PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Probing the structural dynamics of the CRISPR-Cas9 RNA-guided DNA-cleavage system by coarse-grained modeling

Wenjun Zheng*

Department of Physics, State University of New York at Buffalo, Buffalo, New York 14260

ABSTRACT

In the adaptive immune systems of many bacteria and archaea, the Cas9 endonuclease forms a complex with specific guide/scaffold RNA to identify and cleave complementary target sequences in foreign DNA. This DNA targeting machinery has been exploited in numerous applications of genome editing and transcription control. However, the molecular mechanism of the Cas9 system is still obscure. Recently, high-resolution structures have been solved for Cas9 in different structural forms (e.g., unbound forms, RNA-bound binary complexes, and RNA-DNA-bound tertiary complexes, corresponding to an inactive state, a pre-target-bound state, and a cleavage-competent or product state), which offered key structural insights to the Cas9 mechanism. To further probe the structural dynamics of Cas9 interacting with RNA and DNA at the amino-acid level of details, we have performed systematic coarse-grained modeling using an elastic network model and related analyses. Our normal mode analysis predicted a few key modes of collective motions that capture the observed conformational changes featuring large domain motions triggered by binding of RNA and DNA. Our flexibility analysis identified specific regions with high or low flexibility that coincide with key functional sites (such as DNA/RNA-binding sites, nuclease cleavage sites, and key hinges). We also identified a small set of hotspot residues that control the energetics of functional motions, which overlap with known functional sites and offer promising targets for future mutagenesis efforts to improve the specificity of Cas9. Finally, we modeled the conformational transitions of Cas9 from the unbound form to the binary complex and then the tertiary complex, and predicted a distinct sequence of domain motions. In sum, our findings have offered rich structural and dynamic details relevant to the Cas9 machinery, and will guide future investigation and engineering of the Cas9 systems.

Proteins 2017; 85:342–353.

Key words: Cas9; coarse-grained modeling; CRISPR; elastic network model; flexibility; hotspot residue; normal mode analysis; transition pathway.

INTRODUCTION

Bacteria and archaea target and degrade foreign DNA using adaptive immune systems based on the CRISPR (clustered regularly interspaced short palindromic repeats) loci and CRISPR associated (Cas) genes.[1–3] There are three types of CRISPR-Cas systems (I, II, and III). The type I and III CRISPR-Cas systems use multi-protein complexes for RNA-guided DNA targeting.[1–3] In contrast, in the type II CRISPR-Cas systems, a single Cas9 endonuclease associates with a hybrid structure of small CRISPR RNAs (crRNAs) and a trans-activating crRNA (tracrRNA), which together identify and cleave complementary target sequences in foreign DNA.[4,5] A key step of this DNA targeting process is the R-loop formation in which a 20-nucleotide guide RNA segment of the crRNAs invades a double-stranded DNA (dsDNA) to form an RNA-DNA hybrid helix with the target DNA strand while displacing the non-target DNA strand.[6] The accuracy of DNA targeting is ensured by base pairing between the target DNA strand and the guide RNA segment, and recognition of a short protospacer adjacent motif (PAM) with 2–4 base pairs. In particular, the target complementarity of a seed sequence of 10–12 nucleotides

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: American Heart Association; Grant number: #14GRNT18980033; Grant sponsor: National Science Foundation; Grant number: #0952736.

*Correspondence to: Wenjun Zheng; 239 Fronczak Hall, Buffalo, NY 14260. E-mail: wjzheng@buffalo.edu

Received 20 June 2016; Revised 31 October 2016; Accepted 28 November 2016
Published online 9 December 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.25229

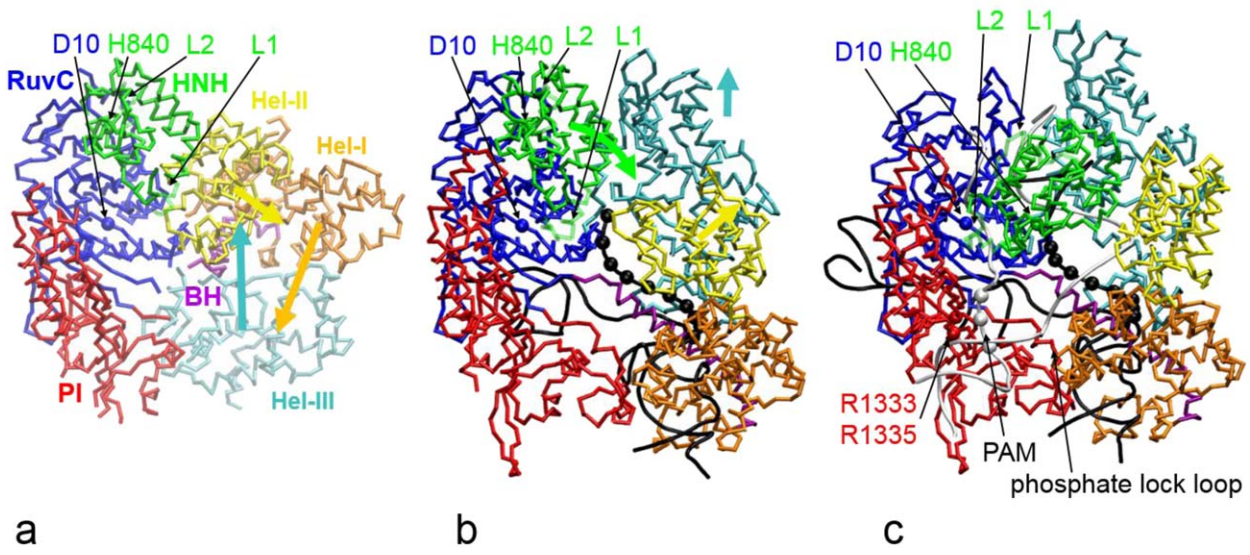

Figure 1

Structural architecture of SpyCas9 in (**a**) the unbound form (PDB id: 4CMP), (**b**) the binary complex with sgRNA (PDB id: 4ZT0), and (**c**) the tertiary complex with sgRNA and dsDNA (PDB id: 5F9R). The various domains are colored as follows: RuvC (blue), HNH (green), PI (red), bridge helix (purple), Hel-I (orange), Hel-II (yellow), and Hel-III (cyan). In (b) and (c), the sgRNA is in black (with 10 seed nucleotides shown as beads). In (c), the dsDNA is in silver (with three PAM nucleotides shown as beads). The cleavage sites of two nuclease domains (D10 of RuvC and H840 of HNH) and the PAM recognition residues (R1333 and R1335 of PI) are shown as beads colored by domain. In (a) and (b), observed domain motions upon forming the binary and the tertiary complex are marked by thick arrows colored by domain. The three structures have been superimposed along the RuvC domain.

