RESEARCH ARTICLE

Structural studies of intrinsically disordered MLL-fusion protein AF9 in complex with peptidomimetic inhibitors

Yuting Yang[1] | Ejaz Ahmad[1] | Vidhya Premkumar[1] | Alicen Liu[1] |
S. M. Ashikur Rahman[1] | Zaneta Nikolovska-Coleska[1,2]

[1]Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan, USA

[2]Rogel Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan, USA

Correspondence
Zaneta Nikolovska-Coleska, Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan, USA.
Email: zanetan@med.umich.edu

Review Editor: Zengyi Chang

Abstract

AF9 (MLLT3) and its paralog ENL(MLLT1) are members of the YEATS family of proteins with important role in transcriptional and epigenetic regulatory complexes. These proteins are two common MLL fusion partners in *MLL*-rearranged leukemias. The oncofusion proteins MLL-AF9/ENL recruit multiple binding partners, including the histone methyltransferase DOT1L, leading to aberrant transcriptional activation and enhancing the expression of a characteristic set of genes that drive leukemogenesis. The interaction between AF9 and DOT1L is mediated by an intrinsically disordered C-terminal ANC1 homology domain (AHD) in AF9, which undergoes folding upon binding of DOT1L and other partner proteins. We have recently reported peptidomimetics that disrupt the recruitment of DOT1L by AF9 and ENL, providing a proof-of-concept for targeting AHD and assessing its druggability. Intrinsically disordered proteins, such as AF9 AHD, are difficult to study and characterize experimentally on a structural level. In this study, we present a successful protein engineering strategy to facilitate structural investigation of the intrinsically disordered AF9 AHD domain in complex with peptidomimetic inhibitors by using maltose binding protein (MBP) as a crystallization chaperone connected with linkers of varying flexibility and length. The strategic incorporation of disulfide bonds provided diffraction-quality crystals of the two disulfide-bridged MBP–AF9 AHD fusion proteins in complex with the peptidomimetics. These successfully determined first series of 2.1–2.6 Å crystal complex structures provide high-resolution insights into the interactions between AHD and its inhibitors, shedding light on the role of AHD in recruiting various binding partner proteins. We show that the overall complex structures closely resemble the reported NMR structure of AF9 AHD/DOT1L with notable difference in the conformation of the β-hairpin region, stabilized through conserved hydrogen bonds network. These first series of AF9 AHD/peptidomimetics complex

structures are providing insights of the protein–inhibitor interactions and will facilitate further development of novel inhibitors targeting the AF9/ENL AHD domain.

1 | INTRODUCTION

In the realm of eukaryotic organisms, an estimated 30–40% of the proteome comprises intrinsically disordered proteins (IDPs) or contains intrinsically disordered regions (IDRs) (Kulkarni et al., 2022). These IDPs and IDRs play pivotal roles across diverse biological processes, serving as central hubs in interaction networks. They orchestrate protein–protein interactions and coordinate a wide range of biological functions (Deiana et al., 2019; Kulkarni et al., 2022; Wright & Dyson, 2015). Consequently, interactions involving IDPs or IDRs are implicated in various diseases, making them promising therapeutic targets.

Proteins AF9 (MLLT3, UniProt P42568, 568 aa) and its paralog ENL (MLLT1, UniProt Q03111, 559 aa) are central players in multiple transcriptional and epigenetic regulatory complexes, including the Super Elongation Complex (SEC) and the Dot1-containing complex (DotCom) (He et al., 2011; Luo et al., 2012; Mohan et al., 2010). These two proteins share a sequence identity of 74% and feature two common domains, an N-terminal YEATS and a C-terminal ANC1 (nuclear anchorage protein) homologous domain (AHD). The YEATS domain and the AHD are connected by a disordered region of unknown function. The highly conserved 150-amino acid N-terminal YEATS domain is named after its presence in five proteins: YAF9, ENL, AF9, TAF14, and SAS5. In humans, four proteins harbor the YEATS domain: AF9, ENL, GAS41 (YEATS4, UniProt O95619, 227 aa), and YEATS2 (UniProt Q9ULM3, 1422 aa). The YEATS domain adopts an immunoglobulin-like fold, acting as an epigenetic "reader" for acetylated (Kac) or crotonylated (Kcr) lysine residues on histone tails, which serve as markers associated with active gene transcription (Li et al., 2014; Martin et al., 2021; Sabari et al., 2015). The 70-amino acid C-terminal AHD domain acts as a "recruiter", facilitating multiple protein–protein interactions. AF9 AHD has been demonstrated to be an intrinsically disordered protein and its interactions with binding partners are orchestrated through coupled binding and folding events (Leach et al., 2013). This is a well-documented phenomenon observed in various other IDPs, where disorder-to-order transitions occur upon binding to partner proteins (Dyson & Wright, 2002). It is notable that in *S. cerevisiae*, three proteins have been identified to possess the YEATS domain: TAF14 (UniProt P35189, 244 aa), SAS5 (UniProt Q99314, 248 aa), and YAF9 (UniProt P53930, 226 aa). Both TAF14 and SAS5 exhibit homology with AF9 and ENL, as they contain both the YEATS domain and the AHD.

The AHD plays a crucial role in establishing direct interactions with a diverse range of proteins involved in the regulation of gene transcription encompassing both activation and repression (Kabra & Bushweller, 2022). Importantly, it recruits AF4 (or AFF1) and DOT1L (disruptor of telomeric silencing 1 like), the only known H3K79 histone methyltransferase. These proteins are key components of the SEC and DotCom complexes, respectively, which support transcriptional activation. In addition, the AHD also interacts with CBX8 (chromobox homolog 8), a component of a Polycomb Group (PcG) multiprotein PRC1-like complex, and BCoR (Bcl-6 corepressor), both of which have been associated with transcriptional repression (Kabra & Bushweller, 2022). Co-expressing AF9 AHD with specific regions of binding partner proteins using a bicistronic vectors allowed the determination of the NMR complex structures of AF9 AHD in association with DOT1L (PDB ID: 2MV7), AF4 (PDB ID: 2LM0), BCOR (PDB ID: 6B7G), and CBX8 (PDB ID: 2N4Q), offering valuable insights into the molecular mechanisms underlying these complexes (Kabra & Bushweller, 2022; Kuntimaddi et al., 2015; Leach et al., 2013; Schmidt et al., 2020).

AF9 and ENL hold important therapeutic implications owing to their involvement in chromosomal translocations with the mixed-lineage leukemia *MLL1* (*KMT2A*) gene, which encodes the MLL1 protein, a histone-lysine (H3K4) methyltransferase. The N-terminal *MLL1* DNA-binding CXXC domain fuses with over 100 different proteins, creating oncogenic MLL-fusion proteins that profoundly impact epigenetic and transcriptional regulations (Marschalek, 2020; Meyer et al., 2018). MLL rearranged (MLLr) leukemia accounts for approximately 10% of all acute lymphoblastic and myeloid leukemias (ALL/AML). Among these, AF9 and ENL stand out

as two of the most common MLL-fusion partners (Meyer et al., 2018). Our recent reports have highlighted that the YEATS domain is differentially retained in MLL-AF9 and MLL-ENL fusion proteins among leukemia patients, whereas the AHD consistently remains present in these oncofusions (Hu et al., 2023). Furthermore, extensive studies have demonstrated that the AHDs are indispensable for the MLL-AF9 and MLL-ENL mediated transformations through their interactions with key components of the SEC and DotCom complexes, such as AF4 and DOT1L (Prange et al., 2017; Wang et al., 2011). These findings underscore the AHD as a promising therapeutic target and inhibiting its interactions offers a novel approach for the treatment of MLL-rearranged leukemia.

Our group conducted comprehensive biochemical, biophysical, and functional characterizations of the interactions between AF9/ENL and DOT1L. We mapped, for the first time, a 10-amino acid DOT1L peptide, LPISIPLSTV (865–874), as the binding site for the AHD, and demonstrated the potential therapeutic benefit of inhibiting the AF9 AHD and DOT1L interaction (Grigsby et al., 2021; Shen et al., 2013). Subsequent investigations confirmed our findings and identified two additional interaction sites within the DOT1L sequence: LELQISIVEL (638–647) and LPVSIPLASV (879–888) (Kuntimaddi et al., 2015). Building upon these findings and the reported NMR structure of AF9 AHD in complex with DOT1L (Kuntimaddi et al., 2015), our group developed a class of peptidomimetic inhibitors and established their structure–activity relationship (Du et al., 2018). This work served as the first proof-of-concept for targeting AHD and assessing its druggability.

The study of intrinsically disordered proteins presents significant challenges owing to their inherent lack of stable structures and propensity to aggregate without binding partners. Consequently, there is an ongoing need for the development of new experimental methodologies in this field. In this study, we introduce a protein engineering strategy designed to facilitates structural investigations of AF9 AHD domain. Our approach involves utilizing the maltose binding protein (MBP) as a crystallization chaperone and assessing the influence of the linker connecting MBP to AF9 AHD on the crystallization of the resulting fusion proteins. This has led to the successful determination of the first series of complex structures of the AF9 AHD and its peptidomimetic inhibitors using X-ray crystallography. The crystal structures closely resemble the reported NMR structures, with the peptidomimetic inhibitors binding to the AHD in a manner reminiscent of the DOT1L peptide. Our findings offer an innovative approach for investigating the structures of the IDPs in the complex form with binding partners and their interactions. Given the critical role of high-

resolution complex structures in structure-based drug development, this work also establishes a foundation for the future development of more potent inhibitors targeting the AF9 AHD.

