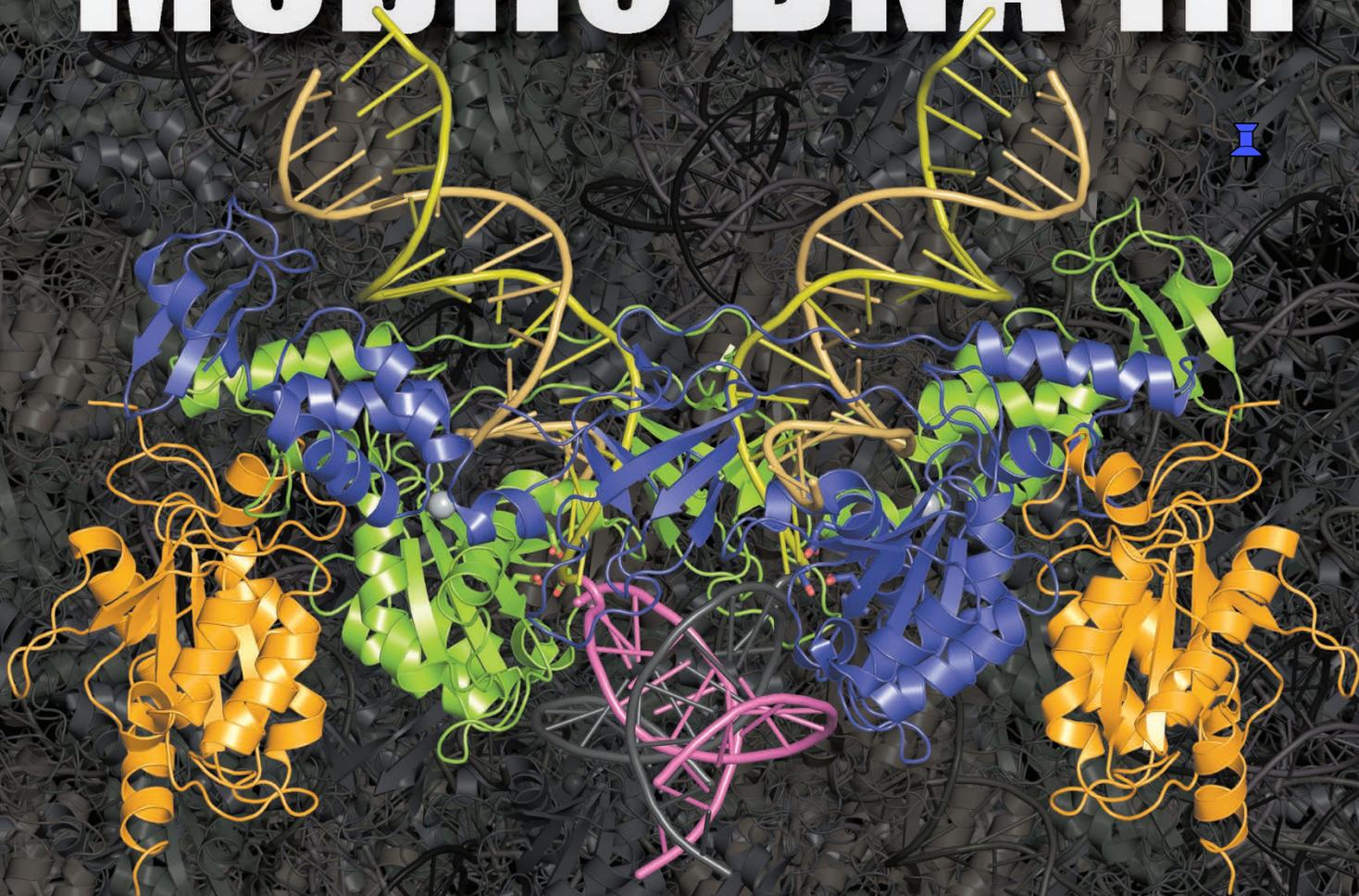


Mobile DNA III



Editor in Chief
Nancy L. Craig

Editors

Michael Chandler
Martin Gellert
Alan M. Lambowitz
Phoebe A. Rice
Suzanne Sandmeyer

Mobile DNA III

Mobile DNA III

Editor in Chief

Nancy L. Craig

Howard Hughes Medical Institute,
Department of Molecular Biology & Genetics,
Johns Hopkins University School of Medicine, Baltimore, MD

