Sequence-Dependent Correlated Segments in the Intrinsically Disordered Region of ChiZ

Alan Hicks [1,2,3,†], Cristian A. Escobar [1,3,4,†], Timothy A. Cross [1,3,4,*] and Huan-Xiang Zhou [5,6,*]

1   Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA;
     ah14k@my.fsu.edu (A.H.); cristian.escobar.b@gmail.com (C.A.E.)
2   Department of Physics, Florida State University, Tallahassee, FL 32306, USA
3   National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32310, USA
4   Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA
5   Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA
6   Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA
*   Correspondence: cross@magnet.fsu.edu (T.A.C.); hzhou43@uic.edu (H.-X.Z.)
†   These authors contribute equally to this work.

check for updates

Abstract: How sequences of intrinsically disordered proteins (IDPs) code for their conformational dynamics is poorly understood. Here, we combined NMR spectroscopy, small-angle X-ray scattering (SAXS), and molecular dynamics (MD) simulations to characterize the conformations and dynamics of ChiZ1-64. MD simulations, first validated by SAXS and secondary chemical shift data, found scant α-helices or β-strands but a considerable propensity for polyproline II (PPII) torsion angles. Importantly, several blocks of residues (e.g., 11–29) emerge as "correlated segments", identified by their frequent formation of PPII stretches, salt bridges, cation-π interactions, and sidechain-backbone hydrogen bonds. NMR relaxation experiments showed non-uniform transverse relaxation rates (R_2s) and nuclear Overhauser enhancements (NOEs) along the sequence (e.g., high R_2s and NOEs for residues 11–14 and 23–28). MD simulations further revealed that the extent of segmental correlation is sequence-dependent; segments where internal interactions are more prevalent manifest elevated "collective" motions on the 5–10 ns timescale and suppressed local motions on the sub-ns timescale. Amide proton exchange rates provides corroboration, with residues in the most correlated segment exhibiting the highest protection factors. We propose the correlated segment as a defining feature for the conformations and dynamics of IDPs.

Keywords: correlated segment; conformational dynamics; intrinsically disordered protein; molecular dynamics; nuclear magnetic resonance (NMR); protein conformation; small-angle X-ray scattering (SAXS)

1. Introduction

Intrinsically disordered proteins (IDPs) and proteins containing intrinsically disordered regions (IDRs) comprise up to 40% of the proteomes in all life forms [1]. They are involved in numerous cellular functions, including regulation and signaling [2,3]. As such, the dysregulation, misfolding, and aggregation of IDPs can lead to many diseases [4,5]. While lacking defined tertiary structures, IDPs can exhibit conformational preferences, such as transient secondary structures and recurrent residue–residue contacts (e.g., salt bridges and cation-π interactions) [6,7]. When binding to partners, transient secondary structures may become stable [8–10], and residue–residue contacts may switch from intramolecular to intermolecular [11]. Conformational dynamics may also play a particularly important role in the competition of IDPs for binding to the same partner [12] and in the binding

kinetics of IDPs with partners by dictating the binding mechanisms and binding and unbinding rate constants [11,13–15]. Clearly, the conformations and dynamics of IDPs are crucial for their cellular functions. Yet, how these properties are coded by the amino acid sequences of IDPs is poorly understood. The present study, using an integrated experimental and computational approach, aimed to address this question for an IDR in Mycobacterium tuberculosis (Mtb) ChiZ, a component of the machinery responsible for cell division.

Sequence analysis and coarse-grained modeling have identified some generic determinants, in particular charged residues, for the disorder and mean sizes of IDPs [16–19]. In recent years, small angle X-ray/neutron scattering (SAXS/SANS) [20]; fluorescence techniques, including nanosecond fluorescence correlation spectroscopy (nsFCS) and single-molecule fluorescence resonance energy transfer (smFRET) [21–23]; nuclear magnetic resonance (NMR) spectroscopy [24]; and all-atom molecular dynamics (MD) simulations [25] have become key biophysical tools in characterizing the conformation ensembles of IDPs. Among these, nsFCS, NMR, and MD can also probe conformational dynamics, each with strengths on particular timescales. Scattering experiments yield information on the overall sizes and extents of disorder [20,26]. By site-specific labeling, fluorescence techniques report on the mean distances between different sites within a protein chain and the reconfiguration dynamics of the chain on the hundreds of ns timescale, as well as interactions between different protein chains [14,27,28].

NMR spectroscopy, based on various types of experiments, remains the only biophysical technique for characterizing both the conformations and dynamics of IDPs at a residue-level resolution across timescales from picosecond (ps) to second [24]. A simple telltale of intrinsic disorder is the narrow hydrogen dispersion in the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra [29]. Secondary chemical shifts, which measure the deviations from random-coil reference values, can indicate the propensities of secondary structures [30,31]. Backbone solvent exposure and interactions can be investigated using hydrogen exchange experiments for both globular and disordered proteins [32–34]. Amide ^{1}H-^{15}N spin relaxation rates report on the ps to supra-ns backbone dynamics [35]. In globular proteins, relaxation data are typically analyzed through the Lipari-Szabo model-free approach, assuming the separability of global tumbling motions from local backbone fluctuations [36]. For disordered proteins, global and local motions are no longer separable, and interpreting relaxation data becomes a challenge. One can still model the relaxation data by fitting the NH bond vector time-correlation functions, $C_{NH}(\tau)$, to a sum of exponentials, but the assignment of the resulting time constants to specific types of motions can be ambiguous [37–39]. MD simulations can help elucidate these connections.

In recent years, it has become evident that MD force fields, traditionally parameterized for structured proteins, when applied to IDPs lead to overly compact conformations [40,41]. Based on benchmarking against experimental data including SAXS profiles, FRET efficiencies, and various NMR parameters, a number of IDP-specific force fields have been proposed, including AMBER03WS/TIP4P2005 [42], various protein force fields in combination with TIP4PD water [43,44], and CHARMM36m/TIP3Pm [45]. Still, the demand for more accurate IDP force fields remains unabated, especially regarding dynamic properties. In several MD simulation studies, the $C_{NH}(\tau)$ correlation functions were fitted to a sum of exponentials, but either the fitting parameters or the trajectories used for the fitting had to be adjusted in order to reach agreement with the experimental data [46–51]. It is thus notable that, using the AMBER14SB [52]/TIP4PD [43] force field and without adjusting $C_{NH}(\tau)$, Kämpf et al. [53] were able to reproduce experimental relaxation data for the 26-residue N-terminal fragment of histone H4. In NMR and MD studies in which $C_{NH}(\tau)$ was fitted to a sum of exponentials, three or four exponentials were typically used, and the time constants ranged from a few ps to 10 ns. While there is significant disagreement as to the nature of the intermediate time scales (0.1 to 1 ns), the fastest of these motions generally is assigned to the libration of the NH bond vector with respect to the peptide plane, and the slowest assigned to some form of segmental motion. One form of segmental motion arises from the simple fact that each residue is part of a polypeptide chain, which has a

certain correlation length along the chain [54,55]. This type of segmental motion lacks strong sequence dependence and can be recognized by reduced transverse relaxation rates (R_2) at the chain termini (or a bell shape for the R_2 vs. sequence curve), as first reported for denatured lysozyme by Schwalbe and co-workers. Although these authors also reported increases in R_2 by tertiary interactions, leading to the apparent sequence dependence of R_2, the latter have not received much attention in studies on IDPs. A notable exception is a recent study of the low-complexity domain of heterogenous nuclear ribonucleoprotein A1 (A1-LCD), where the regions of increased R_2s were attributed to π–π interactions between aromatic residues [56].

The transmembrane protein ChiZ is one of a dozen or so proteins that comprise the Mtb divisome, the machinery responsible for cell division. Mtb is the causative agent of tuberculosis; its cell division has strong implications for both pathogenesis and drug resistance [57]. The structural determination of divisome membrane proteins and their complexes has begun [58], but sequence analysis suggests that many of these proteins, including ChiZ, CrgA, FtsQ, FtsI, and CwsA, have disordered extramembranous regions of various lengths. ChiZ consists of 165 residues (Figure 1a); the cytoplasmic N-terminal 64 residues (ChiZ1-64) are predicted to be disordered (Figure 1b,c), and the next 22 residues form a transmembrane helix; on the periplasmic side, a disordered linker connects the transmembrane helix to a 53-residue LysM domain that binds to peptidoglycans [59]. The exact role of ChiZ in cell division is still an open question. Its full name, cell wall hydrolase interfering with FtsZ ring assembly (gene Rv2719c), may have been a misnomer, as a recent study showed that zymogram assays suggesting cell wall hydrolase activity by Chauhan et al. [59] likely yielded a false positive [60]. On the other hand, the interference with FtsZ ring assembly remains intact. The polymerization of FtsZ (a bacterial homolog of tubulin), forming the FtsZ ring, initiates the septation step of cell division; thus, the correct localization of the FtsZ ring is crucial for proper division [57]. With the increased expression of ChiZ, Mtb cells grown in macrophages were filamentous; promotion of filamentation by ChiZ overexpression in M. smegmatis (a non-pathogenic mycobacterium) affected the mid-cell location of FtsZ rings [59,61]. Importantly, the disordered N-terminal region and transmembrane helix sufficed for cell filamentation and FtsZ ring mislocalization [62]. Bacterial adenylate cyclase-based two-hybrid (BATCH) assays indicated that ChiZ interacts with FtsQ and FtsI but not FtsZ, implicating an indirect mechanism for FtsZ-ring mislocalization [62].
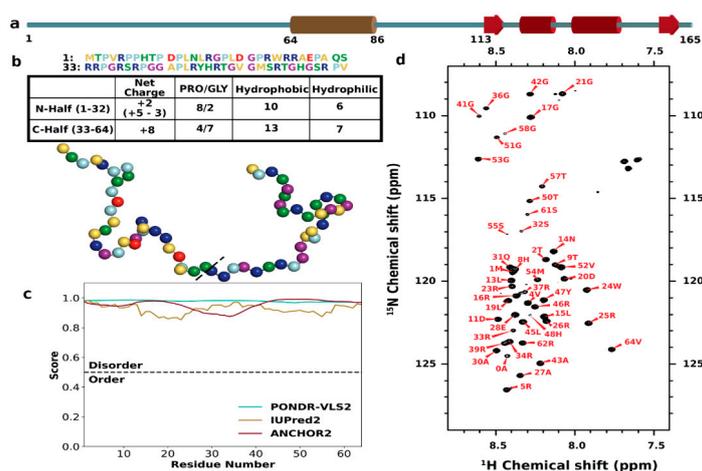

Figure 1. Amino acid sequence and disorder of ChiZ. (**a**) Domain organization of full-length ChiZ, composed of a disordered N-terminal region (residues 1–64), a transmembrane helix (residue 65–86), a disordered linker (residues 87–112), and a LysM domain (residues 113–165). (**b**) Sequence and amino acid composition of ChiZ1-64. In both the sequence and an illustrative conformation, residues are colored in the following scheme: cationic, blue; anionic, red; prolines, light blue; glycines, purple; hydrophilic, green; and hydrophobic, yellow. (**c**) Disorder predictions from three web servers, PONDR-VLS2, IUPred2, and ANCHOR2. (**d**) ^{1}H-^{15}N HSQC spectrum, acquired on an 800 MHz spectrometer at 25 °C in 20 mM of phosphate buffer (pH 7.0) with 25 mM of NaCl.