at the 3' end of the guide RNA segment is critical for DNA recognition and cleavage.[7–10] Following the R-loop formation, the target strand and the non-target strand of the dsDNA are cleaved by the HNH and the RuvC nuclease domain, respectively,[7,11] resulting in double-strand DNA breaks (DSBs). The Cas9 genes from different bacteria and archaea vary widely in sequence and size, and can be grouped into three subfamilies: types II-A, II-B, and II-C.[12] The type II-C Cas9 proteins are ~300 amino acids shorter in sequence than the type II-A ones, which makes it easier to deliver them to somatic tissues for *in vivo* applications. By engineering a synthetic single-guide RNA (sgRNA) that fuses the crRNA and the tracrRNA to target loci in eukaryotes, the Cas9-sgRNA systems can be exploited to generate site-specific DSBs subsequently repaired by either nonhomologous end joining or homologous recombination,[9,13] which allows precise modification of genomic sequences in the vicinity of the Cas9-generated DSBs. Additionally, a catalytically inactive Cas9 fused to transcriptional activation or repression proteins can be used for site-specific transcriptional control.[14–16] Thanks to its simplicity and flexibility, the Cas9-sgRNA system has been widely employed in genome engineering, transcriptional control, and much more.[17–21] Many of these applications require a high specificity in DNA targeting[22] to avoid unwanted off-target cleavage (e.g., in therapeutic applications). To engineer a Cas9-sgRNA system with minimal size and high specificity,[23] it is critical to gain detailed insights to the molecular mechanism of the Cas9-mediated RNA-guided DNA targeting and cleavage.

Toward elucidating the structural basis of the Cas9-sgRNA system, a number of crystal structures were solved for type II-A[24–28] and type II-C[25,29] Cas9 proteins. A majority of them are for the *Streptococcus pyogenes* Cas9 (SpyCas9)—a prototypical type II-A Cas9 protein that recognizes a 5′-NGG-3′ PAM sequence located 3 base pairs from the cleavage site on the non-target DNA strand. These distinct structures trapped unbound SpyCas9,[25] SpyCas9 bound to sgRNA,[26] SpyCas9 bound to sgRNA and a single target DNA strand without[24] or with[28] a partial PAM-containing non-target DNA strand, and SpyCas9 bound to both sgRNA and target dsDNA.[27] In particular, three of them captured SpyCas9 in three key functional states: an unbound SpyCas9 in the inactive state,[25] a binary SpyCas9-sgRNA complex in the pre-target-bound state,[26] and a tertiary SpyCas9-sgRNA-dsDNA complex in the cleavage-competent state,[27] which will be the focus of this study. The Cas9 structure features a bi-lobed architecture comprising an alpha-helical recognition (REC) lobe and a nuclease (NUC) lobe (see Fig. 1). The NUC lobe consists of two well-conserved nuclease domains (HNH and RuvC) and a divergent PAM-interacting (PI) domain.[24,29] The REC lobe contains three alpha-helical domains (Hel-I, Hel-II,

and Hel-III) which are structurally divergent between different Cas9 proteins.[29] In between the two lobes lies a highly conserved Arginine-rich bridge helix (BH) implicated in sgRNA binding by a mutational study.[30] A structural comparison between the unbound SpyCas9 and the binary SpyCas9-sgRNA complex reveals substantial conformational changes in SpyCas9 upon sgRNA loading,[25] including large displacements of the Hel-I (Hel-III) domain toward the PI (HNH) domain, and ordering of the PI domain in the PAM-recognition region [see Fig. 1(a,b)],[26] resulting in the formation of a central channel between the two lobes to accommodate sgRNA. These large structural changes support the key role of sgRNA loading as prerequisite for Cas9 activation. A structural comparison between the binary and the tertiary complex shows further conformational rearrangements in SpyCas9 upon DNA binding and PAM recognition,[27] including modest shifts of the Hel-II and Hel-III domains, a large rotation and displacement of the HNH domain toward the target DNA strand, and the formation of a hybrid duplex between the guide segment of sgRNA and the target DNA strand [see Fig. 1(b,c)], resulting in a more compact Cas9 structure.[27] The above large conformational changes between the two lobes were also visualized by negative-stain electron microscopy (EM)[25] and cryo-EM[27] at lower resolutions ($\geq$4.5 Å).[1–3] These Cas9 structures have opened door for structure-based investigations of Cas9-based DNA targeting and cleavage mechanism and engineering of high-specificity Cas9 systems.[23] Nevertheless, the detailed structural dynamics of Cas9 interacting with sgRNA and dsDNA remain largely unknown, which cannot be directly obtained from these static structures.

Molecular Dynamics (MD) simulation is the method of choice for exploring protein dynamics under physiological conditions at atomic resolution.[31] It has been widely used to study various protein and nucleic acid complexes.[32,33] Despite fast-growing computing power, MD simulation is still highly expensive, especially for large biomolecular complexes in explicit solvent. Typical speeds of MD simulation on a single computer node equipped with graphics processing unit are $\leq$10 nanoseconds per day for a system of 10^5–10^6 atoms, although much higher speeds (e.g., several microseconds per day) can be reached on a massively parallelized or special-purpose supercomputer.[34] Owning to its low speed, MD simulation cannot readily access microseconds–seconds time scales which are relevant to many conformational transitions in biomolecular complexes. As estimated by a recent analysis[35] of an *in vitro* study of DNA interrogation by the Cas9-sgRNA system,[8] the kinetic rates of Cas9-crRNA association and isomerization are 0.1–1/second, which are too slow to be directly simulated by MD.

To overcome the time-scale limit of MD simulation, coarse-grained modeling methods have been developed using reduced protein representations (e.g., one bead per amino acid residue) and simplified energy functions (e.g., harmonic potential).[36,37] As a popular coarse-grained model, the elastic network model (ENM) is constructed by connecting nearby C_α atoms of protein residues with harmonic springs.[38–40] Despite its simplicity, the normal mode analysis (NMA) of ENM can be used to predict a few low-frequency modes of collective motions, which often capture those conformational changes observed between experimentally solved protein structures in different functional states.[41] Numerous studies have established ENM as a useful and efficient means to probe structural dynamics of large biomolecular complexes (including a DNA translocase/helicase[32,42] and protein-RNA complexes like ribosome[43,44]) with virtually no limit in timescale or system size (see reviews[45,46]).

In this study, we will employ ENM-based modeling/analysis tools to gain detailed insights to the structural dynamics of Cas9 in various functional states critical to its DNA targeting and cleavage function. Our structure-based modeling will complement other biophysical modeling studies of Cas9 systems,[35] and guide future experimental investigation and engineering of Cas9 proteins.

MATERIALS AND METHODS

Elastic network model (ENM) and ENM-based flexibility and hotspot analysis

In an ENM, the structure of a protein-DNA/RNA complex is represented as a network of coarse-grained beads corresponding to the C_α atoms of protein residues and the C_4' atoms of DNA/RNA.[32] We also tested alternative DNA/RNA coarse-graining schemes using three beads per nucleic acid, and found very similar results. Harmonic springs link all pairs of beads within a cutoff distance R_c chosen to be 25 Å.[47] A large R_c ensures good local connectivity of the ENM to avoid an unwanted tip effect.[48]

The ENM potential energy is:

$$E=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{i-1}k_{ij}\theta(R_c-d_{ij,0})(d_{ij}-d_{ij,0})^2, \qquad (1)$$

where N is the number of beads, $\theta(x)$ is the Heaviside function, d_{ij} is the distance between bead i and j, $d_{ij,0}$ is the value of d_{ij} as given by a crystallographic structure (e.g., a structure of SpyCas9 in complex with sgRNA and dsDNA). The spring constant k_{ij} is set to be $(4/d_{ij,0})^6$ for non-bonded interactions (following Refs. 47 and 49) and 10 for bonded interactions (in arbitrary unit). We have also tried other ENM schemes (e.g., $k_{ij} \propto d_{ij,0}^{-2}$), and verified that the NMA results are not sensitive to the choice of ENM schemes.

