Integrating NMR, SAXS, and Atomistic Simulations: Structure and Dynamics of a Two-Domain Protein

Karl T. Debiec,[1,2,3] Matthew J. Whitley,[2] Leonardus M. I. Koharudin,[2] Lillian T. Chong,[3] and Angela M. Gronenborn[2,*]

[1]Molecular Biophysics and Structural Biology Graduate Program, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, Pennsylvania; [2]Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; and [3]Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT   Multidomain proteins with two or more independently folded functional domains are prevalent in nature. Whereas most multidomain proteins are linked linearly in sequence, roughly one-tenth possess domain insertions where a guest domain is implanted into a loop of a host domain, such that the two domains are connected by a pair of interdomain linkers. Here, we characterized the influence of the interdomain linkers on the structure and dynamics of a domain-insertion protein in which the guest LysM domain is inserted into a central loop of the host CVNH domain. Expanding upon our previous crystallographic and NMR studies, we applied SAXS in combination with NMR paramagnetic relaxation enhancement to construct a structural model of the overall two-domain system. Although the two domains have no fixed relative orientation, certain orientations were found to be preferred over others. We also assessed the accuracies of molecular mechanics force fields in modeling the structure and dynamics of tethered multidomain proteins by integrating our experimental results with microsecond-scale atomistic molecular dynamics simulations. In particular, our evaluation of two different combinations of the latest force fields and water models revealed that both combinations accurately reproduce certain structural and dynamical properties, but are inaccurate for others. Overall, our study illustrates the value of integrating experimental NMR and SAXS studies with long timescale atomistic simulations for characterizing structural ensembles of flexibly linked multidomain systems.

INTRODUCTION

Multidomain proteins in which the connected domains each fold and function independently are prevalent in nature (1,2). Such proteins, through spatial and temporal coordination of their varied functional units, are capable of executing specific and tailored activities in catalysis, signaling, regulation of gene expression, and other cellular processes (3). The individual domains are connected by interdomain linkers whose length and composition enable them to adopt orientations that have evolved for specific biological activities and functions (4,5). In many cases, the linkers are highly flexible, allowing the domains to adopt numerous interdomain orientations, from which the selection of functional competent conformations may occur (6). Although most multidomain proteins are linked linearly in sequence, roughly one-tenth possess domain insertions where a guest domain is implanted into a loop of a host domain, such that the two domains are connected by a pair of interdomain linkers (7).

Characterization of the relative domain orientations within multidomain proteins has been challenging by traditional structural biology techniques, such as x-ray crystallography, due to the inherent flexibility of interdomain linkers, lack of density for certain segments of the polypeptide chain, and influence of crystal packing on the positioning of domains. On the other hand, multidomain proteins represent intriguing targets for integrative structural biology approaches, which combine results from experiments and computer simulations (5,8) by either 1) computationally generating a large ensemble of potential structural models and subsequently filtering the models based on agreement with the experimental data, or 2) explicitly biasing the generation of structural models in accord with the experimental data. Such approaches have been particularly useful for studying flexibly linked multidomain proteins and protein complexes (9–15), often integrating data from NMR, SAXS, x-ray crystallography, and other experimental techniques into a single structural model. Critically, the validity of any approaches, aimed at bridging the gaps between experimentally accessible and computationally generated data, depends on the accuracy of the

---

biomolecular force fields used in the computations, which dictate sampling of the conformational space for the entire system.

Traditionally, force fields have been parameterized to reproduce the properties of small molecules, with parameters derived from experiment and quantum mechanical calculations, and their accuracy for biomolecules are validated using simulations of well-characterized benchmark systems. Such systems have included small globular proteins (e.g., ubiquitin, GB3, and lysozyme) (16–19), although more recently, simulations of significantly more flexible, disordered peptides and proteins (e.g., the MDM2-binding p53 peptide and α-synuclein) have been carried out (18,20,21). The latter have revealed that most force fields, when paired with their intended explicit water models, suffer from an imbalance between protein-protein and protein-water interactions, yielding conformations of nonglobular proteins that are much more compact than experimentally observed, as well as overstabilizing the folded states of globular proteins (21). Because the conformational space accessible to nonglobular systems is very large, exhaustive sampling is beyond the capabilities of current simulation methods, and experimentally well-characterized systems are necessary to serve as a link between globular and nonglobular proteins. In particular, simultaneous validation of both the interior protein structure and the balance between protein-protein and protein-water interactions is needed. Flexibly linked, multidomain proteins present an ideal opportunity to fulfill this requirement, because the interdomain conformational space of such proteins is large, yet sufficiently restricted to be addressed with current simulation methods. Such affordable, yet complex model systems are becoming increasingly valuable as computational methods shift toward more intricate and expensive algorithms, such as implemented in the AMOEBA and CHARMM Drude polarizable force fields (22,23).

An ideal test system among flexibly linked multidomain proteins is the relatively small, two-domain protein MoCVNH3 that has been structurally characterized by our group using both NMR spectroscopy and x-ray crystallography (24,25). MoCVNH3 is a domain-insertion protein in which a guest LysM domain is inserted into a surface loop of a host Cyanovirin-N Homology (CVNH) domain, positioning the LysM domain between the two pseudosymmetric halves of a two-lobed CVNH domain (26). This protein is found in *Magnaporte oryzae*, an ascomycete fungus that causes rice blast disease, the most devastating infection of cultivated rice, which destroys crops in unprecedented amounts worldwide (27). Functionally, both CVNH and LysM are carbohydrate-binding domains: CVNH binds to mannose sugars, whereas LysM interacts with GlcNAc-containing carbohydrates such as peptidoglycan and chitin (28,29). The binding of carbohydrates by each domain in MoCVNH3 is independent of the other,

with no communication between the domains (30). Although the wild-type protein could not be crystallized, complete removal of the interdomain linkers yielded a construct that crystallized and maintained the ability to bind both carbohydrate ligands. A comparison of the resulting crystal structure with the NMR structure of wild-type MoCVNH3 (Mo-WT) revealed that the absence of the linkers has no effect on the structures of the individual domains (25). However, although the domain structures of wild-type MoCVNH3 were determined to high-resolution by NMR, no fixed relative domain orientations were compatible with the solution data, due to the lack of interdomain restraints (30).

Here, we investigated the influence of interdomain linker length on the overall structure and dynamics of MoCVNH3, as well as the conformational space of interdomain orientations in solution, using an integrated approach that combines biophysical experiments and molecular dynamics (MD) simulations. In particular, we performed SAXS, NMR relaxation, and paramagnetic relaxation enhancement (PRE) experiments along with microsecond (μs)-scale MD simulations in explicit solvent. In carrying out the simulations, we compared the accuracy of two biomolecular force-field/water-model combinations in modeling the structure and dynamics of this tethered two-domain protein, and establish that each combination is accurate for certain properties and inaccurate for others. Overall, we demonstrate that an integrated approach, incorporating different experimental and computational methods, permits characterization of both the interdomain orientations and dynamics of multidomain proteins, using the MoCVNH3 protein as an example.

MATERIALS AND METHODS

Protein expression and purification

Proteins were expressed and purified as described previously for wild-type MoCVNH3 (30). In brief, pET-15b(+) vectors containing the different coding sequences for the individual protein constructs (Fig. 1) were used to transform *Escherichia coli* Rosetta2 (DE3) cells (Novagen, Merck, Darmstadt, Germany). All constructs encoded N-terminal polyhistidine tags, followed by a TEV protease cleavage site. After cleavage, the native protein N-terminus was obtained, removing a four-residue addition that was present in the earlier protein constructs (25). Mutant coding sequences were created using the QuikChange XL II site-directed mutagenesis kit (Stratagene, La Jolla, CA). Cells were initially grown at 37°C, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside at an OD_{600} of ~0.8, and further grown for 18 h at 16°C for protein expression. Cells were harvested by centrifugation, resuspended in lysis buffer (25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM DTT, and 3 mM NaN$_3$) and ruptured by passage through a microfluidizer (MicroFluidics M-110Y; Hyland Scientific, Stanwood, WA). Cell debris was removed by ultracentrifugation (19,000 RPM), and the supernatant was loaded onto an Ni^{2+}-derivatized HisTrap column (GE Healthcare, Little Chalfont, UK), preequilibrated with loading buffer (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM DTT, and 3 mM NaN$_3$). Proteins were eluted using a linear (25–500 mM) imidazole gradient in the same buffer and protein-containing fractions were subjected to TEV digestion in 25 mM Tris-HCl buffer,

FIGURE 1 Mo-WT and reduced linker-length constructs Mo-2G, Mo-0G and Mo-0v. (*A*) ^{1}H-^{15}N HSQC NMR spectra of Mo-WT. Resonances corresponding to residues in the CVNH domain are shown in red, in the LysM domain in blue, and in the interdomain linkers in green. Side-chain resonances are shown in gray. (*B*) Amino acid sequences of the MoCVNH3 constructs, highlighting the interdomain linker regions. Amino acids in the CVNH domain are shown in red, in the LysM domain in blue, and in the interdomain linkers in green. (*C*) Chemical shift differences between Mo-WT and Mo-2G, Mo-0G, and Mo-0v constructs. (*D*) Structures of MoCVNH3 constructs. The Mo-WT structure was previously determined by NMR (30), and Mo-0v by X-ray crystallography (25), whereas Mo-2G and Mo-0G are represented by homology models. Selected residue positions highlighted in the text are labeled in the Mo-WT structure. To see this figure in color, go online.

pH 8.0, 25 mM NaCl, 5 mM DTT, and 3 mM NaN$_3$ for removal of the N-terminal polyhistidine tag. Further purification of the cleaved proteins involved gel filtration on a Superdex75 column (GE Healthcare) in 25 mM sodium acetate buffer, pH 5.0, 25 mM NaCl, 5 mM DTT, and 3 mM NaN$_3$, and cation exchange on an HisTrap SP column (GE Healthcare), preequilibrated with loading buffer (25 mM sodium acetate (pH 5.0), 25 mM NaCl, 5 mM DTT, 3 mM NaN$_3$), and elution by a linear (0–250 mM) NaCl gradient in the same buffer. Protein-containing fractions were collected, buffer exchanged into 25 mM sodium acetate, pH 5.0, 25 mM NaCl, 5 mM DTT, 3 mM NaN$_3$, and concentrated using Centriprep devices (Millipore, Burlington, MA). For ^{15}N isotopic labeling, the bacterial cell culture was grown in modified minimal medium, containing ^{15}NH$_4$Cl as the sole nitrogen source, whereas for ^{13}C isotopic labeling ^{13}C-glucose was provided as the sole carbon source. The purity and identity of all proteins were confirmed by SDS-PAGE and mass spectrometry.

