

Twisting mode of supercoil leucine-rich domain mediates peptide sensing in FLS2-flg22-BAK1 complex

Zhi-Chao Liu(刘志超), Qin Liu(刘琴), Chan-You Chen(陈禅友), Chen Zeng(曾辰), Peng Ran(冉鹏), Yun-Jie Zhao(赵蕴杰), Lei Pan(潘磊)

Citation: Chin. Phys. B . 2020, 29(10): 108709 . **doi:** 10.1088/1674-1056/abaee1
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

A computational study of the chemokine receptor CXCR1 bound with interleukin-8

Yang Wang(王洋), Cecylia Severin Lupala, Ting Wang(王亭), Xuanxuan Li(李选选), Ji-Hye Yun, Jae-hyun Park, Zeyu Jin(金泽宇), Weontae Lee, Leihan Tan(汤雷翰), Haiguang Liu(刘海广)

Chin. Phys. B . 2018, 27(3): 038702 . **doi:** 10.1088/1674-1056/27/3/038702

A novel knowledge-based potential for RNA 3D structure evaluation

Yi Yang(杨毅), Qi Gu(辜琦), Ben-Gong Zhang(张本龚), Ya-Zhou Shi(时亚洲), Zhi-Gang Shao(邵志刚)

Chin. Phys. B . 2018, 27(3): 038701 . **doi:** 10.1088/1674-1056/27/3/038701

Theoretical studies on sRNA-mediated regulation in bacteria

Chang Xiao-Xue, Xu Liu-Fang, Shi Hua-Lin

Chin. Phys. B . 2015, 24(12): 128703 . **doi:** 10.1088/1674-1056/24/12/128703

RNA structure prediction：Progress and perspective

Shi Ya-Zhou, Wu Yuan-Yan, Wang Feng-Hua, Tan Zhi-Jie

Chin. Phys. B . 2014, 23(7): 078701 . **doi:** 10.1088/1674-1056/23/7/078701

SPECIAL TOPIC — Modeling and simulations for the structures and functions of proteins and nucleic acids

Twisting mode of supercoil leucine-rich domain mediates peptide sensing in FLS2–flg22–BAK1 complex*

Zhi-Chao Liu(刘志超)[1,3], Qin Liu(刘琴)[4], Chan-You Chen(陈禅友)[4], Chen Zeng(曾辰)[3], Peng Ran(冉鹏)[1,3], Yun-Jie Zhao(赵蕴杰)[2,†], and Lei Pan(潘磊)[4,‡]

[1] *School of Biological Information, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*
[2] *Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan* 430079, *China*
[3] *Department of Physics, The George Washington University, Washington, DC*, 20052, *USA*
[4] *School of Life Sciences, Jianghan University, Wuhan* 430056, *China*

(Received 23 June 2020; revised manuscript received 7 August 2020; accepted manuscript online 13 August 2020)

Plants and animals recognize microbial invaders by detecting pathogen-associated molecular patterns (PAMPs) through pattern-recognition receptors (PRRs). This recognition plays a crucial role in plant immunity. The newly discovered protein in plants that responds to bacterial flagellin, i.e., flagellin-sensitive 2 (FLS2), is ubiquitously expressed and present in many plants. The association of FLS2 and BAK1, facilitated by a highly conserved epitope flg22 of flagellin, triggers such downstream immune responses as activated MAPK pathway and elevated reactive oxygen species (ROS) for bacterial defense and plant immunity. Here we study the intrinsic dynamics and conformational change of FLS2 upon the formation of the FLS2–flg22–BAK1 complex. The top intrinsic normal modes and principal structural fluctuation components are very similar, showing two bending modes and one twisting mode. The twisting mode alone, however, accounts for most of the conformational change of FLS2 induced by binding with flg22 and BAK1. This study indicates that flg22 binding suppresses FLS2 conformational fluctuation, especially on the twisting motion, thus facilitating FLS2–BAK1 interaction. A detailed analysis of this sensing mechanism may aid better design on both PRR and peptide mimetics for plant immunity.

Keywords: plant immunity, flagellin-sensitive 2, peptide sensing mechanism

PACS: 87.14.gn, 87.15.K–, 87.10.Ca, 87.15.A– **DOI:** 10.1088/1674-1056/abaee1

1. Introduction

In nature, pathogens (including bacteria, viruses, fungal, etc.) cause disease in animals and plants. Within them, the pathogen-associated molecular patterns (PAMPs) or the microbe-associated molecular patterns (MAMPs) are usually conserved and often relatively abundant molecular signatures that are present across a broad range of microbes.[1,2] Upon the interaction with the host cell surface-localized pattern recognition receptors (PRRs), PAMPs trigger defense responses of plant or animal innate immune systems.[2–6] These responses include the production of ethylene, ion fluxes, reactive oxygen species (ROS) burst, and MAP kinase activation.[7,8] They are usually shared by different PAMPs, while the amplitude and duration may differ depending on the host and the specific PAMP.[1,5,9] Many different PRRs for conserved microbial patterns have been identified, most of which are receptor kinases (RKs) or receptor-like proteins in plants.[4,10] In the ectodomains of these PRRs, many of them carry leucine-rich repeats (LRRs). With a spiral-shaped ligand interaction domain, LRRs facilitate PRR recognition of many types of molecules such as peptides, lipids, nucleic acids, and small-molecule hormones.[1,11,12]

One of the most well-known LRR-RKs, flagellin-sensitive 2 (FLS2), is critical for plant antibacterial immunity. It recognizes the conserved epitope flg22 of bacterial flagellin and initiates downstream plant immunity.[4,10,13,14] Flg22 recognition specificities exist widely in higher plants, and it has been shown that the flg22-induced immune response is a direct result of the recognition of flg22 by FLS2.[13] Upon the binding of flg22, FLS2 forms a heterodimer with its coreceptor BRI1-associated kinase 1 (BAK1)[10,15–18] that triggers downstream immune responses and internalizes FLS2 from the plasma membrane (PM) into endosomes.[19] However, the mechanism of flg22-induced FLS2–BAK1 heterodimer formation is still puzzling because the experimentally observed static structures show little conformational change for FLS2 with or without the binding of flg22 and BAK1.[10] How is the recognition specificity of flg22 achieved then? The answer may lie in the detailed dynamics of the complex.

Protein interaction and dynamics have been studied for decades under several different frameworks.[20–29] The first

*Project supported by the National Natural Science Foundation of China (Grant No. 11704140), self-determined research funds of CCNU from the Colleges' Basic Research and Operation of MOE (Grant No. CCNU20TS004) (Y. Z.), and the China Scholarship Council Fund (Grant No. 201708420039) (L. P.).

