

either during ontogeny or phylogenetic diversification, small deviations from quarter-power scaling sometimes occur (3, 23). However, when body sizes vary over many orders of magnitude, these scaling laws are obeyed with remarkable precision. Moreover, the predicted scaling properties do not depend on most details of system design, including the exact branching pattern, provided it has a fractal structure (24). Significantly, nonfractal systems, such as combustion engines and electric motors, exhibit geometric (third-power) rather than quarter-power scaling (1). Because the fractal network must still fill the entire D -dimensional volume, our result generalizes to $a = D/(D + 1)$. Organisms are three-dimensional, which explains the 3 in the numerator of the $3/4$ power law, but it would be instructive to examine nearly two-dimensional organisms such as bryozoans and flatworms. The model can potentially explain how fundamental constraints at the level of individual organisms lead to corresponding quarter-power allometries at other levels. The constraints of body size on the rates at which resources can be taken up from the environment and transported and transformed within the body ramify to cause quarter-power scaling in such diverse phenomena as rate and duration of embryonic and postembryonic growth and development, interval between clutches, age of first reproduction, life span, home range and territory size, population density, and maximal population growth rate (1–3). Because organisms of different body sizes have different requirements for resources and operate on different spatial and temporal scales, quarter-power allometric scaling is perhaps the single most pervasive theme underlying all biological diversity.

REFERENCES AND NOTES

1. T. A. McMahon and J. T. Bonner, *On Size and Life* (Scientific American Library, New York, 1983); J. T. Bonner, *The Evolution of Complexity by Means of Natural Selection* (Princeton Univ. Press, Princeton, NJ, 1983); J. H. Brown, *Macroecology* (Univ. of Chicago Press, Chicago, 1995).
2. K. Schmidt-Nielsen, *Scaling: Why Is Animal Size so Important?* (Cambridge Univ. Press, Cambridge, 1984); W. A. Calder III, *Size, Function and Life History* (Harvard Univ. Press, Cambridge, MA, 1984).
3. R. H. Peters, *The Ecological Implications of Body Size* (Cambridge Univ. Press, Cambridge, 1983).
4. H. A. Feldman and T. A. McMahon, *Respir. Physiol.* **52**, 149 (1983).
5. T. A. McMahon, *Science* **179**, 1201 (1973).
6. M. R. Patterson, *ibid.* **255**, 1421 (1992).
7. K. J. Niklas, *Plant Allometry: The Scaling of Form and Process* (Univ. of Chicago Press, Chicago, 1994); *Am. J. Bot.* **81**, 134 (1994).
8. A. M. Hemmingsen, *Rep. Steno Mem. Hosp. (Copenhagen)* **4**, 1 (1950); *ibid.* **9**, 1 (1960).
9. B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1977).
10. D'A. W. Thompson, *On Growth and Form* (Cambridge Univ. Press, Cambridge, 1942).
11. K. Shinozaki et al., *Jpn. J. Ecol.* **14**, 97 (1964); *ibid.*, p. 133; see also M. H. Zimmerman, *Xylem Structure and the Ascent of Sap* (Springer-Verlag, Berlin, 1983); M. T. Tyree and F. W. Ewers, *New Phytol.* **119**, 345 (1991).
12. The branching of a vessel at level k into n_k smaller vessels (Fig. 1) is assumed to occur over some small, but finite, distance that is much smaller than either l_k or l_{k+1} . This relation is similar to that assumed in the Strahler method [A. N. Strahler, *Trans. Am. Geophys. Union* **34**, 345 (1953); (11, 21)]. A generalization to nonuniform branching, where the radii and lengths at a given level may vary, is straightforward.
13. Normalization factors, such as M_0 , will generally be suppressed, as in Eq. 1. In general, all quantities should be expressed in dimensionless form; note, however, that this does not guarantee that they are size independent and scale as M^0 . For example, the Womersley number, α of Eq. 8, although dimensionless, scales as $M^{1/4}$.
14. This formula is not valid for plant vessel bundles because plants are composed of multiple parallel vessel elements. Their resistance is given by $Z = 8\mu l / N_c \pi r_c^4$, where l is the length of a single vessel element, r_c is its radius, and N_c is their total number.
15. This relation holds for plant vessels from the roots to the leaves, but not within leaves [M. J. Canney, *Philos. Trans. R. Soc. London Ser. B* **341**, 87 (1993)].
16. A. Krogh, *Pfluegers Arch. Gesamte Physiol. Menschen Tiere* **179**, 95 (1920).
17. C. D. Murray, *Proc. Natl. Acad. Sci. U.S.A.* **12**, 207 (1926).
18. C. G. Caro et al., *The Mechanics of Circulation* (Oxford Univ. Press, Oxford, 1978).
19. J. R. Womersley, *Philos. Mag.* **46**, 199 (1955); *J. Physiol. (London)* **127**, 553 (1955).
20. See, for example, A. S. Iberall, *Math. Biosci.* **1**, 375 (1967) and T. F. Sherman, *J. Gen. Physiol.* **78**, 431 (1981), which contain summaries of earlier data; also M. Zamir et al., *J. Biomech.* **25**, 1303 (1992) and J. K.-J. Li, *Comparative Cardiovascular Dynamics of Mammals* (CRC Press, Boca Raton, FL, 1996). Care must be taken in comparing measurements with prediction, particularly if averages over many successive levels are used. For example, if $A_k = \sum_k \pi r_k^2$ is the total cross-sectional area at level k , then for the aorta and major arteries, where $k < \bar{k}$ and the branching is area-preserving, we predict $A_0 = A_k$. Suppose, however, that the first K levels are grouped together. Then, if the resulting measurement gives \bar{A}_K , area-preserving predicts $\bar{A}_K = KA_0$ (but not $\bar{A}_K = A_0$). It also predicts $r_0^3 \approx n^{1/2} \sum N_k r_k^3$. Using results from M. LaBarbera [*Science* **249**, 992 (1990)], who used data averaged over the first 160 vessels (approximately the first 4 levels), gives, for human beings, $A_0 \approx 4.90 \text{ cm}^2$, $\bar{A}_K \approx 19.98 \text{ cm}^2$, $r_0^3 \approx 1.95 \text{ cm}^3$, and $\sum N_k r_k^3 \approx 1.27 \text{ cm}^3$, in agreement with area preservation. LaBarbera, unfortunately, took the fact that $\bar{A}_K \neq A_0$ and $r_0^3 \approx \sum N_k r_k^3$ as evidence for cubic rather than area-preserving branching. For small vessels, where $k > \bar{k}$, convincing evidence for the cubic law can be found in the analysis of the arteriolar system by M. L. Ellsworth et al., *Microvasc. Res.* **34**, 168 (1987).
21. Y. C. Fung, *Biodynamics* (Springer-Verlag, New York, 1984).
22. P. Gehr et al., *Respir. Physiol.* **44**, 61 (1981).
23. P. M. Bennett and P. H. Harvey, *J. Zool.* **213**, 327 (1987); A. F. Bennett, *Am. Zool.* **28**, 699 (1988); P. H. Harvey and M. D. Pagel, *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
24. This is reminiscent of the invariance of scaling exponents to details of the model that follow from renormalization group analyses, which can be viewed as a generalization of classical dimensional analysis.
25. J.H.B. is supported by NSF grant DEB-9318096, B.J.E. by NSF grant GER-9553623 and a Fulbright Fellowship, and G.B.W. by the Department of Energy.