Here, we combined SAXS, NMR, and MD simulations to thoroughly investigate the conformational and dynamic properties of ChiZ1-64. Based on benchmarking against the SAXS profile and secondary chemical shifts, we selected the AMBER14SB/TIP4PD force field among the five tested. Experimental R_2 rates were non-uniform along the sequence, which was recapitulated by MD simulations. The sequence-dependent dynamics can be attributed to the formation of correlated segments, stabilized by polyproline II (PPII) conformation and intra-segmental interactions. In particular, the residues with the largest amplitudes for motions on the slowest timescale (approximately 10 ns), including Asp11, Trp24, Arg25, Arg26, and Tyr47, frequently engaged, with different partners, in salt bridges and cation–π interactions. The linkage of conformation and dynamics to sequence, captured by the formation of correlated segments, will be useful for understanding IDPs and their interactions with partners.

2. Materials and Methods

2.1. Protein Expression and Purification

The expression of ChiZ1-64 containing an N-terminal His-tag with a TEV protease cleavage site was performed in *E. coli* BL21 cells. Cells were grown at 37 °C until the O.D. at 600 nm was 0.7, and then 0.4 mM of IPTG was added to induce expression for 5 h at 37 °C. For ^{13}C-^{15}N uniformly labeled samples, protein expression was performed in M9 media containing 1 g of ^{15}N ammonium chloride and 2 g of ^{13}C-labeled glucose.

The ChiZ1-64 was initially purified using nickel affinity chromatography. The cells were resuspended in a lysis buffer (20 mM of tris-HCl pH 8.0 containing 500 mM of NaCl and 6 M of urea) and lysed by a French Press. The lysate was centrifuged at 12,000× *g* for 40 min to remove insoluble material. After that, the lysate was loaded onto a Ni-NTA resin column (Qiagen). The column was washed with a washing buffer (20 mM of tris-HCl pH 8.0 containing 500 mM of NaCl and 60 mM of imidazole) and then eluted with 400 mM of imidazole.

After nickel affinity chromatography, the fractions containing the protein were pooled and treated with TEV protease. The His-tag and TEV protease were removed by passing the protein sample through a Ni-NTA column. Further purification proceeded by cation exchange chromatography. ChiZ1-64 was dialyzed against the NMR buffer (20 mM of sodium phosphate at pH 7.0 plus 25 mM of NaCl) and then loaded into an SP column (GE Healthcare). The column was washed with the NMR buffer containing 200 mM of NaCl and eluted with 500 mM of NaCl. The fractions containing the protein were concentrated and dialyzed against the NMR buffer for further experiments.

2.2. Small Angle X-ray Scattering

SAXS experiments were performed on the DND-CAT 5ID-D beam line at the Advanced Photon Source of the Argonne National Laboratory. The X-ray wavelength was 1.2398 Å. The X-ray scattering intensities were collected using Rayonix LX170HS CCD detectors positioned at 200.92 mm ($0.0014 < q < 0.08$ Å^{-1}) and 1014.2 mm ($0.077 < q < 0.485$ Å^{-1}). ChiZ1-64 was at 9.1 mg/mL in the NMR buffer. The X-ray exposure time was limited to 5 s to minimize protein degradation. Data processing was performed using the ATSAS software package [63].

2.3. NMR Spectroscopy

Samples for the solution NMR experiments were prepared in the NMR buffer containing 10% D$_2$O and 50 μM of DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid, for referencing). All the experiments were performed at 25 °C in an 800 MHz magnet equipped with cryoprobe. A sequential backbone assignment was performed using standard HNCO, HN(Cα)CO, HNCαCβ, and CαCβ(CO)NH experiments. The data were processed using Topspin 2.1 (Bruker), and analyzed using the CCPNmr software.

The backbone dynamics were characterized by measuring the amide ^{15}N R_1 and R_2 relaxation rates. The R_1 measurements were performed using the Bruker pulse sequence hsqct1etf3gpsi with the

following time delays: 10, 62.5, 125, 250, 500, 750, 1000, and 1500 ms. The R_2 relaxation rates were measured using the Carr–Purcell–Meiboom–Gill experiment (hsqct2etf3gpsi) with the following time delays: 5, 30, 62.5, 125, 187.5, 250, 312.5, and 375 ms. Relaxation delays between scans were 6 s for R_1 and 4 s for R_2. The signal intensities for each residue were fit to an exponential to extract R_1 and R_2; the fitting errors were reported as errors in these parameters. In addition, using the Bruker pulse sequence hsqcnoef3gpsi, the ^{1}H-^{15}N heteronuclear NOE value for each residue was obtained as the ratio of signal intensities collected with and without proton saturation, with a 10 s relaxation delay between scans. The same settings were used for relaxation measurements at pH 4.0. The NOE values were the average of two independent measurements, with the errors corresponding to the standard deviations of those measurements.

The measurement of amide proton exchange rates was carried out using the CLEANEX-PM pulse sequence (fhsqccxf3gpph). The CLEANEX-PM spin lock times were 10, 15, 20, 25, 30, 40, 50, 75, and 100 ms. A fast HSQC reference spectrum was collected using the same pulse sequence parameters. All the experiments were run with a relaxation delay of 3 s. The amide proton exchange rates were calculated by fitting equation 1 in Hwang et al. [64] to the signal intensities for each residue at different spin lock times. The intrinsic exchange rates were calculated using the SPHERE server (https://protocol.fccc.edu/research/labs/roder/sphere/sphere.html) [65]. The protection factors were calculated as $k_{intrinsic}/k_{ex}$ [66].

The significance of the differences in the means of R_1, R_2, NOE, and the protection factor between the two halves of ChiZ1-64 was analyzed by the independent samples t-test assuming unequal variances (Welch's *t*-test) using the scipy stats module in python.

2.4. Molecular Dynamics Simulations

Five force fields were tested on ChiZ1-64 in solution: AMBER14SB [52]/TIP4PD [43] (FF14D), AMBER03WS/TIP4P2005 [42] (FF03WS), AMBER99SB-ILDN [67]/TIP4PD [43] (FF99D), AMBER15IPQ/SPCEb [68] (FF15IPQ), and CHARMM36m/TIP3Pm [45] (C36M). The MD simulations of ChiZ1-64 (with ACE and NME caps), started from an extended conformation generated using tleap in AmberTools16 [69], were performed in AMBER16 [69] (and extended in AMBER18 [70]). The AMBER topology file of the TIP4PD water model was from https://github.com/ajoshpratt/amber16-tip4pd. The FF14D, FF99D, and FF15IPQ simulation systems were set up using tleap, in an orthorhombic box with 12 Å of space to all sides of the protein. 25 mM of NaCl was added to the solution along with 10 neutralizing Cl$^-$ ions. The FF03WS system was built in GROMACS 2016.4 [71,72] to match the AMBER systems and then converted to the AMBER format using GROMBER in PARMED [69]. The C36m system was built using the solution builder in CHARMM-GUI [73], again to match the AMBER systems, and exported to the AMBER format using CHAMBER [74]. The total numbers of atoms in the systems ranged from 112,000 to 162,000.

To start, energy minimization was carried out using sander for 2000 steepest descent steps, followed by 3000 conjugate gradient steps. Subsequently, temperature equilibration, pressure equilibration, and production run were performed using pmemd.cuda on GPUs [75]. Under a constant volume, the temperature was ramped from 0 K to 300 K in 40 ps and then maintained at 300 K for 60 ps, using the Langevin thermostat with a friction coefficient of 3 ps^{-1} at 1 fs timesteps. The simulations then switched to constant pressure (Berendsen thermostat at 1 atm with a pressure relaxation time of 2 ps) and the timestep increased to 2 fs. The first 3 ns was nominally pressure equilibration, and the remaining simulations counted as the production run. The non-bonded cutoff was 10 Å in all simulations, except in C36m, where a force switch was imposed between 10 and 12 Å. All the bonds connected to hydrogens were constrained using the SHAKE algorithm [76].

For each of the force fields tested, 4 replicate simulations started with different random seeds were run for 500 ns each. Simulations in two of the force fields, FF14D and FF03WS, were expanded to 12 replicates, each lasting 3 μs. Snapshots were saved every 20 ps for analysis. In only 1.7% of snapshots ChiZ1-64 came within the non-bonded cutoff (10 Å) from its periodic images.

2.5. Calculation of SAXS Profiles

From the MD conformations, SAXS profiles were calculated using the FOXS code [77]. The optimal parameter for the hydration shell scattering density was selected for each water model according to Henriques et al. [78]. For each trajectory, a SAXS profile was calculated on every 10th saved conformation, and then an average was taken over these conformations. The resulting SAXS profile was linearly scaled to best match the experimental counterpart. The average of the scaled SAXS profiles over the replicate simulations was taken as the MD prediction. In cases where the number of replicates is 12, we also report the standard deviation among the replicates at each q as a measure of the calculation error.

All graphs were plotted using matplotlib and seaborn in python3.

2.6. Calculation of Chemical Shifts

Chemical shifts were calculated using the SHIFTX2 code [79] (at 300 K and pH 7). For each residue, the corresponding random-coil chemical shifts generated from the ncIDP database [80] were subtracted to yield Cα and Cβ secondary chemical shifts (without any scaling). Details for the averages and standard deviations largely followed the protocol for SAXS profiles.

2.7. Radius of Gyration, Secondary Structures, and Hydrogen Bonds

Cpptraj [81] was used for determining the radius of gyration, secondary structures (by implementing DSSP [82]), and hydrogen bonds. DSSP was modified to include PPII, following Mansiaux et al. [83]. Specifically, three or more consecutive residues that were classified as coil and fell into the PPII region of the Ramachandran map were reclassified as PPII.

2.8. Dihedral Principal Component Analysis

Dihedral principal component analysis (dPCA) [84,85] was performed through cpptraj [81], yielding 248 eigenmodes for the backbone φ and ψ angles of the 62 non-terminal residues of ChiZ1-64 (φ and ψ each were represented by their sine and cosine). To display the energy landscape in conformational space, the histogram of the projections of the saved snapshots along the two dPCA eigenmodes with the largest eigenvalues was calculated and then converted to a free energy surface according to the Boltzmann relation. These two projected coordinates were also used to group the snapshots into 16 clusters using the hierarchical Ward agglomerative algorithm. The snapshot that had the highest similarity score to all the members in a cluster was selected as the representative. The similarity score was defined as [86]:

$$S_i = \left\langle \frac{1}{N^2} \sum_{n,m=1}^{N} \frac{1}{1 + (r_{n,m}^i - r_{n,m}^j)^2} \right\rangle_j \tag{1}$$

where $r_{n,m}^i$ denotes the distance between atoms n and m in snapshot i, N is the total number of atoms in ChiZ1-64, and the average over j ran over all the snapshots in the given cluster.

