

E9-Im9 Colicin DNase–Immunity Protein Biomolecular Association in Water: A Multiple-Copy and Accelerated Molecular Dynamics Simulation Study

Riccardo Baron,^{*,†,‡} Sergio E. Wong,^{†,‡} Cesar A. F. de Oliveira,^{†,‡} and J. Andrew McCammon^{†,‡,§,||}

Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, Department of Pharmacology, Howard Hughes Medical Institute, University of California at San Diego, La Jolla, CA 92093-0365

Received: July 11, 2008; Revised Manuscript Received: August 27, 2008

Protein–protein transient and dynamic interactions underlie all biological processes. The molecular dynamics (MD) of the E9 colicin DNase protein, its Im9 inhibitor protein, and their E9-Im9 recognition complex are investigated by combining multiple-copy (MC) MD and accelerated MD (aMD) explicit-solvent simulation approaches, after validation with crystalline-phase and solution experiments. Im9 shows higher flexibility than its E9 counterpart. Im9 displays a significant reduction of backbone flexibility and a remarkable increase in motional correlation upon E9 association. Im9 loops 23–31 and 54–64 open with respect to the E9-Im9 X-ray structure and show high conformational diversity. Upon association a large fraction (~ 20 nm²) of E9 and Im9 protein surfaces become inaccessible to water. Numerous salt bridges transiently occurring throughout our six 50 ns long MC-MD simulations are not present in the X-ray model. Among these Im9 Glu31–E9 Arg96 and Im9 Glu41–Lys89 involve interface interactions. Through the use of 10 ns of Im9 aMD simulation, we reconcile the largest thermodynamic impact measured for Asp51Ala mutation with Im9 structure and dynamics. Lys57 acts as an essential molecular switch to shift Im9 surface loop towards an ideal configuration for E9 inhibition. This is achieved by switching Asp60–Lys57 and Asp62–Lys57 hydrogen bonds to Asp51–Lys57 salt bridge. E9-Im9 recognition involves shifts of conformational distributions, reorganization of intramolecular hydrogen bond patterns, and formation of new inter- and intramolecular interactions. The description of key transient biological interactions can be significantly enriched by the dynamic and atomic-level information provided by computer simulations.

Introduction

Many proteins form complexes with one or more binding partner(s) to fulfill their highly specific, yet multiple tasks in living organisms. This yields a significantly higher number of protein–protein complexes than proteins in the proteome. Aberrant association generally results in disease.^{1,2} Consequently, understanding protein–protein association at the molecular level,^{3–8} discriminating between highly-specific association and nonspecific aggregation processes,^{9–12} designing drugs capable of inhibiting biomolecular association,^{4–6,9,13–15} and developing computational approaches to predict favorable binding partners, dominant physicochemical interactions, and aggregation-prone segments^{16–19} are all targets of intense research. Here we focus on transient protein complexes, those whose unbound partners are stable in solution.

The study of protein–protein interactions (PPI) is several decades old; see for example the early pioneering work by Chothia and Janin.²⁰ Since then much has been learned about PPIs. For example, the fraction of nonpolar surface area at complex interfaces is nearly the same as in the surface of globular proteins.²¹ Furthermore, solute–solute electrostatic interactions are not always exclusively favorable for binding

and their contribution seems highly system dependent.¹⁰ Thus, in a number of cases hydrophobic effects and a thermodynamically favorable solvent–solvent and solute–solvent re-organization seem to dominate the binding process.^{10,22,23} Some evidence suggests that burying backbone amide hydrogen bonds is a significant driving factor in PPIs.²³ However, other studies propose individual interface side-chains as binding determinants.²⁴ On the basis of Ala scanning experiments, only a few (hot spot) residues at the interface of the human growth-hormone and its receptor make a significant contribution to binding. Further studies reported similar results in the context of other systems.^{25–27} Screening of a database combining structural and thermochemical data of Ala mutants suggested that protein interfaces involve a central hydrophobic region surrounded by polar/charged residues that exclude the solvent from the hydrophobic center, that is, the “O-ring” hypothesis.²⁸

Our current knowledge of protein–protein association is largely based on structures inferred by X-ray crystallography^{8,20,21,29} and nuclear magnetic resonance (NMR) in solution.^{30–36} This structural information can be linked to thermodynamic quantities calculated on the basis of association/dissociation kinetics experiments.^{29,31,37,38}

However, elucidating the relationship between protein dynamics and thermodynamic properties at the molecular level is currently hampered by three major limitations. First, the physico-chemical environment experienced during structural determination (i.e., generally the crystalline phase at low temperature) is typically different from the environment

* To whom correspondence should be addressed. E-mail: rbaron@mccammon.ucsd.edu. Phone: +1-858-534-2913. Fax: +1-858-534-4974.

[†] Department of Chemistry and Biochemistry.

[‡] Center for Theoretical Biological Physics.

[§] Department of Pharmacology.

^{||} Howard Hughes Medical Institute.

characterizing calorimetric and kinetic measurements (i.e., generally a solution of protein complexes at room temperature). Second, the interpretation of these (bio)physical experiments is often nontrivial even when including dynamical information due to space/time ensemble-averaging effects (ref 39 and refs therein; refs 40–43). Third, adding to the complexity of performing these experiments, a significant fraction of protein–protein complexes occurring in nature is difficult to study through classical experiments due to sample aggregation. For example, these three limitations may lead to significantly different interpretations of the changes of flexibility and hydration upon protein–protein association. Additionally, due to finite crystallographic resolution and refinement modeling accuracy, they may limit the description of molecular interactions (e.g., salt bridges and hydrogen bonds) essential for biomolecular recognition.

Molecular dynamics (MD) simulation with empirical force fields is a useful theoretical tool to study the properties of (bio)molecular systems with atomic resolution based on classical mechanics.^{43–45} They complement experimental results by providing distributions and time series of, in principle, any physically definable observable, within force field accuracy and finite sampling limitations. When combined with an explicit treatment of solvent molecules, MD simulations capture changes of structure, dynamics, interactions, and thermodynamics of biomolecular association.^{16,46–50}

The separate sampling of the bound and unbound states is the key to describe conformational dynamics differences upon association. While this is possible *in silico*, it may be impossible *in vitro*, because associating partners in protein–protein complexes may be insufficiently stable or difficult to purify in solution when individually unbound. For example, MD simulations can be used to study aspects of protein–protein association, like the conformational-selection hypothesis⁵¹ and the pre-organization⁵² of key residues, based on the independent bound and unbound ensembles of trajectory structures. However, the application of accurate computational chemistry technologies to protein–protein association remains non-standard due to their relatively high computational cost. Here, we overcome this general limitation by means of an integrated multiple-copy (MC; e.g. refs 44, 53, 54) MD and accelerated MD (aMD; refs 55, 56) simulation approach.

The structure and dynamics of the of E9 colicin DNase (deoxyribonuclease) in complex with its Im9 inhibitor protein has been extensively studied via solution NMR^{30–32,57–59} and kinetic experiments.^{29,31,37} A structural ensemble has been reported for Im9 based on NMR experiments.⁵⁷ X-ray crystallographic structures of the E9-Im9 complex⁸ and of the E9 unbound protein (PDB ID: 1fsj) have also been reported.

The major interest in E9-Im9 association dynamics derives from its biological role. *E. coli* under stress conditions secrete colicin DNases (E9) to destroy the DNA of competing bacteria.⁷ In order to protect themselves from the DNase activity, they also produce the Im9 immunity protein that inhibits the colicin DNase. Each DNase in this protein family specifically binds an immunity protein (~ 42 kJ mol⁻¹ partner preference; ~ 84 kJ mol⁻¹ overall binding free energy).^{7,60} However, biomolecular association with non-cognate partners can still be significantly favorable.⁶⁰ Previous studies suggested that the E9-Im9 interface has two hot-spot regions, which would permit cross talk and high specificity among other members of the colicin DNase family.²⁹

The thermodynamic impact on Im9 protein stability and E9-binding has also been investigated based on association/

dissociation kinetics experiments.³¹ Binding seems mainly perturbed by Asp51Ala (~ 25 kJ mol⁻¹) and Tyr54Ala and Tyr55Ala (~ 20 and ~ 19 kJ mol⁻¹) mutations. In the second case, a direct structure/thermodynamics relationship was proposed using the E9-Im9 X-ray structural model.⁸ However, no clear understanding of the reasons for such a large effect on binding after Asp51Ala mutation has been reported in the literature.

In this work, we address this point using the detailed dynamic picture of conformational sampling and molecular interaction changes for the E9-Im9 interface upon complexation. We investigate E9 and Im9 dynamics and their changes upon E9-Im9 colicin DNase association based on a set of six 50 ns long MC-MD simulations in explicit solvent (Table 1). In order to (i) alleviate finite sampling effects, (ii) enhance the conformational sampling statistics, and (iii) assess the convergence of MD simulations of protein systems in general, we employ two independent MD runs, labeled a and b, for each of the key thermodynamic states: unbound E9, unbound Im9, and E9-Im9 recognition complex.

We show that MC-MD allows more accurate monitoring of both global (e.g. protein surface loop motion and backbone motional correlation) and local (e.g., hydrogen-bonding and dihedral) conformational dynamics, by increasing rare events statistics. Additionally, in the case of the more flexible Im9 protein, we use the aMD simulation technique^{55,56} to improve sampling and capture slow protein motion and rare conformational changes. This technique allows us to solve the apparent discrepancies between independent MC-MD trajectories for Im9 protein dynamics. Additionally, we demonstrate how good sampling convergence from independent MC-MD runs can still cover only a limited portion of the conformational space accessible to a solvated protein system at thermodynamic equilibrium, which can instead be visited extensively by aMD simulation.

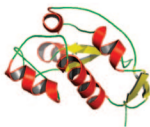
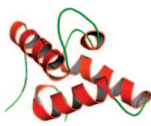
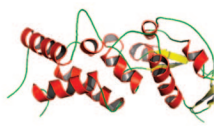
We focus the analysis of E9-Im9 colicin DNase complexation and dynamics in water on a number of key points of biological and biophysical relevance, after validating our simulations using X-ray crystallography and solution NMR experimental data. We describe the overall flexibility, dynamics, motional correlation, and solvent accessibility changes upon E9-Im9 binding. We investigate 23–31 and 54–64 Im9 loops conformational sampling, involved in the recognition process and in Im9 binding specificity. We provide a time-dependent analysis of salt-bridge formation and pre-organization and investigate Im9 Asp60–Lys57, Asp62–Lys57, and Asp51–Lys57 interactions and their role in E9-Im9 recognition through a surface-loop switch. This analysis allows us to capture the link between protein dynamics in solution and thermodynamic data based on kinetics experiments. We characterize the changes of conformational sampling and solute–solute hydrogen bonding of key E9-Im9 interface residues. These findings are discussed in the general context of protein–protein association, protein dynamics, and experiment interpretation.