Site-selective spin-labeling

The purified single-cysteine variant of Mo-WT, Mo-SAVC (C15S, C25A, C82V), was split into two equal portions for parallel spin-labeling at the C-terminal C167 with (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate (MTSL) and the diamagnetic analog of MTSL: (1-acetyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl)methanethiosulfonate (dMTSL; Toronto Research Chemicals, Toronto, ON, Canada), in which the oxygen on the nitroxide of MTSL is replaced with an acetyl group. Both reagents were added from 0.2 mM stocks in 10-fold molar excess to a 65 μM protein solution in 25 mM NaPhosphate buffer, pH 7.0, 25 mM NaCl, 3 mM NaN$_3$, and the reaction mixture was incubated at 4°C for 16 h in the dark. Excess free tags were removed by serial dialysis against 25 mM NaAc buffer, pH 5.0, 25 mM NaCl, 3 mM NaN$_3$. The extent of spin-labeling (>95%) was confirmed by mass spectrometry.

MD simulations

Heavy atom coordinates of the Mo-WT and Mo-0v constructs were extracted from the solution NMR structure of Mo-WT (PDB: 2L9Y) (30) and x-ray crystal structure of Mo-0v (PDB: 58CO) (25), respectively. Coordinates of the reduced linker-length constructs Mo-2G and Mo-0G were generated using the MODELER 9.9 software package (31) based on the coordinates of the CVNH and LysM domains from the Mo-WT NMR structure. Each system was solvated in a $105 \times 105 \times 105$ Å cubic box, which generated a minimum solute-wall distance of 22 Å for the largest two-domain construct (Mo-WT). Amino acid side chains were assigned protonation states consistent with the experimental pH of 5.0; i.e., arginine, lysine, and histidine residues were protonated whereas aspartate and glutamate were deprotonated. The net positive charges on the protein were neutralized through the addition of Cl^- ions; additional Na^+ and Cl^- ions were added to be consistent with the experimental salt concentration of 25 mM. Identical system configurations were used for simulations run using the AMBER software package (32) and on ANTON (33).

Simulations run using AMBER were carried out using the GPU implementation of the pmemd module (32,34,35). Before running production simulations, each system was subjected to energy minimization, followed by a three-stage equilibration. In the first stage, a 20-ps simulation of the energy minimized system was carried out at constant temperature, while restraining the solute heavy atoms to their initial positions using a harmonic potential with a force constant of 1 kcal/mol·Å². In the second stage, a 1-ns simulation was carried out at constant pressure with the same harmonic restraints on atom positions. Finally, an additional 1-ns unrestrained simulation was carried out at constant temperature and pressure. The temperature was maintained at 25°C using a Langevin thermostat (frictional constant of 0.1 ps⁻¹) whereas the pressure was maintained at 1 atm using a Monte Carlo barostat (200 fs between attempts to change the system volume) (36). The van der Waals and short-range electrostatic interactions were truncated at 10 Å; long-range electrostatic interactions were calculated using the particle mesh Ewald method (37). To enable a 4-fs time step, hydrogens were constrained to their equilibrium values using the SHAKE and SETTLE algorithms, and hydrogen mass repartitioning was used (38–40). The masses of solute hydrogen atoms were increased by a factor of three, and that of their attached heavy atoms decreased by the corresponding amount, such that the total mass remained constant; the masses of water molecules were not repartitioned. Coordinates were saved every 100 ps for analysis.

Simulations run on the ANTON special-purpose supercomputer were equilibrated using the Desmond 3.0.1.0 software package (33,41). Each system was subjected to energy minimization followed by a 20-ps equilibration at constant temperature, and a 1-ns equilibration at constant pressure. The temperature was maintained at 25°C and the pressure at 1 atm, using the Martyna-Tobias-Klein thermostat and barostat (time constants of 1 and 2 ps, respectively) (42). To enable a 2.5-fs time step, hydrogen positions were constrained to their equilibrium values using the M-SHAKE algorithm (43). A short-range nonbonded cutoff of 10 Å was used, and long-range electrostatics were calculated using the particle mesh Ewald method (37). Production simulations were carried out at constant pressure using the 512-node ANTON special-purpose supercomputer and the Multigrator integrator (33,44). The temperature was maintained at 25°C using the Nosé-Hoover thermostat and the pressure at 1 atm using the Martyna-Tobias-Klein barostat (time constants of 1 ps) (42,45). To enable a 2.5-fs time step, bonds to hydrogen atoms were constrained to their equilibrium values using the M-SHAKE algorithm (43). The van der Waals and short-range electrostatic interactions were truncated at 10 Å; long-range electrostatic interactions were calculated using the Gaussian split Ewald method (46), and were updated every third time step. Coordinates were saved every 105 ps for analysis.

Analyses of MD simulations were carried out primarily using the AmberTools cpptraj program (47). Secondary structure was assigned using the DSSP method (48), rotational correlation times (τ_c) were calculated using the method of Wong et al. (49), and NMR relaxation rates were calcu-

lated using the iRED method (50). SAXS curves were calculated using two programs: saxs_md, in which water molecules within 5 Å of the protein were explicitly included in the calculation, and CRYSOL 2.8.2, which implicitly accounts for solvent (51,52). Standard errors were estimated using a block averaging method (53).

SAXS

SAXS data were collected for the Mo-WT, Mo-2G, Mo-0G, and Mo-0v constructs at 25°C. Samples were prepared in 25 mM NaAc buffer (pH 5.0), 25 mM NaCl, 5 mM DTT, and 3 mM NaN₃. For the Mo-WT, Mo-2G, and Mo-0G constructs, data were collected using protein concentrations of 5.0, 2.5, and 1.25 mg/mL. No concentration-dependent effects were observed, and the 5.0 mg/mL data are presented here. For the less soluble Mo-0v construct, data was collected at 0.55 mg/mL. All experimental SAXS data were collected at beamline 12-ID-B of the Advanced Photon Source at Argonne National Laboratory (Lemont, IL) using x-rays of energy 14 keV ($\lambda \approx 0.8856$ Å). For each measurement, 30 individual exposures of 1 s each were collected, compared to check for radiation damage, and averaged to yield the final scattering curves. Buffer scattering measurements were performed in an equivalent fashion, using protein-free buffer aliquots from the final purification step and subtracted from the protein scattering data. All data were processed and analyzed using tools from the ATSAS software package including PRIMUS and CRYSOL (52,54,55).

NMR spectroscopy

All spectra were recorded at 25°C on 600, 700, and 800 MHz AVANCE spectrometers (Bruker, Billerica, MA), equipped with 5 mm, triple resonance, three-axis gradient probes, or z-axis gradient cryoprobes. For three-dimensional NMR experiments, the sample contained 300 μM ^{13}C/^{15}N-labeled protein in 25 mM NaAc buffer, pH 5.0, 25 mM NaCl, 5 mM DTT, 3 mM NaN₃, and 5% D₂O. For chemical shift assignments, a series of heteronuclear, multidimensional experiments, routinely used in our laboratory, were recorded (56). Complete ^{1}H, ^{15}N, and ^{13}C backbone resonance assignments were obtained from 3D HNCACB and HN(CO)CACB spectra, using the program CCPNMR (57). Weighted chemical shift differences were calculated using the expression:

$$\Delta\delta = \sqrt{((\Delta\delta_H)^2 + (0.15 \cdot \Delta\delta_N)^2)}.$$

^{15}N R_1 and R_2 relaxation and ^{15}N-{^{1}H} heteronuclear NOE data were collected on a sample of 100 μM ^{15}N-labeled Mo-WT protein in 25 mM NaAc buffer, pH 5.0, 25 mM NaCl, 5 mM DTT, 3 mM NaN₃, and 5% D₂O, using ^{1}H-^{15}N HSQC-based pulse sequences at 600 MHz (58). The R_1 and R_2 experiments employed delays of 0, 100, 150, 200, 300, 500, and 800 ms, and 0, 8, 16, 32, 48, 64, and 80 ms, respectively. Spectra were processed using the software NMRPipe (https://www.ibbr.umd.edu/nmrpipe/install.html). R_1 and R_2 relaxation rates were calculated using single exponential fits, and ^{15}N-{^{1}H} heteronuclear NOE values were calculated using a ratio of experiments recorded with and without ^{1}H saturation (59). Rotational correlation times (τ_c) were calculated using the program relax (60,61). Overlapped resonances and those exhibiting heteronuclear NOE values below 0.7 were omitted from the calculation. 95% confidence intervals of τ_c values were estimated by selecting 1000 subsamples, each with a randomly selected 75% set of rates, calculating τ_c for each.

PRE data were recorded using the single-cysteine variant of Mo-WT, Mo-SAVC (C15S, C25A, C82V). Spectra were recorded at 25°C using 55 μM ^{15}N-labeled protein in 25 mM NaAc buffer, pH 5.0, 25 mM NaCl, 3 mM NaN₃, and 5% D₂O at 800 MHz. Delays of 0.5, 1.0, 2.0, 4.0, and 8.0 ms were employed. ^{15}N R_2 relaxation rates were calculated using single exponential fits, and ^{1}H$_N$-Γ_2 were extracted from the difference between the paramagnetically and diamagnetically tagged samples. Results were visualized using the program Visual Molecular Dynamics (62).

Structure calculation of Mo-WT

Structure calculation of Mo-WT was carried out using XPLOR-NIH version 2.44 (63), subject to restraints from the experimental SAXS and PRE data. Starting models for the calculations were generated by building the MTSL tag onto C167 for each of the 25 solution NMR conformers deposited in the PDB (30), followed by a round of simulated annealing. During the initial structure generation only XPLOR's molecular geometry terms were applied, and the positions of all atoms in the CVNH and LysM domains were kept fixed whereas those of the interdomain linkers (residues 55–61 and 111–117) were unrestrained. The annealing process involved simulating at 10,025°C for 10 ns before ramping down to 25°C in 100°C intervals using a 0.2-ps simulation at each temperature, followed by a 1000-step energy minimization. From each of the 25 original NMR conformers, 25 interdomain orientations were thereby generated, yielding a total of 625 starting models from which calculations were seeded.

Rotational correlation times used in the backcalculation of ^1H_N-Γ_2 rates from molecular coordinates were fixed at the values calculated for the CVNH and LysM domains from the experimental ^{15}N R_1 and R_2 and ^{15}N-^1H heteronuclear NOE data. SAXS restraints were applied by backcalculating scattering intensity from the molecular coordinates (64). PRE ^1H_N-Γ_2 restraints were derived according to the Solomon-Bloembergen equation and weighted based on the experimental error (65). These restraints were grouped into intradomain restraints within the CVNH domain and interdomain restraints with the LysM domain.