The contribution from the fluctuations of torsion angle n to eigenmode k was determined by the amplitude of this eigenmode's components, v_{2n-1}^k and v_{2n}^k, for the sine and cosine of the torsion angle (denoted by indices $2n-1$ and $2n$). Specifically, this contribution was:

$$\Delta_n^k = \left(v_{2n-1}^k\right)^2 + \left(v_{2n}^k\right)^2 \tag{2}$$

Reference [85]. Note that the sum of Δ_n^k over all the torsion angles is 1.

2.9. Contact Maps

Mdtraj [87] was used to load the trajectories, select atoms, and calculate distances between sidechain and sidechain or sidechain and backbone heavy atoms, excluding pairs from the same residue. Two heavy atoms were considered to be in contact if they were within 3.5 Å of each other. For each pair, the fraction of snapshots in which contacts formed was recorded.

The contacts formed by the two aromatic residues, Trp24 and Tyr47, with arginines were further examined to see whether they were cation–π interactions. The distance between the centers of mass of the indole or phenol ring and of the cationic group (including N_ε, C_ξ, $N_{\eta 1}$, and $N_{\eta 2}$), and the angle between the line connecting these two points and the normal of the ring were calculated. The overwhelming majority of Trp24 contacts with Arg16, Arg25, and Arg26 had the above distance < 5 Å and the above angle < 60°, and hence were deemed cation–π interactions. The same was true of Tyr47 contacts with Arg46 and Arg49.

2.10. NMR Relaxation Properties

From each trajectory, the NH bond vector time-correlation function for each non-proline residue was calculated as a time average:

$$C_{NH}(\tau) = \langle P_2[\mathbf{n}(t+\tau) \cdot \mathbf{n}(t)] \rangle_t = \sum_{i=1}^{n} A_i e^{-\tau/\tau_i} \tag{3}$$

where $P_2(x)$ is the second-order Legendre polynomial and $\mathbf{n}(t)$ is the NH bond unit vector at time t. Each correlation function, with τ ranging from 20 ps to 25 ns, was then least-squares fit to a sum of exponentials, using the Levenberg–Marquardt algorithm from the scipy.optimize.curve_fit module in python. Note that the sum of the amplitudes was not restricted to 1, in contrast to most other studies, under the assumption that an ultrafast decay was completed by τ = 20 ps (the time interval at which we saved the snapshots in the MD simulations), thereby accounting for some missing amplitudes (see Section 3.7). To determine the optimal number of exponentials for modeling the simulation data, we compared the chi-squares (χ^2) of the fits with increasing n, starting at n = 2. Any fit with fitting errors higher than 10% of any fitted parameter was rejected. An increase from the n exponentials to $n + 1$ exponentials was accepted specifically if:

$$\frac{\chi_{n+1}^2}{\chi_n^2} < \frac{1}{2}\left[1 - \frac{n}{n+1}\right] \tag{4}$$

This procedure led to n = 3 as the optimum for all residues. The three time constants were ordered as $\tau_1 > \tau_2 > \tau_3$.

After the tri-exponential fit, the spectral density was obtained as:

$$J(\omega) = \sum_{i=1}^{3} \frac{A_i \tau_i}{1 + (\omega \tau_i)^2} \tag{5}$$

Finally, the longitudinal and transverse relaxation times and the NOE were obtained as [35]:

$$R_1 = f_{DD}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + f_{CSA}J(\omega_N) \tag{6}$$

$$R_2 = \frac{R_1}{2} + f_{DD}[2J(0) + 3J(\omega_H)] + \frac{2}{3}f_{CSA}J(0) \tag{7}$$

$$\text{NOE} = 1 + \frac{f_{DD}\gamma_H}{R_1\gamma_N}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \tag{8}$$

Here, $f_{DD} = \frac{1}{10}\left(\frac{\mu_0\hbar\gamma_N\gamma_H}{4\pi r_{NH}^3}\right)^2$ and $f_{CSA} = \frac{2}{15}\omega_N^2\Delta_{CSA}^2$ represent contributions to ^{15}N spin relaxation by NH dipole–dipole interactions and the nitrogen chemical shift anisotropy, respectively. The meanings of the other symbols are: μ_0, permittivity of free space; $\hbar$, reduced Plank constant; γ_N and γ_H, gyromagnetic ratios of nitrogen and hydrogen; $\omega_H = \gamma_H B_0$, Larmor frequency of hydrogen (800 MHz in our case); ω_N, counterpart of nitrogen; r_{NH}, NH bond length (set to 1.02 Å); and Δ_{CSA} (= −170 ppm), chemical shift anisotropy of nitrogen. For each relaxation property, we report the discrepancies between the measured and predicted values as the root mean squared error (RMSE), calculated over the entire ChiZ1-64 sequence except for the first and last residues. A bootstrapped 95% confidence interval was obtained to determine the error in the calculated R_1, R_2, and NOE.

We also considered two modifications to the tri-exponential NH bond vector time-correlation functions. The first was to include an ultrafast decay component, with time constant τ_f (< 20 ps) and an amplitude of 1—A_{sum}. The second was to account for the possibility that the longest timescale was exaggerated by the AMBER14SB/TIP4PD force field selected here. We hence tested scaling down the three time constants from the tri-exponential fits by a factor $1 + \tau_i/\tau_s$, with τ_s of the order of 10 ns. This scaling has little effect on τ_2 and τ_3, but reduced the longest time constant τ_1 by roughly half, from 7–17 ns to 5–8 ns.

2.11. Data Availability

The chemical shifts of ChiZ1-64 have been deposited in BMRB (accession # 50115). Python scripts written for the NMR relaxation analysis are available on GitHub at https://github.com/achicks15/CorrFunction_NMRRelaxation.

3. Results

3.1. Sequence Characteristics and Disorder of ChiZ1-64

ChiZ1-64 has disparate amino acid compositions between the first 32 residues (N-half) and the last 32 residues (C-half), in particular concerning prolines, glycines, and charged residues (Figure 1b). Of the 12 prolines (19% of the sequence), two thirds are in the N-half. In contrast, of the nine glycines (14%), seven, or nearly 80%, are in the C-half. Prolines are known to break α-helices and β-strands but promote PPII helices, whereas glycines break all secondary structures. There is a significant net charge, +10e, coming from 13 arginines, two aspartates, and one glutamate. All the three anionic residues are in the N-half, whereas eight, or 62%, of the cationic residues are in the C-half, resulting in the contrast between a near balance of opposite charges in the N-half and total imbalance in the C-half. Lastly, we note that each half contains an aromatic residue, Trp24 in the N-half and Tyr47 in the C-half.

Giving the abundance of prolines and glycines and the high net charge, it is not surprising that the entire sequence of ChiZ1-64 is predicted to be disordered with high confidence [88–90] (Figure 1c). The ^{1}H-^{15}N HSQC spectrum confirms the disorder, with proton chemical shifts confined to the narrow range of 7.7 to 8.6 ppm (Figure 1d).

3.2. SAXS Profile and Secondary Chemical Shifts

The SAXS profile (Figure 2a), i.e., the scattering intensity $I(q)$ as a function of q, the magnitude of the scattering vector, shows ChiZ1-64 as a typical IDP, especially when presented as a Kratky plot (Figure 2b). The radius of gyration R_g obtained from a fit to the Debye approximation (Figure S1a) is 24.17 ± 0.05 Å. This value is slightly largely than that, 22.3 Å, predicted by a scaling relation,

$$R_g = 2.54N^{0.522} \tag{9}$$

deduced from a set of IDPs [91]. A modest degree of expansion is also indicated by an upward tilt of the Kratky plot at high q.
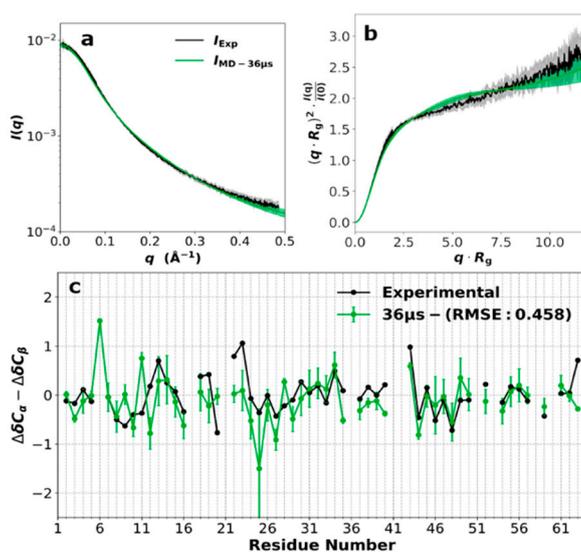
Figure 2. Small-angle X-ray scattering (SAXS) profile and secondary chemical shifts. (**a**) Scattering intensity $I(q)$; (**b**) Kratky plot; (**c**) secondary chemical shifts. Experimental data are shown in black, while the predictions from the 36-μs AMBER14SB/TIP4PD simulations are in green.

Secondary chemical shifts for C_α and C_β can indicate the presence of α-helices and β-strands (corresponding to $\Delta\delta C_\alpha - \Delta\delta C_\beta > 2$ ppm and < -2 ppm, respectively) [30]. For ChiZ1-64, only a single residue had $|\Delta\delta C_\alpha - \Delta\delta C_\beta| > 1$ (Figure 2c), indicating a lack of α-helices and β-strands for the entire sequence.

3.3. Force Field Validation

We used the measured SAXS profile and secondary chemical shifts to test five force fields: AMBER14SB [52]/TIP4PD [43], AMBER03WS/TIP4P2005 [42], AMBER99SB-ILDN [67]/TIP4PD [43], AMBER15IPQ/SPCEb [68], and CHARMM36m/TIP3Pm [45]. In simulations totaling 2 μs, AMBER14SB/TIP4PD and AMBER03WS/TIP4P2005 did equally well and outperformed the other three force fields in matching the SAXS profile (Figure S1b–f). For secondary chemical shifts, AMBER14SB/TIP4PD was ahead of all the other four force fields (Figure S2). We further expanded the AMBER14SB/TIP4PD and AMBER03WS/TIP4P2005 simulations to 36 μs (among 12 replicate trajectories). In the expanded simulations, the AMBER14SB/TIP4PD results moved even closer toward the experimental counterparts (Figure 2, Figures S1b and S2a), whereas the AMBER03WS/TIP4P2005 results did not see any improvement (Figures S1c and S2b). From here on, we will focus on the AMBER14SB/TIP4PD simulations and no longer state the name of the force field.

To check the convergence of the 36 μs simulations, we calculated the R_g histograms from the 12 replicate trajectories (Figure S3). The histograms all showed broad distributions, with significant frequencies for R_g between 15 to 40 Å and mean R_g values ranging from 22.44 to 26.44 Å. Combining the 12 replicate simulations, the overall mean R_g was 24.4 Å, with a standard deviation of 1.4 Å among the replicates. The mean R_g agrees well with the experimental results. Overall, the selected force field reproduced the experimental data well for both residue-specific properties and global conformational properties.

