Title

Retraction: Site-specific recombination of nitrogen-fixation genes in cyanobacteria by XisF–XisH–XisI complex: Structures and models, William C. Hwang, James W. Golden, Jaime Pascual, Dong Xu, Anton Cheltsov, Adam Godzik

Journal

Proteins, 86(2)

ISSN

0887-3585

Authors

Hwang, William C
Golden, James W
Pascual, Jaime
et al.

Publication Date

2018-02-01

DOI

10.1002/prot.24679

Published in final edited form as:

Proteins. 2018 February ; 86(2): 268. doi:10.1002/prot.24679.

Site-specific recombination of nitrogen-fixation genes in cyanobacteria by XisF-XisH-XisI complex: structures and models

William C. Hwang[1,2], **James W. Golden**[3], **Jaime Pascual**[4], **Dong Xu**[2], **Anton Cheltsov**[2], **Adam Godzik**[1,2,5]

[1]Joint Center for Structural Genomics, http://www.jcsg.org, USA

[2]Bioinformatics and Systems Biology Program, Sanford Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

[3]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

[4]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

[5]Center for Research in Biological Systems, University of California, San Diego, La Jolla, CA 92093–0446, USA

Abstract

Nitrogen fixation is an important process that converts atmospheric gaseous nitrogen, a form plants cannot utilize, into ammonia that can be easily assimilated. Large serine recombinase XisF (*fdxN* element site-specific recombinase), together with controlling factors XisH and XisI, plays a critical role in the expression of nitrogen fixation genes of certain *Anabaena* and *Nostoc* species of cyanobacteria. All three proteins are required to excise the *fdxN* DNA element from the chromosome in differentiating heterocysts for the expression of nitrogen fixation related genes. We report the first crystal structures of XisH and XisI proteins, both adopting novel protein folds. Based on the analysis of their sequences and structures, we propose that XisH and XisI proteins function as endonucleases and recombination directionality factors (RDFs), respectively.

Keywords

fdxN element; recombinase; endonuclease; excisionase; recombination directionality factor

Correspondence to: William C. Hwang; Adam Godzik.

Competing interests

None declared.

Authors' contributions

WH conceived the article and prepared the manuscript. JG wrote part of the introduction and conclusion. JG, JP, AG commented on the manuscript. DX built the XisF model structure with I-TASSER. JP and AC annotated the XisH and XisI structures on TOPSAN.

Supplementary Data

Supplementary Data are available at "Proteins: Structure, Function, and Bioinformatics" online. Supplementary Data I: XisH, XisI data collection and refinement statistics. Supplementary Data 2: Electrostatic potential surfaces; conservations of XisH-like and XisI-like proteins.

INTRODUCTION

Nitrogen, one of the constituent elements in amino acids and nucleotides, is vital to life. Although the atmosphere is replete with gaseous nitrogen (N_2), in this form it is, however, inaccessible to plants, animals, and most microbes. The covalent triple bond between nitrogen atoms in N_2 makes it chemically inert and recalcitrant to being incorporated into other biological compounds. Only in its reduced form as ammonia (NH_3), nitrites (NO_2), or nitrates (NO_3), can nitrogen be readily assimilated by plants and other organisms.

Biological nitrogen fixation (BNF), the enzymatic conversion of nitrogen to ammonia, is carried out mostly by specialized groups of bacteria. Some of them are free-living bacteria, while others are symbionts such as bacteria found in root nodules of leguminous plants. Some archaea are also capable of fixing nitrogen [1,2], but the overall contribution of archaea to the nitrogen cycle is difficult to gauge. Non-biological nitrogen fixation, including the industrial Haber-Bosch synthetic process and natural lightning combustion, both requiring high temperature and high pressure, are sources of about 30% of available ammonia.

Cyanobacteria are photosynthetic bacteria found in many, mostly aquatic, environments. They can be free living, but they are often found in microbial assemblages, or in symbiotic associations with multicellular organisms. Some cyanobacteria are capable of nitrogen fixation. In the event of a shortage of reduced nitrogen compounds, certain photosynthetic vegetative cells of the filamentous cyanobacteria *Anabaena* and *Nostoc sp.* are terminally differentiated into diazotrophic heterocysts located every ten to twenty vegetative cells along the filament [3]. Differentiation is necessary to separate oxygen generated in photosynthetic vegetative cells from nitrogenase produced in the nitrogen-fixing heterocysts because nitrogenases are readily inactivated by oxygen. During the late stages of heterocysts differentiation, in addition to changes in global gene expression, three different site-specific DNA rearrangements of heterocyst-specific genes occur. The 55-kb *fdxN* element, the 11-kb *nifD* element, and the 9.5-kb *hupL* element are excised from within genes on the chromosome during the late stages of heterocyst differentiation by site-specific recombinases encoded on the DNA elements (Figure 1) [3]. Precise excision of these elements is required to restore the open reading frames of the heterocyst-specific genes that are interrupted by these elements.

The excision of the *fdxN* element is accomplished by the recombinase XisF, which requires at least two additional controlling factors XisH and XisI [4,5]. Deletion of a region containing the *xisH* and *xisI* genes blocks excision of the *fdxN* element, and both *xisH* and *xisI* are required to complement the deletion mutant [5]. A plasmid containing both *xisH* and *xisI* causes *xisF*-dependent excision of the *fdxN* element from the chromosome in vegetative cells growing in the presence of a nitrogen source, but expression of the recombinase *xisF* gene alone does not cause the rearrangement in vegetative cells. These data indicate that XisH and XisI are involved in controlling the cell-type specificity of the excision event. The *nifD* and *hupL* DNA rearrangements were not effected in either the *xisH* and *xisI* deletion or overexpression strains. Homologs of XisH and XisI have been found in the BOGUAY

(*Beggiatoa sp. Orange Guaymas*) genome [6]. They were likely acquired through lateral gene transfer.

Recombinases catalyze the process of conservative site-specific recombination (CSSR), which in some cases require additional accessory proteins. Based on amino acid sequence similarities and the catalytic residues involved, all known recombinases can be classified into two families: the tyrosine recombinases (also known as the λ integrase family) [7] and the serine recombinases (also known as resolvase family) [8]. These two recombinase families appear to be the result of convergent evolution as they differ in sequences and mechanisms [9]. Based on the lengths of their C-terminal domains (CTD), serine recombinase families can be further divided into two subgroups: small serine recombinase and large serine recombinase (LSR). Recombinase XisF, homologous to *B. subtilis* SpoIVCA recombinase [4], belongs to the large serine recombinase family.

