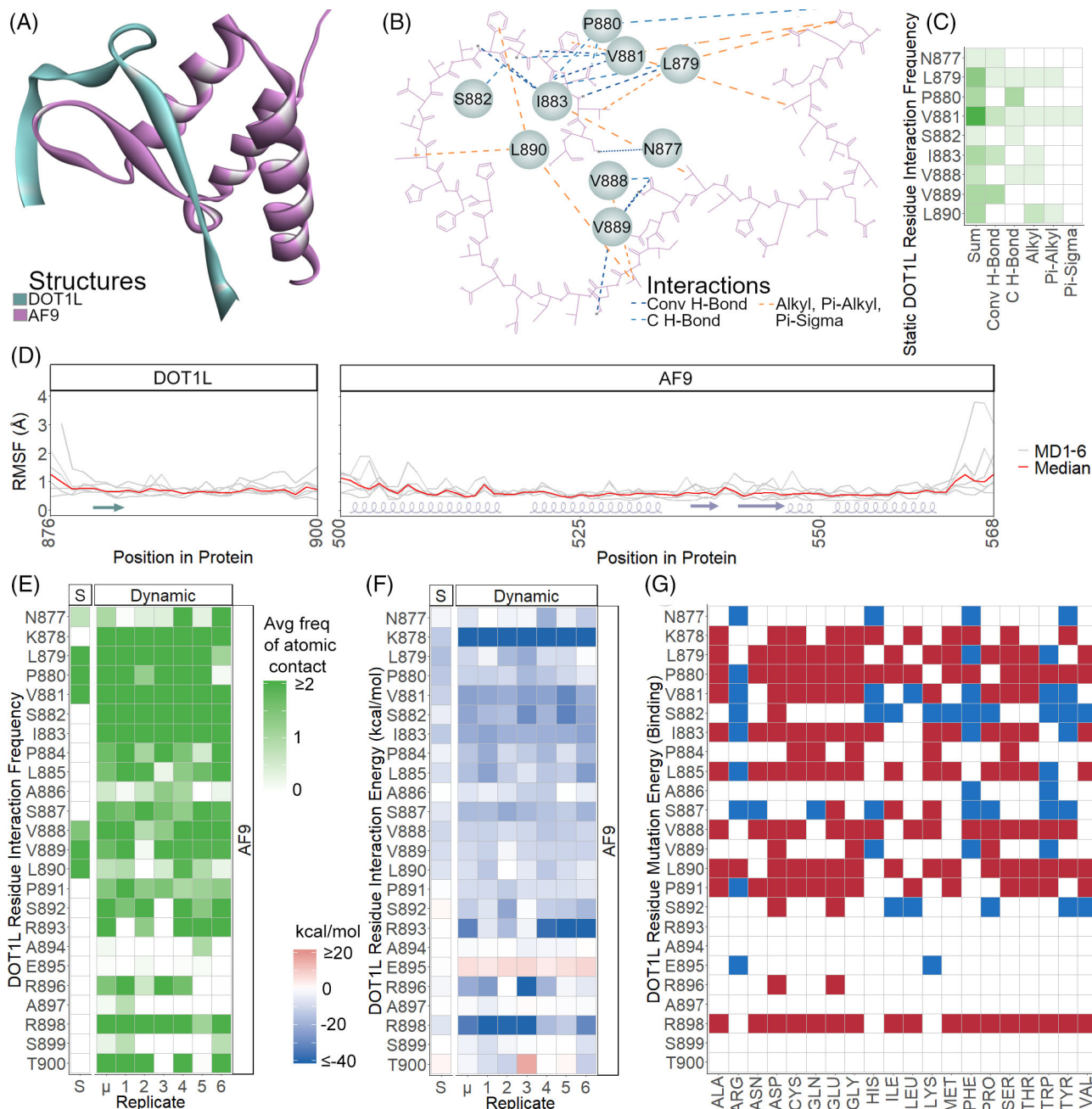
FIGURES 6 Interface and coordinated motion of the DOT1L–AF9 alpha-beta fold complex. (A) DOT1L (light blue) interaction with AF9 (pink) (PDB 2MV7). (B) Two-dimensional projection of DOT1L/AF9 interaction, DOT1L residues blue circles, AF9 pink. (C) Seven DOT1L residues make hydrogen bonds or hydrophobic interactions with AF9 in the static structure. (D) RMSF plot for DOT1L and AF9 shows mobile *N*-terminal end of DOT1L and *C*-terminal end of AF9, with the majority of the molecule stable. Individual replicates are shown in gray, group median is shown in red. (E) Dynamics reveal 18 noncovalent interactions with consistently high average frequencies, with the L890 interaction not consistent. (F) There is quantitatively small interaction energy between any of the DOT1L residues with AF9 in the static structure, while with dynamics residues K878, R896, and R898 have the lowest interaction energies (stabilizing, blue). (G) Mutational energy (binding) between AF9 and specific DOT1L residues show interactions with K878-P880, I883, L885, V888, L890, P891, and R898 destabilized (red) by most mutations