Editors

Michael Chandler

Laboratoire de Microbiologie et Génétique Moléculaire,
C.N.R.S., F-31062 Toulouse Cedex, France

Martin Gellert

National Institutes of Health,
Molecular Genetics Section, NIDDK, Bethesda, MD

Alan M. Lambowitz

Institute for Cellular Molecular Biology and
Department of Molecular Biosciences,
University of Texas at Austin, Austin, TX

Phoebe A. Rice

Department of Biochemistry and Molecular Biology,
University of Chicago, Chicago, IL

Suzanne B. Sandmeyer

Departments of Biological Chemistry and
Microbiology and Molecular Genetics,
University of California, Irvine, Irvine, CA

ASM Press, Washington, DC

Copyright © 2015 by ASM Press. ASM Press is a registered trademark of the American Society for Microbiology. All rights reserved. No part of this publication may be reproduced or transmitted in whole or in part or reutilized in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Disclaimer: To the best of the publisher's knowledge, this publication provides information concerning the subject matter covered that is accurate as of the date of publication. The publisher is not providing legal, medical, or other professional services. Any reference herein to any specific commercial products, procedures, or services by trade name, trademark, manufacturer, or otherwise does not constitute or imply endorsement, recommendation, or favored status by the American Society for Microbiology (ASM). The views and opinions of the author(s) expressed in this publication do not necessarily state or reflect those of ASM, and they shall not be used to advertise or endorse any product.

Library of Congress Cataloging-in-Publication Data

Mobile DNA III / editor in chief, Nancy L. Craig, Johns Hopkins University School of Medicine, Baltimore, MD ; editors, Michael Chandler, Université Paul Sabatier, Toulouse, France, Martin Gellert, National Institutes of Health, Bethesda, MD, Alan M. Lambowitz, University of Texas, Austin, TX, Phoebe A. Rice, University of Chicago, Chicago, IL, Suzanne Sandmeyer, University of California, Irvine, CA.

pages cm

Includes bibliographical references and index.

ISBN 978-1-55581-920-0 (alk. paper)

1. Mobile genetic elements. I. Craig, Nancy Lynn, 1952- editor. II. Chandler, Michael (Molecular microbiologist), editor. III. Gellert, Martin, editor. IV. Lambowitz, Alan, editor. V. Rice, Phoebe A., editor. VI. Sandmeyer, Suzanne, editor. VII. Title: Mobile DNA 3. VIII. Title: Mobile DNA three.

QH452.3.M633 2015

572.8'69--dc23

2015011126

doi:10.1128/9781555819217

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Address editorial correspondence to: ASM Press, 1752 N St., N.W., Washington, DC 20036-2904, USA.

Send orders to: ASM Press, P.O. Box 605, Herndon, VA 20172, USA.

Phone: 800-546-2416; 703-661-1593. Fax: 703-661-1501.

E-mail: books@asmusa.org

Online: <http://estore.asm.org>

Cover: An artistic representation of the retroviral intasome engaged with target DNA in the nucleus of a host cell. The illustration is based on the crystal structure of the prototype foamy virus strand transfer complex (Protein Databank ID 3OS0; for details see Maertens *et al.*, Nature, 2010, 468, 326-9). Image provided by Dr. Peter Cherepanov, Cancer Research UK, London Research Institute, London, EN6 3LD United Kingdom.

Contents

Contributors *xi*

Preface *xxi*

I. INTRODUCTION

1. A Moveable Feast: An Introduction to Mobile DNA 3
NANCY L. CRAIG

II. CONSERVATIVE SITE-SPECIFIC RECOMBINATION

2. An Overview of Tyrosine Site-specific Recombination:
From an F₁p Perspective 43
MAKKUMI JAYARAM
3. The Serine Recombinases 73
W. MARSHALL STARK
4. The λ Integrase Site-specific Recombination Pathway 91
ARTHUR LANDY
5. Cre Recombinase 119
GREGORY D. VAN DUYN
6. The Integron: Adaptation On Demand 139
JOSÉ ANTONIO ESCUDERO, CÉLINE LOOT, AKEKSANDRA NIVINA, AND
DIDIER MAZEL

