Structural basis of a novel repressor, SghR, controlling *Agrobacterium* infection by cross-talking to plants

Received for publication, February 4, 2020, and in revised form, July 3, 2020 Published, Papers in Press, July 10, 2020, DOI 10.1074/jbc.RA120.012908

Fuzhou Ye, Chao Wang, Qinqin Fu, Xin-Fu Yan, Sakshibeedu R. Bharath, Arnau Casanas, Meitian Wang, Haiwei Song, Lian-Hui Zhang, and Yong-Gui Gao

From the 1School of Biological Sciences, Nanyang Technological University, Singapore, 2State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou, China, 3Institute of Molecular and Cell Biology, Singapore, and 4Swiss Light Source at Paul Scherrer Institut, Villigen, Switzerland

Edited by Joseph M. Jez

*Agrobacterium tumefaciens* infects various plants and causes crown gall diseases involving temporal expression of virulence factors. SghA is a newly identified virulence factor enzymatically releasing salicylic acid from its glucoside conjugate and controlling plant tumor development. Here, we report the structural basis of SghA, a LacI-type transcription factor highly conserved in *Rhizobiales* family, regulating the expression of SghA and involved in tumorigenesis. We identified and characterized the binding site of SghR on the promoter region of *sghA* and then determined the crystal structures of apo-SghR, SghR complexed with its operator DNA, and ligand sucrose, respectively. These results provide detailed insights into how SghR recognizes its cognate DNA and shed a mechanistic light on how sucrose attenuates the affinity of SghR with DNA to modulate the expression of SghA. Given the important role of SghR in mediating the signaling cross-talk during *Agrobacterium* infection, our results pave the way for structure-based inducer analog design, which has potential applications for agricultural industry.

Regulation of gene transcription is essential for bacteria to maintain their cellular function and respond to environmental changes. In bacteria, a large number of transcription factors have been identified and shown to form tremendously complicated regulation networks that control gene expression (1). Transcription factors are generally classified into repressors and activators, which block and promote gene transcription, respectively. Most transcription factors are composed of two domains: a regulatory domain (core domain) for sensing the external or internal signals and triggering the conformational change and a DNA-binding domain (DBD) for associating/dissociating with its cognate DNA in response to the conformational change originated from the regulatory domain. These two domains function allosterically to play a central role in gene regulation.

As one of the most well studied transcription factors, LacI family regulators have been discovered in numerous bacteria to coordinate a wide range of cellular processes (2, 3). Since the first lac repressor from *Escherichia coli* identified in the 1960s (4), more than 1000 bacterial members of the LacI family have been characterized with diverse functions (2). Therefore, the LacI family has been of considerable interest in molecular biology, microbiology, and structural biology. Detailed studies of these LacI-type regulators not only have advanced our understanding of the genetic control of a variety of metabolic and signaling pathways but also have revealed huge potentials in practical applications, such as redesigning of existing proteins in synthetic biology (5).

As the causative agent of crown gall disease, *A. tumefaciens* infects a wide range of plants, which is of great concern to the agricultural industry. This bacterial pathogen invades plants by transferring its tumor-inducing plasmid into the plant genomes, which also renders it a useful tool in plant genetic engineering. Infection by *A. tumefaciens* is mainly regulated by the *Agrobacterium* VirA/VirG two-component regulatory system (6). Independent of VirA/VirG, we have recently found that a pair of novel proteins, SghR and SghA, regulate plant tumor growth through controlled release of the host defense signal salicylic acid (SA) by a previously uncharacterized mechanism. SghA is a glucosidase, controlling the release of SA from its inactive storage form, SA 2-O-β-d-glucoside, in host plants, which then activates plant immune response and shuts down bacterial virulence when the infection is successfully established (7).

Sucrose is a major photosynthetic product representing >95% transported carbohydrate in plant phloem, and it could reach a concentration of up to 1 M in the conducting vascular cells and 2–7 m in the extracellular space (8). In addition, sucrose would be unloaded from the sieve elements of phloem, leading to its accumulation to reach even higher concentrations while tissue wounding (9, 10). It has also been reported that sucrose plays an important role in wound signaling (11) and endogenous signaling to induce defense responses against pathogens in rice (12). We found that the accumulated sucrose at the wound sites during healing appears to serve as an environmental stimulus involved in the SghR-SghA pair mediating the signaling cross-talk at the late stage of *Agrobacterium* infection (7).

Here, we first investigated the functional regulation of SghR on SghA *in vivo* by performing a tumorigenesis assay and RT-PCR. The results demonstrated the involvement of SghR/SghA in tumorigenesis, and transcription regulation of SghA is...
modulated by SghR. Given that the expression of SghA is negatively regulated by SghR and SghR is a LacI family transcription factor and highly conserved in the Rhizobiaceae family, we further identified and characterized the cognate DNA sequence for SghR binding on the sghA promoter region. Subsequently, we determined the crystal structures of isolated SghR, SghR–DNA, and SghR–sucrose complexes. Structural analyses unraveled the detailed interactions of SghR with its cognate DNA and its effector, sucrose, as well as uncovered the allosteric mechanism underlying how sucrose relieves the repression of SghA upon its binding to SghR. Taken together, SghR controls Agrobacterium infection through its regulation of SghR in response to host signal sucrose, and our results are of considerable importance for establishing a detailed understanding of gene regulation during the process of host–pathogen interaction.

Results

The regulation of SghA by SghR

SghA, an SA-releasing enzyme, has been implicated in immune response and tumorogenesis during Agrobacterium infection, which is likely relevant to SghR (7). To further reveal that SghA is regulated by SghR and that this pair, SghR-SghA, is involved in regulating plant tumor development, we examined the tumor occurrence and growth on carrot disks that are infected by either WT or deletion mutant Agrobacterium strains (ΔsghA, ΔsghR, and ΔsghRA). First, we recorded the number of tumors incited on carrot disks and found no significant difference among the WT and these mutant strains, suggesting that SghA does not involve the initiation of infection. However, the tumor weight, which has been used to measure the severity of crown galls (13, 14), appeared to be ~300 mg for the ΔsghA and ~120 mg for the ΔsghR mutant, in contrast to ~200 mg for the WT A6. Statistical analyses show significant difference between WT A6 and either the ΔsghA or ΔsghR mutant (Fig. 1A). These results appeared consistent with our previous data with the Tn5 transposon mutants (7), further demonstrating that the transcription factor SghR negatively regulates the expression of sghA (Fig. 1B) and the SghR-SghA pair is involved in tumor growth after agrobacterial infection.

Characterization of the cognate DNA sequence of SghR on sghA promoter region

To identify the exact DNA-binding site of SghR on the sghA promoter, a DNase I protection assay was carried out. The sequencing result using a FAM-labeled forward primer indicated that SghR protects a DNA region covering the sequence 5′-GCTGAAAAACGTGCAATTTTGGCT-3′ from digestion (Fig. 1C and D, and Fig. S1A), thereby suggesting that this region is the DNA-binding site for SghR. This binding site was also confirmed by sequencing using an FAM-labeled reverse primer in the DNase I protection assay (Fig. S1B). Sequence analyses showed this binding region comprised a pseudopalindromic sequence (Fig. 1D). Furthermore, EMSA results showed that the binding ratio of SghR dimer protein to DNA appears to be 1:1, as they completely bind to each other and shift together without free DNA or free protein observed in the case of mixing SghR dimer and DNA in equal molar amounts, whereas free DNA or free protein bands could be observed given an excess amount added (Fig. 1F). Isothermal titration calorimetry (ITC) assay further demonstrated that the stoichiometry of protein and DNA complex was one SghR dimer to one double-stranded DNA (dsDNA), with a dissociation constant (Kd) value of 60 nM (Fig. 2B), although there is no interaction observed between SghR and Ppa0305 promoter DNA from Pseudomonas aeruginosa (Fig. 2A). When incubating SghR protein with sucrose before the titration experiment, the Kd value increased up to 1.7 μM (Fig. 2D), implying that sucrose attenuates the repression effect of SghR on the sghA promoter.