Here we report the first high-resolution crystal structures of XisH and XisI, and propose that XisH and XisI likely function as endonucleases and recombination directionality factors (RDFs), respectively, based on the analysis of their sequences and structures.

MATERIAL AND METHODS

Structure Determination

Atomic coordinates and experimental structure factors have been deposited into the Protein Data Bank (http://www.rcsb.org) with PDB ID: 2INB, 2OKF, 2NLV, 2NWV, 2NVM and 3D7Q.

Structure determination of XisH-like proteins and XisI-like proteins were carried out by the JCSG (Joint Center for Structural Genomics) high-throughput structural biology pipeline [10]. Detailed protocols used for protein expression and purification can be found in the links therein. Briefly, genes encoding XisH-like proteins and XisI-like proteins were cloned into kanamycin-resistant, arabinose and T7 dual-promoter-based bacteria expression vector pSpeedET, which has an N-terminal expression and purification tag (MGSDKIHHHHHH) followed by a tobacco etch virus (TEV) protease cleavage site (ENLYFQG). Target proteins were expressed in *E. coli* HK100 cells in selenomethionine-containing medium with suppression of normal methionine synthesis. Cells were pelleted by centrifugation, resuspended in lysis buffer [50 mM HEPES, 50 mM NaCl, 10 mM imidazole, 1 mM tris(2-carboxyethyl) phosphine-HCl (TCEP), pH 8.0], and frozen. After one freeze-thaw cycle, the cells were lysed by Tissue Homogenizer (Omni International, Part Number: TH115) coupled with Microfluidizer Processor (Microfluidics, Part Number: M-110L). Target proteins were purified with Ni-NTA metal chelating resins (GE Life Sciences). The eluate was buffer-exchanged to TEV buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) using a PD-10 column (GE Life Sciences), treated (1mg of AcTEV per 15 mg of protein) with AcTEV protease (Life Technologies), leaving a lone residue “G” cloning artifact at the N-terminus of the target protein. The cleaved N-terminal 6xHis tags were removed by Ni-NTA metal chelating resins, while the cleaved proteins were recovered from the flowthrough. Proteins were further purified by HiLoad 16/60 Superdex 200 or 75 columns (GE Life Sciences) to remove remaining impurities. Purified proteins were buffer

exchanged to HEPES crystallization buffer (20 mM HEPES, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP, pH 8.0), and concentrated with Amicon Ultra-15 (Millipore) to 14.3 (mg/mL), 16.0 (mg/mL), 14.4 (mg/mL), 16.8 (mg/mL), 13.0 (mg/mL), 14.6 (mg/mL), for 2INB, 2NLV, 2NWV, 3D7Q, 2OKF, 2NVM, respectively. Crystallization trials were conducted using sitting drop vapor-diffusion method. Equal volume (200 nL) of protein and well solution were mixed and equilibrated against 50 uL of reservoir solutions at 277 K. Proteins were crystallized in the following conditions: 0.2M $(NH_4)_2SO_4$, 10.0% Glycerol, 20.0% PEG-300, 0.1M Phosphate Citrate, pH 4.2, for 2INB; 1.0M LiCl, 20.0% PEG-6000, 0.1M TRIS, pH 8.0, for 2NLV; 0.2M NaCl, 30.0% PEG-3000, 0.1M TRIS, pH 7.0, for 2NWV; 0.15M lithium sulfate, 26.0% PEG-4000, 0.1M TRIS pH 9.0, for 3D7Q; 0.2M $(NH_4)_2SO_4$, 25.0% PEG-4000, 0.1M Acetate, pH 4.6 for 2OKF; 30.0% PEG-6000, 0.1M MES, pH 6.0, for 2NVM. Cryoprotectants used for data collection were 10% Glycerol, 10% PEG-200, 10% PEG-200, 20% Glycerol, 12% Ethylene Glycol, 15% PEG-200, for 2INB, 2NLV, 2NWV, 3D7Q, 2OKF, 2NVM, respectively. Diffraction data were collected at Advanced Light Source (ALS) beamline 8.2.2 for 2INB, 2NLV, 2NWV, 2NVM, and at Stanford Synchrotron Radiation Lightsource (SSRL) beamlines 11-1 for 3D7Q and 2OKF. All diffraction data were processed with MOSFLM [11] and scaled with SCALA in the CCP4 package [12], except for 2OKF, which was processed with XDS [13] and scaled with XSCALE in the XDS suite. The crystal structures were solved by selenium MAD phasing with SHELXD [14] and autoSHARP [15] for 3D7Q and 2NVM, and with SOLVE/RESOLVE [16] for 2INB, 2NLV, 2OKF, and 2NWV. Automatic model building was performed with ARP/wARP [17]. Model completion and refinement was performed with coot [18] and REFMAC5.2 [19]. Structural and refinement statistics can be found in the supplementary data 1. The structures were validated using the JCSG Quality Control server (http://smb.slac.stanford.edu/jcsg/QC) [10].

Structure Analysis

2INB and 3D7Q dimers were generated by symmetry-related positions in PyMOL [20]. Dimer interfaces were assessed by PISA [21]. Structure alignments were performed with MultiProt [22] or POSA [23]. Conservation of amino acid residues was assessed by ConSurf [24]. The electrostatic potential surfaces were calculated with PDB2PQR [25] and APBS [26]. Structure graphics were prepared in Chimera [27] or PyMOL.

Structure Modeling

Model of XisF was built with I-TASSER [28], using the profile-profile sequence alignments from FFAS [29,30] as guides. Recent structure of the N-terminal and recombinase domains of the *Streptomyces* temperate phage serine recombinase, phiC31 integrase (PDB ID 4BQQ), provides a template for modeling the N-terminal 2/3 of XisF. The C-terminus of XisF can be modeled from the recent structure of the DNA binding domain of an integrase from *Listeria innocua* Clip11262 (PDB ID 4KIS). The tetrameric XisF synapse model was constructed using tetrameric serine recombinase catalytic domain structures (PDB ID 1ZR4 and 3PKZ) as guides. The tetrameric XisF synapse model was subjected to energy minimization using YASARA [31] to relieve steric clashes. Dockings were performed with HexDock [32].

