Supplementary material

Assessing the direct binding of Ark-like E3 RING ligases to ubiquitin and its implication on their protein interaction network.

Dimitris G. Mintis [1], Anastasia Chasapi [2], Konstantinos Poulas [3,4], George Lagoumintzis [3,4,*] and Christos T. Chasapis [5,6,*]

[1] Laboratory of Statistical Thermodynamics and Macromolecules, Department of Chemical Engineering, University of Patras & FORTH/ICE-HT, Patras GR26504, Greece; dimitris.g.mintis@gmail.com
[2] Biological Computation & Process Lab, Chemical Process & Energy Resources Institute, Centre for Research & Technology Hellas, GR-57001 Thessaloniki, Greece; chasapi@certh.gr
[3] Laboratory of Molecular Biology and Immunology, Department of Pharmacy, University of Patras, 26504 Patras, Greece; kpoulas@upatras.gr (K.P.); glagoum@upatras.gr (G.L.)
[4] Institute of Research and Innovation-IRIS, Patras Science Park SA, Stadiou, Platani, Rio, 26504 Patras, Greece
[5] NMR Center, Instrumental Analysis Laboratory, School of Natural Sciences, University of Patras, 26504 Patras, Greece
[6] Institute of Chemical Engineering Sciences, Foundation for Research and Technology, Hellas (FORTH/ICE-HT), 26504 Patras, Greece

***** Correspondence: cchasapis@upatras.gr (C.T.C.); glagoum@upatras.gr (G.L.); Tel.: +30-2610-996-261 (C.T.C.) ;+30-2610-996-312 (G.L.)

1. Simulation Details

All MD simulations conducted in this work were performed using the GROMACS software (version 2016.3) (1). A similar protocol was followed for all simulations performed in this work. The protein structure was initially assigned the AMBER-99SB-STAR-ILDNP molecular mechanics force field supported in GROMACS. Zinc ions were also assigned the AMBER-99SB-STAR-ILDNP molecular mechanics force field. The protein complex was then immersed in a fully periodic cubic box with sufficiently large dimensions to avoid periodic artifacts from occurring. The cubic cell was then filled with explicit water molecules, based on the TIP3P water model, and the appropriate number of counter ions were inserted to ensure the electroneutrality of the simulated system. Energy minimization was then followed by employing the steepest descent algorithm to relax the system to the closest local energy minimum with the energy minimization adjusted to be converged when the maximum force was smaller than 30 kJ mol^{-1} nm^{-1}. Equilibration of the protein complex was then followed by carrying out a short simulation of 1 ns in the nVT statistical ensemble (n denotes the total number of interacting units in the simulation cell) using the leap-frog integration algorithm to solve Newton's equations of motion with integration time step of 2 fs whereas the Nosé-Hoover (2, 3) thermostat was utilized to couple temperature at 300 K with the time constant adjusted at 2.5 ps. All bonds were constrained according to the LINCS (4) protocol. Electrostatic interactions were treated with the Particle Mesh Ewald (5) (PME) method with a real-space cutoff of 1.2 nm, and Van der Waals interactions were treated with a cutoff of 1.2 nm. Production MD runs were conducted in the nPT statistical ensemble for 150 ns with a timestep of 2 fs at constant temperature 300 K and constant pressure 1 atm by employing the Nosé-Hoover thermostat and Parrinello-Rahman (6) barostat with the time constant for coupling adjusted at 2.5 ps and 1 ps accordingly.

2. Analysis of MD trajectories

Root Mean Square Deviation (RMSD) – The RMSD of certain atoms in a molecule with respect to a reference structure $\mathbf{r}^{\text{ref}}$ is calculated as:

$$\text{RMSD}(t) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \left| \mathbf{r}_i(t) - \mathbf{r}_i^{\text{ref}} \right|^2 \right]^{0.5} \quad (1)$$

Where M is the sum of atom's masses, and $\mathbf{r}_i(t)$ denotes the position vector of atom i at time t.

Root Mean Square Fluctuation (RMSF) – The RMSF describes the deviation between particle's position from some reference position and is calculated as:

$$\text{RMSF}_i = \left[\frac{1}{T} \sum_{t_j=1}^{T} \left| \mathbf{r}_i(t_j) - \mathbf{r}_i^{\text{ref}} \right|^2 \right]^{0.5} \quad (2)$$

Where T is the time (frames) over which someone is interested to average.