†Corresponding author. E-mail: yjzhaowh@mail.ccnu.edu.cn
‡Corresponding author. E-mail: leipan@jhun.edu.cn

© 2020 Chinese Physical Society and IOP Publishing Ltd

one is the "lock-and-key" model proposed by E. Fischer.[20] This model emphasizes the complementary shapes between the two interacting proteins. This rigid framework, however, ignores the conformational changes associated with the protein–protein interaction. To account for this type of protein plasticity, the "induced fit" model[21] has been proposed to incorporate some structural change of the receptor induced by substrate binding. For some protein complexes that undergo significant structural changes upon the association, a more recent model of "conformational selection" has emerged using the free energy landscape concept of protein folding theories.[30] According to this model, the native state of a protein is not a single conformation, but rather an ensemble of closely related conformations that coexist in equilibrium. The most suitable conformation will bind the substrate and reach a new equilibrium for the complex.[22] Recent protein dynamics studies by Bahar *et al*. have shown that the protein structural changes during complex formation correlate with its intrinsic dynamics and fluctuations near its equilibrium state. Again, this suggests that the conformational fluctuations and intrinsic dynamics of a protein complex can be essential for its biological functions.[22–24]

Repeat proteins made of tandem copies of similar structural elements are widespread in viruses, bacteria, and eukaryotes.[31] They play various roles in molecular recognition signal transduction, cell adhesion, RNA processing, plant immune response, and other biological processes. Specifically, LRR proteins usually consist of 2 to 45 tandem repeats, each of which contains 20–30 amino acids with a high leucine content,[32] and a short helix. For example, FLS2 has 28 small helix units. Interestingly, these short helical LRRs also form a helical structure overall, the so-called superhelix. The interactions between adjacent repeats dictate the precise shape and curvature of the superhelical structure, a fact that has been exploited in designing different families of repeat proteins of identical repeats. For example, in the ribonuclease inhibitor (RI) family, the curvature, length, and helical twist can be adjusted for specific functions.[33–35] Rämisch *et al.* have shown that the global dynamics can be modulated by single amino acid alteration; for example, a buried cysteine controls the stability and folding cooperativity in RI-type LRR proteins.[34] Similar to the principal component analysis by Emberly *et al*.[36] on helical fragments extracted from all proteins in the PDB database that shed light on their elasticity and functionality, a detailed analysis of the dynamics of superhelical LRR proteins may also help decipher some of their functions.

Here, we study the flg22 recognition mechanism by probing the dynamics of the FLS2–flg22–BAK1 complex. The intrinsic dynamics of FLS2 is characterized using normal mode analysis (NMA) and principal component analysis (PCA) of molecular dynamics (MD) simulations on its crystal structure.[10] Both analyses clearly show that the top three dynamical modes for the superhelical structure of FLS2 consist of two bending modes and one twisting mode similar to a standard helix. However, only the projection on the twisting mode can successfully distinguish three FLS2 complexes of unbound, bound with flg22, and bound with both flg22 and BAK1, respectively. This suggests that flg22 suppresses the conformational fluctuation of FLS2, especially on the twisting mode to facilitate FLS2–BAK1 heterodimer formation. A detailed analysis of this sensing mechanism may aid better design on both LRR proteins and peptide mimetics for plant immunity.

2. Methods

2.1. Molecular dynamics simulation

The MD simulations were carried out using the GROMACS software package[37] and the AMBER03 force field.[38] A water solvent box of 12 Å was created between the outside of the protein and the edge of the box. All the structures were simulated at the room temperature (300 K). The initial states (FLS2, FLS2–BAK, and FLS2–BAK1–flg22 complex structures) were extracted from the PDB database (PDB code: 4MN8)[10] and solvated with water molecules in a periodic rectangular box with a normal saline condition. The non-bonded (electrostatic and VDW) cutoff range was 8 Å. A time step of 2 fs was used for numerical integration. Before the MD simulation, the entire system was first minimized by the steepest descent calculation for 1000 steps followed by 300 ps equilibration. After that, 200 ns molecular dynamics simulation was conducted for each of the states. The small root mean standard deviation (RMSD, provided in SI) of the MD topology indicates little configuration destabilization. Besides, previous research showed that a 20 ns molecular dynamics simulation could identify the correlation and reveal the motions of the protein.[39] Due to the limitation of computational time and resources, we performed all the downstream analyses with a configuration of 30 ns for each trajectory. We saved the structural snapshots for every 15 ps during the 30 ns trajectory. The structures were visualized and analyzed by VMD and PyMOL.[27,40]

2.2. Interface correlation analysis

The interface residues of FLS2–BAK1 are chosen based on the following criteria. If the distance of any heavy atom of a residue in FLS2 to any heavy atom of any residue in BAK1 is less than 4.0 Å, the two residues are defined as in contact[27] and belong to the interface residue set S_{FLS2} and S_{BAK1}, respectively. Given the dynamics trajectory of the molecular structure, the residue–residue Pearson correlation C_{mn} can be

calculated by

$$C_{mn} = \frac{\langle \Delta \boldsymbol{r}_m(t) \cdot \Delta \boldsymbol{r}_n(t) \rangle}{\langle \Delta \boldsymbol{r}_m(t)^2 \rangle \langle \Delta \boldsymbol{r}_n(t)^2 \rangle}, \quad m \in S_{\mathrm{FLS2}}, \quad n \in S_{\mathrm{BAK1}},$$

where $\Delta \boldsymbol{r}_m(t) = \boldsymbol{r}_m(t) - \langle \boldsymbol{r}_m(t) \rangle$ and the position $\boldsymbol{r}_m(t)$ of the residue m is represented by its C_α atom. The brackets stand for average over time.

2.3. Quantitative analysis of superhelix structure and dynamics

The superhelical structure of FLS2 can be simplified as a helix with each LRR unit viewed as one node of the helix. To quantify changes in structure and dynamics, we approximated the superhelix as a helix composed of the innermost concave surface residue of each LRR. These representative residues (one from each of 28 LRRs) together with one more residue before the first LRR roughly delineate the superhelical structure of FLS2 (Table S3). We will use FLS2LRR to denote this simplified helix thereafter. Radius and pitch length of the simplified helix were calculated by fitting the points to a standard helix (which can be parameterized by helix axis, radius, pitch, and the number of points per turn) with a total least squares method. The fitting analysis was implemented with HELFIT.[41] The calculation was performed on the structures of the MD trajectory under various conditions.

2.4. Normal mode analysis of protein structure

NMA is a classical approach to retrieve the dominant modes of motion based on native contact topology only.[23,42] Anisotropic network model (ANM) is one of the most common models of NMA. The Hessian matrix $\boldsymbol{H}$ forms the basis of the ANM approach. It is a $3N \times 3N$ matrix and can be seen as $N \times N$ submatrices of size 3×3. The ij-th submatrix $H^{(ij)}$ is given by

$$H^{(ij)} = \frac{\Gamma_{ij}}{\left(R_{ij}^0\right)^2} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix}$$

for $i \neq j$. For diagonal elements, $H^{(ii)} = -\sum_{j \neq i} H^{(ij)}$. R_{ij}^0 is the equilibrium distance (from PDB structure) between residues i and j and X_{ij}, Y_{ij}, and Z_{ij} are its components in 3D space. Γ_{ij} equals to 1 if residues i and j are in contact (defined by a cutoff distance of r_{cut}), or 0 otherwise. We implemented the NMA with an online server ANM.[43] We have tried several cutoffs for the ANM calculation. The results of different cutoffs around 15 Å (the default cutoff) show the similar top normal modes of FLS2. Therefore, we used 15 Å as the cutoff in the calculations. The 15 Å cutoff is a compromise between better agreement with isotropic B-factors obtained in larger cutoffs and more realistic anisotropic displacement parameters (ADPs) obtained using lower cutoffs.[43]

2.5. Principal component analysis of the MD trajectory

Principal modes of motion for the MD trajectory were obtained by decomposing the covariance matrix $\boldsymbol{C}$. For a protein of N residues in 3D space, there are $3N$ dimensions in total to describe the protein structure. The covariance matrix $\boldsymbol{C}$ is a $3N \times 3N$ matrix, each element being the covariance of two of the $3N$ dimensions for an MD trajectory. Similar to above, the position of each residue is represented by its C_α atom. According to linear algebra,

$$\boldsymbol{C} = \sum_{i=1}^{3N} \sigma_i \boldsymbol{P}_i \boldsymbol{P}_i^{\mathrm{T}},$$

where σ_i and $\boldsymbol{P}_i$ are the i-th eigenvalue and eigenvector of $\boldsymbol{C}$. The fractional contribution of $\boldsymbol{P}_i$ to structural variance in the dataset is given by $f_i = \sigma_i / \sum_k \sigma_k$, where the summation is performed over all $3N$ components. Usually the eigenvectors are ranked according to their relative contribution to the total variance, and the top 2 or 3 are adopted to approximate the whole $3N$ space. PCA was implemented with Python package *sklearn* and customized code.