13 September 1996; accepted 12 February 1997

Flexibility in DNA Recombination: Structure of the Lambda Integrase Catalytic Core

Hyock Joo Kwon, Radhakrishna Tirumalai, Arthur Landy,*
Tom Ellenberger*

Lambda integrase is archetypic of site-specific recombinases that catalyze intermolecular DNA rearrangements without energetic input. DNA cleavage, strand exchange, and religation steps are linked by a covalent phosphotyrosine intermediate in which Tyr³⁴² is attached to the 3'-phosphate of the DNA cut site. The 1.9 angstrom crystal structure of the integrase catalytic domain reveals a protein fold that is conserved in organisms ranging from archaeobacteria to yeast and that suggests a model for interaction with target DNA. The attacking Tyr³⁴² nucleophile is located on a flexible loop about 20 angstroms from a basic groove that contains all the other catalytically essential residues. This bipartite active site can account for several apparently paradoxical features of integrase family recombinases, including the capacity for both cis and trans cleavage of DNA.

The integrase protein (Int) of *Escherichia coli* phage lambda (λ) belongs to a large family of site-specific DNA recombinases from archaeobacteria, eubacteria, and yeast (1–3) that catalyze rearrangements be-

tween DNA sequences with little or no sequence homology to each other (4–8). Like λ Int, many of these recombinases function in the integration and excision of viral genomes into and out of the chromosomes of their respective hosts. Others function in the decatenation or segregation of newly replicated chromosomes, conjugative transposition, regulation of plasmid copy number, or expression of cell surface proteins. Integrase family members have the distinctive ability to carry out a complete site-specific recombination reac-

H. J. Kwon and T. Ellenberger, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston MA 02115, and the Graduate Program in Biophysics, Harvard University, Cambridge, MA 02138, USA.

R. Tirumalai and A. Landy, Division of Biology and Medicine, Brown University, Providence, RI 02912, USA.

*Corresponding authors.

tion between two DNAs in the absence of high energy cofactors. DNA cleavage and rejoining is accomplished in two steps. First, a tyrosine hydroxyl attacks the scissile phosphate, nicking the DNA and forming a 3' phosphotyrosine-linked DNA complex. This covalent protein-DNA intermediate is resolved when the 5' terminal hydroxyl of the invading DNA strand attacks the phosphotyrosine linkage and displaces the protein, forming a Holliday junction. The reaction is repeated for the other strand of each DNA partner, generating the recombinant DNA duplexes. It is the transient covalent linkage of protein and DNA that conserves the energy of the broken phosphodiester bond, enabling a pair of reciprocal strand exchanges to proceed.

The 40-kD Int protein, which was first purified by Kikuchi and Nash (9), has been separated into discrete domains by limited proteolysis (10, 11). On the basis of these results, a minimal catalytic domain, termed Int c170 (residues 170 to 356), was cloned and purified. The c170 domain is approximately the same size as the smallest Int family members; it can cleave and ligate DNA and it functions as a type I topoisomerase (11). This catalytic domain encompasses both of the highly conserved sequence motifs of the integrase family (1, 2), including an invariant arginine-histidine-arginine catalytic triad, and the tyrosine nucleophile (12). Amino acid substitutions within the catalytic triad severely impair phage recombination, and they compromise DNA-binding and cleavage *in vitro*. The results of mutational analyses (13–15) and phosphate interference footprinting (16) of various recombinases belonging to the integrase family are consistent with the suggestion (17) that the Arg-His-Arg triad activates the scissile phosphate for cleavage of DNA. The location of the attacking tyrosine is carboxyl terminal to the catalytic triad in a less conserved region that is rich in acidic residues. Despite their sequence similarity, Int family proteins exhibit a remarkable mechanistic duality. In some cases, the tyrosine nucleophile of one subunit cleaves DNA that is bound by an adjacent subunit (trans cleavage) (17), whereas in other cases the tyrosine cleaves the site bound by the same subunit (cis cleavage) (18). These distinctive modes of DNA cleavage by conserved catalytic residues may reflect intrinsic differences among Int family proteins, or the tailoring of the cleavage mechanism in response to reaction conditions.

The crystal structure of the Int catalytic domain was determined by multiple isomorphous replacement and refined at 1.9 Å resolution (Table 1). Int c170 has a mixed α - β structure consisting of seven α helices

and seven β strands (Fig. 1). An α -helical bundle with an unusual packing geometry is cradled by two antiparallel β hairpins, which together form a globular structure of roughly 25 Å by 38 Å by 50 Å. Helices α A, α B, and α C form the core of the domain and they are circumscribed by a helical collar consisting of α D, α E, and α F. The packing angles between these helices are irregular and apparently unprecedented; an automated search of structures in the Protein Data Bank (19) did not identify any proteins with a similar fold. More specifically, the Int c170 domain does not resemble the catalytic domains of other site-specific recombinases of known structure, including the HIV-1 and ASV retroviral integrases (20, 21), MuA transposase (22), and $\gamma\delta$ resolvase (23). The Int catalytic domain also differs from those of topoisomerase I and topoisomerase II, both of which have buried tyrosine nucleophiles (24, 25).