Overall structure of apo SghR

The crystal structure of isolated full-length SghR was refined at 2.1-Å resolution to a working R value of 20.6% and a free R value of 25.9% (Table 1). There are two copies of SghR in one asymmetric unit, with a root mean square deviation (r.m.s.d.) of 0.4 Å for 275 aligned Ca atoms, suggesting the two SghR molecules are almost identical. The final model (chain A, as an example) comprises 275 out of 350 residues, corresponding to the C-terminal core domain, whereas the N-terminal region (residues 1–74), which presumably functions as a DNA-binding domain, was not observed in the resultant electron density, implying that this region is disordered. The stereochemical quality of the final model was validated with PROCHECK (15), and the Ramachandran plot indicated that 96.3% of the residues lie in the most favored regions, with the remaining 3.7% in the allowed region (Table 1).

The C-terminal domain, the core of SghR protein, comprises two subdomains: the N- and C-terminal subdomains. The N-terminal subdomain is composed of a core of six parallel β strands (β1–β5 and β11) flanked by two α helices (α6 and α7) on one side and three (α5, α12, and α13) on the other side (Fig. 3A). Meanwhile, the C-terminal subdomain consists of a core of five parallel β strands (β6–β10), which is also flanked by two α helices on each side, namely, α8, α9, α10, and α11, respectively (Fig. 3A). The two subdomains are connected by three noncontiguous peptide crossovers, which not only form a cleft to allow ligand binding but also create a linker region serving as a hinge to permit the movements of the subdomains relative to each other upon ligand association and dissociation.

Dimerization of SghR

Analysis of the apo SghR structure revealed that two SghR molecules in an asymmetric unit form a dimer (Fig. 3B), which was further supported by the molecular weight calibration result indicating that SghR exists as a homodimer in solution (16). Upon dimerization, the buried interface is 2870 Å², accounting for ~11.5% of the total solvent-accessible area. The dimerization interface involves helix α5 and strand β2 in the N-terminal subdomains of both monomers as well as helix α11 and the loop connecting β9 and β10 in the C-terminal subdomains of both monomers, where they pack against each other. The formation of dimer is mainly through hydrogen-bonding interactions formed between residues Ser95-Glu88′, Val114-Leu112′, and their dyad symmetry mates (where prime refers to the second monomer within a dimer, same as below). In
Structural insights into the regulation mechanism by SghR

A

B

C

D

E

free D  free Pro  ProD=1:1  ProD=1:2  ProD=2:1

free D  free Pro  ProD=1:1  ProD=1:2  ProD=2:1
addition, hydrophobic interactions involving residues Met91, Thr102, Pro116, Ile270, Phe297, Trp300, and Ile301 also contribute to the dimer assembly (Fig. 3, C and D).

Overall structure of SghR–dsDNA

To gain structural insights into how SghR recognizes the sghA operator DNA, an intense effort has been put into improving the diffraction quality of SghR–dsDNA complex crystals. Different dsDNAs encompassing the core region (partial palindrome, as indicated in Fig. 1C) with varying lengths and sequences, as well as different SghR truncations containing both the N-terminal DBD and the C-terminal core, were tested for crystallization. Ultimately, well diffracting crystals of SghR–dsDNA complex were obtained by cocRYSTALLization of the truncated SghR (without the first N-terminal 17 residues) and a 16-bp dsDNA with a two-nucleotide overhang at the 5’ end (Fig. 4, B and D). Notably, two base pairs on the left half-site of the natural pseudo-palindrome dsDNA were mutated to make a perfect palindrome sequence (Fig. 4, B and D). Indeed, this strategy was frequently used to obtain the crystal structures of lac repressor–DNA complexes, and it was proposed that it would not change the overall conformation by comparing the lac repressor-natural operator structure with the lac repressor-symmetric operator structure (17, 18).

The crystal structure of SghR–dsDNA was solved by the molecular replacement method using the C-terminal domain from our apo SghR structure combining the cognate DNA-bound N-terminal DBD from the lac repressor-DNA complex structure (PDB entry 1EFA) as a search model. The structure was refined to 3.2-Å resolution with crystallographic statistics shown in Table 1. The final model contains 6 dimers and 6 dsDNA molecules per asymmetric unit. Each dimer binds one dsDNA, consistent with our ITC and EMSA results that the stoichiometry of SghR dimer to dsDNA is 1 to 1 (Fig. 1D and 2B). Comparison of the six SghR dimers revealed that their overall structures are quite similar, with an average r.m.s.d. of 0.75 Å for the aligned Ca atoms. Therefore, we used the SghR–dsDNA complex structure to follow analysis (Fig. 4A). Structural analysis of the SghR–dsDNA complex revealed that SghR contains an additional headpiece, referred to as the N-terminal DBD, besides the large C-terminal core adopting a similar conformation as observed in the isolated SghR structure (Fig. 4A). The N-terminal DBD (residues 18–73) consists of four helices and can be divided into two functionally important parts. The first part contains a helix-turn-helix (HTH) motif (α1-turn-α2, residues 22–40), followed by a short loop (residues 41–47), the helix α3 (residues 48–61), and its C-terminal loop (residues 62–66) (Fig. 4A). Connecting the globular subdomain involving the first three helices to the C-terminal core are the fourth helix α4 (residues 67–73) and the following loop (residues 73–78) (Fig. 4A). The α4 helix, usually named the hinge helix, constitutes the second functionally important part of the N-terminal DBD. The first part within the N-terminal domain is responsible for the dsDNA major groove interaction, whereas the hinge helix binds with its 2-fold related mate to the minor groove of the dsDNA (Fig. 4, A and C). It was proposed that the formation of hinge helix was conducted by both the C-terminal core dimerization and the DNA minor groove recognition (19). Note that in our structure, stabilization of such a helix conformation was mediated by hydrophobic interaction involving the residues Gly69-Leu72’ and van der Waals forces between residues Ala68 and Arg71’ and their dyad symmetry mates (Fig. 52). In addition to the C-terminal core domain and the hinge helix, hydrogen bonds formed between Arg19 in the N-terminal DBD and Thr133’ in the N-subdomain also contribute to the dimerization of SghR in our complex (see Fig. 6C). Interestingly, the dsDNA in this structure is kinked ~45° at the central C10-G11’ base pair toward the major groove and pulled away from the protein, which is largely because of the insertion of the two hinge helices α4 and α4’ into the minor groove (Fig. 4A).

SghR–DNA interaction

In our SghR–dsDNA structure, each SghR monomer binds to a semisymmetrical half of the dsDNA via extensive hydrogen-bonding interactions (Fig. 4C). In particular, the HTH (α1-turn-α2) motifs within the N-terminal DBD of the two monomers make extensive and symmetrically related contacts with the bases and phosphate backbone in the major grooves located at both half-sites of the dsDNA. Taking one monomer (chain A) as an example, Leu22 in α1, Thr35 and Ser37 in α2, Thr50 in α3, as well as Gly32, Lys41, and Gly47, proximal to the HTH motif, interact with the phosphate backbone of DNA via hydrogen bonds (Fig. 4, C and D). Notably, the canonical recognition helix α2 is well positioned, such that its N terminus projects into the major groove of dsDNA for specific interaction, namely, hydrogen bond between Thr34 and the base of G-6 in DNA chain b. Additionally, the residue Asp45, located in the loop connecting α2 and α3, forms a base-specific hydrogen-