Methods

Molecular Model and Multiple-Copy Molecular Dynamics. Initial coordinates (PDB ID: 1bxi) for the E9-Im9 complex⁸ including 104 crystallographic water molecules were used after removing Ni⁺ and PO₄²⁻ ions. The same coordinates for the individual monomers (excluding water oxygens) were used to initialize E9 and Im9 simulations (see Table 1 for reference codes).

System topologies and coordinate files were generated with the Amber 9 software.⁶¹ The parm99SB parameter set was

TABLE 1: Simulated Systems

reference code	E9_a / E9_b	Im9_a / Im9_b / Im9_aMD	E9-Im9_a / E9- Im9_b
MC-MD: a or b; accelerated MD: acc			
<i>T</i> [K]	300	300	300
nr. Na ⁺ ions	0	9	4
nr. Cl ⁻ ions	5	0	0
total system charge	0	0	0
nr. solute atoms	2071	1257	3328
nr. solvent molecules ^a (from X-ray water sites)	19959 (0)	13242 (0)	31194 (104)
nr. atoms in the system	22035	14511	34526
nr. solute-solute H-bonds in X-ray structure ^b	178	101	300
equilibration period [ns]	4.5	4.5	4.5
equilibrium period [ns]	50.0	50.0 / 10.0 (Im9_aMD)	50.0

^a From X-ray crystal structure as in ref 8. ^b After energy minimization to relax the solvent molecules around the protein.

employed in all simulations.^{62,63} Alternatively E9-Im9, E9, or Im9 systems were solvated in pre-equilibrated rectangular boxes of Amber-compatible TIP3P water,⁶⁴ large enough to avoid any interaction between images under periodic boundary conditions. Water molecules not corresponding to X-ray model water sites were deleted if overlapping to any solute atom based on van der Waals radii. Na⁺ and Cl⁻ ions were added at random positions to neutralize the overall system charge (Table 1) and treated with parameters derived for the SPC water⁶⁵ and adapted for the TIP3P water model.⁶¹

For each system, successive energy minimization, heat-up, and equilibration were performed employing different atom-positional harmonic restraining potentials. During the first stage all protein atoms were restrained to their crystallographic coordinates with a force constant of 209 kJ mol⁻¹ nm⁻² (0.5 kcal mol⁻¹ Å⁻²) to relax the solvent. The second stage involved the same potential applied on C^α atoms only. The third stage did not involve any restraint in order to remove any residual strain. The heat-up procedure followed a similar scheme. In the first stage (50 ps) all atoms were restrained to their X-ray coordinates (force constant as above), the temperature was linearly increased, from 0.1 to 300 K, over the first 30 ps. A weak-coupling algorithm⁶⁶ was employed using a single thermal bath to the overall system (0.1 ps relaxation time). During equilibration and production stages electrostatic interactions were evaluated using the Particle Mesh Ewald (PME) approximation⁶⁷⁻⁶⁹ under periodic boundary conditions (Ewald cutoff for direct space sum of 0.8 nm). Independent copies for MC-MD simulation were initialized using identical set-up, but different initial velocities randomly assigned from Maxwell-Boltzmann distributions at 50 K. At each 1 fs long MD integration step the bonds to hydrogen atoms were kept fixed using the SHAKE

algorithm.⁷⁰ The second 50 ps stage involved a (NVT) simulation at 300 K using a weaker restraining potential on all atoms (force constant of 104 kJ mol⁻¹ nm⁻²; 0.25 kcal mol⁻¹ Å⁻²). The third 100 ps stage with no restraint was performed by coupling the system to a 1 bar pressure bath⁶⁶ via isotropic coordinate scaling (relaxation time 0.5 ps; isothermal compressibility 4.6 10⁻⁴ [kJ mol⁻¹ nm⁻³]⁻¹). A general approach for enhancing phase-space sampling in computer simulation is the use of simultaneous or parallel MD simulations. Multiple-copy approaches allows to speed-up phase-space sampling while splitting the calculation into a number of independent runs.^{44,53,54} This is particularly appealing considering the rapid and steady increase of computational power in the form of multiple CPU clusters vs. single CPU supercomputers. In this study two independent MC-MD runs at 300 K were extended for 50 ns with the set up of stage three, as summarized in Table 1.

Accelerated Molecular Dynamics. Current MD simulations cover timescales often too short to capture biological processes dominated by rare events.^{43,44} In addition, regular atomistic MD simulations of biomolecules may spend most of the simulation time trapped in one or another local minimum.^{71,72} Therefore, a number of alternative techniques have been developed to enhance sampling.^{55,71-75}

We employed the accelerated MD approach (aMD) that deforms the energy landscape by adding a boost potential, $\Delta V(r)$, to the original potential surface every time $V(r)$ is below a pre-defined energy level E , as⁵⁵

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \frac{(E - V(r))^2}{\alpha + (E - V(r))}, & V(r) < E \end{cases} \quad (1)$$

where α modulates both depth and local roughness of the energy basins in the modified potential. In principle, this approach also allows the correct quasi-canonical averages of an observable, calculated from configurations sampled on the modified potential energy surface, to be recovered from aMD simulations.⁵⁵ In order to simultaneously enhance the sampling of internal and diffusive degrees of freedom, a dual boosting approach was employed.⁷⁶ In this case, separate torsional and total boost potentials are defined as

$$V(r) = V_0(r) + V_t(r) \quad (2)$$

$$V^*(r) = \{V_0(r) + [V_t(r) + \Delta V_t(r)]\} + V_T(r) \quad (3)$$

where $V_t(r)$ is the total potential of the torsional terms; $\Delta V_t(r)$ and $\Delta V_T(r)$ are the boost potentials applied respectively to the torsional terms $V_t(r)$ and the total potential energy $V_T(r)$, and both follow the same definition as in eq 1 in combination with α_t and α_T parameters. The parameters of eqs 2 and 3 were set as follows. $E_t = 1168$ kcal mol⁻¹, that is, ~ 300 kcal mol⁻¹ higher than the ensemble-averaged torsional potential energy from simulation Im9_b. $\alpha_t \approx 60$ kcal mol⁻¹.⁷⁷ $E_T = 0.1$ kcal mol⁻¹ (nr. atoms)⁻¹ plus the ensemble-averaged total potential energy from Im9_b. $\alpha_T \approx 0.1$ kcal mol⁻¹ (nr. atoms)⁻¹.⁷⁸ These E_T and α_T values allow one to reproduce the most relevant structural and energetic properties of liquid water while increasing the water self-diffusion coefficient by $\sim 15\%$.^{76,79} The quasi-canonical re-weighting procedure to recover thermodynamically meaningful ensemble configurations is described in ref 76. Snapshots were extracted every 1 ps along the 10-ns Im9_aMD trajectory for analysis.

Analysis Procedure. Trajectory snapshots were extracted every 5 ps along each of the six equilibrated 50 ns MC-MD trajectories E9_a, E9_b, Im9_a, Im9_b, E9-Im9_a, E9-Im9_b for analysis.

Structural Fitting. Structural fitting was performed by (i) superimposing their centers of mass (to remove overall translation) and (ii) performing an atom-positional least-square fitting procedure (to remove overall rotation)⁸⁰ using all C α atoms, except those belonging to the first and last five residues, unless otherwise specified.

Secondary-Structure Monitoring. Secondary structure elements formation was monitored according to the definition by Kabsch and Sander.⁸¹

Chemical-Shift Calculation. Im9 C α chemical shifts (δ) were calculated using the SHIFTX program⁸² on individual structures of each of the separate six MC-MD ensembles, using either all 50 ns equilibrium trajectories (10⁴ snapshots) or only their final 10 ns periods (2×10^3 snapshots). Next, their ensemble averages ($\langle \delta \rangle$) were calculated and used to estimate the chemical-shift perturbations upon protein-protein association ($\Delta \delta$). Time/ensemble averaged values from NMR experiments³⁰ have been previously deposited (BioMagRes ID: 4115 and 4114 for bound and free Im9). We calculated corresponding values also on the basis of the E9-Im9 X-ray crystallography structure⁸ as additional validation of our six MC-MD simulations. The same calculations were performed for E9 C α $\langle \delta \rangle$ and for both E9 and Im9 ¹H, ¹³C, ¹⁵N (not reported; available upon request to the authors).

Correlation, Pair Distance, and Standard Deviation Maps. The normalized cross-correlation maps (NCCM) were calculated as previously proposed by Hünenebrger et al.⁸³ using the Cartesian coordinates of all pairs of C α atoms. The maps of the

average C α -atom pair distances, $\langle d \rangle$, and of their standard deviations, σ , were calculated using the ensemble-averaged distances of all pairs of C α atoms. In all cases independent E9_a, E9_b, Im9_a, Im9_b, E9-Im9_a, and E9-Im9_b MC-MD runs were separately employed.

Surface-Loop Conformational Clustering. Trajectory structures were clustered into batches of similar configurations using the backbone atom-positional RMSD of Im9 residues 50–65 as the similarity criterion and the clustering algorithm previously described⁸⁴ and applied to protein surface loops.⁴² A cutoff of 0.2 nm was chosen after evaluation of the dependence of cluster populations against the total number of clusters found for each simulation (0.05–0.25 nm cutoff values range).

Hydrogen-Bond Monitoring. Hydrogen bonds were defined to have a maximum hydrogen-acceptor distance of 0.35 nm and a minimum donor-hydrogen-acceptor angle of 120°. Only those occurring at least 10% of the time along at least one of the two independent MC-MD simulations were reported.

Solvent Accessibility. Solvent-accessible surface area (SASA) values were calculated using the program NACCESS⁸⁵ for individual protein or complex structures after removing water molecules and ions. Partial contributions to the total SASA value were calculated⁴² for five classes of particles: (i) all atoms, (ii) polar side chains (all oxygens and nitrogens in the side chains) or (iii) nonpolar side chains (all non-oxygens and non-nitrogens in the side chains), and (iv) side chains (all side-chain atoms) or (v) main chain (all backbone N, C, and O atoms). For this purpose C α atoms were classified as side-chain atoms to assign a side-chain solvent accessibility also to Gly residues.

Salt-Bridge Monitoring. Salt bridges were monitored for all residues throughout the seven trajectories and considered formed when the (shortest possible) carboxylic oxygen–amino nitrogen distance fall below 0.35 nm. This threshold was defined after analyzing the distance probability distributions for all bridging residues. Only salt bridges occurring at least 10% of the time along at least one of the two independent MC-MD simulations are reported. A subgroup of these, displaying distance probability distributions with more than one dominant peak, were defined as water-mediated salt bridges if the above-defined distance fall below 0.55 nm and simultaneously at least one (hydrogen-bonding) water-oxygen atom was found within a sphere of 0.55 nm diameter centered at the middle point between the carboxylic oxygen and the amino nitrogen.

Results and Discussion

MD Simulation Validation: Stability and Comparison with Experiment. To assess the stability of the six 50 ns long MC-MD simulations a collection of properties was monitored as a function of time.