Radis of Gyration and End-to-End distance – To measure the compactness and size of the polypeptide chain, the square root of the mean-square radius of gyration $\left\langle R_{\mathrm{g}}^2 \right\rangle^{0.5}$ and end-to-end distance $\left\langle R_{\mathrm{ee}}^2 \right\rangle^{0.5}$ are calculated, defined as:

$$\left\langle R_{\mathrm{g}}^2 \right\rangle^{0.5} = \frac{1}{N}\left\langle \sum_{i=1}^{N} \left(\mathbf{r}_i - \mathbf{r}_{\mathrm{c.m.}}\right)^2 \right\rangle^{0.5} \tag{3}$$

$$\left\langle R_{\mathrm{ee}}^2 \right\rangle^{0.5} = \left\langle \left(\mathbf{r}_N - \mathbf{r}_1\right)^2 \right\rangle^{0.5} \tag{4}$$

where $\langle ... \rangle$ denotes the average over all possible configurations, $\mathbf{r}_N$ denotes the position vector of the last atom in the chain backbone and $\mathbf{r}_{\mathrm{c.m.}}$ denotes the position vector of the polypeptide's chain center of mass defined as:

$$\mathbf{r}_{\mathrm{c.m.}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{r}_i \tag{5}$$

Local (internal) dynamics analysis – The local dynamics of the protein is assessed by measuring the order parameter S^2 based on the model-free approach of Lipari and Szabo (6) by averaging the normalized time-correlation functions, $C(t)$, of the NH bond vectors in a protein after overall rotation is removed (and, after it has reached a plateau) as presented in below equation:

$$C(t) = \left\langle \frac{3}{2}\cos^2 \theta_{t'+t} - \frac{1}{2} \right\rangle_{t'} \tag{6}$$

where $\theta_{t'+t}$ is the NH bond angle between times t' and $t'+t$. The angular brackets refer to the equilibrium ensemble average. This expression yields to the order parameter S^2.

Secondary structure analysis – The DSSP (Dictionary of Secondary Structure for Proteins) algorithm (7) was employed to analyze the secondary structure of the polypeptide chain over the time of the simulation. The DSSP output assigns a letter that corresponds to the secondary structure of each residue, such as H for alpha helix, B for residue in isolated beta-bridge, E for extended strand, G for 3-helix, I for 5 helix, T for a hydrogen-bonded turn, and S for a bend.

Solvent accessible surface area – The solvent accessible surface area (SASA) is computed to characterize the hydration behavior of the polypeptide chain. A sphere of water (of the solvent) with a radius of 1.4 Å probes (is rolled over) the surface of a molecule, which is represented by a set of interlocking spheres with the appropriate van der Waals radii assigned to each atom to calculate the surface area of the molecule (solute) accessible to the solvent (8).

3. Force Field validation based on the comparison of MDs for Arkadia RING$_{927\text{-}994}$ domain against its NMR study.

Comparison of the RMSD – **Figure S1A** presents the comparison of the RMSD of the backbone atoms computed for the residues Asp937-Ala988 of the Arkadia RING$_{927\text{-}994}$ domain between the experimentally obtained value (NMR 31 models, pdb id: 2KIZ) (9) and the value obtained from the three different molecular mechanics force fields. The short dashed red line and the red point indicates the mean value of the reference (experimental) value and the statistical error, respectively. **Figure** S1B presents the comparison of the RMSD of the heavy atoms for the residues Asp937-Ala988 of the Arkadia RING$_{927\text{-}994}$ domain.

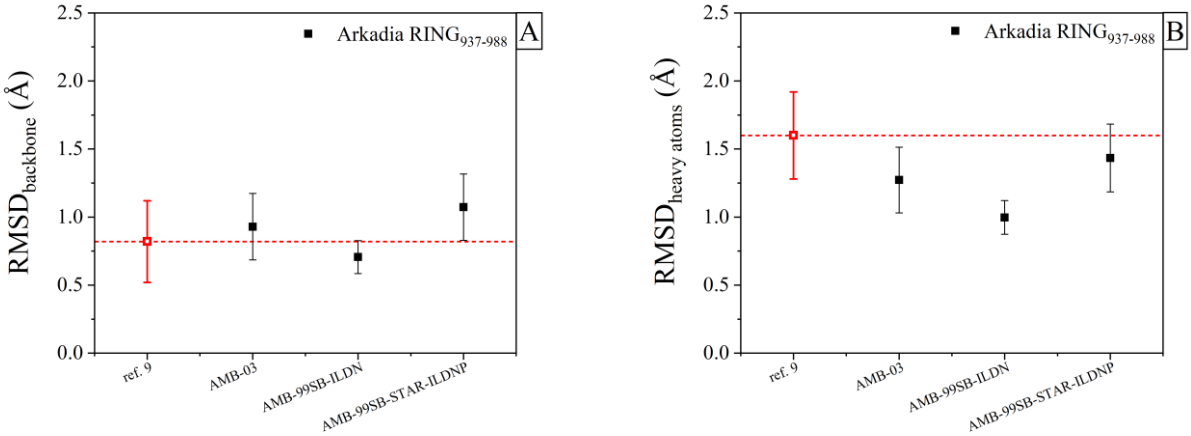


Figure S1. The comparison of the mean value of the RMSD of the A) backbone and B) heavy atoms of the Arkadia RING$_{927\text{-}994}$ domain (for residues 937-988) with statistical errors as predicted from the three different molecular mechanics force fields and the experiment.(9)

The time evolution of the RMSD as computed for the heavy atoms and the backbone atoms for the reference/experimental value is given in **Figure S2**. To further assess the convergences of the simulations, the time evolution of the $\left\langle R_g^2 \right\rangle^{0.5}$ and $\left\langle R_{ee}^2 \right\rangle^{0.5}$ of the protein complex are computed, as shown in **Figure S3**.

