SUPPLEMENTARY MATERIAL

for

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[1,2], Mark J. Young[1,3,4], and C. Martin Lawrence[1,2*]

[1]Thermal Biology Institute, [2]Department of Chemistry and Biochemistry, [3]Department of Plant Sciences and Plant Pathology and [4] Department of Microbiology, Montana State University, Bozeman, MT 59717, USA

Supplemental Methods

Int^{SSV} Structure Determination. Int^{SSV} was cloned into pDEST14 (Invitrogen) as previously described (8), yielding the expression vector pEXP14-Int^{SSV}. Internal forward and reverse primers were GTGATGGTGATGGTGATGATGACGAAAGATAAGAC and CAAGAAAGCTGGGTCCTAGACCCCTTTTAGCCATT, respectively. External primers were reported previously (11), as was the protein expression protocol with the following modifications (10). Lysis buffer was 20 mM Tris, 1M NaCl, 10 mM imidazole, pH 8.0, the Ni-NTA wash buffer was 20 mM Tris, 1 M NaCl, 20 mM imidizole, pH 8.0, Ni-NTA elution buffer was 10 mM Tris (pH 8.0), 300 mM NaCl and 200 mM imidizole, and size exclusion chromatography buffer was 10 mM Tris (pH 8.0), 20mM $NaPO_4$ and 300 mM NaCl. Purified D335 was dialyazed against 10 mM Tris (pH 8.0), 150 mM NaCl, 5mM $NaPO_4$ and concentrated to 10 mg/ml. The protein was crystallized by hanging drop vapour-diffusion at 22 ^oC using 2 µl of SSV^{Int} and 2µl of well solution (15-20% PEG 3350, 0-1% Tryptone, and 0.1 M HEPES at pH 7.0). Crystals up to 0.20 x 0.20 x 0.05 mm in size were obtained in 7-10 days time. Native and a three-wavelength anomalous diffraction dataset centered on the Se-K edge were collected at the Stanford Synchrotron Radiation Laboratory (beamline 9-2). Data were processed in space group $P6_522$ with HKL2000 (14). Crystal parameters, data and model quality are presented in Supplemental Tables 1 and 2. PHENIX (1) was used to identify the selenium atom substructure and to calculate initial phases. The structure was completed with ARP/WARP (9), Coot (6) REFMAC5 (12, 15) and Molprobity (4), and deposited in the Protein Data Bank (3UXU). Figures were generated with PYMOL (5)

Supplemental Tables

Supplemental Table 1. Data collection

Data Set	Se-Edge	Se-Peak	Se-Remote	Native
Wavelength (Å)	0.97939	0.97891	0.91837	0.84917
Space Group	$P6_522$			
Cell Constants (a,b,c; Å) α=90, β=90, γ =120 ^{o}	74.24, 74.24, 176.10			73.96, 73.96, 176.25
Resolution Range[a] (Å)	50-3.2 (3.31-3.2)	50-2.8 (2.91-2.80)	50-3.1 (3.21-3.1)	50-2.7 (2.75-2.70)
Unique Reflections[a]	5,162	7,561	5,662	8,414
Average Redundancy[a]	10.1(10.5)	20.1(21.2)	10.1(10.2)	13.5(13.5)
I/σ[a]	39.7(5.3)	39.3(10.2)	45.6(6.7)	58.3(6.5)
Completeness (%)	(99.6)	98.1(99.7)	99.5 (100)	99.9(100)
R_{sym}[a,b] (%)	6.0 (38.5)	8.8 (31.3)	5.1 (30.0)	4.3 (31.5)

[a]Numbers in parenthesis refer to the highest resolution shell.

[b]$R_{sym}=100*\Sigma_h\Sigma_i|I_i(h)-<I(h)>|/\Sigma_h I(h)$ where $I_i(h)$ is the i^{th} measurement of reflection h and <I(h)> is the average value of the reflection intensity.