The NMA solves the following eigen equation for a Hessian matrix H which is obtained by calculating the

second derivatives of ENM potential energy (see Ref. 50):

$$HV_m = \lambda_m V_m, \tag{2}$$

where λ_m and V_m represent the eigenvalue and eigenvector of mode m, respectively. After excluding six zero modes corresponding to three rotations and three translations, we number non-zero modes starting from 1 in the order of ascending eigenvalue.

For mode m, we use a perturbation analysis to assess how much the eigenvalue changes (represented as $\delta\lambda_m$) in response to a perturbation at a chosen residue position[51–53] (i.e., by uniformly weakening the springs connected to this position to mimic an Alanine substitution). Then we average $\delta\lambda_m/\lambda_m$ over the lowest $M = 20$ modes to assess the overall dynamic importance of this residue position[54]:

$$\langle \delta\lambda/\lambda \rangle = \frac{1}{M}\sum_{m=1}^{M} \delta\lambda_m/\lambda_m. \tag{3}$$

To validate ENM-based NMA, we compare each mode (i.e., mode m) with the observed structural change X_{obs} between two superimposed protein structures by calculating the following overlap[55]:

$$I_m = X_{\text{obs}} \cdot V_m / |X_{\text{obs}}|, \tag{4}$$

where $|I_m|$ varies between 0 and 1 with higher value meaning greater similarity. I_m^2 gives the fractional contribution of mode m to X_{obs}. The cumulative squared overlap $C_M = \sum_{m=1}^{M} I_m^2$ gives the fractional contribution of the lowest $M = 20$ modes to X_{obs}.[55] The linear combination of the lowest $M = 20$ modes predicts the following conformational changes: $X_M = |X_{\text{obs}}| \sum_{m=1}^{M} I_m V_m$.

To assess the local flexibility at individual residue positions as described by the lowest $M = 20$ modes, we define the following cummulative flexibility (CF)[54,56]:

$$\text{CF}_n = \sqrt{\sum_{m=1}^{M} \left(|V_{m,nx}|^2 + |V_{m,ny}|^2 + |V_{m,nz}|^2 \right)}, \tag{5}$$

where $V_{m,nx}$, $V_{m,ny}$, and $V_{m,nz}$ are the x, y, and z component of mode m's eigenvector at residue position n.

Coarse-grained transition pathway modeling by iENM

We previously developed an interpolated ENM (iENM) protocol to construct a transition pathway (i.e., a series of C_α-only intermediate conformations) between a beginning and an end protein conformation by solving the saddle points of a double-well potential built from these two conformations.[57] This protocol is available at the following webserver: http://enm.lobos.nih.gov/start_ienm.html. Here, we apply this method to the unbound-to-binary and the binary-to-tertiary conformational transitions of SpyCas9. To include sgRNA of the binary complex (sgRNA and dsDNA of the tertiary complex) in the modeling of the first (second) transition, we initiate the pathway construction from the end conformation of each transition toward the beginning conformation, and then reverse the pathways to obtain the forward transition pathways (the iENM-predicted pathways are reversible). In the beginning conformation of the unbound-to-binary transition, we assume the sgRNA adopts the same conformation as in the binary complex but without interacting with SpyCas9. In the beginning conformation of the binary-to-tertiary transition, we assume the dsDNA adopts a B-form double-helix conformation (modeled by the 3DNA server[58] at http://w3dna.rutgers.edu/) but without interacting with SpyCas9.

Using the iENM-predicted transition pathway, we can determine the motional order of various protein domains by calculating and comparing the reaction coordinates for different domains (denoted RC_S for domain S, for details see Ref. 59). RC_S varies from 0 to 1 as the transition proceeds from the beginning to the end conformation. For two domains S and S', if $\text{RC}_S < \text{RC}_{S'}$ along the pathway, we infer that the motion of S' precedes that of S. Here we plot RC_S as a function of RC_{all} (i.e., RC of all residues) to track the evolution of RC_S along the transition pathway.

RESULTS AND DISCUSSION

First, we present results of using ENM-based NMA to analyze key modes of collective motions, flexibility, and hotspot residues in the unbound SpyCas9, the binary complex, and the tertiary complex of SpyCas9 with sgRNA and dsDNA. Then we model the two conformational transitions from the unbound SpyCas9 to the binary complex and then the tertiary complex.

NMA predicts key collective motions, flexible regions, and hotspot residues in SpyCas9

As revealed by previous structural studies,[25–28] SpyCas9 undergoes large conformational changes upon binding with sgRNA and then dsDNA (see Fig. 1). Here, we have analyzed these experimentally observed structural changes in terms of collective motions as predicted by the ENM-based NMA (see Methods).

Starting from the unbound structure of SpyCas9 (PDB id: 4CMP, chain B), we constructed a C_α-only ENM by linking all pairs of residues within a cutoff distance with harmonic springs (see Methods). Then we performed NMA (see Methods) to obtain a spectrum of total 3498 modes, and focused on the lowest 20 modes, each