3.4. High Poly-Proline II Propensities

Consistent with the lack of α-helices and β-strands indicated by secondary chemical shifts, the contents of these secondary structures were minimal in the MD simulations (Figure 3). Two stretches of residues in the N-half, 13–15 and 30–32, formed 3_{10} helices with a moderate frequency (~7%). Note that 3_{10} helices have a much lower intrinsic stability than α-helices. In addition, 3_{10} helices and anti-parallel β-sheets were formed infrequently by C-half residues (45 to 61). On the other hand, ChiZ1-64 exhibited

high PPII propensities, which only the MD simulations were able to reveal. Here, PPII was counted when contiguous residues (minimum of three) fell in the PPII region on the Ramachandran map (Figure S4). Three stretches of residues sampled PPII over 50% of the time. All of them are in the N-half: residues 4–6, 10–12, and 27–29. In comparison, the highest PPII frequency in the C-half was only 36% for residue 44. Prolines are the most direct reason for the high PPII propensities, as the high-PPII stretches in the N-half contain or border prolines at positions 3, 6, 7, 10, 12, and 29; in the C-half, residue 44 is also a proline.
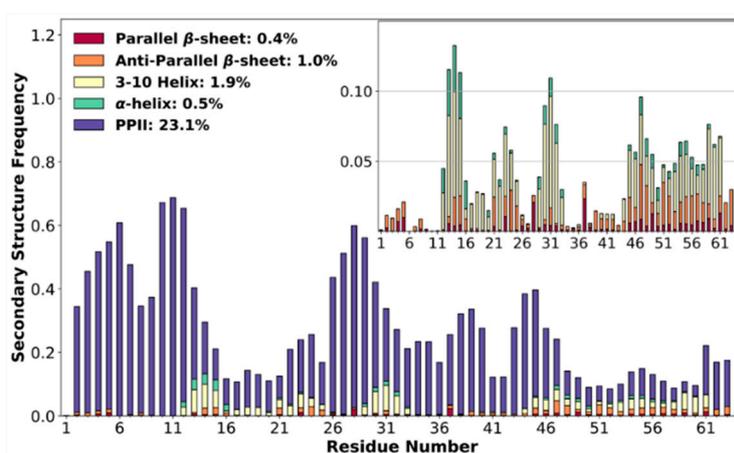
Figure 3. Normalized frequencies of individual residues in various types of secondary structures according to the molecular dynamics (MD) simulations. The inset zooms into the low populations of helices and β-sheets.

Two of the eight N-half prolines, at positions 18 and 22, were not found in or next to high-PPII stretches, possibly because each is next to a glycine (at positions 17 and 21). Glycines may also partly explain the much lower PPII propensities of the C-half, by being next to Pro35 (at position 36), Pro40 (at positions 41 and 42), and between Pro44 and Pro63 (at positions 51, 53, 58, and 60).

Proline strongly prefers the PPII region on the Ramachandran map (Figure S4). This preference extends to the preceding residue, unless it is a glycine. However, PPII helices are only marginally stable. Unlike α-helices and β-sheets, PPII helices are not stabilized by backbone hydrogen bonds. Although prolines provide some impetus, PPII stretches may not form unless stabilized by other interactions (see below).

3.5. Flat Energy Landscape in Conformational Space

The lack of stable secondary structures portended a high degree of diversity in the conformations sampled by ChiZ1-64. To quantify this aspect, we performed backbone dihedral principal component analysis (dPCA) [84,85] on conformations saved from the MD simulations. Each conformation was projected onto the first two eigenmodes with the largest eigenvalues, and the distribution of the conformations in this two-dimensional subspace was obtained. The resulting free energy surface shows a broad, shallow basin, with local barriers all less than 2 k_BT (k_B: Boltzmann constant; T: absolute temperature) (Figure S5a).

Another indication of the conformational diversity is provided by the closely spaced eigenvalues (Figure S6a; in a contrasting scenario where a few large eigenvalues are separated from many small eigenvalues, the former would correspond to modes involving the concerted motions of a large portion of the protein, whereas the latter would correspond to localized motions). When normalized by the sum of all eigenvalues, the four largest eigenvalues were 0.029, 0.025, 0.023, and 0.021; the eigenvalues decreased smoothly with an increasing mode number. The first four eigenmodes, represented by the fluctuation amplitudes of individual torsion angles, are displayed in Figure S6b–e. The amplitudes of the φ angles, with the exception for those of a few glycines, were low, reflecting the fact that φ

was mostly confined to the range of $-50°$ to $-150°$ (Figure S4). The ψ values spanned a wide range, covering different secondary structures ($-100°$ to $0°$ for α- and 3_{10} helices, and $100°$ to $180°$ for PPII helices and β-strands). Residues with high ψ amplitudes in the first three modes mostly were found in the two N-half stretches, 13–15 and 30–32, with a moderate 3_{10} propensity. The fourth mode mostly involved C-half residues (45 to 61) that formed 3_{10} helices and anti-parallel β-sheets infrequently.

To find a minimal set of conformations that still conveyed the overall sense of conformational diversity, we used the projections of the MD conformations in the subspace of the first two eigenmodes to group them into 16 clusters (Figure S5b) and selected one conformation from each cluster. The selection was based on a similarity score, which measured the extent of similarity of a given conformation to all the other conformations in the same cluster. The highest similarity score for any conformation with all the other cluster members ranged from 0.15 to 0.19, about the same as that between two randomly chosen conformations, again highlighting the conformational diversity.

The set of 16 conformations, one from each cluster with the highest similarity score, illustrates the conformational diversity in the MD simulations (Figure 4). All these conformations contained at least one PPII stretch; five of them contained a 3_{10} helix; two contained a hybrid 3_{10}-α helix (featuring both i to $i + 3$ and i to $i + 4$ hydrogen bonds); one contained an antiparallel β-sheet. Visual inspection also revealed that arginines frequently formed salt bridges with the aspartates and glutamates as well as cation–π interactions with the tryptophan and tyrosine. Furthermore, the cationic and anionic side chains frequently formed hydrogen bonds with backbone carbonyls and amides, respectively. Sometimes these interactions grew into a network. Thus, while the backbone conformations were diverse, the salt bridges, cation–π interactions, and side chain-backbone hydrogen bonds were pervasive, albeit formed by different partners at different times.
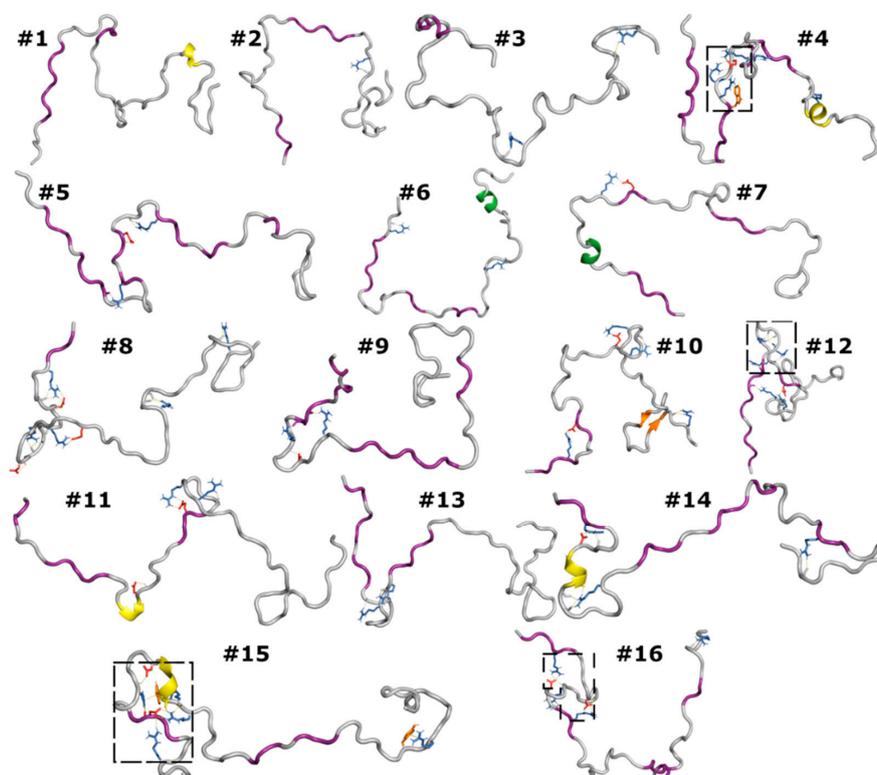

Figure 4. Representative conformations from the MD simulations. Secondary structures are shown by the color of the backbone (PPII, 3_{10}, and hybrid 3_{10}-α helices in purple, yellow, and green respectively; β-sheet, orange). Cationic, anionic, and aromatic side chains involved in salt bridges, cation–π interactions, and SC-BB hydrogen bonds are shown in blue, red, and orange, respectively. Boxed regions, after enlargement, are shown in Figure 5.
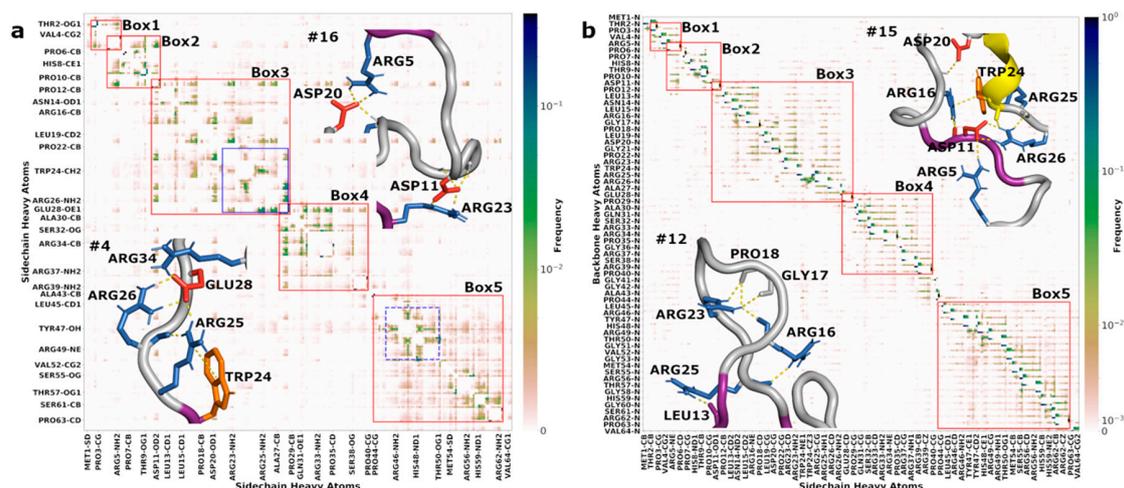
Figure 5. Contact maps between (**a**) sidechain–sidechain and (**b**) sidechain–backbone atom pairs. Normalized contact frequencies are shown as color gradients, which follow a linear scale for contact frequencies between 0 and 0.01 but a logarithmic scale above 0.01. Red boxes indicate segments of residues that frequently form contacts. Blue boxes (solid, residues 23–28; dash, 46–50) highlight residues with particularly high contact frequencies. See an enlarged view of Box 3 and Box 5 in Figure S7. Enlarged views of regions from four conformations (#4, #12, #15, and #16) in Figure 4 are shown as insets to illustrate SC-SC and SC-BB interactions.