Supplemental Table 2. Model Refinement

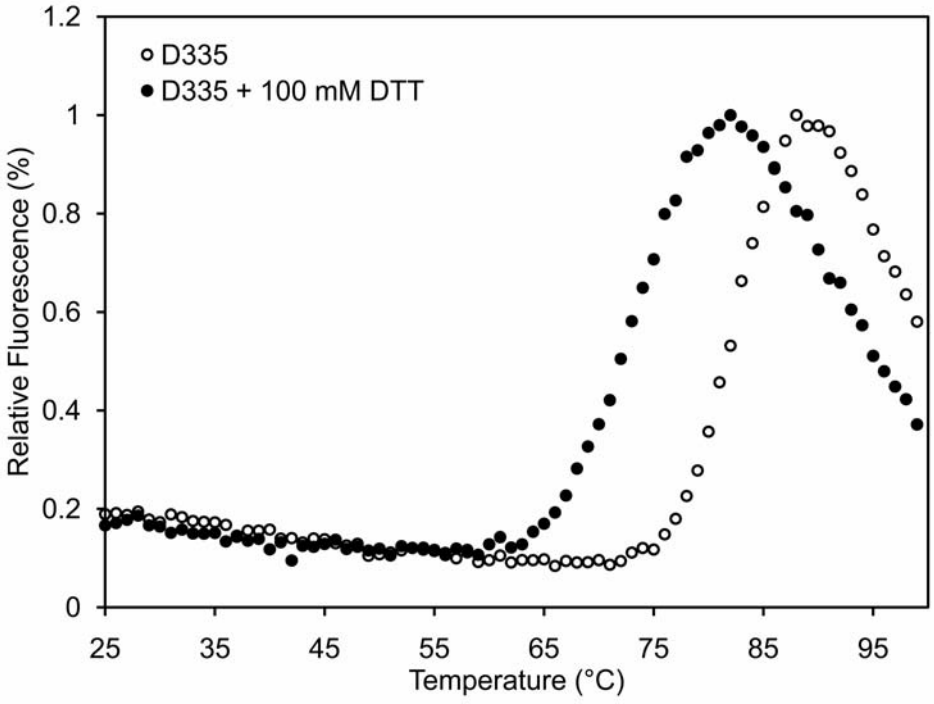
R_{work}[c] (%)	19.0 (25.1)
R_{free}[c] (%)	21.4 (24.7)
Real Space CC[d] (%)	95.9
Mean B Value (overall; Å^2)	24.9
Coordinate Error (based on maximum likelihood, Å)	0.177
RMSD from ideality:	
Bonds (Å)	0.016
Angles (°)	1.445
Ramachandran Plot[e]:	
Most Favored (%)	98.1
Additional Allowed (%)	1.9
PDB Accession Code	3UXU

[c]$R_{work} = \Sigma||F_o|-F_c||/\Sigma F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes used in refinement. R_{free} is calculated as R_{work}, but using the "test" set of structure factor amplitudes that were withheld from refinement (4.9%).

[d]Correlation coefficient (CC) is agreement between the model and $2mF_o$-DF_c electron density map.