RESULTS

Structures of XisH Proteins

Structures of two XisH proteins, from *Nostoc punctiforme* PCC 73102 and from *Anabaena variabilis* ATCC 29413 (PDB ID 2INB and 2OKF, respectively) were solved at JCSG as part of an effort to provide structural coverage for novel protein families [33]. They represent the first and, so far, only structures available of the XisH-like protein family (PF08814) [34]. This family is a member of PD-(D/E)XK nuclease superfamily and shows distant homology to other nucleases. These structures were classified by SCOP [35] as having restriction endonuclease-like folds. Each monomer contains four α helices, six β strands, and one 3/10 helix. Monomers of 2INB and 2OKF superposed well for their backbone atoms (RMSD 0.34 Å over 123 cα atoms), except for the loop connecting the β3 and αB, in which they deviate by as much as 7Å (Figure 2a).

The XisH proteins form dimers in their native state, as observed in the crystal structures and also as classified by PPI Prediction Server (http://ppi.zbh.uni-hamburg.de/). The dimer interfaces, formed mainly by αB and αD between monomers, consist of a buried surface area of 1022 and 1001 $Å^2$ for 2INB and 2OKF, respectively. The large buried surface of the dimer interface lends support for dimer being the native state of XisH, however no experimental evidence is yet available. Dimers of 2INB and 2OKF superposed well with each other for their backbone atoms (RMSD 0.52 Å over 246 cα atoms), except the loop connecting the β3 and αB (Figure 2b).

A salient feature of the XisH dimers is the uneven electrostatic charge distribution (supplementary data 2 Figure S1 and S2). In both 2INB and 2OKF, one side of the dimer has a large positive charge patch atop the dimer interface. These positive charge patches are good candidates for DNA binding. Intriguingly, in the complex structure of the structurally similar restriction endonuclease BamHI-DNA complex (PDB ID 1BHM), the nucleic acid also binds to the positively charged groove at the protein dimer interface.

Structures of XisI Protein

As part of the same effort, JCSG also solved structures (PDB ID 2NLV, 2NWV, 2NVM and 3D7Q) of several proteins from the XisI-like protein family (PF08869). They also belong to the α/β protein class, but have a novel fold, classified by SCOP as XisI-like fold. They are the first and only solved structures of the XisI-like protein family to date. Each monomer contains three α helices, one 3–10 helix, and five β strands (Figure 3a). Monomers of 2NLV, 2NWV, 2NVM and 3D7Q superposed well, with an RMSD of 1.0 Å for their backbone atoms over 89 cα atoms.

Dimers are the native form of XisI-like proteins, as seen in their crystal structures and also as classified by PPI Prediction Server (http://ppi.zbh.uni-hamburg.de/). The dimerization interface involves the β sheets of the two interacting XisI-like protein monomers (Figure 3b). The buried surface areas are 1047, 948, 1332.6, 1032.7 $Å^2$ for 2NLV, 2NWV, 2NVM, 3D7Q, respectively. Although no experimental evidence is yet available, the large buried surface of the dimer interface as observed in the crystal structure suggests XisI may natively form dimers. Unlike XisH-like proteins, which have a distinct cluster of positively charged

residues, the surfaces of XisI-like proteins, except 2NLV and 2NVM, are dominated by residues with negative electrostatic potentials, as shown in supplementary data 2 Figure S3, S4, S5, and S6.

Model of Large Serine Recombinase XisF Structure

Experimental structure of recombinase XisF structure is not yet available, and experimental structure for a full length large serine recombinase is not available either, but a high quality model can be built by a multi-template modeling (Figure 4). A model of XisF was built with I-TASSER [15] using PDB ID 4BQQ and 4KIS as templates. XisF is a multi-domain large serine recombinase (LSR), noted for its large C-terminal domain. The N-terminus of XisF harbors the catalytic domain, similar to that of the N-terminal domain of serine resolvase (PF00239). The center of XisF contains a DNA binding recombinase domain (PF07508), which is often associated with the serine resolvase catalytic domain. A conserved DNA-binding zinc beta ribbon domain (PF13408) is found after the recombinase domain in XisF. The zinc beta ribbon domain contains a DNA binding motif, corresponding to the motif that binds to the major groove of bound DNA in the DNA binding domain of integrase from *Listeria innocua* Clip11262 (PDB ID 4KIS). The linker region between the recombinase domain and the zinc beta ribbon domain is likely a minor groove binding loop, as suggested by superposition with the DNA binding domain of integrase (PDB ID 4KIS). The zinc beta ribbon domain is intercalated with a coil-coil motif region, which extends over 80 angstroms in length.

DISCUSSION

Cyanobacteria, during heterocysts differentiation, require excision of specific DNA elements by site-specific recombinases to express heterocyst-specific proteins. The excision of the *fdxN* element requires both recombinase XisF and regulating factors XisH and XisI. All three proteins are required to perform site-specific recombination [5]. Deletion of XisH and XisI abolish the excision of the heterocyst-specific *fdxN* element. The detailed mechanism of XisF-XisH-XisI mediated site-specific recombination remains unknown. However, availability of XisF models and experimental structures of XisH and XisI allude to their probable functions.

XisH-like proteins are likely endonucleases

XisH-like proteins represent a novel protein fold, as they are significantly different from know structures. Structure similarity search for XisH-like protein 2INB using the Dali server [36] yields hits mostly with less than 18% sequence identities when compared with XisH-like protein 2INB. The top hit of 2INB from the Dali server is an uncharacterized protein AF1548 from *Archaeoglobus fulgidus* (PDB 1Y88), despite sharing only 14% sequence identity. Both structures aligned with an RMSD of 2.9 Å for their backbone atoms over 105 Cα atoms. The two beta strands at the C-terminus of XisH-like protein 2INB are replaced with a shorter beta strand and 6 alpha helices (including one 3/10 helix), which form a subdomain, in the corresponding positions of AF1548 structure. The top Dali hit of known structure is a CDI toxin/immunity complex (PDB 4G6V) from *Burkholderia pseudomallei* 1026b. The CDI toxin and XisH-like protein 2INB aligned with an RMSD of 3.3 Å for their

backbone atoms over 104 Cα atoms. The N-terminus of the CDI toxin has an extended alpha helix and a beta strand connected by a nine-residue loop, when compared with the XisH-like protein 2INB. The CDI toxin was reported to be structurally similar to endonucleases [37].