In each production run, an ensemble of structures was subjected to simulated annealing, with the average backcalculated PRE ^1H_N-Γ_2 rates and SAXS intensity restrained to their experimental values (64,65). The ensemble size was chosen, based on a series of test calculations in which progressively larger ensembles were generated and evaluated, using PRE ^1H_N-Γ_2 rate Q-factors for CVNH and LysM domain residues and X^2 values for the SAXS intensity. Ensembles of 1, 8, 16, 24, and 48 members all reproduced CVNH domain ^1H_N-Γ_2 rates and SAXS X^2 satisfactorily (Table S1). Ensembles of 1, 8, and 16 members reproduced LysM domain PRE ^1H_N-Γ_2 rates poorly, whereas ensembles of 24 members yielded good results, with some further improvement noted for 48-member ensembles. However, calculations with 48-member ensembles were found to be time-inefficient under our current conditions, which made us settle on an ensemble size of 24 for the production calculation.

During the production calculations, the CVNH and LysM domain backbone coordinates were fixed, whereas side-chain atoms and all residues in the interdomain linkers were unrestrained. Because residues very close to the MTSL label lack visible amide resonances, ^1H_N-Γ_2 rates could not be measured for all residues in the CVNH domain. To derive acceptable conformations of the solvent exposed MTSL tag, which is surrounded by relatively few restraints, simulated annealing was performed in two stages. In the first stage, XPLOR's molecular geometry terms were applied alongside 67 PRE ^1H_N-Γ_2 restraints between the MTSL label and backbone amide hydrogens of the CVNH domain. The system was equilibrated at 3000°C for 100 ps to allow different interdomain orientations to emerge, compared to those present in the initial structures. Subsequent simulated annealing of the equilibrated system involved cooling down from 3000 to 25°C in 25°C intervals, with 0.2 ps of simulation at each temperature, followed by a 1000-step energy minimization, after which the coordinates of C167 and the attached MTSL tag were fixed. During the second stage of simulated annealing, XPLOR's molecular geometry terms were applied alongside 39 PRE ^1H_N-Γ_2 restraints between the MTSL tag and residues on the LysM domain, as well as a SAXS intensity restraint on the overall system. The system was reequilibrated at 3000°C for 100 ps, followed by simulated annealing and energy minimization as described above. Overall, 625 ensembles of 24 structures each were calculated, yielding a total of 15,000 structures. To quantify the influence of the interdomain PRE and SAXS restraints on the resulting structural ensemble, three control calculations were carried out, omitting 1) interdomain PRE restraints, 2) SAXS restraints, or 3) both interdomain PRE and SAXS restraints from the second stage of simulated annealing.

RESULTS AND DISCUSSION

The structure and dynamics of flexibly linked multidomain proteins are particularly challenging to characterize by experimental methods such as x-ray crystallography. Instead, they are well suited as targets for integrated methods that combine experimental data with computer simulations. Here, we apply such methodology to the two-domain protein MoCVNH3, whose domain-insertion topology sets it apart from linearly connected multidomain proteins. Our prior structural work by solution NMR and x-ray crystallography (25,30) demonstrated that the two domains have no fixed interdomain orientation and did not provide details about the nature or distribution of sampled orientations. Here, we used a combination of SAXS, NMR, and MD simulation to characterize the interdomain orientations of MoCVNH3 as well as the influence of the interdomain linker lengths on the overall structure and dynamics of the protein. Our results provide extensive data for evaluating the accuracy of simulation models, utilizing this unique system for validating the structure and dynamics of the individual domains in tandem with the overall interdomain dynamics.

Accessible interdomain orientations

In our previous work, we investigated the wild-type MoCVNH3 construct, called Mo-WT throughout this article, and several reduced linker-length constructs. Here, we study two of these constructs, Mo-0G and Mo-0v, as well as a newly created construct, Mo-2G, in detail. In the Mo-0G and Mo-2G constructs, each interdomain linker is shortened to zero and two glycine residues, respectively (Fig. 1 B). In the Mo-0v construct, three additional residues adjacent to the second linker are replaced by a single glycine. Mo-0v was successfully crystallized, which proved impossible for the Mo-WT and Mo-0G constructs, despite considerable effort (25). To eliminate potential confounding factors on the global structure of this two-domain system, we also deleted the four- or six-residue N-terminal cloning artifacts that were present in the proteins studied previously (25,30), preserving the native amino acid sequence. Consistent with our prior work, a comparison of the ^1H-^{15}N HSQC NMR spectra of our newly created, native, Mo-2G, Mo-0G, and Mo-0v constructs with native Mo-WT revealed only very small chemical shift changes for residues distant from the linkers, demonstrating that the structures of the individual domains are retained in all constructs (Fig. 1).

To characterize the influence of the lengths of the interdomain linkers on the global structure and dynamics of MoCVNH3, we carried out MD simulations of the WT and each of the three reduced linker-length variants (Mo-2G, Mo-0G, and Mo-0v), using two current force-field/water-model combinations: 1) the AMBER ff15ipq force field (from this point onwards, we will refer to this force field as "ff15ipq") with the SPC/E_b water model (18,66), and 2)

the CHARMM22* force field with the TIP4P-D water model (21,67). Both the ff15ipq and CHARMM22* force fields were parameterized to address the issue of overstabilizing salt bridges—a limitation of many other contemporary force fields (68). The ff15ipq force field is a complete reparameterization, which includes implicitly polarized atomic charges, new angle parameters, new atomic radii for polar hydrogens, and a greatly expanded set of torsion terms. This force field was developed for use with the SPC/E$_b$ water model, which more accurately reproduces the experimental rotational diffusion of globular proteins, compared to earlier water models (66). The CHARMM22* force field is a modification of CHARMM22 that includes adjustments to the atomic charges of arginine, aspartate, and glutamate side chains and updates to the backbone torsion parameters. This force field was paired with the TIP4P-D water model, which reduces the oversampling of compact conformations of nonglobular proteins by earlier water models. In total, eight MD simulations were each run for 5–10 μs, with an aggregate simulation time of >60 μs (Table 1). These simulations provide an excellent opportunity to validate the accuracy with which the two force field/water model combinations reproduce 1) the rotational diffusion of a two-domain protein, expanding on efforts involving single-domain proteins (18,49,66) and 2) the compactness of a flexibly linked, globular two-domain protein, expanding on efforts involving nonglobular proteins (21).

Whereas our earlier solution NMR results lacked detailed information about the interdomain orientations of the proteins, MD simulations afford the opportunity to efficiently generate large ensembles of orientations. As shown in Fig. 2, our simulations reveal that ff15ipq/SPC/E$_b$ and CHARMM22*/TIP4P-D yield very different ensembles of interdomain orientations: whereas the ff15ipq/SPC/E$_b$ simulations remained in a single interdomain orientation, the CHARMM22*/TIP4P-D simulations of Mo-WT and Mo-2G sampled a range of interdomain orientations with more extended conformations. Smaller differences between the two force field/water model combinations are observed for the Mo-0G construct. In contrast, both simulations of Mo-0v yielded nearly identical, more restricted sets of accessible interdomain orientations.

TABLE 1   MD Simulations of MoCVNH3 Constructs

				SAXS	
Construct	Force Field	Water Model	Duration (μs)	saxs_md X^2	CRYSOL X^2
WT	AMBER ff15ipq	SPC/E$_b$	10.0	67.1	32.8
	CHARMM22*	TIP4P-D	7.3	9.1	13.2
2G	AMBER ff15ipq	SPC/E$_b$	10.0	28.9	13.0
	CHARMM22*	TIP4P-D	5.5	9.9	11.0
0G	AMBER ff15ipq	SPC/E$_b$	10.0	21.9	10.9
	CHARMM22*	TIP4P-D	5.5	11.7	14.7
XRAY	AMBER ff15ipq	SPC/E$_b$	10.0	0.1	0.5
	CHARMM22*	TIP4P-D	5.0	0.1	0.6

AMBER ff15ipq, SPC/E$_b$ **CHARMM22*, TIP4P-D**

Mo-WT

Mo-2G

Mo-0G

Mo-0v
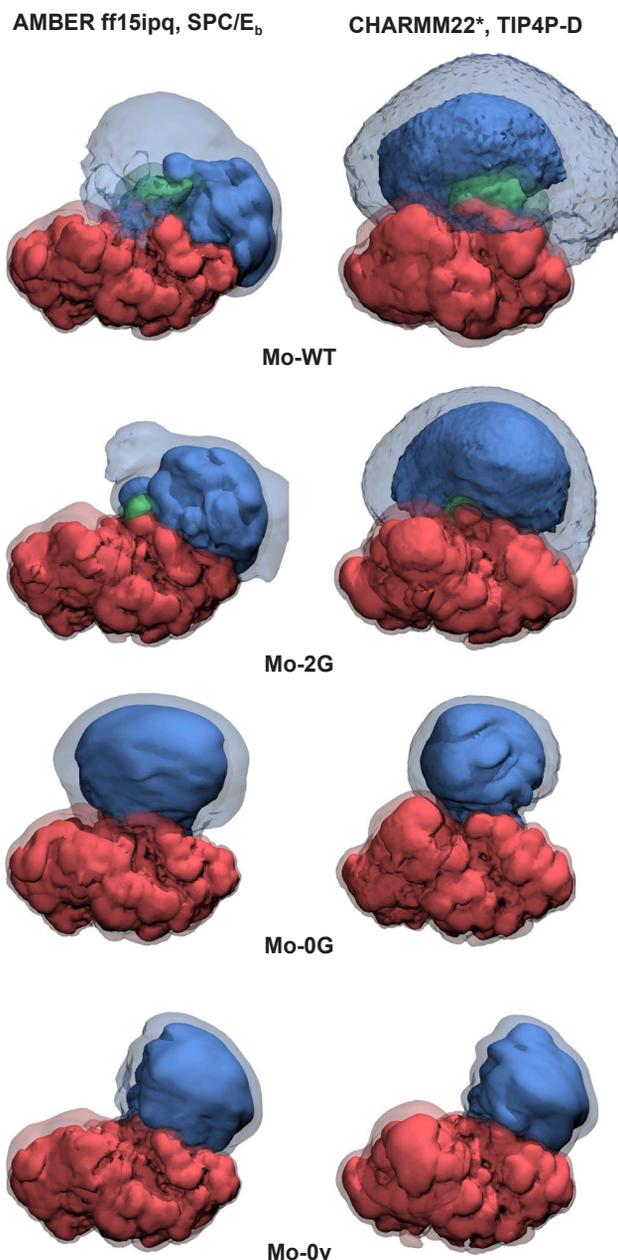
FIGURE 2   Probability distributions of interdomain orientations sampled in MD simulations for four MoCVNH3 constructs. The CVNH domain is shown in red, the LysM domain in blue, and the interdomain linkers in green. Trajectories were best fit to the CVNH domain coordinates and the simulation cell was divided into 1-Å^3 bins; solid contours represent bins occupied by a heavy atom for at least 1% of the simulation, whereas transparent contours represent bins occupied for at least 0.1% of the simulation. To see this figure in color, go online.

