The Structure of an Archaeal Viral Integrase Reveals an Evolutionarily Conserved Catalytic Core yet Supports a Mechanism of DNA Cleavage in trans

Brian J. Eilers, a,b Mark J. Young, a,c,d and C. Martin Lawrence a,b

Thermal Biology Institute, a Department of Chemistry and Biochemistry, b Department of Plant Sciences and Plant Pathology, c and Department of Microbiology, d Montana State University, Bozeman, Montana, USA

The first structure of a catalytic domain from a hyperthermophilic archaeal viral integrase reveals a minimal fold similar to that of bacterial HP1 integrase and defines structural elements conserved across three domains of life. However, structural superposition on bacterial Holliday junction complexes and similarities in the C-terminal tail with that of eukaryotic Flp suggest that the catalytic tyrosine and an additional active-site lysine are delivered to neighboring subunits in trans. An intramolecular disulfide bond contributes significant thermostability in vitro.

Lysogenic viral life cycles in which viral DNA integrates within the host genome are employed in each of the three domains of life: Eukarya, Bacteria, and Archaea. Mechanistically, the integration of viral DNA is an example of site-specific recombination, and enzymes responsible for integration fall into two categories, tyrosine and serine recombinases. Despite similar activities, these recombinase families are unrelated in protein sequence and structure and employ different recombination mechanisms (12, 25).

Tyrosine recombinases like the well-known integrase from bacteriophage λ, or Cre, which is involved in dimer reduction of phage P1 plasmids, are composed of a variable N-terminal DNA binding domain and a C-terminal catalytic domain. Though unexpectedly diverse, the catalytic domains generally contain 6 well-conserved active-site residues within a 40-residue C-terminal segment (3). This includes a nucleophilic tyrosine that attacks DNA, expelling a 5'-hydroxyl group to form a covalent adduct. This is followed by the reverse reaction, in which an incoming DNA strand displaces the tyrosine, religating the DNA to form a Holliday junction (HJ) intermediate. This reaction cycle is then repeated to complete DNA integration, or excision.

Structural studies of phage and bacterial tyrosine recombinases have provided detailed information on the structure of the catalytic domain alone and for complete enzymes complexed with DNA in several mechanistically important states, including tetrameric recombinase assemblies on the HJ intermediate (1, 7, 11, 13–15, 17, 20). Structural studies of the tetrameric HJ complex have also been completed for Flp, a eukaryotic tyrosine recombinase. For these reasons, the structure of IntSSV is of considerable interest.

Structure of the IntSSV catalytic domain. With minor modifications described in the supplemental material, IntSSV was cloned (21), expressed, and purified and its structure (Protein Data Bank identifier [PDB ID], 3UXU) determined by multiwavelength anomalous diffraction at the Se edge as previously described (19). Full-length IntSSV runs as a monomer on a Superdex S75 column and has a calculated mass of 39.8 kDa. However, N-terminal sequencing and mass spectrometry of dissolved crystals indicated subsequent, serendipitous in situ proteolysis (4) leading to crystallization of only the C-terminal domain, residues 173 to 335. BLAST searches against the PDB do not identify similarity to any integrase, but consistent with the results of previous bioinformatics and biochemical work (3, 18, 24), the structure revealed an unembellished integrase catalytic core composed of three successive alpha helices (α1 to α3) followed by a three-stranded, up-and-down, antiparallel β-sheet (β1 to β3) and four additional helices (α4 to α7) (Fig. 1).

A disulfide bond between Cys227 and Cys232 cross-links β1 and β2. Because disulfides in cellular proteins of crenarchaeal viruses can enhance thermostability (17a, 17b, 17c, 21a), we determined the melting temperatures (Tm) of full-length IntSSV in the absence and presence of a reducing agent using differential scanning fluorometry and found a ΔTm of 9°C (see Fig. S1 in the supplemental material). As they are the only cysteines in IntSSV, the increased thermostability is likely due to this disulfide, sug-

Received 2 March 2012 Accepted 8 May 2012
Published ahead of print 16 May 2012
Address correspondence to C. Martin Lawrence, lawrence@chemistry.montana.edu.
Supplemental material for this article may be found at http://jvi.asm.org/
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JVI.00547-12

Published ahead of print on September 23, 2017 by guest
gesting that introduction of Int\textsuperscript{SSV}-like disulfides into other recombinases might enhance their many applications in biotechnology.