2.6. Sequence evolution analysis

The sequence evolution analysis measures residue conservation by the ConSurf program.[44] First, we extracted the sequence information from the protein (PDB code: 4MN8 for chain A and chain C). Second, we performed homology search using HMMER algorithm (number of iterations = 1, E-value cutoff = 0.0001) from UNIREF-90 protein database. Third, we performed multiple sequence alignment by MAFFT. Then, the conservation scores were calculated by ConSurf with default parameters. The continuous conservation scores were divided into a discrete scale of 9 grades with grades 1–3 for the most variable positions and grades 7–9 for the most conserved ones. The binding sites of FLS2–flg22 on N- and C-terminus were extracted from the known crystal structure (PDB code: 4MN8).

3. Results

3.1. FLS2–BAK1 interface correlation is enhanced with the binding of flg22

The tertiary structure of the FLS2–flg22–BAK1 complex was extracted from the PDB database with a resolution of 3.1 Å (PDB code: 4MN8).[10] To probe the interaction and dynamics responsible for the sensing of flg22 by FLS2–BAK1, we conducted MD simulations based on the known structures for the following four complexes: (1) FLS2 (4MN8 chain A); (2) FLS2–flg22 (4MN8 chain A, C); (3) FLS2–BAK1 (4MN8 chain A, B); and (4) FLS2–flg22–BAK1 (4MN8 chain A, B, C). MD simulations for each structure were performed three times independently, and each MD trajectory contains 2002 frames.

Based on the criterion for being in contact, the interface interacting pairs (36 pairs, Table S1) and residue sets (21 and 20 residues for FLS2 and BAK1, respectively, Table S2) were generated. To evaluate and compare the interface interaction between FLS2 and BAK1 with and without flg22, the residue–residue Pearson correlation was calculated from the MD simulations, and the results were shown in Fig. 1. The left panel shows the FLS2–BAK1 correlation matrix without flg22 and right panel with flg22. There is a clear trend that the interface correlation is enhanced globally with the binding of flg22. This matches with the experimental fact that BAK1 ectodomain interacts with FLS2 only after the FLS2–flg22 binding.

In the network calculation, if two residues move in the same (opposite) direction in most snapshots of the MD simulation trajectory, the residues are defined as correlated (anti-correlated) with positive (negative) correlation values. A correlation value close to zero indicates uncorrelated motion. We divided the FLS2–BAK1 interface into two regions: up (red box in Fig. 1) and down (gray box in Fig. 1) interface areas. The up interface is close to the flg22 binding area. The interactions in the down interface region are weak. There is a significant enhancement of correlations in both up and down interface regions upon flg22 binding. Before the flg22 binding, the correlation of the up interface is not very strong due to the weak interactions between FLS2–BAK1 without flg22. For the down interface, FLS2 and BAK1 are fluctuating in the opposite directions (for the conservation of global momentum), leading to a slight negative correlation. After the flg22 binding, the up interface is enhanced with larger correlations. However, the correlations of the down interface are close to zero because of the reduced dynamics of both FLS2 (also discussed in later sections) and BAK1.

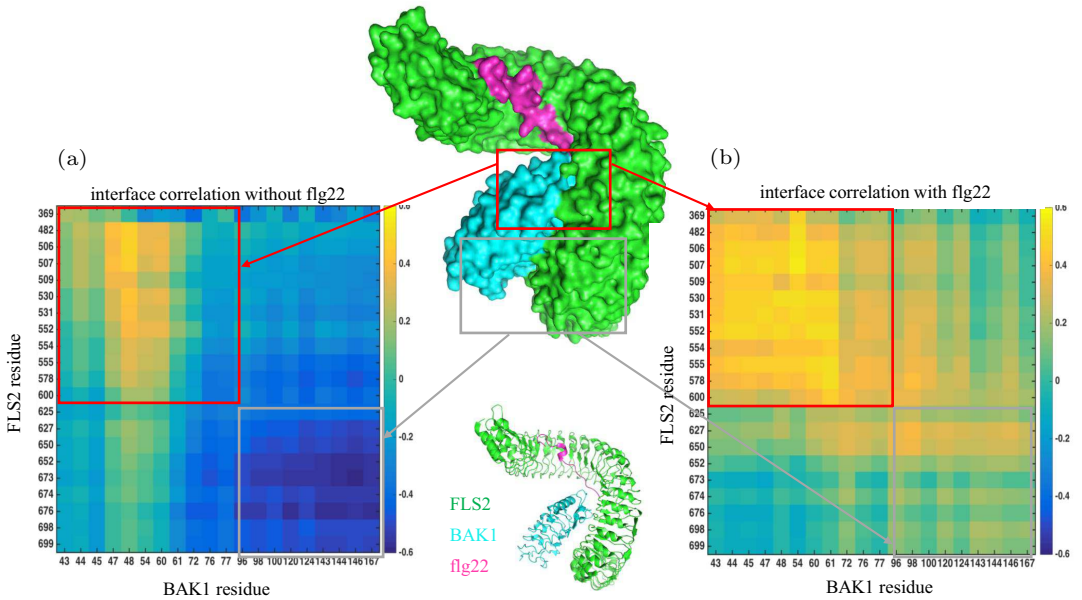


Fig. 1. FLS2–BAK1 interface correlation is enhanced upon the flg22 binding. The horizontal axis indicates the interface residues in BAK1, and the vertical axis shows the interface residues in FLS2. (a) Interface correlation of FLS2–BAK1 without flg22; (b) with flg22.

3.2. The dynamics of FLS2 is suppressed by flg22

To understand why the FLS2–BAK1 association only appears in the presence of flg22, we evaluated the structural and dynamical change of FLS2. FLS2LRR is a superhelical structure and similar to a spiral spring. As described in Section 2, we approximated the FLS2LRR superhelix as a helix by extracting the innermost concave surface residue for each LRR. We chose 29 residues (Table S3) for the fitting. A standard helix can be parameterized by helix axis, radius, pitch length, and the number of points per turn.[41] Since the number of residues per LRR in FLS2 is roughly the same and unchanged, and the direction of the protein structure is arbitrary, we focused on radius and pitch to evaluate the structure of FLS2. The fitting was optimized with a total least squares method and implemented with the open-source software HELFIT.[41]

We calculated the radius and pitch of the simplified FLS2LRR for FLS2, FLS2–flg22, and FLS2–flg22–BAK1 states. We have performed 200 ns simulations for the FLS2 structure. We analyzed both start and end 30 ns episodes. The radius and pitch results indicate a similar trend for the two episodes. More specifically, after the flg22 binding, the pitch length of FLS2 is decreased with less variance (Figs. 2(a) and 2(c)). The FLS2 radius is also decreased (Figs. 2(b) and 2(d)). The dynamics of FLS2–flg22–BAK1 is much more complicated. The PCA and NMA analysis results show a similar conclusion for the two episodes, too. We also analyzed the structural differences during the 200 ns simulations. The small RMSDs (provided in Fig. S1) indicate little configuration destabilization during the simulations. Due to the compu-

tational time and resources limitations, we performed all the downstream analyses with 30 ns for each trajectory.