Figure 1. SghR–SghA involved in plant tumorigenesis and SghR regulating SghA via binding to its promoter region. A. tumefaciens A6 and its derivatives using carrot discs. Representative images (bottom) were taken and fresh weight of tumors (top) were recorded at 6 weeks after inoculation. ΔsghR, ΔsghA, and ΔsghAΔΔsghR indicate the in-frame deletions of sghR, sghA, and both genes, respectively. The error bars represent SD from 6 repeats. A nonparametric 1-way analysis of variance was performed in GraphPad for statistical analyses, *p < 0.01; **p < 0.001; ***p < 0.0001. B. SghR represses the transcription of the gene indicated by RT-PCR. The transcription level of 16S RNA in two strains is also presented. C. Identification of SghR-binding site in sghA promoter region using Dye Primer sequencing on an automated capillary DNA analyzer. The double-ended arrow shows the region in the sghA promoter protected by SghR during DNase I digestion. The sghA promoter region and DNase I are incubated either in the presence (−SghR) or absence (+SghR) of SghR protein. Electropherograms of ddNTP (ddATP, ddTTP, ddCTP, and ddGTP) panels displayed in the figure are the sequencing results using forward primer (6-FAM labeling). D. Depiction of the SghR-binding site in the promoter region of sghA. The binding site is shown in red letters. The two opposite red arrows indicate the partial palindrome feature presented in the binding site. −35 and −10 regions of sghA were predicted by the online program BPRoM, and sequences were labeled with blue lines. The pentacle indicates the transcription start site, predicted by an online server (http://www.fruitfly.org/seq_tools/promoter.html). The translation start site of SghA is also indicated by a blue arrow. E. In vitro gel shift assay of SghR with the identified binding site. Various ratios of protein to DNA were used. Left, native-PAGE gel stained by DNA-staining dye. Right, the same gel stained by Coomassie brilliant blue (CBB). D and Pro represent DNA and protein, respectively. ProD = 1:1, 1:2, and 2:1 indicate that the molar ratios of protein dimer to dsDNA are 1 to 1, 1 to 2, and 2 to 1, respectively. This is native PAGE; no protein marker was included in the gel, as the charge of protein also contributes to its migration rate in addition to its molecular weight.
Figure 2. Binding affinity of SghR with Ppa0305 promoter DNA (A), dsDNA from sghA promoter (B), sucrose (C), and sucrose-binding SghR with dsDNA (D). Panel A is a control experiment for panel B, B and D, ITC data of titrating dsDNA into SghR and SghR–sucrose mixture, respectively. Raw data are shown in the upper panels, whereas the fitted curves are shown in the lower panels. Both profiles were fitted into one set of the site model. C, quantification of binding affinity between SghR and sucrose by microscale thermophoresis. Data points indicate the difference in normalized fluorescence (%ΔF/norm) generated by SghR protein, and the curve indicates the calculated fit.
bonding interaction with the base of C-4 in DNA chain b (Fig. 4, C and D). As aforementioned, the hinge helices α4 and α4′ insert into the minor groove of the dsDNA, involving both specific and nonspecific interactions. The insertion of these two helices into the narrow minor groove is accomplished by the side chains of residues Tyr63 and Asn66, which intercalate into the central GC pairs (C10–G11′ and G11–C10′) and act as a lever to pry open the minor groove. As a result, the main chain O of both Gly69 and Leu72 forms hydrogen bonds with the base of G-11 in DNA a and the base of C-10 in DNA chain b, respectively (Fig. 4, C and D). Remarkably, a single residue in α4, Arg73, makes a complicated hydrogen-bonding network involving several nucleotides, namely, the base of T-13 and the deoxyribose of G-14 in DNA chain a, as well as the bases of A-9 and C-10 in DNA chain b (Fig. 4, C and D). In proximity to the hinge helix α4, the residues Tyr63 and Asn66 contact the phosphate backbone of the DNA chain a (Fig. 4, C and D). Therefore, the residues Thr34, Asp45, Gly69, Leu72, and Arg73 from each monomer contribute to sequence-specific recognition of each half-site of its cognate DNA, which validates the necessity of palindrome DNA sequence. As expected, neither mutated nucleotide pair (T3-A18′ and C7-G14′) is involved in specific protein–DNA interaction, implying that our SghR–dsDNA structure represents an authentic operator DNA-bound complex. Sequence alignment analysis revealed that those residues, such as Thr35, Ser37, Lys41, Thr50, and Tyr63, which are involved in phosphate backbone interactions by their side chains, are conserved among the LacI family (Fig. S3). Consistently, their counterparts in lac repressor and PurR protein sequences also have been shown to participate in DNA binding (20–23). However, residues that contact DNA bases through their side chains, such as Thr34, Asp45, and Arg73, are varied in these sequences (Fig. S3). Taken together, our results suggest that the interactions between SghR and DNA are in a sequence-specific manner despite sharing a similar DNA-binding mode with other LacI family proteins.

### Table 1

Summary of crystallographic data and refinement statistics

| Parameter        | Value(s) for Apo | Value(s) for SghR-sucrose | Value(s) for SghR-DNA |
|------------------|------------------|---------------------------|----------------------|
| Data collection  |                  |                           |                      |
| Space group      | P2;2;21          | P2;2;21                   | P2;2;21              |
| Unit cell dimensions (Å) | 35.54 | 35.7 | 107.90 |
| a                | 119.45           | 120.2                     | 218.86               |
| b                | 121.94           | 122.2                     | 284.14               |
| c                | 90.0             | 90.0                      | 90.0                 |
| Resolution (Å)   | 50–2.1 (2.21–2.1) | 50–2.1 (2.23–2.1)         | 50–2.1 (2.25–2.2)    |
| Rmerge (%)       | 8 (32.7)         | 9.2 (40.7)                | 9.3 (116.4)          |
| Rfree (%)        | 9.6 (3.7)        | 14.7 (2.4)                | 12.9 (13.3)          |
| Completeness (%) | 99.7 (99.9)      | 97.7 (98.5)               | 99.8 (99.9)          |
| Redundancy       | 3.5 (3.6)        | 6.3 (5.9)                 | 5.2 (5.2)            |
| Refinement       |                  |                           |                      |
| Resolution (Å)   | 50–2.1           | 50–2.1                    | 50–2.1               |
| No. of reflections | 31,106          | 58,464                    | 111,573              |
| Rmerge (%)       | 20.6/25.9        | 22.5/26.9                 | 24.0/28.8            |
| Avg B factor (Å²) | 28.9            | 23.6                      | 66.9                 |
| Protein          | 34.9             | 28.1                      | 71.8                 |
| Sucrose          | 0.005            | 0.010                     | 0.010                |
| Water            | 0.86             | 1.27                      | 1.408                |
| DNA bases r.m.s.d. | 96.3            | 96.1                      | 96.1                 |
| Bond length (Å)  | 3.7              | 3.9                       | 3.1                  |
| Bond angle (°)   | 0.8              | 0.8                       | 0.8                  |
| Ramachandran plot | 0.8             | 0.8                       | 0.8                  |

*Numbers in parentheses refer to outer resolution shell.

**Structure of sucrose-bound SghR**

Given our functional characterization of sucrose as a ligand for SghR (7), we carried out microscale thermophoresis experiments and obtained the binding affinity of sucrose to SghR protein ($K_d = 42.7 ± 3.3 \mu M$) (Fig. 2C). Next, we cocrystallized SghR with sucrose and determined its complex structure by a molecule replacement method using apo SghR as a search model, and the crystallographic statistics are listed in Table 1. As expected, sucrose, clearly indicated by the electron density map, binds to the cleft created by the two C-terminal subdomains (Fig. 5, A and B) in a position similar to that observed for
Figure 4. Crystal structure of SghR in complex with its operator DNA. A, cartoon representation of the structure of SghR-DNA in two views. The chains A and B of SghR protein are colored cyan and light magenta, whereas the dsDNA duplex is colored gold (a) and light green (b), respectively (same as below). The first four N-terminal helices are labeled. B, comparison of the natural operator DNA sequence with the modified symmetric DNA sequence. The upper panel shows the natural operator DNA sequence determined by the DNase I footprinting assay, which is 24 bp long, containing the pseudopalindromic region indicated by the blue inverted arrows. The lower panel shows the symmetric DNA sequence used for the crystallization of the SghR-DNA complex, which is a symmetric (16 bp long) DNA with a two-nucleotide 5'-overhang on each strand. The palindrome region is also indicated by the inverted blue arrows, and the nucleotides that were mutated to generate a perfect palindrome are colored blue.