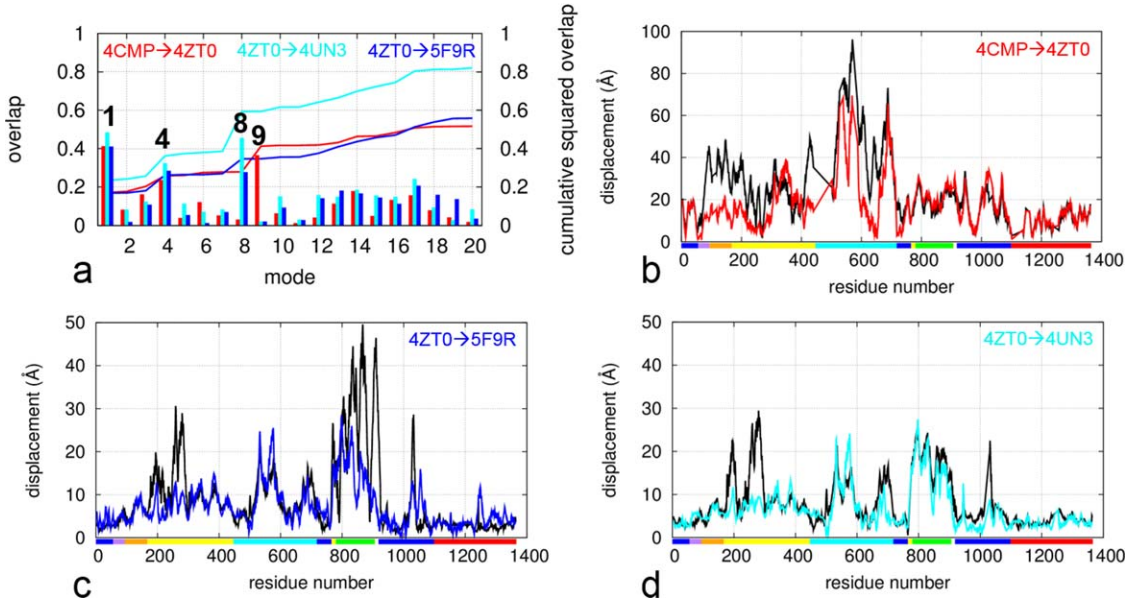

Figure 2

Results of NMA for SpyCas9: (**a**) Overlap and cumulative squared overlap as a function of mode for the following observed conformational changes: from the unbound structure (PDB id: 4CMP) to the binary complex (PDB id: 4ZT0) colored in red, from the binary complex to the cleavage-competent tertiary complex (PDB id: 5F9R) colored in blue, from the binary complex to another tertiary complex (PDB id: 4UN3) colored in cyan. The top-contributing modes are labeled. (**b**), (**c**), and (**d**) show the per-residue displacement amplitude for the conformational changes X_{20} predicted by combining the lowest 20 modes (red: 4CMP → 4ZT0, blue: 4ZT0→5F9R, cyan: 4ZT0 → 4UN3) in comparison with the observed conformational changes X_{obs} (black). Various SpyCas9 domains are marked by horizontal bars with the same color coding as Figure 1.

describing a specific pattern of collective motions involving various SpyCas9 domains (see Fig. 3). The functional importance of these domain motions was validated by many previous studies (see reviews[45,46]). Similarly, we performed ENM-based NMA for the binary SpyCas9-sgRNA complex (PDB id: 4ZT0, chains A and B), and the tertiary SpyCas9-sgRNA-dsDNA complex (PDB id: 5F9R, chains A–D).

For validation of the lowest 20 modes solved from the unbound structure, we assessed how well they collectively capture the observed conformational changes from the unbound structure to the binary complex. These structural changes are thought to underlie the SpyCas9 activation by sgRNA loading via either direct SpyCas9-sgRNA interactions or indirect allosteric couplings.[25] We calculated the overlap between each mode and the observed conformational changes and the cumulative squared overlap of the lowest 20 modes (see Methods). Encouragingly, ~52% of the observed large conformational changes (RMSD = 29 Å) are captured by the lowest 20 modes, with mode 1 and 9 contributing the most [see Fig. 2(a)]. The collective motions predicted by these modes are energetically favored by the unbound structure and can be driven by binding of SpyCas9 to sgRNA. Mode 1 describes a downward rotation of the Hel-I domain together with an upward rotation of the Hel-III domain [see Fig. 3(a)], which partially account for the large displacements of these alpha-helical domains from the unbound structure to the binary complex [see Fig. 1(a,b)]. Mode 9 describes large concerted rotations of the alpha-helical domains, accompanied by smaller displacements of the PI and the HNH domain [see Fig. 3(b)]. The combination of the lowest 20 modes accurately describes the amplitude of motions in the RuvC, the HNH, and the PI domain of the NUC lobe, while the large motions in the three alpha-helical domains of the REC lobe are underestimated [see Fig. 2(b)].

Similarly, we evaluated the lowest 20 modes solved from the binary complex in comparison with the conformational changes observed from the binary complex to two tertiary complexes—one in a cleavage-competent state (PDB id: 5F9R) and another in a post-cleavage product state (PDB id: 4UN3). Reassuringly, the lowest 20 modes capture 56% (82%) of the observed conformational changes in SpyCas9 from 4ZT0 to 5F9R (4UN3) with RMSD = 13 Å (10 Å). The following three top-contributing modes [see Fig. 2(a)] were identified:

Mode 1 predicts concerted upward rotations of the Hel-I, the Hel-II, and the lower PI domain toward the HNH domain, leading to further closing of the central channel between the NUC and the REC lobe [see Fig. 3(c)];

Mode 4 predicts concerted upward rotations of the three alpha-helical domains, a downward rotation of the

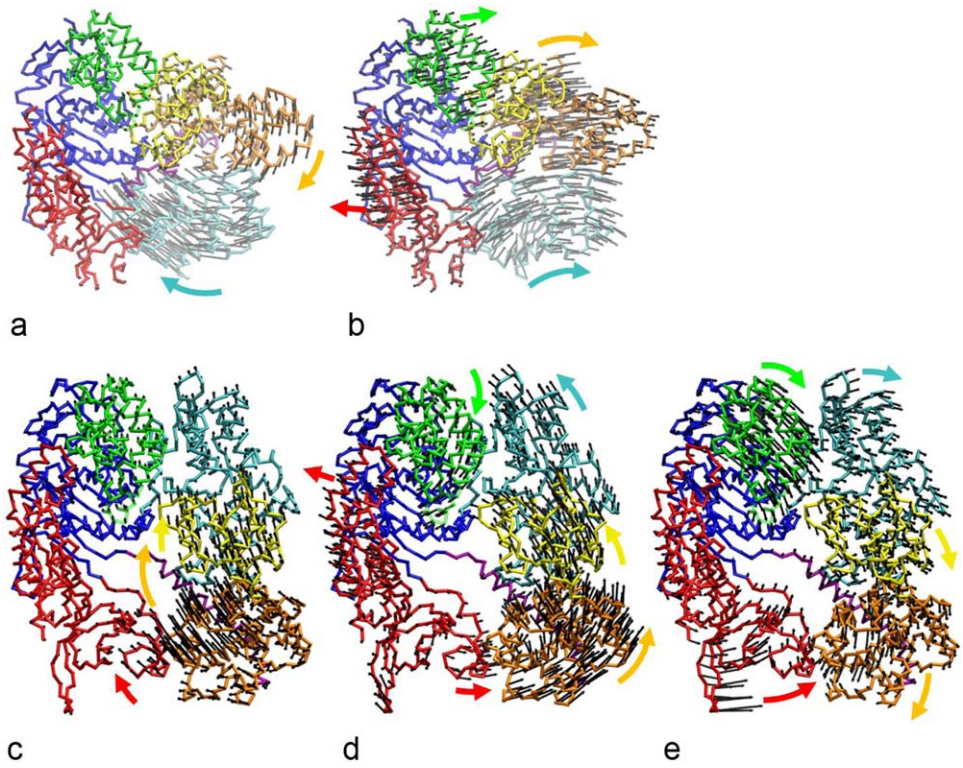

Figure 3

Structural motions predicted by top-contributing modes: (**a**) and (**b**) show the vector plots for motions of residues which were predicted by mode 1 and mode 9 of the unbound structure (PDB id: 4CMP). (**c**), (**d**), and (**e**) show the vector plots for motions of residues which were predicted by mode 1, 4, and 8 of the binary complex (PDB id: 4ZT0). The color coding for SpyCas9 domains is the same as Figure 1. Key domain motions are shown by thick arrows colored by domain.