Upon peptide binding (the start episode), the fluctuation on the radius of FLS2 reduces because of the constraint (flg22) placed on the free supercoil of FLS2 (Fig. 2(b)). However, after the formation of the whole complex, while the interface between FLS2 and BAK1 becomes tighter with stronger interaction (Fig. 1), the impact on FLS2 motion is less clear in the longer simulation. Instead, there is a slight decrease in the radius of the supercoil (Fig. 2(d)) due to the decreased twist mode. The fact that the difference in pitch has already emerged even with only flg22 but no BAK1 implies that flg22 alone is able to suppress the intrinsic motion of FLS2. More details about the intrinsic dynamics of FLS2 are examined below.

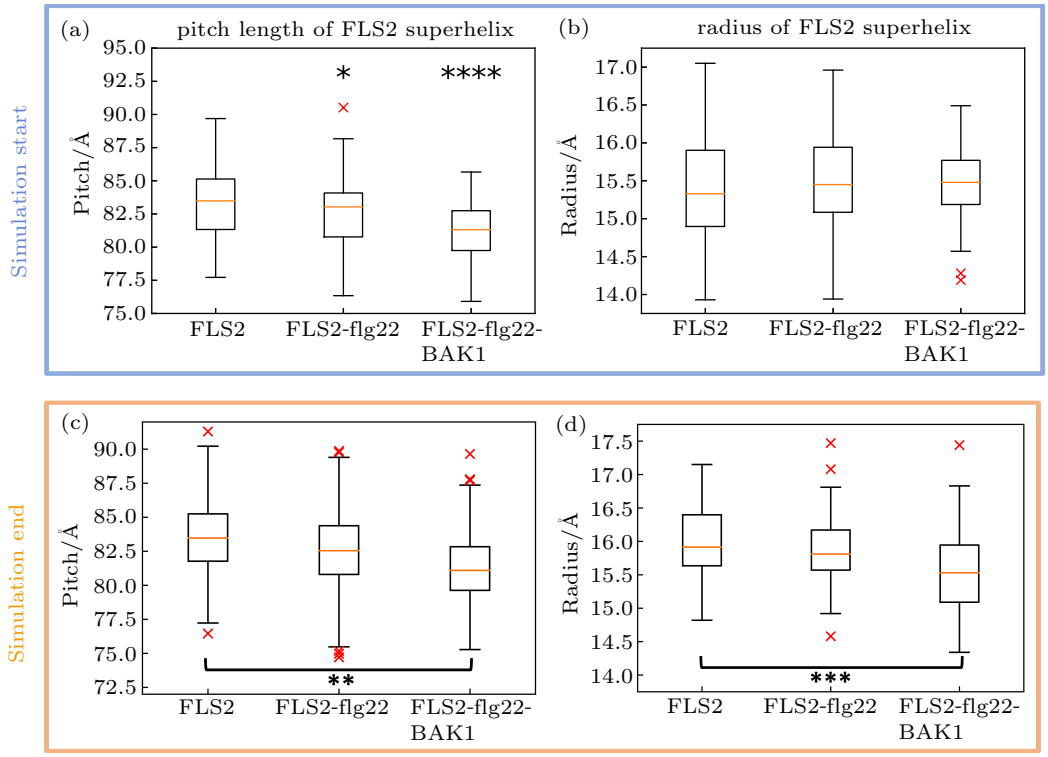


Fig. 2. (a), (c) Pitch length and (b), (d) radius distribution of FLS2 superhelix structure for FLS2, FLS2–flg22, and FLS2–flg22–BAK1 at the start (upper panel) and end (lower panel) episodes of the whole 200 ns MD simulations. From left to right, pitch length and its variance decrease. The radius also fluctuates less and slightly decreases.

3.3. The dominant dynamics of FLS2 can be described by two bending modes and one twisting mode

From the above analysis, the dynamics of FLS2 is suppressed after binding with flg22 and BAK1. More specifically, the change of pitch length and radius of a helix can be visualized as a spring under bending and twisting motion. Does the change of the FLS2 structure reflect its intrinsic dominant dynamics? To answer this, we performed the normal mode analysis for FLS2.[43] For a protein of N residues in 3D space, there are $(3N-6)$ normal modes that describe the intrinsic motion patterns. Figure 3 shows the first 3 dominant normal modes of FLS2.

As indicated by the gray arrows in Fig. 3(a), the first mode (side view) is the bending motion with two ends moving in opposite directions, while the third mode is another bending mode with two ends moving in the same direction. The second mode (top view) is the twisting mode with two ends rotating around the center axis in opposite directions. The two bending modes and one twisting mode lead to a change in the pitch length and radius of the FLS2 superhelix. Figure 3(b) shows the X, Y, and Z components of the first three normal modes.

3.4. PCA of FLS2 MD trajectory reveals the dominant components of structural variance

The above NMA is a theoretical prediction on the dominant motion of FLS2 based on an empirical elastic network model using its static structure information. To examine if FLS2 also shows such dynamics in a more *ab initio* calculation, we performed PCA on the MD trajectory for FLS2. The first three components ranked by their corresponding eigenvalues are shown in Fig. 4(a).

Comparing the shape of each principal component (PC) in Fig. 4(a) with the first three normal modes in Fig. 3(b), PC1 is nearly identical to NM1, PC3 to NM2, and PC2 to NM3 except with a minus sign. Further comparison is shown in the next section. It can be inferred that PC1 and PC2 indicate the bending modes and PC3, the twisting mode of the helix. This is consistent with the work of Emberly *et al.*[36] on the three dominant modes of flexibility for α-helix: two bending modes and one twisting mode. Emberly *et al.*[36] observed

these three modes from the dataset of helical fragments extracted from PDB and analyzed these modes using an elastic spring model. Our first three components show similar patterns to their modes. The twisting mode (PC3) is, however, less pronounced since FLS2 superhelix contains less than 1.5 turns (29 'residues' with $\sim$ 24 'residues' per turn). At the same time, the typical length of helix fragments in Emberly *et al.*'s study is 18 with about 3.6 residues per turn. Note that a 'residue' in FLS2 superhelix stands for an LRR rather than a real residue.