Figure 1a shows the backbone C α atom-positional root-mean-square deviation (RMSD) between the trajectory structures and the corresponding X-ray structure of the E9-Im9 complex.⁸ These fluctuations are of small magnitude overall and remain stable along the simulation time (standard deviations for E9-Im9_a, 0.01; E9-Im9_b, 0.01; E9_a, 0.01; E9_b, 0.01 nm). Im9 fluctuations are comparatively larger (standard deviations for Im9_a and Im9_b: 0.03 and 0.05 nm) and demonstrate higher structural deviations for unbound Im9 in the liquid phase. Most of the conformational diversity observed in Im9 dynamics can be attributed to the high flexibility of its 54–64 loop (Supporting Information, Figure S1).

In the case of unbound E9 an additional X-ray structure is available (PDB ID: 1fsj; primary citation to be published) and we additionally employed it to monitor corresponding RMSD

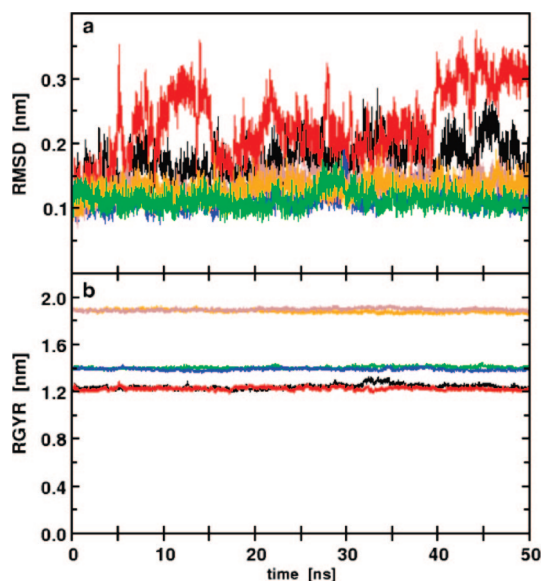


Figure 1. Time series of backbone (a) root-mean-square deviation, RMSD, from the X-ray structure of the E9-Im9 complex and (b) radius of gyration, RGYR, for MC-MD simulations Im9_a (black), Im9_b (red), E9_a (green), E9_b (blue), E9-Im9_a (orange), and E9-Im9_b (brown). All C $^{\alpha}$ -atoms excluding first five and last five residues were used. See Table 1 for reference codes and Methods for computational details.

time series for the two independent E9_a and E9_b simulations. The RMSD time series throughout E9_a and E9_b simulations based on the two X-ray models are almost identical; however, values from the unbound E9 crystal structure are shifted ~ 0.025 nm to lower RMSD values (not shown).

Figure 1b displays the corresponding backbone C $^{\alpha}$ atom-positional radius of gyration (RGYR). These values fluctuate stably along the 50 ns periods, and the corresponding mean values (E9, 1.2 and 1.2 nm; Im9, 1.4 and 1.4 nm; E9-Im9, 1.9 and 1.9 nm, respectively) show good agreement with values from the X-ray model structure (1.2, 1.4, and 1.8 nm, respectively). A somewhat more compact E9-Im9 complex structure in the crystalline than liquid phase is expected and found due to crystal-packing forces.

The formation of secondary structure elements was monitored along each of our 50-ns long MC-MD runs according to the definition by Kabsch and Sander⁸¹ and compared with the experimental X-ray models (Supporting Information, Table S1). All secondary structure elements are stable (not shown) and their ensemble-averaged relative occurrences are similar between independent MC-MD trajectories. These values are overall also close to values from the E9 X-ray structure. On the basis of X-ray crystal experiments E9 increases α -helix (28 to 32%) and decreases bend (19 to 16%) formation upon Im9 binding. However, corresponding values calculated from our simulations in the liquid phase do not show such a behavior. This can be explained by considering that crystal-packing forces may favor helical vs. flexible-bend elements formation.

We additionally compared Im9 conformational sampling from our Im9_a, Im9_b, E9-Im9_a, and E9-Im9_b simulations with available NMR experimental data in solution. The time series of the RMSD from any of the ten reported dominant structures in solution display values always lower than 0.31 nm (not shown) with fluctuation behavior similar to those calculated from the X-ray model (Figure 1). The Im9 backbone C $^{\alpha}$ -atom ensemble-averaged chemical shifts, $\langle\delta\rangle$, from NMR experiments and our MD simulations also show excellent agreement

along the entire Im9 residue sequence, as calculated by using the SHIFTX program⁸² on all individual structures of each of the separate six MD ensembles (Supporting Information, Figure S2a–d). A systematic overlap is found between values estimated from the entire 50 ns trajectories or using the last 10 ns periods only, showing that our simulations are well converged. However, $\Delta\delta$ perturbations upon E9-Im9 binding are small relative to the chemical shift data ($<0.1\%$ of magnitude) and within the average (over all residues) of the standard deviation of our estimated $\langle\delta\rangle$ values (~ 1.1 ppm). Similar observations could be drawn on the basis of Im9 ^1H , ^{13}C , ^{15}N chemical shifts (not reported). Overall these results show good agreement with experiment.

E9 and Im9 Backbone Dynamics and Motional Correlation Changes upon Association. Figure 2 reports the normalized-cross-correlation maps (NCCM), the maps of the average distance ($\langle d \rangle$), and of its standard deviation (σ) for all C $^{\alpha}$ atom pairs of E9 and Im9 in their free (E9_a, E9_b, Im9_a, and Im9_b) and bound (E9-Im9_a and E9-Im9_b) states. NCCM maps allow to capture pairwise motional correlation along the protein backbone; their comparison with $\langle d \rangle$ maps evidences if lowly or highly correlated motion is experienced between regions close or far in space on average; corresponding σ maps show deviations from the average backbone pair distances, pointing at pairwise flexibility along the backbone. Such a novel combination of these standard analyses provides a useful tool to interpret overall backbone flexibility and motional correlation for biomolecular simulations.

Both E9 and Im9 proteins experience an increased overall motional correlation upon E9-Im9 binding. This effect is particularly evident for sequence regions E9 70–100 (3- or α -helix secondary structure elements sampled 95% of time), Im9 30–44 (α -helix secondary structure element sampled 98% of time), and Im9 50–65. These regions largely encompass the E9-Im9 binding interface residues.

The E9 NCCM from our MC-MD runs E9_a and E9_b or E9-Im9_a and E9-Im9_b simulations are substantially similar (Figure 2). NCCM are a sensitive measure of convergence for MD simulation,⁸³ thus these results confirm the good convergence of E9 backbone conformational sampling. Additionally, the corresponding ensemble-averaged distance $\langle d \rangle$ maps show that (i) these regions of highly-increased E9 motional correlation are largely associated with relatively small C $^{\alpha}$ -atom pair $\langle d \rangle$ values (< 1.0 nm) and (ii) a very limited perturbation of $\langle d \rangle$ values occurs upon association (Figure 2). σ maps for free and bound E9 are also similar. These observations are in line with the observed limited changes of backbone flexibility of E9 upon binding (Figure 3).

The Im9 NCCM maps from Im9_a and Im9_b or E9-Im9_a and E9-Im9_b simulations are also largely similar (Figure 2). However, an interesting exception can be noticed for residues Im9 54–64 comparing Im9_b with E9-Im9_a and E9-Im9_b simulations. This residue sequence corresponds to the flexible Im9 54–64 surface loop, which shows different dynamic behaviors in the two independent MD runs due to hinged-loop opening (see next section). From the NCCM map is clearly demonstrated that the 54–64 loop highly anti-correlated motion (Im9_b: correlation < -0.5). A comparison of Im9 CCM and $\langle d \rangle$ maps for the 54–64 loop in its free and bound states shows that major changes in motional correlations are associated with sizeable C $^{\alpha}$ atom pair $\langle d \rangle$ values (> 1.5 nm), and that a significant increase of motional correlation is experienced upon E9-Im9 association.

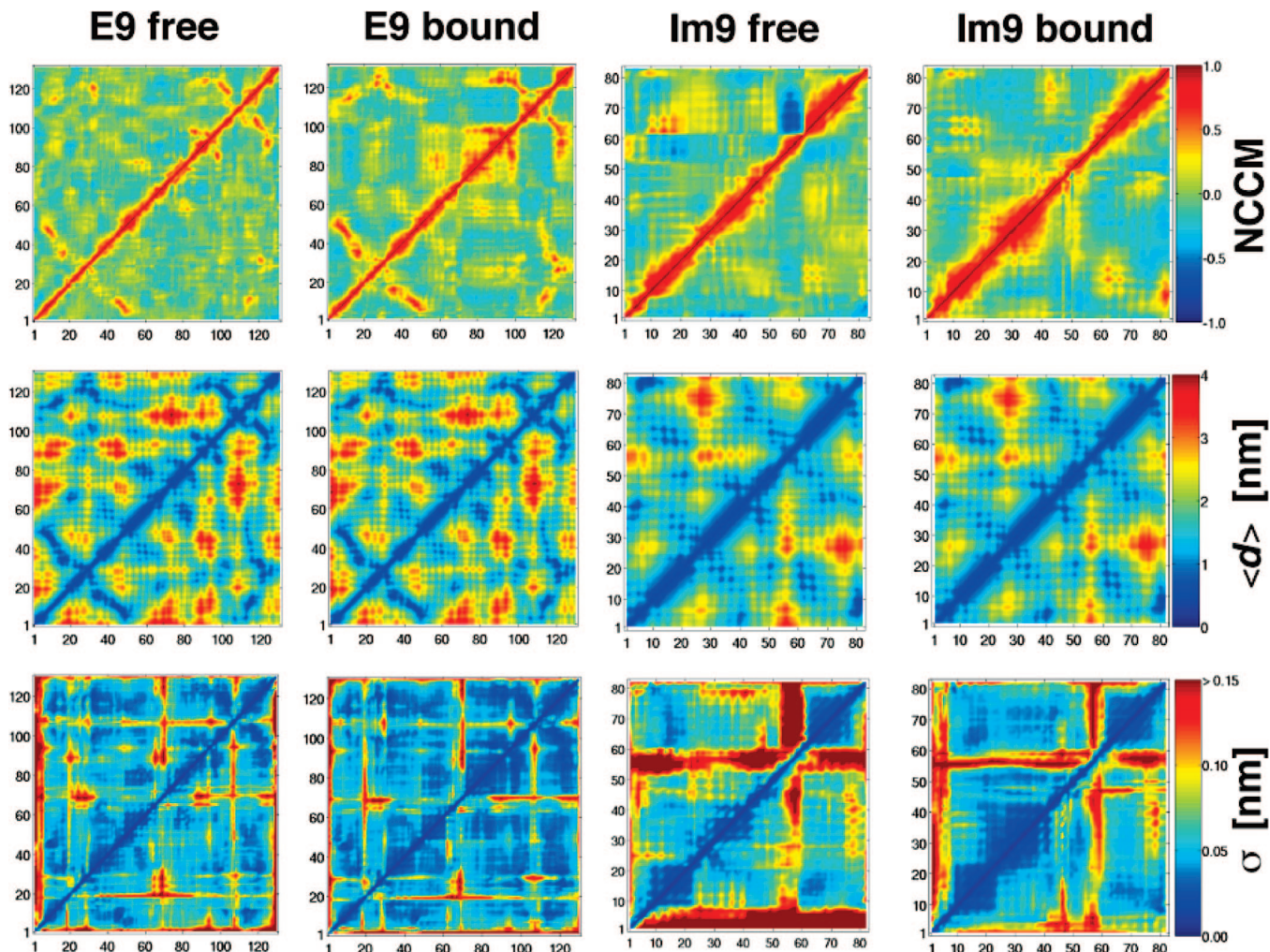


Figure 2. Normalized cross-correlation maps, NCCM, and maps of the average distance, $\langle d \rangle$, and its standard deviation, σ , for the C^α atom pairs of E9 and Im9 in their free (E9_a, E9_b, Im9_a, and Im9_b) and bound (E9-Im9_a and E9-Im9_b) states. For each map the upper left and lower right triangles report values from MC-MD runs a or b, respectively. Map axes run along the residue nr. sequences. See Table 1 for reference codes and Methods Section for computational details.