3.4.3 | Dynamics interaction between DOT1L and AF9

With dynamics, DOT1L residues K878, S882, I883, P884, L885, S887, P891, S892, R893, R896, R898, and T900 contributed high frequencies of noncovalent interactions with AF9 residues in more than 50%

of the MD simulations. These interactions were not present in the static structure (Figure 6E). The residue interaction energies were stabilizing across these residues, with DOT1L K878 more stabilizing than the others (Figure 6F). Similarly, even though aromatic residues take part in binding, only three π interactions were formed to stabilize the complex. Detailed information on the type of interactions formed, the

distance of their influence, and their chemical properties are listed in Table 1. Overall, dynamics revealed 2.7-fold increase in DOT1L residues interacting AF9 relative to the static structure.

Scanning mutagenesis revealed that most mutations in residues K878, P880, V888, L890, P891, and R898 destabilized the DOT1L–AF9 interaction, while S882 mutations frequently stabilized the DOT1L–AF9 interaction. Mutations at V881, I883, L885, S887, V889, and S892 had variable responses (Figure 6G). In summary, dynamics analysis revealed highly coordinated movement between DOT1L and AF9 in this alpha-beta fold complex, and that DOT1L residues K878 and R898, which do not interact with AF9 in the static structure, contribute both high frequency and energy lowering interactions in dynamics at positions in the protein that are sensitive to destabilization by mutations.

4 | DISCUSSION

The current study extends our insights into the biochemistry of DOT1L, a histone MTase, known for catalyzing the H3K79 methylation mark and its mistargeting to aberrant regions of the genome by MLL-AF fusion proteins to participate in leukemogenesis. Currently, 42 DOT1L structures have been deposited, alone and in combination with substrates, cofactors, inhibitors, and interacting proteins. However, details on the time-dependent changes in conformations, as well as intramolecular and intermolecular interactions, have remained poorly understood. Therefore, we used molecular mechanics and dynamics simulations to fill this important knowledge gap, which is not only important for basic biochemistry and epigenomic mechanisms, but also for understanding the genomics underpinning the development of leukemia.

We initiated this study by analyzing the *N*-terminal domain of DOT1L and its binding to the SAM co-substrate. For this purpose, we used two distinct structures solved by different laboratories with identical results (1NW3[28] and 3QOW[29]), significantly increasing the reliability of our results. Our findings indicate that, when considered in isolation, there is a time-dependent interaction of this region of the KMT with SAM, which cannot be completely defined from static structures. In fact, while in static conditions, one cannot confer differential weight to the contribution of residues to the binding surface at the lowest energy conformation. On the other hand, after MD simulations, we identified residues that are in contact most often with SAM relative to other residues. This result is important as several previous studies demonstrated in different contexts that MD simulations either refined the binding site or improved predictions of binding properties to small molecule fragments, which ultimately led to the development of specific inhibitors.[43–46] Our dynamics analysis determined which regions of protein complexes appear most stable and those most flexible, providing knowledge for future development of small molecules inhibitors.