2 | RESULTS

2.1 | Design of MBP and AF9 AHD fusion proteins to facilitate crystallization

Various strategies have been employed to aid in the crystallization and determination of challenging protein structures (Bergfors, 2022; Rhodes, 2006). One such approach involves fusing the protein of interest with crystallization chaperones, such as antibodies, lysozyme, glutathione S-transferase (GST), MBP, and other highly soluble and stable proteins. This fusing technique introduces new surface interactions and shape complementarity, thereby enhancing the likelihood of successful crystal formation (Bukowska & Grutter, 2013; Derewenda, 2010; Jin et al., 2017; Kotschy et al., 2016; Moon et al., 2010; Ullah et al., 2008).

Maltose binding protein (MBP), which consists of 370 amino acids, is frequently employed as a tag in *E. coli* expression systems. The primary purpose is to enhance the solubility of the target proteins during protein expression and purification process (Bregegere et al., 1994; di Guan et al., 1988). It functions as a "holdase" by preventing partially folded proteins from forming insoluble aggregates before undergoing either spontaneous or chaperone-assisted folding (Raran-Kurussi & Waugh, 2012). In the previously reported NMR solution structures, the AF9 AHD was either fused or co-expressed with its binding partners to create stable complexes, preventing aggregation of the AHD (Kabra & Bushweller, 2022; Kuntimaddi et al., 2015; Leach et al., 2013; Schmidt et al., 2020). In our study, we employed a different approach by fusing MBP to the N-terminus of the AF9 AHD. This strategy was designed to facilitate the crystallization of the AF9 AHD in complex with peptidomimetics that we had developed and recently reported (Du et al., 2018). Given the substantial size difference between MBP and the AHD, we anticipated that MBP molecules would play a dominant role in crystal lattice formation. Consequently, the primary interactions driving crystal contacts were expected to occur between MBP molecules and between MBP and the AF9 AHD (Smyth et al., 2003). To further enhance the diversity of crystal contacts and minimize conformational heterogeneity while maintaining an unobstructed inhibitor binding site on the AHD, we introduced linkers with varying lengths and flexibility. Our expectation was

TABLE 1 Binding affinity of different MBP–AF9 AHD fusion proteins to DOT1L and AF4 derived fluorescent labeled peptides.

Fusion protein	Abbreviation	Linker	K_d [nM] DOT1L (879–888 aa)	K_d [nM] AF4 (761–770 aa)
MBP-GS-AF9 AHD	**GS**	GS	17.91 ± 3.49	3.7 ± 0.94
MBP-AAA-AF9 AHD	**AAA**	AAA	8.79 ± 0.41	2.01 ± 0.22
MBP-AAAA-AF9 AHD	**AAAA**	AAAA	15.56 ± 1.93	3.66 ± 0.31
MBP-RITK-AF9 AHD	**RITK**	RITK	7.86 ± 0.76	1.31 ± 0.05
MBP-AAA-AF9 AHD Q336C/A502C	**3AQA**	AAA	20.93 ± 2.59	3.49 ± 0.18
MBP-AAAA-AF9 AHD Q336C/K501C	**4AQK**	AAAA	10.65 ± 2.84	1.26 ± 0.11

that these linkers would adopt a helical conformation, leading to the formation of a fused inter-domain helix that connects the C-terminal helix of MBP with the N-terminal helix of the AHD.

In total, six linkers were designed and tested, as detailed in Table 1. The choice of using the GS as linker was inspired by the crystal structure of MBP-fused Mcl-1 in complex with a highly selective and potent inhibitor, S63845 (PDB ID: 5LOF) (Kotschy et al., 2016). In this structure, the GS residues were observed to form an approximate helical connection, bridging MBP and Mcl-1. In addition linkers composed of glycine and serine residues are known for their flexibility. To evaluate the effect of the linker length, **AAA** and **AAAA** constructs were created. Considering that an α-helical turn spans 3.6 amino acids, these alanine-based linkers would cause the AF9 AHD to undergo a rotation of approximately 100° and 200°, respectively, relative to the AHD in the **GS** construct. To limit linker flexibility, construct **RITK** was created. Residues RITK are the four C-terminal residues of the wild-type MBP, which are typically absent in standard commercial MBP vectors. They were reintroduced as a linker between MBP and the AHD, with the expectation that they would impose greater rigidity compared with linker AAAA. Using the structures of MBP (PDB ID: 5LOF) and AF9 AHD (PDB ID: 2MV7), we generated a model for this construct, with residues RITK modeled as an α-helix (model not shown). It was observed that these residues established both hydrophobic interactions and hydrogen bonding with adjacent residues in the MBP. This suggests that RITK may reduce linker mobility compared to the AAAA linker.

To ensure the adoption of an α-helical conformation at the linker and to further enhance the rigidity of the fusion proteins, we introduced intramolecular disulfide bonds by incorporating cysteine mutations in both the **AAA** and **AAAA** constructs. To determine the sites for introducing the disulfide bonds, we modeled these constructs using the same approach as we created the **RITK** model, assuming helical conformations for the alanine linkers (Figure S1). Our analysis revealed that introducing a cysteine mutation at Q336 within MBP could

facilitate the formation of disulfide bonds with the beginning of the N-terminal helix of the AF9 AHD. Specifically, Q336C mutation is expected to form disulfide bond with A502 (A502C) in the **AAA** construct and K501 (K501C) in the **AAAA** construct, giving rise to two different constructs, **3AQA** and **4AQK** (Table 1). Importantly, the mutated residues in the AHD do not participate in binding with the DOT1L peptide; therefore, the disulfide bonds should not interfere with peptidomimetic binding.

All six fusion proteins of MBP and AF9 AHD were successfully expressed and purified to satisfactory qualities through a two-step procedure, which involved an initial Ni-NTA affinity column followed by size-exclusion chromatography (Figure S2). To determine their binding affinities to different protein partners DOT1L and AF4, fluorescence polarization assay was conducted using fluorescent-labeled peptides derived from DOT1L (879–888 aa) and AF4 (761–770 aa) proteins (Table S1). All six MBP and AF9 AHD fusion proteins exhibited comparable binding affinities, with K_d values ranging from 7.86 ± 0.76 to 20.93 ± 2.59 nM for the DOT1L-derived peptide and from 1.26 ± 0.11 to 3.7 ± 0.94 nM for the AF4-derived peptide (Table 1, Figures S3 and S4). These results align with prior reports indicating that AF9 AHD exhibits higher affinity for the AF4 peptide than for the DOT1L peptide (Kabra & Bushweller, 2022; Shen et al., 2013). These findings also demonstrate that the AHD peptide binding site remains accessible, and modifications in the linker region do not impact peptide binding. In addition, the two disulfide-bridged fusion proteins, **3AQA** and **4AQK**, exhibited K_d values comparable to those of the other fusion proteins lacking disulfide bonds. This suggests that the formation of disulfide bonds does not impede the coupled folding and binding process of the intrinsically disordered AF9 AHD.

2.2 | Crystal structures of AF9 AHD complexes with DOT1L peptidomimetics

Crystallization experiments were conducted using inhibitor **28**, identified as the most potent DOT1L peptidomimetic

developed in our previous studies (Du et al., 2018). To assess the adaptability of the fusion proteins in facilitating crystallization of the AHD in complex with chemically diverse AHD inhibitors, we synthesized a derivative of the previously reported peptidomimetic **21**. This derivative, referred to as **21a**, features distinct chemical modifications at the N- and C-termini when compared to **28**. The chemical structures of **28** and **21a** are shown in Figure S5. The binding affinities of these peptidomimetics for AF9 AHD were consistent with the previously reported data, determined through competitive fluorescence polarization binding assay (Figure S6) (Du et al., 2018).

Crystals were obtained for the two proteins containing disulfide-bridges, namely **3AQA** and **4AQK**, when in complex with **28**. Interestingly, only protein **3AQA** yielded crystals when in complex with **21a**. Detailed information regarding the crystallization procedure is discussed in Section 4. The overall structures of these complexes are illustrated in Figure 1a.

In all three structures, continuous α-helices extending from the MBP to the AF9 AHD were observed, confirming our earlier predictions (Figure 1a). Difference electron density maps covering the peptidomimetics are shown in Figure 1b. Disulfide bridges between AHD C502 and MBP C336 in **3AQA**, as well as between AHD C501 and MBP C336 in **4AQK** can be clearly observed in the electron density maps shown in Figure 1c. The distances between the Cα atoms of the two cysteine residues are 5.31, 5.16, and 5.15 Å in the complex structures of **4AQK:28**, **3AQA:28**, and **3AQA:21a**, respectively.