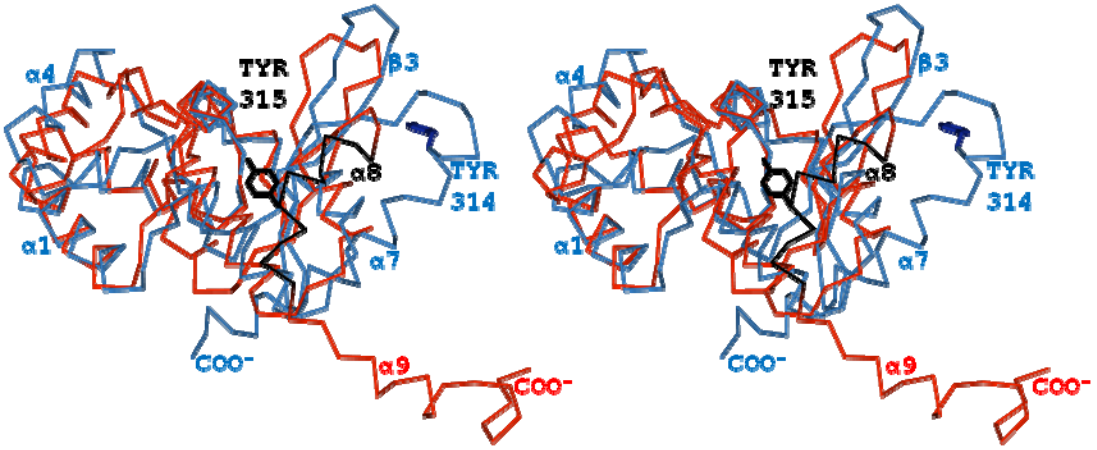
[e]Calculated using Molprobity (4)

Supplemental Figures

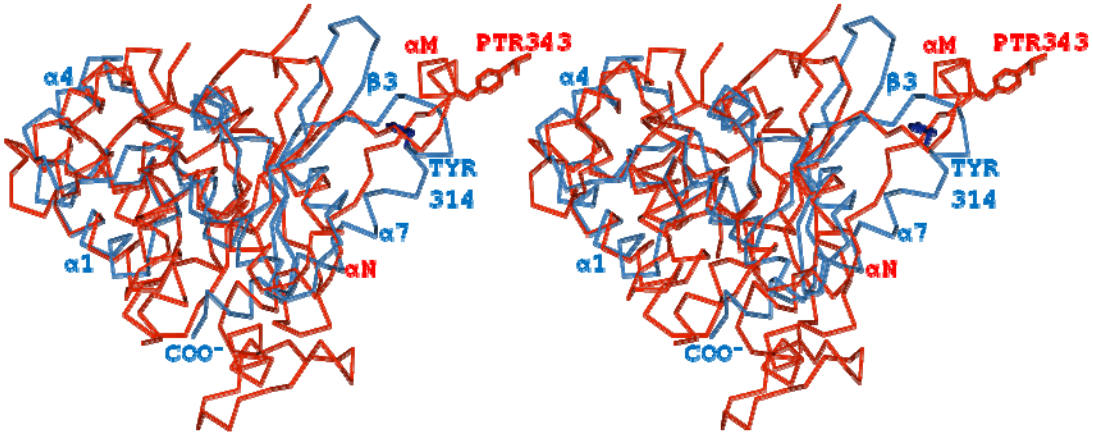


Supplemental Figure 1. Thermostability of full-length Int^{SSV}. The thermostability of freshly purified full length Int^{SSV} was examined using differential scanning fluorometry (13). In the absence of reducing agent (open circles), the inflection point of the fluorescent melt curve indicates a T_m of 82 ^oC, while in 100 mM DTT (closed circles) this fell to 73 ^oC, a ΔT_m of 9 ^oC by differential scanning fluorometry. Because they are the only cysteines in the full length Int^{SSV} protein, the difference in thermostability is probably due to the disulfide bond between Cys227 and Cys232. Introducing IntSSV-like disulfide bonds at equivalent positions in other tyrosine recombinases could increase their stability and thus enhance their many applications in biotech.