Subsequently, we also applied molecular mechanics and dynamic simulations to gain insight into the interaction of DOT1L with both ubiquitin and the H4 tail. These interactions are critical for the full

function of DOT1L in vivo.[31,42,47–49] We defined the molecular surfaces of interactions, demonstrating that DOT1L makes significant nonpolar contact with ubiquitin. Conversely, the DOT1L–H4 tail surface interaction contacts are relatively small. These data support the fact that contact with ubiquitin is necessary for full DOT1L KMT activity, whereas the H4 tail is transiently used when the enzyme docks to nucleosomes. Thus, the small interaction surface between DOT1L and H4 is likely required to rapidly associate and dissociate from the nucleosome as needed. On the other hand, the more extensive nonpolar surface that supports the DOT1L–ubiquitin complex is likely responsible for maintaining the close association between the enzyme and this cofactor/activator. Thus, the number of residues and extent of the interaction surface contacts likely reflect the time and strength of association among all these key components for DOT1L function.

We also report results of additional novelty regarding the structural and dynamic analyses of the complexes formed by DOT1L with AF proteins. First, we examined the nature of the interaction and dynamics of the DOT1L–AF10 dimer, which is likely an initial step toward assembly of larger complexes. Subsequently, we also studied the DOT1L–AF10 tetramer of heterodimers. Finally, we investigated the complex formed by the DOT1L recruitment domain of AF9. These interactions are responsible for the mistargeting of DOT1L by the MLL-AF10 and MLL-AF9 oncogenic fusion proteins found in leukemias.[12,22,50,51] The study on the heterodimers yields useful information regarding the interactions that stabilize the coiled-coiled or leucine-zipper interactions. However, this dimeric species may not be highly represented *in vivo*, since most epigenomic regulators reside in larger multi-molecular complexes.[52] Therefore, we turned our attention to the analyses of the heterotetramer. The structure and dynamics of this complex reveal that DOT1L works with AF10 to generate a bowtie structure which moves in a fish-like fashion, transmitting kinetic forces to the DOT1L enzymatic domains that are located at the ends of the structure. Interestingly, 150 residues from the *N*-terminal where the solved structure of the heterotetramer terminates, DOT1L contains the site that binds nucleosomal DNA. Therefore, the movements caused by the natural dynamics of the complex, as described here, likely position the DOT1L enzymatic region in the correct proximity to nucleosomes, where DOT1L binds ubiquitin and transfers the methyl group from SAM to H3K79. Regarding the recruitment of DOT1L by AF9, we find both are highly dynamic components of the dimer. However, we show the intrinsic dynamics of the AF9 domain are responsible for the mechanic movement given to the intrinsically disordered region of DOT1L. Analyses of the contact interface between the domains of both proteins demonstrate that AF9 offers the larger contact surface, with a sizable number of interactions, while DOT1L contributes a relatively smaller surface. Thus, the description of this dynamic interaction indicates that targeting the AF9 domain will likely be feasible and effective due to its major role in establishing and maintaining the binding and dynamics of the complex.

In summary, these investigations fill a gap in the existing knowledge on the dynamics of DOT1L and its complexes. Our main finding

is that MDs extend prior information provided by structure-based experiments. Taken together, our results illuminate the interactions that stabilize the complexes that DOT1L forms with SAM, histone H4, ubiquitin, and AF proteins, which due to their intrinsic dynamics may impact the localization and function of this enzyme. We are optimistic that this data will further inform several scientific fields, including biochemistry, epigenomics, and leukemia research. In addition, the data presented here can be used to refine the design of novel drugs targeting DOT1L directly or its complexes. Therefore, combined, our study bears both basic mechanistic and biomedical relevance.