The crystal structures of AF9 AHD in **3AQA** and **4AQK** consist of three helices and two β-strands, which are identical to the NMR structures of AF9 AHD in complex with DOT1L (PDB ID: 2MV7) and AF4 (PDB ID: 2LM0) (Kuntimaddi et al., 2015; Leach et al., 2013). The three helices are arranged to create a shallow groove, while the two β-strands form a beta hairpin. The peptidomimetics **28** and **21a** make direct contact with α1 and are orientated anti-parallel to β2. The C-terminus of these compounds is positioned between the helical groove and the β-hairpin (Figure 1a).

By superimposing the MBP molecules in the two complex structures bound to **28** in **3AQA:28** and **4AQK:28**, it was observed that the AHD in **3AQA** undergoes a counterclockwise rotation of approximately
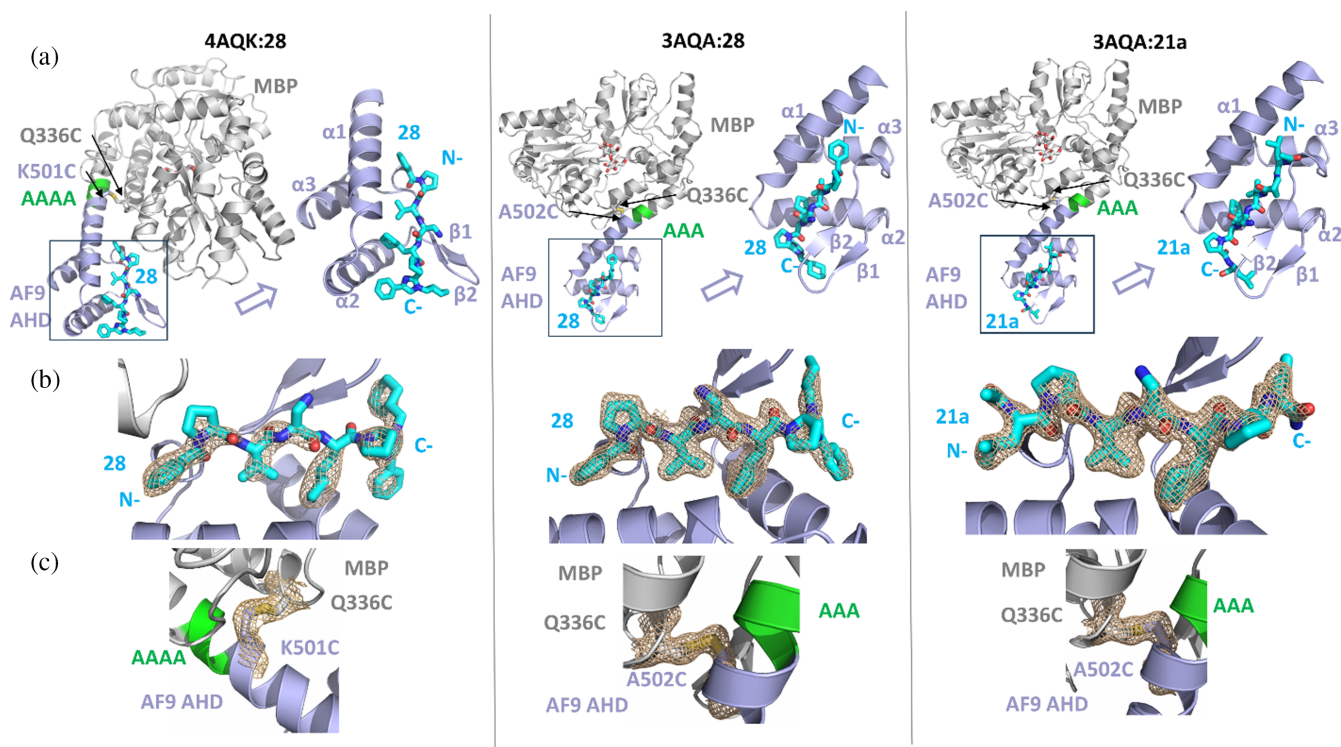

FIGURE 1 The crystal structures of two MBP–AF9 AHD fusion proteins containing disulfide-bridges, **4AQK** and **3AQA**, in complex with peptidomimetics **28** and **21a**. (a) The overall structures of MBP-AAAA-AF9 Q336C/K501C (**4AQK**) in association with **28** and MBP-AAA-AF9 Q336C/A502C (**3AQA**) in complex with **28** and **21a**. MBP is shown in gray, AF9 AHD in light blue, linkers in green, and the peptidomimetic inhibitors in cyan. (b) The F_o–F_c electron density maps contoured at 3σ, calculated with the ligand removed during autoBUSTER refinements, exhibit a positive density at the DOT1L binding site of the AF9 AHD in the structure of **4AQK:28**, **3AQA:28**, and **3AQA:21a**. (c) The $2F_o$–F_c electron density maps, contoured at 1σ, encompass the disulfide bonds established between Q336C of MBP and K501C and A502C of AF9 AHD, in **4AQK** and **3AQA**, respectively.

96° along the common helix compared to that in **4AQK** (Figure 2a). This rotation exposes the binding site of the AHD to the solvent in **3AQA**, where **28** does not participate in crystal contacts, except for the C-terminal phenyl-cyclohexyl-imidazole scaffold, which makes contact with the top of a symmetry-related MBP subdomain 2. In this scenario, the phenyl group of **28** establishes hydrophobic interactions with residues F195, F251, K252, and A191, while the cyclohexyl group is in contact with the carbonyl groups of residues K252, G253, and E131 (Figure 2b) from the symmetry related MBP in **3AQA**.

In contrast to the orientation of the MBP molecule and the AHD observed in **3AQA:28**, the structure of **4AQK:28** exhibits a distinct arrangement where the ligand binding pocket of AHD is directed toward MBP subdomain 1. This positioning partially shields the inhibitor from solvent exposure by MBP, particularly with respects to the benzamide and proline moieties at the N-terminus of **28**. In **4AQK:28**, the phenyl ring of the benzamide establishes hydrophobic interactions with AHD residues L504, L507, V508, L547, and C548. It also forms Van der Waals contacts with two hydrophilic residues of MBP, Q73 and S74 (Figure 2c). The proline moiety of **28** is positioned in close proximity to S74 of MBP. This suggests that MBP in the **4AQK** protein may hinder the binding of inhibitors with bulkier functional groups at the N-terminus. Moreover, differences in the crystal contacts between **3AQA:28** and **4AQK:28** also influence the conformation of **28** at the C-terminus. Specifically, in the **4AQK** complex, **28** shifts by approximately 0.9 Å toward β2. In addition, the β1–β2 hairpin and the α1–α2 loop move in the opposite direction in **4AQK**, resulting in a slightly wider groove in this complex (Figure 2d).
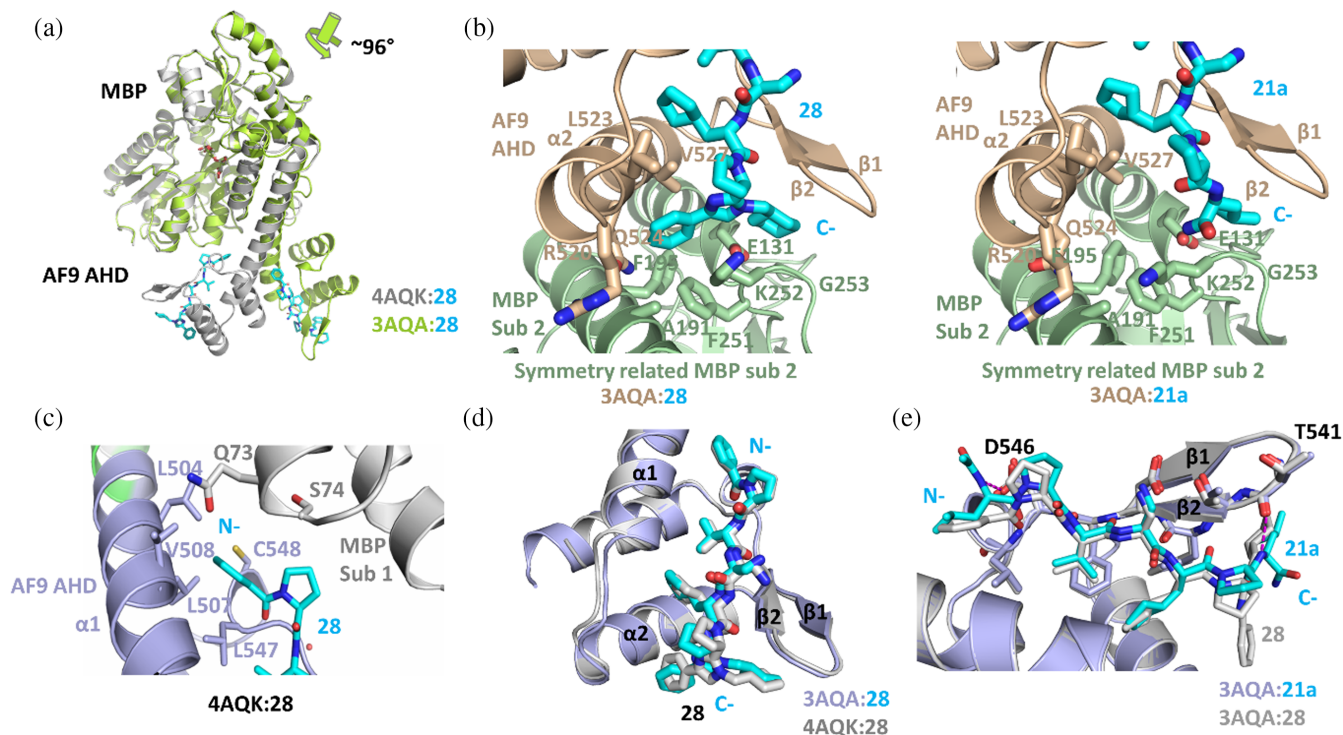
FIGURE 2 Structural details of MBP-AAAA-AF9 AHD (500–568) Q336C/K501C in complex with **28** (**4AQK:28**) and MBP-AAA-AF9 AHD (500–568) Q336C/A502C in association with **28** (**3AQA:28**) and **21a** (**3AQA:21a**). (a) The superimposition of MBP in **3AQA** and **4AQK** reveals that AF9 AHD in **3AQA** rotates ∼96°counterclockwise relative to the AHD of **4AQK**, exposing the inhibitor binding pocket of **3AQA** to the solvent. The structure of **4AQK** is shown in gray and **3AQA** is in green. P **28** is shown in cyan in both structures. (b) Side-by-side comparison of the C-termini of **28** and **21a** in contact with AHD and in crystal contact with a symmetry related MBP in **3AQA**. **28** and **21a** are shown in cyan, AF9 AHD in light brown, and the symmetry related MBP subdomain 2 in light green. The phenyl of **28** interacts with hydrophobic residues of both the AHD and MBP. Despite lacking the phenyl involved interactions, **3AQA:21a** maintains the same crystal lattice as **3AQA:28**, and **21a** maintains the same conformation as **28** near the tip of the β-hairpin of AHD. (c) The N-terminus of **28** (cyan) is in a position sandwiched between MBP (light gray) and AF9 AHD (light blue) in **4AQK:28**, showing interactions with both molecules. (d) The superposition of **3AQA:28** (light blue) and **4AQK:28** (light gray) on AHD demonstrates that the conformation of **28** in the two structures differs at the C-terminus because of crystal contacts; the inhibitor binding groove in **4AQK** is slightly wider. (e) Analysis of **3AQA** in complex with peptidomimetics: **21a** establishes two additional hydrogen bonds (shown in magenta) in comparison to **28**, one at the N-terminus of the peptidomimetic with the D546 side chain and the other at the C-terminus with the T541 carbonyl.