In addition, a heat map that indicates the deviation from the value of the RMSD of the Arkadia RING$_{927\text{-}994}$ domain as simulation evolves in time for each molecular mechanics force field studied in this work is given in **Figure S4**.

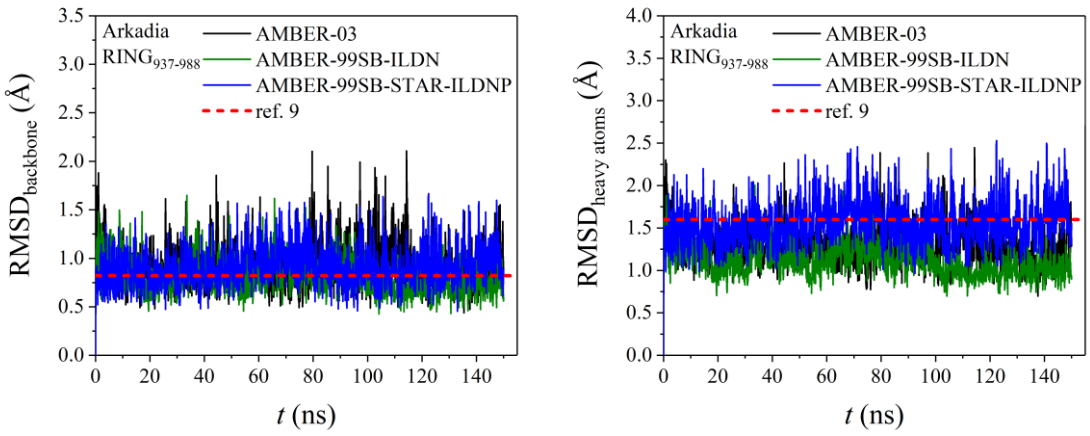


Figure S2. Time evolution of the RMSD of the backbone atoms and of the heavy atoms of the Arkadia $RING_{927-994}$ (for residues 937-988) as predicted from the three different molecular mechanics force fields with respect to previous NMR-based study(9) indicated by the short dashed red line.

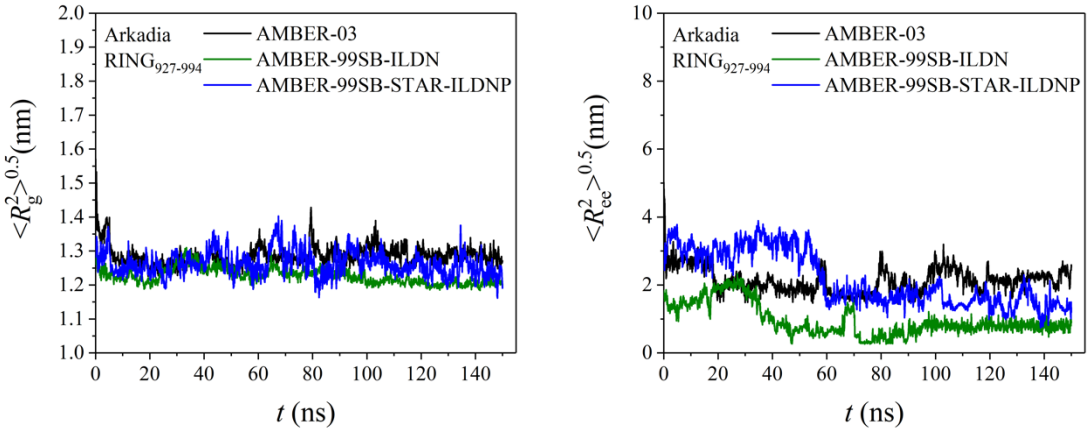


Figure S3. Time evolution of the $\left\langle R_g^2 \right\rangle^{0.5}$ and $\left\langle R_{ee}^2 \right\rangle^{0.5}$ of the Arkadia $RING_{927-994}$ based on three different molecular mechanics force fields.