ACKNOWLEDGMENTS

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/prot.26219.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Brian F. Volkman https://orcid.org/0000-0002-6681-5179
Brian C. Smith https://orcid.org/0000-0001-6330-2768
Raul Urrutia https://orcid.org/0000-0002-1640-6780

REFERENCES

1. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004;429(6990):457-463.
2. Lund AH, van Lohuizen M. Epigenetics and cancer. *Genes Dev*. 2004;18(19):2315-2335.
3. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis*. 2010;31(1):27-36.
4. Boukas L, Havrilla JM, Hickey PF, Quinlan AR, Bjornsson HT, Hansen KD. Coexpression patterns define epigenetic regulators associated with neurological dysfunction. *Genome Res*. 2019;29(4):532-542.
5. Mazor T, Pankov A, Johnson Brett E, et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell*. 2015;28(3):307-317.
6. Dobrovic A, Kristensen LS. DNA methylation, epimutations and cancer predisposition. *Int J Biochem Cell Biol*. 2009;41(1):34-39.
7. Lawlor ER, Thiele CJ. Epigenetic changes in pediatric solid tumors: promising new targets. *Clin Cancer Res*. 2012;18(10):2768-2779.

8. Ng HH, Feng Q, Wang H, et al. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev*. 2002;16(12):1518-1527.
9. Steger DJ, Lefterova MI, Ying L, et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol*. 2008;28(8):2825-2839.
10. Wang Z, Zang C, Rosenfeld JA, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*. 2008;40(7):897-903.
11. Nguyen AT, Zhang Y. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev*. 2011;25(13):1345-1358.
12. Sarno F, Nebbioso A, Altucci L. DOT1L: a key target in normal chromatin remodelling and in mixed-lineage leukaemia treatment. *Epigenetics*. 2020;15(5):439-453.
13. Godfrey L, Crump NT, Thorne R, et al. DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat Commun*. 2019;10(1):2803.
14. Rao ST, Rossmann MG. Comparison of super-secondary structures in proteins. *J Mol Biol*. 1973;76(2):241-256.
15. Laurino P, Tóth-Petróczy Á, Meana-Pañeda R, Lin W, Truhlar DG, Tawfik DS. An ancient fingerprint indicates the common ancestry of Rossmann-fold enzymes utilizing different ribose-based cofactors. *PLoS Biol*. 2016;14(3):e1002396.
16. Tóth-Petróczy A, Tawfik DS. The robustness and innovability of protein folds. *Curr Opin Struct Biol*. 2014;26:131-138.
17. Petrossian TC, Clarke SG. Uncovering the human methyltransferasome. *Mol Cell Proteom*. 2011;10(1):M110.000976.
18. Falnes P, Jakobsson ME, Davydova E, Ho A, Małecki J. Protein lysine methylation by seven-β-strand methyltransferases. *Biochem J*. 2016;473(14):1995-2009.
19. Hamey JJ, Hart-Smith G, Erce MA, Wilkins MR. The activity of a yeast family 16 methyltransferase, Efm2, is affected by a conserved tryptophan and its N-terminal region. *FEBS Open Bio*. 2016;6(12):1320-1330.
20. Kaye FJ. Mutation-associated fusion cancer genes in solid tumors. *Mol Cancer Ther*. 2009;8(6):1399-1408.
21. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer*. 2007;7(4):233-245.
22. Pui C-H, Gaynon PS, Boyett JM, et al. Outcome of treatment in childhood acute lymphoblastic leukaemia with rearrangements of the 11q23 chromosomal region. *The Lancet*. 2002;359(9321):1909-1915.
23. Balgobind BV, Raimondi SC, Harbott J, et al. Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study. *Blood*. 2009;114(12):2489-2496.
24. Szczepański T, Harrison CJ, van Dongen JJM. Genetic aberrations in paediatric acute leukaemias and implications for management of patients. *Lancet Oncol*. 2010;11(9):880-889.
25. Meyer C, Hofmann J, Burmeister T, et al. The MLL recombinome of acute leukemias in 2013. *Leukemia*. 2013;27(11):2165-2176.
26. Feng Q, Wang H, Ng HH, et al. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol*. 2002;12(12):1052-1058.
27. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242.
28. Min J, Feng Q, Li Z, Zhang Y, Xu R-M. Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. *Cell*. 2003;112(5):711-723.
29. Richon VM, Johnston D, Sneeringer CJ, et al. Chemogenetic analysis of human protein methyltransferases. *Chem Biol Drug des*. 2011;78(2):199-210.
30. Worden EJ, Hoffmann NA, Hicks CW, Wolberger C. Mechanism of cross-talk between H2B Ubiquitination and H3 methylation by Dot1L. *Cell*. 2019;176(6):1490-1501.e1412.