Structural characterization of the CVNH and LysM domains

To assess the ability of the two force-field/water model combinations to maintain the integrity of the individual CVNH and LysM domain structures, we monitored the backbone root mean standard deviations (RMSD) from

the initial structures over the course of the simulations. All simulations yielded average RMSD values <3.0 Å for both domains. However, inspection of the distributions of RMSD values (Fig. 3) reveals that significantly larger variations in the RMSD occur over the course of each simulation. In particular, in all of the simulations that were run, and most pronounced for the simulations run with CHARMM22*/TIP4P-D, the RMSD distribution for the CVNH domain is bimodal, sampling two minima with small and large deviations from the starting structure, respectively. In general, the RMSD values increased as the simulations progressed (Fig. S3), e.g., for Mo-WT and Mo-0G, the RMSD of the CVNH domain remained <3 Å up until 4 μs, after which the deviations increased to ~4 Å. These results underscore the importance of reaching the multimicrosecond timescale when simulating complex systems such as the MoCVNH3 protein. Although the ff15ipq/SPC/E$_b$ simulations also show a bimodal distribution for the backbone RMSD of the CVNH domain, the RMSD values were consistently <3.0 Å, with the trajectories for Mo-WT, Mo-2G, and Mo-0v settling at lower

RMSD values, with few excursions to higher RMSD values.

For the LysM domain, lower and more tightly distributed RMSD values were observed for the Mo-WT, Mo-2G, and Mo-0G for both field/water model combinations, whereas RMSD values were more variable for Mo-0v. The overall higher RMSDs observed for the CHARMM22*/TIP4P-D simulations are consistent with observations made during the development of the TIP4P-D water model, which suggested that the implemented increase in the protein-water interaction strength may destabilize the folded states of proteins (21). Compared to SPC/E$_b$ and most other fixed-charge water models, TIP4P-D increases the relative strength of protein-water interactions versus protein-protein interactions to reduce the oversampling of compact states of nonglobular proteins that has resulted from the use of other water models. Our simulations of MoCVNH3 with CHARMM22*/TIP4P-D suggest that this water model may also have the unintentional effect of reducing the sampling of compact (i.e., folded) states of folded proteins, resulting in both a more expanded overall system (Fig. 2) and some changes in the native folded structure. Indeed, the increase in RMSD over the course of the CHARMM22*/TIP4P-D simulations suggests that further loss of native structure could ensue, if the simulations were extended for even longer times.

Some curious simulation artifacts were observed for the CVNH domain at the residue level (see Figs. S4–S20, and Fig. 1 D for the position of highlighted residues within the domain structures). In all eight of our simulations, the N-terminal residues G1 through N6 of the CVNH domain sampled diverse conformations, whereas from F7 onwards, distributions around a single conformation consistent with the experimental structures dominated along the chain. However, in some of the simulations, a few deviations from the experimental structures were observed as far into the sequence as S10. Also, for the first half of the CVNH domain, the loop spanning residues L16 through A19 sampled multiple conformations in all simulations, whereas N45 and D46 mostly retained the experimentally determined conformations in simulations run with ff15ipq/SPC/E$_b$ but not with CHARMM22*/TIP4P-D. Within the second half of the CVNH domain, the loop comprising S143 and G144 did not maintain the native conformation in any of our simulations. The greatest differences in CVNH domain coordinates were observed in the simulation of Mo-WT with CHARMM22*/TIP4P-D in which a conformational change occurred in the β-strand that connects the end of the second interdomain linker to the second half of the CVNH domain, although the antiparallel β-sheet (residues 125–151) that makes up most of the second half of CVNH stayed intact (Figs. S5 and S14).

In all eight simulations, the overall structure of the LysM domain was retained more faithfully than that of the CVNH domain. Interestingly, the experimental conformations of
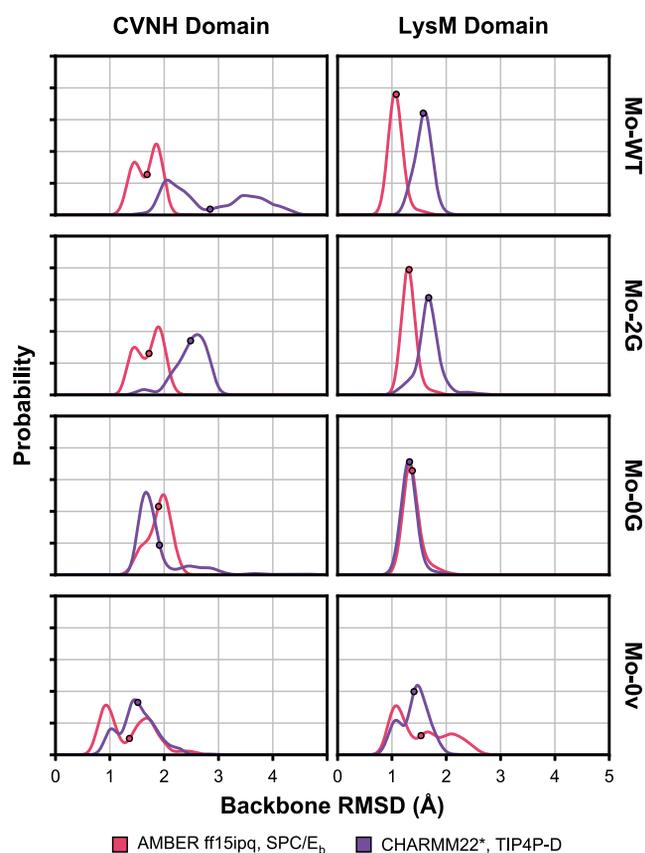
FIGURE 3 Fluctuations in the CVNH and LysM domain coordinates in MoCVNH3 constructs over the course of MD simulations using the AMBER ff15ipq force-field/SPC/E$_b$ water model (*red*), and the CHARMM22* force-field/TIP4P-D water model (*purple*), as represented by the distribution of backbone RMSD relative to their initial structures. Average backbone RMSD values are indicated by circles. To see this figure in color, go online.

loop residues F80 and D81 in the LysM domain were better retained in simulations with CHARMM22*/TIP4P-D than those with ff15ipq/SPC/E$_b$. This better retention is most likely due to D81's native left-handed α-helical conformation, which is disfavored by ff15ipq (18). Interestingly, in seven of the eight simulations, I99 primarily sampled a rare conformation centered at $\Phi \approx 60°$, $\Psi \approx 150°$ whereas this residue exhibits a PPII conformation in the experimental structures of Mo-WT and Mo-0v. The occurrence of this conformation, which is essentially the inverse of the rare plateau conformation, may simply be a consequence of the limited functional form of the ff15ipq and CHARMM22* force fields.

To provide insight into why it was possible to crystallize Mo-0v, but not Mo-0G (25), we compared the interdomain linker regions of Mo-0G and Mo-0v. In Mo-0G, the first linker between the CVNH and LysM domains, comprising consecutive residues V53, S54, T62, and A63, exhibited diverse conformations in the AMBER ff15ipq/SPC/E$_b$ simulation, whereas in the CHARMM22*/TIP4P-D simulation, these residues occupied a single conformation. However, in the second interdomain junction, involving residues P108, T109, K110, G118, and N119, both force fields resulted in diverse conformations for Mo-0G. In contrast, for Mo-0v, in which P108, T109, and K110 have been replaced by a single G117, simulations with both force fields stably retained the conformations that were seen in the crystal structure for the interdomain junctions. This result suggests that the changes that were introduced into the second linker of Mo-0v resulted in a conformationally more restricted system and may therefore be responsible for its successful crystallization.

Compactness of the two-domain systems

To quantify the influence of the interdomain linker-lengths on the overall structure of the two-domain MoCVNH3 system, SAXS curves were measured for Mo-WT and the three reduced linker-length constructs (Fig. 4). During sample preparation of the Mo-0v protein we noted that this protein was less soluble than the other three constructs, requiring data collection at a lower concentration (0.55 mg/mL compared to 5.0 mg/mL), therefore resulting in noisier data. The lower solubility of Mo-0v may relate to its reduced net charge, as this construct contains one fewer lysine than the larger constructs (Fig. 1 B). Earlier work on Mo-0v with the construct, which contained a four-residue N-terminal cloning artifact, including a histidine (positively charged at the experimental pH), did not exhibit reduced solubility (25), illustrating that small amino acid changes can greatly influence a protein's behavior. In this study, we removed the nonnative N-terminal amino acids from our protein constructs to eliminate their potential contributions to interdomain interactions. Although the removal of these amino acids had the unfortunate consequence of lowering the solubility of Mo-0v, the data obtained with
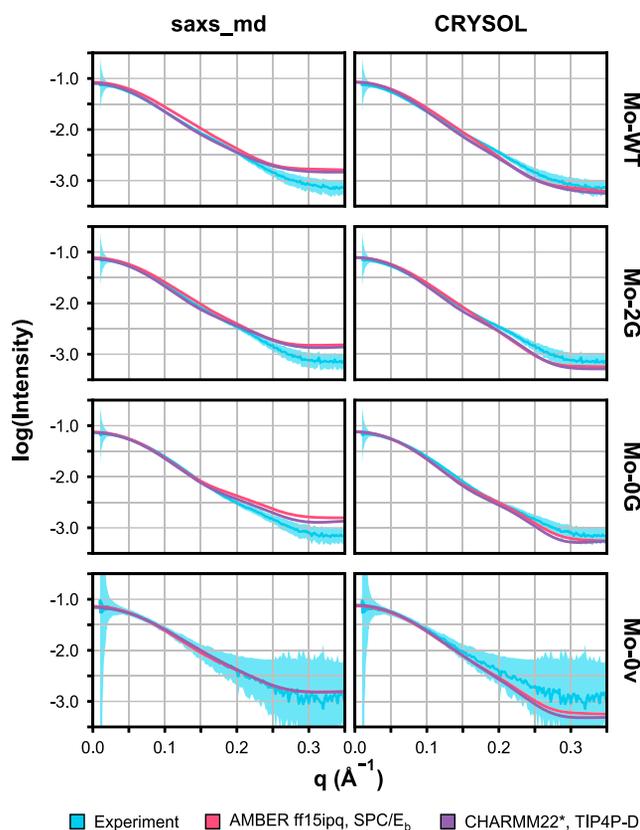
FIGURE 4 SAXS intensity of MoCVNH3 constructs measured by experiment (*cyan*) and backcalculated from MD simulations (*red, purple*). The left panel depicts simulated curves backcalculated using AmberTools' saxs_md program (51), which includes explicit water molecules for calculating the scattering, and the right panel shows simulated curves backcalculated with ATSAS' CRYSOL program (52), which represents solvent implicitly. The scale of the *y* axis is arbitrary; shaded regions represent 95% confidence intervals for the experimental and simulated values; the larger uncertainty in the experimental scattering for the Mo-0v construct arises from the need to collect the data at lower concentration due to the protein's lower solubility. To see this figure in color, go online.

all the proteins are of sufficient quality for a valid comparison as described below.