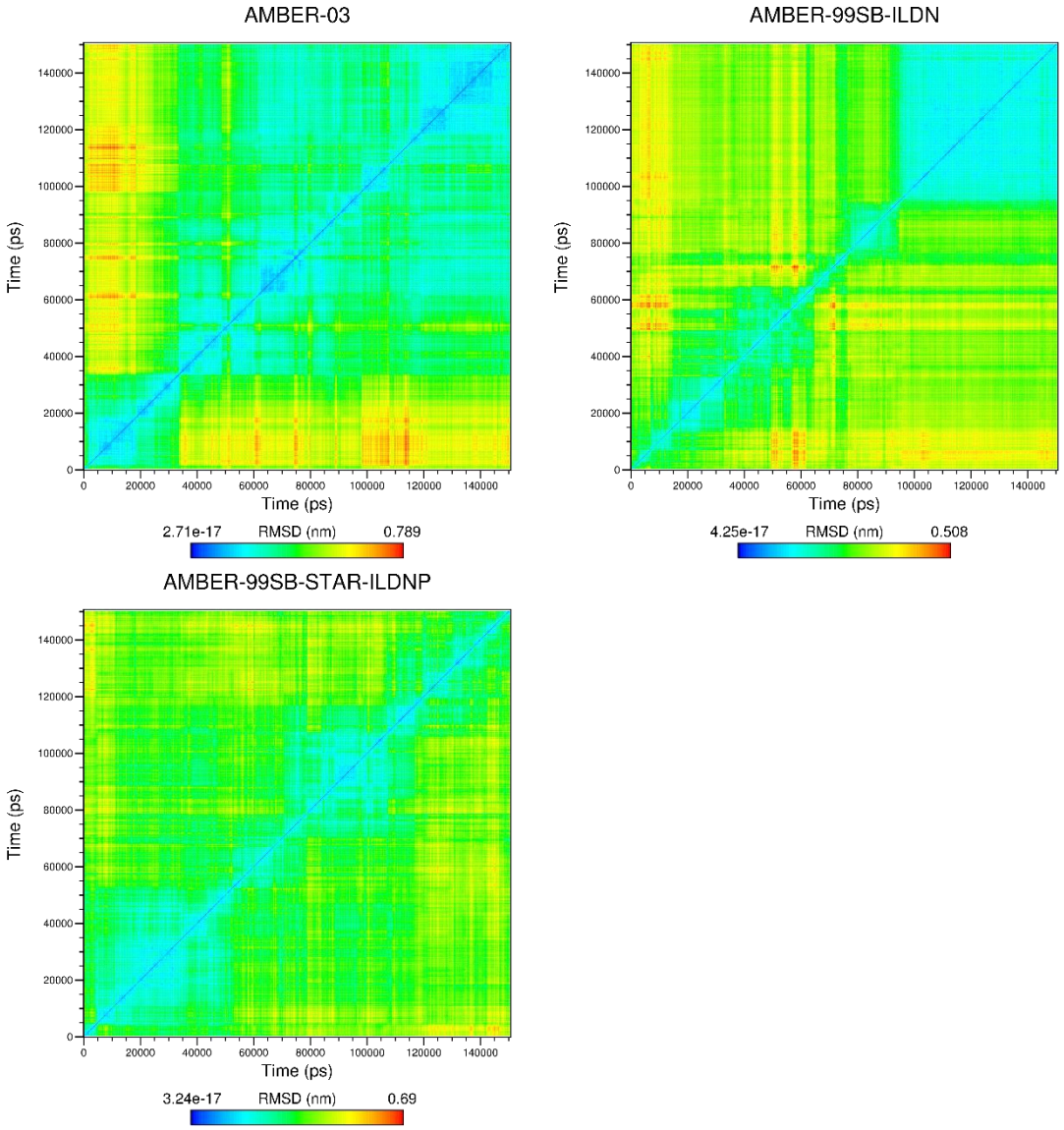


Figure S4. Heat map of the deviation of the value of the RMSD for the Arkadia $RING_{927-994}$ as simulation evolves in time for three different systems assigned with a different molecular mechanics force field.

Comparison of the secondary structure – **Figure S5**, presents the secondary structure of the Arkadia $RING_{927-994}$ domain as a function of the simulation time based on the AMBER-99SB-STAR-ILDNP molecular mechanics force field, and the right panel shows the secondary structure elements as obtained from the previous experimental study (9). In **Figure S6** and **Figure S7**, relevant information are given based on the AMBER-03 and AMBER-99SB-ILDN molecular mechanics force field, respectively.

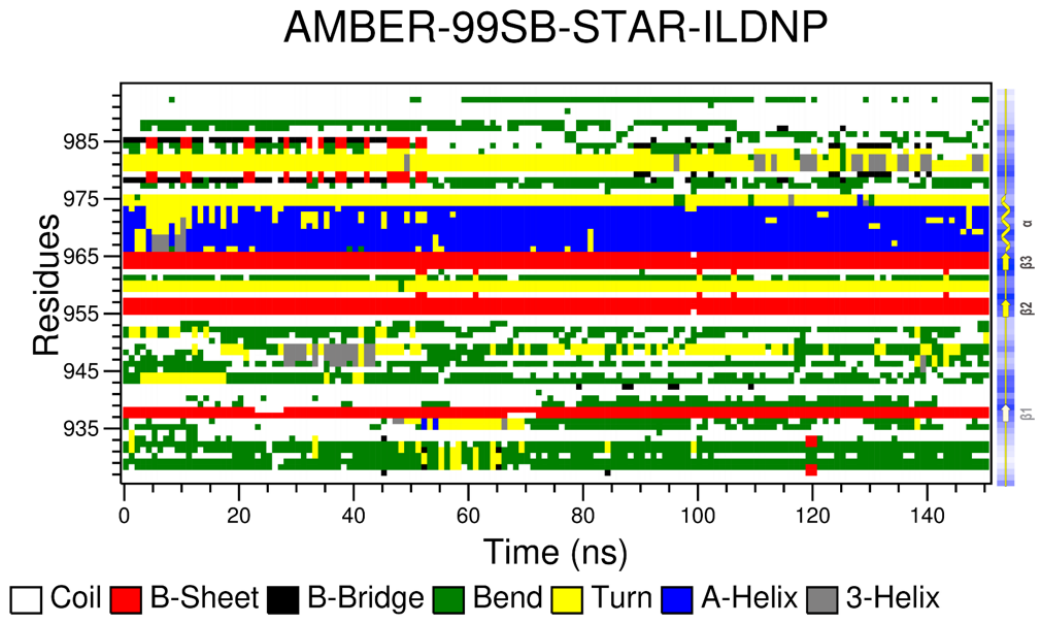


Figure S5. Secondary structure with simulation time predicted based on the AMBER-99SB-STAR-ILDNP molecular mechanics force field. The right panel is obtained from the previous NMR-based study (9) and indicates the secondary structure elements with respect to the residues of the Arkadia $RING_{927-994}$.