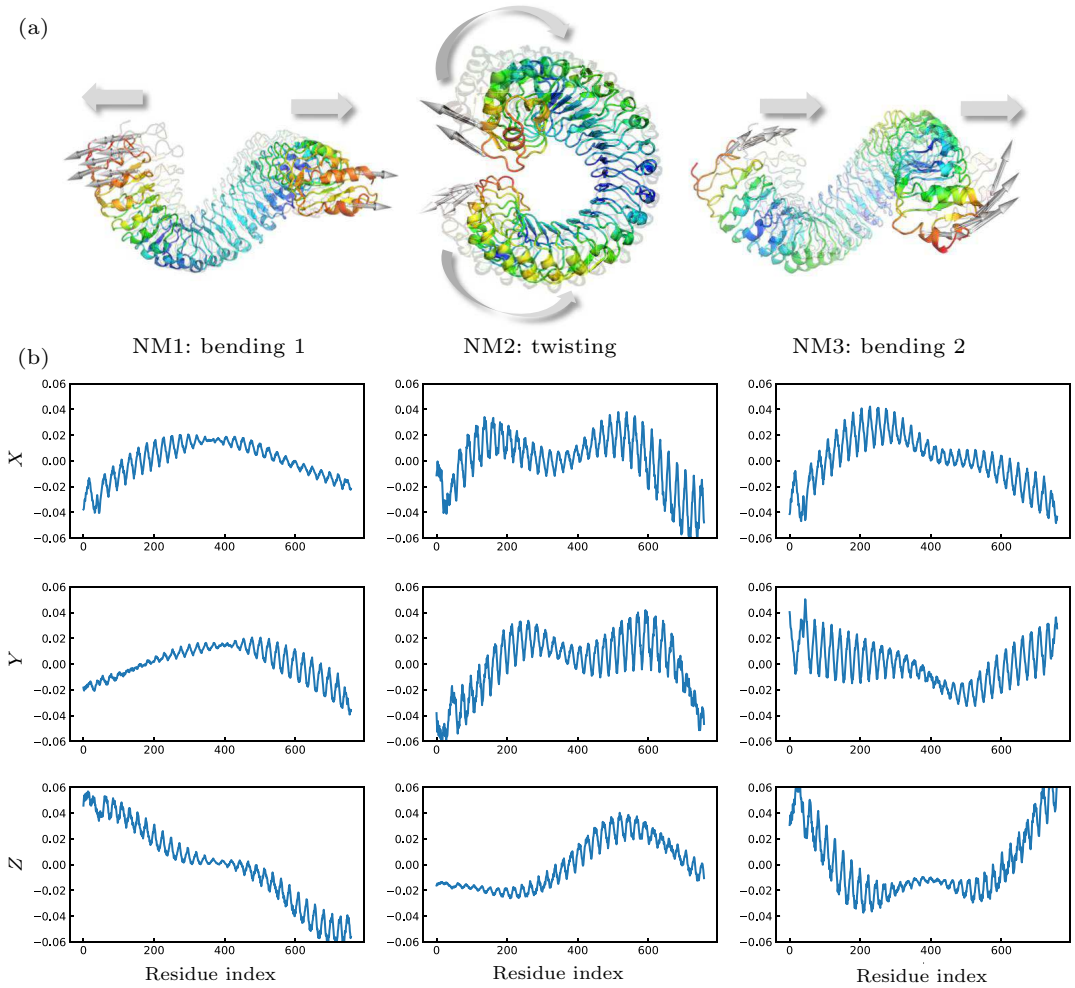


Fig. 3. First 3 dominant normal modes of FLS2. (a) Left: mode 1 (side view) shows the bending motion with two ends moving asynchronously; Middle: mode 2 (top view) shows the twisting motion with two ends rotating around the center axis in opposite directions; Right: mode 3 (side view) shows another bending motion with two synchronous ends. (b) The x, y, z components (in normalized scale) of the first three normal modes (NMs).

Given that PC1 and PC2 account for most of the structural variance (36% and 20% for PC1 and PC2, respectively, as also shown in Fig. 5(a)), the distribution density of the FLS2 MD structure is shown in 2D space spanned by PC1 and PC2 as in Fig. 4(b). Darker color indicates a higher density of structure in the MD trajectory. While some differences among the three distributions can be seen, the two bending modes alone cannot clearly separate the FLS2 structure and dynamics in three different complexes of FLS2, FLS2–flg22, and FLS2–flg22–BAK1.

3.5. Principal modes by NMA and PCA show high correlations

Next, we compared the top 3 PCA modes, which stand for the variations in structures observed in MD simulations, and the top 3 ANM modes obtained from basic elastic theory. The explained variance ratio of the top 10 principal components in PCA is shown in Fig. 5(a). The top three principal components explain around 70% of the variance of all structures in the MD simulations. The correlation of the top 3 PCs and the top 3 NMs is shown in Fig. 5(b). It can be seen that the top PCA modes and ANM modes match well with high correlation coefficients. Note that while PC1 matches best with NM1, PC2 corresponds to NM3 and PC3 to NM2. This consistency in results between PCA and NMA validates that the first two PC modes (corresponding to NM1 and NM3) are bending modes, and PC3 (corresponding to NM2) is a twisting mode. The consistency in results between PCA and NMA indicates that the dominant FLS2 dynamics in MD simulations indeed follows its intrinsic elastic modes. The result coincides with

the previously established overlap between the low-frequency modes in protein fluctuations and large-scale conformational changes in allosteric transitions.[45–47] Besides, the short interaction between FLS2 and flg22 leads to a long-range and global impact of dynamical motions. This matches the idea of the criticality of protein native states.[45]