HNH domain, and a leftward shift of the upper PI domain [see Fig. 3(d)];

Mode 8 predicts concerted downward rotations of the three alpha-helical domains, a downward rotation of the HNH domain, and closing of the lower PI domain toward the Hel-I domain, leading to further closing of the central channel [see Fig. 3(e)].

For the conformational changes from 4ZT0 to 5F9R, the combination of the lowest 20 modes accurately describes the amplitude of motions in most domains except the N-terminal half of the Hel-II domain and the HNH domain [see Fig. 2(c)]. For the conformational changes from 4ZT0 to 4UN3, the combination of the lowest 20 modes nicely describes the amplitude of motions in all domains but the N-terminal half of the Hel-II domain [see Fig. 2(d)].

In sum, the 20 lowest modes (i.e., <0.6% of all modes) capture 52–82% of the observed functionally important structural changes in SpyCas9 upon binding with sgRNA and dsDNA (see Fig. 1). These encouraging findings strongly support the use of ENM-based NMA to analyze functional motions in SpyCas9. Despite overall agreement, the motions in some domains are not well described by the lowest 20 modes. We reason that these motions are energetically unfavorable and may therefore occur later during the conformational transitions (see below).

To further analyze the flexibility of different domains in SpyCas9, we used the lowest 20 modes to calculate the cumulative flexibility (CF, see Methods) at individual residue positions of SpyCas9 in the unbound form, the binary complex, and the tertiary complex [See Fig. 4(a)]. High/low CF implies high/low flexibility. The following domains exhibit high flexibility in all three structures: the C-terminal section of Hel-I domain (residues 315 to 447), the Hel-III domain, the HNH domain, and the PI domain [see Fig. 4(a)]. Among them, the HNH domain is known to be highly dynamic, which is disordered in a crystal structure of SpyCas9,[24] and exists in a dynamic equilibrium between an inactive and an active configuration as shown by a single-molecule study.[60] The PI domain is involved in binding of dsDNA and recognition of PAM,[28] and is partially disordered in the unbound structure of SpyCas9.[25] The intrinsic flexibility of the PI domain may enable it to interact readily with specific nucleotides of the target DNA. Overall, the unbound form exhibits higher flexibility than the binary and the tertiary complex [see Fig. 4(a)], particularly in the Hel-I,

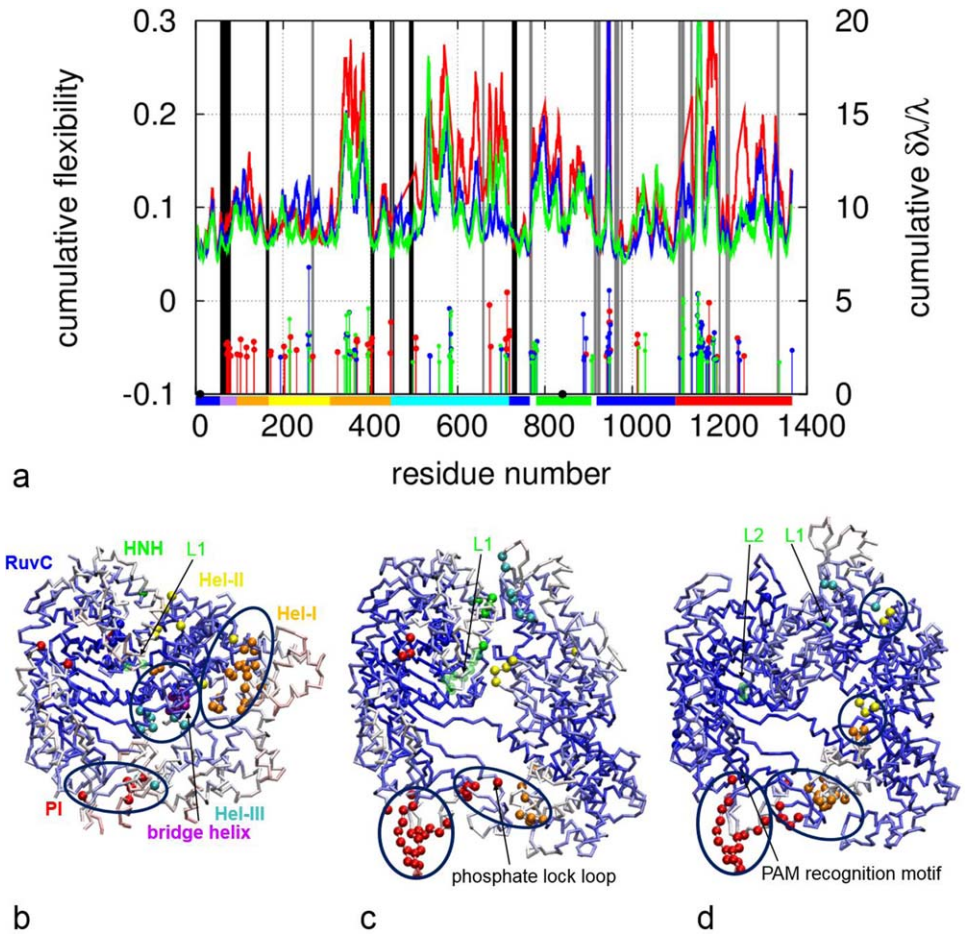

Figure 4

Results of flexibility and hotspot analysis: (**a**) Cumulative flexibility (CF) and $\langle \delta\lambda/\lambda \rangle$ as a function of residue position for the unbound structure (red), the binary complex (blue), and the cleavage-competent tertiary complex (green), where only $\langle \delta\lambda/\lambda \rangle$ values for the top 5% hotspot residues are shown as impulses topped with a dot, the positions of sgRNA-binding and dsDNA-binding residues are marked by vertical lines colored in black and gray, the positions of two nuclease cleavage sites (D10 and H840) are marked by two black dots, and various SpyCas9 domains are marked by horizontal bars with the same color coding as Figure 1. (**b**), (**c**), and (**d**) show the three SpyCas9 structures colored by CF (red/blue for high/low CF) with hotspot residues shown as beads colored by domain (with the same color coding as Fig. 1). Some clusters of hotspot residues are circled.

the Hel-III, and the PI domain, owning to closure of the central channel and binding of sgRNA and dsDNA to these domains.[25–27]