The base of the molecule is formed by a four-stranded mixed β sheet (β 1, β 2, β 3, β 7) with two β -hairpin extensions (β 2- β 3 and β 4- β 5) that wrap around the α -helical bundle. The connection between α G and β 6 (residues Lys³³⁴ to Gln³⁴¹) is disordered in both molecules of the crystallographic asymmetric unit, implying that this segment is flexible (Fig. 1B). In one Int c170 protomer the catalytic Tyr³⁴² (phenylalanine in the crystal structure) is adjacent to a short β strand (β 6) that is hydrogen-bonded to the COOH-terminal strand β 7. Tyr³⁴² is not visible in the electron density of the other protomer, and the disordered (flexible) segment extends from Lys³³⁴ to Trp³⁵⁰. In both molecules strand β 7 is hydrogen-bonded in a parallel orientation to strand β 3 at one edge of the β sheet. It is fixed in this orientation by several buried hydrophobic residues including Trp³⁵⁰ and Ile³⁵³. These contacts anchor β 7 to the rest of the

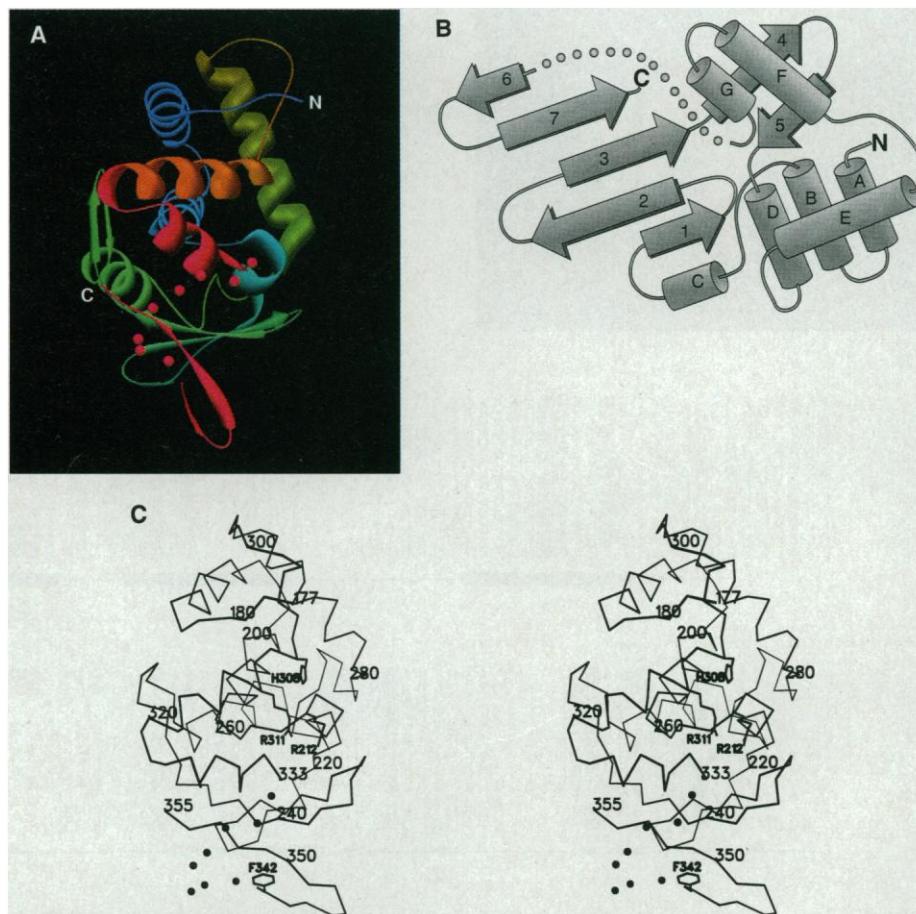


Fig. 1. The lambda integrase catalytic domain is a seven-helix bundle cradled by a β sheet and two β hairpins. The ribbon diagram (A) is colored with a gradient from the NH₂-terminus (blue) to the COOH-terminus (red) with the program SETOR (58). (B) Topology diagram of the Int catalytic domain with the flexible loop connecting helix G and strand β 6 depicted by spheres. The α helices of Int c170 are intertwined in an irregular packing arrangement that, to our knowledge, has not been seen before. (C) The catalytically essential residues Arg²¹², His³⁰⁰, and Arg³¹¹ face a small depression in the protein surface, located far from the attacking nucleophile Tyr³⁴². Tyr³⁴² is at the junction of the flexible loop (spheres) and strand β 6, which forms a hairpin structure with β 7 that is, in turn, anchored to the protein core by hydrophobic and electrostatic interactions.

protein, thereby defining the distal boundary of the flexible loop and restricting the range of motion of the attacking nucleophile Tyr³⁴².

The amino acid sequences of 66 members of the integrase family were aligned with a tree-based algorithm (26) and several striking features emerge (Fig. 2B) (27). First, the invariant residues of the integrase family (1, 2) are all located on the proposed DNA interaction surface of λ Int (Fig. 2, A and B) and many of these residues are essential for catalytic function. In contrast,

few of the exposed residues away from the enzyme active site are conserved in the integrase family. Furthermore, it is found that many of the buried residues of Int c170 are conserved. In particular, Leu²⁰⁵, Val²⁰⁷, Val²⁰⁸, Leu²¹⁶, Met²¹⁹, Leu²²⁹, Val²³¹, Ile²⁴², Pro²⁴³, Met²⁵⁵, and Leu³³⁰ form a highly conserved core surrounding Thr²⁰⁹. This clustering of conserved, functionally important residues around the active site implies that the catalytic domains of integrase family members have similar folds.