Despite weak sequence level similarities, there are parallels between XisH-like proteins and Holliday junction resolvase Hjc (PDB ID 1GEF) [38,39], an archaeal endonuclease. First, they both adopt restriction endonuclease-like folds. Second, their surfaces both feature asymmetric electrostatic potential distribution, reminiscent of those in type II restriction endonucleases. Third, the XisH-like protein crystal structure shows a dimer with a cluster of positively charged residues lining the dimer interface, likely a DNA binding site. The shallowness of the positively charged groove in XisH-like proteins suggests that substantial conformational change is necessary to accommodate the DNA. Indeed, in the structure of phage T7 endonuclease I, a Holliday junction-resolving enzyme, in complex with a synthetic four-way DNA junction (PDB ID 2PFJ), the T7 endonuclease I monomers were separated but remained tethered through an extended loop to accommodate the Holliday junction DNA [40]. It remains to be verified whether XisH-like proteins function as Holliday junction-resolving enzymes, or simply as restriction endonucleases. Conserved residues of XisH-like protein dimer (PDB ID 2INB) are shown in supplementary data 2 Figure S8.

XisI-like proteins are likely recombination directionality factors

While small serine recombinases generally function without recombination directionality factors (RDFs), also known as excisionases, large serine recombinases, in contrast, are often found to require RDFs. Excisionase was suggested to prevent the reintegration of the excised phage genome during the lambda bacteriophage lytic cycle[41]. XisI is similar sequence-wise to some RDFs associated with large serine recombinases, e.g. excisionase from *Mahella australiensis* 50-1 BON (YP_004462138.1), and Rv1584c (NP_216100.1) [42], an RDF for the large serine recombinase phiRv1 integrase in *Mycobacterium tuberculosis* H37Rv. Sequence alignments between XisI and selected RDFs are shown in Figure 5.

DNA binding is a common feature for excisionases in tyrosine recombinase mediated site-specific recombinations. Excisionases in large serine recombinase mediated site-specific recombinations appear to use different mechanisms, since they do not show appreciable binding to attachment sites DNA [42–45]. It was suggested that RDF Rv1584c, instead, interacts directly with phiRv1 integrase [42]. XisI (PDB ID 2NVM) preferentially docked to the cavity between the catalytic domain and coil-coil motifs in the modeled XisF synapse (Figure 6). Parts of the XisI dimer are in direct steric clashes with the coil-coil motifs in the modeled XisF synapse, suggesting XisI could disrupt the interactions of coil-coil motifs. Consistent with this notion, a hyperactive phiC31 integrase mutant (E449K) can catalyze the attL x attR reaction in the absence of RDF[46]. The equivalent position on XisF (E407) is located in the coil-coil hinge region, where the docked XisI has steric contact with the coil-coil motifs (Figure 6). Sequence alignment between XisF and phiC31 integrase is shown in Figure 7. It is also consistent with the current model for the LSR mediated site-specific recombination [32,37,38], based on the recent crystal structure of the C-terminus of a LSR bound to DNA [37]. In this model, the interactions between coil-coil motifs are thought to inhibit the reverse excision reaction, in the absence of additional controlling factors.

Unlike the well-studied lambda excisionases, which are shorter in length and adopt the winged-helix fold, XisI-like proteins as described here are longer and adopt a unique XisI-like fold. It is likely that XisI-like proteins represent a new family of recombination directionality factors. To our knowledge, XisI-like protein structures are the first and only available recombination directionality factors structures for large serine recombinases to date.

Structure similarity search with Dali shows no significant similarity with known protein structures and thus represents a novel protein fold. The top hit of XisI-like protein (PDB 2NLV) from Dali search is Frataxin (PDB 1EKG), which shares only 9% sequence identity. The two structures aligned with an RMSD of 3.9 Å for their backbone atoms over 77 Cα atoms. Frataxin is a protein known to cause the neurodegenerative disease Friedreich's ataxia when its expression level reduced [47]. Conserved residues of the XisI-like protein dimer (PDB ID 2NLV) are shown in supplementary data 2 Figure S9.

Conclusion

Cyanobacteria have been the focus of an increasing number of studies because of their roles in carbon and nitrogen fixation, and as a potential source of renewable products such as biofuels. Furthermore, site-specific DNA recombination, as found in cyanobacteria differentiation, could be a valuable tool in molecular cloning, and gene delivery and therapy [48,49]. Similar site-specific DNA recombination from bacteriophage lambda has been successfully applied as an efficient cloning technology (Gateway cloning) in the biotech industry[50,51]. Structure determination of XisH-like proteins and XisI-like proteins, both the first structures in their respective protein families as presented in this report, will advance our understanding of cyanobacteria differentiation as well as the mechanism and regulation of site-specific recombination.

Lastly, the cyanobacterial "excision elements" represent a type of mobile DNA related to transposons, cryptic prophage, DNA islands, and integrative and conjugative elements (ICEs) [52]. It is unclear if these elements are simply parasitic or have a role in gene regulation. However, it is clear that excision of the elements has evolved to be tightly regulated by heterocyst differentiation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the members of the JCSG high-throughput structural biology pipeline for their contribution to this work. Gye Won Han refined structures of 2INB, 2NLV, 2NWV, 3D7Q; Kevin K. Jin refined the structure of 2OKF; Abhinav Kumar refined the structure of 2NVM.

FUNDING

The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and

Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS, NCRR or NIH. This work was supported in part by National Institutes of Health Grant U54 GM094586 from the NIGMS Protein Structure Initiative to the Joint Center for Structural Genomics.