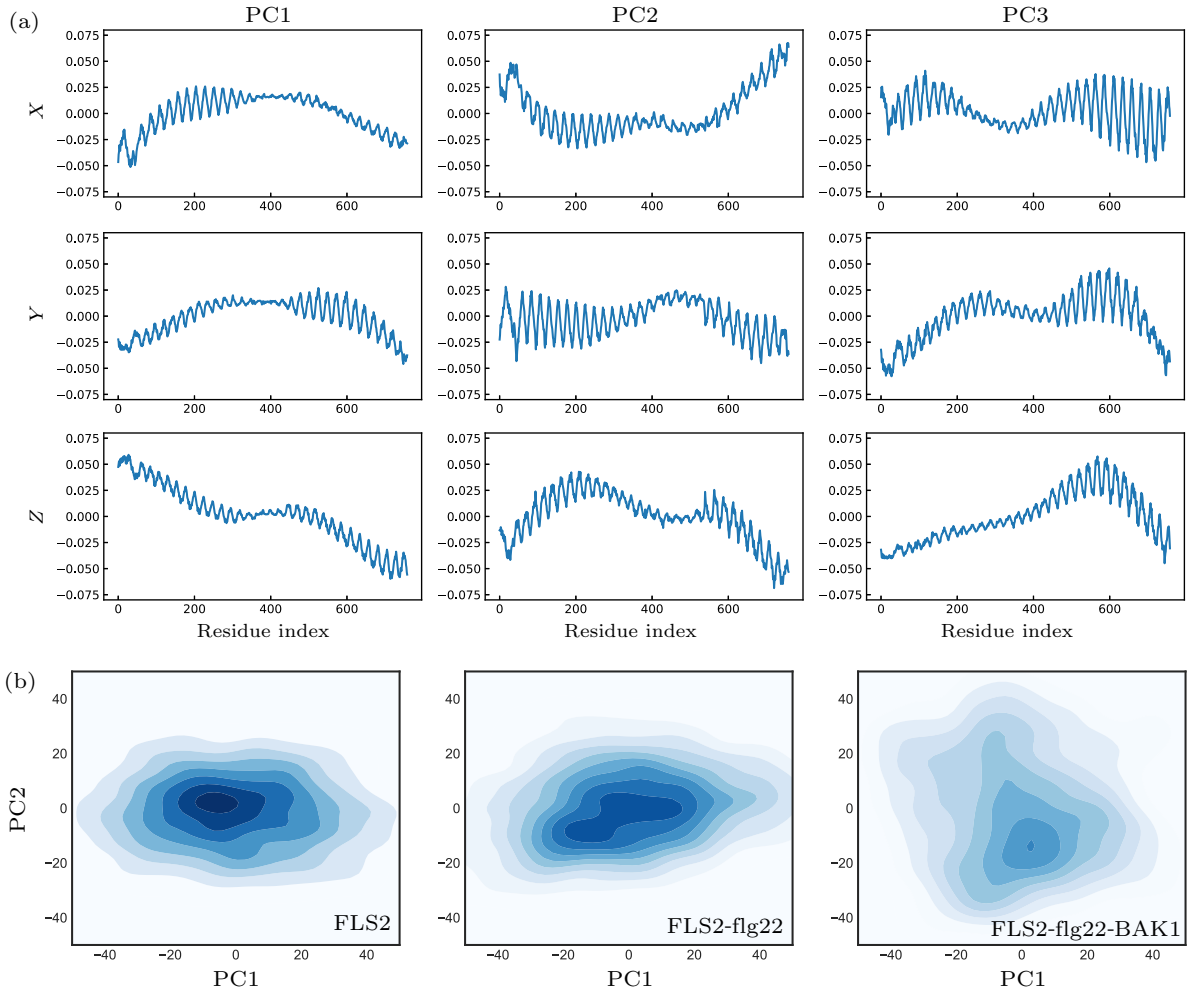


Fig. 4. Principal component analysis of FLS2 MD trajectory. (a) The x, y, z dimensions of the first three principal components. PC1 and PC2 indicate bending modes and PC3 twisting mode of helix. (b) The FLS2 structure distribution density of the MD trajectories in the 2D space spanned by PC1 and PC2 for FLS2-only (lower left), FLS2–flg22 (lower middle), and FLS2–flg22–BAK1 (lower right).

To further check the correlation of the top PCs and NMs, we projected the MD structures into these top mode directions and compared their correlation, as shown in Fig. 5(c). As expected, the projection coefficients of the structure into PC1 and NM1 align very well (with a correlation coefficient as high as 0.961, in Fig. 5(c) left), demonstrating the strong equivalence of the two modes. Similar results were also obtained for PC2–NM3 projection (Fig. 5(c) center) and PC3–NM2 projection (Fig. 5(c) right).

3.6. Twisting mode projection completely separates the structure of FLS2 in the three different complexes

We noticed in Fig. 5(c) that the FLS2 structures in the three different complexes show a well-separated distribution in the PC3 (or NM2) direction. Therefore, the distribution of the FLS2 structure in each dominant mode is examined separately in Fig. 6. As already shown in Fig. 4(b), the structure in the three different complexes cannot be distinguished in PC1 or PC2 direction. However, they show a clear separated distribution in the PC3 direction (Fig. 6 right), i.e., along with the twisting motion. Considering the overlap percentage of projection, the FLS2 structure in the FLS2–flg22–BAK1 complex shows a twisting motion with much smaller amplitude compared with that of FLS2-only. Moreover, the distribution of FLS2 in FLS2–flg22 lies in between those in FLS2–only and FLS2–flg22–BAK1, suggesting that FLS2–flg22 is an intermediate state between FLS2-only and the full FLS2–flg22–BAK1 complex. This reinforces our results in Fig. 2, suggesting that the binding of flg22 to FLS2 suppresses the dynamics of FLS2 and contributes to the FLS2–BAK1 association.

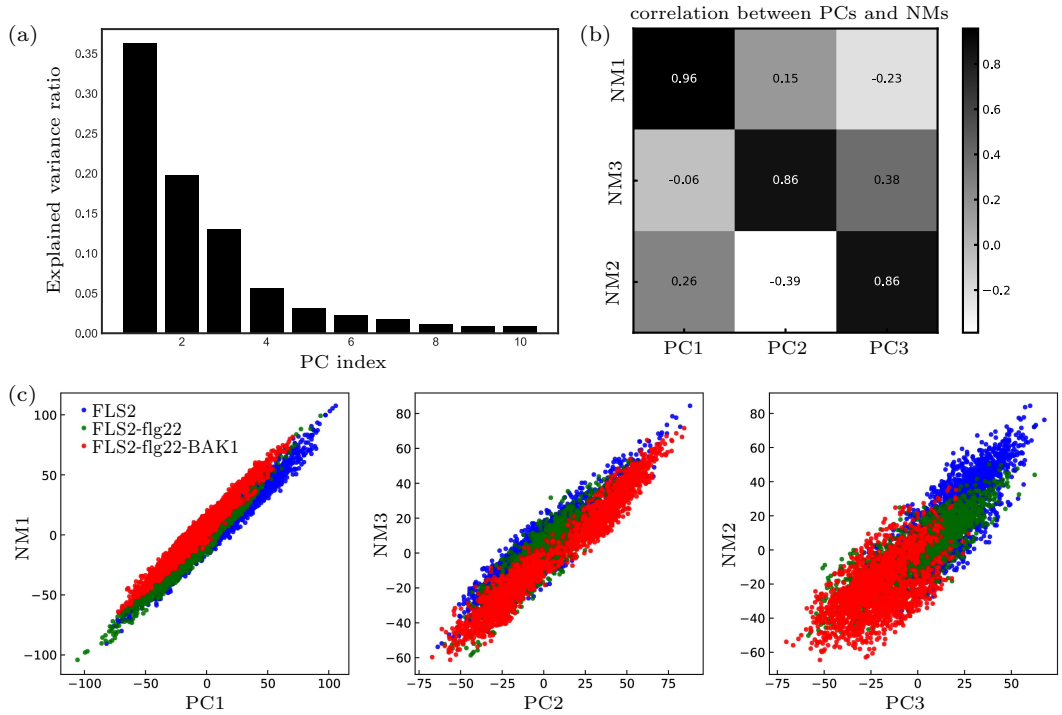


Fig. 5. (a) Explained variance ratio by the top 10 PCs in PCA. Top 3 components (bending1, bending2, and twisting mode) contribute to around 70% of the variance; (b) Correlation of the top 3 components of PCA and the top 3 modes of ANM. PC1 matches best with NM1 (Pearson correlation coefficient 0.96); PC2 corresponds to NM3 (0.93); PC3 to NM2 (0.86). (c) Top PCs and NMs projection correlation. The projection coefficients of the FLS2 MD structures correlate well for PC1–NM1 (left), PC2–NM3 (center), and PC3–NM2 (right), with the correlation coefficients indicated.

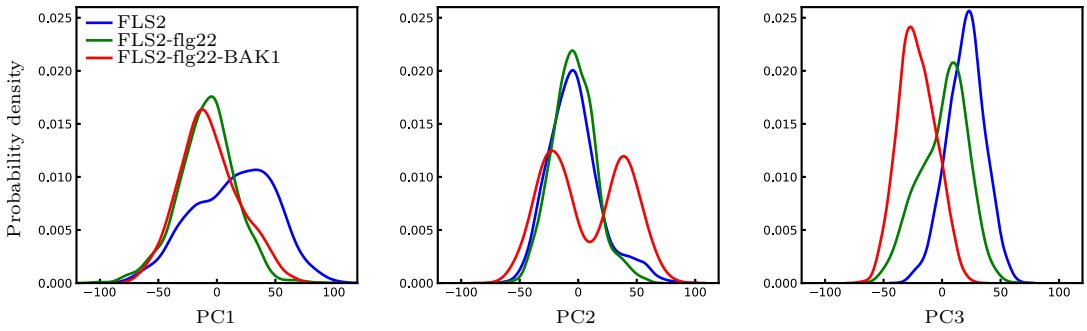


Fig. 6. Top PCs and NMs projection distribution. The distribution of projection coefficients of FLS2 structure in (a) PC1, (b) PC2, and (c) PC3 direction. The FLS2 structure in the three different complexes shows no obvious difference in PC1 and PC2 directions while shows an obvious global shift for FLS2, FLS2–flg22, and FLS2–flg22–BAK1 conditions, with FLS2–flg22 being the intermediate state.