A shallow groove that is approximately 25

Å wide runs along one face of the protein (Figs. 1 and 2). It is circumscribed by α C, α D, α E, α F, and the β 2- β 3 hairpin. Conserved, polar residues facing this groove include Arg²¹², Asp²¹⁵, His³⁰⁸, Arg³¹¹, Ser³¹², Gln³²⁸, and His³³³ (Fig. 2). The catalytic triad of Arg²¹², His³⁰⁸, and Arg³¹¹ are within 7 Å of each other at the base of the groove (Arg²¹² and Arg³¹¹ are within 3.4 Å, and His³⁰⁸ is 6.4 Å from Arg³¹¹ and 7.2 Å from Arg²¹²); His³⁰⁸ and Arg³¹¹ extend from the face of α F (Fig. 1), and Arg²¹² is located at the NH₂-terminus of the short α C helix and it forms a

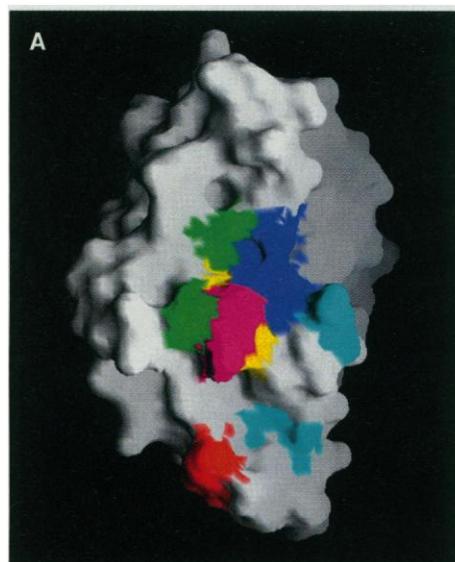
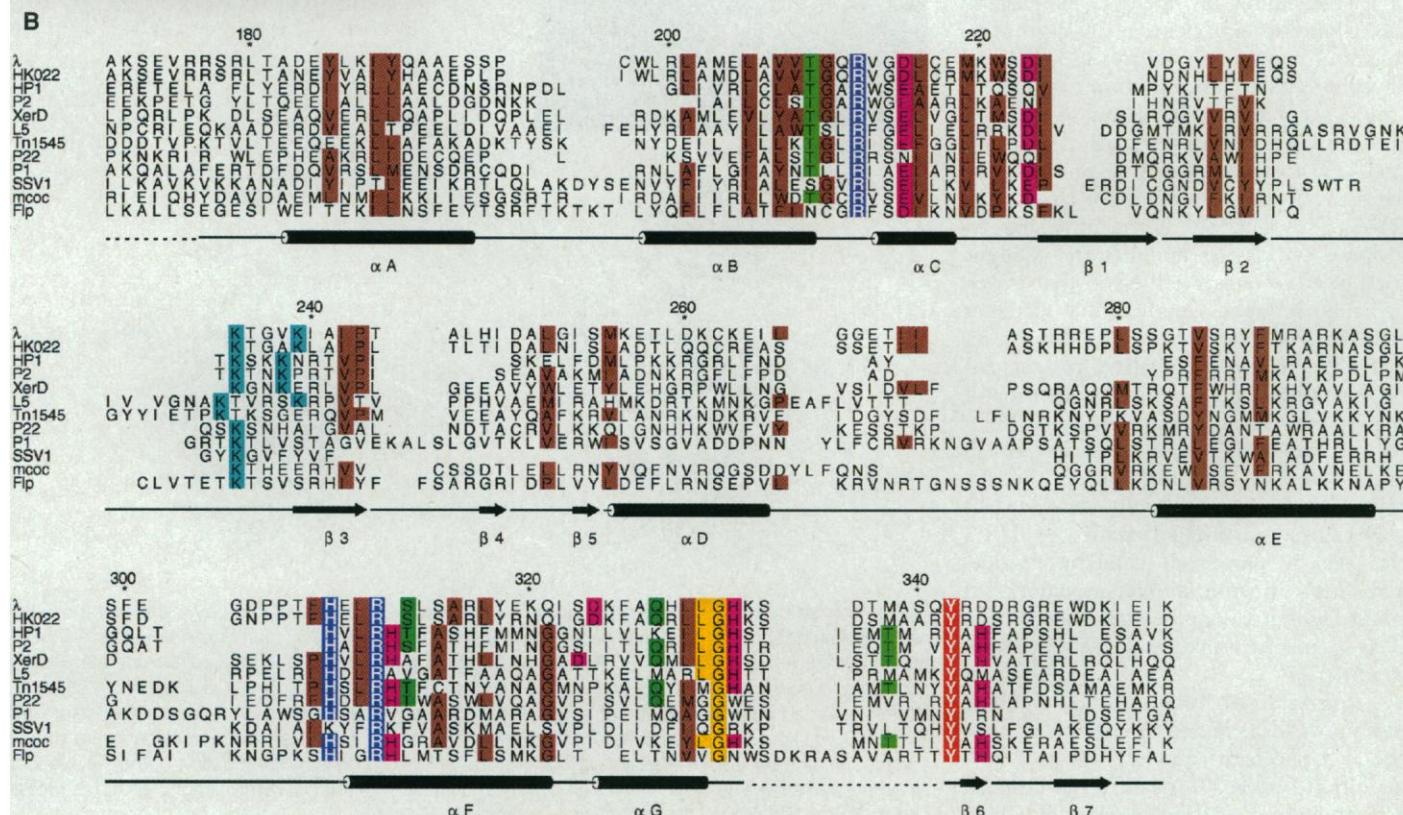


Fig. 2. (A) Conserved residues cluster around the active site pocket of lambda integrase. These include the Arg²¹²-His³⁰⁸-Arg³¹¹ triad (navy), Lys²³⁵ and Lys²³⁹ (cyan), His³³³ (magenta), Ser³¹²-Gln³²⁸ (green), and Leu³³¹-Gly³³² (yellow). Tyr³⁴² (red) is located some distance from the other catalytic residues. **(B)** Sequence alignment of selected integrase family members showing the conserved hydrophobic residues (brown) that form the core of the Int catalytic domain. This conservation of buried residues strongly implies that integrase family recombinases have similar folds. The other highly conserved motifs (1, 2) are predominantly surface residues that cluster around the enzyme active site, and they are color-coded as for (A). The AMPS program suite (26) was used for alignment of integrase sequences, and (B) was prepared with ALSCRIPT (59). Residue numbers refer to λ Int residues and the secondary structure of the Int catalytic domain is shown. The aligned sequences are: λ , phage lambda Int; HK022, phage HK022 Int; HP1, *Haemophilus influenzae* phage HP1 Int; P2, phage P2 Int; XerD, *E. coli* XerD; L5, *Mycobacterium* spp. phage L5 Int; Tn1545, *Streptococcus pneumoniae* transposase Tn1545; P22, phage P22 Int; P1, Cre of phage P1; SSV1, *Sulfolobus* phage SSV1; mccc, *Methanococcus jannaschii* putative Int; Flp, *Saccharomyces cerevisiae* Flp.