3.6. Correlated Segments Revealed by Contact Maps

To quantify these prevailing interactions formed in the MD simulations, we calculated the contact frequencies between heavy atoms on any two side chains (SC-SC; Figure 5a) or between a heavy atom on any side chain and a heavy atom on the backbone of any other residue (SC-BB; Figure 5b). A contact was formed when two heavy atoms were less than 3.5 Å apart.

Overall, the N-half formed nonlocal SC-SC contacts much more frequently than the C-half. To quantify this difference, we took the highest contact frequency among the SC heavy atoms of two residues to represent that residue pair and, for each residue, defined its nonlocal contact number as the average of the contact frequencies among all the partner residues except for the three nearest neighbors in either direction. The mean of the nonlocal contact numbers for the N-half residues was 0.00112, nearly twice of the counterpart, 0.00069, for the C-half residues. Residues forming SC-SC contacts with significant frequencies could roughly be grouped into five segments along the sequence (indicated by red boxes in Figure 5a). The N-half broke into three segments: Thr2 to Pro6, Arg5 to Pro12, and Asp11 to Pro29. The fourth segment, Glu28 to Pro40, straddled the two halves. The rest of the C-half contained one more segment, Pro44 to Pro63. For several residues, including Arg5, Asp11, and Glu28 (Glu28 is illustrated in Figure 5a inset #4), the contacts extended beyond a single segment, explaining why every two adjacent segments in the N-half had a two-residue overlap. Contacts made by the three anionic residues, Asp11, Asp20, and Glu28, traversed the N-half (illustrated by an Asp20-Arg5 salt bridge in Figure 5a inset #16) and even extended into the entire C-half.

The most extensive interaction network was formed with Trp24, Arg25, Arg26, Glu28 at the core (Figure 5a blue solid box and inset #4; see also Figure 5b inset #15; Figure S7a). Trp24 formed cation–π interactions with Arg25 and other arginines, whereas Glu28 formed multiple salt bridges with Arg25, Arg26, and other arginines. In the C-half, the most extensive interaction network (Figure 5a blue dash box; Figure S7b) had cation–π interactions of Tyr47 with Arg46 and Arg49 at the core (e.g., Tyr47-Arg49 in Figure 4 #15). We will see that these salt bridges and π interactions align with the regions of slow backbone dynamics when presenting Figures 6 and 7.
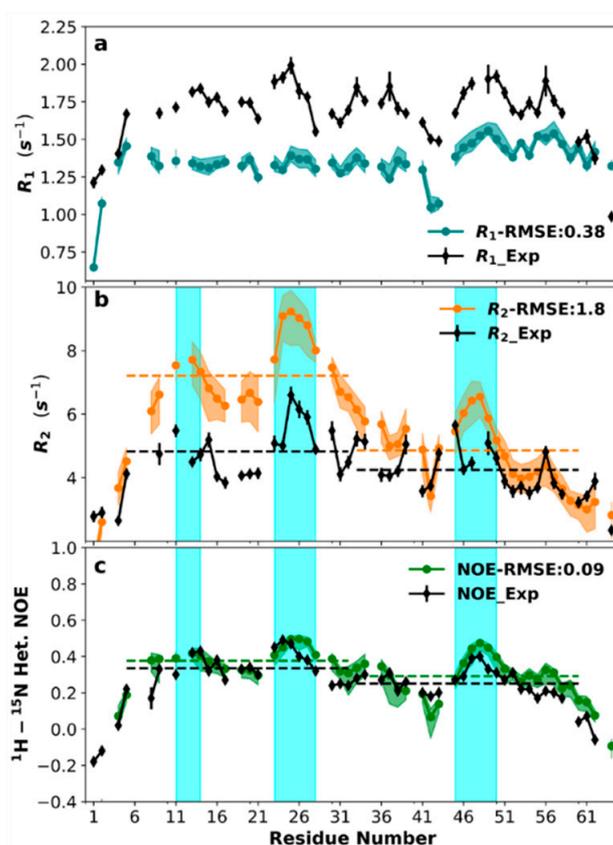
Figure 6. Backbone ^{15}N relaxation properties. (**a**) R_1, (**b**) R_2, and (**c**) nuclear Overhauser enhancements (NOE). Experimental data (pH 7.0) are in black; MD results are in blue, orange, and green, with shaded bands indicating 95% confidence intervals. Dashed lines indicate averages over residues 5–32 and 33–60; shaded cyan regions highlight residues (11–14, 23–28, and 46–50) with higher-than-average R_2s and NOEs. Gaps in the plots are due to prolines (and unresolved residues in the case of experimental data).

The five correlated segments each contain one or more transiently formed PPII stretches (Figure S8). The three most prevalent PPII stretches (residues 4–6, 10–12, and 27–29) identified above fall right into Boxes 1, 2, and 3. It is thus evident that SC-SC contacts contribute to the prevalence of the PPII stretches. This point is clearly illustrated by the contrast between Pro29 and Pro63. These prolines are both free from the direct influence of neighboring prolines or glycines and yet differ significantly in PPII frequencies (52% for Pro29 vs. 15% for Pro63). The most likely reason for the much higher PPII frequency of Pro29 is that it is next to a stretch of residues (Trp25 to Glu28) that form extensive interactions. Pro44 is close to a stretch of residues (Arg46 to Arg48) that form less extensive interactions and has an intermediate PPII frequency (36%).

The patterns of SC-BB contacts largely mirrored those of the SC-SC contacts. The SC-BB contacts segregated into the same five segments. Most frequent were contacts between adjacent residues, in particular hydrogen bonds between arginines and backbone carbonyls (e.g., Arg25 with the carbonyls of residues 24 and 25, as shown in Figure 5b inset #15) and between anionic residues and backbone amides. Still, nonlocal SC-BB hydrogen bonds occurred with significant frequencies in the N-half segments. For instance, Arg16 hydrogen bonded with the carbonyl of residue 25; Arg23 with residues 16, 17, and 18; and Arg25 with residue 13, as shown in Figure 5b inset #12; Arg16 with residue 11; and Asp20 with the backbone amide of residue 16, as shown in Figure 5b inset #15. There were relatively fewer nonlocal SC-BB hydrogen bonds in the C-half. All in all, the SC-BB hydrogen bonds contribute to the stability of the correlated segments identified by the SC-SC contacts and, at the same time, also directly influence the backbone ^{15}N relaxation and amide proton exchange rates.
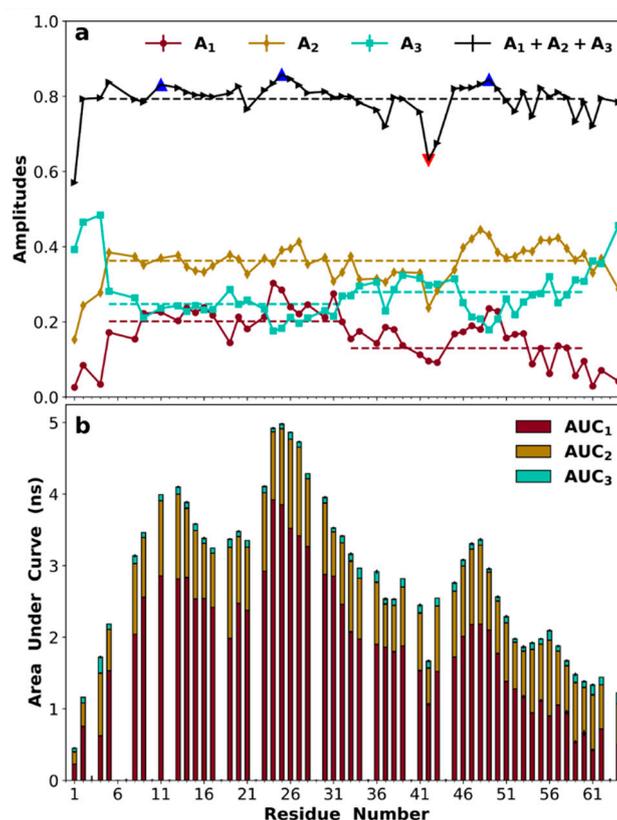
Figure 7. Amplitudes of backbone dynamics on three timescales. (**a**) Amplitudes A_1, A_2, A_3 for exponentials with time constants in the 7–17 ns, 1.5–3.5 ns, and 0.2–0.5 ns ranges. The sum of the three amplitudes is also shown, with blue triangles (at Asp11, Arg25, and Arg49) indicating local maxima and a red triangle (at Gly42) indicating a local minimum. Dashed lines show averages over either the entire sequence (residues 5–60) or the two halves (5–32 and 33–60); (**b**) area under the correlation function (AUC). The contributions of the three exponentials are indicated by bars in different colors. Fitting errors for amplitudes and propagated errors for AUC are plotted but are smaller than the size of the symbols.

3.7. Sequence-Specific Backbone Dynamics

In Figure 6 (black solid curves), we display the longitudinal and transverse relaxation rates (R_1 and R_2) and nuclear Overhauser enhancements (NOEs) of individual backbone ^{15}N sites at pH 7.0. At first glance, the relaxation properties are relatively uniform across the sequence, except for the extreme four residues at each terminus, with reduced R_1, R_2, and NOE. The resulting "bell" shape for R_2 has been suggested as arising from the residue–residue connectivity of a (denatured or disordered) polypeptide chain [54,55]. The average R_1 for residues 5–60 was 1.74 s^{-1}; the only pronounced deviation was a local minimum at residues Gly42 and Ala43.

Closer inspection revealed a small but systematic difference in R_2 between the N- and C-halves, with mean values for residues 5–32 and 33–60 at 4.76 and 4.17 s^{-1}, respectively (black dashed lines in Figure 6b). There was also a distinction in NOE between the two halves, with mean values at 0.34 for residues 5–32 and 0.25 for residues 33–60 (black dashed lines in Figure 6c). A t-test treating the N-half and C-half as two independent samples found the *p*-values for the differences in mean R_2 and in the mean NOE between the two halves to be both below 0.05 (Table S1), therefore indicating statistical significance. The overall low NOE values once again corroborate the lack of stable backbone structures. Still, the R_2 and NOE data together suggest that the N-half overall has larger amplitudes for motions on the slower (e.g., 10-ns) timescale but smaller amplitudes on the faster (sub-ns) timescale than the

C-half. Also worth noting are three stretches of residues, 11–14 and 23–28 in the N-half and 45–50 in the C-half (blue shading in Figure 6b,c), that had higher-than-average R_2s and NOEs.

The relaxation properties at pH 4.0 showed an even stronger disparity between the N- and C-halves (Figure S9). The mean R_2s in the two halves were 3.10 and 2.26 s^{-1}, and the mean NOEs had a wide gap, with values of 0.24 and 0.03 for the two halves. A distinction in the mean R_1 also became apparent between the N- and C-halves. The *p*-values for the differences in mean values were below 0.001 for all the three relaxation parameters, indicating a strong statistical significance. A likely consequence of the decrease in pH to 4.0 is the protonation of the three histidines (at positions 8, 48, and 59), which would amplify the charge imbalance in the C-half and thereby increase its disorder.