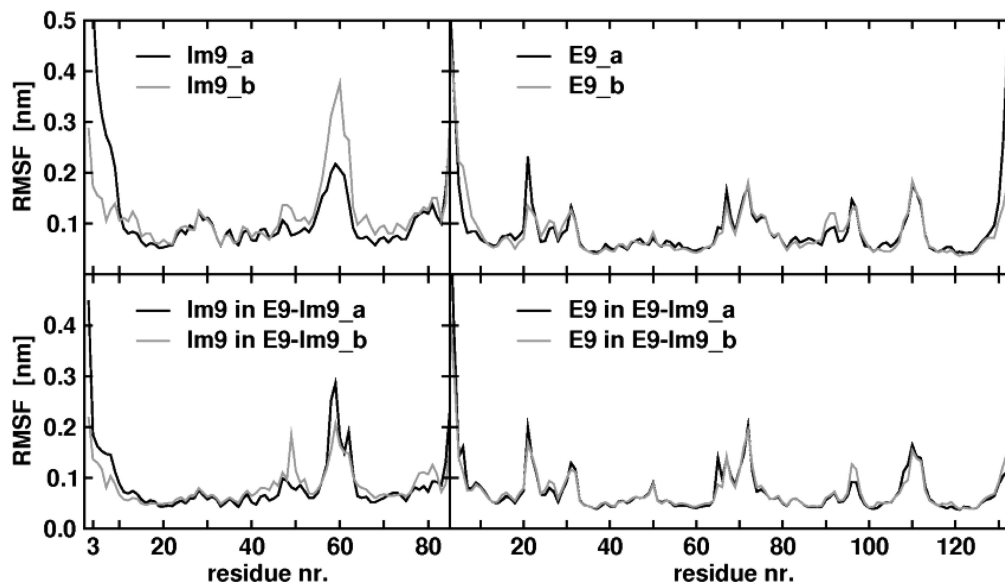


Figure 3. Backbone C^α atom-positional root-mean-square fluctuations, RMSF, along the amino acid sequence for the Im9 and E9 unbound proteins (top panels) and the E9-Im9 complex (bottom panels) based on independent MC-MD runs a (black) or b (grey).

We note that all analyses are consistent with the general picture of Im9 being significantly more flexible than its E9 counterpart (see Figures 1, 2, and 3). The σ maps for free

and bound Im9 show that the amplitude of the deviations from the average $\langle d \rangle$ values is significantly reduced upon E9-Im9 binding, particularly for the Im9 54–64 loop region.

The different dynamic behaviors of the E9 and Im9 recognition partners additionally suggest that a remarkable loss of Im9 configurational entropy occurs upon E9-Im9 binding, while E9 dynamics is substantially affected at the E9-Im9 interface only.

Figure 3 shows the overall backbone flexibility along the amino acid sequences for comparison between the unbound (E9 and Im9) and the E9-Im9 complex ensembles. A comparison of RMSF values between corresponding sequence regions of the E9 or Im9 unbound proteins and the E9-Im9 complex demonstrates a reduction of backbone flexibility upon association. Sequence regions Im9 25–45, Im9 50–70, and E9 75–95 are the most affected. These regions encompass the E9-Im9 interface.

In general, we note peaks of comparable intensities between corresponding copies of MC-MD runs. Thus, we conclude that the most flexible regions of the proteins are sampling their configurational spaces to a similar extent throughout independent MD runs. This latter statement does not hold when comparing Im9_a and Im9_b simulations. In fact, the region of residues 50–64 shows a pronounced mobility in both cases, but the magnitude of the peaks in these regions is significantly higher in the Im9_b than Im9_a simulation (see next section).

Im9 54–64 Flexible Hinged-Loop Conformational Sampling. The conformational sampling of Im9 54–64 hinged-loop, located at the protein–protein interface in the E9-Im9 complex, may play a fundamental role in the highly specific DNase-immunity proteins recognition process. Wallis et al. and Kleanthous et al. proposed that residues Tyr54 and Tyr55 are particularly relevant in this context.^{8,31}

Clustering trajectory snapshots into batches of highly similar configurations is a useful tool to capture the dominant configurations sampled by flexible protein regions along a MD trajectory.⁴² Figure 4a shows the dominant conformations adopted, in our two independent Im9_a and Im9_b simulations, by the Im9 loop region of residues 54–64 and allows a comparison with the available E9-Im9 X-ray crystal structure.⁸ The dominant loop configuration represents 89% of the Im9_a trajectory structures. However, the central members of the most populated first, second, and third cluster of structures represent 41, 29, and 12% of the Im9_b ensemble, respectively. E9-Im9_a and E9-Im9_b complex simulations display similar dominant loop configurations representing 90 and 97% of the structures, respectively. For simulations Im9_a, E9-Im9_a, and E9-Im9_b there are 4, 3, and 3 clusters populated by at least 1%, totaling 98, 99, and 99% of the entire ensemble, respectively. Instead, for simulation Im9_b we find that 8 clusters are populated by at least 1%, totaling 96% of the ensemble. The higher diversity of loop-region conformations for simulation Im9_b reflects the higher flexibility of the surface loop in the Im9 unbound state.

The intramolecular hydrogen-bonding and salt-bridge interactions characterizing each loop dominant configuration suggest that these are stabilized by specific interactions (Supporting Information, Table S2). The X-ray crystal structure⁸ displays 10 hydrogen bonds in the 50–65 region, and includes the Asp60–Lys57 salt bridge (Figure 4b). All of these hydrogen bonds form transiently throughout the E9-Im9_a and E9-Im9_b trajectories. Additionally, our simulations show that the Asp60–Lys57 salt bridge is the key interaction determining the shape of the 54–64 loop backbone (Figure 4c–e). During the 50 ns period, the Asp60–Lys57 salt bridge is substituted tens of times by the Asp62–Lys57 one and is transiently water-mediated. This emerges in the second most populated cluster structure (Figure 4d).

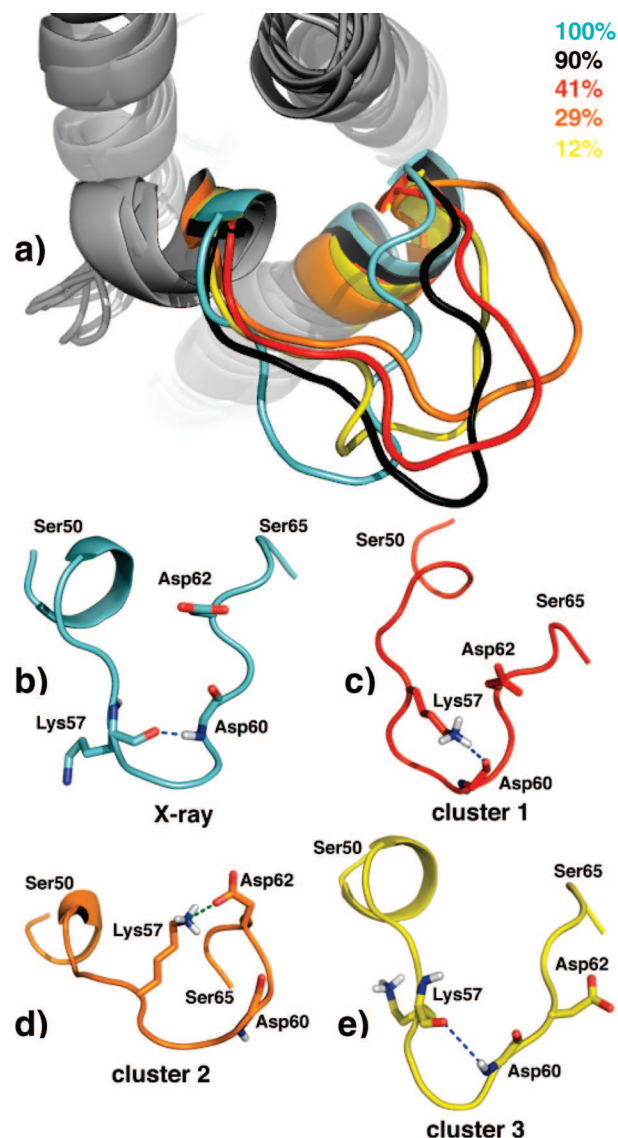


Figure 4. Dominant conformations for the Im9 loop-gating region (residues 50–65) from X-ray crystallography experiments and Im9_a or Im9_b simulations. (a) The E9-Im9 X-ray structure (cyan tube) is superimposed to the central members of the most populated clusters of structures from simulations Im9_a (black tube) and Im9_b (red, orange, and yellow tubes for first, second, third clusters, respectively). Cluster populations are reported using the corresponding colors. Superimposition of all backbone C α atoms was used for all representations. The following panels summarize the key interactions within the loop region for (b) the E9-Im9 X-ray structure and based on the central member structures of the (c) first, (d) second, and (e) third most populated clusters of the Im9_b simulation. See Table 1 for reference codes and Methods for computational details.

Im9 Conformational Sampling from Accelerated Molecular Dynamics. Enhanced sampling was achieved for the most flexible Im9 protein by using the accelerated MD (aMD) simulation protocol by Hamelberg et al.⁵⁵ (see Methods). Protein stability along 10 ns of Im9_aMD trajectory is consistent with what is described above for MC-MD 50-ns runs Im9_a and Im9_b. However, aMD allows sampling a significantly larger configurational space, which is comparatively larger for the most flexible secondary structure elements.