7. Xer Site-Specific Recombination: Promoting Vertical and Horizontal Transmission of Genetic Information 163
CAROLINE MIDONET AND FRANSOIS-XAVIER BARRE
8. The Integration and Excision of CTnDOT 183
MARGARET M. WOOD AND JEFFREY F. GARDNER
9. Site-specific DNA Inversion by Serine Recombinases 199
REID C. JOHNSON
10. Serine Resolvases 237
PHOEBE A. RICE
11. Phage-encoded Serine Integrases and Other Large Serine Recombinases 253
MARGARET C.M. SMITH
12. Hairpin Telomere Resolvases 273
KERRI KOBRYN AND GEORGE CHACONAS
13. Biology of Three ICE Families: SXT/R391, ICEBs1, and ICES $t1$ /ICES $t3$ 289
NICHOLAS CARRARO AND VINCENT BURRUS

III. PROGRAMMED REARRANGEMENTS

14. V(D)J Recombination: Mechanism, Errors, and Fidelity 313
DAVID B. ROTH
15. Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination 325
JOYCE K. HWANG, FREDERICK W. ALT, AND LENG-SIEW YEAP
16. Programmed Genome Rearrangements in *Tetrahymena* 349
MENG CHAO YAO, JU-IAN CHAO, AND CHAO-YIN CHENG
17. Programmed Rearrangement in Ciliates: *Paramecium* 369
MIREILLE BETERMIER AND SANDRA DUHARCOURT
18. Programmed Genome Rearrangements in the Ciliate *Oxytricha* 389
V. TALYA YERLICI AND LAURA F. LANDWEBER
19. DNA Recombination Strategies During Antigenic Variation in the African Trypanosome 409
RICHARD McCULLOCH, LIAM J. MORRISON, AND JAMES P.J. HALL
20. Recombination and Diversification of the Variant Antigen Encoding Genes in the Malaria Parasite *Plasmodium falciparum* 437
LAURA A. KIRKMAN AND KIRK W. DEITSCH

21. Mobile DNA in the Pathogenic *Neisseria* 451
KYLE P. OBERGFELL AND H. STEVEN SEIFERT
22. *vls* Antigenic Variation Systems of Lyme Disease *Borrelia*: Eluding Host Immunity through both Random, Segmental Gene Conversion and Framework Heterogeneity 471
STEVEN J. NORRIS
23. Mating-type Gene Switching in *Saccharomyces cerevisiae* 491
CHENG-SHENG LEE AND JAMES HABER
24. A Unique DNA Recombination Mechanism of the Mating/Cell-type Switching of Fission Yeasts: a Review 515
AMAR KLAR

IV. DNA-ONLY TRANSPOSONS

25. Mechanisms of DNA Transposition 531
ALLISON BURGESS HICKMAN AND FRED DYDA
26. Everyman's Guide to Bacterial Insertion Sequences 555
PATRICIA SIGUIER, EDITH GOURBEYRE, ALESSANDRO VARANI,
BAO TON-HOANG, AND MICHAEL CHANDLER
27. Copy-out-Paste-in Transposition of IS911: A Major Transposition Pathway 591
MICHAEL CHANDLER, OLIVIER FAYET, PHILIPPE ROUSSEAU,
BAO TON HOANG, AND GUY DUVAL-VALENTIN
28. The IS200/IS605 Family and "Peel and Paste" Single-strand Transposition Mechanism 609
SUSU HE, ALIX CORNELOUP, CATHERINE GUYNET, LAURE LAVATINE,
ANNE CAUMONT-SARCOS, PATRICIA SIGUIER, BRIGITTE MARTY,
FRED DYDA, MICHAEL CHANDLER, AND BOA TON HOANG
29. Transposons Tn10 and Tn5 631
DAVID B. HANIFORD AND MICHAEL J. ELLIS
30. Tn7 647
JOSEPH E. PETERS
31. Transposable Phage Mu 669
RASIKA M. HARSHEY
32. The Tn3-family of Replicative Transposons 693
EMILIEN NICOLAS, MICHAEL LAMBIN, DAMIEN DANDROY, CHRISTINE GALLOY, NATHAN NGUYEN, CEDRIC A. OGER, AND BERNARD HALLET
33. P Transposable Elements in *Drosophila* and other Eukaryotic Organisms 727
SHARMISTHA MAJUMDAR AND DONALD C. RIO