The MD simulations afforded the opportunity for a detailed interpretation of the NMR relaxation data. After evaluating the NH bond vector time-correlation functions, $C_{NH}(\tau)$, from the MD trajectories (at 20 ps time intervals) and fitting them to a sum of three exponentials, the resulting spectral densities were used, without any modification, to calculate the relaxation properties. The results were close to the experimental counterparts but with systematic underestimates in R_1 and overestimates in R_2 (colored solid curves). The root mean square errors (RMSEs) relative to the experimental data (excluding the extreme residue at each terminus) were 0.38 s^{-1} for R_1, 1.8 s^{-1} for R_2, and 0.09 for NOE. Importantly, the MD simulations recapitulated the sequence-dependent features of the experimental data, including: (1) the overall differences in R_2 and NOE between the N- and C-halves (as indicated by disparate mean values in the two halves, shown as colored dashed lines in Figure 6b,c); (2) the three stretches of residues showing the local maxima in R_2 and NOE (blue shading in Figure 6b,c); and (3) the local minimum in R_1 at residues 42 and 43 (Figure 6a).

3.8. Amplitudes of Backbone Dynamics on Different Timescales

Given the above qualitative agreement with the experimental data, we now report on the MD results for $C_{NH}(\tau)$, specifically their tri-exponential fits (with time constants τ_1, τ_2, and τ_3, ordered from large to small, and amplitudes A_1, A_2, and A_3; see Figure S10 for representative fits). It is important to note that we did not restrain the sum of the amplitudes, $A_{sum} = A_1 + A_2 + A_3$, to be 1. Implicitly, we assumed that the missing amplitude, $1 - A_{sum}$, represented an ultrafast decay that occurred before the first time point, 20 ps, at which we evaluated $C_{NH}(\tau)$. Indeed, adding an ultrafast decay component with an amplitude of $1 - A_{sum}$ and a time constant of $\tau_f = 10$ ps largely made up for some underestimates of the tri-exponential fits at short times (Figure S10 insets). Data between 0 and 20 ps would be required for a precise fit of the ultrafast component for each residue, but a global value of 10 ps for τ_f apparently worked well for most residues. The mean ± standard deviation of A_{sum} for non-terminal residues was 0.80 ± 0.04 (black dashed line in Figure 7a). In comparison, the order parameters for NH libration calculated after superimposing the peptide plane were 0.933 ± 0.001, implicating additional contributions (e.g., rapid fluctuations in the φ and ψ angles adjoining the peptide plane [53]) to the ultrafast decay. Interestingly, the three local maxima in A_{sum}, at Asp11, Arg25, and Arg49, apparently coincided with the residues showing higher-than-average R_2s and NOEs (triangles filled in blue in Figure 7a), whereas the minimum in A_{sum} at Gly42 (triangles filled in red in Figure 7a) coincided with the residues showing lower-than-average R_1s.

The means ± standard deviations of the three time constants were 11.5 ± 2.4, 2.4 ± 0.5, and 0.34 ± 0.06 ns for the non-terminal residues. The three exponentials with these time constants each contribute most to a different relaxation property, specifically, with the slow, intermediate, and fast timescales controlling R_2, R_1, and NOE, respectively. The amplitudes (A_2) associated with the intermediate time constant were nearly uniform along the sequence (at 0.36 ± 0.04), except for two very low values at Gly42 and Ala43 (Figure 7a). These results for A_2 largely explain the corresponding behavior of R_1 presented above, i.e., near constancy except for higher-than-average values for Gly42 and Ala43. On the other hand, A_1 and A_3 showed disparities between the N- and C-halves. In the N-half, the A_1 and A_3 averages were nearly the same, at 0.22 and 0.23, respectively. In the C-half, the A_1 average moved down to 0.15, while the A_3 average moved up to 0.27. Given the near constancy of A_2

and A_{sum}, the opposite movements of A_1 and A_3 were inevitable. The disparity in A_1 between the two halves explains the corresponding disparity in R_2, with lower A_1 values in the C-half accounting for the lower R_2s (i.e., weaker transverse relaxation) in that half. Likewise, the disparity in A_3 between the two halves explains the corresponding disparity in NOE, with higher A_3 values in the C-half accounting for the lower NOEs (i.e., higher flexibility) in that half.

The area under the $C_{NH}(\tau)$ curve (AUC) equals the spectral density, $J(0)$, at zero frequency, to which R_2 is particularly sensitive. The AUC values (and their contributions from the three exponentials) are displayed in Figure 7b. Two patterns are apparent (which are also true of the A_1 component). First, the N-half overall had higher AUCs than the C-half (with averages at 3.8 and 2.4 ns, respectively). Second, there were three local maxima at residues 11–13, 24–27, and 47–48. These were the same maxima as identified based on A_{sum} (Figure 7a), but were now much more conspicuous. They explain the higher-than-average R_2s of the involved residues (Figure 6b).

Ultimately, the higher amplitudes (A_1) for the slow timescale (and higher AUCs) of the N-half came from the more frequent SC-SC and SC-BB contacts in this half, in particular salt bridges, cation–π interactions, and SC-BB hydrogen bonds mediated by charged residues, resulting in correlated segmental motions. Indeed, the two most extensive interaction networks, centered around residues 24–27 and 46–49, were directly responsible for the local maxima in AUC and the corresponding local maxima in R_2 at these residues. In contrast, for Gly41 and Gly42, the absence of a sidechain not only allows them to access the left-handed side of the Ramachandran map (Figure S4), but also precludes them from forming any SC-SC or SC-BB contacts (Figure 5), resulting in much faster backbone dynamics.

3.9. Non-Uniform Amide Proton Exchange Rates along the Sequence

Amide proton exchange rates (k_{ex}; Figure 8a) further corroborated the presence of correlated segments suggested by the NMR relaxation experiments and MD simulations. The average k_{ex} of the N-half, 3.9 s^{-1}, was less than one third of the counterpart of the C-half, 12.4 s^{-1}.
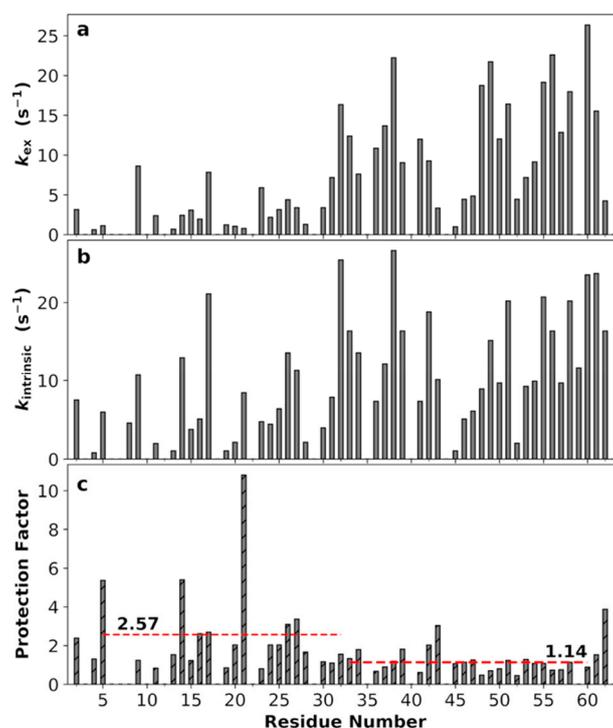

Figure 8. Backbone amide proton exchange rates (k_{ex}; pH 7.0), intrinsic exchange rates ($k_{intrinsic}$), and protection factors. (**a**) k_{ex}, (**b**) $k_{intrinsic}$, (**c**) protection factors. The mean protection factors for residues 5–32 and 33–60 are shown by red dashes.

k_{ex} is strongly dependent on the amino acid sequence [66]. We calculated the intrinsic exchange rates ($k_{intrinsic}$) from the sequence using the SPHERE server (Figure 8b) [65]. By taking the ratio $k_{intrinsic}/k_{ex}$, we obtained the protection factors (Figure 8c). Interestingly, for the C-half residues the k_{ex} values were all close to the $k_{intrinsic}$ values, indicating there is very little influence beyond the immediate amino acid sequence (average protection factor at 1.14). In contrast, for most of the N-half residues the protection factors were higher than 1, averaging 2.57. A t-test showed that the difference in mean protection factor between the N- and C-halves is statistically significant (*p*-value at 0.015; Table S1). This difference can be nicely explained by the disparities in the SC-SC and SC-BB contacts between the two halves. In particular, the two residues with the highest protection factors (residue 14 at 5.4 and residue 21 at 10.8) are located in the most correlated segment (residues 11–29, identified by extensive interactions and high R_2s and NOEs).

4. Discussion

By combining NMR, SAXS, and MD simulations, we have characterized the conformations and dynamics of ChiZ1-64 and delineated their linkage to the amino acid sequence. The conformations of ChiZ1-64 were diverse, with the only notable feature being high propensities of PPII stretches, especially in the N-half. Backbone ^{15}N relaxation experiments revealed non-uniform R_2s and NOEs along the sequence, with high values for residues 11–14 and 23–28. These or neighboring residues also have high protections factors for amide proton exchange. MD simulations recapitulated these observations and suggest that the reason for the non-uniform dynamics is the formation of correlated segments, which are stabilized by PPII stretches, salt bridges, cation–π interactions, and sidechain–backbone hydrogen bonds. Moreover, the extent of segmental correlation is sequence-dependent: segments where internal interactions are more prevalent manifest elevated "collective" motions and suppressed local motions.

Similar to ChiZ1-64, sequence-specific backbone dynamics have been reported on a number of other IDPs using NMR, some in combination with MD simulations [38,48–50,53,55,92,93]. Whereas stable secondary structures such as α-helices and β-hairpins can certainly lead to slow backbone dynamics [38,48,50,53], as demonstrated here for ChiZ1-64, interaction networks, in particular those mediated by charged and aromatic residues, can lead to the formation of correlated segments, which can have slow dynamics even when the backbone remains disordered. We propose the correlated segment as a defining feature for the conformation and dynamics of IDPs. Contact maps provide a way to identify correlated segments and characterize their stabilizing interactions. For example, it is interesting to investigate whether cation–π or other types of interactions contribute to the slow dynamics of two tryptophans in the C-terminal domain of the nucleoprotein of Sendai virus [38,48], or the precise interactions that may be responsible for the slow dynamics of two stretches of residues in HOX transcription factors [93]. π–π interactions between aromatic residues have been proposed to produce elevated R_2s in A1-LCD [56], though it remains to be seen whether explicit-solvent MD simulations can quantitatively explain the NMR data. The accumulation of this type of knowledge over a large number of IDPs will advance our understanding of how amino acid sequences of IDPs, through the formation of correlated segments, code for dynamics.