Figure 5 summarizes Im9 dynamics from snapshots of our 10 ns long aMD simulation. Im9 loops 23–31 and 54–64 significantly open with respect to the E9-Im9 X-ray structure model⁸ (Figure 5a). Both loops can also transiently twist farther

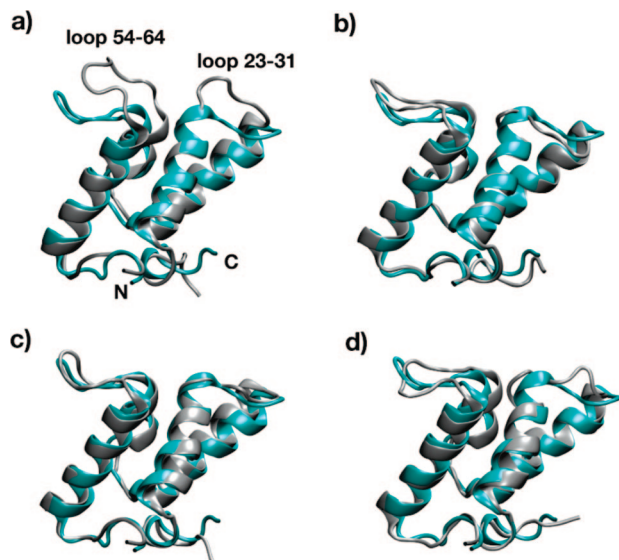


Figure 5. Conformational sampling of Im9 23–31 and 54–64 loops from key snapshots of the accelerated Im9_aMD simulation (grey cartoon) and corresponding model from E9-Im9 crystal X-ray diffraction experiments (cyan cartoon). The two loops are shown when (a) both open, (b) both closed, and (c) close to the X-ray model configuration. Panel (d) shows an example with 23–31 loop open while 54–64 loop closed. See Table 1 for Im9_aMD simulation set-up and Methods for computational details.

apart than observed in the crystal (Figure 5b). Yet, repeatedly the aMD simulation visits configurations close to the X-ray model (Figure 5c). This demonstrates that the dynamic behavior observed in our explicit solvent simulations is consistent with the ensemble-averaged E9-Im9 crystal X-ray model. Interestingly, the motion of 23–31 and 54–64 loops is not strictly correlated, and alternatively one only of the two loops display larger mobility (see example snapshot in Figure 5d).

A more detailed characterization of loops 23–31 and 54–64 sampling is summarized in Figure 6, comparing our MC-MD and aMD results for key ϕ and ψ backbone dihedral angles. Loop 23–31 backbone sampling is substantially identical from Im9_a and Im9_b runs. The most highly visited regions of the Ramachandran plot can be observed for $(\phi;\psi)$ values around $-70;50$ and $-60;-15$ degrees for Asp26 and Ser28, respectively. Asp26 conformational sampling from the Im9_aMD simulation is significantly extended towards the $-60;120$ region. Remarkably, Ser28 extensively samples the region around $-45;130$ in addition to that around $-60;-15$ during the aMD run. Loop 54–64 backbone sampling differs between the Im9_a and Im9_b runs, as discussed above. Such a different dynamic behavior from independent MD copies can be reconciled on the basis of our aMD simulation. The most highly visited regions of the Ramachandran's plot for Lys57 can be observed for $(\phi;\psi)$ values around $-120;160$, $-50;150$, and $-90;155$ degrees, for Im9_a, Im9_b, and Im9_aMD, respectively. Additionally, the broader sampling of Im9_aMD extends to the additional region around $-55;-45$. Gly59, the central residue of the 54–64 loop, experiences the highest sampling variability between Im9_a and Im9_b simulations, and between the dominant configurations of this loop (see Figure 4). The most highly visited region is around $80;0$ from all three Im9 simulations. Im9_a samples as well around $-60;0$, Im9_b as well around $90;-170$ and $-70;-170$, while aMD fully accesses all these regions. Ser61 shows a similar behavior: Im9_a samples mainly around $-60;140$, Im9_b additionally around $-150;-150$, while Im9_aMD covers additionally the region around $-70;0$.

In summary, a good correspondence is found between Im9_a and Im9_b runs. However, these results point out the difference between apparent (microstate) convergence and complete (macrostate) sampling in MD simulations. Sufficient sampling (>100 ns) in MD simulation is essential to capture slow protein dynamics, albeit these simulations seem well converged when trapped in specific microstates. We note that employing independent MC-MD replicas of a same system may be an effective way to enhance the statistics of such events as the motion of protein surface loops. However, the aMD technique seems a more powerful tool to explore additional regions of the global phase space for solvated macromolecules.

Change of Surface Solvent Accessibility upon E9-Im9 Association. Table 2 reports the SASA values for the reference unbound E9 (PDB ID: 1fsj) and complex E9-Im9⁸ X-ray crystal structures, as well as the ensemble-averaged values for our six MC-MD simulations. The corresponding 50-ns long time series along the six MD trajectories show relatively small fluctuations (maximum standard deviation from the mean values of 1.7, 1.8, and 2.1 nm² for E9_a, Im9_a, and E9-Im9_a, respectively; not shown) in agreement with the observed stability. Instead, smaller values systematically characterize the reference X-ray crystal structures and the E9 protein in the E9-Im9 complex vs. its unbound structure (Table 2). Interestingly, our E9_a and E9_b unbound state simulations display SASA values closer to the E9 unbound than to the E9-Im9 complex X-ray model, albeit the latter was used to initialize the two independent E9_a and E9_b simulations.

The origin of a different solvent accessibility between X-ray crystal structures and explicit solvent MD simulations can be detailed considering the partial contributions to the overall (all) SASA values based separately on the polar or nonpolar, and the side-chains or main-chain subgroups of atoms only.⁴² Considering either the polar or the sidechains subgroups systematically larger average values are found for MD trajectories with respect to the corresponding based on the E9-Im9 X-ray structure (see Table 2). On the other hand, considering either the nonpolar or the main-chain subgroups similar values are found for the same trajectories structures with respect to the X-ray structures. The major deviations seem to be dominated by the polar and side-chain subgroups. These trends can be explained considering that crystal-packing forces restrict protein motion, especially for those surface regions mainly contributing to the solvent-accessible surface. Instead, MD simulations in explicit solvent reproduce more closely the physical phase of solvated proteins. Similar trends have been reported using a different MD force field for a different system,⁴² supporting these general conclusions. A consequence of this behavior is a substantial reduction of the number of solute–solute hydrogen bonds observed in the liquid (MD) versus crystal (X-ray) phase (Supporting Information, Figure S3).

The changes of surface solvent accessibility upon E9-Im9 complex formation can be quantified by comparison of all SASA values for the E9-Im9 complex, and the corresponding sums of unbound E9 and Im9 partners (Table 2). We observe that upon E9-Im9 association more than 20 nm² of E9 and Im9 protein surfaces become inaccessible to the solvent, corresponding to $\sim 17\%$ of the overall E9-Im9 complex all SASA values. A smaller value of ~ 14 nm² would be estimated from the E9-Im9 X-ray crystal structure, assuming that no significant conformational change occurs for the individual E9 and Im9 unbound partners. Unbound E9 and Im9 proteins lose a significant surface accessible to water due to E9-Im9 association. Additionally, both protein surfaces experience different solvent

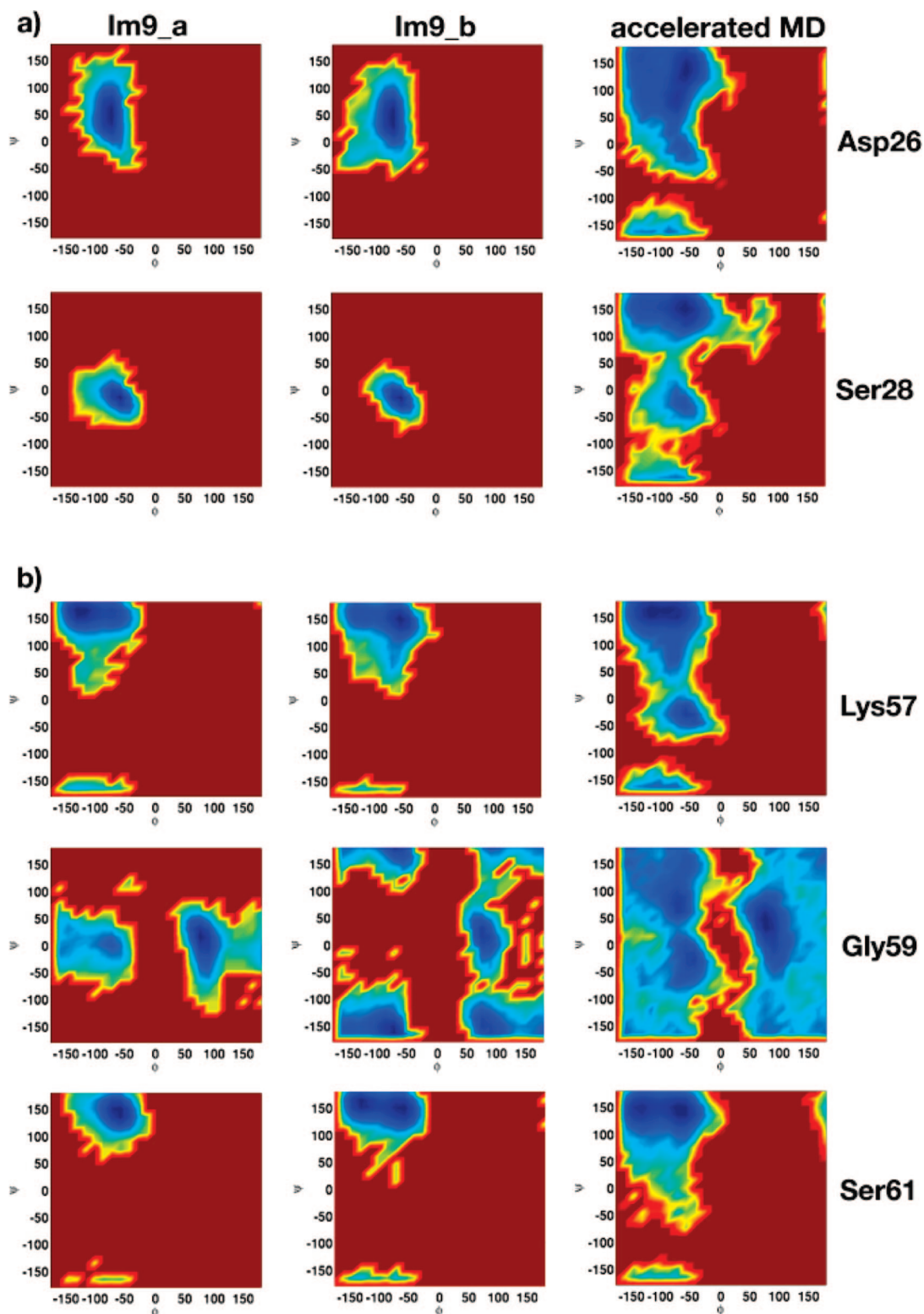


Figure 6. Ramachandran's plots of Im9 backbone angles ϕ and ψ for key residues of (a) 23–31 and (b) 54–64 loops. Color coding goes from red (low population) to blue (high population). Convergence between independent MC-MD runs can be addressed by comparison between Im9_a and Im9_b simulations (left and middle columns). Corresponding results, are shown on the basis of the accelerated Im9_aMD simulation (right column), after quasi-canonical re-weighting. See Table 1 for reference codes and Methods for computational details.

accessibility depending on crystalline-like or liquid-like physico-chemical environments.