34. Mariner and the ITm Superfamily of Transposons 753
MICHAEL TELLIER, CORENTIN CLAEYS BOUUAERT, AND RONALD CHALMERS
 35. *hAT* Transposable Elements 773
PETER W. ATKINSON
 36. *Mutator* and *MULE* Transposons 801
DAMON LISCH
 37. Adeno-associated Virus as a Mammalian DNA Vector 827
MAX SALGANIK, MATTHEW HIRSCH, AND R. JUDE SAMULSKI
 38. *Sleeping Beauty* Transposition 851
ZOLTÁN IVICS AND ZSUZSANNA IZSVÁK
 39. *piggyBac* Transposon 873
KOSUKE YUSA
 40. *Helitrons*, the Eukaryotic Rolling-circle Transposable Elements 891
JAINY THOMAS AND ELLEN J. PRITHAM
- V. LTR RETROTRANSPOSONS**
41. Ty1 LTR-retrotransposon of Budding Yeast, *Saccharomyces cerevisiae* 927
M. JOAN CURCIO, SHEIL LUTZ, AND PASCALE LESAGE
 42. Ty3, a Position-specific Retrotransposon in Budding Yeast 965
SUZANNE SANDMEYER, KURT PATTERSON AND VIRGINIA BILANCHONE
 43. The Long Terminal Repeat Retrotransposons Tf1 and Tf2 of *Schizosaccharomyces pombe* 997
CAROLINE ESNAULT AND HENRY L. LEVIN
 44. Retroviral Integrase Structure and DNA Recombination Mechanism 1011
ALAN ENGELMAN AND PETER CHEREPANOV
 45. Host Factors in Retroviral Integration and Selection of Integration Target Sites 1035
ROBERT CRAIGIE AND FREDERIC D. BUSHMAN
 46. Reverse Transcription of Retroviruses and LTR Retrotransposons STEVEN H. HUGHES 1051
 47. Mammalian Endogenous Retroviruses 1079
DIXIE L. MAGER AND JONATHAN P. STOYE
 48. Retroviral DNA Transposition: Themes and Variations 1101
ANNA MARIE SKALKA

VI. NON-LTR RETROTRANSPOSONS

49. Integration, Regulation, and Long-Term Stability of R2 Retrotransposons 1127
THOMAS H. EICKBUSH AND DANNA G. EICKBUSH
50. Site-Specific non-LTR retrotransposons 1147
HARUHIKO FUJIWARA
51. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes 1165
SANDRA R. RICHARDSON, AURÉLIEN J. DOUCET, HUIRA C. KOPERA, JOHN B. MOLDOVAN, JOSÉ LUIS GARCIA-PEREZ, AND JOHN V. MORAN
52. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution 1209
ALAN LAMBOWITZ AND MARLENE BELFORT
53. Diversity-generating Retroelements in Phage and Bacterial Genomes 1237
HUATAO GUO, DIEGO ARAMBULA, PARTHO GHOSH, AND JEFF F. MILLER
54. An Unexplored Diversity of Reverse Transcriptases in Bacteria 1253
STEVEN ZIMMERLY AND LI WU
55. Tyrosine Recombinase Retrotransposons and Transposons 1271
RUSSELL T.M. POULTER AND MARGI I. BUTLER
- Index* 1293

Contributors

FREDERICK W. ALT

Department of Genetics, Harvard Medical School, Boston, MA 02115

DIEGO ARAMBULA

Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095

PETER W. ATKINSON

Department of Entomology and Institute for Integrative Genome Biology, University of California, Riverside, CA 92521

FRANCOIS-XAVIER BARRE

Centre de Genetique Moleculaire, C.N.R.S., Gif-sur Yvette, France 91198

MARLENE BELFORT

Department of Biological Sciences and RNA Institute, University of Albany, State University of New York, Albany, NY 12222

MIREILLE BETERMIER

Centre de Génétique Moléculaire, C.N.R.S., Gif-sur Yvette, France 91198

VIRGINIA W. BILANCHONE

Departments of Biological Chemistry and Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697

VINCENT BURRUS

Departement de biologie, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada

FREDERIC D. BUSHMAN

Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

MARGI I. BUTLER

Department of Biochemistry, University of Otago, Dunedin, New Zealand 09054

NICOLAS CARRARO

Departement de biologie, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada

GEORGE CHACONAS

Departments of Biochemistry and Molecular Biology; Microbiology, Immunology & Infectious Diseases, The University of Calgary, Calgary, ALB, Canada AB T2N 4N1