water-mediated interaction with Asp²¹⁵. The strict conservation of the Arg-His-Arg triad within the Int family, in conjunction with their demonstrated roles in binding and cleavage of DNA or DNA ligation (13–15, 18), implicate this surface as part of the enzyme active site. Other basic residues facing the groove of Int c170 include Arg¹⁷⁷, Arg¹⁷⁹, Lys²³⁵, Arg²⁸⁷, Arg²⁹¹, Arg²⁹³, Lys²⁹⁴, and Arg³¹⁷, and these bolster the positive electrostatic character of this surface. The shape and charge of this surface of the protein, together with the presence of essential catalytic residues, suggest that this groove is the DNA-binding surface of the Int catalytic domain. This is consistent with the identification of six amino acid substitutions in λ integrase that decrease recombinase activity for the phage λ attachment site (*att* site) DNA sequence and promote activity on the *att* sites of the closely related phage HK022 (28, 29). All but one of these six residues are located in the Int c170 domain, where they face the surface of the proposed DNA-binding groove. The phenotypes of other mutants, along with the pattern of conserved amino acids, suggest that the

highly conserved Gly³³²-His³³³ residues preceding the loop and the Tyr³⁴² nucleophile in the middle of the loop (Fig. 2, A and B) also interact with DNA during catalysis. Residue substitutions involving Gly³³² and Tyr³⁴² have varied effects on DNA binding, DNA bending, and catalysis by λ Int and FLP (14, 30).

The catalytic Tyr³⁴² (phenylalanine in the crystal structure), which covalently links to DNA during Int-mediated recombination, must be activated for nucleophilic attack of the scissile phosphate. This most likely occurs by donating the tyrosine hydroxyl proton to a nearby Lewis base. In the Int c170 structure, the phenyl ring of Tyr³⁴² is sandwiched between Glu³⁴⁹ and Asp³⁵¹. These or other nearby acidic residues like Asp³⁴⁴ (His in some Int family proteins) (Fig. 2B), Asp³⁴⁵, or Glu³⁵⁴ could activate Tyr³⁴² or aid in positioning it on DNA. These acidic residues are not strictly conserved in the integrase-related proteins, but the acidic character of this segment (Fig. 2) is a consistent feature of these recombinases.

One striking feature of the Int catalytic

domain structure is the location of the catalytic Tyr³⁴² on an exposed 17 amino acid loop extending from Lys³³⁴ to Trp³⁵⁰ between α G and β 7 (Fig. 1D). In one protomer no electron density is evident for strand β 6. In the other protomer, a crystal contact stabilizes the orientation of β 6 and the β 6 to β 7 turn and the atomic B factors of this segment are high, averaging 58 Å² as opposed to 25 Å² elsewhere in the protein. Downstream of the flexible segment, β 7 contacts the protein core in both protomers and thereby restricts the range of motion of Tyr³⁴². The finding that Tyr³⁴² is flexibly tethered to the rest of the protein by a peptide loop is consistent with the proteolytic sensitivity of this segment in both λ Int and yeast FLP recombinases (11, 31). Three- or four-residue insertions in this region of FLP inactivate strand cleavage, although DNA-binding activity is retained (31).

Flexible loops are a common feature of proteins that bind to the phosphates of DNA or nucleotides. Examples include the glycine-rich loop of nucleotide-binding folds (32), the active site loop of pro-

Table 1. Summary of crystallographic structure determination. The Int c170 domain of lambda integrase (residues 170 to 356), in which the active site residue Tyr342 is substituted with phenylalanine, was prepared as described (11). Crystals of Int c170 were grown by mixing an equal volume of Int c170 (26 mg/ml) with a well solution containing 50 mM MES (pH 6.15), 75 mM NaCl, 7 mM MgCl₂, 40 mM sodium citrate, 1 mM DTT, 0.1 mM EDTA, and 1 mM spermine-HCl. Crystals were grown by vapor diffusion at 22°C during the course of several days, and larger cube-like crystals were obtained by seeding with microcrystals. The Int c170 crystals belong to space group *R*3 (*a* = *b* = 107.3 Å, *c* = 108.7 Å; triply-primitive hexagonal indexing) with two Int c170 protomers occupying the asymmetric unit. The crystals were flash-frozen in a -160°C nitrogen stream after being stabilized in a well solution containing 30% ethylene glycol. Diffraction data were integrated and scaled with DENZO and SCALEPACK (49). Heavy atom derivatives were obtained by soaking crystals for 12 to 18 hours in well solution containing 1 mM chloro (2,2':6',2''-terpyridine)platinum(II) plus 10 mM 2-mercaptoethanol, or 0.5 mM PtCl₄, or 1 mM Au(CN)₂. Heavy atom parameters were refined with the

programs HEAVY (50) and MLPHARE (CCP4, 1979). The initial MIRAS map was subjected to solvent flattening, twofold averaging, and histogram matching (CCP4, 1979) (51–53). A model was built with the program O (54), and electron-density maps were further improved by incorporating SIGMAA-weighted model phases (55). Model refinement by energy minimization and simulated annealing was performed with X-PLOR (56). An explicit bulk solvent correction was applied to the *F*_{calc} values so that low resolution reflections could be included in the refinement. The final structure was checked by inspection of simulated annealing-omit electron density (57). Noncrystallographic symmetry restraints were imposed during the early stages of model refinement, then released as high-resolution data were incorporated into the refinement. The final model consists of residues 177 to 333 and residues 342 to 355 for protomer A, residues 177 to 334 and 350 to 355 for protomer B, and 195 bound water molecules. Side chain density is lacking for residues Lys²⁵⁶, Glu²⁶⁹, Glu³⁰¹, Glu³¹⁹, Lys³²⁰, Asp³²⁴, Lys³²⁵, and Arg³⁴³ in protomer A and Glu²⁶⁹, Glu³⁰¹, and Asp³²⁴ in protomer B, and these residues are modeled as alanines.