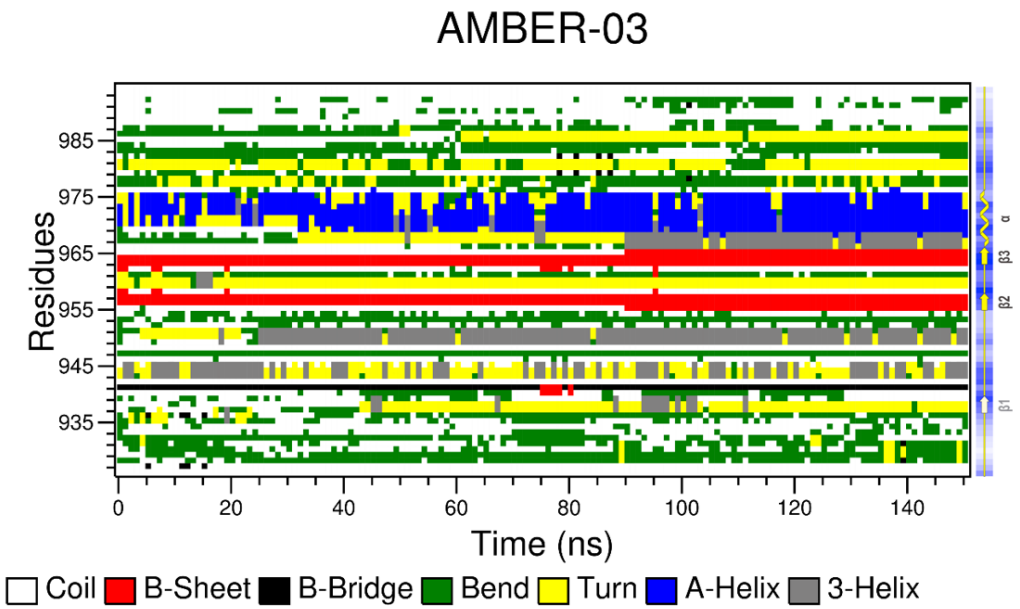


Figure S6. Secondary structure with simulation time predicted based on the AMBER-03 molecular mechanics force field. The right panel is obtained from a previous NMR-based study (9) and indicates the secondary structure elements with respect to the residues of the Arkadia $RING_{927-994}$.

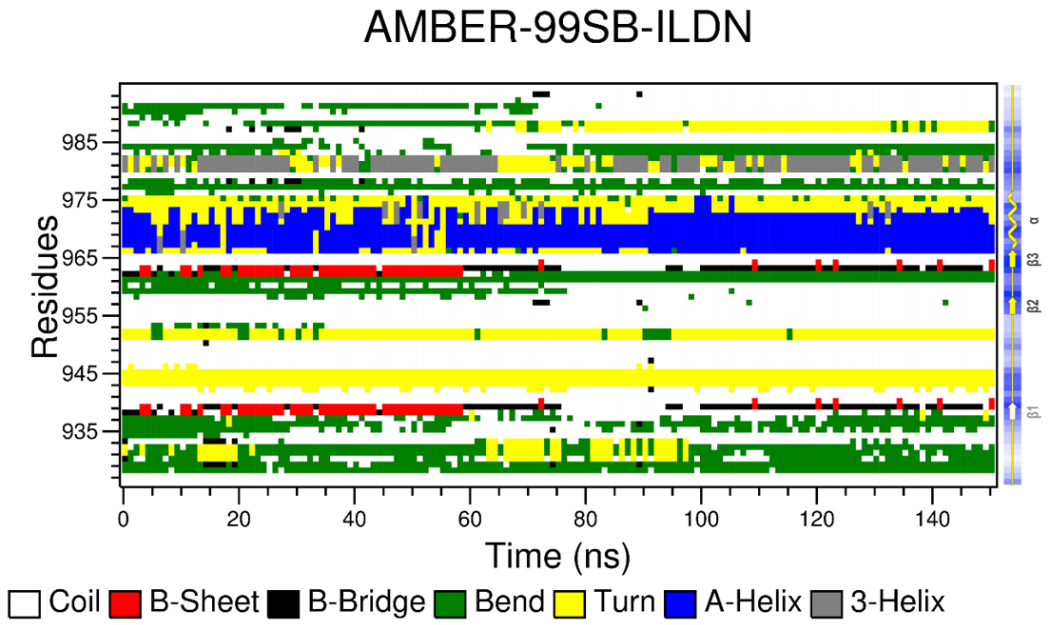


Figure S7. Secondary structure with simulation time predicted based on the AMBER-99SB-ILDN molecular mechanics force field. The right panel is obtained from a previous NMR-based study (9) and indicates the secondary structure elements in respect to the residues of the Arkadia $RING_{927-994}$