3.7. The sensing mechanism of flg22 by FLS2 is consistent with experiments

From the quantitative analysis of both the conformational change and dominant dynamics modes of FLS2, we conclude that the binding with BAK1 and flg22 immune response of FLS2 are enhanced by the suppressed intrinsic dynamics after the binding of flg22. Flg22 interacts with FLS2 in its concave surface on both halves, acting as an inhibitory rod that effectively restricts the twisting and bending motion of FLS2. The suppressed dynamics of FLS2, together with the molecular "glue" of the C-terminal segment of flg22, contribute to the interaction of FLS2 with BAK1. This proposed sensing mechanism is schematically shown in Fig. 7(a).

This proposed model explains that flg15, a variant of flg22 with seven residues at the N-terminal deleted, is fully active in tomato (*Solanum lycopersicum*, Sl),[48] but displays an extremely low activity in *Arabidopsis thaliana* (1000-fold difference) even if it can still bind with FLS2.[48] Mueller *et al*. designed a chimeric FLS2 by substituting a subset of the 28 LRRs in At-FLS2 with the corresponding LRRs from Sl-FLS2. They demonstrated that the different effects of flg15 in At-FLS2 and Sl-FLS2 are due to the heterogeneity of LRR7-10, which is slightly downward from the original binding sites of FLS2 with N-terminus of flg22. This makes sense because as the flagellin peptide shortens, the binding sites of its two halves have to be closer and more towards the center kink of FLS2, moving from LRR4-6 towards LRR7-10. The new N-terminus of flg15 fails to bind with At-FLS2, thus resulting in little restriction on the twisting dynamics of FLS2 and a much weaker FLS2–BAK1 association or the corresponding downstream immune response. The heterogeneity of FLS2 LRR7-10 may explain the significantly different effect of flg15 in *Arabidopsis* and tomato.

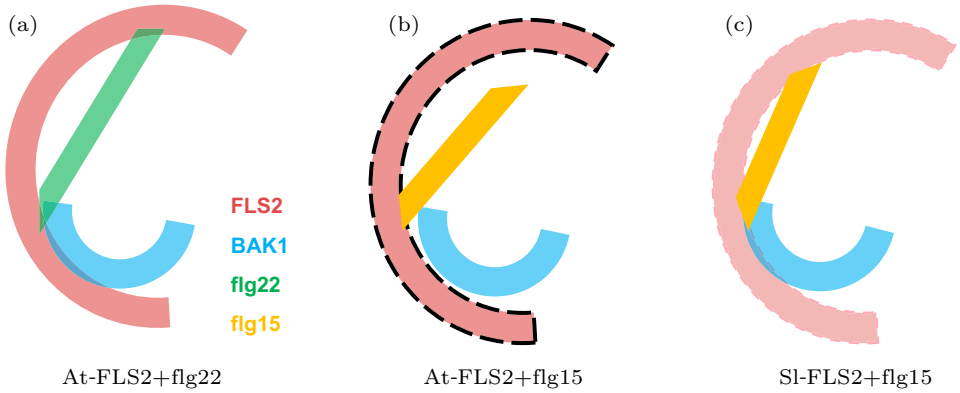


Fig. 7. Proposed interaction mechanism of the FLS2–flg22–BAK1 complex. (a) Interaction of FLS2–flg22–BAK1. N- and C-terminal ends of flg22 interact with At-FLS2 and promote the binding of BAK1. Interaction of FLS2–flg15–BAK1 in (b) *Arabidopsis thaliana* (At) and (c) *Solanum lycopersicum* (Sl). The new N-terminus of flg15 fails to bind with At-FLS2, thus resulting in little restriction on the twisting dynamics of FLS2 and, therefore much weaker FLS2–BAK1 association or correspondingly, much weaker downstream immune response. In Sl, the N-terminus of flg15 binds with Sl-FLS2 to induce a tight binding of BAK1.

We further performed a sequence evolution analysis to investigate the conservation property of the binding sites of FLS2–flg22. As shown in Fig. 8(a), the C-terminus of flg22 is highly conserved, while the N-terminus is more variable. Interestingly, this difference is also present in the corresponding sites in FLS2. The binding sites of FLS2 with N-terminal flg22 are less conserved than those binding with flg22 C-terminus (Fig. 8(b)).

This matching conservation patterns may reflect the co-evolution, i.e., two sites must evolve accordingly to maintain their interaction. The relatively lower conservation of these sites may be presumably for adapting to the variable flg22 N-terminus and sensing-specificity in different plant species.

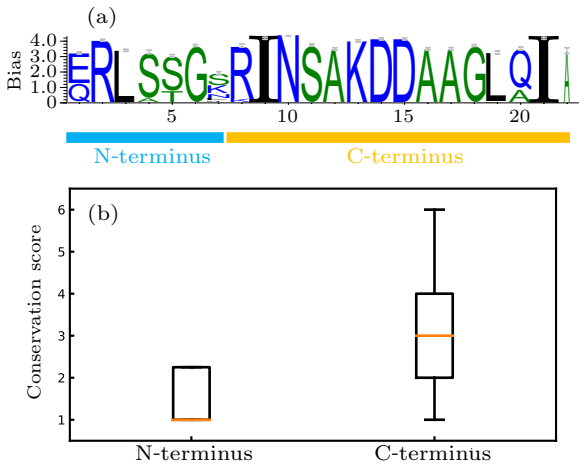


Fig. 8. Conservation analysis of FLS2–flg22 binding sites. (a) Conservation of flg22. The N-terminus (first 7 residues) is less conserved than the C-terminus. (b) The corresponding binding sites of FLS2 with flg22 N-terminus are also less conserved than those with flg22 C-terminus.

4. Discussion and summary

While the two bending modes are the most dominant (with 36% and 20% variance ratio) for FLS2 intrinsic dynamics and MD simulation variance, the component that actually best separates the structure of FLS2 before and after binding with flg22 and BAK1 is the twisting mode. One reason may be that the displacement induced by the bending modes is quadratic with the helix length, while linear for the twisting mode. So the variance contribution is proportional to the 4th and 2nd power of helix length for bending and twist modes, respectively. This means that the variance in the bending modes is mostly caused by the two ends. However, the interaction domain of FLS2–flg22–BAK1 is in the center portion of FLS2, where the twist plays a more important role in determining the local conformation. Besides, compared with a simple helix, the FLS2 superhelical structure is softer, especially in the twisting mode, because its building block is an entire LRR unit instead of a single amino acid. As a result, the twisting mode better separates different states of FLS2.

PRRs and their sensing of PAMPs are essential for plant immunity. While plants can recognize many different microbial invaders through certain highly conserved patterns, the faster mutation rate of the microbes can result in some PAMPs that fail to be recognized by plants. On the other hand, different plants may contain the same PRR with slightly different sequences and structures for functional specialization. For FLS2, while both At-FLS2 and Sl-FLS2 respond to the genuine flg22 at picomolar concentration, considerable differences occur in recognition of shortened or modified flg22 ligands such as flg15, CLV3p, and flg22-AYA.[48] Indeed, the double-Ala scanning mutagenesis experiments by Dunning *et al.*[49] have identified some sites that can directly influence the perception function of At-FLS2 to flg22.