The experimental SAXS data offer the opportunity to validate our MD simulations' modeling of the overall structural of the two-domain system. To this end, we backcalculated SAXS curves from the simulation coordinates, using two different methods: 1) AmberTools' saxs_md program (51), which explicitly includes the coordinates of surrounding water molecules in the calculation, and 2) ATSAS' CRYSOL program (52), which implicitly accounts for the scattering of surrounding water molecules. As shown in Fig. 4, for *q* values below 0.2 Å^{-1}, the two methods yield results that are broadly similar to one another and to experiment. Beyond 0.2 Å^{-1}, the saxs_md program consistently yields higher scattering intensity than observed experimentally, whereas CRYSOL yields lower scattering intensity. Because we were interested in capturing differences between the four constructs, we calculated the differences in

the corresponding scattering intensities (Fig. S21). The small, systematic difference between the two calculation methods is of little consequence, because for $q > 0.2$ Å^{-1} the curves for the MoCVNH3 constructs are indistinguishable. For $q < 0.2$ Å^{-1}, the two methods of backcalculation yield similar results, suggesting that, for the region of q measured here, the computationally more expensive explicit-solvent saxs_md calculation does not provide a tangible benefit over the less demanding CRYSOL calculation. Qualitatively, the CHARMM22*/TIP4P-D simulations reproduce the experimental trends more accurately than ff15ipq/SPC/E$_b$. Overall, the above results suggest that the global structure of the MoCVNH3 two-domain system is more accurately represented by using CHARMM22*/TIP4P-D than ff15ipq/SPC/E$_b$.

A key structural parameter that can be calculated from the measured SAXS curves is the radius of gyration (R_g), which reflects the compactness of the protein. Because all the MoCVNH3 constructs are similar in overall mass, R_g provides a means by which the effect of the different linker lengths on the population of extended versus collapsed conformations in the two-domain system can be assessed. As shown in Fig. 5, the R_g for Mo-WT, Mo-2G, and Mo-0G,
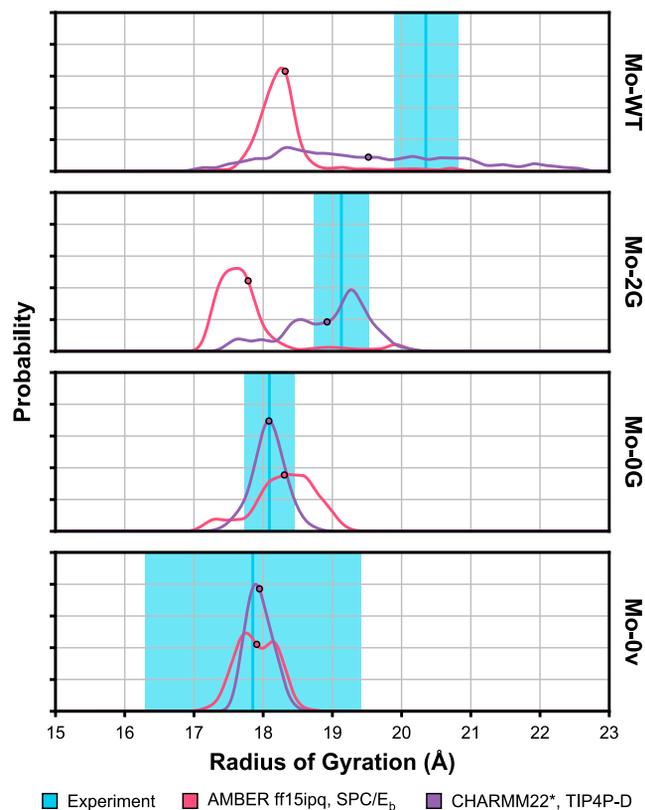
calculated from the experimental SAXS curves, exhibit a clear and intuitively expected decrease of R_g, with reduced interdomain linker length. Although the data for Mo-0v are too noisy to confidently differentiate its R_g from that of Mo-0G and Mo-2G, the R_g of Mo-0v is statistically distinguishable from that of Mo-WT, and the center of the confidence interval lies just below that of Mo-0G, consistent with its slightly shorter second interdomain linker.

R_g can be calculated directly from the MD simulation coordinates, providing further validation of the accuracy of the constructs' global simulated ensembles. The distributions of R_g, sampled over the course of each simulation (Fig. 5), are consistent with the observations about the backcalculated SAXS curves: the CHARMM22*/TIP4P-D simulations more accurately reproduce the experimental R_g value, compared to the ff15ipq/SPC/E$_b$ simulations. Both the ff15ipq/SPC/E$_b$ and CHARMM22*/TIP4P-D trajectories yielded R_g values of ~18 Å for Mo-0v and Mo-0G, in excellent agreement with experiment. The simulations of Mo-2G and Mo-WT, run with ff15ipq/SPC/E$_b$, yielded R_g values similar to those of Mo-0G and Mo-0v, whereas those run with CHARMM22*/TIP4P-D resulted in larger R_g values, in much better agreement with experiment. However, the R_g value of Mo-WT obtained with CHARMM22*/TIP4P-D is still ~1 Å below the experimental value and fluctuated significantly over the course of the simulation (Fig. S22). The observed broad distribution suggests that longer simulations than performed here (7.3 μs) may be necessary to achieve convergence for Mo-WT. Both experiment and simulation yielded R_g values of ~18 Å for Mo-0G and Mo-0v; our simulations of Mo-WT sampled compact conformations, when the domains were in contact, as well as conformations with R_g of up to 23 Å, when the domains were not in contact. The experimental data for Mo-WT shows an R_g value of 20.4 Å, which is 12% larger than those of Mo-0G and Mo-0v as a result of more frequent sampling of extended conformations.

Dynamical properties of Mo-WT

The global structural information obtained from SAXS can be complemented with single-residue and single-domain dynamics information, accessible by NMR relaxation approaches. In particular, the ratio between the ^{15}N transverse (R_2) and longitudinal (R_1) relaxation of the backbone amide resonances provides a measure of the system's rotational diffusion in solution, the isotropic rotational correlation time, τ_c. NMR relaxation data were collected for the Mo-WT construct, and the two domains exhibited characteristic R_1/R_2 ratios of ~9 and ~5 for the CVNH and LysM domains, respectively (Fig. 6), corresponding to τ_c values of 8.7 and 6.7 ns. This difference in correlation time confirms that both domains tumble essentially independently in solution.

FIGURE 5 Radius of gyration (R_g) of MoCVNH3 constructs calculated from experimental SAXS intensity (*cyan*) and directly from conformations sampled in MD simulations (*red, purple*). Cyan shaded regions represent 95% confidence intervals on the experimental values. Average R_g values are indicated by circles. To see this figure in color, go online.

CVNH Domain   **LysM Domain**

FIGURE 6 Rotational diffusion of CVNH and LysM domains of Mo-WT calculated from experimental NMR relaxation (*cyan*) and MD simulation (*red, purple*). Top, distribution of R_1/R_2 relaxation ratios for residues in the CVNH (*left*) and LysM domain (*right*). Bottom, distribution of calculated rotational correlation times (τ_c) for the CVNH (*left*) and LysM domains (*right*). Cyan shaded regions represent 95% confidence intervals of the experimental values. Average values are indicated by circles. To see this figure in color, go online.

To validate the accuracy of the two MD simulations for Mo-WT, ^{15}N R_1 and R_2 relaxation rates were backcalculated from the motions of the backbone amide N-H vectors over the course of the simulations. We previously noted that the ff15ipq/SPC/E_b force-field/water model combination yielded accurate rotational diffusion times for single-domain globular proteins, whereas the CHARMM22*/TIP4P-D combination yielded less accurate results (18). In contrast to the well-defined distributions of the experimental R_2/R_1 ratios, the R_2/R_1 ratios backcalculated from the simulations exhibited much greater variability for individual residues within each domain (Fig. S23). Inspection of the rotational correlation times extracted from the simulation of Mo-WT with ff15ipq/SPC/E_b resulted in similar τ_c values of 9–10 ns for both domains. This implies that in the ff15ipq/SPC/E_b simulations the dynamics of the two domains are too tightly coupled and is consistent with our observation that the ff15ipq/SPC/E_b conformational ensemble is too compact (Fig. 5), adopting only a single interdomain orientation (Fig. 1). The CHARMM22*/TIP4P-D simulation yielded τ_c values of ~11 ns and 7 ns, capturing the difference in rotational diffusion between the domains. However, the τ_c value of the CVNH domain is somewhat higher than the experimentally measured one, consistent with our prior results for single-domain proteins using this force-field/water model combination (18).

Tracking τ_c over the course of the simulation reveals that for the first microsecond of the ff15ipq/SPC/E_b simulation the τ_c of the LysM domain was relatively accurate, but

became worse at ~700 ns when the two domains collapsed onto each other and remained in a single interdomain orientation for the remainder of the simulation (Fig. S24). It therefore appears that the ff15ipq/SPC/E_b combination provided accurate rotational diffusion for the conformations that were sampled, but that the distribution of sampled conformations was inaccurate.

Preferred interdomain orientations of Mo-WT

To determine the distribution of interdomain orientations within the Mo-WT protein, PRE measurements were carried out with a single-cysteine variant of Mo-WT. PREs represent the increased relaxation of nearby nuclear spins around a paramagnetic moiety, resulting in broadening of the associated resonances. The large magnetic moment of the unpaired electron causes a large effect, and PREs can be observed over distances up to 35 Å. The PRE effect scales as $1/r^6$, with r the distance between the unpaired electron in the paramagnetic center and the affected nucleus. Further, in systems that exchange rapidly between different conformations, the measured PREs are the population-weighted averages of the PREs for all sampled conformations, allowing even transient, low-population contacts to be captured (69). Finally, PREs can be backcalculated from known structures, permitting integration into structure calculations and validation of candidate models (65), rendering the PRE approach the ideal methodology for the characterization of distributions.

To probe whether contacts between the two domains of Mo-WT can be captured by PREs, we attached the paramagnetic MTSL tag to a cysteine residue in the CVNH domain and measured PREs on residues in the LysM domain. Because Mo-WT contains four cysteines, none of which are involved in disulphide bonds in the native structure (30), it was necessary to remove all but one cysteine for single site spin-labeling. This was achieved by introducing C15S and C25A mutations into the CVNH domain and the C82V mutation into the LysM domain, leaving C167 at the C terminus exclusively available for attachment of the paramagnetic tag (Figs. S1 and S25). To verify that attachment of the MTSL tag did not affect the structure of the protein, we recorded the ^{1}H-^{15}N HSQC spectrum of Mo-SAVC tagged with the diamagnetic analog of MTSL at C167 (Fig. 7 A). Compared to the spectrum of untagged Mo-SAVC, only very small chemical shift changes were noted for amino acids close to the C167 attachment site, but not elsewhere (Fig. 7, B and D). Interestingly, several of the affected resonances exhibited doubling, suggesting that attachment of the tag results in two slightly different conformations in its vicinity. In the paramagnetically tagged species, however, the equivalent resonances were broadened beyond detection and were therefore not included in the analysis.