Overall, the configuration of the **3AQA** protein, combined with the crystal contact arrangement in its structure, allows **3AQA** to accommodate inhibitors featuring a wide range of chemical modifications without compromising the integrity of the crystal lattice. This makes it particularly well-suited for crystallization studies of the AHD in complex with diverse inhibitors. Consistent with these observations, our attempts to crystallize the **21a** complex resulted in crystal formation only with **3AQA**, while no crystals were observed in case of **4AQK**. Furthermore, the **3AQA:21a** complex was crystallized under the same conditions and in the same space group, exhibiting comparable cell parameters to the structure of **3AQA:28**.

The structures of **3AQA:28** and **3AQA:21a** exhibit a high degree of similarity, with an RMSD of approximately 0.3 Å for 435 Cα atoms spanning both MBP and AF9. In contrast to **28**, the amide of the N-terminal Leu in **21a** establishes a hydrogen bond with the side chain carboxylate of D546, resulting in a 1.4 Å displacement of the Cα of Leu from the groove, in comparison with that in **28**. This conformational change also affects the nearby proline residue of **21a**, albeit to a lesser extent (0.8 Å). The three middle residues of **28** and **21a** are nearly identical, with the valine side chain and cyclopentyl moiety buried in the hydrophobic pockets, mimicking the conserved hydrophobic interactions of the AHD binding partners, as demonstrated in our previous studies that I867 and I869 of DOT1L are critical hydrophobic residues buried within the DOT1L−AF9 interface (Du et al., 2018; Shen et al., 2013). The amide of the C-terminal Leu in **21a** establishes an additional hydrogen bond with the carbonyl of T541, inducing conformational changes in the backbone of Leu and the preceding Pro of **21a** (Figure 2e). The side chain of the C-terminal Leu of **21a** rests against the β-hairpin, which is occupied by the cyclohexyl group in **28**. The absence of a phenyl group in **21a** creates a void near α2 (Figure 2b) and causing the loss of hydrophobic interactions with residues L523, V527, and aliphatic side chain of R520 and Q524, which are observed in **28** through the phenyl group. Intriguingly, the β-hairpin does not undergo reorganization to accommodate this change, suggesting that the conformation of the β-hairpin remains stable and unaffected by the ligand. On the other hand, as **21a** lacks the C-terminal phenyl group of **28**, the hydrophobic interactions with a symmetry-related MBP observed in the structure of **3AQA:28** are no longer present. This implies that these hydrophobic interactions are not critical for the formation of the crystal lattice of **3AQA** (Figure 2b), and the crystal lattice of **3AQA** exhibits a high degree of flexibility. In summary, when compared to **28**, **21a** forms two additional hydrogen bonds with the AHD but displays reduced hydrophobic interactions at the N- and C-termini. This leads to a five-fold decrease in its affinity for AF9 AHD (Figure S6).

2.3 | The conformation of the AHD β-hairpin is stabilized by conserved hydrogen bonds

The overall complex structures of AF9 AHD with DOT1L peptidomimetics in **3AQA:28**, **3AQA:21a**, and **4AQK:28** closely resemble the previously reported NMR structure of AF9 AHD/DOT1L (PDB ID: 2MV7) (Kuntimaddi et al., 2015), with RMSD values of 1.17, 1.24, and 1.21 Å for superpositions of 63, 65, and 66 Cα atoms, respectively. Differences are observed in the region surrounding the ligand binding groove. In all three structures, β2 shifts away from α1 at the N-terminus of the groove, resulting in an expansion of the groove by approximately 1 Å compared to the NMR structure (Figure 3a). Notable differences are also observed in the β-hairpin region at the C-terminus of the inhibitors (Figure 3a). In the three structures, the side chain of T539 establishes two hydrogen bonds, one with the backbone amide and the other with the side chain of T542. In contrast, a single hydrogen bond between the side chains of T539 and T542 is observed in the NMR structure (Figure 3b and Figure S7). This induces a distinct backbone conformation at T541 in the crystal structures, with its Cα shifted approximately 2.5 Å in **3AQA:28** from its position in the NMR structure.