References

1. Leigh JA. 2000; Nitrogen fixation in methanogens: the archaeal perspective. Curr Issues Mol Biol. 2:125–131. [PubMed: 11471757]
2. Cabello P, Roldan MD, Moreno-Vivian C. 2004; Nitrate reduction and the nitrogen cycle in archaea. Microbiology. 150:3527–3546. [PubMed: 15528644]
3. Kumar K, Mella-Herrera RA, Golden JW. 2010; Cyanobacterial heterocysts. Cold Spring Harb Perspect Biol. 2:a000315. [PubMed: 20452939]
4. Carrasco CD, Ramaswamy KS, Ramasubramanian TS, Golden JW. 1994; Anabaena xisF gene encodes a developmentally regulated site-specific recombinase. Genes Dev. 8:74–83. [PubMed: 8288129]
5. Ramaswamy KS, Carrasco CD, Fatma T, Golden JW. 1997; Cell-type specificity of the Anabaena fdxN-element rearrangement requires xisH and xisI. Mol Microbiol. 23:1241–1249. [PubMed: 9106215]
6. MacGregor BJ, Biddle JF, Teske A. 2013; Mobile elements in a single-filament orange Guaymas Basin Beggiatoa (“Candidatus Maribeggiatoa”) sp. draft genome: evidence for genetic exchange with cyanobacteria. Appl Environ Microbiol. 79:3974–3985. [PubMed: 23603674]
7. Nunes-Duby SE, Kwon HJ, Tirumalai RS, Ellenberger T, Landy A. 1998; Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res. 26:391–406. [PubMed: 9421491]
8. Smith MC, Thorpe HM. 2002; Diversity in the serine recombinases. Mol Microbiol. 44:299–307. [PubMed: 11972771]
9. Grindley ND, Whiteson KL, Rice PA. 2006; Mechanisms of site-specific recombination. Annu Rev Biochem. 75:567–605. [PubMed: 16756503]
10. Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wooley J, et al. 2010; The JCSG high-throughput structural biology pipeline. Acta Crystallogr Sect F Struct Biol Cryst Commun. 66:1137–1142.
11. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. 2011; iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr. 67:271–281. [PubMed: 21460445]
12. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. 2011; Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 67:235–242. [PubMed: 21460441]
13. Kabsch W. 2010; Xds. Acta Crystallogr D Biol Crystallogr. 66:125–132. [PubMed: 20124692]
14. Sheldrick GM. 2008; A short history of SHELX. Acta Crystallogr A. 64:112–122. [PubMed: 18156677]
15. Vonrhein C, Blanc E, Roversi P, Bricogne G. 2007; Automated structure solution with autoSHARP. Methods Mol Biol. 364:215–230. [PubMed: 17172768]
16. Terwilliger T. 2004; SOLVE and RESOLVE: automated structure solution, density modification and model building. J Synchrotron Radiat. 11:49–52. [PubMed: 14646132]
17. Langer G, Cohen SX, Lamzin VS, Perrakis A. 2008; Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc. 3:1171–1179. [PubMed: 18600222]
18. Emsley P, Cowtan K. 2004; Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 60:2126–2132. [PubMed: 15572765]
19. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, et al. 2011; REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 67:355–367. [PubMed: 21460454]
20. The PyMOL Molecular Graphics System, Version 1.2r3pre. Schrödinger, LLC;

21. Krissinel E, Henrick K. 2007; Inference of macromolecular assemblies from crystalline state. J Mol Biol. 372:774–797. [PubMed: 17681537]
22. Shatsky M, Nussinov R, Wolfson HJ. 2004; A method for simultaneous alignment of multiple protein structures. Proteins. 56:143–156. [PubMed: 15162494]
23. Ye Y, Godzik A. 2005; Multiple flexible structure alignment using partial order graphs. Bioinformatics. 21:2362–2369. [PubMed: 15746292]
24. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010; ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38:W529–533. [PubMed: 20478830]
25. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, et al. 2007; PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 35:W522–525. [PubMed: 17488841]
26. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001; Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 98:10037–10041. [PubMed: 11517324]
27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. 2004; UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 25:1605–1612. [PubMed: 15264254]
28. Roy A, Kucukural A, Zhang Y. 2010; I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 5:725–738. [PubMed: 20360767]
29. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A. 2005; FFAS03: a server for profile--profile sequence alignments. Nucleic Acids Res. 33:W284–288. [PubMed: 15980471]
30. Jaroszewski L, Li Z, Cai XH, Weber C, Godzik A. 2011; FFAS server: novel features and applications. Nucleic Acids Res. 39:W38–44. [PubMed: 21715387]
31. Krieger E, Joo K, Lee J, Lee J, Raman S, et al. 2009; Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins. 77(Suppl 9):114–122. [PubMed: 19768677]
32. Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW. 2010; HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Res. 38:W445–449. [PubMed: 20444869]
33. Nair R, Liu J, Soong TT, Acton TB, Everett JK, et al. 2009; Structural genomics is the largest contributor of novel structural leverage. J Struct Funct Genomics. 10:181–191. [PubMed: 19194785]
34. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. 2012; The Pfam protein families database. Nucleic Acids Res. 40:D290–301. [PubMed: 22127870]
35. Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995; SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 247:536–540. [PubMed: 7723011]
36. Holm L, Rosenstrom P. 2010; Dali server: conservation mapping in 3D. Nucleic Acids Res. 38:W545–549. [PubMed: 20457744]
37. Morse RP, Nikolakakis KC, Willett JL, Gerrick E, Low DA, et al. 2012; Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. Proc Natl Acad Sci U S A. 109:21480–21485. [PubMed: 23236156]
38. Nishino T, Komori K, Tsuchiya D, Ishino Y, Morikawa K. 2001; Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition. Structure. 9:197–204. [PubMed: 11286886]
39. Bond CS, Kvaratskhelia M, Richard D, White MF, Hunter WN. 2001; Structure of Hjc, a Holliday junction resolvase, from Sulfolobus solfataricus. Proc Natl Acad Sci U S A. 98:5509–5514. [PubMed: 11331763]
40. Hadden JM, Declais AC, Carr SB, Lilley DM, Phillips SE. 2007; The structural basis of Holliday junction resolution by T7 endonuclease I. Nature. 449:621–624. [PubMed: 17873858]
41. Sam MD, Papagiannis CV, Connolly KM, Corselli L, Iwahara J, et al. 2002; Regulation of directionality in bacteriophage lambda site-specific recombination: structure of the Xis protein. J Mol Biol. 324:791–805. [PubMed: 12460578]

42. Bibb LA, Hancox MI, Hatfull GF. 2005; Integration and excision by the large serine recombinase phiRv1 integrase. Mol Microbiol. 55:1896–1910. [PubMed: 15752208]
43. Ghosh P, Wasil LR, Hatfull GF. 2006; Control of phage Bxb1 excision by a novel recombination directionality factor. PLoS Biol. 4:e186. [PubMed: 16719562]
44. Khaleel T, Younger E, McEwan AR, Varghese AS, Smith MC. 2011; A phage protein that binds phiC31 integrase to switch its directionality. Mol Microbiol. 80:1450–1463. [PubMed: 21564337]
45. Zhang L, Zhu B, Dai R, Zhao G, Ding X. 2013; Control of Directionality in Streptomyces Phage phiBT1 Integrase-Mediated Site-Specific Recombination. PLoS One. 8:e80434. [PubMed: 24278283]
46. Rowley PA, Smith MC, Younger E, Smith MC. 2008; A motif in the C-terminal domain of phiC31 integrase controls the directionality of recombination. Nucleic Acids Res. 36:3879–3891. [PubMed: 18502775]
47. Pastore A, Puccio H. 2013; Frataxin: a protein in search for a function. J Neurochem. 126(Suppl 1):43–52. [PubMed: 23859340]
48. Kolb AF. 2002; Genome engineering using site-specific recombinases. Cloning Stem Cells. 4:65–80. [PubMed: 12006158]
49. Kilby NJ, Snaith MR, Murray JA. 1993; Site-specific recombinases: tools for genome engineering. Trends Genet. 9:413–421. [PubMed: 8122308]
50. Hartley JL, Temple GF, Brasch MA. 2000; DNA cloning using in vitro site-specific recombination. Genome Res. 10:1788–1795. [PubMed: 11076863]
51. Katzen F. 2007; Gateway((R)) recombinational cloning: a biological operating system. Expert Opin Drug Discov. 2:571–589. [PubMed: 23484762]
52. Wozniak RA, Waldor MK. 2010; Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol. 8:552–563. [PubMed: 20601965]