Comparison of the solvation – The SASA with respect to the residues of the Arkadia $RING_{927-994}$ domain based on the AMBER-99SB-STAR-ILDNP is presented in **Figure S9**. **Figure S10** shows the time evolution of the SASA for the three different molecular mechanics force fields as well as the SASA with respect to the residues of the Arkadia $RING_{927-994}$ domain based on the AMBER-03 and AMBER-99SB-ILDN, respectively. For each SASA plot vs. residues, a prediction as obtained from the previous NMR-based study (9) is given on the top plane for which the light blue denotes that residues have higher solvent accessibility and dark blue denotes that residues have very low solvent accessibility. In addition, a red dashed line is constructed to observe which residues have an accessible surface area less or equal to 0.25 nm^2.

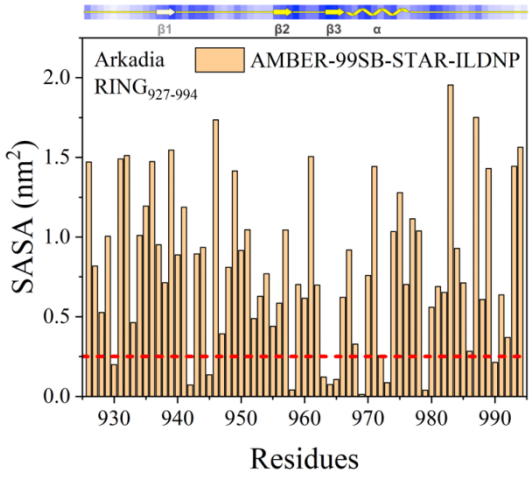
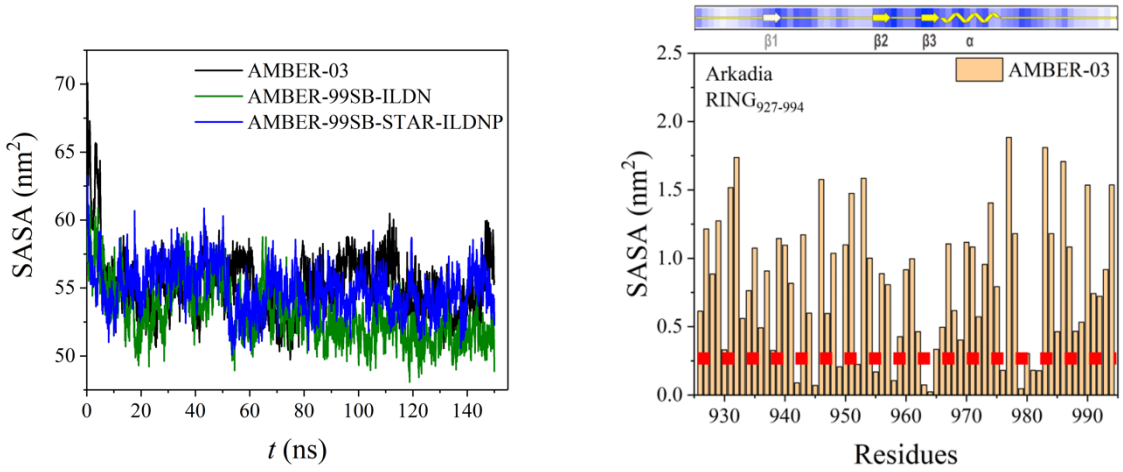


Figure S8. SASA plot of the Arkadia $RING_{927-994}$ domain is based on the AMBER-99SB-STAR-ILDNP molecular mechanics force field as a function of its Residues. The top plane indicates observations made from a previous NMR-based study (9) where the light-blue denotes residues with higher solvent accessibility and with dark blue residues with very low solvent accessibility.



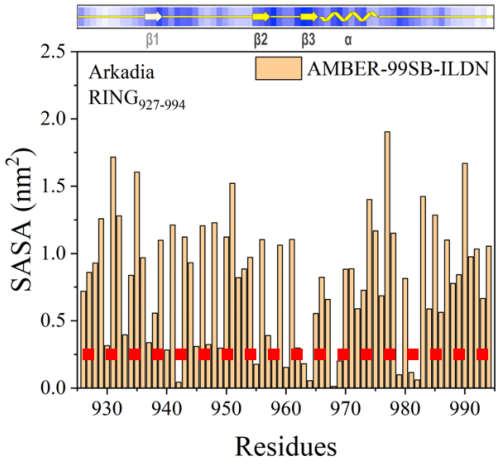


Figure S9. The top left plot shows the time evolution of SASA for Arkadia $RING_{927-994}$ based on three different molecular mechanics force fields. The other two plots show the SASA of the Arkadia $RING_{927-994}$ based on the AMBER-03 and AMBER-99SB-ILDN molecular mechanics force field as a function of its residues respectively. The top plane indicates observations made from a previous NMR-based study (9) where the light-blue denotes residues with higher solvent accessibility and with dark blue residues with very low solvent accessibility.