PREs were quantified by comparing resonance intensities in the spectra of paramagnetically and diamagnetically
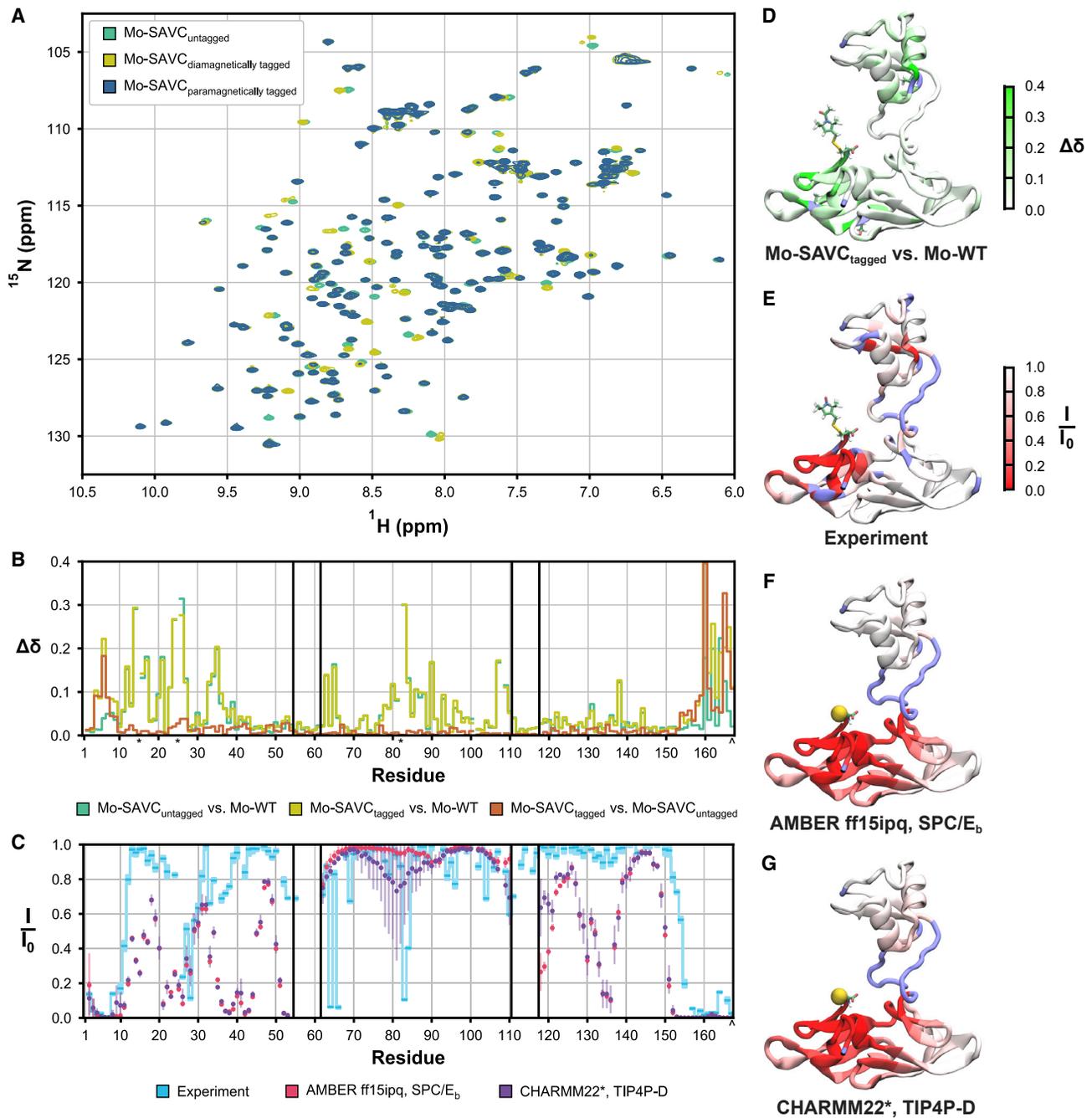
FIGURE 7 PRE data for Mo-WT. (A) Superposition of the ^{1}H-^{15}N HSQC NMR spectra of untagged Mo-SAVC (*green*) and Mo-SAVC with paramagnetic (*teal*) or diamagnetic (*gold*) tags attached to C167. (B) Chemical shift differences between Mo-WT and Mo-SAVC, and between tagged and untagged Mo-SAVC. The positions of the changed cysteines C15S, C25A, and C82S are marked with asterisks, and C167 to which the tags are attached is marked with a caret. (C) Ratio of peak intensities for the paramagnetically and diamagnetically tagged proteins, measured experimentally (*cyan*) and calculated from MD simulations (*red*, *purple*). Cyan lines represent 95% confidence interval of the experimental ratios. Because the MTSL tag was not present in the MD simulations, simulated intensities were calculated based on the distances between each backbone amide hydrogen and Sγ of C167. (D) Chemical shift differences between Mo-WT and diamagnetically tagged Mo-SAVC, mapped onto the structure of MoCVNH3, using a white to green gradient from low to high. Residues for which chemical shift differences are not available are shown in blue. The substituted amino acids C15, C25, and C82, as well as C167 with the tag are shown in licorice representation. (E) Experimental and backcalculated (F) AMBER ff15ipq/SPC/E$_b$ and (G) CHARMM22*/TIP4P-D ratios of peak intensities mapped onto the experimental NMR structure, using a red to white gradient from low to high. Residues for which intensity ratios are not available are shown in blue. C167 with the tag is shown in licorice representation. To see this figure in color, go online.

tagged samples (Fig. 7 C). For the CVNH domain, the experimentally determined intensity ratios exhibited a clear dependence on the distance between each residue and the MTSL tag (Fig. 7 E). For residues in the LysM domain, the amide resonances of T64, T66, D83, and F84 show strikingly lower intensity ratios than resonances of other residues in this domain. Given that T64 and T66 lie on one side of the LysM domain and D83 and F84 lie on the other side of the domain, it is impossible for all four of these residues to be simultaneously close to the MTSL tag. Thus, the Mo-WT protein must be exchanging between two different orientations with different sides of the LysM domain transiently approaching the MTSL tag.

To determine whether our MD simulations of Mo-WT captured these two different orientations, PRE intensity ratios were backcalculated from the sampled conformations. Because the MTSL tag was not included in the simulation model, the $S\gamma$ atom of C167 was used as a proxy. Within the CVNH domain, agreement between the simulated and experimental intensity ratios is observed only for residues either close enough to C167 to yield ratios near 0, or distant enough from C167 to yield ratios near 1 (Fig. 7, C, F, and G). The intensity ratios of residues at intermediate distances from C167 are poorly reproduced. This indicates that the approximation of the paramagnetic group's location in our simulations is insufficiently accurate for residues at intermediate distances, where small inaccuracies in distance and orientation have large effects on the backcalculated ratios. However, the imprecision of the paramagnetic tag's location has a smaller effect on ratios calculated for residues in the LysM domain, which are on average further away from C167. For the first contact site on the LysM domain, containing T64 and T66, neither simulation reproduces the experimental ratios, and none of these two amino acids gets close to C167. However, for the second contact site, including D83 and F84, the simulation with CHARMM22*/TIP4P-D results in smaller intensity ratios for these and several nearby residues centered around D81. Although it is possible that the simulated conformations in which D81 approaches C167 are representative of the experimental conformations that are responsible for the low intensity ratios of D83 and F84, the large standard errors of the backcalculated interdomain PREs suggest

that even longer simulations may be required to obtain reliable distributions of interdomain orientations.

To quantitatively link the two contact sites on the LysM domain to the global structure of Mo-WT, we measured ^1H_N-Γ_2 PRE rates, representing the R_2 relaxation induced by the paramagnetic tag, and incorporated them alongside our SAXS data as restraints in the calculation of a structural ensemble using XPLOR-NIH (63,70). The calculation was seeded from the individual domain structures of the CVNH and LysM domains of the NMR structure (26). A total of 15,000 structures were calculated, and the results yielded good agreement with the experimental restraints. To dissect the influence of the individual interdomain PRE and SAXS restraint terms on the distribution of interdomain orientations in the ensemble, we calculated three control ensembles, in which the interdomain orientations were restrained by 1) only interdomain PRE restraints, 2) only the SAXS restraint, or 3) neither. The Q-factors and X^2 values for all three ensembles are summarized in Table 2, whereas the distributions of interdomain orientations are depicted in Fig. S26, and the backcalculated SAXS curves in Fig. S27.

The average R_g of the calculated structural ensemble is 20.0 Å, which is within the 95% confidence interval of the value calculated from the experimental SAXS data using Guinier analysis (20.4 Å). The ensemble generated using the full complement of interdomain PRE and SAXS restraints yielded Q-factors for the PRE ^1H_N-Γ_2 rates of 0.41 for CVNH domain residues and 0.56 for LysM domain residues, and a X^2 of 0.04 for the SAXS intensity. The extremely low SAXS X^2 value of 0.04 may imply that the calculation was restrained too tightly to reproduce the experimental SAXS intensity, given that a correctly calibrated SAXS restraint is expected to yield a X^2 value of ~1, which implies that on average the deviation between the calculated model and experimental target is equal to the standard error in the target. However, the ensemble that included neither the interdomain PRE- nor the SAXS restraints also yielded a surprisingly low SAXS X^2 value of 0.13, already below the ideal target. This suggests that the conformational space accessible to the domains is relatively well modeled by including a repulsion term, preventing spatial domain overlap, and geometric terms, which represent the linker lengths. In essence, the overall size of

TABLE 2 Structure Calculations of Mo-WT

	CVNH Domain PRE		LysM Domain PRE		SAXS		
Restraints Applied	Q-Factor	R^2	Q-Factor	R^2	X^2	$P_{T64 \text{ and } T66 \leq 13 \text{ Å}}$[a] (%)	$P_{D83 \text{ and } F84 \leq 13 \text{ Å}}$[b] (%)
PRE[c] and SAXS	0.41	0.94	0.56	0.67	0.04	4.1	6.9
PRE[c] only	0.41	0.94	0.52	0.70	0.95	4.1	7.0
SAXS only	0.41	0.94	0.67	0.53	0.04	0.3	1.5
Unrestrained	0.41	0.94	0.76	0.47	0.13	0.2	1.3

[a]Percentage of structures within the ensemble, in which the backbone amide hydrogens of both T64 and T66 are within 13 Å of MTSL's nitroxide group.
[b]Percentage of structures within the ensemble, in which the backbone amide hydrogens of both D83 and F84 are within 13 Å of MTSL's nitroxide group.
[c]Refers to interdomain PRE restraints only.

the system is dominated by geometric restrictions. However, an unrestrained ensemble does not satisfactorily reproduce the LysM domain $^1H^N$-Γ_2 values, resulting in a Q-factor of 0.76, and does not capture the two contact sites on the LysM domain. Therefore, although no fixed interdomain orientation of the two domains is present, measurable differences in populations of the accessible orientations in the ensembles can be discerned.