We then related the findings of our flexibility analysis to the following functional motifs/sites of SpyCas9 [see Fig. 4(a)]:

a. The two nuclease cleavage sites (D10 of the RuvC domain and H840 of the HNH domain) are both near CF minima with low flexibility, which may confer local stability and precise coordination needed for a catalytic reaction as observed in other enzymes.[61]
b. The L1 and L2 linkers (two key hinge regions of the HNH domain[29] near its N and C terminus, see Fig 1) are both near CF minima, with L1 having greater CF than L2. This is consistent with the observation that L2 is ordered (with low flexibility) and L1 is disordered (with high flexibility) in the unbound structure and the binary complex. L1 and L2 are in contact with dsDNA in the tertiary complex,[27] hinting for a possible coupling of the HNH domain motion [see Fig. 1(b,c)] with binding and melting of dsDNA.
c. Six clusters of sgRNA-binding residues [see Fig. 2(f) of Ref. 26] are near CF minima [residues R71, F164, N407, R447, N501, and T740 in the binary complex; residues R71, F164, I410, I448, A488, and T740 in the tertiary complex, see Fig. 4(a)], which may confer local stability needed for binding of sgRNA.
d. Most DNA-binding residues in the tertiary complex are near CF minima [e.g., residues T270, I448, R655, A764, T924, and A984, see Fig. 4(a)]. However, some DNA-binding residues in the PI domain are near local CF maxima, including the PAM recognition motif 1332DRKRY1336, and the phosphate lock loop (K1107-S1109), supporting a key role of flexibility in PAM recognition (via residues R1333 and R1335[28])

Table I
Hotspot Residues in Three SpyCas9 Structures

PDB id (chain ids)	Hotspot residues
4CMP (B)	R70* Y72 T73 R74* K76 N77 C80 D95 E102 E103 H116 P117* P133 T134* E171* G172 N202 S204 S217* G231 K268 Y325 I363* S368 Q369* E370 N394 E396 K401* Q402* R403* T404* F446* R447* N504 E505 K673 G676 K710 A711 Q712* V713 D718 S719 G773* K775 Q894 T941 K942 Y943 N946 D947 K948 L949 I950 G1011* D1012 S1172 E1175* K1176 N1177 L1194 H1241 Q1256
4ZT0 (AB)	Q194 D257* L258 A259 E260* P344 E345 K348 F352 D353 D364* G365* G366* S368* K536 P537 S581 G582 E584 D585 D700 V713 Q768 T769 T770* K772 G773* K775 R778* E779* K782* N888* A889 K890 Q894 Y943* D944* E945* N946* D947* K948* L949* I950* S1109 P1137 K1148 V1149 E1150 K1151 G1152* K1155 K1156 L1157 K1158 S1159 V1160 K1161 E1162 M1169 E1170 S1172 S1173 K1176 K1185 G1186 Y1187 K1188* E1189 D1193 E1243 K1244 K1246 G1366
5F9R (ABCD)	A214 R215 L216 A259 E260* D261 R340 P344 E345 K348 F352 D353 Q354 S355 D364* G365* G366* N394 R395 E396 N497 R557* K558 S581 G582 V583 E584 D585 R586 N588 H698* Q709* V713 T770* Q771 K772 Q774* K775 R778* E779* K782* G906 S909* E910* D947* K948* V1015 I1029* G1030* R1114* N1115* S1116* D1117* K1118* V1149 E1150 K1151 G1152* K1153* K1155 K1156 L1157 K1158 S1159 V1160 K1161 E1162 S1172 K1185 G1186 K1188* D1193 T1337

Note: The conserved residues (identified by the ConSurf server[67] with a grade $\geq$8) are marked with star.

and DNA strands separation (via the phosphate lock loop[28]).

In sum, our CF analysis revealed a number of flexible domains (e.g., the HNH domain, the PI domain, etc) consistent with previous structural and functional studies, along with several low/high-flexibility regions involved in DNA cleavage and binding of sgRNA/dsDNA. The flexibility of these key domains/regions is fine-tuned according to their functional roles.

Next, we used an ENM-based perturbation analysis (see Methods) to identify a small set of hotspot residues that control the collective motions described by the lowest 20 modes (see Fig. 4 and Table I). To this end, we calculated for each residue position an average score $\langle \delta\lambda/\lambda \rangle$ that assesses the overall importance of this residue position to the energetics of the lowest 20 modes (see Methods). We selected top 5% hotspot residue positions as ranked by $\langle \delta\lambda/\lambda \rangle$.

In the unbound structure, hotspot residues are clustered in the bridge helix and nearby domains [such as Hel-I and Hel-III, as circled in Fig. 4(b)], at the interface between the Hel-I domain and the Hel-II domain [as circled in Fig. 4(b)], at the interfaces between the HNH domain and neighboring domains (including PI, Hel-II, and the L1 linker), and at the interface between the PI domain and the Hel-III domain [as circled in Fig. 4(b)].

In the binary complex, hotspot residues are distributed at the interfaces between the HNH domain and nearby domains (including Hel-II, Hel-III, PI, and the L1 linker), within the PI domain [as circled in Fig. 4(c)], and at the interface between the PI domain and the Hel-I domain [as circled in Fig. 4(c)].

In the tertiary complex, hotspot residues are located at the interfaces between the HNH domain and adjacent domains (e.g., Hel-II, Hel-III, L1, and L2), between neighboring alpha-helical domains [as circled in Fig. 4(d)], within the PI domain [as circled in Fig. 4(d)], and at the interface between the PI domain and the Hel-I domain [as circled in Fig. 4(d)].

In sum, the predicted hotspot residues overlap well with known functional motifs/domains involved in sgRNA/dsDNA binding (including the bridge helix,[30] the Hel-I domain,[24] the PAM recognition motif,[28] and the phosphate lock loop[28]), and other key hinge regions that control large domain motions in SpyCas9 (such as the bridge helix,[25] the L1, and the L2 linker[27,29]). These agreements support the functional importance of the above hotspot residues. These hotspot residues may allosterically couple the binding of sgRNA and dsDNA to large conformational changes required for R-loop formation and DNA cleavage. Therefore, by perturbing these key residues with point mutations, one can fine-tune the energetics of these conformational changes and improve the DNA targeting specificity of the Cas9 systems. An increase in the energy cost is expected to lower the probability of off-target cleavage of DNA sequences which are imperfectly based-paired with the guide RNA.

Transition pathway modeling predicts a sequence of domain motions in SpyCas9 during its transitions to the binary and the tertiary complex

Finally, we have modeled two structural transitions from the unbound SpyCas9 in the inactive state[25] to a binary SpyCas9-sgRNA complex in the pre-target-bound state,[26] and then to a tertiary SpyCas9-sgRNA-dsDNA complex in the cleavage-competent state.[27] To fully explore these transitions, we must go beyond NMA which is limited to structural fluctuations in the vicinity of the known structures, and sample large conformational changes far away from the crystal structures of SpyCas9.[25–28] To this end, we employed a transition pathway modeling method based on the interpolated ENM (iENM[57], see Methods) to predict a pathway of intermediate conformations that connect the unbound SpyCas9 structure to the binary complex, and then to the tertiary complex. This method was previously applied to various large protein complexes to delineate a sequence of