The per-residue SASA differences upon E9_Im9 binding were also calculated from our six MC-MD simulations (Supporting Information, Figure S4). From this analysis we find that a larger number of interface residues decrease their solvent accessibility

upon E9-Im9 binding, in line with the behavior described above. The residues affected by major changes are E9 Ser74, Tyr83, Phe86, Lys89, and Lys97 (decreasing by about -1.1 , -1.4 , -1.5 , -1.6 , and -1.0 nm², respectively) and Im9 Ala23, Glu28, Val32, Tyr52, and Tyr53 (decreasing by about -0.9 , -1.7 , -1.1 , -1.6 , and -1.4 nm², respectively). Previous NMR

TABLE 2: Comparison of Ensemble-Averaged Solvent-Accessible Surface Areas^a from X-Ray Crystal Structures and MC-MD Simulation in Water

ref code	< SASA > [nm ²]				
	all	polar	non-polar	side chains	main chain
X-ray structure					
E9 ^b	80	33 (44)	47 (56)	70 (87)	10 (13)
E9 ^c	74	32 (43)	42 (57)	66 (89)	8 (11)
Im9 ^c	49	22 (45)	27 (55)	43 (88)	6 (12)
E9-Im9 ^c	109	48 (42)	61 (58)	96 (86)	13 (14)
explicit solvent					
MC-MD simulation					
E9_a	82	37 (45)	45 (55)	71 (86)	11 (14)
E9_b	80	36 (45)	44 (55)	70 (87)	10 (13)
Im9_a	59	28 (47)	31 (53)	50 (85)	9 (15)
Im9_b	58	28 (48)	30 (52)	49 (84)	9 (16)
E9-Im9_a	118	55 (47)	63 (53)	101 (86)	17 (14)
E9-Im9_b	120	56 (47)	64 (53)	103 (86)	17 (14)

^aThe values calculated using all, polar or nonpolar, and side-chains or main-chain atoms subgroups are reported separately. Relative percent values are given in parentheses (with respect to all values). See Methods for computational details subgroups definition. ^bE9 unbound from PDB ID: 1fsj; same sequence as in ref 8 and MD model; energy minimized. ^cE9-Im9 X-ray structure (ref 8); energy minimized.

experiments showed that Im9 Glu28 and Tyr53 are among the critical residues experiencing the largest peptide ¹⁵N amide $\Delta\delta$ perturbations upon E9-Im9 complexation.⁵⁸ The comparison of per-residue SASA change with NMR experimental data additionally supports the observation that a determinant of complex-induced chemical shifts is the change of solvent exposure experienced by specific N-H groups.⁸⁶

Salt-Bridge Formation and Pre-Organization: MD versus X-ray Crystallography. Analysis of the salt bridges was carried out along the six MC-MD trajectories (i) to investigate the correspondence between simulations and X-ray crystallography data, (ii) to evaluate the changes of salt-bridge formation upon E9-Im9 complexation, and (iii) to elucidate to what extent residues involved in salt bridges are pre-organized in the unbound states.

Figure 7 summarizes the results for the unbound E9 and Im9 states and the E9-Im9 complex using the statistics over pairs of independent MD runs. All possible salt-bridging residue pairs were taken into account. Only those with a salt-bridge occurrence of at least 10% of the individual simulation time are reported (see Methods Section).

A good correspondence is found comparing the E9-Im9_a and E9-Im9_b simulations with the E9-Im9 complex X-ray crystal structure.⁸ Out of seven salt-bridges in the X-ray model five are also highly formed during our simulations. Im9 Glu30–E9 Arg54 and Im9 Glu41–E9 Lys97 involve hot-spot residues. E9 Asp64–E9 Arg37, E9 Glu59–E9 Lys55, and E9 Glu 100–E9 Arg96 (the latter water-mediated) are formed on the E9 surface. Only E9 Asp40–E9 Arg43 and Im9 Glu45–Im9 Arg75, found in the E9-Im9 complex X-ray crystal structure,⁸ occur less than 2% of the time during our MD simulations. Considering these latter two are on the protein surface their higher formation in the E9-Im9 crystal may be explained by crystal-packing effects. No salt-bridge occurring between Im9 residues has occurrence of at least 10% of the time throughout the E9-Im9_a and E9-Im9_b simulations.

The same analysis shows that a significant number of salt bridges revealed by MD simulations in explicit water would

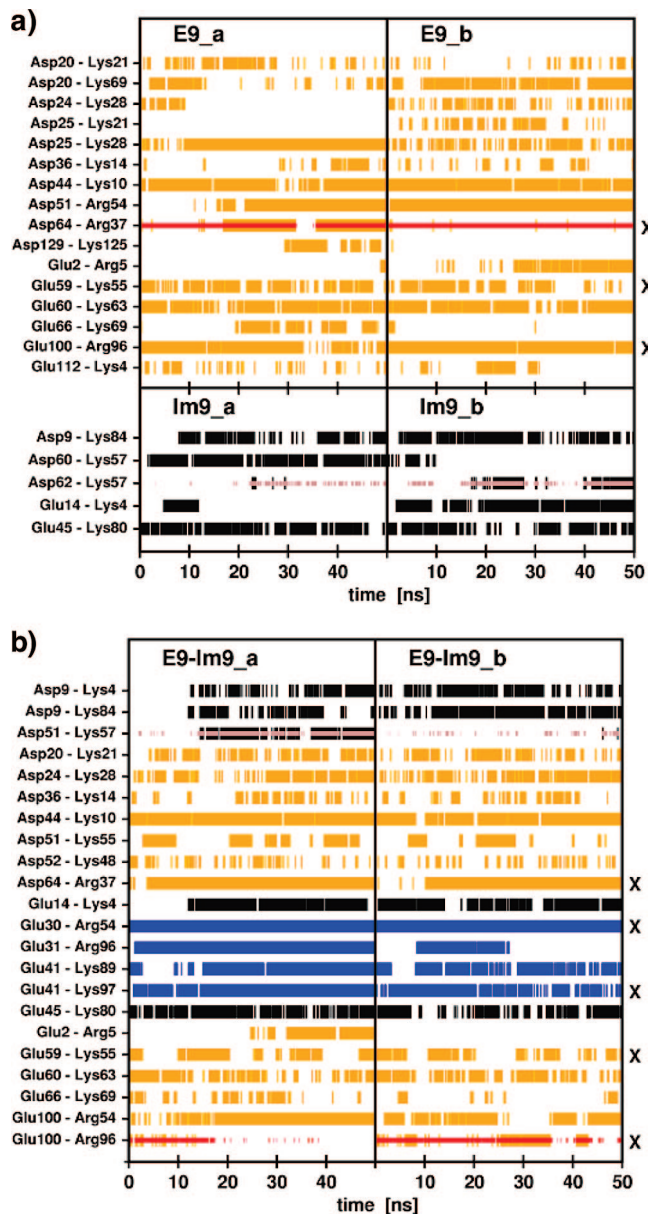


Figure 7. Salt-bridge formation along (a) the unbound states E9_a, E9_b, Im9_a, and Im9_b simulations and (b) the complex E9-Im9_a and E9-Im9_b simulations. Orange and black denote salt bridges between residues of E9 or Im9, respectively. Red and brown denote water-mediated salt bridges between residues of E9 or Im9, respectively. Blue denotes salt bridges between Im9–E9 complex partners. Only salt bridges occurring >10% in at least one of the two MC-MD runs are reported. X labels salt bridges also occurring in the X-ray model. See Table 1 for reference codes and Methods for computational details.

not emerge from an analysis limited to the X-ray crystal model (E9-Im9 complex: 17 out of 22). Remarkably, Im9 Glu31–E9 Arg96 and Im9 Glu41–E9 Lys89 revealed by E9-Im9_a and E9-Im9_b simulations, not present in the X-ray E9-Im9 complex structure, involve interface interactions. The second of these E9-Im9 interface salt bridges can explain the substantial free-energy change upon binding (~ 9 kJ mol⁻¹) calculated based on kinetic and fluorescence experiments.³¹ Sinha et al. recently reported intra- and intermolecular salt bridges, not detected by X-ray crystallography, using short MD simulations of antibody-antigen association.⁸⁷ We suggest that MD simulations may in general improve the description of PPIs based on X-ray models.

The comparison of salt-bridge formation between unbound (Figure 7a) and bound (Figure 7b) states provides information

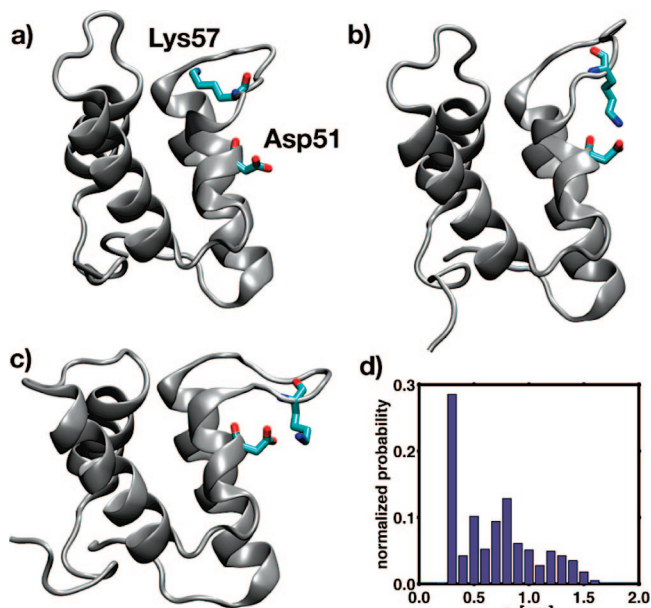


Figure 8. Conformational sampling of the Lys57–Asp51 salt bridge from key snapshots of the accelerated Im9_aMD simulation. Panel (a) shows Lys57 interacting with loop 54–64 residues when the salt bridge is not formed. Alternatively, the Lys57–Asp51 salt bridges can be formed when the 54–64 loop is (b) open or (c) closed. Panel (d) displays the corresponding quasi-canonical re-weighted normalized distribution of the salt-bridge distance.

on the lack of pre-organization of salt-bridge residues. Four intercomplex salt bridges are overall highly formed: Im9 Glu30–E9 Arg54, Im9 Glu31–E9 Arg96, Im9 Glu41–E9 Lys 89, and Im9 Glu41–E9 Lys97. In the first case, we find that E9 Arg 54 is involved in a stable salt-bridge with E9 Asp51 in the unbound state. In the second case, E9 Arg96 forms a stable salt bridge with E9 Glu100 in the unbound state. The third and fourth cases involve a multiple salt-bridge interactions between Im9 Glu41 and both E9 Lys89 and E9 Lys97. In all four cases, it seems evident that no static preorganization occurs; instead, dynamic sampling of favorable binding configurations may be transiently experienced. This is a crucial point to the more fundamental understanding of biomolecular association.^{51,52}

Im9 Asp51–Lys57 Salt-Bridge Dynamics and Its Role in E9–Im9 Recognition As a Surface-Loop Switch. Wallis et al. reported a less favorable (~ 25 kJ mol⁻¹) E9–Im9 binding free energy due to Im9 Asp51Ala mutation, calculated based on a stopped-flow fluorescence experiment.³¹ Asp51Ala mutation is the one affecting E9–Im9 binding the most out of 34 reported. Kleanthous et al.⁸ proposed that mutations Tyr54Ala and Tyr55Ala, among others, are essential for binding affinity and specificity. Instead, it has not yet been clarified the outstanding effect of Asp51Ala mutation in E9–Im9 recognition. We used our MC-MD Im9_a, Im9_b, E9–Im9_a, E9–Im9_b runs and accelerated Im9_aMD simulations to elucidate this point.