RONALD CHALMERS

School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH United Kingdom

MICHAEL CHANDLER

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

JU-LAN CHAO

Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan 11529

CHAO-YIN CHENG

Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan 11529

PETER CHEREPANOV

Cancer Research UK, London Research Institute, London, EN6 3LD United Kingdom

CORENTIN CLAEYS BOUUAERT

School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH United Kingdom

ALIX COUNELOUP

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

ROBERT CRAIGIE

Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892

M. JOAN CURCIO

Laboratory of Molecular Genetics, University of Albany, State University of New York, Albany, NY 12208

DAMIEN DANDROY

Institut des Sciences de la Vie, Université Catholique de Louvain, Louvain-la-Neuve, B-1348, Belgium

KIRK W. DEITSCH

Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10065

AURELIEN J DOUCET

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109

SANDRA DUHARCOURT

Institut Jacques Monod, C.N.R.S., Paris, France 75205

GUY DUVAL-VALENTIN

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

FRED DYDA

Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20814

DANNA G. EICKBUSH

Department of Biology, University of Rochester, Rochester, NY 14627

THOMAS H. EICKBUSH

Department of Biology, University of Rochester, Rochester, NY 14627

MICHAEL JOSEPH ELLIS

Department of Biochemistry, University of Western Ontario, London, Ontario, N6A 5C1, Canada

ALAN ENGELMAN

Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, MA 02215

JOSE ANTONIO ESCUDERO

Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

CAROLINE ESNAULT

Program in Cellular Regulation and Metabolism, NICHD, National Institutes of Health, Bethesda, MD 20892

HSIU-FANG FAN

Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan 00112

OLIVIER FAYET

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse Cedex, France 31062

HARUHIKO FUJIWARA

Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan 277-8562

CHRISTINE GALLOY

Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

JOSE L. GARCIA-PEREZ

Department of Human DNA Variability, GENYO (Pfizer-University of Granada & Andalusian Regional Government Genomics & Oncology Center), 18016 Granada, Spain

JEFFREY F. GARDNER

Department of Microbiology, University of Illinois at Urbana-Champaign,
Urbana, IL 61801

PARTHO GHOSH

Department of Chemistry and Biochemistry, University of California at San
Diego, La Jolla, CA 92093

EDITH GOURBEYRE

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse
Cedex, France 31062

PIOTR GUGA

Department of Bio-organic Chemistry, Polish Academy of Sciences, Lodz,
POLAND

HUATAO GUO

Department of Molecular Microbiology and Immunology, University of
Missouri School of Medicine, Columbia, MO 65212

CATHERINE GUYNET

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse
Cedex, France 31062

JAMES HABER

Department of Biology and Rosenstiel Basic Medical Sciences Research Center,
Brandeis University, Waltham, MA 02454

JAMES HALL

Department of Biology, University of York, York, YO10 5DD United Kingdom

BERNARD HALLET

Institut des Sciences de la Vie, Université Catholique de Louvain, Louvain-la-
Neuve, BELGIUM B-1348

DAVID B. HANIFORD

Department of Biochemistry, The University of Western Ontario, London,
Ontario, N6A 5C1 Canada

RASIKA HARSHEY

Department of Molecular Biosciences, Institute of Cellular and Molecular
Biology, University of Texas at Austin, Austin, TX 78712

SUSU HE

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse
Cedex, France 31062

ALISON BURGESS HICKMAN

Laboratory of Molecular Biology, NIDDK, National Institutes of Health,
Bethesda, MD 20814

MATTHEW HIRSCH

Gene Therapy Center, Department of Ophthalmology, University of North
Carolina, Chapel Hill, Chapel Hill, NC 27599

STEPHEN H. HUGHES

HIV Drug Resistance Program, National Cancer Institute at Frederick, NIH,
Frederick, MD 21702

JOYCE K. HWANG

Program in Cellular and Molecular Medicine; Department of Genetics, Harvard Medical School, Boston, MA 02115

KEN ISHIKAWA

Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute at Frederick, NIH, Frederick, MD 21702

ZOLTAN IVICS

Division of Medical Biotechnology, Paul Ehrlich Institute, Langen, Germany 63225

ZSUZSANNA IZSVAK

Max Delbrück Center for Molecular Medicine, Max Delbrück Center for Molecular Medicine, Berlin, Germany 13092