Item	Native-1*	Native-2†	Pt-terpyridine†	PtCl ₄ †	Au(CN) ₂ †
Resolution (Å)	1.9	2.5	2.5	2.5	2.5
Reflections					
Total (<i>N</i>)	110468	84568	93737	117656	91214
Unique (<i>N</i>)	36385	15855	15899	15950	16068
Completeness (%)	98.9	99.4	98.7	99.8	100.0
$I/\sigma(I)$	12.9	14.7	12.6	14.0	7.8
<i>R</i> _{sym}	4.2	4.2	5.2	5.4	10.0
		MIRAS phasing (15–2.5 Å)			
Number of sites			2	3	2
<i>R</i> _{iso} (%)			20.5	24.8	16.1
<i>R</i> _{Cullis}					
Acentric			0.71	0.76	0.80
Anomalous			0.90	0.89	
Phasing power			1.60	1.41	1.16
Figure of merit = 0.63					
		Model refinement			
Resolution (Å)	16–1.9				
<i>R</i> _{cryst} (%)	19.9				
<i>R</i> _{free} (%)	23.2				
rms deviation bond (Å)	0.007				
rms deviation angle (°)	1.178				

*Beamline x12c, Brookhaven National Laboratory, Mar research image plate.

†Elliot GX-13, Mar research image plate.

tein kinases (33, 34), the DNA-binding loop of deoxyribonuclease (DNase I) (35), the surface loops of the $\gamma\delta$ resolvase catalytic domain (36, 37), and the loop harboring the conserved glutamic acid of the D-D-E motif in HIV-1 integrase (20) and MuA transposase (22). The movable loop of λ Int could shield the phosphotyrosine intermediate from solvent-mediated hydrolysis, or it could contribute to the correct positioning of the scissile phosphate for cleavage. Serine and threonine are prominent in the loops of the integrase-related proteins, and these residues could bind to DNA phosphates.

The relatively long flexible loop that loosely tethers the active site Tyr³⁴² to the body of the protein explains one of the major mechanistic paradoxes concerning DNA cleavage by Int family recombinases. Conservative site-specific recombination involves the cleavage, exchange, and rejoining of four DNA strands. Although it is generally recognized that this reaction is executed by four protomers (one for each DNA strand cleaved), experiments addressing the roles of individual protomers support both cis cleavage and trans cleavage of DNA by Int family proteins. In the cis cleavage model, a single protomer provides both the catalytic tyrosine and the Arg-His-Arg triad for cleavage of one DNA strand (8, 18). In the trans cleavage model, the tyrosine from one protomer cleaves a DNA strand that is bound and activated by the Arg-His-Arg triad of a neighboring protomer (17). A substantial body of evidence indicates that FLP cleaves DNA in trans, and current data suggest that the two collaborating protomers are bound to the same DNA (as opposed to synapsing partners) during at least one step of recombination (38). Trans cleavage has also been reported for λ Int and Cre (39, 40). However, other experiments point to a cis cleavage mecha-

nism for λ Int (18) and the related XerC-XerD recombinases (15). These findings have raised fundamental questions of how closely related proteins, and possibly the same protein, can cleave DNA either in cis or in trans (3, 8, 38, 41, 42).

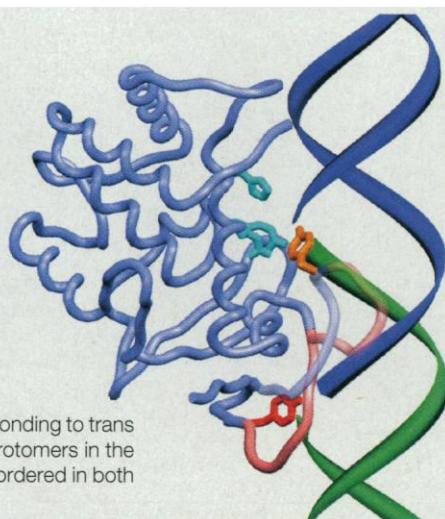
In considering how the λ Int catalytic domain interacts with *att* site DNA, we are guided by the experimental observations that Tyr³⁴² covalently attaches to the scissile phosphate and that the Arg³¹¹ homologue of the closely related XerD recombinase directly contacts the scissile phosphate (16). We have positioned two Int c170 protomers on an idealized *att* site in a configuration that accounts for the interactions described above and that places the positively charged groove of the protein in contact with DNA. For clarity, we show only one of the two Int protomers modeled on DNA (Fig. 3). Both subunits lie to the 3'-side of their respective cut sites with the dimer interface located over the major groove surface of the overlap region between cut sites. The NH₂-terminal end of each c170 protomer lies over the minor groove at the 3'-end of the consensus *att* core sequence 5'-CAACTT-3', adjacent to its major groove surface. This orientation is consistent with evidence that residues upstream of the c170 fragment are important for core site recognition by lambda Int and related proteins (11, 28, 29, 43). Our theoretical model of the Int c170-DNA complex is also consistent with the pattern of methylation protection for Int-DNA complexes, in which the minor groove adjacent to the cut site and the major groove surfaces of the overlap region and of the core binding sites are protected by Int along one face of the helix (44).

The active site loop of λ Int is disordered in the unliganded structure of Int c170, and it could deliver Tyr³⁴² to the Arg-His-Arg catalytic triad in either a cis

or a trans configuration. The cis-cleavage mode requires the disruption of the $\beta 6$ - $\beta 7$ hairpin, which is only visible in one of the two protomers in the crystal structure, and the movement of Tyr³⁴² toward the active site center. This is consistent with the observation that the topoisomerase activity of λ Int is enhanced by deletion of residues 350 to 356 ($\beta 7$) (14). Alterations that remove or destabilize the $\beta 6$ - $\beta 7$ hairpin are likely to release the active site Tyr³⁴², and these findings support the possibility that a fixed orientation of Tyr³⁴² may not be required or preferred for enzymatic activity. The λ Int loop can readily be modeled in a cis configuration that orients Tyr³⁴² for an in-line attack of the scissile phosphate coordinated by the Arg-His-Arg triad (Fig. 3). In the trans cleavage mode, Tyr³⁴² would engage one scissile phosphate and the catalytic triad of the same protomer would bind to the opposite scissile phosphate located some 25 Å away in a B-form *att* site. Indeed, the Tyr³⁴² of one Int protomer in the asymmetric unit extends away from the protein core and it is located 23 Å from His³⁰⁸ and about 17 Å from Arg²¹² and Arg³¹¹ (Figs. 1 and 3). In this orientation, each Int protomer can bridge the scissile phosphates of the *att* site in a trans cleavage mode. It is to be noted that FLP, for which trans cleavage has been most thoroughly documented, and all five of the other eukaryotic recombinases, have loop segments that are five to eight residues longer than those of other Int family proteins (3). These additional residues may restrict the tyrosine-containing loop of FLPs to the trans orientation. The structure of the λ Int catalytic domain provides an explanation for both cis and trans DNA cleavage reactions catalyzed by closely related recombinases and, perhaps, catalyzed by the same protein.