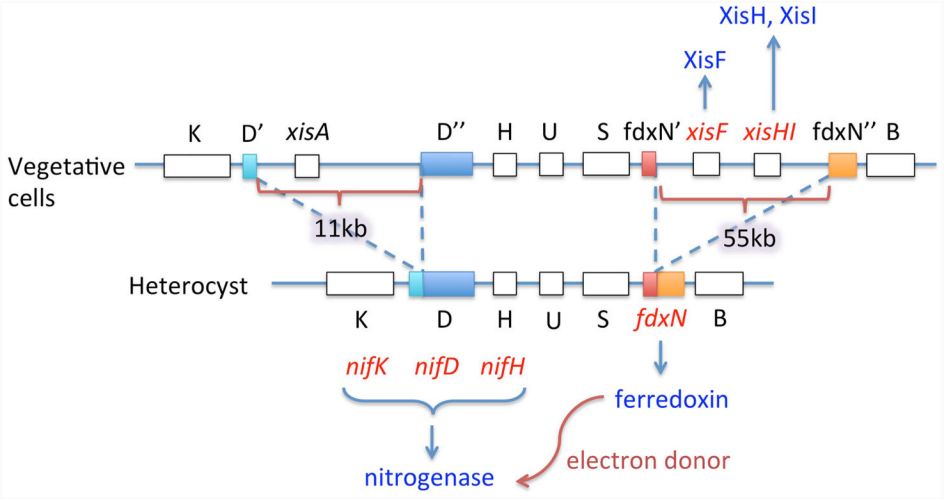

Figure 1.

Site-specific recombination in heterocysts. The DNA elements are named for the genes within which they reside, *fdxN*, *nifD,* and *hupL*, and are excised by specific recombinases (XisF, XisA, and XisC, respectively) from the chromosome during heterocyst differentiation. XisF requires at least two additional regulating factors, XisH and XisI.

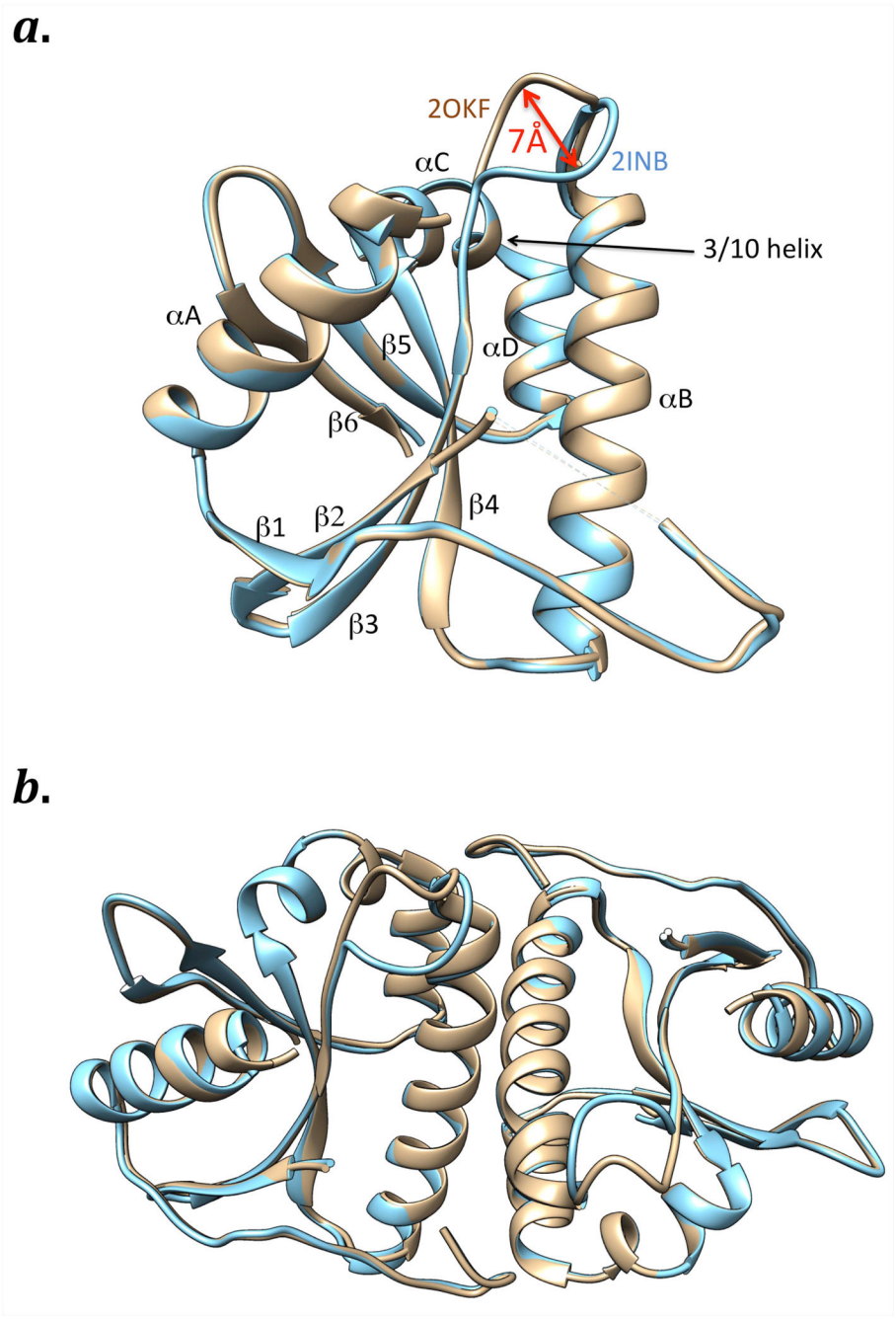

Figure 2.
Superposition of XisH-like proteins. 2INB in cyan and 2OKF in brown. ***a***. monomers. ***b***. dimers.