The preferred conformations of loop 54–64 are significantly dominated by the interactions formed by Lys57 within the loop region itself. From our MD simulations we observe loop configurations similar to the X-ray model (Figure 4). Additionally, loop dynamics events occurring on longer time scales can be observed on the basis of the Im9_aMD simulation (Figure 5). Our simulations reveal a larger number of key salt-bridge interactions than the X-ray model (Figure 7) also based on Im9_aMD (not shown). Lys57 forms transient hydrogen bonds to Asp60 and Asp62 throughout the Im9_aMD trajectory (Figure 8a). This corresponds to what observed from regular Im9_a and

Im9_b simulations (Figure 4). Interestingly, Im9 displays also two alternative dominant configurations during the 10-ns aMD period, both characterized by a stable Asp51–Lys57 salt bridge (Figures 7 and 8b,c). In the first case, the 54–64 loop is open, in the second further closed than the X-ray model. The re-weighted normalized probability distribution of the Asp51:CG–Lys57:NZ distance confirms that the formation of this salt bridge, observed using aMD only, is also energetically stabilized (see largest peak at ~ 0.35 nm, Figure 8d). Asp51Ala mutation prevents the Asp51–Lys57 salt-bridge anchoring of the 54–64 loop in its ideal configuration for E9 binding, with a consequent decrease of E9–Im9 binding free energy. We summarize this as a molecular switch that shifts Lys57 intraloop interactions to interactions with the more rigid helix III through a Asp51–Lys57 salt bridge.

E9–Im9 Interface Residues: Conformational Spaces and Hydrogen Bonding. We monitored the conformational sampling of side-chain χ_1 and χ_2 dihedral angles and solute–solute hydrogen bonding of E9–Im9 interface residues using simulations E9_a, E9_b, Im9_a, and Im9_b of the free states and simulations E9–Im9_a and E9–Im9_b of the bound state.

Im9 residues Ile22, Cys23, Ser28, Glu30, Leu33, Ser50, Asp51, Ile53, Tyr54, Tyr55, and Asp62 show significant conformational changes upon binding. The largest shifts in dihedral sampling were observed for Glu30, Asp51, and Tyr55. Of these 11 residues, based on NMR studies,^{30,58} 5 are in the group of those displaying the largest $\Delta\delta$ perturbations upon binding (Ser28, Ser50, Ile53, Tyr54, and Tyr55), 2 in the intermediate group (Cys23 and Leu33), and 2 experience small perturbation (Ile22 and Asp51). No significant perturbation could be revealed for Asp62, while no chemical shift was measured for Glu30.

E9 residues Arg54, Asn72, Ser77, Tyr83, Ser84, Lys89, and Gln92 display similar large conformational changes upon binding (not shown). Overall, good correspondence is observed between independent MC-MD runs and a reduction of accessed dihedral space upon binding is expected and found. Remarkable differences between two MC-MD replicas are only evident for distributions of Im9 Asp51 and Tyr55 and can be explained considering the different 54–64 loop sampling for simulations Im9_a and Im9_b (see Figures 2, 3, and 4).

We also investigated the dynamics of the E9–Im9 interface analyzing solute–solute hydrogen bonds. We report separately these results considering all possible E9 and Im9 intramolecular hydrogen bonds highly formed ($>10\%$ time occurrences) in both free and bound states, and E9–Im9 intra- and intermolecular hydrogen bonds formed upon association (Supporting Information, Figure S5). Out of 17 newly formed upon E9–Im9 recognition, 9 involve intermolecular interactions between the E9 and Im9 partners (Supporting Information, Figure S5c). A number of new hydrogen bonds are formed upon E9–Im9 binding (Supporting Information, Figure S5c). Interestingly, this E9–Im9 hydrogen bond network favors the occurrence of those intramolecular hydrogen bonds also present in E9 and Im9 unbound states (compare Supporting Information, Figure S5a,b versus Figure S5c). This is shown for E9 and Im9 in 16 and 18 cases out of 22 and 21, respectively (Supporting Information, Figure S5a,b). Im9 intramolecular Ile53:NH–Ser50:O and Tyr54:NH–Ser50:O and E9–Im9 intermolecular hydrogen bonds Tyr55:OH–Thr87:N and Tyr55:OH–Phe86:O play an important role in E9–Im9 binding thermodynamics.²⁹

These results suggest that in general no preorganization of the residues involved in E9–Im9 recognition is strictly required. Instead, E9–Im9 binding involves shifts of conformational

distributions, reorganization of E9 and Im9 intramolecular hydrogen-bond patterns, and formation of new intermolecular E9-Im9 interactions. However, a few residues at the interface (e.g., E9 Lys97; Im9 Glu41) seems to experience very limited rearrangements upon binding.

Conclusions

Changes of E9 and Im9 dynamics in water upon E9-Im9 colicin DNase-Immunity protein recognition were investigated using a combination of multiple-copy (MC) MD and accelerated MD (aMD) simulation approaches. Overall, our simulations show good general stability and are consistent with the available experimental data. After validation with X-ray crystallography and solution NMR experiments, these simulations provide a dynamic picture at the atomic level of E9-Im9 recognition in water and allow a number of conclusions.

First, Im9 is significantly more flexible than its E9 counterpart, both in its free and bound states. However, due to a decrease in flexibility, the backbone motional correlation increases for both E9 and Im9 proteins upon binding. This effect is larger for Im9, and involves also highly correlated motions for Im9 backbone atom pairs at high distances (> 1.5 nm). Sequence regions Im9 25–45, Im9 50–70, and E9 75–95, encompassing the critical residues at the E9-Im9 interface, display the largest reduction of flexibility upon association.

Second, flexible protein surface loops are a determinant of E9-Im9 recognition specificity in the liquid phase. The Im9 54–64 hinged loop shows high conformational diversity. Through the use of conformational-clustering, salt-bridge, and hydrogen-bonding analyses we propose that Lys57-Asp60 salt bridge is the key intraloop interaction determining the shape of the 54–64 loop, particularly that associated with free Im9 for subsequent complex formation. The dominant loop configurations are modulated by specific hydrogen-bond patterns, which play an essential role in the DNase-immunity protein recognition process. We observe that Im9 loops 23–31 and 54–64 in solution open significantly with respect to the X-ray crystal model.

Third, we observe that upon E9-Im9 association a large part of E9 and Im9 protein surfaces (~ 20 nm²) become inaccessible to the solvent, corresponding to $\sim 17\%$ of the overall solvent accessible surface area of the E9-Im9 complex. A significantly smaller value (~ 14 nm²) would be estimated based on the E9-Im9 and E9 X-ray crystal structures currently available, assuming a rigid-body behavior. This observation points to the remarkable plasticity of the Im9 protein and to its key role in E9-Im9 binding.

Fourth, a significantly higher number of salt bridges is revealed by MD simulations in explicit water than from the X-ray crystal model (17 out of 22 for the E9-Im9 complex). Importantly, among these salt bridges, Im9 Glu31 - E9 Arg96 and Im9 Glu41-E9 Lys89 involve important interface interactions. Based on our simulations, we can thus explain the binding free-energy contribution of Glu41-Lys89 salt bridge (~ 9 kJ mol⁻¹) calculated based on previously reported kinetic and fluorescence experiments.³¹

Fifth, a comparison of conformational sampling of Im9-E9 salt bridges Glu30-Arg54, Glu31-Arg96, Glu41-Lys89, and Glu41-Lys97 with the corresponding unbound Im9 and E9 sampling of these residues shows that no static pre-organization occurs. Instead, dynamic sampling of tens of favorable binding configurations likely occurs during the existence of the initial encounter complex.

Sixth, the atomic-level dynamic information provided by simulations in the liquid phase can be used to follow the conformational sampling of specific residues and their transient

interactions. For example, in relation to Im9 Asp51Ala mutagenesis and its effect on E9-Im9 binding, this allowed us to reconcile Im9 dynamics with liquid-phase kinetic experiments and corresponding calculated thermodynamic data.³¹ Importantly, the same observations would not emerge from an analysis limited to the ensemble-averaged X-ray model.⁸ These results are a first important step to explain the link between the free-energy changes upon binding and Im9 dynamics. We find that the large thermodynamic impact of Asp51Ala mutation (~ 25 kJ mol⁻¹) is due to the resulting loss of anchoring of the 54–64 loop in its ideal closed position for E9 binding. Thus, we suggest that Lys57 acts as an essential molecular switch to shift this Im9 surface loop towards an ideal configuration for E9 recognition. This is achieved at the molecular level by switching Asp 62-Lys57 and Asp60-Lys57 intraloop hydrogen bonds to a Asp51-Lys57 salt bridge. This likely contributes to the high affinity and specificity of E9-Im9 binding.

Seventh, a comparison of dihedral sampling and hydrogen bonding for interface residues of the E9, Im9, and E9-Im9 ensembles shows that no preorganization of residues in E9-Im9 recognition is strictly required. Instead, E9-Im9 binding involves shifts of conformational distributions, reorganization of intramolecular hydrogen bond patterns, and formation of new intermolecular E9-Im9 interactions.

Overall, we demonstrate the importance of an atomic-level and dynamic view of E9-Im9 recognition in solution. Three major general conclusions are supported by our study. First, protein surfaces are significantly perturbed depending on different crystalline-like or liquid-like physico-chemical environments. Second, sufficient sampling (> 100 ns) in MD simulation is essential to capture slow protein dynamics, albeit simulations of such a timescale may seem to be well converged when trapped in specific microstates. Therefore, assessing the convergence of protein simulations should distinguish between local phase-space property-convergence (apparent convergence) from global sampling of the potential energy surface (actual convergence). Indeed, employing independent MC-MD replicas of a same system may be an effective way to enhance the statistics of rare events, as the motion of protein surface loops. However, after quasi-canonical re-weighting of the configurational ensembles, the aMD technique seems a more powerful tool to explore additional regions of the global phase space for solvated macromolecules. We find that aMD (10 ns) outperforms two independent copies of MC-MD (50 ns each) in sampling the biologically-relevant low-frequency motions of the flexible Im9 inhibitor protein. Third, MD simulations provide atomic-level and time-dependent information of biomolecular systems otherwise inaccessible to experiments that can significantly enrich our current view of biomolecular association and assist the essential development of modeling and bioinformatics technology.