Within the calculated, experimental data-derived ensemble, the probability distribution of interdomain orientations shows that a wide range of orientations is sampled by the system (Fig. 8 A). Comparison of this experiment-driven ensemble with the two ensembles derived from MD simulations alone (Fig. 2) revealed that CHARMM22*/TIP4P-D reproduced the sampled interdomain orientations in the experiment-driven ensemble more accurately than ff15ipq/SPC/E_b. This
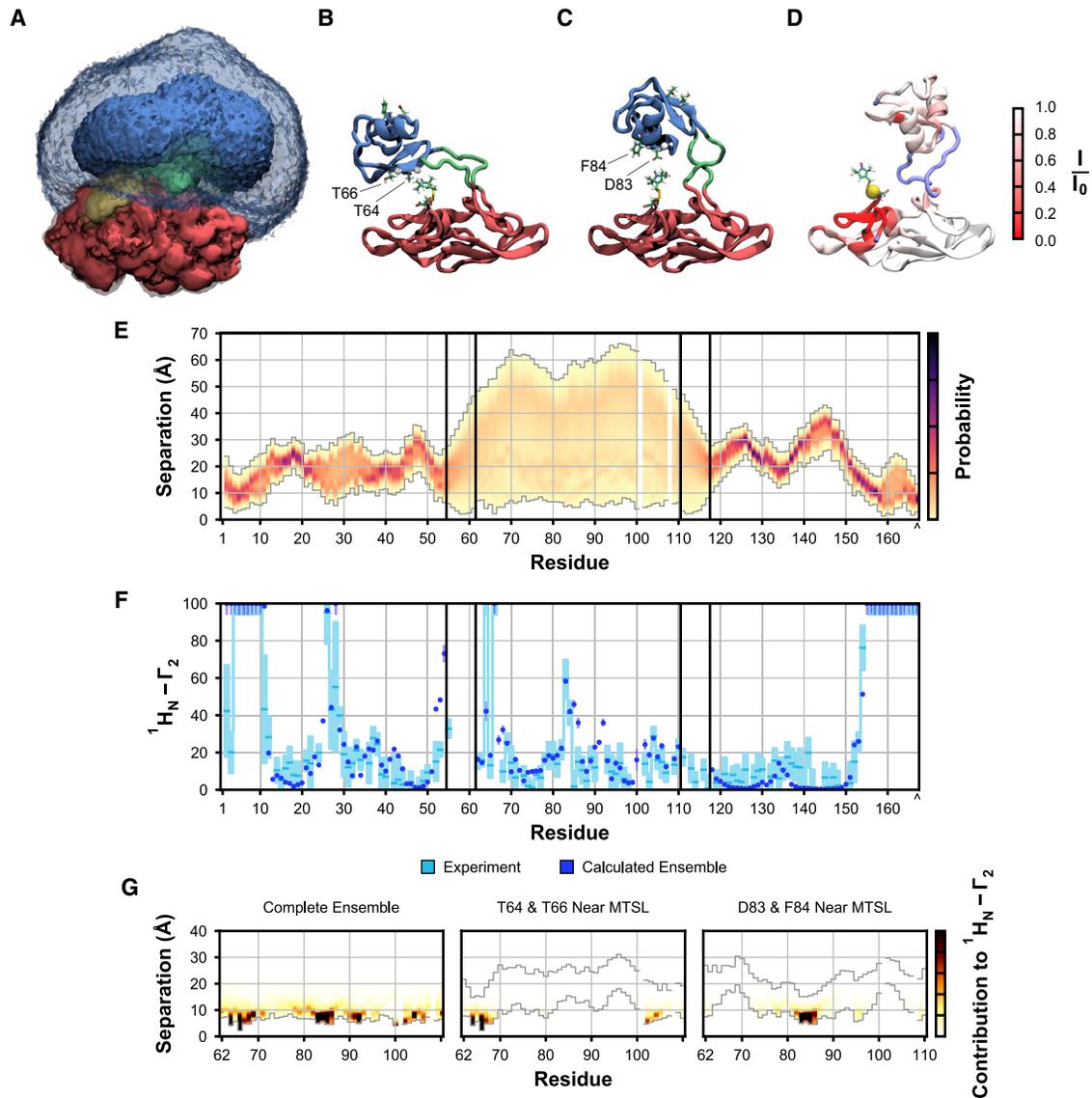
FIGURE 8 Structural ensemble of Mo-WT based on experimental SAXS and PRE data. (A) Probability distributions of interdomain orientations in the calculated structural ensemble. The CVNH domain is shown in red, the LysM domain in blue, the interdomain linkers in green, and MTSL paramagnetic tag in yellow. Structures were best fit to the CVNH domain coordinates, and the simulation cell was divided into 1-Å^3 bins; solid contours represent bins occupied by a heavy atom for at least 1% of the ensemble, whereas transparent contours represent bins occupied for at least 0.1% of the ensemble. (B) A representative structure, illustrating the approach between the MTSL tag and T64 and T66. (C) A representative structure, illustrating the approach between the MTSL tag and D83 and F84. (D) Peak intensity ratios, backcalculated from the ensemble and mapped onto the experimental NMR structure, using a red to white gradient from low to high. (E) Distributions of interatomic separation between the backbone amide hydrogen of each residue and the nitroxide radical of MTSL. (F) ^1H_N-Γ_2 relaxation measured experimentally (cyan horizontal dash) and backcalculated from the structural ensemble (blue). 95% confidence intervals of the experimental ratios are shown in cyan vertical lines. For clarity, residues whose ^1H_N-Γ_2 exceeded 100 are shown as 100. (G) Contribution of mutually exclusive subsets of conformations to backcalculated ^1H_N-Γ_2. For both the first (T64 and T66) and second contact sites (D83 and F84), nearly all contribution to the backcalculated ^1H_N-Γ_2 comes from a subset of structures, comprising ~5% of the total ensemble. To see this figure in color, go online.

demonstrates that the rebalancing of protein/protein and protein/water nonbonded dispersion interactions in the TIP4P-D water model has benefits for systems beyond disordered peptides and proteins, such as the one investigated here.

The distributions of sampled interatomic distances between each backbone amide hydrogen and the nitroxide group of MTSL (Fig. 8 E) reveal that in most structures, amino acids in the LysM domain are relatively far away from MTSL. However, for distances <25 Å, it is possible to discern subsets of structures for which parts of the LysM domain are closer to the MTSL tag. Indeed, the back-calculated PRE ^1H_N-Γ_2 rates (Fig. 8 F) reveal that the two contact sites on the LysM domain, identified above from the paramagnetic/diamagnetic intensity ratios (Fig. 7, C and E), are captured in the ensemble. The high ^1H_N-Γ_2 of residues D83 and F84 are accurately reproduced, and residues T64 and T66 also yield high ^1H_N-Γ_2. This finding is particularly gratifying, because no restraints on the ^1H_N-Γ_2 of T64 and T66 were applied during the calculation. Although their low intensity ratio indicated that their ^1H_N-Γ_2 rate had to be high, the peak intensities were too low to confidently measure a ^1H_N-Γ_2 rate, which could be converted into a restraint. However, the ^1H_N-Γ_2 restraints of surrounding residues clearly were sufficient to capture this contact site in the calculated ensemble. The subset of structures in which the backbone amide hydrogens of both T64 and T66 are within 13 Å of MTSL's nitroxide group comprises 4.0% of all structures and is responsible for over 80% of the ensemble's calculated ^1H_N-Γ_2 for these two residues (Fig. 8, B and G). The analogous subset for D83 and F84 (Fig. 8, C and G) comprises 6.6% of structures, which are responsible for >90% of the calculated ^1H_N-Γ_2 of these residues. The two subsets are mutually exclusive, and each one does not contribute to the ^1H_N-Γ_2 values of the other set's contact site. This result vividly demonstrates and supports previous findings about the significant influence of low-population conformational substates on measured PRE ^1H_N-Γ_2 rates (69). However, it is important to recognize the limitations of the data: although our results demonstrate the existence of two mutually exclusive low-population subsets of conformations, the accuracy of their relative amounts are limited by the quantity of input data and subject to the configuration of the ensemble calculation. In essence, using a limited ensemble size per calculation places a lower bound on the representable population. To reproduce the large PRE ^1H_N-Γ_2 rates at a given site, at least one structure in the 24-structure ensemble, or ~4%, must have the paramagnetic label close to the site. It is therefore possible that the true populations of the two subsets may be somewhat different than noted here.

CONCLUSIONS

In this work, we have characterized the global structure and dynamics of the flexibly linked domain-insertion protein MoCVNH3, using experimental NMR and SAXS studies in combination with microsecond timescale MD simulations. To evaluate the influence of interdomain linker length on the properties of the system, we studied a series of reduced linker-length constructs, in which the two domains ultimately become locked into a single interdomain orientation. For four tested linker lengths, the global structural properties of the systems were measured by SAXS, and results were compared to MD simulations, testing the ff15ipq/SPC/E$_b$ and CHARMM22*/TIP4P-D force-field/water-model combinations. We found that whereas ff15ipq/SPC/E$_b$ accurately retained the experimental structures of the individual domains, it did not capture the observed increase in the radii of gyration (R_g) for increasing linker lengths, resulting in interdomain orientations that rendered to overall structure too compact. CHARMM22*/TIP4P-D did capture the increase in R_g, but resulted in structural changes in the individual domain structures. Together, our results indicate that none of the tested force-field/water-model combinations is able to simultaneous maintain an accurate intra- and interdomain structure. The interdomain orientations of the wild-type protein were evaluated using PRE measurements, which identified two mutually exclusive contact sites between the CVNH and LysM domains. Ultimately, we used the SAXS and PRE data in combination for calculating an overall structural ensemble. Although the currently available experimental data are likely insufficient to fully define the native conformational ensemble, our results show that measurable differences in population for the accessible orientations can be discerned using PREs, even if no fixed interdomain orientation exists for the two domains in MoCVNH3.

The above findings are valuable in the context of integrative structural biology, where, through combinations of various experimental data and computer models, one aspires to derive a more comprehensive view of structure and dynamics than is accessible from either experiment or computation alone. Naturally, all computational models are subject to the accuracy of the selected molecular mechanical force fields, which, although quite robust, still possess considerable room for improvement. Recent advancements in model development, such as implemented in the Implicitly Polarized Q and Force Balance approaches, reduce the necessary time-consuming efforts in each round of improvement (18,19). Coupled to advances in computer hardware and algorithms, which increasingly enable longer simulations of larger systems (33,71–74), ever more complex systems will become accessible to simulation. We suggest that the joint simulation/experimental study of MoCVNH3 reported here provides a valuable benchmark toward this end, in particular for the characterization of structural and dynamical properties of multidomain proteins.