There is pressing need for the continued development of IDP force fields. The AMBER14SB/TIP4PD force field selected here based on benchmarking against SAXS profiles and chemical shifts also performed reasonably well for dynamic properties. Coincidentally, the same force field was also selected by Kämpf et al. [53] from comparison with backbone ^{15}N relaxation data. Still, for ChiZ1-64, the MD results had an apparent systematic underestimation of R_1 and an overestimation of R_2. The opposite deviations suggest an exaggeration of the longest timescale in the NH bond vector time-correlation functions. To test this idea, we scaled down the three time constants from the tri-exponential fits by a factor $1 + \tau_i /\tau_s$, with τ_s on the order of 10 ns, along with the addition of an ultrafast decay component with the time constant $\tau_f = 10$ ps and amplitude $1 - A_{sum}$ noted above. With $\tau_s = 16.75$ ns, the systematic errors were reduced for R_1 and almost eliminated for R_2; the NOE

calculations maintained a good agreement with the experimental data (Figure S11). Whether TIP4PD indeed makes ns dynamics too slow and, if so, how to improve this promising water model warrants further studies. It is also possible that the Langevin thermostat affects the dynamics of ChiZ1-64. Dynamic properties of IDPs have much to contribute in force field validation and improvements.

Although the functional role of ChiZ in Mtb cell division remains open, it may involve interactions with other divisome proteins, including FtsQ and FtsI [62]. Like ChiZ, both of the latter proteins contain disordered cytoplasmic regions high in charged amino acids. The interactions between all these IDRs may lead to fuzzy complexes. Moreover, these highly charged IDRs are also very likely to associate with the highly anionic Mtb membranes. The conformational and dynamic characterization of the ChiZ IDR in isolation done here will set the stage for studying these more complex systems. Given the disparity between the two halves of the ChiZ IDR, we expect the N-half to be more calcitrant and the C-half more adaptive in interacting with the various partners. In the full-length protein, the C-terminus of the IDR would be tethered to the membrane via its linkage to the transmembrane helix. The very C-terminal residues of the IDR would thus be restricted, though the rest of the C-half could still be free to sample its conformational space.

5. Conclusions

It is becoming evident that conformational dynamics play crucial roles in the functionality of IDPs. A number of experimental techniques can characterize IDP dynamics on different timescales, but in many cases the interpretations of such data are not straightforward. Computational methods such as MD simulations can help with the interpretations and with elucidating the link between sequence, dynamics, and function. In this study, we combined small angle X-ray scattering, NMR spectroscopy, and MD simulations to characterize a newly identified disordered region of ChiZ, ChiZ1-64. Overcoming the traditional limitations of MD simulations of IDPs with regard to force fields and sampling, we determined that the backbone dynamics of ChiZ1-64 are sequence-dependent, with several segments, mostly in the first 32 residues, showing high amplitudes in correlated motions. These correlated segmental dynamics are promoted by PPII formation and side chain–side chain and side chain–backbone interactions. The overexpression of ChiZ has been shown to halt Mtb cell division, potentially through interactions with FtsI and FtsQ, two other Mtb divisome proteins with disordered regions. Although we cannot absolutely determine ChiZ's mechanistic role in Mtb cell division, we hypothesize that sequence-dependent dynamics will be critical for this understanding. Potentially, the intrinsic fast dynamics of the C-half would allow it to readily adapt to binding partners, including the Mtb membrane and other divisome proteins, while the N-half rich in correlated segments may adopt nascent conformations that become stabilized by binding with partners. The characterization and methods illustrated here will also provide a framework for future studies to investigate the roles of dynamics in IDP functions.

Author Contributions: Conceptualization, A.H., C.A.E., T.A.C., and H.-X.Z.; methodology, A.H., C.A.E., T.A.C., and H.-X.Z.; software, A.H.; validation, A.H., C.A.E., T.A.C. and H.-X.Z.; formal analysis, A.H. and C.A.E.; investigation, A.H., C.A.E., T.A.C., and H.-X.Z.; resources, T.A.C. and H.-X.Z.; data curation, A.H., C.A.E., T.A.C., and H.-X.Z.; writing—original draft preparation, A.H. and H.-X.Z.; writing—review and editing, A.H., C.A.E., T.A.C., and H.-X.Z.; visualization, A.H., C.A.E., T.A.C., and H.-X.Z.; supervision, T.A.C. and H.-X.Z.; project administration, T.A.C. and H.-X.Z.; funding acquisition, T.A.C. and H.-X.Z. All authors have read and agreed to the published version of the manuscript.

References

1. Xue, B.; Dunker, A.K.; Uversky, V.N. Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **2012**, *30*, 137–149. [CrossRef] [PubMed]

2. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Boil.* **2014**, *16*, 18–29. [CrossRef] [PubMed]

3. Kjaergaard, M.; Kragelund, B.B. Functions of intrinsic disorder in transmembrane proteins. *Cell. Mol. Life Sci.* **2017**, *74*, 3205–3224. [CrossRef]

4. Babu, M.M.; van der Lee, R.; de Groot, N.S.; Gsponer, J. Intrinsically disordered proteins: Regulation and disease. *Curr. Opin. Struct. Boil.* **2011**, *21*, 432–440. [CrossRef] [PubMed]

5. Martinelli, A.H.S.; Lopes, F.C.; John, E.B.D.O.; Carlini, C.R.; Ligabue-Braun, R. Modulation of Disordered Proteins with a Focus on Neurodegenerative Diseases and Other Pathologies. *Int. J. Mol. Sci.* **2019**, *20*, 1322. [CrossRef]

6. Cook, E.C.; Sahu, D.; Bastidas, M.; Showalter, S.A. Solution Ensemble of the C-Terminal Domain from the Transcription Factor Pdx1 Resembles an Excluded Volume Polymer. *J. Phys. Chem. B* **2018**, *123*, 106–116. [CrossRef]

7. Morgan, J.L.; Jensen, M.R.; Ozenne, V.; Blackledge, M.; Barbar, E. The LC8 Recognition Motif Preferentially Samples Polyproline II Structure in Its Free State. *Biochemistry* **2017**, *56*, 4656–4666. [CrossRef]

8. Schneider, R.; Maurin, D.; Communie, G.; Kragelj, J.; Hansen, D.F.; Ruigrok, R.W.H.; Jensen, M.R.; Blackledge, M. Visualizing the Molecular Recognition Trajectory of an Intrinsically Disordered Protein Using Multinuclear Relaxation Dispersion NMR. *J. Am. Chem. Soc.* **2015**, *137*, 1220–1229. [CrossRef]

9. Arai, M.; Sugase, K.; Dyson, H.J.; Wright, P.E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9614–9619. [CrossRef]

10. Karlsson, E.; Andersson, E.; Dogan, J.; Gianni, S.; Jemth, P.; Camilloni, C. A structurally heterogeneous transition state underlies coupled binding and folding of disordered proteins. *J. Boil. Chem.* **2018**, *294*, 1230–1239. [CrossRef]

11. Ou, L.; Matthews, M.; Pang, X.; Zhou, H.-X. The dock-and-coalesce mechanism for the association of a WASP disordered region with the Cdc42 GTPase. *FEBS J.* **2017**, *284*, 3381–3391. [CrossRef] [PubMed]

12. Berlow, R.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Role of Backbone Dynamics in Modulating the Interactions of Disordered Ligands with the TAZ1 Domain of the CREB-Binding Protein. *Biochemistry* **2019**, *58*, 1354–1362. [CrossRef] [PubMed]

13. Zhou, H.-X. Intrinsic disorder: Signaling via highly specific but short-lived association. *Trends Biochem. Sci.* **2011**, *37*, 43–48. [CrossRef] [PubMed]

14. Borgia, A.; Borgia, M.B.; Bugge, K.; Kissling, V.M.; Heidarsson, P.O.; Fernandes, C.B.; Sottini, A.; Soranno, A.; Buholzer, K.J.; Nettels, D.; et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61–66. [CrossRef]

15. Pang, X.; Zhou, H.-X. Rate Constants and Mechanisms of Protein-Ligand Binding. *Annu. Rev. Biophys.* **2017**, *46*, 105–130. [CrossRef]

16. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Bioinform.* **2000**, *41*, 415–427. [CrossRef]

17. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13392–13397. [CrossRef]

18. Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691–1741. [CrossRef]

19. Baul, U.; Chakraborty, D.; Mugnai, M.L.; Straub, J.E.; Thirumalai, D. Sequence Effects on Size, Shape, and Structural Heterogeneity in Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2019**, *123*, 3462–3474. [CrossRef]

20. Kikhney, A.G.; Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570–2577. [CrossRef]

21. Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16155–16160. [CrossRef]

22. Soranno, A.; Stucki-Buchli, B.; Nettels, D.; Cheng, R.R.; Müller-Späth, S.; Pfeil, S.H.; Hoffmann, A.; Lipman, E.A.; Makarov, D.E.; Schuler, B. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17800–17806. [CrossRef] [PubMed]

23. Soranno, A.; Holla, A.; Dingfelder, F.; Nettels, D.; Makarov, D.E.; Schuler, B. Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1833–E1839. [CrossRef]

24. Jensen, M.R.; Zweckstetter, M.; Huang, J.-R.; Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* **2014**, *114*, 6632–6660. [CrossRef] [PubMed]

25. Best, R.B. Computational, and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Boil.* **2017**, *42*, 147–154. [CrossRef]

26. Banks, A.; Qin, S.; Weiss, K.; Stanley, C.; Zhou, H.-X. Intrinsically Disordered Protein Exhibits Both Compaction and Expansion under Macromolecular Crowding. *Biophys. J.* **2018**, *114*, 1067–1079. [CrossRef]

27. Sturzenegger, F.; Zosel, F.; Holmstrom, E.D.; Buholzer, K.J.; Makarov, D.E.; Nettels, D.; Schuler, B. Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat. Commun.* **2018**, *9*, 4708. [CrossRef] [PubMed]

28. Kim, J.-Y.; Meng, F.; Yoo, J.; Chung, H.S. Diffusion-limited association of disordered protein by non-native electrostatic interactions. *Nat. Commun.* **2018**, *9*, 4707. [CrossRef]

29. Dyson, H.J.; Wright, P.E. Equilibrium NMR studies of unfolded and partially folded proteins. *Nat. Genet.* **1998**, *5*, 499–503. [CrossRef]

30. Marsh, J.A.; Singh, V.K.; Jia, Z.; Forman-Kay, J.D. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci.* **2006**, *15*, 2795–2804. [CrossRef]

31. Mittag, T.; Forman-Kay, J.D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Boil.* **2007**, *17*, 3–14. [CrossRef] [PubMed]

32. Dass, R.; Corlianò, E.; Mulder, F.A.A. Measurement of Very Fast Exchange Rates of Individual Amide Protons in Proteins by NMR Spectroscopy. *ChemPhysChem* **2018**, *20*, 231–235. [CrossRef] [PubMed]

33. McAllister, R.G.; Konermann, L. Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective. *Biochemistry* **2015**, *54*, 2683–2692. [CrossRef] [PubMed]