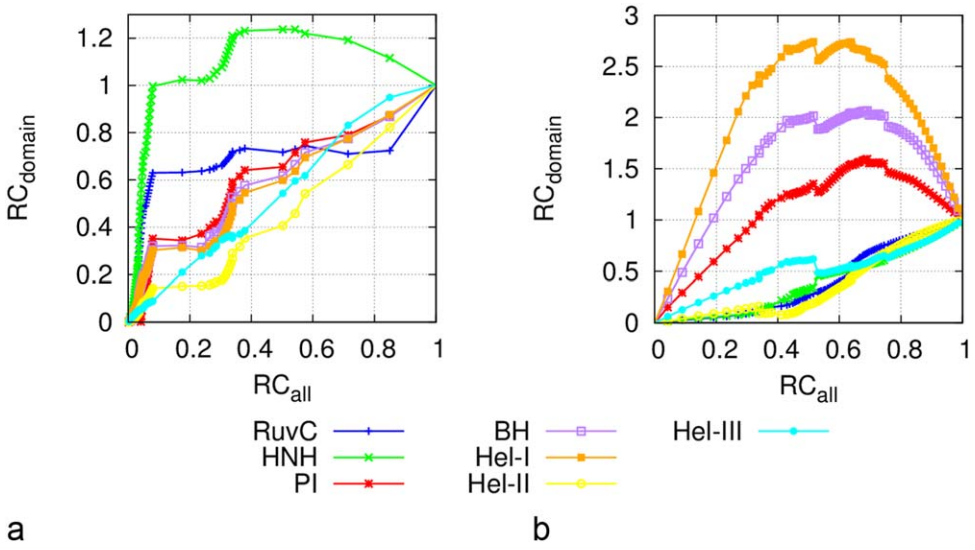

Figure 5

Results of transition pathway modeling for (**a**) the unbound-to-binary transition and (**b**) the binary-to-tertiary transition. The reaction coordinates (RC) of various domains are plotted as a function of the overall RC for the entire SpyCas9. The lines and points are colored by domain using the same color coding as Figure 1. See Supporting Information Movies S1 and S2 for structural details of the predicted pathways. [Color figure can be viewed at wileyonlinelibrary.com]

structural motions involving different functional domains during a conformational transition.[32,50,56,59,62–66] By performing a reaction-coordinate (RC) analysis of the predicted transition pathway (see Methods), we can determine the motional order of various domains (with early/late moving domains having high/low RC values, see Fig. 5).

First, we constructed a transition pathway starting from the sgRNA-bound structure of SpyCas9 (PDB id: 4ZT0, chains A and B) toward the unbound structure (PDB id: 4CMP, chain B). Then we reversed this pathway to obtain a series of conformational changes caused by sgRNA loading, leading to the central channel formation (see Supporting Information Movie S1). By comparing the RCs of various SpyCas9 domains at the middle of the transition pathway [near RC_{all} =0.5, see Fig. 5(a)], we found $RC_{HNH} > RC_{RuvC} > RC_{PI} \sim RC_{BH} \sim RC_{Hel\text{-}I} > RC_{Hel\text{-}III} > RC_{Hel\text{-}II}$, which implies the following motional order: HNH → RuvC → PI, BH, and Hel-I → Hel-III → Hel-II. This order is consistent with the finding that the lowest 20 modes of the unbound structure describe the motions in the NUC lobe better than the REC lobe [see Fig. 2(b)]. Interestingly, the motions of BH, Hel-I, and PI follow very similar RC curves [see Fig. 5(a)], suggesting that they undergo coupled motions upon sgRNA binding. The late motions of Hel-II and Hel-III are likely driven by the earlier motions of BH and Hel-I which directly interact with the sgRNA [see Fig. 1(b)]. Our findings highlight the allosteric nature of the sgRNA-loading-induced conformational changes that propagate from sgRNA-interacting domains (such as BH and Hel-I) to distant domains (such as Hel-III and Hel-II) during the transition.

Next, we built a transition pathway beginning from the sgRNA-dsDNA-bound structure of SpyCas9 (PDB id: 5F9R, chains A to D) toward the binary complex (PDB id: 4ZT0, chains A and B), and then reversed this pathway to obtain a sequence of domain motions induced by dsDNA binding that lead to the dsDNA strands separation and the R-loop formation (see Supporting Information Movie S2). By comparing the RCs of the SpyCas9 domains at the middle of the transition pathway, we found $RC_{Hel\text{-}I} > RC_{BH} > RC_{PI} > RC_{Hel\text{-}III} > RC_{HNH} \sim RC_{RuvC} \sim RC_{Hel\text{-}II}$, which implies the following motional order: Hel-I → BH → PI → Hel-III → HNH, RuvC, and Hel-II [see Fig. 5(b)]. The above order is consistent with the finding that the lowest 20 modes of the binary complex did not well describe the motions in Hel-II and HNH [see Fig. 2(c)]. The early motions of Hel-I, BH, and PI allow further closing of the central channel for dsDNA binding and PAM recognition (see Supporting Information Movie S2), which is consistent with the domain motions predicted by mode 1 [see Fig. 3(c)]. Subsequently, the central channel re-opens, and the two DNA strands separate with the target strand forming a hybrid duplex with the seed region of sgRNA and the non-target strand moving toward the RuvC domain (see Supporting Information Movie S2). Remarkably, the HNH domain undergoes a large rotation to move its catalytic residue H840 toward the cleavage site (at the nucleotide G13) on the target strand by ~30 Å (see

Supporting Information Movie S2), while the catalytic residue D10 of the RuvC domain moves by ~20 Å to the cleavage site (at the nucleotide A17) on the non-target strand. Although the HNH and the RuvC domain move at similar pace [see Fig. 5(b)], after the transition the H840-G13 distance is ~10 Å greater than the D10-A17 distance, which explains why the cleavage reactions occur earlier in the RuvC nuclease than the HNH nuclease.[26] The above motional order is consistent with the functional implication that dsDNA binding and PAM recognition are prerequisite to enable subsequent DNA strands separation, R-loop formation (in the seed region), and DNA cleavage.[8] Our predicted pathway is consistent with the following model of Cas9 proposed in Ref. 26: the pre-ordered PAM recognition site initiates DNA interrogation, followed by base pairing between a PAM-proximal segment of the target DNA strand and the pre-ordered seed region in the sgRNA accompanied by dsDNA melting, and then the hybrid duplex between the guide RNA and the target DNA strand extends further, triggering additional displacements of the RuvC and the HNH domain toward their cleavage-competent positions.

In sum, our transition pathway modeling has predicted distinct sequences of domain motions during the two key conformational transitions in SpyCas9, offering new structural insights to these transitions which are not available from structural comparison or linear morphing between known structures. A prominent feature of the binary-to-tertiary transition is a transient closing followed by reopening between the two lobes, suggesting that Cas9 may adopt an even more closed conformation than observed in the tertiary complex.

CONCLUSIONS

In summary, we have performed a comprehensive coarse-grained modeling of the Cas9 dynamics using an array of ENM-based analyses. Our findings are in broad agreement with previous structural and functional findings of Cas9: a. Our NMA predicted 2–3 key modes of domain motions that capture >50% of the crystallographically observed structural changes from the unbound SpyCas9 to the binary complex and then the tertiary complex, which are critical to formation of the central channel, recognition of the PAM, and the DNA strands separation. b. Our flexibility analysis identified specific regions with high/low flexibility that coincide with known functional sites (including the sgRNA/dsDNA-binding sites, the nuclease cleavage sites, the PAM recognition motif, the phosphate lock loop, and the L1 and the L2 linker of the HNH domain).