C, structural illustration of residues on the N-terminal DNA-binding domain of monomer A interacting with one-half of the DNA duplex. The interacting residues are labeled and shown in stick model with nitrogen and oxygen atoms displayed in blue and red, respectively. The electron densities (Fo-Fc omit map contours to 3σ) of these residues are shown as a blue mesh as well. D, illustration of the interactions between SghR and the DNA duplex. The DNA bases are labeled (guanine = G; adenine = A; cytosine = C; thymine = T), and the pentose sugars are numbered from the 5’ to 3’ end. The black dotted lines represent hydrogen bonds between the respective SghR residues and DNA phosphate backbone or DNA bases. The residues of SghR involved in the interactions are colored cyan for chain A and light magenta for chain B. The black dot in the middle of nucleotides 10 and 11 represents the 2-fold symmetry of the DNA in the crystal structure.
other effectors, such as IPTG (Fig. 5D) and guanine (24, 25). The orientation of sucrose, surrounded by polar and aromatic residues, was nearly parallel to the strands β5 and β11 (Fig. 5A). The fructose ring is flanked by the phenyl group of Phe210 in the C-terminal subdomain and the imidazole group of His164 in the N-terminal subdomain, thereby establishing a sandwich arrangement to stabilize the orientation of sucrose (Fig. 5C). The sucrose is recognized by several specific hydrogen bonds formed between its hydroxyl groups in the sugar ring and the residues in both the N- and the C-terminal subdomains, including Glu89, Ser142, Lys143, His164, Asp177, Asn180, Lys292, and Glu310 (Fig. 5C). The binding site for the glucosyl and fructosyl moiety in sucrose corresponds to −1 and +1 subsites, respectively. Interestingly, the side chains of Asn180 and Ser142, as well as the main-chain N of Lys143, form hydrogen bonds with the two hydroxymethyl groups of fructosyl moiety, appearing to be crucial for defining the +1 subsite. The binding of the +1 subsite was further stabilized by the hydrogen bonds between its hydroxyl groups and the main-chain N of Lys143 and Ile144, the main-chain O of His164, and the side chain of Asp177. Simultaneously, water-mediated interactions with the sugar hydroxymethyl group, as well as the side chains of Glu89 and Lys292, contacting hydroxyl groups in glucosyl moiety, contribute to the recognition of the −1 subsite. At the kinked link of sucrose (which is located at the half gateway of the pocket), residue Glu310 established bilateral hydrogen-bonding interactions with both moieties, thereby determining the two subsites. Remarkably, complicated hydrogen-bonding networks mediated by several water molecules, both sugar moieties and residues Thr94, Ser142, Arg166, Ser294, and Phe210, cross-link the C-terminal subdomains and define their relative orientation to each other (Fig. 5C). In contrast to the LacI repressor and its gratuitous inducer IPTG binding pocket, depicted in Fig. 5D (25), there are also mainly hydrogen bonds formed between sugar ring hydroxyl groups and LacI to bind the IPTG in the middle of N-terminal and C-terminal subdomains. Namely, the residues Ser69, Asp149, Arg197, Asn246, and Asp274 form direct hy-

Figure 5. Crystal structure of SghR in complex with sucrose and inducer binding pocket comparison between LacI and SghR. A. cartoon representation of the structure of SghR–sucrose complex. SghR is colored light orange. Sucrose is shown as a stick, with carbon and oxygen atoms displayed as cyan and red, respectively. The color codes are the same below. B, unbiased difference Fourier electron density map (Fo-Fc) of sucrose (Fo-Fc map contours to 3σ). The interacting residue E310 of SghR and one water molecule are shown as well. C, detailed interaction of SghR with sucrose. The residues involved in sucrose binding are shown as sticks. The water molecules involved in the hydrogen-bonding network are shown as pink spheres. The residues involved in sucrose binding are labeled and shown as sticks, with electron densities (2Fo-Fc map contours to 1σ) indicated as blue mesh. The hydrogen bonds formed between residues/water molecules with sucrose are shown as dashed lines. D, detailed interactions of LacI with IPTG (PDB code 2P9H). For comparison, the residues involved in interacting with IPTG were shown as sticks. The hydrogen bonds between IPTG and LacI or the water-mediated hydrogen bond are shown as dash lines.
Structural insights into the regulation mechanism by SghR

SghA from Agrobacterium tumefaciens is a virulence factor that confers its host plant Agrobacterium-mediated transformation (1). Upon infection, SghA promotes its own expression through inducing the sucrose gene (sgh) expression (2,3). Positive regulation of this gene involves the transcription factor SghR (4,5). Although SghR has lost its DNA binding domain, it binds to a sucrose-responsive promoter to induce its cognate gene expression (6,7). In this work, we describe the structural basis of how an effector molecule alters the affinity of SghR for its cognate DNA.

**Insight into allosteric mechanism by structural comparison of SghR in apo, sucrose- and DNA-bound forms**

The structural basis of how an effector molecule alters the affinity of the transcription factor for its cognate DNA is of significant importance for a detailed understanding of the molecular mechanism of gene regulation. Here, the structures of SghR in three states (apo, sucrose–, and DNA–SghR) allow us to exploit the structural basis of how sucrose modulates the affinity of SghR for its cognate DNA.

Consistent with the notion that the hinge helix of LacI family protein is unfolded and the N-terminal DBD moves freely with respect to the C-terminal core in the absence of their cognate DNA (26–28), the N-terminal DBD in both apo and sucrose-bound SghR are disordered. Nevertheless, comparison of their C-terminal domains reveals virtually identical structures, with r.m.s.d value of 0.15 Å for the aligned Ca atoms. In addition, the conformation of residues Glu89, Ser142, Lys292, and Glu310, which are involved in hydrogen-bonding networks for sucrose recognition in SghR–sucrose, are almost identical to that in apo SghR (Fig. S4). This is in line with LacI from E. coli. The inducer (IPTG) binding structure (PDB entry 2P9H) is extremely similar to apo LacI (PDB entry 1LBI). The structure comparison of LacI–IPTG with apo LacI (PDB entry 1LBI) gave an r.m.s.d. value of 0.5 Å (Fig. S4B). In addition, there is little conformation change for the residues that are involved in IPTG binding. A detailed analysis revealed that many water molecules in apo SghR structure were observed in the sucrose-binding pocket (Fig. S4A). Notably, some of these water molecules occupy the positions of hydroxyl groups of the sugar ring in the SghR–sucrose structure, thereby forming extensive hydrogen-bonding networks with the residues of SghR (Fig. S4A). These interactions are similar to that observed between sucrose and SghR in the SghR–sucrose complex, which perhaps explains the similarity of sucrose-binding pocket in these two structures. Moreover, weak binding of sucrose to SghR was demonstrated by microscale thermophoresis assay (Fig. 2C). Consistent with this weak binding, relatively high sucrose concentration (1000-fold molar excess of SghR) is required to reduce the SghR–DNA binding (Fig. 2D). Nevertheless, we could postulate that the side chains of the residues involving sucrose binding have undergone certain conformational changes during sucrose diffusing into the binding pocket. Such a conformational change of these local residues induced upon sucrose binding is so transient that it cannot be observed in the present structure.