The folding of disordered segments within the DNA-binding domains of many proteins accompanies their binding to specific DNA targets, and such an "induced fit" may enhance discrimination against nonspecific sites (45). The productive interaction of a site-specific recombinase with DNA generates double-stranded breaks, so it is essential that DNA-binding or cleavage activity is stringently controlled. In that λ Int and other recombinases show relatively high nonspecific DNA binding (44), strand cleavage must be tightly controlled. The flexible active site loop of λ Int may provide such a control mechanism. In the case of cis cleavage the flexible loop may adopt the catalytically competent orientation only in the context of the correct multiprotein-DNA complex, thereby keeping the active site tyrosine at bay until the recombi-

Fig. 3. Theoretical model of the λ Int catalytic core bound to a B-form half-*att* site. A full *att* site contains a pair of inverted core-type Int binding sites. An Int protomer at each site is responsible for cleaving one DNA strand via formation of a covalent 3' phospho-tyrosine linkage and a free 5'-hydroxyl. The two nicks are staggered by seven base pairs with a 5' overhang. For clarity, only one subunit of the Int c170 dimer that was modeled on DNA is shown. The catalytic Arg-His-Arg triad (cyan) of Int is docked over one of the scissile phosphates (shown as breaks in the DNA ribbon). The C α trace of Int c170 (blue) is displayed with the active site loop containing the Tyr³⁴² nucleophile shown in two alternative conformations. The orientation corresponding to cis cleavage (orange tyrosine) is a theoretical model, whereas that corresponding to trans cleavage (red tyrosine) is present in one of two Int protomers in the crystal structure. The segment of the loop that is disordered in both protomers (Lys³³⁴ to Gln³⁴¹) is modeled in pink.



tion apparatus is poised to proceed. In the case of trans cleavage, the requirement for two precisely aligned protomers could serve to restrain inappropriate cleavages, as proposed for FLP (17). A second potential role of the flexible catalytic loop may be to accommodate strand transfer during synapsis. Movement of the protein-DNA complex may be required for access of the invading strand to the catalytic center, and this movement might be facilitated by adjustment of the active site loop.

The structure of λ Int c170 with its flexibly tethered tyrosine nucleophile also suggests a mechanism to explain another variable property of Int family recombinases. Some members of the Int family are, like λ Int, quite fastidious in their requirement for DNA-DNA homology within the overlap regions of two recombining partners (46). Other Int family recombinases, such as the Tn1545 and Tn916 transposases, are more relaxed in their response to heterologies between recombining partners (47, 48). This difference might be due in part to the variety of amino acids occupying this segment of Int family recombinases (Fig. 2B). Alterations in the delivery arm of Tyr³⁴² could be expected to significantly influence such parameters as the rate, reversibility, and specificity of the DNA cleavage and ligation reactions.

REFERENCES AND NOTES

1. J. M. Leong *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **49**, 707 (1984).
2. K. E. Abremski and R. H. Hoess, *Protein Eng.* **5**, 87 (1992).
3. G. W. Blakely and D. J. Sherratt, *Mol. Microbiol.* **20**, 234 (1996).
4. A. Landy, *Annu. Rev. Biochem.* **58**, 913 (1989).
5. N. L. Craig, *Annu. Rev. Genet.* **22**, 77 (1988).
6. P. D. Sadowski, *FASEB J.* **7**, 760 (1993).
7. W. M. Stark, M. R. Boocock, D. J. Sherratt, *Trends Genet.* **8**, 432 (1992).
8. A. Landy, *Curr. Opin. Genet. Dev.* **3**, 699 (1993).
9. Y. Kikuchi and H. A. Nash, *J. Biol. Chem.* **253**, 7149 (1978).
10. L. Moitoso de Vargas, C. A. Pargellis, N. M. Hasan, E. W. Bushman, A. Landy, *Cell* **54**, 923 (1988).
11. R. S. Tirumalai, E. Healey, A. Landy, in preparation.
12. C. A. Pargellis, S. E. Nunes-Düby, L. Moitoso de Vargas, A. Landy, *J. Biol. Chem.* **263**, 7678 (1988).
13. R. L. Parsons, P. V. Prasad, R. M. Harshey, M. Jayaram, *Mol. Cell. Biol.* **8**, 3303 (1988).
14. Y. W. Han, R. I. Gumpport, J. F. Gardner, *J. Mol. Biol.* **235**, 908 (1994).
15. L. K. Aroiszewska and D. J. Sherratt, *EMBO J.* **14**, 2112 (1995).
16. G. W. Blakely, A. O. Davidson, D. J. Sherratt, *J. Mol. Biol.* **265**, 30 (1997).
17. J. Chen, J. Lee, M. Jayaram, *Cell* **69**, 647 (1992).
18. S. E. Nunes-Düby *et al.*, *EMBO J.* **13**, 4421 (1994).
19. L. Holm and C. Sander, *J. Mol. Biol.* **233**, 123 (1993).
20. F. Dyda *et al.*, *Science* **266**, 1981 (1994).
21. G. Bujacz *et al.*, *J. Mol. Biol.* **253**, 333 (1995).
22. P. Rice and K. Mizuuchi, *Cell* **82**, 209 (1995).
23. M. R. Sanderson *et al.*, *ibid.* **63**, 1323 (1990).
24. C. D. Lima, J. C. Wang, A. Mondragon, *Nature* **367**, 138 (1994).
25. J. M. Berger, S. J. Gamblin, S. C. Harrison, J. C. Wang, *ibid.* **379**, 225 (1996).
26. G. J. Barton, *Methods Enzymol.* **183**, 403 (1990).