MAKKUNI JAYARAM

Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

REID C JOHNSON

Department of Biological Chemistry, University of California, Los Angeles School of Medicine, Los Angeles, CA 90095

AASHIQ H. KACHROO

Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

LAURA KIRKMAN

Department of Internal Medicine, Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10065

AMAR J. S. KLAR

Gene Regulation and Chromosome Biology Laboratory,, National Cancer Institute at Frederick, NIH, Frederick, MD 21702

KERRI KOBRYN

Department of Microbiology and Immunology, College of Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5E5 Canada

HUIRA C. KOPERA

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109

MICHAEL LAMBIN

Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

ALAN M. LAMBOWITZ

Institute for Cellular Molecular Biology and Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

LAURA F. LANDWEBER

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544

ARTHUR LANDY

Division of Biology and Medicine, Brown University, Providence, RI 02912

LAURE LAVATINE

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062
Toulouse Cedex, France

CHENG-SHENG LEE

Department of Biology and Rosenstiel Basic Medical Sciences Research Center,
Brandeis University, Waltham, MA 02454

PASCALE LESAGE

Institut Universitaire d'hématologie, C.N.R.S., F-75475 Paris Cedex 10, France

HENRY L. LEVIN

Program in Cellular Regulation and Metabolism, NICHD, National Institutes of
Health, Bethesda, MD 20892

DAMON LISCH

Department of Botany and Plant Pathology, University of California, Berkeley,
Berkeley, CA 94720

CELINE LOOT

Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

SHEILA LUTZ

Laboratory of Molecular Genetics, University of Albany, State University of
New York, Albany, NY 12208

CHIEN-HUI MA

Department of Molecular Biosciences, University of Texas at Austin, Austin, TX
78712

DIXIE L. MAGER

Terry Fox Laboratory, British Columbia Cancer Agency and Dept. of Medical
Genetics, University of British Columbia, Vancouver, British Columbia, V5Z
1L3, Canada

SHARMISTHA MAJUMDAR

Department of Molecular and Cell Biology, University of California, Berkeley,
Berkeley, CA 94720

BRIGITTE MARTY

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062
Toulouse Cedex, France

DIDIER MAZEL

Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

RICHARD McCULLOCH

Wellcome Trust Centre for Molecular Parasitology, University of Glasgow,
Glasgow, G12 8TA, United Kingdom

CAROLINE MIDONET

Centre de Génétique Moléculaire, C.N.R.S., Gif-sur Yvette, 91198 France

JEFF F. MILLER

Department of Microbiology, Immunology and Molecular Genetics; The California Nano Systems Institute, University of California at Los Angeles, Los Angeles, CA 90095

JOHN B. MOLDOVAN

Cellular and Molecular Biology Graduate Program, University of Michigan Medical School, Ann Arbor, MI 48109

SHARON MOORE

Gene Regulation and Chromosome Biology Laboratory,, National Cancer Institute at Frederick, NIH, Frederick, MD 21702

JOHN V. MORAN

Departments of Human Genetics; Cellular and Molecular Biology, University of Michigan Medical School, Ann Arbor, Michigan 48109

LIAM MORRISON

Roslin Institute, University of Edinburgh, Midlothian, EH25 9RG United Kingdom

NATHAN NGUYEN

Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

EMILIEN NICOLAS

Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

ALEKSANDRA NIVINA

Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

STEVEN J NORRIS

Department of Pathology and Laboratory Medicine, University of Texas Health Medical School, Houston, TX 77225

KYLE P. OBERGFELL

Department of Microbiology and Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611

CEDRIC A. OGER

Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

KURT PATTERSON

Departments of Biological Chemistry and Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697

JOSEPH E. PETERS

Department of Microbiology, Cornell University, Ithaca, NY 14853

RUSSELL POULTER

Department of Biochemistry, University of Otago, Dunedin, 09054 New Zealand

ELLEN J. PRITHAM

Department of Human Genetics, University of Utah, Salt Lake City, UT 84112

PHOEBE A. RICE

Department of Biochemistry and Molecular Biophysics, University of Chicago,
Chicago, IL 60637

SANDRA R RICHARDSON

Department of Human Genetics, University of Michigan Medical School, Ann
Arbor, MI 48109

DONALD RIO

Department of Molecular and Cell Biology, University of California, Berkeley,
Berkeley, CA 94720