Similar to LacI and PurR (18, 29) (Fig. S7), the C-terminal domain of SghR is organized into two structurally similar subdomains: an N-subdomain containing residues 79–178 and 312–339 and a C-subdomain covering residues 180–309 and 340–345. Superposition of the individual N-subdomains was based on the alignment matrix of individual C-subdomains from both SghR–sucrose and SghR–DNA complexes using the program ALIGN (30). Comparison of both N- and C-subdomains in sucrose-bound dimeric SghR with that in DNA-bound dimeric SghR showed r.m.s.d. values of 0.79 Å and 0.66 Å, respectively, for Ca atoms. In contrast, comparison of the N-subdomains of these two structures based on the alignment of the C-subdomains gave an r.m.s.d. of 1.31 Å for Ca atoms, suggesting that the relative orientation of the N- and C-subdomains in SghR is changed upon sucrose binding. Furthermore, detailed plotting of the mean Ca deviation for each residue within the N-subdomain revealed that those residues with relatively larger variation are mapped to two regions: residues 145–174 and 321–337 (Fig. 6, A and B). Note that these regions are closely linked to the residues involved in sucrose binding, such as residues Ser142, Lys143, Ile144, Ile145, Asp177, Asn180, Lys292, and Glu310, which are involved in hydrogen-bonding networks with the residues of SghR (Fig. S4). This is in line with LacI from E. coli. The inducer (IPTG) binding structure (PDB entry 2P9H) is extremely similar to apo LacI (PDB entry 1LBI). The structure comparison of LacI–IPTG with apo LacI (PDB entry 1LBI) gave an r.m.s.d. value of 0.5 Å (Fig. S4B). In addition, there is little conformation change for the residues that are involved in IPTG binding. A detailed analysis revealed that many water molecules in apo SghR structure were observed in the sucrose-binding pocket (Fig. S4A). Notably, some of these water molecules occupy the positions of hydroxyl groups of the sugar ring in the SghR–sucrose structure, thereby forming extensive hydrogen-bonding networks with the residues of SghR (Fig. S4A). These interactions are similar to that observed between sucrose and SghR in the SghR–sucrose complex, which perhaps explains the similarity of sucrose-binding pocket in these two structures. Moreover, weak binding of sucrose to SghR was demonstrated by microscale thermophoresis assay (Fig. 2C). Consistent with this weak binding, relatively high sucrose concentration (1000-fold molar excess of SghR) is required to reduce the SghR–DNA binding (Fig. 2D). Nevertheless, we could postulate that the side chains of the residues involving sucrose binding have undergone certain conformational changes during sucrose diffusing into the binding pocket. Such a conformational change of these local residues induced upon sucrose binding is so transient that it cannot be observed in the present structure.

**Discussion**

Transcriptional regulation of specific genes through signaling molecules plays a central role for bacteria to adapt to environmental change. Given that the ubiquitous soil agrobacteria infect a variety of plant species, it is of great interest to investigate its gene regulation during the process of infection. We have identified a novel virulence factor, SghA, from A.
that temporally regulates plant tumor development by controlled release of host-conjugated defense signal, in response to sucrose accumulated in plant wounding sites at the late stage of infection (7). In this paper, we reported that the expression of SghA is negatively controlled by a putative LacI family transcription factor, SghR, thereby involving tumor growth postinfection. How SghR specifically recognizes the promoter of \textit{sghA} to repress gene transcription and how SghR specifically senses the signaling molecule sucrose to make responses are central to elucidate the molecular mechanism of this regulation process. To address these issues, our structural findings unraveled the detailed molecular mechanism of SghR controlling \textit{Agrobacterium} infection by cross-talking to plants.

Our \textit{in vivo} tumorigenesis assay together with RT-PCR analyses revealed that both SghA and SghR are involved in plant tumor growth in a manner that the mRNA level of \textit{sghA} is negatively controlled by SghR (Fig. 1, \textit{A} and \textit{B}). Given that SghA is involved in the enzymatic releases of the plant defense signal SA from its conjugated form, SA 2-O-\textit{\beta}-D-glucoside, after successful infection, this pair of novel virulence factors, SghA/SghR, plays a critical role in tumor development in agrobacterium-infected plants. The DNA-binding site of SghR on \textit{sghA} promoter region covers \(2^{24}\) nt to \(2^{1}\) nt upstream of the \textit{sghA} transcription start site (defined as \(2^{1}\)), based on DNase I footprinting assay (Fig. 1\textit{C}). Note that the binding site presents the common feature of inverted repeat sequence (partial palindromic consensus sequences), as observed for other LacI family members (2, 3).

**Figure 6. Structural comparison of SghR-sucrose and SghR-DNA complexes.**\(A\), superposition of the individual N-subdomains based on the alignment matrix of individual C-subdomains from both SghR-sucrose and SghR-DNA complexes. The N-subdomains of SghR-sucrose are colored light blue, whereas the N-subdomains of SghR-DNA complex are colored cyan for chain A and light magenta for chain B. The dashed-lined boxes demonstrate the regions (residues 145–174 and 321–337, as indicated by the double-ended arrows in panel \(B\)) exhibiting relatively large conformational changes. The color codes for chain A and chain B are the same below. \(B\), the \(\sigma\) deviation for each residue on the N-subdomains based on the alignment methods in panel \(A\) for SghR-sucrose and SghR-DNA complexes. The two double-ended arrows indicate the relatively large conformational change regions between the two complex structures. \(C\) and \(D\), conformational changes of the key residues involved in domain interactions based on the comparison of SghR-DNA complex with SghR-sucrose complex.

Comparison of our apo SghR structure with apo \textit{lac} repressor Atu1522 (PDB entry 3GV0) from \textit{A. tumefaciens} C58 gave an r.m.s.d. value of 0.62 for C\(\alpha\) atom, demonstrating high structural similarity (Fig. S6\textit{A}). However, for Atu1522, no other information is available. In addition, structural comparison of apo SghR with Lacl from \textit{E. coli} (PDB entry 1LBI, excluding the C-terminal residues 325–357, involved in tetramerization) and PurR from \textit{E. coli} (purine repressor, PDB entry 1DBQ) gave...
Structural insights into the regulation mechanism by SghR

r.m.s.d. values of 2.39 and 3.29, respectively, for the Cα atom (Fig. S6, B and C). The relatively larger conformational difference might be explained by their roles in sensing different signal molecules during transcription regulation. Nevertheless, all four structures show a similar overall topology for their C-terminal core domains (18, 19).

Similar to our SghR structures in apo (Fig. 3A) and sucrose-bound form (Fig. 5A), the N-terminal DBDs in apo Atu1522, LacI, and PurR structures are all disordered (Fig. S6), which reinforces the high flexibility of the N-terminal DBD of the LacI family in the absence of their cognate DNAs. As expected, DNA binding would stabilize this flexible domain, leading to an ordered model in our SghR-DNA complex structure (Fig. 4A). The SghR-DNA complex structure provides the detailed interactions of how SghR specifically recognizes the operator DNA in a sequence-dependent manner, as well as rationalizes the requirement of the partial palindrome feature for its cognate dsDNA binding (Fig. 4, C and D). Moreover, similarities in protein topology, DNA deformability, and protein-DNA binding mode could also be observed for other LacI family members in complex with their cognate DNAs, such as LacI-DNA (PDB entry 1EFA) and PurR-DNA corepressors (PDB entry 1PNT and 1WET) (17, 18, 29), suggesting a universal dsDNA-binding mode applies to all LacI family members.