To elucidate the differences between the three different molecular mechanics force fields and how they compare against the previous NMR-based study (9) the table S1 is constructed in which someone can see on the first column the residues of the Arkadia $RING_{927-994}$ domain for which very low solvent accessibility (or equivalently lower than 0.25 nm^2 of accessible surface area) is reported in the previous NMR-based study (9) and in the next columns the comparison is shown based on the three different molecular mechanics force fields.

Table S1. The first column shows the residues of Arkadia $RING_{927-994}$ domain for which low solvent accessibility is observed from a previous NMR-based study (9) and the next columns show the comparison of the three different molecular mechanics force fields employed in this work.

Residues	AMBER-03	AMBER-99SB-ILDN	AMBER-99SB-STAR-ILDNP
938			✓
939			✓
940			
942	✓	✓	✓
943			✓
944			✓
945	✓	✓	✓
949	✓	✓	✓
955	✓	✓	✓
956			
957			
958	✓	✓	✓
959			✓
962		✓	✓
963	✓	✓	✓
964	✓	✓	✓

965			✓
966			
968		✓	✓
969		✓	✓
970			
972			
973			✓
979	✓	✓	✓
980	✓	✓	✓
986			✓

It is worth noting that a systematic study of several molecular mechanics force fields carried out by previous MD studies (10, 11) revealed that the AMBER-99SB-STAR-ILDN force field is capable of efficiently and accurately predicting the structural behavior of ubiquitin-based systems. The only difference between the AMBER-99SB-STAR-ILDNP employed in this study and the AMBER-99SB-STAR-ILDN force field is that the Pro and Hyp parameters were improved by fittings of experimental correlation times and NMR J-couplings (12).

Internal dynamics (*order parameter analysis)* – The comparison of the order parameter between experiment (9) and observations made from this MD study by employing the AMBER-03 and the AMBER-99SB-ILDN molecular mechanics force field is given in **Figure S8**. The comparison between experiments and predictions made by employing the AMBER-99SB-STAR-ILDNP molecular mechanics force field is shown in **Figure 4 (**in the main text).

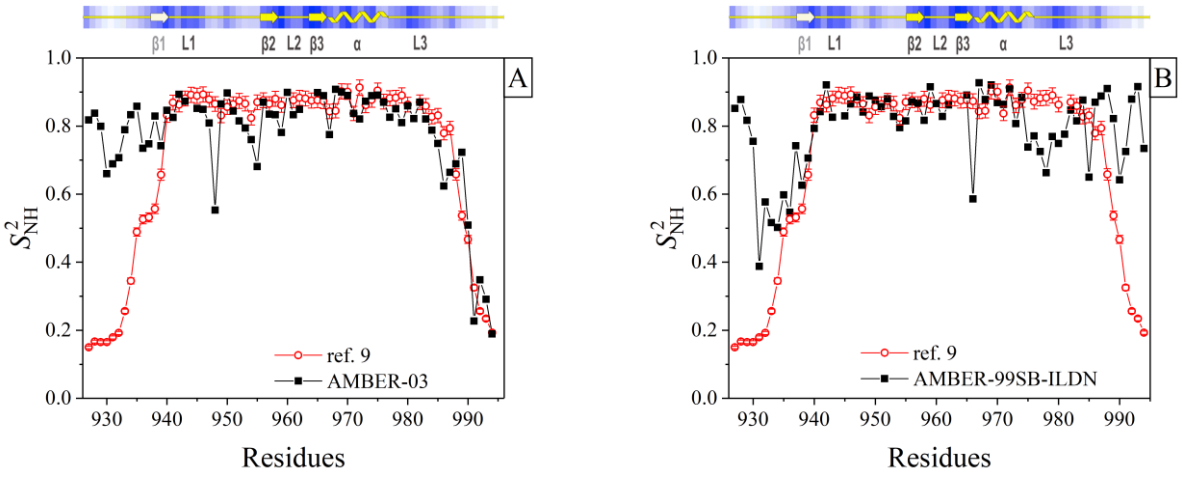


Figure S10. Order parameter, S^2, of the NH bonds of the Arkadia $RING_{927-994}$ as predicted from A) the AMBER-03 and B) the AMBER-99SB-ILDN molecular mechanics force field and their comparison against findings regarding the order parameter of the NH bonds and of the secondary structure (top plane) obtained from previous NMR based study (9).