The C-terminal non-peptidic modification of **28**, that is, the phenyl-cyclohexylmethyl-imidazole group, occupies a pocket that is previously occupied by a short helix of the DOT1L peptide between residues L885 and V888. The cyclohexyl group of **28** is positioned at a site corresponding to L885, while the phenyl group is located at V888 of the DOT1L peptide (Figure 3c). Notably, even in the absence of the bulky non-peptidic modifications at the C-terminus of **21a**, the β-hairpin maintains a conformation resembling that observed in the complex structures of **4AQK:28** and **3AQA:28** (Figure 3a). The presence of the hydrogen bonds between T539 and T542 likely contribute to stabilizing the conformation of the β-hairpin (Figure 3b and Figure S7). These hydrogen bonds involving the two threonine residues are part of a broader hydrogen bonding network that also includes residues D544 and H537 (Figure 3b and Figure S7). D544 makes direct contact with the DOT1L peptide, and previous studies have shown that the AF9 D544R mutation negatively affects the recruitment of DOT1L by MLL-AF9, leading to the loss of H3K79me2 and H3K79me3 at MLL-AF9 target genes (Kuntimaddi et al., 2015).
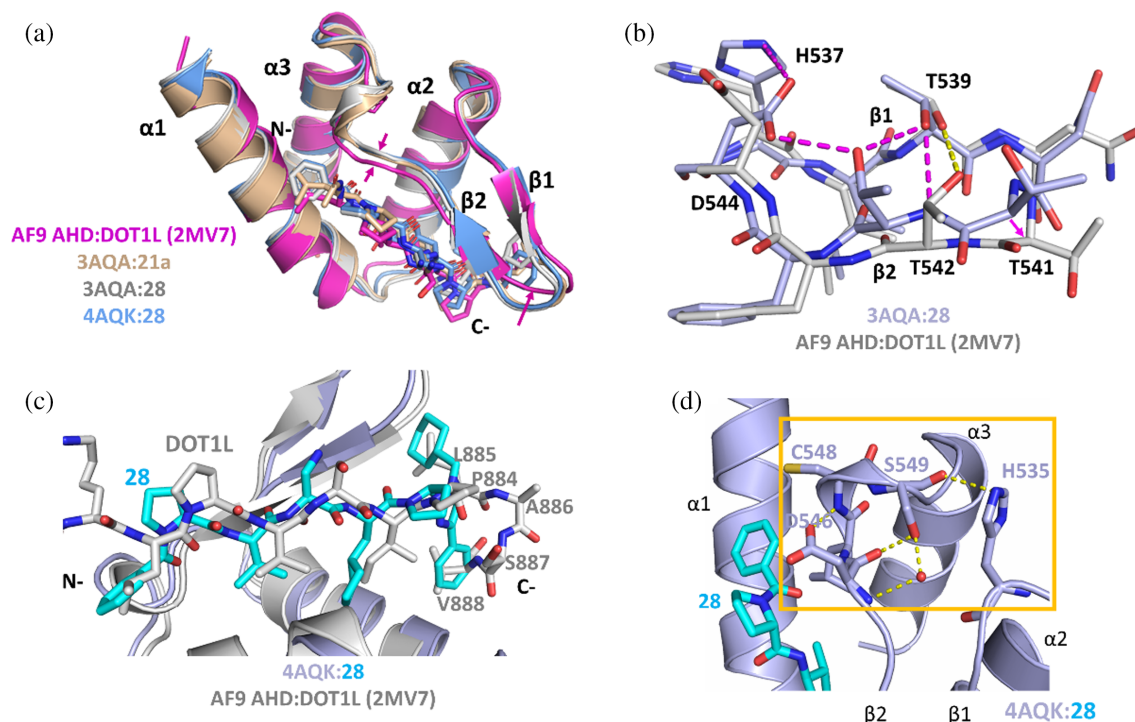
FIGURE 3 Comparison of the inhibitor bound crystal structures and the NMR structure of AF9 AHD in association with DOT1L (PDB ID: 2MV7). (a) The three crystal structures of **4AQK:28**, **3AQA:28**, and **3AQA:21a** exhibit consistent conformational differences on the β2 of the AHD from the NMR structure (magenta). (b) The tip of the β1-2 hairpin shows a unique hydrogen-bonding network in all three crystal structures, represented here using **3AQA:28** (light blue) when compared to the NMR structure PDB IB: 2MV7 (silver). The hydrogen bonds induce a backbone conformational change at T541. The hydrogen bonds in **3AQA:28** are shown in magenta and the hydrogen bond in the NMR structure 2MV7 is shown in yellow. (c) The superposition of AF9 AHD in **4AQK:28** (light blue) and the NMR structure PDB IB: 2MV7 (silver). The C-terminus of DOT1L folds into a helical turn and is substituted in **28** (cyan) by phenyl-cyclohexylmethyl-imidazole group. (d) An extensive hydrogen-bonding network is shown in **4AQK:28**, involving residues between D546 and S549 on a short helical turn connecting β2 and α3, that is conserved throughout AHD domains.
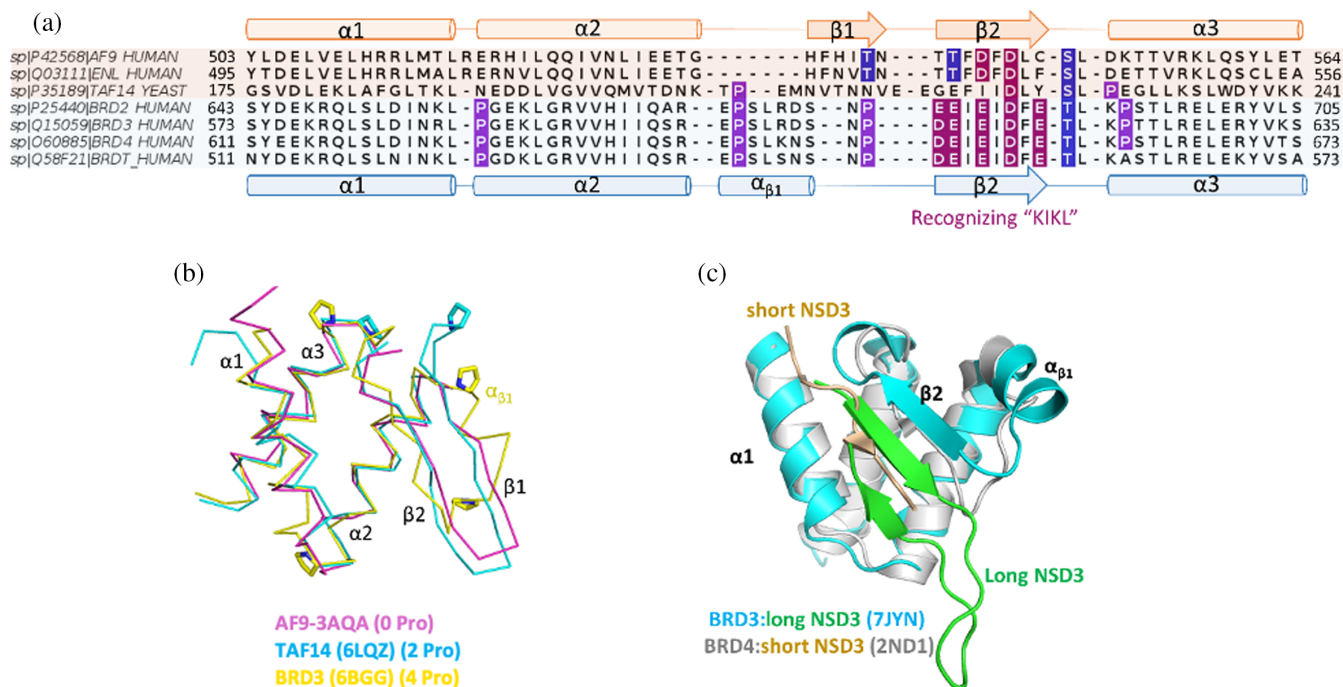
Additionally, we identified a conserved short helical turn connecting β2 and α3 at the N-terminus of the inhibitor binding site, involving residues from D546 to L550. Specifically, the carboxylate side-chain of D546 forms a hydrogen bond with the C548 backbone amide, while the side chain of S549 forms a hydrogen bond with the carbonyl and, through a water molecule, with the amide of D546 on β2. Furthermore, the carbonyl of S549 forms a hydrogen bond with the side chain of H535 on β1 (Figure 3d). Notably, D546 and S549 are highly conserved hydrophilic residues in the AHD (Figure 4a). Collectively, these hydrogen bonds contribute to stabilizing the structure of the AHD.

3 | DISCUSSION

X-ray crystallography is a powerful tool for elucidating the three-dimensional structures of biological macromolecules. It offers valuable insights into the function of a protein by uncovering its structure and interactions with other proteins and small molecules. Complex structures of proteins and their inhibitors are vital for advancing structure-based drug discovery and development (Mazzorana et al., 2020; Zheng et al., 2014). The process of protein crystallization, which entails ordered arrangement of protein molecules, demands a thorough exploration of various protein constructs and meticulous screening of diverse buffers, precipitants, salts, and additives, to pinpoint the ideal conditions for crystal formation. Obtaining crystals suitable for X-ray diffraction and, subsequently, achieving high-resolution structures can be highly unpredictable, which is widely recognized as one of the most challenging aspects of X-ray crystallography. While several alternative techniques exist for studying protein structures, including NMR (Hu et al., 2021; Purslow et al., 2020), cryo-EM (Baker, 2018; Wu & Rapoport, 2021), micro electron diffraction (Xu et al., 2019; Zhu et al., 2020), and X-ray free electron laser (XFEL) (Hough & Owen, 2021; Pandey et al., 2020), X-ray crystallography stands out in providing high resolution structures. The recent breakthrough of AI-assisted

(a)

(b)

AF9-3AQA (0 Pro)
TAF14 (6LQZ) (2 Pro)
BRD3 (6BGG) (4 Pro)

(c)

BRD3:long NSD3 (7JYN)
BRD4:short NSD3 (2ND1)

FIGURE 4 Structural analysis of AF9 AHD in comparison with the ET domains of the BET family of proteins. (a) Sequence alignment of the AHDs of AF9, ENL, and TAF14, as well as the ET domains of BRD2–4 and BRDT. (b) The AF9 AHD (magenta) possesses no proline residues in the protein sequence, whereas the TAF14 AHD (cyan) has two and the BRD3 ET (yellow) has four proline residues. The proline residues in TAF14 and BRD3 are found at the N-cap of a helix or helical turn. (c) The superposition of ET domains in complex with NSD3 peptides. Short NSD3 binds to the ET domain of BRD4 (silver) as a single β-strand (light brown), while the long NSD3 peptide binds to the BRD3 ET domain (cyan) as a β-hairpin (green), widening the peptide binding groove.

AlphaFold has shown promising advancements in predicting protein structures (alphafold.ebi.ac.uk) (Callaway, 2022; Perrakis & Sixma, 2021). However, its accuracy remains limited when predicting the structures of intrinsically disordered proteins/regions (IDPs/IDRs), which lack stable structures under physiological conditions. This study demonstrates a successful protein engineering approach that effectively addresses obstacles in obtaining high-quality crystals of the intrinsically disordered AF9 AHD in complexes with peptidomimetic inhibitors. The methodology presented here offers a strategy for characterizing complex structures of challenging protein targets, such as IDPs and IDRs. Reported three-dimensional structural information will aid the structure-based discovery strategies for targeting protein–protein interactions of AF9 AHD and its binding partners.