LacI/GalR family proteins play a vital role in regulating metabolic pathways through sensing and responding to a large variety of internal and external signal molecules, such as lactose, fructose, galactose, and many others, to adapt to constant environmental changes (2, 31). Detailed studies of how LacI proteins respond to these effectors are of considerable importance to elucidate their subtle specificity as well as to design anti-infection agents. Once infection is established, sucrose could lower the binding affinity of SghR for its cognate DNA, thereby facilitating the transcription of Agrobacterium infection. At the initial stage of infection, SghR physically binds to the promoter region of sghA and tightly represses its transcription to avoid the release of SA, which consequently triggers host plant defense. Once Agrobacterium infection is established, sucrose accumulated with high concentration at the wounding sites during healing and served as an environmental stimulus to alleviate the repression of SghA by SghR, thereby facilitating
the expression of SghA and leading to increased defense signal in plants. As a pathogen, *Agrobacterium* causes crown gall disease in a wide variety of crops. Given the role of SghR in regulating the virulence factor SghA in tumorigenesis in response to the signal effector sucrose originated from the host plant, our results offer a structure-based inducer analog design that has potential applications for the agricultural industry.

**Experimental procedures**

**Tumorigenesis assay**

The deletion mutants of *A. tumefaciens* A6 with genes *sghR* and *sghA* both in-frame deleted were generated according to the protocol described previously (36). The virulence assays were performed on carrot disks. Briefly, fresh carrot roots obtained from the supermarket were surface sterilized by scrubbing under running water. After immersing in 2% Clorox for 5 min, the carrot roots were then cut into ~1-cm-thick slices and immersed in 2% Clorox for 30 min. After rinsing thrice in autoclaved distilled water for 15 min, the discs were placed apical side up on agar plates. *A. tumefaciens* strains freshly grown in BM medium were resuspended in sterile DPBS buffer with an optical density at 600 nm of 1.0, and for each disk, 10-μl agrobacteria suspensions were applied to the top surface of the carrot disk. They were then sealed with parafilm, and the Petri plates with the inoculated disks were incubated in a growth chamber for tumor development. After a 4-week incubation, results were recorded by photography and then the fresh weight of tumor was measured.

**RT-PCR**

Total RNAs were extracted from bacterial cells cultivated in BM minimal medium and VIB medium using an RNaseasy® mini kit (Qiagen). To minimize data variation, total RNAs were prepared from three independent repeats and pooled together for RT-PCR analysis. RT-PCR was conducted by following the one-step strategy (Qiagen). An aliquot of 0.2 μg total RNAs was used as the template for RT-PCR to amplify a portion of the target genes. For *sghA*, the PCR primers are 5’-GGCGACCGGTGTGATG-3’ and 5’-CGGGGCTTGTT-TGGTGGG-3’. In addition, a fragment of 16s rRNA was also amplified in each RT-PCR reaction as an internal control using primers of 5’-TGACGAGTGGCGGACGGTTG-3’ and 5’-ATGCGATTTCCCAGGTGAC-3’.

**Gene cloning, protein expression, and purification**

Generally, the gene encoding the full-length target protein SghR (residues 1–350) from *A. tumefaciens* was amplified by PCR and cloned into pET-14b vector. The resultant plasmid was verified by DNA sequencing and then transformed into the *E. coli* BL21 CodonPlus- (DE3) RIL cell strain. The culture, expression, and purification of full-length SghR are the same as methods reported previously (16). In general, two sequential purification steps (nickel affinity and gel filtration chromatography) were applied to produce homogeneous SghR protein. The purified protein was in the final buffer of 50 mM HEPES, pH 7.0, 50 mM NaCl, and 2 mM tris(2-carboxyethyl)-phosphine and concentrated to 6.4 mg/ml (the protein concentration was determined by NanoDrop™ 2000/2000c spectrophotometers, applying SghR protein molecular weight [including extra affinity tag] and its extinction coefficients during the measurement, which is the same as that below for determining protein concentrations of all the other different SghR constructs), and then frozen into liquid nitrogen and stored at ~−80°C for later experiments.

For cocrystallizing SghR–DNA complex, the gene encoding the SghR fragment spanning from residue 18 to 350 was cloned into the NdeI and XhoI restriction sites of a modified pET-26b vector. The recombinant SghR truncation with a C-terminal 6-×His tag (LEHHHHHHH) was expressed in the *E. coli* BL21 CodonPlus- (DE3) RIL cell strain. To purify this truncated SghR protein, procedures similar those reported previously (16) were used, except for the introduction of a high salt concentration (1 M NaCl) into the lysis buffer and the addition of a heparin chromatography step between the nickel affinity chromatography and gel filtration chromatography steps. This additional step was revealed to remove the junk nucleic acids that are non-specifically bound to the target protein. Finally, the purified protein was dialyzed against buffer containing 10 mM Tris pH 8.5, 100 mM NaCl, 0.1 mM EDTA, and 4 mM β-mercaptoethanol (β-ME) and concentrated to 1.67 mg/ml for preparing protein–DNA complex. The DNA oligonucleotide with the sequence 5’-TATCTGCAACGTGGCAGA-3’, used for cocrystallization, was purchased from Sigma Aldrich with HPLC purification grade.

**DNase I footprinting assay**

The *sghA* promoter region was amplified by PCR using the FAM-labeled forward primers 5’- (6-FAM)-CAGATGCAA-GACTTTCACCAC-3’ and reverse primers 5’-CGTACGAGTGGCGGACGGTTG-3’ from the genomic DNA of *A. tumefaciens* A6. The experiment was performed referring to previous published methods (37, 38) with some modifications. Briefly, 500 ng of FAM-labeled DNA was diluted with 20 μl of gel shift buffer (20 mM NaH2PO4 pH 8.5, 200 mM NaCl, and 5 mM β-ME), and then 3 μl of purified SghR (18.6 mg/ml) and 16 μl of water were added. After incubation at room temperature for 20 min, 10 U of DNase I (Roche) was added, followed by further incubation for 5 min. Subsequently, the reaction was stopped by heating the sample at 75°C for 10 min. The control experiment was performed in parallel by using the labeled probe in the absence of SghR protein. The digested DNA products were purified with the PCR purification kit (Qiagen) and eluted in 20 μl of water for further analysis. Dideoxynucleotide-based sequencing was performed using Thermo Sequence Dye Primer manual cycle sequencing kit (USB, Inc., Cleveland) by following the manufacturer’s instructions, and the samples were analyzed with the Applied Biosystems 3730 DNA Analyzer. To confirm the forward FAM-labeled primer sequencing result, the FAM-labeled reverse primer sequencing was also carried out in parallel. The two primers used for sequencing are 5’-CAGATGCAA-GACTTTCACCAC-3’ and 5’- (6-FAM)-
Structural insights into the regulation mechanism by SghR

CCTTTCGTCA TTCGACGT-3'. The procedures are the same as described for forward primer sequencing.

EMS A

Based on the results of our DNase I footprinting assay, the DNA oligonucleotides (forward, 5’-GCTGAAACGTGCTG-AGATTTTGCGT-3'; reverse, 5’-ACGCAAAAATCTGACAG- TTACGC-3') were synthesized to test binding with SghR. Equimolar complementary strands were mixed in the annealing buffer (20 mM NaH2PO4, pH 8.5, 200 mM NaCl, 5 mM β-ME), with a final concentration of 0.5 mM. Subsequently, the DNA solution was heated at 95°C for 3 min, followed by gradually cooling down to room temperature.

For the in vitro DNA-binding assay, full-length SghR was cloned into a modified pET-26b vector between restriction enzyme sites Ndel and HindIII. The final construct containing a C-terminal 6×His tag was prepared according to the same procedures used for purifying the SghR truncation (18–350 aa). The purified SghR (0.5 μg; if the molar ratio SghR to DNA was 2:1, then 1 μg SghR protein was added) and dsDNA (amount added based on the ratio with SghR) were mixed at varied molar ratios, as indicated in Fig. 1E, and incubated at room temperature for 30 min. The samples were then resolved on 10% native PAGE, followed by gel staining with ethidium bromide and Coomassie brilliant blue. The staining results were recorded by Gel Doc™ XR (Bio-Rad).