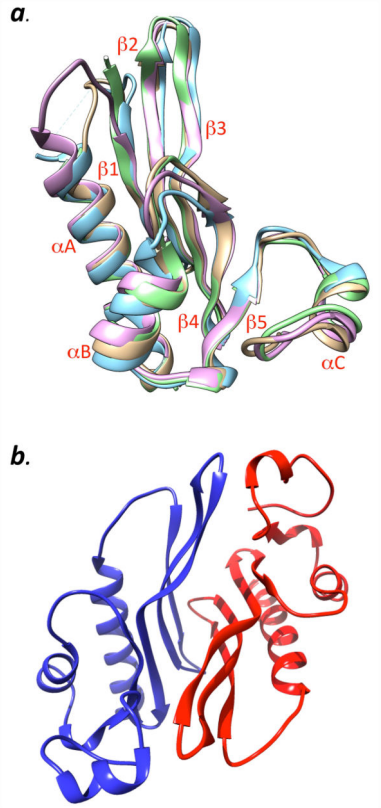

Figure 3.

a. Superposition of XisI-like proteins. 2NLV in brown, 2NVM in cyan, 2NWV in light purple, 3D7Q in green. ***b.*** Dimer of XisI-like protein 2NLV.

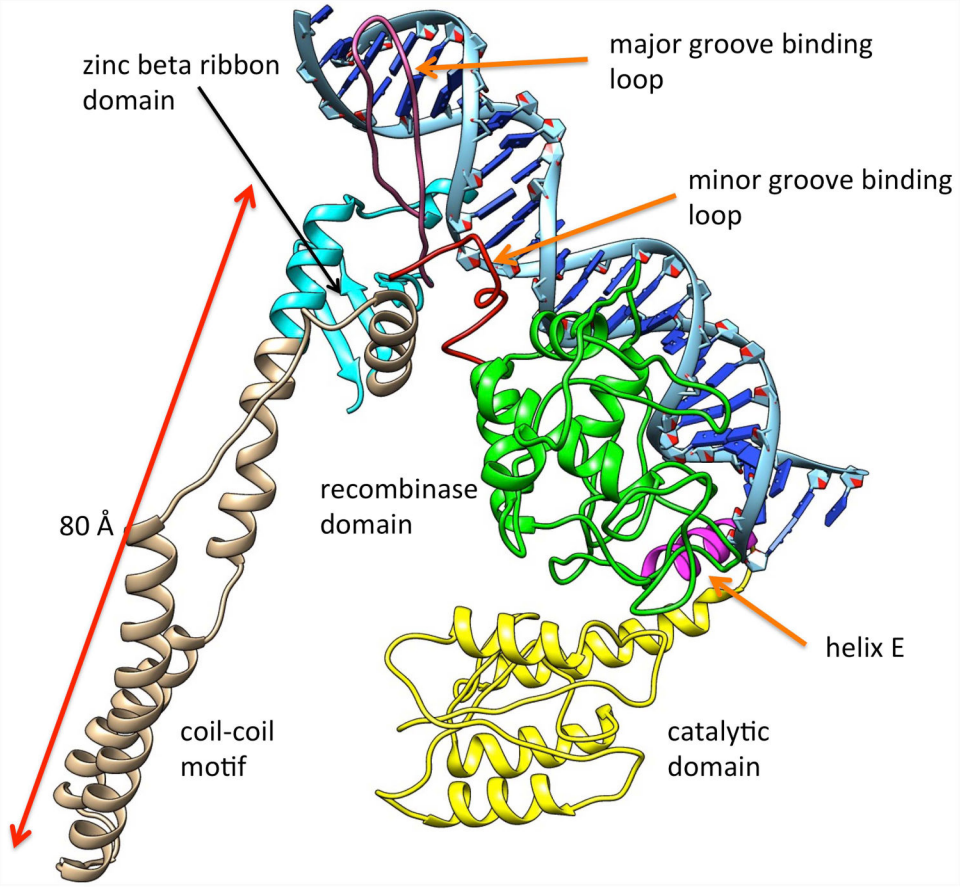

Figure 4.

XisF model. The XisF is composed of multiple domains. Catalytic domain (yellow) is located at the N-terminus, followed by recombinase domain (green), and zinc beta ribbon domain (cyan). A coil-coil motif (brown) extends away from the zinc beta ribbon domain. A major groove binding loop (pink) is found within the zinc beta ribbon domain. The linker region (red) between the recombinase domain and zinc beta ribbon domain intercalates into the minor groove of bound DNA. Helix E is shown in magenta.

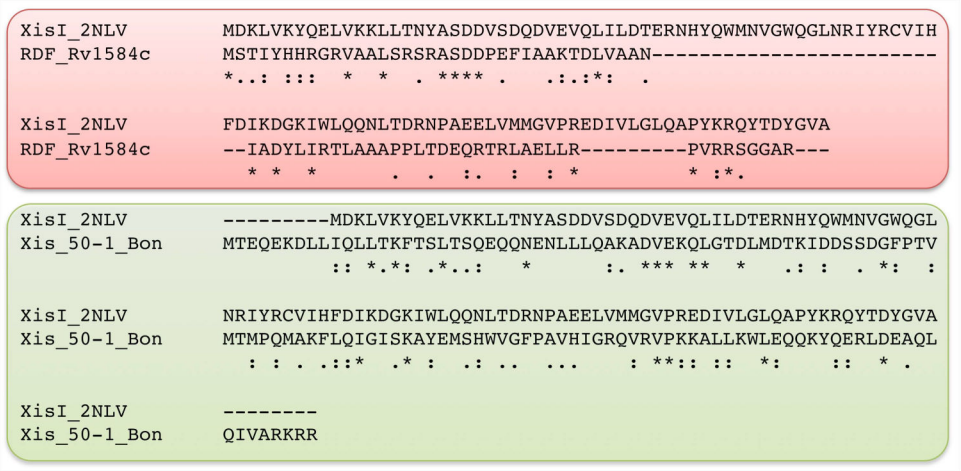

Figure 5.
Sequence alignments between XisI and selected RDFs.