31. Anderson CJ, Baird MR, Hsu A, et al. Structural basis for recognition of ubiquitylated nucleosome by Dot1L methyltransferase. *Cell Rep.* 2019;26(7):1681-1690.e1685.

32. Valencia-Sánchez MI, De Ioannes P, Wang M, et al. Structural basis of Dot1L stimulation by histone H2B lysine 120 ubiquitination. *Mol Cell.* 2019;74(5):1010-1019.e1016.

33. Jang S, Kang C, Yang H-S, et al. Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase. *Genes Dev.* 2019; 33(11–12):620-625.

34. Yao T, Jing W, Hu Z, et al. Structural basis of the crosstalk between histone H2B monoubiquitination and H3 lysine 79 methylation on nucleosome. *Cell Res.* 2019;29(4):330-333.

35. Song X, Yang L, Wang M, et al. A higher-order configuration of the heterodimeric DOT1L–AF10 coiled-coil domains potentiates their leukemogenenic activity. *Proc Natl Acad Sci.* 2019;116(40):19917-19923.

36. Kuntimaddi A, Achille Nicholas J, Thorpe J, et al. Degree of recruitment of DOT1L to MLL-AF9 defines level of H3K79 Di- and tri-methylation on target genes and transformation potential. *Cell Rep.* 2015;11(5):808-820.

37. Velez G, Lin M, Christensen T, Faubion WA, Lomberk G, Urrutia RA-O. Evidence supporting a critical contribution of intrinsically disordered regions to the biochemical behavior of full-length human HP1γ. (0948-5023 [Electronic]).

38. Momany FA, Rone R. Validation of the general purpose QUANTA ®3.2/CHARMm® force field. *J Comput Chem.* 1992; 13(7):888-900.

39. Phillips JC, Braun R Fau-Wang W, et al. Scalable molecular dynamics with NAMD. (0192-8651 [Print]).

40. *BIOVIA, Dassault Systèmes, Discovery Studio, version 19.1.0.18287, San Diego: Dassault Systèmes* [computer program]. 2018.

41. Sawada K, Yang Z, Horton JR, Collins RE, Zhang X, Cheng X. Structure of the conserved Core of the yeast Dot1p, a nucleosomal histone H3 lysine 79 Methyltransferase. *J Biol Chem.* 2004;279(41):43296-43306.

42. Ng HH, Xu RM, Zhang Y, Struhl K. Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. *J Biol Chem.* 2002;277(38):34655-34657.

43. Hazuda DJ, Anthony NJ, Gomez RP, et al. A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc Natl Acad Sci U S A.* 2004;101(31):11233-11238.

44. Grant BJ, Lukman S, Hocker HJ, et al. Novel allosteric sites on Ras for lead generation. *PLOS One.* 2011;6(10):e25711.

45. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a novel binding trench in HIV integrase. *J Med Chem.* 2004;47(8):1879-1881.

46. Durrant JD, Keränen H, Wilson BA, McCammon JA. Computational identification of uncharacterized Cruzain binding sites. *PLoS Negl Trop Dis.* 2010;4(5):e676.

47. Briggs SD, Xiao T, Sun ZW, et al. Gene silencing: trans-histone regulatory pathway in chromatin. *Nature.* 2002;418(6897):498.

48. Kim J, Hake SB, Roeder RG. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol Cell.* 2005;20(5):759-770.

49. McGinty RK, Kim J, Chatterjee C, Roeder RG, Muir TW. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature.* 2008;453(7196):812-816.

50. Pui CH, Chessells JM, Camitta B, et al. Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. *Leukemia.* 2003;17(4):700-706.

51. Meyer C, Burmeister T, Gröger D, et al. The MLL recombinome of acute leukemias in 2017. *Leukemia.* 2018;32(2):273-284.

52. Medvedeva YA, Lennartsson A, Ehsani R, et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database.* 2015;2015:bav067.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.