34. Croke, R.L.; Sallum, C.O.; Watson, E.; Watt, E.D.; Alexandrescu, A.T. Hydrogen exchange of monomeric α-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in Escherichia coli. *Protein Sci.* **2008**, *17*, 1434–1445. [CrossRef]

35. Palmer, A.G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **2004**, *104*, 3623–3640. [CrossRef]

36. Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559. [CrossRef]

37. Khan, S.N.; Charlier, C.; Augustyniak, R.; Salvi, N.; Déjean, V.; Bodenhausen, G.; Lequin, O.; Pelupessy, P.; Ferrage, F. Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophys. J.* **2015**, *109*, 988–999. [CrossRef]

38. Abyzov, A.; Salvi, N.; Schneider, R.; Maurin, D.; Ruigrok, R.W.H.; Jensen, M.R.; Blackledge, M. Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation. *J. Am. Chem. Soc.* **2016**, *138*, 6240–6251. [CrossRef]

39. Gill, M.; Byrd, R.A.; Palmer, A.G. Dynamics of GCN4 facilitate DNA interaction: A model-free analysis of an intrinsically disordered region. *Phys. Chem. Chem. Phys.* **2016**, *18*, 5839–5849. [CrossRef]

40. Nettels, D.; Müller-Späth, S.; Küster, F.; Hofmann, H.; Haenni, M.; Rüegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; et al. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20740–20745. [CrossRef]

41. Rauscher, S.; Gapsys, V.; Gajda, M.J.; Zweckstetter, M.; de Groot, B.L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524. [CrossRef]

42. Best, R.B.; Zheng, W.; Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124. [CrossRef]

43. Piana, S.; Donchev, A.G.; Robustelli, P.; Shaw, D.E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123. [CrossRef] [PubMed]

44. Robustelli, P.; Piana, S.; Shaw, D.E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4758–E4766. [CrossRef]

45. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B.L.; Grubmüller, H.; MacKerell, A.D. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2016**, *14*, 71–73. [CrossRef] [PubMed]

46. Xue, Y.; Skrynnikov, N.R. Motion of a Disordered Polypeptide Chain as Studied by Paramagnetic Relaxation Enhancements,15N Relaxation, and Molecular Dynamics Simulations: How Fast Is Segmental Diffusion in Denatured Ubiquitin? *J. Am. Chem. Soc.* **2011**, *133*, 14614–14628. [CrossRef] [PubMed]

47. Salvi, N.; Abyzov, A.; Blackledge, M. Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation. *J. Phys. Chem. Lett.* **2016**, *7*, 2483–2489. [CrossRef]

48. Salvi, N.; Abyzov, A.; Blackledge, M. Analytical Description of NMR Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. *Angew. Chem. Int. Ed.* **2017**, *56*, 14020–14024. [CrossRef]

49. Rezaei-Ghaleh, N.; Parigi, G.; Soranno, A.; Holla, A.; Becker, S.; Schuler, B.; Luchinat, C.; Zweckstetter, M. Local and Global Dynamics in Intrinsically Disordered Synuclein. *Angew. Chem. Int. Ed.* **2018**, *57*, 15262–15266. [CrossRef]

50. Rezaei-Ghaleh, N.; Parigi, G.; Zweckstetter, M. Reorientational Dynamics of Amyloid-β from NMR Spin Relaxation and Molecular Simulation. *J. Phys. Chem. Lett.* **2019**, *10*, 3369–3375. [CrossRef]

51. Salvi, N.; Abyzov, A.; Blackledge, M. Solvent-dependent segmental dynamics in intrinsically disordered proteins. *Sci. Adv.* **2019**, *5*, eaax2348. [CrossRef] [PubMed]

52. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef] [PubMed]

53. Kämpf, K.; Izmailov, S.A.; Rabdano, S.O.; Groves, A.T.; Podkorytov, I.S.; Skrynnikov, N.R. What Drives 15N Spin Relaxation in Disordered Proteins? Combined NMR/MD Study of the H4 Histone Tail. *Biophys. J.* **2018**, *115*, 2348–2367. [CrossRef] [PubMed]

54. Schwalbe, H.; Fiebig, K.M.; Buck, M.; Jones, J.A.; Grimshaw, S.B.; Spencer, A.; Glaser, S.J.; Smith, L.J.; Dobson, C.M. Structural and Dynamical Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8 M Urea. *Biochemistry* **1997**, *36*, 8977–8991. [CrossRef]

55. Klein-Seetharaman, J. Long-Range Interactions Within a Nonnative Protein. *Science* **2002**, *295*, 1719–1722. [CrossRef]

56. Martin, E.W.; Holehouse, A.S.; Peran, I.; Farag, M.; Incicco, J.J.; Bremer, A.; Grace, C.R.; Soranno, A.; Pappu, R.V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694–699. [CrossRef]

57. Kieser, K.J.; Rubin, E.J. How sisters grow apart: Mycobacterial growth and division. *Nat. Rev. Genet.* **2014**, *12*, 550–562. [CrossRef]

58. Das, N.; Dai, J.; Hung, I.; Rajagopalan, M.R.; Zhou, H.-X.; Cross, T.A. Structure of CrgA, a cell division structural and regulatory protein from Mycobacterium tuberculosis, in lipid bilayers. *Proc. Natl. Acad. Sci. USA* **2014**, *112*, E119–E126. [CrossRef] [PubMed]

59. Chauhan, A.; Lofton, H.; Maloney, E.; Moore, J.; Fol, M.; Madiraju, M.V.V.S.; Rajagopalan, M. Interference of Mycobacterium tuberculosis cell division by Rv2719c, a cell wall hydrolase. *Mol. Microbiol.* **2006**, *62*, 132–147. [CrossRef] [PubMed]

60. Escobar, C.A.; Cross, T.A. False positives in using the zymogram assay for identification of peptidoglycan hydrolases. *Anal. Biochem.* **2018**, *543*, 162–166. [CrossRef] [PubMed]

61. Chauhan, A.; Madiraju, M.V.V.S.; Fol, M.; Lofton, H.; Maloney, E.; Reynolds, R.; Rajagopalan, M. Mycobacterium tuberculosis Cells Growing in Macrophages Are Filamentous and Deficient in FtsZ Rings. *J. Bacteriol.* **2006**, *188*, 1856–1865. [CrossRef]

62. Vadrevu, I.S.; Lofton, H.; Sarva, K.; Blasczyk, E.; Plocinska, R.; Chinnaswamy, J.; Madiraju, M.; Rajagopalan, M. ChiZ levels modulate cell division process in mycobacteria. *Tuberculosis* **2011**, *91*, S128–S135. [CrossRef] [PubMed]

63. Franke, D.; Petoukhov, M.V.; Konarev, P.V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.D.T.; Kikhney, A.G.; Hajizadeh, N.R.; Franklin, J.M.; Jeffries, C.M.; et al. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225. [CrossRef] [PubMed]

64. Hwang, T.-L.; van Zijl, P.C.; Mori, S. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **1998**, *11*, 221–226. [CrossRef] [PubMed]

65. Zhang, Y.-Z. Protein and Peptide Structure and Interactions Studied by Hydrogen Exchanger and NMR. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA, USA, 1995.

66. Bai, Y.; Milne, J.S.; Mayne, L.; Englander, S.W. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct. Funct. Bioinform.* **1993**, *17*, 75–86. [CrossRef]

67. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 1950–1958. [CrossRef]

68. Debiec, K.T.; Cerutti, D.S.; Baker, L.R.; Gronenborn, A.M.; Case, D.A.; Chong, L.T. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12*, 3926–3947. [CrossRef]

69. Case, D.A.; Betz, R.M.; Cerutti, D.S.; Cheatham, T.E.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; et al. *AMBER 2016*; University of California: Oakland, CA, USA, 2016.

70. Case, D.A.; Ben-Shalom, I.Y.; Brozell, S.R.; Cerutti, D.S.; Cheatham, T.E.; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Gilson, M.K.; et al. *AMBER 2018*; University of California: Oakland, CA, USA, 2018.

71. Abraham, M.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25. [CrossRef]

72. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [CrossRef]

73. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865. [CrossRef]

74. Crowley, M.F.; Williamson, M.J.; Walker, R.C. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. *Int. J. Quantum Chem.* **2009**, *109*, 3767–3772. [CrossRef]

75. Salomon-Ferrer, R.; Götz, A.W.; Poole, D.; le Grand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888. [CrossRef] [PubMed]

76. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]

77. Schneidman-Duhovny, D.; Hammel, M.; Sali, A. FoXS: A web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **2010**, *38*, W540–W544. [CrossRef] [PubMed]

78. Henriques, J.; Arleth, L.; Lindorff-Larsen, K.; Skepö, M. On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Boil.* **2018**, *430*, 2521–2539. [CrossRef] [PubMed]

79. Han, B.; Liu, Y.; Ginzinger, S.W.; Wishart, D.S. SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR* **2011**, *50*, 43–57. [CrossRef]

80. Tamiola, K.; Acar, B.; Mulder, F.A.A. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132*, 18000–18003. [CrossRef]

81. Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [CrossRef]

82. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637. [CrossRef]

83. Mansiaux, Y.; Joseph, A.P.; Gelly, J.-C.; de Brevern, A.G. Assignment of PolyProline II Conformation and Analysis of Sequence—Structure Relationship. *PLoS ONE* **2011**, *6*, e18401. [CrossRef]

84. Mu, Y.; Nguyen, P.H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins Struct. Funct. Bioinform.* **2004**, *58*, 45–52. [CrossRef] [PubMed]

85. Altis, A.; Hegger, R.; Nguyen, P.H.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 244111. [CrossRef]

86. Best, R.B.; Hummer, G.; Eaton, W.A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17874–17879. [CrossRef]

87. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.; Klein, C.; Swails, J.M.; Hernández, C.X.; Schwantes, C.R.; Wang, L.-P.; Lane, T.; Pande, V.S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532. [CrossRef] [PubMed]

88. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids. *Biochim. Biophys. Acta Proteins Proteom.* **2010**, *1804*, 996–1010. [CrossRef] [PubMed]

89. Mészáros, B.; Erdos, G.; Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [CrossRef] [PubMed]

90. Dosztányi, Z.; Mészáros, B.; Simon, I. ANCHOR: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **2009**, *25*, 2745–2746. [CrossRef] [PubMed]

91. Bernadó, P.; Blackledge, M. A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering. *Biophys. J.* **2009**, *97*, 2839–2845. [CrossRef] [PubMed]

92. Cho, M.-K.; Kim, H.-Y.; Bernadó, P.; Fernández, C.O.; Blackledge, M.; Zweckstetter, M. Amino Acid Bulkiness Defines the Local Conformations and Dynamics of Natively Unfolded α-Synuclein and Tau. *J. Am. Chem. Soc.* **2007**, *129*, 3032–3033. [CrossRef]

93. Maiti, S.; Acharya, B.; Boorla, V.S.; Manna, B.; Ghosh, A.; De, S. Dynamic Studies on Intrinsically Disordered Regions of Two Paralogous Transcription Factors Reveal Rigid Segments with Important Biological Functions. *J. Mol. Boil.* **2019**, *431*, 1353–1369. [CrossRef]