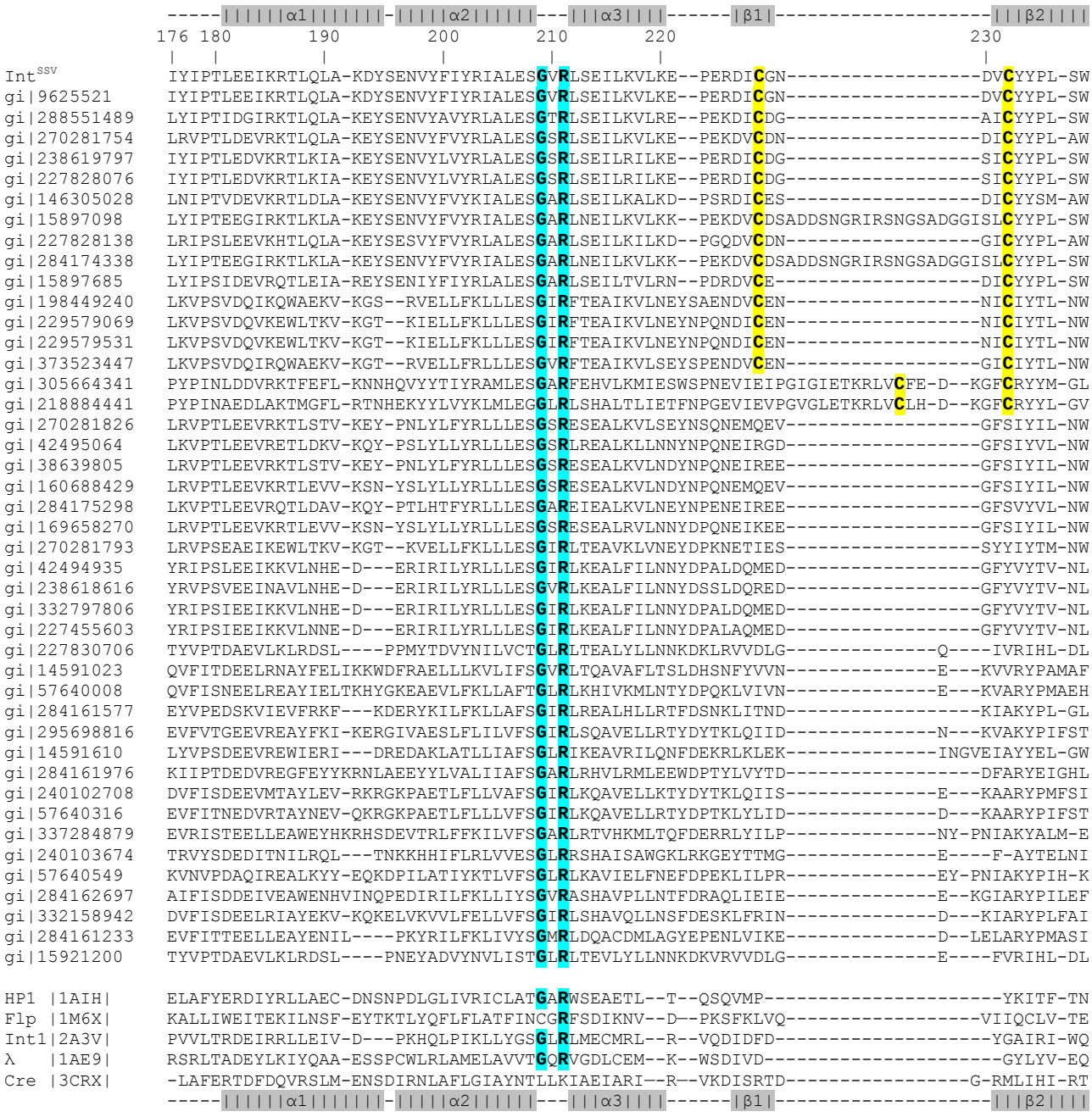
A



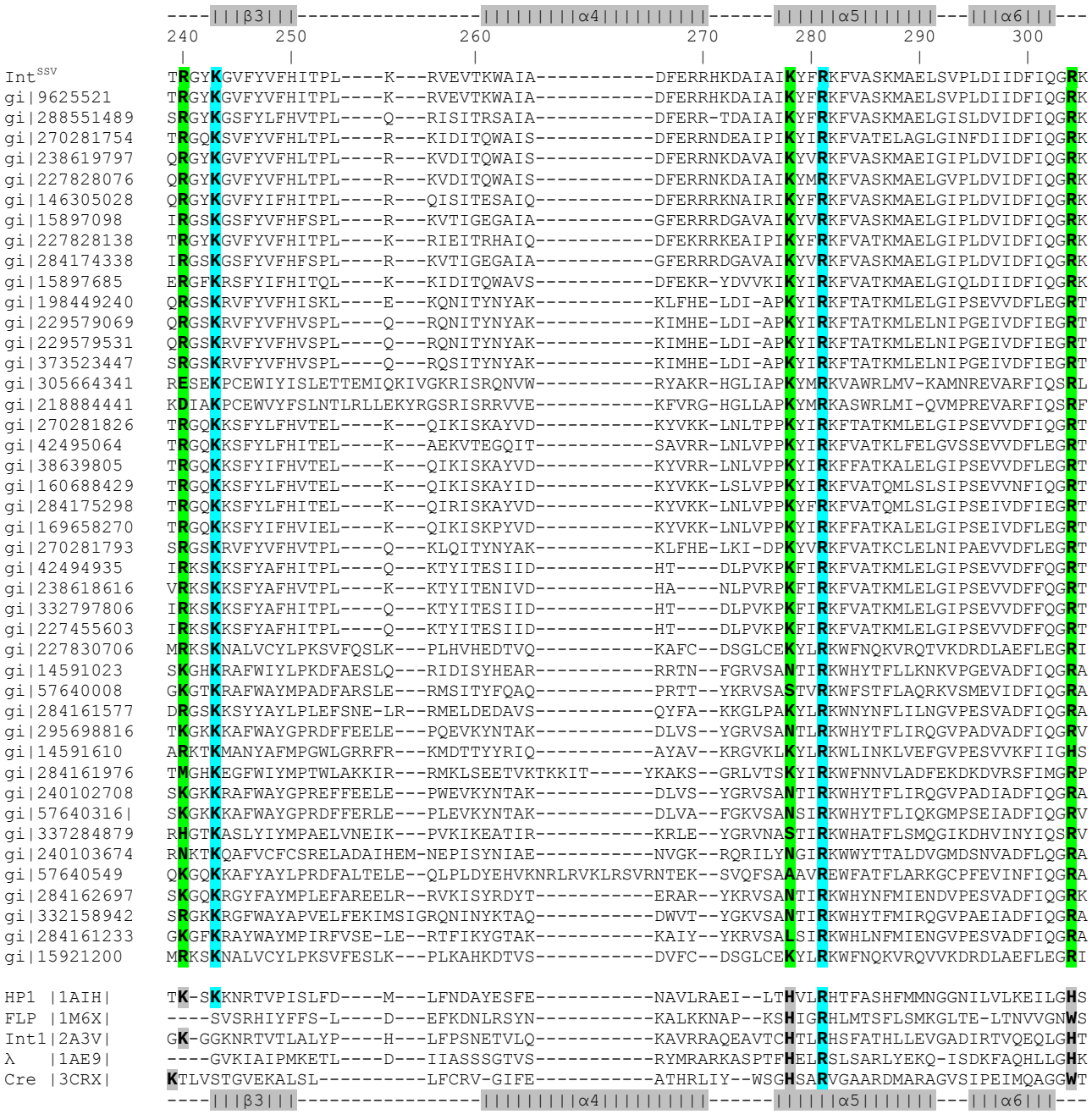
B



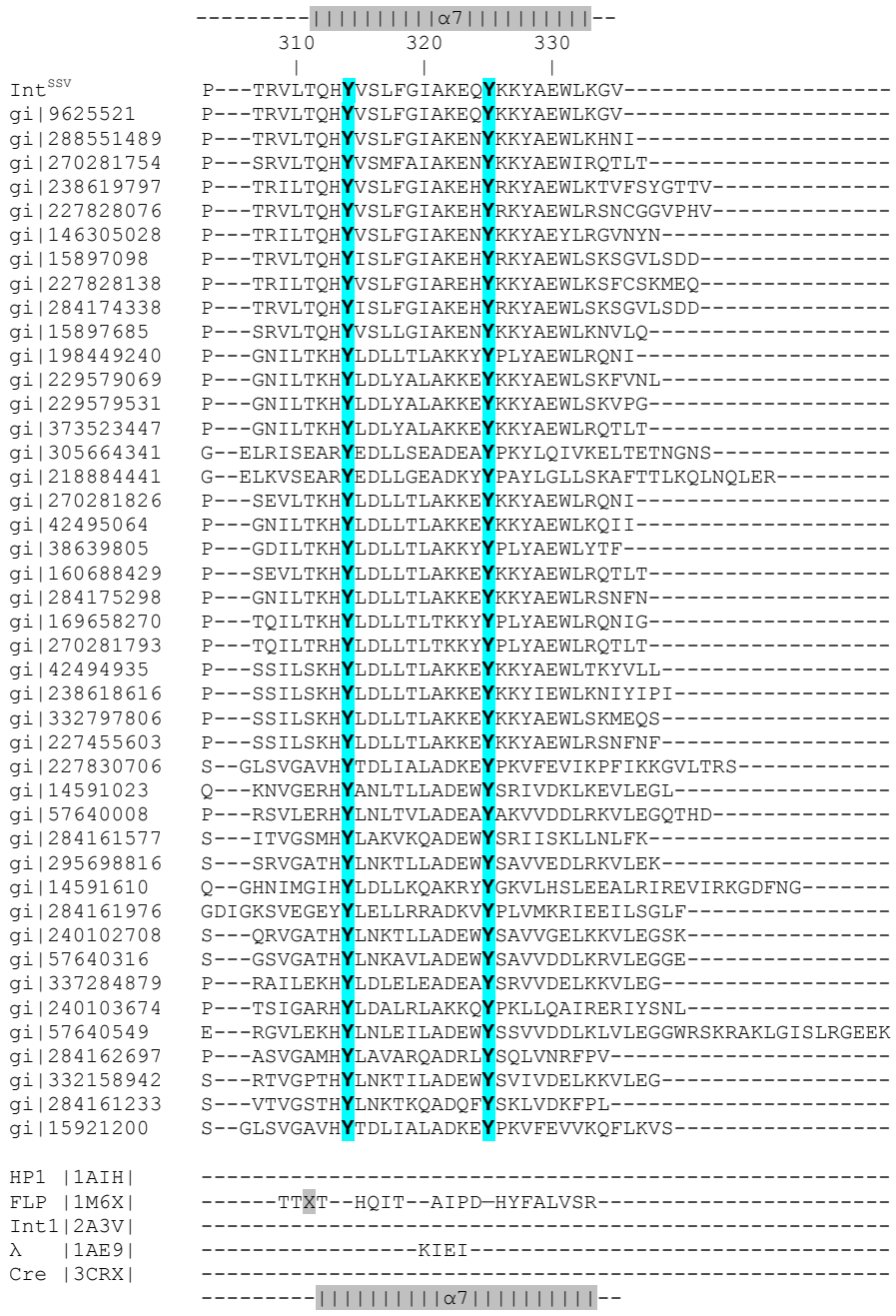
Supplemental Figure 2. A) Int^{SSV} superpositioned on HP1. The orientation is identical to that in Figure 1. Int^{SSV} is shown as a C_α-trace in blue and HP1 as a C_α-trace in red, except for HP1 helix α8, which is highlighted in black. The catalytic tyrosine residues [Int^{SSV} Tyr314 (blue), HP1 Tyr315 (black)] are also shown, and fall on different faces of the domain. The C_α-$C_\alpha$distance between the tyrosines is 25 Å. B) Int^{SSV} superpositioned on Flp. Int^{SSV} is in blue and Flp in red. The nucleophilic tyrosine residues (Int^{SSV} TYR314 and Flp PTR343), which in the Flp structure is was caputured as a phosphotyrosine intermediate (3), fall on the same face of the domain with the C_α-$C_\alpha$distance falling to 8 Å.