To crystallize AF9 AHD, we designed six MBP–AF9 AHD fusion proteins. We investigated how the length and rigidity of the linkers connecting MBP and the AHD affected the crystallization of these proteins. These linkers were designed to position AF9 AHD at various distances and orientations relative to MBP. All six proteins were successfully expressed and purified, highlighting the role of MBP as a molecular chaperone that enhances the solubility and stability of the AHD.

Importantly, MBP did not interfere with the binding of AHD with its partner molecules, namely DOT1L and AF4 (Table 1). Among the six fusion proteins, **3AQA** and **4AQK** were strategically engineered with the introduction of disulfide bonds. These bonds were formed between the C-terminus of MBP (Q336C) and the N-terminus of AHD, specifically at A502C in the fusion protein linked by AAA and K501A in the fusion protein linked by AAAA. The presence of these disulfide bridges contributed to the formation of a rigid and fused interdomain helix extending from MBP to the AHD, decreased conformational heterogeneity of the fusion proteins, and ultimately led to the successful crystallization of their complexes with peptidomimetic inhibitors (Figure 1a). Both constructs maintained accessible inhibitor binding pockets. In particular, the orientation of the AHD in the **3AQA** protein (Figure 2a) was found to be favorable for crystallization with ligands featuring diverse functional and chemical modifications, forming complex structures with both tested peptidomimetics (Table 2).

The crystal structures of AF9 AHD in complex with DOT1L peptidomimetics **28** and **21a** showed remarkable similarity in their overall structures to the previously reported NMR structure of AF9 AHD in complex with DOT1L peptide (PDB ID: 2MV7) (Kuntimaddi

TABLE 2 Crystallographic data collection and refinement statistics.

Data collection	4AQK:28	3AQA:28	3AQA:21a
PDB ID	8TLV	8TLW	8TLX
Space group	$C222_1$	$P4_12_12$	$P4_12_12$
Unit cell a, b, c (Å)	101.906, 156.894, 63.660	75.701, 75.701, 170.789	76.101, 76.101, 172.183
Wavelength (Å)	1.12705	1.12704	1.12704
Resolution[a] (Å)	2.65 (2.70–2.65)	2.10	2.10
R_{merge}[b]	0.090 (0.576)	0.085 (0.355)	0.105 (0.659)
$<I/\sigma I>$[c]	19.9 (2.1)	36.3 (6.3)	24.2 (3.5)
Completeness (%)[d]	99.3 (98.6)	97.2 (95.5)	99.6 (100.0)
Redundancy	8.6 (6.2)	17.4 (17.4)	11.0 (11.1)
Refinement			
Resolution (Å)	2.66	2.11	2.10
R_{work}[e]	0.2018	0.1757	0.1893
R_{free}[f]	0.2461	0.2132	0.2268
Protein atoms	3273	3347	3376
Water molecules	70	356	354
Ligand	61	61	57
Unique reflections	14,920	28,784	30,402
r.m.s.d.[g]			
Bonds (Å)	0.007	0.008	0.008
Angles (°)	0.89	0.90	0.92
Molecules/ASU	1	1	1
MolProbity score[h]	1.64	1.04	1.08
Clash score[h]	2.46	1.63	1.31
RSCC (%)[i]	0.92	0.93	0.92
RSR (%)[i]	0.2	0.11	0.12

[a]Statistics for highest resolution bin of reflections in parentheses.

[b]$R_{merge} = \Sigma_h\Sigma_j|I_{hj} - <I_h>|/\Sigma_h\Sigma_jI_{hj}$, where I_{hj} is the intensity of observation j of reflection h and $<I_h>$ is the mean intensity for multiply recorded reflections.

[c]Intensity signal-to-noise ratio.

[d]Completeness of the unique diffraction data.

[e]R-factor $= \Sigma_h||F_o| - |F_c||/\Sigma_h|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes for reflection h.

[f]R_{free} is calculated against a 10% random sampling of the reflections that were removed before structure refinement.

[g]Root mean square deviation of bond lengths and bond angles.

[h]Chen et al. (2010).

[i]wwPDB Validation Server.

et al., 2015). The backbone of the peptidomimetics overlaps with that of the DOT1L peptide. A key component of the interaction between the AF9 AHD domain and DOT1L is the formation of a hydrophobic core that stabilizes this interaction. Substitution of a single hydrophobic amino acid in DOT1L binding motif can disrupt this interaction (Kuntimaddi et al., 2015; Shen et al., 2013). Consistently with these findings, valine and cyclopentyl methyl group in peptidomimetics are buried within the hydrophobic pocket of the protein, mimicking the conserved residues, V881 and I883 of DOT1L (Figures 1b). These complex structures provided further insights to the reported structure–activity relationships of these peptidomimetics, confirming the importance of the modifications at the N- and C-termini aimed at increasing the binding affinity (Du et al., 2018). In the complex structure of peptidomimetic **21a** it was observed reduced hydrophobic interactions at both termini sites, consistent with a lower affinity for the AHD compared to **28** (Figure 2e). Importantly, it was identified a conformational disparity at the tip of the AHD β1-2 (from T539 to T542) in comparison with the NMR structure (PDB ID: 2MV7) in all three crystal structures, revealing the conserved hydrogen bonds between T539 and T542 in

stabilizing the β-hairpin conformation (Figure 3b and Figure S7). It is worth mentioning that these residues are conserved in ENL but not in Taf14, a yeast AF9/ENL homolog, suggesting that they might also contribute to substrate selectivity (Figure 4a).

Using DALI search (Holm, 2020), it was revealed that the folding of the AF9/ENL AHD resembles the extra terminal (ET) domain of the BET (bromodomain and extra-terminal domain) family of proteins (Figure S8). Like AHDs, ET domains play a crucial role as signaling hubs, mediating protein–protein interactions in several chromatin remodeling complexes (Faure & Callebaut, 2013; Wai et al., 2018). The ET domains selectively bind to peptides with a motif "KIKL" suggesting functional diversity between these two families of proteins (Figure 4a). In addition, the AHD and ET domains exhibit contrasting degrees of stability in their unbound state. Specifically, the AHD in AF9 and ENL is intrinsically disordered, while the ET domains are folded proteins and form stable structures (Aiyer et al., 2021). Of note, the AHD of yeast TAF14 is a structured protein (Aiyer et al., 2021; Chen et al., 2020). Sequence alignments of the AHDs and ET domains have shown that within the ET domain there are four conserved Proline residues (Figure 4a), which are located at the N-caps of α-helices or helical turns (PDB ID: 6BGG) (Figure 4b). Proline substitution at the N-cap of helices has been reported to enhance protein thermostability (Watanabe et al., 1994). TAF14 possesses two proline residues, one at the N-cap of α3 and the other within the helical turn connecting α2 and β1, contributing to the formation of stable structure (Figure 4b). However, although apo TAF14 is a folded protein, it undergoes concentration-dependent aggregation in vitro (Chen et al., 2020). In contrast, the AF9 AHD lacks Pro residues (Figure 4a,b), probably contributing to its structural plasticity. Understanding the factors that influence the structural flexibility of these domains and their impacts on chromatin transcription remains an ongoing pursuit.

Key questions regarding the folding and function of intrinsically disordered proteins remain unresolved (Dyson & Wright, 2002). These inquires include identifying the specific initiation site and mechanism governing the transition from disorder to order when IDP encounters partner molecules. Protein folding often involves the creation of a hydrophobic core driven by the hydrophobic side chains of amino acids (Baldwin & Matthews, 1994; Banach et al., 2020). The interaction of DOT1L with AHD may trigger the formation of localized hydrophobic clusters that subsequently propagate to form the hydrophobic core of the complex protein. It is worth noting that hydrophobic interactions are characterized by their plasticity and lack of directional preferences. The role of hydrophilic interactions, involving both hydrogen bonds and interactions between oppositely charged side chains, is also important during protein folding (Korkmaz et al., 2015). In contrast to relatively long-range hydrophobic and charge interactions, hydrogen bonds are sensitive to both direction and distance. They play a crucial role in stabilizing the secondary structures of a protein through characteristic interactions between the backbone amide and carbonyl groups. Notably, both the unbound AHD and binding region of DOT1L lack apparent secondary structures (Leach et al., 2013). The fundamental question of why, amidst numerous potential side chain and main chain interactions, a protein typically converges on a singular, definitive structural conformation remains a compelling subject of investigation.

Questions regarding the structural dynamics of the intrinsically disordered proteins are also under investigation (De Guzman et al., 2004; Kabra & Bushweller, 2022). The ET domains of BRD3/4 exhibit a highly dynamic substrate binding groove that can accommodate long peptides in the form of a double-stranded anti-parallel β-hairpin, as opposed to a single beta strand. This phenomenon is observed in two complex structures with murine leukemia virus (MLV) integrase C-terminal domain (PDB ID: 7JQ8) and histone-lysine N-methyltransferase NSD3 (148–184) (PDB ID: 7JYN) (Aiyer et al., 2021). By comparing the ET complex structures with longer construct NSD3 (148–184) and shorter NSD3 peptide (152–163) which forms a single β-strand (PDB ID: 2ND1), it was demonstrated that the double-stranded β-hairpin of the longer NSD3 creates a broader substrate binding groove, which might result in tighter binding (Figure 4c). It has been reported that a longer BCOR peptide (1176–1228) exhibits higher binding affinity than a short peptide (1176–1207) to both AF9 and ENL AHDs (Kabra & Bushweller, 2022). This suggests the possibility that the AHD may associate with its binding partners via more than one β-strand. While the precise conformational dynamics of the AHD have not been fully elucidated, it is likely that a highly dynamic structure plays a role in facilitating exchange among binding partner proteins. Conformational changes induced by their association and dissociation can be further complicated by competitions from local concentration and variations in binding affinity (De Guzman et al., 2004). The mechanism of displacement and molecular recognition between AF9 AHD domain and its binding partner proteins is particularly of interest, especially in the context of being an intrinsically disordered protein.