ITC

ITC measurements were carried out at 15°C using a Micro ITC200 (MicroCal Inc., Northampton, MA, USA). SghR was loaded into the sample cell at a concentration of 15 μM (dimer), and the dsDNA (the same sequence as that in the above EMSA experiment from the sghA promoter) was loaded into the syringe at a concentration of 200 μM. To investigate the effect of sucrose on SghR-DNA binding, the SghR protein at a concentration of 18.7 μM was premixed with 1000-fold sucrose for 1 h at room temperature before loading to the sample cell. As a control, Ppa0305 promoter DNA from P. aeruginosa, used to test its binding with SghR, was carried out in parallel. To correct the dilution and mixing effect, blank titrations were conducted by titrating DNA buffer into either protein or protein-sucrose solution in the sample cell, which was subtracted from the raw data. All data were analyzed with MicroCal Origin 7.0 (Microcal Software, Inc., MA, USA).

MST

The microscale thermophoresis (MST) experiment was performed using a Monolith NT.115 instrument (NanoTemper Technologies, Germany) at 24°C. The PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4) supplemented with 0.05% Tween-20 was used as the MST-binding buffer. Premium-coated capillaries were used. SghR protein was purified to homogeneity as described above. The purified SghR was labeled with the Monolith protein labeling kit RED-NHS 2nd generation (Amine Reactive). The labeled SghR was diluted to a concentration of 250 nM, mixed with an equal volume of a serial dilution series of sucrose, and then incubated at room temperature for 20 min before loading into MST capillaries. A single MST experiment was performed using 60% LED power and 20% MST power with a wait time of 5 s, laser-on time of 30 s, and a back-diffusion time of 5 s. MST data were analyzed in GraphPad Prism, and the data were fitted with the Hill equation. The mean 50% effective concentration (or Kd) values were calculated with standard error (S.E.). Each experiment was repeated at least three times.

Crystallization, data collection, and structure determination

The crystallization procedures for apo SghR are the same as those reported in our previous paper (16). In general, for each condition, three ratios of protein and reservoir solution (0.2:0.1 nl, 0.15:0.15 nl, and 0.1:0.2 nl) were screened by the sitting-drop vapor-diffusion method using a Phoenix robot (Art Robbins Instruments) at 20°C. Screening kits from Hampton Research, including Crystal Screen, Crystal Screen 2, Index, PEG/Ion, PEGRx, and SaltRx, were used. The apo SghR crystal hits were observed under conditions consisting of 0.05 M HEPES-Na, pH 7.0, 20% PEG 3350. The crystallization condition was optimized by varying the PEG concentration (17–22%) and increasing the drop size to a 2-μl protein solution and 1-μl reservoir solution using the hanging-drop method at 20°C. The crystals were cryoprotected with a cryoprotectant consisting of 35% PEG 3350 in the reservoir and were then flash-cooled in liquid nitrogen. To cocrystallize SghR-sucrose complex, sucrose with a final concentration of 60 mM was mixed with SghR (6 mg/ml) and incubated for 30 min at room temperature before setting up the trays. After crystallization screening, crystals were observed under the same condition as that of apo-SghR. A similar optimization strategy was also applied to obtain diffraction-quality crystals.

To cocrystallize the SghR–DNA complex, dsDNA was prepared using the above-mentioned annealing method in buffer containing 10 mM Tris-HCl, pH 8.5, 100 mM NaCl, and 1 mM EDTA. The annealed DNA duplexes were directly added to the concentrated truncated SghR (residues 18–350) with a molar ratio of 1:1.1 (DNA:protein). The mixture was incubated on ice for at least 3 h before setting up crystallization trays. The final protein concentration in the mixed solution was ~5.5 mg/ml. The initial screening (sitting-drop method at 20°C) and crystal optimization (hanging-drop method at 20°C) for this complex were performed similarly to those for apo-SghR. Well-diffracting SghR-DNA complex crystals were obtained in condition consisting of 19–22% PEG 3350 and 0.2 mM ammonium citrate tribasic, pH 7.0. Crystals were then flash-frozen in the above reservoir solution supplemented with 30% glycerol.

Diffraction data were collected at 100K on beamline X06SA of the Swiss light source (SLS) or I04 (Diamond). The data were processed using XDS (39). All three structures were determined by molecular replacement using the program Phaser (40). The apo-SghR structure was solved using the structure of a homologous protein (PDB entry 3GV0) as a search model. It showed that there are two copies of SghR in an asymmetric unit. Subsequently, automatic model building was employed using ARP/wARP (41). The model was further improved by manual model building with Coot (42) and refined by applying
TLS refinement combing Phenix (43) and REFMAC5 (44). The apo SghR structure was then used as a search model to obtain the initial phase of the SghR–sucrose complex. The procedure for manual model building and refinement was the same as that for apo SghR. As for the SghR–DNA complex, initial molecular replacement trials with either monomer or dimer of apo SghR (residues 74–349) did not yield unambiguous solutions. However, there were 1233 partial solutions, and the best one showed four monomers in the space group P2₁2₁2₁. Using these four monomers (forming two physiological dimers) as a fixed solution, a second round of molecular replacement was carried out using the dsDNA-bound N-terminal domain of the lac repressor (PDB entry 1EFA). The N-terminal domain of lac repressor is equivalent to the region 1–73 of SghR. This allowed the identification of the orientation and positioning of the dsDNA-bound N-terminal region (residues 1–73) of SghR with respect to the C-terminal domain (residues 74–349). Thus, a dimer of SghR was correctly placed. Further molecular replacement with the newly obtained dsDNA-bound SghR dimer allowed the identification of all six copies in one asymmetric unit. The procedure for manual model building and refinement was the same as that of apo SghR. The final models of the three structures determined in this study were checked by PROCHECK (15). The crystallographic data and refinement statistics are listed in Table 1, and all figures were generated with PyMol (Delano Scientific). The area of protein interface was calculated using the PISA program (45).

**Data availability**

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 7CDV, 7CDX, and 7CE1. All remaining data are included within the paper or are in the supporting information.

**Acknowledgments**—We thank S. Chew and X. Zhang for their technical assistance in protein purification and crystallization.

**Author contributions**—F. Y., C. W., Q. F., X.-F. Y., S. R. B., H. S., L.-H. Z., and Y.-G. G. conceptualization; F. Y., C. W., Q. F., X.-F. Y., S. R. B., A. C., and M. W. data curation; F. Y., C. W., Q. F., X.-F. Y., A. C., and M. W. software; F. Y., C. W., Q. F., X.-F. Y., S. R. B., and Y.-G. G. formal analysis; F. Y., H. S., L.-H. Z., and Y.-G. G. supervision; F. Y. and Y.-G. G. validation; F. Y., C. W., Q. F., X.-F. Y., and S. R. B. visualization; F. Y., C. W., Q. F., X.-F. Y., S. R. B., A. C., and M. W. methodology; F. Y., C. W., Q. F., X.-F. Y., and Y.-G. G. writing–original draft; F. Y., C. W., X.-F. Y., H. S., L.-H. Z., and Y.-G. G. writing–review and editing; Q. F., X.-F. Y., H. S., L.-H. Z., and Y.-G. G. investigation; H. S., L.-H. Z., and Y.-G. G. project administration; L.-H. Z. and Y.-G. G. funding acquisition.

**Funding and additional information**—This work was partially supported by the Singapore National Research Foundation NRF-RF2009-RF001-267 (Y.-G. G.), the NIMBELS grant NIM/01/2018 (to Y.-G. G.), and the Medical Research Council, the Agency for Science, Technology and Research in Singapore and National Program on Key Basic Research Project (973 Program) of China 2015CB150600 (to L.-H. Z.).