27. S. Nunes-Düby, unpublished results.
28. L. Dorgai, E. Yagil, R. A. Weisberg, *J. Mol. Biol.* **252**, 178 (1995).
29. E. Yagil, L. Dorgai, R. Weisberg, *ibid.*, p. 163.
30. C. J. E. Schwartz and P. D. Sadowski, *ibid.* **205**, 647 (1989).
31. B. R. Evans *et al.*, *J. Biol. Chem.* **265**, 18504 (1990).
32. M. G. Rossmann, A. Liljas, C. Branden, L. J. Banaszak, in *The Enzymes* (Academic Press, New York, 1975), pp. 61–102.
33. J. Zheng *et al.*, *Acta Crystallogr.* **D49**, 362 (1993).
34. H. L. DeBontd *et al.*, *Nature* **363**, 595 (1993).
35. S. A. Weston, A. Lahm, D. Suck, *J. Mol. Biol.* **226**, 1237 (1992).
36. P. Rice and T. A. Steitz, *Structure* **2**, 371 (1994).
37. W. Yang and T. A. Steitz, *Cell* **82**, 193 (1995).
38. J. Lee, I. Whang, M. Jayaram, *EMBO J.* **13**, 5346 (1994).
39. Y. W. Han, R. I. Gumpport, J. F. Gardner, *ibid.* **12**, 4577 (1993).
40. A. C. Shaikh and P. D. Sadowski, *J. Biol. Chem.* **272**, 5695 (1997).
41. X. Qian and M. M. Cox, *Genes Dev.* **9**, 2053 (1995).
42. W. M. Stark and M. R. Boocock, *Trends Genet.* **11**, 121 (1995); M. Jayaram and J. Lee, *ibid.*, p. 432.
43. G. B. Panigrahi and P. D. Sadowski, *J. Biol. Chem.* **269**, 10940 (1994).
44. W. Ross and A. Landy, *Cell* **33**, 261 (1983).
45. R. S. Spolar and T. Record, Jr., *Science* **263**, 777 (1994).
46. R. A. Weisberg, L. W. Enquist, C. Foeller, A. Landy, *J. Mol. Biol.* **170**, 319 (1983).
47. M. G. Caparon and J. R. Scott, *Cell* **59**, 1027 (1989).
48. P. Trieu-Cuot, C. Poyart-Salmeron, C. Carlier, P. Courvalin, *Mol. Microbiol.* **8**, 179 (1994).
49. Z. Otwinowski, *Data Collection and Processing* (SERC Daresbury Laboratory, Warrington, UK, 1993), pp. 56–62.
50. T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr.* **A39**, 813 (1983).
51. K. D. Cowtan and P. Main, *ibid.* **D49**, 148 (1993).
52. K. Zhang and P. Main, *ibid.* **A46**, 41 (1990).
53. T. A. Jones, in *Molecular Replacement* (SERC Daresbury Laboratory, Warrington, UK, 1992), pp. 91–105.
54. T. A. Jones, *O-The Manual* (Uppsala, Sweden, 1992).
55. R. J. Read, *Acta Crystallogr.* **A42**, 140 (1986).
56. A. T. Brunger, *X-PLOR Version 3.1, A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
57. A. T. Brunger, *Nature* **355**, 472 (1992).
58. S. V. Evans, *J. Mol. Graphics* **11**, 134 (1990).
59. G. J. Barton, *Protein Eng.* **6**, 37 (1993).
60. We thank E. Healey for purified protein, S. Nunes-Düby for help with sequence alignments, R. Sweet for assistance with data collection at beamline X-12C, National Synchrotron Light Source (Upton, NY), T. Oliveira for technical assistance, J. Boyles for assistance with manuscript preparation, and J. Cheah, S. Doublie, S. Nunes-Düby and other members of our research groups for their assistance and comments. Supported by the Lucille P. Markey Charitable Trust (TE) and NIH grants AI13544 and GM33928 (AL) and a Howard Hughes Medical Institute predoctoral fellowship (HJK). The coordinates have been deposited in the Brookhaven Protein Data Base with accession number 1AE9.

22 November 1996; accepted 25 February 1997

A Transmembrane Helix Dimer: Structure and Implications

Kevin R. MacKenzie, James H. Prestegard, Donald M. Engelman*

The three-dimensional structure of the dimeric transmembrane domain of glycoprotein A (GpA) was determined by solution nuclear magnetic resonance spectroscopy of a 40-residue peptide solubilized in aqueous detergent micelles. The GpA membrane-spanning α helices cross at an angle of -40 degrees and form a small but well-packed interface that lacks intermonomer hydrogen bonds. The structure provides an explanation for the previously characterized sequence dependence of GpA dimerization and demonstrates that van der Waals interactions alone can mediate stable and specific associations between transmembrane helices.

Glycoprotein A, the primary sialoglycoprotein of human erythrocyte membranes, forms noncovalent dimers by sequence-specific, reversible association of its single hydrophobic membrane-spanning domain (1–3). We have determined the dimeric structure of GpA, a peptide corresponding to residues 62 to 101 of human GpA, by heteronuclear nuclear magnetic resonance (NMR) methods (Fig. 1). Nuclear Overhauser effect (NOE) distance restraints and J coupling–derived dihedral restraints were used to generate a family of 20

K. R. MacKenzie and D. M. Engelman, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520–8114, USA.
J. H. Prestegard, Department of Chemistry, Yale University, New Haven, CT 06520–8107, USA.

*To whom correspondence should be addressed.

structures (4). Residues 62 to 71, which lie outside the membrane (5) and are unnecessary for dimerization (2), are poorly defined by this procedure, but within the transmembrane region the structures differ from the average by only 0.4 Å [root mean square (rms)] for backbone and 0.8 Å (rms) for all nonhydrogen atoms. The well-defined interface between the hydrophobic helices allows us to describe the physical basis for helix-helix association, the second stage of the proposed “two-stage model” for membrane protein folding (6).

Residues 73 to 96 of GpA form α helices that interact at a -40° (right-handed) crossing angle in a symmetric GpA dimer (Fig. 2). Side chain and backbone atoms of seven residues make favorable van der Waals inter-