There is a growing interest in the AHDs of AF9 and ENL as potential targets for cancer therapy (Wu et al., 2021). Our structural investigations of the intrinsically disordered AF9 AHD in complex with peptidomimetic inhibitors will facilitate drug development targeting this domain. These structural insights have also

raised important questions regarding the folding and structural dynamics of IDP/IDRs. Overall, this work holds promise for advancing cancer therapy and enhancing our understanding of intrinsically disordered proteins.

4 | MATERIALS AND METHODS

4.1 | Cloning of MBP and AF9 AHD fusion proteins with various linkers and mutations

All six constructs of the MBP (2–366) and AF9 AHD (500–568) fusion proteins were generated using a multi-step polymerase chain reaction (PCR) and ligated into a modified pET28-6xHis-TEV (Tobacco Etch Virus) vector using BamHI/XhoI sites. The mutations of respective MBP Q336C, AF9 AHD A502C, and AF9 AHD K501C were incorporated in the primers. All plasmids were transformed into Rosetta™ (DE3) cells for protein expression.

4.2 | Expression and purification of the MBP and AF9 AHD fusion proteins

Each protein was expressed at 18°C in 0.5 L terrific broth (TB) medium. Pellets were re-suspended in 50 mL of lysis buffer containing 50 mM Tris 7.4, 150 mM NaCl, 10% glycerol, 1:100 protease inhibitor cocktail (Bimake, B14002), 100 μg/mL lysozyme, 20 mM imidazole, 3 mM tris(2-carboxyethyl)phosphine (TCEP, pH 7) followed by sonication on ice. The cell lysate was centrifuged at $38,000 \times g$ for 1 h at 4°C. The supernatant was then incubated for 2 h at 4°C with 3 mL of Ni^{2+}-nitrilotriacetic acid (Ni-NTA, HisPurTM Ni-NTA, Thermo Scientific 88222). The protein was eluted with buffer containing 50 mM Tris 7.4, 150 mM NaCl, 10% glycerol, and gradient concentrations of imidazole from 50 to 250 mM. The eluted protein was treated with TEV at a TEV: protein weight ratio of 1:100 for overnight with tumbling. The protein was subsequently purified on Superdex 200 using size exclusion chromatography (Figure S2). The concentrated protein was stored at −80°C in a buffer containing 25 mM Tris 7.4, 150 mM NaCl, and 1 mM TCEP.

4.3 | Crystallization of MBP-AAAA-AF9 Q336C/K501C in complex with 28 (4AQK:28)

Co-crystallization experiments were conducted using Crystal Screen HT and Index Screen kits from Hampton Research as well as Wizard Classic 1&2 from Rigaku. Crystals were optimized by gradient screen. The fusion protein **4AQK** at an approximate concentration of 20 mg/mL was mixed with **28** at a protein-inhibitor molar ratio of 1:2 and incubated for 2 h on ice with the addition of 5 mM maltose and 3 mM glutathione disulfide. The complex was crystallized using a protein: well volume ratio of 1:2 in a sitting drop vapor diffusion experiment against reservoir solution of 20%–30% PEG 3350, 0.1 M HEPES pH 7.5, and 0.2 M ammonium acetate. Crystals grew within 1 day at 20°C and were cryoprotected in the crystallization solution with the addition of 20% glycerol before being flash-frozen in liquid N_2 prior to data collection.

4.4 | Crystallization of MBP-AAA-AF9 Q336C/A502C in complex with compound 28 (3AQA:28) and 21a (3AQA:21a)

Co-crystallization experiments were conducted using Crystal Screen HT and Index Screen kits from Hampton Research as well as Wizard Classic 1&2 from Rigaku. Crystals were optimized by gradient screen. The fusion protein **3AQA** and inhibitor protein complexes were prepared according to the above **4AQK:28** complexes. In a sitting drop vapor diffusion experiment, 0.4 μL of the protein complex was mixed with 0.8 μL reservoir solution containing 20%–30% PEG 3350 and 0.1 M Bis-Tris pH 5.5, and incubated for 3 h at 20°C. The drop was then seeded with 0.2 μL of previously prepared seeds. Crystals grew within 1 day at 20°C and were cryoprotected in the crystallization solution with the addition of 20% glycerol before being flash-frozen in liquid N_2 prior to data collection.

4.5 | Crystal structure determination and refinements

Diffraction data were collected on LS-CAT 21-ID-D beamline at the Advanced Photon Source at Argonne National Laboratory and processed with HKL2000 (Otwinowski & Minor, 1997). The structures were solved by molecular replacement with Phaser (McCoy et al., 2007) in the CCP4i suite (Potterton et al., 2003) using search models for MBP from the reported structure PDB ID: 4WMS and AF9 AHD from PDB ID: 2MV7. Iterative rounds of electron density fitting and refinement were completed using Coot (Emsley et al., 2010) and Buster (Smart et al., 2012) respectively. The coordinates and geometric restraints for each inhibitor were created from SMILES using AceDRG (Long et al., 2017) in CCP4i

suite. The coordinates were validated with MolProbity (Williams et al., 2018). Data collection and refinement statistics were listed in Table 2. Difference electron density maps (Figure 1b) showed that one inhibitor was bound in the DOT1L binding pocket of AF9 AHD. In the **4AQK:28** crystal, MBP residues 2–367 and AF9 AHD residues 500–562 linked by four Ala residues were present in the electron density; in the **3AQA:28** crystal, MBP 2–367 and AF9 AHD 500–565 linked by three Ala were present in the electron density; in the **3AQA:21a** crystal, MBP residues 1–367 and AF9 AHD residues 500–565 linked by three Ala are present in the electron density.

ACCESS CODE

The crystal structures of **4AQK:28**, **3AQA:28**, and **3AQA:21a** were deposited to Protein Data Bank under accession codes (PDB ID) **8TLV**, **8TLW**, and **8TLX**, respectively.

AUTHOR CONTRIBUTIONS

Yuting Yang: Writing – original draft; methodology; formal analysis; data curation; writing – review and editing; investigation. **Ejaz Ahmad:** Methodology; data curation; formal analysis. **Vidhya Premkumar:** Methodology. **Alicen Liu:** Methodology. **S. M. Ashikur Rahman:** Data curation and analysis. **Zaneta Nikolovska-Coleska:** Conceptualization; funding acquisition; writing – review and editing; supervision; resources; project administration; investigation; formal analysis; methodology; validation.

ORCID

Yuting Yang https://orcid.org/0000-0003-2077-2780
Ejaz Ahmad https://orcid.org/0000-0002-5713-6713
S. M. Ashikur Rahman https://orcid.org/0000-0002-3791-977X
Zaneta Nikolovska-Coleska https://orcid.org/0000-0001-6688-4206

REFERENCES

Aiyer S, Swapna GVT, Ma LC, Liu G, Hao J, Chalmers G, et al. A common binding motif in the ET domain of BRD3 forms polymorphic structural interfaces with host and viral proteins. Structure. 2021;29(8):886–898 e886.

Baker M. Cryo-electron microscopy shapes up. Nature. 2018; 561(7724):565–7.

Baldwin EP, Matthews BW. Core-packing constraints, hydrophobicity and protein design. Curr Opin Biotechnol. 1994;5(4):396–402.

Banach M, Fabian P, Stapor K, Konieczny L, Roterman AI. Structure of the hydrophobic Core determines the 3D protein structure-verification by single mutation proteins. Biomolecules. 2020;10(5):767.

Bergfors TM. Protein crystallization. La Jolla: International University Line; 2022.

Bregegere F, Schwartz J, Bedouelle H. Bifunctional hybrids between the variable domains of an immunoglobulin and the maltose-binding protein of Escherichia coli: production, purification and antigen binding. Protein Eng. 1994;7(2):271–80.

Bukowska MA, Grutter MG. New concepts and aids to facilitate crystallization. Curr Opin Struct Biol. 2013;23(3):409–16.

Callaway E. What's next for AlphaFold and the AI protein-folding revolution. Nature. 2022;604(7905):234–8.

Chen VB, Bryan Arendall W, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. 2010;D66:12–21.

Chen G, Wang D, Wu B, Yan F, Xue H, Wang Q, et al. Taf14 recognizes a common motif in transcriptional machineries and facilitates their clustering by phase separation. Nat Commun. 2020; 11(1):4206.

De Guzman RN, Martinez-Yamout MA, Dyson HJ, Wright PE. Interaction of the TAZ1 domain of the CREB-binding protein with the activation domain of CITED2: regulation by competition between intrinsically unstructured ligands for non-identical binding sites. J Biol Chem. 2004;279(4):3042–9.