DAVID B. ROTH

Department Pathology and Laboratory Medicine and the Center for
Personalized Diagnostics, Perelman School of Medicine, University of
Pennsylvania, Philadelphia, PA 19104

PHILIPPE ROUSSEAU

Laboratoire de Microbiologie et Génétique Moléculaires, C.N.R.S., F-31062
Toulouse Cedex, France

PAUL A. ROWLEY

Department of Molecular Biosciences, University of Texas at Austin, Austin, TX
78712

MAXIM SALGANIK

Gene Therapy Center/Lineberger Comprehensive Cancer Center, University of
North Carolina, Chapel Hill, Chapel Hill, NC 27599

RICHARD JUDE SAMULSKI

Gene Therapy Center, University of North Carolina, Chapel Hill, Chapel Hill,
NC 27599

SUZANNE B. SANDMEYER

Departments of Biological Chemistry and Microbiology and Molecular
Genetics, University of California, Irvine, Irvine, CA 92697

ANNE SARCOS

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062
Toulouse Cedex, France

H. STEVEN SEIFERT

Department of Microbiology and Immunology, Northwestern University
Feinberg School of Medicine, Chicago, IL 60611

PATRICIA SIGUIER

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062
Toulouse Cedex, France

ANNA MARIE SKALKA

Department of Microbiology, Fox Chase Cancer Center, Philadelphia, PA 19111

MARGARET C.M. SMITH

Department of Biology, University of York, York, YO10 5DD United Kingdom

W. MARSHALL STARK

Institute of Molecular Cell and Systems Biology, University of Glasgow,
Glasgow, G12 8WW, United Kingdom

JONATHAN P. STOYE

Division of Virology, MRC National Institute for Medical Research, London,
NW7 1AA, United Kingdom

MICHAEL TELLIER

School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH,
United Kingdom

JAINY THOMAS

Department of Human Genetic, University of Utah, Salt Lake City, UT 84112

BAO TON HOANG

Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062
Toulouse Cedex, France

GREGORY D. VAN DUYNÉ

Department of Biochemistry and Biophysics, Perelman School of Medicine,
University of Pennsylvania, Philadelphia, PA 19104

ALESSANDRO VARANI

Departamento de Tecnología, UNESP - Univ. Estadual Paulista, Jaboticaba. SP,
BRAZIL 14870

YURI VOZIYANOV

School of Biosciences, Louisiana Tech University, Ruston, LA 71272

MARGARET M. WOOD

Academic Advising Center, Simmons College, Boston, MA 02115

LI WU

Department of Biological Sciences, University of Calgary, Calgary, Alberta, T2N
1N4, Canada

MENG-CHAO YAO

Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan 11529

LENG SIEW YEAP

Program in Cellular and Molecular Medicine; Department of Genetics, Harvard
Medical School, Boston, MA 02115

V. TALYA YERLICI

Department of Molecular Biology, Princeton University, Princeton, NJ 8544

KOSUKE YUSA

Stem Cell Genetics, Wellcome Trust Sanger Institute, Cambridge, CB10 1SA,
United Kingdom

STEVEN ZIMMERLY

Department of Biological Sciences, University of Calgary, Calgary, Alberta, T2N
1N4, Canada

Preface

Nowhere do the cooperative powers of DNA sequencing, high-resolution protein structure, biochemistry and molecular genetics shine more intensely than on Mobile DNAs. In *Mobile DNA II*, we knew that almost half of the human genome is comprised of retroelements. What discoveries since *Mobile DNA II* could surpass that claim? Very simply: everywhere DNA is dynamic, and we now meet the elegant protein machines, co-evolved DNA partners, and diverse RNA choreographers. These pages hold something for every reader, beginning with the introductory overview of mechanisms. Novices will find some of the most lucid reviews of these complex topics available anywhere. Specialists will be able to pick and choose advanced reviews of specific elements, but will be drawn in by unexpected parallels and contrasts among the elements in diverse organisms. Biomedical researchers will find documentation of recent advances in understanding immune-antigen conflict between host and pathogen. Biotechnicians will be introduced to amazing tools for *in vivo* control of designer DNAs. And long-time aficionados will simply fall in love all over again.