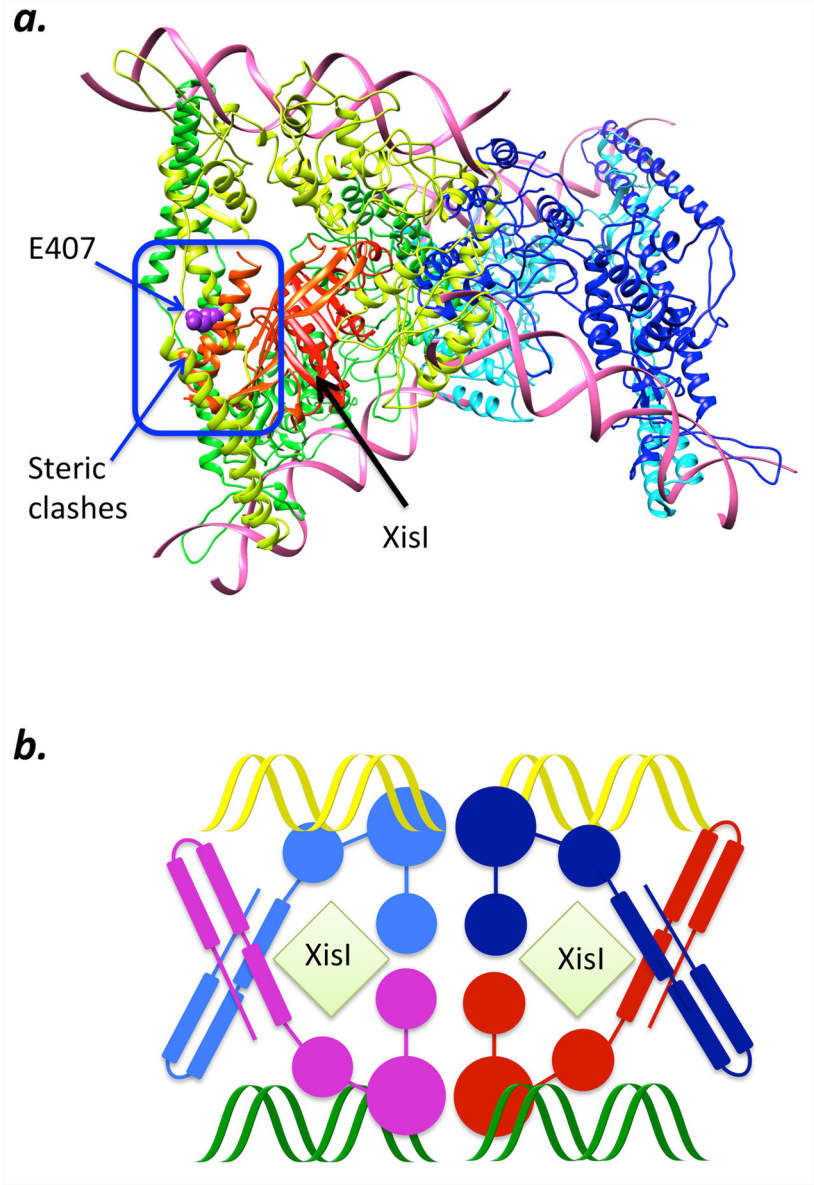

Figure 6.

Model of XisI bound to XisF synapse. ***a.*** Docking of XisI to XisF synapse. XisF is shown in blue, light blue, green, and yellow. XisI is shown in red and orange. Nucleic acids are shown in pink. The position of E407 from one of the XisF monomers is shown in purple. ***b.*** Schematic drawing shows the XisI docked to the XisF synapse.

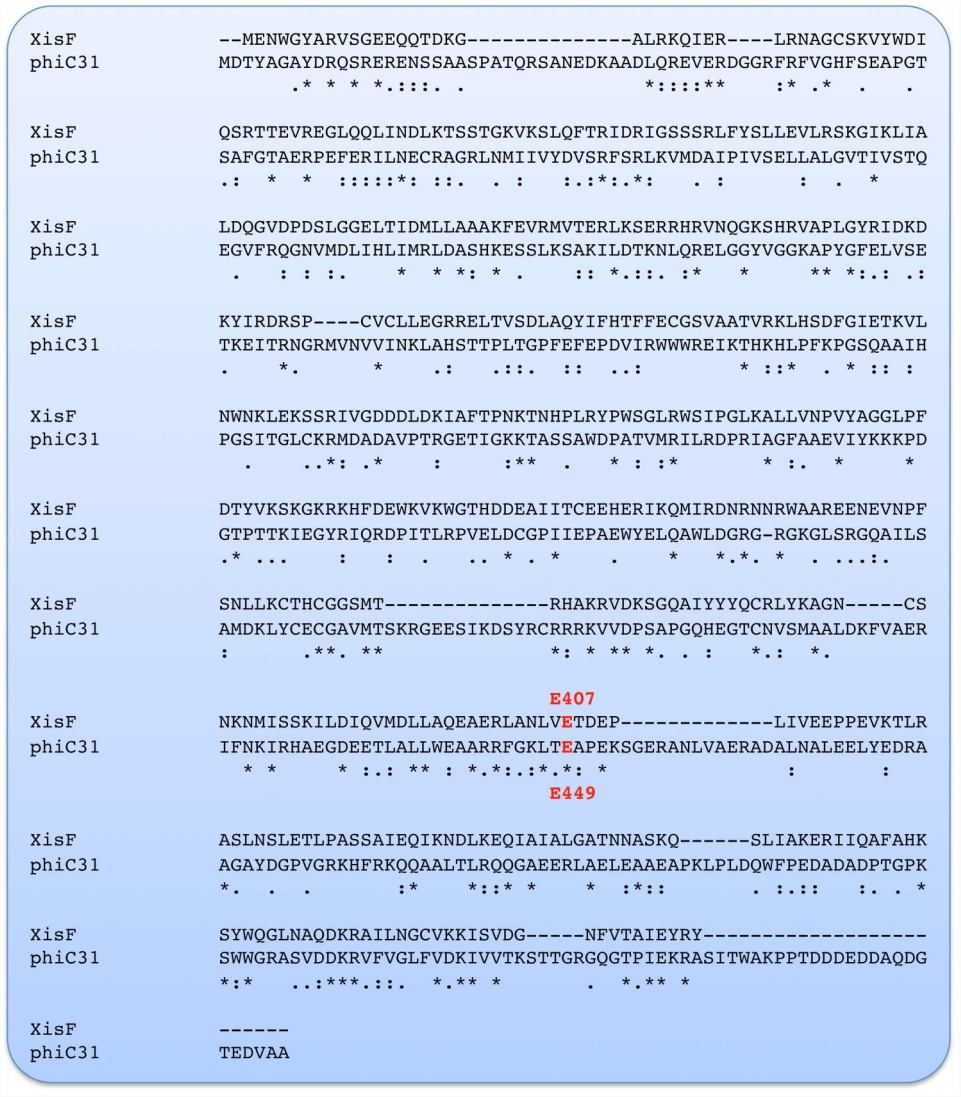

Figure 7.
Sequence alignment between XisF and phiC31 integrase.