4. Order parameter, S^2, of the NH bonds of Ark2C RING domain, E2 and Ub in their complexes.

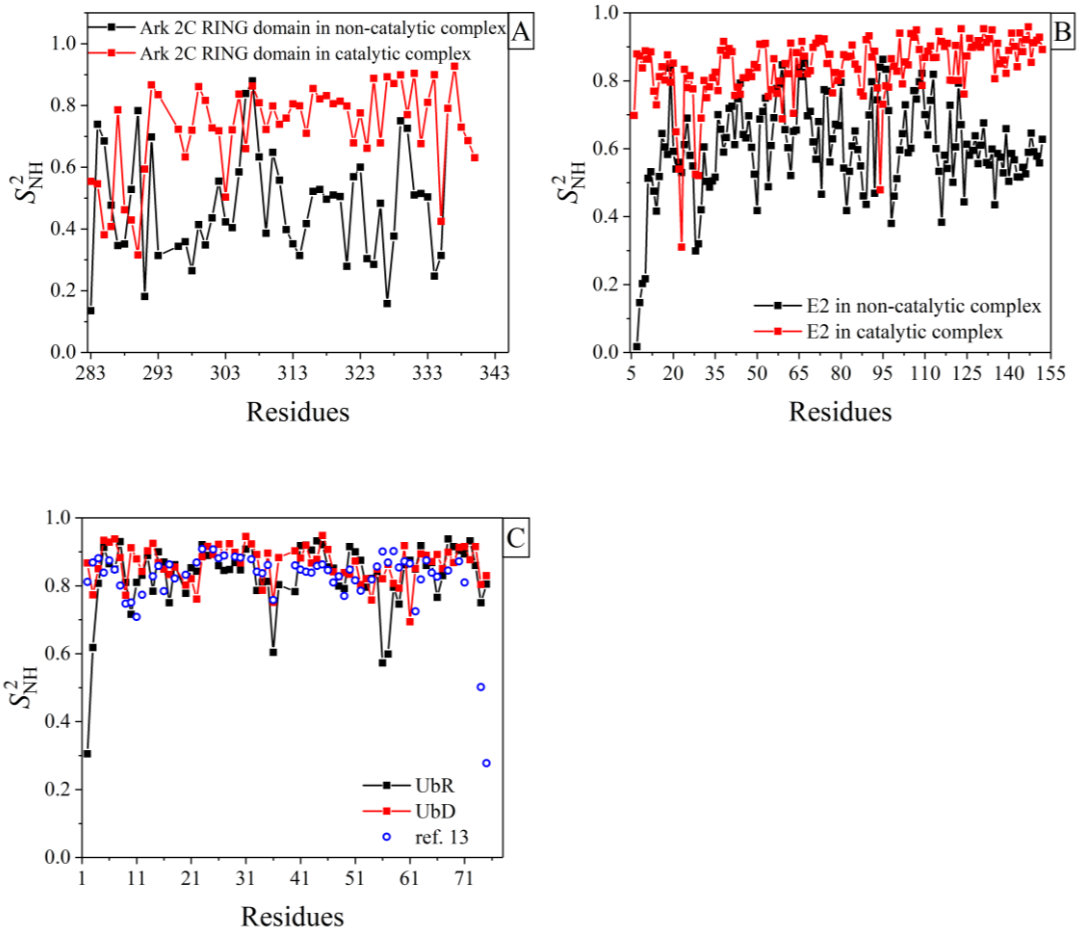


Figure S11: A) Order parameter, S^2, of the NH bonds of the monomeric RING domain of the human Ark2C E3 ligase complexed with E2-Ub in non-catalytic conformation (RING-E2-Ub, black) and with E2-Ub in catalytic conformation (Ub–RING-E2-Ub, red), B) Order parameter, S^2, of the NH bonds or the E2 enzyme in the closed conformation of the E3-E2- UbD complex (inactive, black) and in the open conformation of the UbR-E3-E2- UbD complex (active, red). C) Order parameter, S^2, of the NH bonds for the UbR bound to RING domain (black), and UbD bound to E2 (red) in the open conformation of the UbR-E3-E2- UbD complex. Open circles with blue color denote the comparison to experimentally obtained values (13).

References

1. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19-25.
2. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. Molecular Physics. 2006;52(2):255-68.
3. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. Physical Review A. 1985;31(3):1695-7.
4. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry. 1997;18(12):1463-72.
5. Darden T, York D, Pedersen L. Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 1993;98(12):10089-92.
6. 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(17):4546-59.
7. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577-637.
8. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. J Comput Chem. 1995;16(3):273-84.
9. Chasapis CT, Kandias NG, Episkopou V, Bentrop D, Spyroulias GA. NMR-based insights into the conformational and interaction properties of Arkadia RING-H2 E3 Ub ligase. Proteins: Struct Funct Bioinform. 2012;80(5):1484-9.
10. Muller DJ, Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, et al. Correction: Systematic Validation of Protein Force Fields against Experimental Data. PLoS ONE. 2013;8(4).
11. Bowman GR. Accurately modeling nanosecond protein dynamics requires at least microseconds of simulation. Journal of Computational Chemistry. 2016;37(6):558-66.
12. Aliev AE, Kulke M, Khaneja HS, Chudasama V, Sheppard TD, Lanigan RM. Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics. Proteins: Structure, Function, and Bioinformatics. 2014;82(2):195-215.
13. Nederveen AJ, Bonvin AMJJ. NMR Relaxation and Internal Dynamics of Ubiquitin from a 0.2 μs MD Simulation. Journal of Chemical Theory and Computation. 2005;1(3):363-74.