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: DBD, DNA-binding domain; SA, salicylic acid; ITC, isothermal titration calorimetry; r.m.s.d., root mean square deviation; β-ME, β-mercaptoethanol; MST, microscale thermophoresis.

**References**

1. Ishihama, A. (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol. Rev.* **34**, 628–645 CrossRef Medline
2. Swint-Kruse, L., and Matthews, K. S. (2009) Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol.* **12**, 129–137 CrossRef Medline
3. Weickert, M. J., and Adhya, S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* **267**, 15869–15874 Medline
4. Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 CrossRef
5. Wilson, C. J., Zhan, H., Swint-Kruse, L., and Matthews, K. S. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol. Life Sci.* **64**, 3–16 CrossRef Medline
6. Mukhopadhyay, A., Gao, R., and Lynn, D. G. (2004) Integrating input from multiple signals: the VirA/VirG two-component system of Agrobacterium tumefaciens. *ChemBioChem* **5**, 1535–4227 CrossRef
7. Wang, C., Ye, F., Chang, C., Liu, X., Wang, J., Wang, J., Yan, X. F., Fu, Q., Zhou, J., Chen, S., Gao, Y. G., and Zhang, L. H. (2019) Agrobacteria reprogram virulence gene expression by controlled release of host-conjugated signals. *Proc. Natl. Acad. Sci. U S A* **116**, 22331–22340 CrossRef Medline
8. Koch, K. (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr. Opin. Plant Biol.* **7**, 235–246 CrossRef Medline
9. Ibraheem, O., Hove, R. M., and Bradley, G. (2008) Sucrose assimilation and the role of sucrose transporters in plant water response. *African J. Biotech.* **7**, 4850–4855
10. Wachter, R., Langhans, M., Aloni, R., Gotz, S., Weilmunster, A., Koops, A., Tenguiua, L., Mistrik, I., Pavlovkin, J., Rascher, U., Schwalm, K., Koch, K. E., and Ullrich, C. I. (2003) Vascularization, high-volume solution flow, and localized roles for enzymes of sucrose metabolism during tumorigenesis by Agrobacterium tumefaciens. *Plant Physiol.* **133**, 1024–1037 CrossRef Medline
11. Davis, J. M., Gordon, P. M., and Smit, B. A. (1991) Allostery in the LacI/GalR homologous to Gal and Lac repressors. *J. Biol. Chem.* **266**, 1139–1145 CrossRef Medline
12. Ishihama, A. (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol. Rev.* **34**, 628–645 CrossRef Medline
13. Haudecoeur, E., Planamente, S., Cirou, A., Tannieres, M., Shelp, B. J., Morera, S., and Faure, D. (2009) Proline antagonizes GABA-induced quenching of quorum-sensing in Agrobacterium tumefaciens. *Proc. Natl. Acad. Sci. U S A* **106**, 14587–14592 CrossRef Medline
14. Chevrot, R., Rosen, R., Haudecoeur, E., Cirou, A., Shelp, B. I., Ron, E., and Faure, D. (2006) GABA controls the level of quorum-sensing signal in Agrobacterium tumefaciens. *Proc. Natl. Acad. Sci. U S A* **103**, 7460–7464 CrossRef Medline
15. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 9
16. Ye, F., Wang, C., Fu, Q., Zhang, L., and Gao, Y. G. (2015) Cloning, expression, purification and crystallization of a pair of novel virulence factors, SghA and SghR, from Agrobacterium tumefaciens. *Acta Crystallogr. Sect. F* **71**, 1139–1145 CrossRef
Structural insights into the regulation mechanism by SghR

17. Bell, C. E., and Lewis, M. (2001) Crystallographic analysis of Lac repressor bound to natural operator O1. J. Mol. Biol. 312, 921–926 CrossRef Medline

18. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science 271, 1247–1254 CrossRef Medline

19. Nagadai, A., Morikawa, S., Nakamura, H., Enari, M., Kobayashi, K., Yamamoto, H., Sampegi, G., Mizobuchi, K., Schumacher, M. A., and Brennan, R. G. (1995) Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. Structure 3, 1217–1224 CrossRef Medline

20. Bell, C. E., and Lewis, M. (2000) A closer view of the conformation of the Lac repressor bound to operator. Nat. Struct. Biol. 7, 209–214 CrossRef Medline

21. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305, 386–389 CrossRef Medline

22. Lewis, M. (2011) A tale of two repressors. J. Mol. Biol. 409, 14–27 CrossRef Medline

23. Sprock, C. A., Bonvin, A. M., Radha, P. K., Melacini, G., Boelens, R., and Kaptein, R. (1999) The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. Structure 7, 1483–1492 CrossRef Medline

24. Schumacher, M. A., Glasfeld, A., Zalkin, H., and Brennan, R. G. (1997) The X-ray structure of the PurR-guanine-purF operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity. J. Biol. Chem. 272, 22648–22653 CrossRef Medline

25. Daber, R., Stayrook, S., Rosenberg, A., and Lewis, M. (2007) Structural analysis of lac repressor bound to allosteric effectors. J. Mol. Biol. 370, 609–619 CrossRef Medline

26. Kalodinos, C. G., Folkers, G. E., Boelens, R., and Kaptein, R. (2001) Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. Proc. Natl. Acad. Sci. USA 98, 6039–6044 CrossRef Medline

27. Spolar, R. S., and Record, M. T., Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science 263, 777–784 CrossRef Medline

28. Sprock, C. A., Sliper, M., van Boom, J. H., Kaptein, R., and Boelens, R. (1996) Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. Nat. Struct. Biol. 3, 916–919 CrossRef Medline

29. Schumacher, M. A., Choi, K. Y., Zalkin, H., and Brennan, R. G. (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science 266, 763–770 CrossRef Medline

30. Cohen, G. H. (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. J. Appl. Cryst. 30, CrossRef

31. Jue, D. H., Matthews, B. W., and Huber, R. E. (2012) LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. Protein Sci. 21, 1792–1807 CrossRef Medline

32. Taraban, M., Zhan, H., Whitten, A. E., Langley, D. B., Matthews, K. S., Swint-Kruse, L., and Trewella, J. (2008) Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. J. Mol. Biol. 376, 466–481 CrossRef Medline

33. Matthews, K. S., and Nichols, J. C. (1998) Lactose repressor protein: functional properties and structure. Proc. Nucleic Acids Res. Mol. Biol. 58, 127–164 CrossRef Medline

34. Jobe, A., and Bourgeois, S. (1972) The natural inducer of the lac operon. J. Mol. Biol. 69, 397–408 CrossRef Medline

35. Wilson, C. J., Zhan, H., Swint-Kruse, L., and Matthews, K. S. (2007) Ligand interactions with lactose repressor protein and the repressor-operator complex: the effects of ionization and oligomerization on binding. Biophs. Chem. 126, 94–105 CrossRef Medline

36. Wang, C., Tang, D., Gao, Y. G., and Zhang, L. H. (2016) Succinic semialdehyde promotes prosurvival capability of Agrobacterium tumefaciens. J. Bacteriol. 198, 930–940 CrossRef Medline

37. Zianini, M., Tessanne, K., Merighi, M., Lagana, R., and Tabita, F. R. (2006) Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument. J. Biomol. Tech. 17, 103–113 Medline

38. Wang, C., Ye, F., Kumar, V., Gao, Y. G., and Zhang, L. H. (2014) BswR controls bacterial motility and biofilm formation in Pseudomonas aeruginosa through modulation of the small RNA rsmZ. Nucleic Acids Res. 42, 4563–4576 CrossRef Medline

39. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. 26, 795 CrossRef

40. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. 63, 32–41 CrossRef Medline

41. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. 3, 1171–1179 CrossRef Medline

42. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 CrossRef Medline

43. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. 68, 352–367 CrossRef Medline

44. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. D Biol. Crystallogr. 60, 2184–2195 CrossRef Medline

45. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 CrossRef Medline