Our modeling has made the following new predictions for future experiments to test: a. We predicted a small set of dynamically important hotspot residues at key domain interfaces and functional sites, which offer promising targets for engineering Cas9 mutants with improved specificity. b. By modeling the two conformational transitions of SpyCas9 from the unbound form to the binary complex and then the tertiary complex, we predicted a sequence of domain motions and intermediate conformations (e.g., a more closed conformation of the two lobes than observed in the tertiary complex), which can be tested by FRET-based dynamic distance measurement or cross-linking experiment.

Taken together, our findings have offered new structural and dynamic insights to the Cas9-mediated DNA targeting and cleavage mechanism (i.e., how sgRNA and dsDNA binding allosterically trigger subsequent structural changes in proximal and distal domains). In future studies, it will be interesting to perform direct MD simulation of the structural dynamics of Cas9 at atomic resolutions, which will require overcoming the time-scale limit of such simulations via software and hardware innovations.

REFERENCES

1. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature 2012;482:331–338.
2. Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem 2013; 82:237–266.
3. Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature 2015; 526:55–61.
4. Jiang F, Doudna JA. The structural biology of CRISPR-Cas systems. Curr Opin Struct Biol 2015;30:100–111.
5. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol 2014;12:479–492.
6. Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, Siksnys V, Seidel R. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci USA 2014;111:9798–9803.
7. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012;337:816–821.
8. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 2014;507:62–67.
9. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339:819–823.
10. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 2013;31:233–239.
11. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA 2012;109:E2579–E2586.
12. Makarova KS, Aravind L, Wolf YI, Koonin EV. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol Direct 2011;6:38.
13. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. Science 2013;339:823–826.
14. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for

sequence-specific control of gene expression. Cell 2013;152:1173–1183.
15. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 2013;154:442–451.
16. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 2013;31:833–838.
17. Wright AV, Nunez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell 2016;164:29–44.
18. Sternberg SH, Doudna JA. Expanding the Biologist's Toolkit with CRISPR-Cas9. Mol Cell 2015;58:568–574.
19. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 2014;157:1262–1278.
20. Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: progress, implications and challenges. Human Mol Genet 2014;23:R40–R46.
21. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 2013;8:2281–2308.
22. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013;31:827–832.
23. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science 2016;351:84–88.
24. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 2014;156:935–949.
25. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 2014;343:1247997.
26. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science 2015;348:1477–1481.
27. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science 2016;351:867–871.
28. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 2014;513:569–573.
29. Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O. Crystal structure of Staphylococcus aureus Cas9. Cell 2015;162:1113–1126.
30. Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 2013;497:254–257.
31. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–652.
32. Zheng W, Tekpinar M. Structure-based simulations of the translocation mechanism of the hepatitis C virus NS3 helicase along single-stranded nucleic acid. Biophys J 2012;103:1343–1353.
33. Sanbonmatsu KY. Computational studies of molecular machines: the ribosome. Curr Opin Struct Biol 2012;22:168–174.
34. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol 2009;19:120–127.
35. Farasat I, Salis HM. A Biophysical Model of CRISPR/Cas9 Activity for Rational Design of Genome Editing and Gene Regulation. PLoS Comput Biol 2016;12:e1004724.
36. Tozzini V. Coarse-grained models for proteins. Curr Opin Struct Biol 2005;15:144–150.
37. Tozzini V. Minimalist models for proteins: a comparative analysis. Q Rev Biophys 2010;43:333–371.
38. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
39. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng 2001;14:1–6.
40. Zheng W, Doniach S. A comparative study of motor-protein motions by using a simple elastic-network model. Proc Natl Acad Sci USA 2003;100:13253–13258.
41. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. Proteins 2002;48:682–695.
42. Zheng W, Liao JC, Brooks BR, Doniach S. Toward the mechanism of dynamical couplings and translocation in hepatitis C virus NS3 helicase using elastic network model. Proteins 2007;67:886–896.
43. Wang Y, Rader AJ, Bahar I, Jernigan RL. Global ribosome motions revealed with elastic network model. J Struct Biol 2004;147:302–314.
44. Zimmermann MT, Jia K, Jernigan RL. Ribosome Mechanics Informs about Mechanism. J Mol Biol 2016;428:802–810.
45. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.
46. Tama F, Brooks CL. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu Rev Biophys Biomol Struct 2006;35:115–133.
47. Hinsen K, Petrescu A-J, Dellerue S, Bellissent-Funel M-C, Kneller GR. Harmonicity in slow protein dynamics. Chem Phys 2000;261:25–37.
48. Lu M, Poon B, Ma J. A New Method for Coarse-Grained Elastic Normal-Mode Analysis. J Chem Theory Comput 2006;2:464–471.
49. Yang L, Song G, Jernigan RL. Protein elastic network models and the ranges of cooperativity. Proc Natl Acad Sci USA 2009;106:12347–12352.
50. Zheng W, Auerbach A. Decrypting the sequence of structural events during the gating transition of pentameric ligand-gated ion channels based on an interpolated elastic network model. PLoS Comput Biol 2011;7:e1001046.
51. Zheng W, Brooks BR, Doniach S, Thirumalai D. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure 2005;13:565–577.
52. Zheng W, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. Proc Natl Acad Sci USA 2006;103:7664–7669.
53. Zheng W, Tekpinar M. Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model. BMC Struct Biol 2009;9:45.
54. Zheng W. Probing the structural dynamics of the SNARE recycling machine based on coarse-grained modeling. Proteins 2016;84:1055–1066
55. Zheng W. Coarse-grained modeling of the structural states and transition underlying the powerstroke of dynein motor domain. J Chem Phys 2012;136:155103.
56. Zheng W, Qin F. A combined coarse-grained and all-atom simulation of TRPV1 channel gating and heat activation. J Gen Physiol 2015;145:443–456.
57. Tekpinar M, Zheng W. Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. Proteins 2010;78:2469–2481.
58. Zheng G, Lu XJ, Olson WK. Web 3DNA–a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. Nucleic Acids Res 2009;37(Web Server issue):W240–W246.
59. Zheng W. Multiscale modeling of structural dynamics underlying force generation and product release in actomyosin complex. Proteins 2010;78:638–660.
60. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 2015;527:110–113.

61. Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 2005;13:893–904.
62. Zheng W, Brooks BR, Hummer G. Protein conformational transitions explored by mixed elastic network models. Proteins 2007;69:43–57.
63. Zheng W, Tekpinar M. Analysis of protein conformational transitions using elastic network model. Methods Mol Biol 2014;1084:159–172.
64. Zheng W. Coarse-grained modeling of the structural states and transition underlying the powerstroke of dynein motor domain. J Chem Phys 2012;136:155103.
65. Zheng W. Coarse-grained modeling of conformational transitions underlying the processive stepping of myosin V dimer along filamentous actin. Proteins 2011;79:2291–2305.
66. Tekpinar M, Zheng W. Coarse-grained and all-atom modeling of structural states and transitions in hemoglobin. Proteins 2013;81: 240–252.
67. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 2016;44: W344–W350.