A structural homology search with the program DALI (16) (Table 1) identifies greatest similarity to the catalytic domain of bacteriophage HP1 integrase (15), which also contains a catalytic core that is minimal relative to those of more-elaborate structures, like β-integrase (17), Cre (13), Vibrio cholerae IntI (VcIntI) (20), and eukaryotic Flp (8,10). However, the proposed catalytic Tyr (Tyr\textsubscript{314}) is distant from these first 5 active-site residues and the catalytic tyrosine in HP1, suggesting that like the eukaryotic Flp enzyme, Int\textsuperscript{SSV} may provide the catalytic tyrosine to a neighboring subunit in trans. (B) Superpositional (stereo pair) docking of Int\textsuperscript{SSV} on one arm of the eukaryotic Flp HJ complex (Z score, 9.1) (Table 1) suggests potential interactions for the first 5 active-site residues with the substrate DNA. Again, however, the putative active-site Tyr is distant from the ribose-phosphate backbone and is found on the back face of the subunit.

FIG 1 (A) Int\textsuperscript{SSV} catalytic domain. The structure reveals a minimal integrase fold with similarity to that of bacterial HP1 (Dali Z score, 10.6) (Table 1). Superposition on HP1 indicates that 5 (Arg\textsubscript{211}, Lys\textsubscript{278}, His\textsubscript{281}, Arg\textsubscript{304}) of 6 hallmark active-sites residues correspond to Arg\textsubscript{211}, Arg\textsubscript{240}, Lys\textsubscript{278}, His\textsubscript{281}, Arg\textsubscript{304} (His/Trp). However, the proposed catalytic Tyr (Tyr\textsubscript{314}) is distant from these first 5 active-site residues and the catalytic tyrosine in HP1, suggesting that like the eukaryotic Flp enzyme, Int\textsuperscript{SSV} may provide the catalytic tyrosine to a neighboring subunit in trans. (B) Superpositional (stereo pair) docking of Int\textsuperscript{SSV} on one arm of the eukaryotic Flp HJ complex (Z score, 9.1) (Table 1) suggests potential interactions for the first 5 active-site residues with the substrate DNA. Again, however, the putative active-site Tyr is distant from the ribose-phosphate backbone and is found on the back face of the subunit.

### TABLE 1 Int\textsuperscript{SSV} structural superpositions

<table>
<thead>
<tr>
<th>Tyr recombinase</th>
<th>PDB ID</th>
<th>Z score</th>
<th>RMSD (Å)</th>
<th>No. of equivalent residues (of 159 in Int\textsuperscript{SSV})</th>
<th>% identity to equivalent residues</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1 integrase</td>
<td>1AIH</td>
<td>10.6</td>
<td>2.6</td>
<td>121</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Flp recombinase</td>
<td>1M6X</td>
<td>9.1</td>
<td>3.1</td>
<td>142</td>
<td>9</td>
<td>8, 10</td>
</tr>
<tr>
<td>V. cholera IntI</td>
<td>2A1V</td>
<td>8.9</td>
<td>2.5</td>
<td>121</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>λ integrase</td>
<td>1AE9</td>
<td>8.7</td>
<td>2.9</td>
<td>123</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Cre</td>
<td>3CRX</td>
<td>7.8</td>
<td>3.3</td>
<td>121</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

\( ^a \text{RMSD, root mean square deviation.} \)
studies suggest that these correspond to Arg211 (Argi), Arg240 (Lysβ), Lys278 (Hisii), Arg281 (Argii), Arg304 (His/Trp), and Tyr314 (Tyr) in IntSSV and that 3 of the 6 residues are thus divergent from the consensus motif (18, 24). Importantly, while the superposition of IntSSV on HP1 confirms the predicted structural equivalence for the first 5 residues, the nucleophilic tyrosine present at the N terminus of IntSSV α7 is found on a different face of the domain, displaced from its counterpart in HP1 (Tyr 315) by ~25 Å (see Fig. S2A in the supplemental material).