Acknowledgment. This work has been funded by the National Science Foundation, the Howard Hughes Medical Institute, and the National Institutes of Health. Additional support from the San Diego Supercomputer Center, Accelrys, Inc., the W. M. Keck Foundation, the National Biomedical Computational Resource and the Center for Theoretical Biological Physics is gratefully acknowledged.

Supporting Information Available: Atom-positional RMSD time series for the backbone excluding the 50–65 loop, the 50–65 loop, the overall backbone (Figure S1); Im9 C^α atom ensemble-averaged chemical shifts (Figure S2); Im9 time series of total solute-solute hydrogen bonds (Figure S3); SASA change upon E9-Im9 complexation of critical interface residues (Figure S4); solute-solute hydrogen bonds occurrences for

interface residues (Figure S5); secondary structure formation (Table S1); and intramolecular hydrogen bonds in the 54–64 loop region (Table S2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- Fink, A. L. *Folding Des.* **1998**, *3*, R9.
- Dobson, C. M. *Methods* **2004**, *34*, 4.
- Wu, T. Y. H.; Wagner, K. W.; Bursulaya, B.; Schultz, P. G.; Deveraux, Q. L. *Chem. Biol.* **2003**, *10*, 759.
- Fry, D. C. *Biopolymers* **2006**, *84*, 535.
- Loregian, A.; Palu, G. *J. Cell. Physiol.* **2005**, *204*, 750.
- Sharma, S. K.; Ramsey, T. M.; Bair, K. W. *Curr. Med. Chem. Anticancer Agents* **2002**, *2*, 311.
- Kleanthous, C.; Hemmings, A. M.; Moore, G. R.; James, R. *Mol. Microbiol.* **1998**, *28*, 227.
- Kleanthous, C.; Kuhlmann, U. C.; Pommer, A. J.; Ferguson, N.; Radford, S. E.; Moore, G. R.; James, R.; Hemmings, A. M. *Nat. Struct. Biol.* **1999**, *6*, 243.
- Kleanthous, C. *Protein-Protein Recognition*; Oxford University Press: Oxford, 2000.
- Sheinerman, F. B.; Honig, B. *J. Mol. Biol.* **2002**, *318*, 161.
- Kortemme, T.; Baker, D. *Curr. Opin. Chem. Biol.* **2004**, *8*, 91.
- Reichmann, D.; Rahat, O.; Cohen, M.; Neuvirth, H.; Schreiber, G. *Curr. Opin. Struct. Biol.* **2007**, *17*, 67.
- Arkin, M. R.; Wells, J. A. *Nat. Rev. Drug Discovery* **2004**, *3*, 301.
- Gonzalez-Ruiz, D.; Gohlke, H. *Curr. Med. Chem.* **2006**, *13*, 2607.
- Rajapakse, H. *Curr. Top. Med. Chem.* **2007**, *7*, 966.
- Elcock, A.; Sept, D.; McCammon, J. *J. Phys. Chem. B* **2001**, *105*, 1504.
- van Dijk, A. D.; Bonvin, A. M. *Bioinformatics* **2006**, *22*, 2340.
- Humphris, E. L.; Kortemme, T. *PLoS Comput. Biol.* **2007**, *3*, e164.
- Conchillo-Sole, O.; de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Daura, X.; Ventura, S. *BMC Bioinf.* **2007**, *8*, 65.
- Chothia, C.; Janin, J. *Nature* **1975**, *256*, 705.
- Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13.
- Sundberg, E. J.; Urrutia, M.; Braden, B. C.; Isern, J.; Tsuchiya, D.; Fields, B. A.; Malchiodi, E. L.; Tormo, J.; Schwarz, F. P.; Mariuzza, R. A. *Biochemistry* **2000**, *39*, 15375.
- Fernandez, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 113.
- Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383.
- Kim, M.; Sun, Z. Y. J.; Byron, O.; Campbell, G.; Wagner, G.; Wang, J. H.; Reinherz, E. L. *J. Mol. Biol.* **2001**, *312*, 711.
- Hancox, E. L.; Halford, S. E. *Biochemistry* **1997**, *36*, 7577.
- Delano, W. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14.
- Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1.
- Li, W.; Keeble, A. H.; Giffard, C.; James, R.; Moore, G. R.; Kleanthous, C. *J. Mol. Biol.* **2004**, *337*, 743.
- Boetzel, R.; Czisch, M.; MacDonald, C. J.; Kaptein, R.; Hemmings, A. M.; James, R.; Kleanthous, C.; Moore, G. R. *J. Biomol. NMR* **1998**, *12*, 567.
- Wallis, R.; Leung, K. Y.; Osborne, M. J.; James, R.; Moore, G. R.; Kleanthous, C. *Biochemistry* **1998**, *37*, 476.
- Whittaker, S. B.; Boetzel, R.; MacDonald, C.; Lian, L. Y.; James, R.; Kleanthous, C.; Moore, G. R. *J. Biomol. NMR* **1999**, *14*, 201.
- Takahashi, H.; Nakanishi, T.; Kami, K.; Arata, Y.; Shimada, I. *Nat. Struct. Biol.* **2000**, *7*, 220.
- Bonvin, A. M.; Boelens, R.; Kaptein, R. *Curr. Opin. Chem. Biol.* **2005**, *9*, 501.
- Burz, D. S.; Dutta, K.; Cowburn, D.; Shekhtman, A. *Nat. Protoc.* **2006**, *1*, 146.
- Baminger, B.; Ludwiczek, M. L.; Kontaxis, G.; Knapp, S.; Konrat, R. *J. Biomol. NMR* **2007**, *38*, 133.
- van den Bremer, E. T.; Keeble, A. H.; Visser, A. J.; van Hoek, A.; Kleanthous, C.; Heck, A. J.; Jiskoot, W. *Biochemistry* **2004**, *43*, 4347.
- Tang, C.; Iwahara, J.; Cloutier, G. M. *Nature* **2006**, *444*, 383.
- van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. *Curr. Opin. Struct. Biol.* **2008**.
- Zagrovic, B.; Pande, V. S. *Biophys. J.* **2004**, *87*, 2240.
- Baron, R.; Bakowies, D.; van Gunsteren, W. F. *Angew. Chem. Int. Ed.* **2004**, *43*, 4055.
- Baron, R.; McCammon, J. A. *Biochemistry* **2007**, *46*, 10629.
- van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 4064.
- Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589.
- Berendsen, H. *Simulating the Physical World*; Cambridge University Press: Cambridge, U. K., 2007.
- Minh, D. D.; Bui, J. M.; Chang, C. E.; Jain, T.; Swanson, J. M.; McCammon, J. A. *Biophys. J.* **2005**, *89*, L25.
- Missimer, J. H.; Steinmetz, M. O.; Baron, R.; Winkler, F. K.; Kammerer, R. A.; Daura, X.; van Gunsteren, W. F. *Protein Sci.* **2007**, *16*, 1349.
- Zoete, V.; Meuwly, M.; Karplus, M. *Proteins* **2005**, *61*, 79.
- Chang, C. E.; McLaughlin, W. A.; Baron, R.; Wang, W.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7456.
- Wong, S. E.; Baron, R.; McCammon, J. A. *Biopolymers* **2008**, *89*, 916.
- Tsai, C. J.; Ma, B.; Sham, Y. Y.; Kumar, S.; Nussinov, R. *Proteins* **2001**, *44*, 418.
- Keskin, O.; Ma, B. Y.; Nussinov, R. *J. Mol. Biol.* **2005**, *345*, 1281.
- Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 927.
- Shirts, M. R.; Pande, V. S. *Phys. Rev. Lett.* **2001**, *86*, 4983.
- Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919.
- Hamelberg, D.; McCammon, J. A. *Ann. Rep. Comput. Chem.* **2006**, *2*, 221.
- Osborne, M. J.; Breeze, A. L.; Lian, L. Y.; Reilly, A.; James, R.; Kleanthous, C.; Moore, G. R. *Biochemistry* **1996**, *35*, 9505.
- Osborne, M. J.; Wallis, R.; Leung, K. Y.; Williams, G.; Lian, L. Y.; James, R.; Kleanthous, C.; Moore, G. R. *Biochem. J.* **1997**, *323*, 823.
- Whittaker, S. B.; Czisch, M.; Wechselberger, R.; Kaptein, R.; Hemmings, A. M.; James, R.; Kleanthous, C.; Moore, G. R. *Protein Sci.* **2000**, *9*, 713.
- Keeble, A. H.; Kirkpatrick, N.; Shimizu, S.; Kleanthous, C. *Biochemistry* **2006**, *45*, 3243.
- Case, D. A.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Merz, K.; Pearlman, D.; Crowley, M.; Walker, R.; Zhang, W.; Wang, B.; Hayik, A.; Roiberg, A.; Seabra, G.; Wong, K.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Morgan, J.; Hornak, V.; Cui, G.; Beroza, P.; Matthews, D.; Schffmeister, C.; Ross, W.; Kollman, P. *AMBER*, 9th ed.; University of California, San Francisco, 2006.
- Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
- Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712.
- Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021.
- Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
- Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. *J. Chem. Phys.* **2000**, *113*, 10913.
- Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- Christen, M.; van Gunsteren, W. F. *J. Comput. Chem.* **2008**, *29*, 157.
- Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput. Aided Mol. Des.* **1994**, *8*, 695.
- Crippen, G. M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *64*, 42.
- Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.
- Voter, A. F. *J. Chem. Phys.* **1997**, *106*, 4665.
- Hamelberg, D.; de Oliveira, C. A.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 155102.
- Markwick, P. R.; Bouvignies, G.; Blackledge, M. J. *Am. Chem. Soc.* **2007**, *129*, 4724.
- de Oliveira, C. A. F.; Hamelberg, D.; McCammon, J. A. *J. Phys. Chem. B* **2006**, *110*, 22695.
- de Oliveira, C. A.; Hamelberg, D.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 175105.
- McLachlan, A. D. *J. Mol. Biol.* **1979**, *128*, 49.
- Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
- Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215.
- Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1995**, *252*, 492.
- Daura, X.; van Gunsteren, W. F.; Mark, A. E. *Proteins* **1999**, *34*, 269.
- Hubbard, S. J.; Thornton, J. M. *NACCESS*; University College, London, U. K., 1993.
- Spitzfaden, C.; Weber, H. P.; Braun, W.; Kallen, J.; Wider, G.; Widmer, H.; Walkinshaw, M. D.; Wuthrich, K. *FEBS Lett.* **1992**, *300*, 291.
- Sinha, N.; Li, Y. L.; Lipschultz, C. A.; Smith-Gill, S. J. *Cell Biochem. Biophys.* **2007**, *47*, 361.