Supplemental Figure 3. Sequence alignment for the Int^{SSV} catalytic domains (3 pages). The secondary structural elements in Int^{SSV} are depicted above the sequence of the Int^{SSV}. A BLAST search (2) seeded with Int^{SSV} identified 43 additional Int^{SSV}-like sequences that were subsequently aligned with ClustalW (7, 16). Below these, ungapped structure based sequence alignments with selected tyrosine recombinase structures (HP1, Flp, Int1, λ-integrase and Cre) are also presented. The Clustal W alignment of the 44 Int^{SSV}-like sequences identified only 6 invariant residues, each of which fell within the catalytic domain. The strictly conserved residues are Gly209, Arg211 (ArgI), Lys243, Arg281 (ArgII), the catalytic tyrosine (Tyr314), and Tyr325, each of which is in blue. Additional active sites residues [Arg240 (Lysβ), Lys278 (HisII) and Arg304 (His/Trp)] are shown in green, and the cysteine pair forming a putative disulfide bond in yellow. Seventeen of the 44 Int^{SSV}-like sequences contain Cys at position 232. When Cys232 is present, a preceding cysteine by 5, 7 or 22 residues is also found. In contrast, when Cys232 is absent, so too is the preceding cysteine. Thus, while the cysteine residues are not strictly conserved, their presence is strictly correlated and consistent with a role in disulfide enhanced thermostability.



Supplemental Figure 3. Continued.



Supplemental Figure 3. Concluded.

Supplemental References

1. **Adams, P. D., P. V. Afonine, G. BunkOczi, V. B. Chen, I. W. Davis, D. N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart.** 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr **D66:**213-221.
2. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-3402.
3. **Conway, A. B., Y. Chen, and P. A. Rice.** 2003. Structural plasticity of the Flp-Holliday junction complex. J Mol Biol **326:**425-34.
4. **Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson, and D. C. Richardson.** 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res **35:**W375-83.
5. **DeLano, W. L.** 2002. The PyMOL Molecular Graphics System. http://www.pymol.org.
6. **Emsley, P., and K. Cowtan.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D **60:**2126-2132.
7. **Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson.** 1998. Multiple sequence alignment with Clustal x. Trends in Biochemical Sciences **23:**403-405.
8. **Kraft, P., D. Kummel, A. Oeckinghaus, G. H. Gauss, B. Wiedenheft, M. Young, and C. M. Lawrence.** 2004. Structure of D-63 from *Sulfolobus* Spindle-Shaped Virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J. Virol. **78:**7438-7442.

9. **Langer, G., S. X. Cohen, V. S. Lamzin, and A. Perrakis.** 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc **3:**1171-9.

10. **Lintner, N. G., K. A. Frankel, S. E. Tsutakawa, D. L. Alsbury, V. Copie, M. J. Young, J. A. Tainer, and C. M. Lawrence.** 2011. The Structure of the CRISPR-Associated Protein Csa3 Provides Insight into the Regulation of the CRISPR/Cas System. J Mol Biol **405:**939-55.

11. **Menon, S. K., B. J. Eilers, M. J. Young, and C. M. Lawrence.** 2010. The crystal structure of D212 from sulfolobus spindle-shaped virus ragged hills reveals a new member of the PD-(D/E)XK nuclease superfamily. J Virol **84:**5890-7.

12. **Murshudov, G. N., A. A. Vagin, and E. J. Dodson.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst. **D53:**240-255.

13. **Niesen, F. H., H. Berglund, and M. Vedadi.** 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc **2:**2212-21.

14. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode, p. 307-326. *In* C. Carter and R. Sweet (ed.), Macromolecular Crystallography, Part A, vol. 276. Academic Press, New York, NY.

15. **Painter, J., and E. A. Merritt.** 2006. TLSMD web server for the generation of multi-group TLS models. Journal of Applied Crystallography **39:**109-111.

16. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22:**4673-4680.