In contrast, IntSSV α7 does superpose on helix N of the eukaryotic Flp HJ complex (see Fig. S2B in the supplemental material), where the N-terminal end of IntSSV α7 and Tyr314 approach within 8 Å of Flp’s catalytic Tyr343, present at the C terminus of helix M. This suggests that, like Flp, IntSSV may assemble its active site in trans, donating the catalytic Tyr to a neighboring subunit as the synaptosome assembles the expected IntSSV tetramer on target DNA (8, 10). Further support for this comes from the tetrameric bacterial HJ complexes, particularly Cre, where IntSSV superposition on protomer A positions the IntSSV catalytic tyrosine within 4.5 Å of Cre Tyr324 in protomer D (Fig. 2). Similarly, when superposed on subunits A or D of VcIntI, the IntSSV catalytic tyrosine is intermediate between the positions of VcIntI Tyr324 in the D and C subunits. Importantly, this interpretation is also consistent with the mutational analysis of IntSSV by Letzelter et al., who also concluded that DNA cleavage occurs in trans, at least when topoisomerase IB activity was used to report activity (18). Delivery of the active-site tyrosine in trans is an attractive mechanism for controlling the requisite “half-of-sites” reactivity that is integral to DNA integration and excision and appears to be utilized in both the archaeal and eukaryotic domains of life.

A multiple-sequence alignment of 44 IntSSV-like sequences identifies 6 invariant residues (see Fig. S3 in the supplemental material), each within the catalytic domain. Arg211 (Argi), Arg281 (Argii), and the catalytic tyrosine (Tyr314) have been described above; the others are Gly209, which is next to Arg211 and appears to be structural, Lys243, and Tyr325. Though side-chain density for Lys243 is poor, its location suggests that this strictly conserved residue is also delivered in trans to the active site of a neighboring subunit (Fig. 2), where it might participate in DNA recognition or catalysis.

The invariant Tyr325 is found in the C-terminal half of helix α7, which occupies a structurally conserved cleft (Fig. 3). One wall of this cleft is formed by portions of strands β2 and β3, a second by helices α5 and α6, and a third by the C terminus of α2. As the α6-α7 loop leaves the active site, it makes a reverse turn that places the α7 helix on the back side of the α5 and α6 helices, filling this cleft. The importance of this interaction is highlighted by Tyr325, which forms a hydrogen bond that caps the C-terminal end of helix α2, helping to anchor α7 within the structurally conserved cleft.

Returning to the bacterial complexes, when IntSSV is superpositioned on the tetrameric Cre HJ complex, we find that αN from a neighboring Cre subunit is superposed on IntSSV α7 (Fig. 3). Thus, at the quaternary level, this extension from the neighboring Cre subunit is accommodated by the same structurally conserved cleft that we see in IntSSV and Flp. Similarly, C-terminal extensions from a neighboring subunit in the λ integrase and VcIntI HJ complexes also position within this cleft, as does HP1 α9 when IntSSV is superpositioned on the HP1 dimer in the absence of DNA. Thus, considering tertiary- and quaternary-level structures, these cleft-helix interactions are conserved across the three domains of life, although the lengths and conformations of the helices do vary. This helix swapping is directly analogous to domain swapping, a generally
recognized mechanism for forming oligomeric proteins from individual subunits (5, 6). From this point of view, it is actually the bacterial tyrosine recombinases, as opposed to IntSSV and Flp, that are helix or domain swapped. Overall, donation of the active-site tyrosine in trans and the lack of helix swapping suggest that IntSSV is more similar to the eukaryotic rather than the bacterial enzymes. This is consistent with the observation that archaeal enzymes in nucleic acid metabolism are frequently more similar to their eukaryotic than to their bacterial counterparts (2). However, the minimal sequence identity between IntSSV and the bacterial and eukaryotic enzymes also suggests that IntSSV and the fuselloviridae in general are deeply rooted within the archaeal domain.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (grant MCB-0920312). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory.

REFERENCES

1 and extension to hyperthermophilic viruses in general. Virology. 376: 270–278.