Overall, the characterization of the structure and dynamics of the two-domain MoCVNH3 protein is, to our

knowledge, the most in-depth biophysical characterization of a domain-insertion protein system and illustrates the value of integrating a synergistic combination of NMR, SAXS, and long timescale atomistic simulations for characterizing structural ensembles of flexibly linked multidomain systems.

SUPPORTING MATERIAL

Twenty-seven figures and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)30059-6.

AUTHOR CONTRIBUTIONS

K.T.D., L.T.C., and A.M.G. conceived the study and planned the experiments. L.M.I.K. provided clones and K.T.D. expressed and purified the proteins. K.T.D. acquired and analyzed the NMR data. M.J.W. collected and analyzed the SAXS data. K.T.D. performed and analyzed the MD simulations and structural ensemble calculations. K.T.D., L.T.C., and A.M.G. wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

1. Levitt, M. 2009. Nature of the protein universe. *Proc. Natl. Acad. Sci. USA.* 106:11079–11084.

2. Ekman, D., Å. K. Björklund, …, A. Elofsson. 2005. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol.* 348:231–243.

3. Bashton, M., and C. Chothia. 2007. The generation of new protein functions by the combination of domains. *Structure.* 15:85–99.

4. Bhaskara, R. M., A. G. de Brevern, and N. Srinivasan. 2013. Understanding the role of domain-domain linkers in the spatial orientation of domains in multi-domain proteins. *J. Biomol. Struct. Dyn.* 31:1467–1480.

5. Papaleo, E., G. Saladino, …, R. Nussinov. 2016. The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.* 116:6391–6423.

6. Hammes, G. G., Y.-C. Chang, and T. G. Oas. 2009. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. USA.* 106:13737–13741.

7. Aroul-Selvam, R., T. Hubbard, and R. Sasidharan. 2004. Domain insertions in protein structures. *J. Mol. Biol.* 338:633–641.

8. van den Bedem, H., and J. S. Fraser. 2015. Integrative, dynamic structural biology at atomic resolution–it's about time. *Nat. Methods.* 12:307–318.

9. Bernadó, P. 2010. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. Biophys. J.* 39:769–780.

10. Madl, T., F. Gabel, and M. Sattler. 2011. NMR and small-angle scattering-based structural analysis of protein complexes in solution. *J. Struct. Biol.* 173:472–482.

11. Perkins, S. J., D. W. Wright, …, J. E. Curtis. 2016. Atomistic modelling of scattering data in the collaborative computational project for small angle scattering (CCP-SAS). *J. Appl. Cryst.* 49:1861–1875.

12. Zhao, D., X. Wang, …, Q. Gong. 2014. Structural investigation of the interaction between the tandem SH3 domains of c-Cbl-associated protein and vinculin. *J. Struct. Biol.* 187:194–205.

13. Kikhney, A. G., and D. I. Svergun. 2015. A practical guide to small angle x-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* 589:2570–2577.

14. Schneidman-Duhovny, D., A. Rossi, …, A. Sali. 2012. A method for integrative structure determination of protein-protein complexes. *Bioinformatics.* 28:3282–3289.

15. Tainer, J. A. 2011. X-ray scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Biophys. J.* 100:38a.

16. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.

17. Maier, J. A., C. Martinez, …, C. Simmerling. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.

18. Debiec, K. T., D. S. Cerutti, …, L. T. Chong. 2016. Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model. *J. Chem. Theory Comput.* 12:3926–3947.

19. Wang, L.-P., K. A. McKiernan, …, V. S. Pande. 2017. Building a more predictive protein force field: A systematic and reproducible route to AMBER-FB15. *J. Phys. Chem. B.* 121:4023–4039.

20. Mittal, J., T. H. Yoo, …, T. M. Truskett. 2013. Structural ensemble of an intrinsically disordered polypeptide. *J. Phys. Chem. B.* 117:118–124.

21. Piana, S., A. G. Donchev, …, D. E. Shaw. 2015. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B.* 119:5113–5123.

22. Shi, Y., Z. Xia, …, P. Ren. 2013. The polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* 9:4046–4063.

23. Lopes, P. E. M., J. Huang, …, A. D. Mackerell, Jr. 2013. Force field for peptides and proteins based on the classical Drude oscillator. *J. Chem. Theory Comput.* 9:5430–5449.

24. Koharudin, L. M. I., W. Furey, and A. M. Gronenborn. 2011. Novel fold and carbohydrate specificity of the potent anti-HIV cyanobacterial lectin from *Oscillatoria agardhii*. *J. Biol. Chem.* 286:1588–1597.

25. Koharudin, L. M. I., K. T. Debiec, and A. M. Gronenborn. 2015. Structural insight into fungal cell wall recognition by a CVNH protein with a single LysM domain. *Structure.* 23:2143–2154.

26. Koharudin, L. M. I., and A. M. Gronenborn. 2011. Structural basis of the anti-HIV activity of the cyanobacterial Oscillatoria Agardhii agglutinin. *Structure.* 19:1170–1181.

27. Martin-Urdiroz, M., M. Oses-Ruiz, ..., N. J. Talbot. 2016. Investigating the biology of plant infection by the rice blast fungus Magnaporthe oryzae. *Fungal Genet. Biol.* 90:61–68.

28. Percudani, R., B. Montanini, and S. Ottonello. 2005. The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins.* 60:670–678.

29. de Jonge, R., and B. P. H. J. Thomma. 2009. Fungal LysM effectors: extinguishers of host immunity? *Trends Microbiol.* 17:151–157.

30. Koharudin, L. M. I., A. R. Viscomi, ..., A. M. Gronenborn. 2011. Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus Magnaporthe oryzae. *Structure.* 19:662–674.

31. Eswar, N., B. Webb, ..., A. Sali. 2006. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics.*, Chapter 5:Unit-5.6. https://doi.org/10.1002/0471250953.bi0506s15.

32. Case, D. A., J. T. Berryman, ..., P. A. Kollman. 2015. AMBER 2015. http://ambermd.org/.

33. Shaw, D. E., M. M. Deneroff, ..., S. C. Wang. 2008. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM.* 51:91–97.

34. Götz, A. W., M. J. Williamson, ..., R. C. Walker. 2012. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1 Generalized Born. *J. Chem. Theory Comput.* 8:1542–1555.

35. Salomon-Ferrer, R., A. W. Götz, ..., R. C. Walker. 2013. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* 9:3878–3888.

36. Allen, M. P., and D. J. Tildesley. 1989. Computer Simulations of Liquids. Clarendon Press/Oxford University Press, Oxford, United Kingdom.

37. Essmann, U., L. Perera, ..., L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

38. Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.

39. Miyamoto, S., and P. A. Kollman. 1992. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962.

40. Hopkins, C. W., S. Le Grand, ..., A. E. Roitberg. 2015. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* 11:1864–1874.

41. Bowers, K. J., E. Chow, ..., D. E. Shaw. 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters. *In* Proceedings of the 2006 ACM/IEEE SC|06 Conference. pp. 1–13.

42. Martyna, G. J., D. J. Tobias, and M. L. Klein. 1994. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101:4177–4189.

43. Krautler, V., W. F. van Gunsteren, and P. H. Hunenberger. 2001. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* 22:501–508.

44. Lippert, R. A., C. Predescu, ..., D. E. Shaw. 2013. Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J. Chem. Phys.* 139:164106.

45. Martyna, G. J., M. L. Klein, and M. E. Tuckerman. 1992. Nosé-Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* 97:2635–2643.

46. Shan, Y., J. L. Klepeis, ..., D. E. Shaw. 2005. Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* 122:54101.

47. Roe, D. R., and T. E. Cheatham, 3rd. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9:3084–3095.

48. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

49. Wong, V., and D. A. Case. 2008. Evaluating rotational diffusion from protein MD simulations. *J. Phys. Chem. B.* 112:6013–6024.

50. Prompers, J. J., and R. Brüschweiler. 2002. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124:4522–4534.

51. Nguyen, H. T., S. A. Pabit, ..., D. A. Case. 2014. Accurate small and wide angle x-ray scattering profiles from atomic models of proteins and nucleic acids. *J. Chem. Phys.* 141:22D508.

52. Svergun, D. I., C. Barberato, and M. H. J. Koch. 1995. CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28:768–773.

53. Flyvbjerg, H., and H. G. Petersen. 1989. Error estimates on averages of correlated data. *J. Chem. Phys.* 91:461–466.

54. Petoukhov, M. V., D. Franke, ..., D. I. Svergun. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Cryst.* 45:342–350.

55. Konarev, P. V., V. V. Volkov, ..., D. I. Svergun. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* 36:1277–1282.

56. Bax, A., and S. Grzesiek. 1993. Methodological advances in protein NMR. *Acc. Chem. Res.* 26:131–138.

57. Vranken, W. F., W. Boucher, ..., E. D. Laue. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* 59:687–696.

58. Palmer, A. G., 3rd 2001. Nmr probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* 30:129–155.

59. Delaglio, F., S. Grzesiek, ..., A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6:277–293.

60. d'Auvergne, E. J., and P. R. Gooley. 2008. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR.* 40:107–119.

61. d'Auvergne, E. J., and P. R. Gooley. 2008. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J. Biomol. NMR.* 40:121–133.

62. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

63. Schwieters, C. D., J. J. Kuszewski, ..., G. M. Clore. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 160:65–73.

64. Schwieters, C. D., and G. M. Clore. 2014. Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog. Nucl. Magn. Reson. Spectrosc.* 80:1–11.

65. Iwahara, J., C. D. Schwieters, and G. M. Clore. 2004. Ensemble approach for NMR structure refinement against 1H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* 126:5879–5896.

66. Takemura, K., and A. Kitao. 2012. Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. *J. Phys. Chem. B.* 116:6279–6287.

67. Piana, S., K. Lindorff-Larsen, and D. E. Shaw. 2011. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100:L47–L49.

68. Debiec, K. T., A. M. Gronenborn, and L. T. Chong. 2014. Evaluating the strength of salt bridges: a comparison of current biomolecular force fields. *J. Phys. Chem. B.* 118:6561–6569.

69. Clore, G. M., and J. Iwahara. 2009. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* 109:4108–4139.

70. Schwieters, C. D., J. J. Kuszewski, and G. Marius Clore. 2006. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* 48:47–62.

71. Shaw, D. E., J. P. Grossman, …, C. Young. 2014. ANTON 2 : Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. *In* SC14: International Conference for High Performance Computing, Networking, Storage, and Analysis. pp. 41–53.

72. Stone, J. E., D. J. Hardy, …, K. Schulten. 2010. GPU-accelerated molecular modeling coming of age. *J. Mol. Graph. Model.* 29:116–125.

73. Le Grand, S., A. W. Götz, and R. C. Walker. 2013. SPFP: speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* 184:374–380.

74. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.