While many such cases have been studied and experimentally validated, the causal mechanism between the minor variance or even single amino acid variants (SAVs) with the diverse functional behaviors is mostly unknown. Currently, sequence conservation and structural properties are the two main attributes of this puzzle. Although they prove to be effective, some shortcomings still exist. Most recent studies by Bahar *et al.* have shown that the structural dynamical implication of SAVs can be a determinant of functional significance in human

proteins.[24] The more interference of a site mutation can create in the structural flexibility and conformational stability, the higher chance it will lead to a function alteration or even dysfunction. In this paper, the suppression of dominant dynamical modes, especially the twisting mode of FLS2 has been demonstrated to be crucial for the flg22 sensing and FLS2–BAK1 association. Taking these all together, it is our conjecture that these "sensitive" sites for FLS2 functionality may impact its global dynamics, which remains to be validated by detailed dynamic analysis with more solved structures.

In summary, we evaluated the intrinsic normal modes of FLS2 superhelical structure and showed that the three principal modes – two bending and one twisting – coincide with the normal modes of a standard helix shown before. Moreover, we demonstrated that the process of flg22 sensing by FLS2 with BAK1 could mainly be attributed to the suppressed twisting dynamics of FLS2. A better understanding of this sensing mechanism of FLS2 provides a potential solution for custom-made PRR and drug design for plant immunity.

References

[1] Helft L, Thompson M and Bent A F 2016 *PLoS One* **11** e0157155
[2] Saijo Y, Loo E P and Yasuda S 2018 *Plant J.* **93** 592
[3] Medzhitov R and Janeway C A, Jr. 1997 *Cell* **91** 295
[4] Boller T and Felix G 2009 *Annu. Rev. Plant Biol.* **60** 379
[5] Zipfel C 2014 *Trends Immunol* **35** 345
[6] Boutrot F and Zipfel C 2017 *Annu Rev. Phytopathol* **55** 257
[7] Zipfel C 2009 *Curr. Opin. Plant. Biol.* **12** 414
[8] Tintor N, Ross A, Kanehara K, Yamada K, Fan L, Kemmerling B, Nurnberger T, Tsuda K and Saijo Y 2013 *Proc. Natl. Acad. Sci. USA* **110** 6211
[9] Robatzek S and Wirthmueller L 2013 *Protoplasma* **250** 671
[10] Sun Y, Li L, Macho A P, Han Z, Hu Z, Zipfel C, Zhou J M and Chai J 2013 *Science* **342** 624
[11] Padmanabhan M, Cournoyer P and Dinesh-Kumar S P 2009 *Cell Microbiol* **11** 191
[12] Zhang Z and Thomma B P 2013 *J. Integr. Plant Biol.* **55** 1212
[13] Chinchilla D, Bauer Z, Regenass M, Boller T and Felix G 2006 *Plant Cell* **18** 465
[14] Gomez-Gomez L and Boller T 2000 *Mol. Cell* **5** 1003
[15] Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones J D, Boller T and Felix G 2006 *Cell* **125** 749
[16] Mithoe S C and Menke F L 2018 *Curr. Opin. Plant Biol.* **45** 162
[17] Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nurnberger T, Jones J D, Felix G and Boller T 2007 *Nature* **448** 497
[18] Somssich M, Ma Q, Weidtkamp-Peters S, Stahl Y, Felekyan S, Bleckmann A, Seidel C A and Simon R 2015 *Sci. Signal* **8** ra76
[19] Beck M, Zhou J, Faulkner C, MacLean D and Robatzek S 2012 *Plant Cell* **24** 4205
[20] Fischer E 1894 *Berichte der deutschen chemischen Gesellschaft* **27** 2985
[21] Koshland D E 1958 *Proc. Natl. Acad. Sci. USA* **44** 98
[22] Tobi D and Bahar I 2005 *Proc. Natl. Acad. Sci. USA* **102** 18908
[23] Bakan A and Bahar I 2009 *Proc. Natl. Acad. Sci. USA* **106** 14349
[24] Ponzoni L and Bahar I 2018 *Proc. Natl. Acad. Sci. USA* **115** 4164
[25] Wang K L, Jian Y R, Wang H W, Zeng C and Zhao Y J 2018 *Bioinformatics* **34** 3131
[26] Wang H W, Wang K L, Guan Z Y, Jian Y R, Jia Y, Kashanchi F, Zeng C and Zhao Y J 2017 *Chin. Phys. B* **26** 128702
[27] Zhao Y, Jian Y, Liu Z, Liu H, Liu Q, Chen C, Li Z, Wang L, Huang H H and Zeng C 2017 *Sci. Rep.* **7** 2876
[28] Chen H, Zhao Y J, Li H T, Zhang D Y, Huang Y Z, Shen Q, Van Duyne R, Kashanchi F, Zeng C and Liu S Y 2014 *Plos One* **9**
[29] Yi Yang Q G, Zhang B G, Shi Y Z and Shao Z G 2018 *Chin. Phys. B* **27** 38701
[30] Dill K A and Chan H S 1997 *Nat. Struct. Biol.* **4** 10
[31] Brunette T J, Parmeggiani F, Huang P S, Bhabha G, Ekiert D C, Tsutakawa S E, Hura G L, Tainer J A and Baker D 2015 *Nature* **528** 580
[32] Enkhbayar P, Kamiya M, Osaki M, Matsumoto T and Matsushima N 2004 *Proteins* **54** 394
[33] Stumpp M T, Forrer P, Binz H K and Pluckthun A 2003 *J. Mol. Biol.* **332** 471
[34] Ramisch S, Weininger U, Martinsson J, Akke M and Andre I 2014 *Proc. Natl. Acad Sci. USA* **111** 17875
[35] Doyle L, Hallinan J, Bolduc J, Parmeggiani F, Baker D, Stoddard B L and Bradley P 2015 *Nature* **528** 585
[36] Emberly E G, Mukhopadhyay R, Wingreen N S and Tang C 2003 *J. Mol. Biol.* **327** 229
[37] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A E and Berendsen H J 2005 *J. Comput. Chem.* **26** 1701
[38] Duan Y, Wu C, Chowdhury S, Lee M C, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J and Kollman P 2003 *J. Comput. Chem.* **24** 1999
[39] Sethi A, Eargle J, Black A A and Luthey-Schulten Z 2009 *Proc. Natl. Acad. Sci. USA* **106** 6620
[40] Humphrey W, Dalke A and Schulten K 1996 *J. Mol. Graph* **14** 33
[41] Enkhbayar P, Damdinsuren S, Osaki M and Matsushima N 2008 *Comput. Biol. Chem.* **32** 307
[42] Atilgan A R, Durell S R, Jernigan R L, Demirel M C, Keskin O and Bahar I 2001 *Biophys. J.* **80** 505
[43] Eyal E, Lum G and Bahar I 2015 *Bioinformatics* **31** 1487
[44] Ashkenazy H, Erez E, Martz E, Pupko T and Ben-Tal N 2010 *Nucleic Acids Res.* **38** W529
[45] Tang Q Y, Zhang Y Y, Wang J, Wang W and Chialvo D R 2017 *Phys. Rev. Lett.* **118** 088102
[46] Tang Q Y and Kaneko K 2020 *PLoS Comput. Biol.* **16** e1007670
[47] Bahar I, Lezon T R, Yang L W and Eyal E 2010 *Ann. Rev. Biophys.* **39** 23
[48] Mueller K, Bittel P, Chinchilla D, Jehle A K, Albert M, Boller T and Felix G 2012 *Plant Cell* **24** 2213
[49] Dunning F M, Sun W, Jansen K L, Helft L and Bent A F 2007 *Plant Cell* **19** 3297