Deiana A, Forcelloni S, Porrello A, Giansanti A. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. PLoS One. 2019;14(8):e0217889.

Derewenda ZS. Application of protein engineering to enhance crystallizability and improve crystal properties. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 5):604–15.

di Guan C, Li P, Riggs PD, Inouye H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene. 1988;67(1):21–30.

Du L, Grigsby SM, Yao A, Chang Y, Johnson G, Sun H, et al. Peptidomimetics for targeting protein-protein interactions between DOT1L and MLL oncofusion proteins AF9 and ENL. ACS Med Chem Lett. 2018;9(9):895–900.

Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol. 2002;12(1):54–60.

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4): 486–501.

Faure G, Callebaut I. Identification of hidden relationships from the coupling of hydrophobic cluster analysis and domain architecture information. Bioinformatics. 2013;29(14):1726–33.

Grigsby SM, Friedman A, Chase J, Waas B, Ropa J, Serio J, et al. Elucidating the importance of DOT1L recruitment in MLL-AF9 leukemia and hematopoiesis. Cancers (Basel). 2021;13(4):642.

He N, Chan CK, Sobhian B, Chou S, Xue Y, Liu M, et al. Human polymerase-associated factor complex (PAFc) connects the super elongation complex (SEC) to RNA polymerase II on chromatin. Proc Natl Acad Sci U S A. 2011;108(36):E636–45.

Holm L. DALI and the persistence of protein shape. Protein Sci. 2020;29(1):128–40.

Hough MA, Owen RL. Serial synchrotron and XFEL crystallography for studies of metalloprotein catalysis. Curr Opin Struct Biol. 2021;71:232–8.

Hu Y, Cheng K, He L, Zhang X, Jiang B, Jiang L, et al. NMR-based methods for protein analysis. Anal Chem. 2021;93(4):1866–79.

Hu H, Saha N, Yang Y, Ahmad E, Lachowski L, Shrestha U, et al. The ENL YEATS epigenetic reader domain critically links MLL-ENL to leukemic stem cell frequency in t(11;19) leukemia. Leukemia. 2023;37(1):190–201.

Jin T, Chuenchor W, Jiang J, Cheng J, Li Y, Fang K, et al. Design of an expression system to enhance MBP-mediated crystallization. Sci Rep. 2017;7:40991.

Kabra A, Bushweller J. The intrinsically disordered proteins MLLT3 (AF9) and MLLT1 (ENL) – multimodal transcriptional switches with roles in normal hematopoiesis, MLL fusion leukemia, and kidney cancer. J Mol Biol. 2022;434(1):167117.

Korkmaz EN, Volkman BF, Cui Q. Interplay of electrostatics and hydrophobic effects in the metamorphic protein human lymphotactin. J Phys Chem B. 2015;119(30):9547–58.

Kotschy A, Szlavik Z, Murray J, Davidson J, Maragno AL, Le Toumelin-Braizat G, et al. The MCL1 inhibitor S63845 is tolerable and effective in diverse cancer models. Nature. 2016; 538(7626):477–82.

Kulkarni P, Bhattacharya S, Achuthan S, Behal A, Jolly MK, Kotnala S, et al. Intrinsically disordered proteins: critical components of the wetware. Chem Rev. 2022;122(6):6614–33.

Kuntimaddi A, Achille NJ, Thorpe J, Lokken AA, Singh R, Hemenway CS, 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): 808–20.

Leach BI, Kuntimaddi A, Schmidt CR, Cierpicki T, Johnson SA, Bushweller JH. Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure. 2013;21(1): 176–83.

Li Y, Wen H, Xi Y, Tanaka K, Wang H, Peng D, et al. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. Cell. 2014;159(3):558–71.

Long F, Nicholls RA, Emsley P, Graaeulis S, Merkys A, Vaitkus A, et al. AceDRG: a stereochemical description generator for ligands. Acta Crystallogr D Struct Biol. 2017;73(Pt 2):112–22.

Luo Z, Lin C, Shilatifard A. The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol. 2012; 13(9):543–7.

Marschalek R. The reciprocal world of MLL fusions: a personal view. Biochim Biophys Acta Gene Regul Mech. 2020;1863(7): 194547.

Martin BJE, Brind'Amour J, Kuzmin A, Jensen KN, Liu ZC, Lorincz M, et al. Transcription shapes genome-wide histone acetylation patterns. Nat Commun. 2021;12(1):210.

Mazzorana M, Shotton EJ, Hall DR. A comprehensive approach to X-ray crystallography for drug discovery at a synchrotron facility – the example of diamond light source. Drug Discov Today Technol. 2020;37:83–92.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007;40(Pt 4):658–74.

Meyer C, Burmeister T, Groger D, Tsaur G, Fechina L, Renneville A, et al. The MLL recombinome of acute leukemias in 2017. Leukemia. 2018;32(2):273–84.

Mohan M, Herz HM, Takahashi YH, Lin C, Lai KC, Zhang Y, et al. Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes Dev. 2010; 24(6):574–89.

Moon AF, Mueller GA, Zhong X, Pedersen LC. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci. 2010;19(5): 901–13.

Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997;276:307–26.

Pandey S, Bean R, Sato T, Poudyal I, Bielecki J, Cruz Villarreal J, et al. Time-resolved serial femtosecond crystallography at the European XFEL. Nat Methods. 2020;17(1):73–8.

Perrakis A, Sixma TK. AI revolutions in biology: the joys and perils of AlphaFold. EMBO Rep. 2021;22(11):e54046.

Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. Acta Crystallogr D Biol Crystallogr. 2003;59(Pt 7):1131–7.

Prange KHM, Mandoli A, Kuznetsova T, Wang SY, Sotoca AM, Marneth AE, et al. MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia. Oncogene. 2017; 36(23):3346–56.

Purslow JA, Khatiwada B, Bayro MJ, Venditti V. NMR methods for structural characterization of protein-protein complexes. Front Mol Biosci. 2020;7:9.

Raran-Kurussi S, Waugh DS. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. PLoS One. 2012;7(11):e49589.

Rhodes G. Other kinds of macromolecular models. Crystallography made crystal clear: a guide for users of macromolecular models. Cambridge: Academic Press; 2006.

Sabari BR, Tang Z, Huang H, Yong-Gonzalez V, Molina H, Kong HE, et al. Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. Mol Cell. 2015;58(2):203–15.

Schmidt CR, Achille NJ, Kuntimaddi A, Boulton AM, Leach BI, Zhang S, et al. BCOR binding to MLL-AF9 is essential for leukemia via altered EYA1, SIX, and MYC activity. Blood Cancer Discov. 2020;1(2):162–77.

Shen C, Jo SY, Liao C, Hess JL, Nikolovska-Coleska Z. Targeting recruitment of disruptor of telomeric silencing 1-like (DOT1L): characterizing the interactions between DOT1L and mixed lineage leukemia (MLL) fusion proteins. J Biol Chem. 2013; 288(42):30585–96.

Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. Acta Crystallogr D Biol Crystallogr. 2012;68(Pt 4):368–80.

Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. Protein Sci. 2003;12(7):1313–22.

Ullah H, Scappini EL, Moon AF, Williams LV, Armstrong DL, Pedersen LC. Structure of a signal transduction regulator, RACK1, from Arabidopsis thaliana. Protein Sci. 2008;17(10):1771–80.

Wai DCC, Szyszka TN, Campbell AE, Kwong C, Wilkinson-White LE, Silva APG, et al. The BRD3 ET domain recognizes a short peptide motif through a mechanism that is conserved across chromatin remodelers and transcriptional regulators. J Biol Chem. 2018;293(19):7160–75.

Wang QF, Wu G, Mi S, He F, Wu J, Dong J, et al. MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. Blood. 2011;117(25):6895–905.

Watanabe K, Masuda T, Ohashi H, Mihara H, Suzuki Y. Multiple proline substitutions cumulatively thermostabilize Bacillus cereus ATCC7064 oligo-1,6-glucosidase. Irrefragable proof supporting the proline rule. Eur J Biochem. 1994;226(2):277–83.

Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 2018;27(1):293–315.

Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16(1):18–29.

Wu F, Nie S, Yao Y, Huo T, Li X, Wu X, et al. Small-molecule inhibitor of AF9/ENL-DOT1L/AF4/AFF4 interactions suppresses malignant gene expression and tumor growth. Theranostics. 2021;11(17):8172–84.

Wu X, Rapoport TA. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). Proc Natl Acad Sci U S A. 2021;118(41):e2115001118.

Xu H, Lebrette H, Clabbers MTB, Zhao J, Griese JJ, Zou X, et al. Solving a new R2lox protein structure by microcrystal electron diffraction. Sci Adv. 2019;5(8):eaax4621.

Zheng H, Hou J, Zimmerman MD, Wlodawer A, Minor W. The future of crystallography in drug discovery. Expert Opin Drug Discovery. 2014;9(2):125–37.

Zhu L, Bu G, Jing L, Shi D, Lee MY, Gonen T, et al. Structure determination from lipidic cubic phase embedded microcrystals by MicroED. Structure. 2020;28(10):1149–1159 e1144.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.