Questions still abound about the Transposable Elements (TE) described in this volume. Perhaps none is more profound than the basis and consequences of TE diversity even among related genomes. Active DNA TE show perhaps the most disparate distribution among organisms, being dominant in prokaryotes, and in some animals, including some insects and fish, but with the exception of certain bats, virtually absent in mammals. Plants illustrate expansion of genomes, mediated not only by increasing ploidy, but also by expansion of DNA-based TE and Long Terminal Repeat (LTR) retrotransposons. Although reverse transcriptases are found throughout all kingdoms, autonomous retroelements simply explode together with their non-autonomous partners in mammals with remarkably species-specific types. These differences in mobile DNAs define self and mate, sister species, host and pathogen.

The most striking impression from these pages must be the raw power of genetic material to refashion itself to any purpose. DNA exchange between bacteria and their environment blurs the boundaries between host, transposon, and phage, as organisms secrete and take up DNA, stash genetic material in integrons for future use, conjugate, are attacked by phage and fight back. Delving into mechanisms, we see single-stranded hairpin structures and G quartets that anchor rearrangements in multiple ways; chemically diverse nucleophilic centers—hydroxyls couched in pentose, tyrosine or serine moieties—that covalently bond or attack directly in strand-transfer reactions. Proteins act as clamps to topologically constrain DNA or act as mechanical swivels, linking and unlinking mobilizing strands. Resolution of transposition intermediates might also involve host replication or recombination machinery. More recently discovered helitrons offer unexpected opportunities for expansion of DNA-based elements by rolling-circle replication.

RNA, the primal, catalytic nucleic acid, is in evidence everywhere. In retroelements, RNA partners with reverse transcriptase to deliver on transcriptional expansion of autonomous and non-autonomous TE sequences. Group II introns in bacteria likely gave rise to eukaryotic organelle group II self-splicing, retro-homing introns, Long INterspersed Elements (LINEs), telomerase reverse transcriptases and in addition, spliceosomal introns. Phylogenetic analysis of bacterial genomes previously revealed group II introns, diversity-generating retroelements Diversity-Generating Retroelements (DGR), and retrons, but next generation sequencing now identifies a multitude of novel reverse transcriptases of unknown function.

In ciliates, *Paramecium*, *Tetrahymena* and *Oxytrichia*, RNA directs massive genome reduction between germ-line and somatic nuclei, mediated by ancient transposase-like enzymes. LINEs containing restriction-enzyme like- or AP-endonucleases dominate in some eukaryotic cells. Others are dominated by LTR retrotransposons and their offspring, the retroviruses; stripped down Penelope-like elements with GIY-YIG endonucleases; DIRS elements with tyrosine recombinases: and attendant non-autonomous elements.

Exceptional elements provide evidence for the interaction of domains over evolutionary time, including LTR retrotransposons encoding envelope proteins, retroviruses replicating intracellularly, and DIRS elements in which retroelement RT/RNaseH is associated with a Crypton-type DNA element tyrosine recombinase.

Nowhere is the sharp focus of structural biology and biochemistry more apparent than in studies of key retroelement enzymes reverse transcriptase and integrase motivated by the quest for inhibitors of human immunodeficiency virus (HIV) replication. Reverse transcriptase structures for multiple retroviruses, as well as now one retrotransposon, demonstrate the robustness of the palm, thumb, fingers model. However, as a caution against too much generalization, subdomains are re-arranged in monomeric, homodimeric, and heterodimeric forms in different enzymes, and catalytic activities operate in *cis* or *trans* within the complex, depending on the enzyme. The structure of full length retrovirus integrase notoriously resisted high resolution structural analysis, but now has rewarded efforts of many labs with key insights (cover of this volume). These include a surprising dimer-dimer interface where active sites are juxtaposed to a trapped, and dramatically bent and widened, major groove target. Whereas LTR retrotransposons target integration to transcriptionally-repressed regions through interactions with heterochromatic protein domains or Pol III-transcribed genes thought to repress Pol II transcription, next generation sequencing has surfaced less dramatic, but significant, retrovirus integration bias, favoring transcriptionally-active regions.

This distribution has been shown now in two cases to be mediated by interactions between integrase and epigenetic mark-associated proteins.

While it has been argued that mobile elements are “selfish DNAs”, these pages are replete with examples of the positive contributions of mobile elements to host genome function. Bacterial transposons encode and mobilize selectable markers including antibiotic resistance, detoxifying enzymes, and conjugation and virulence functions. In eukaryotic cells, mobile elements contribute to chromosome structure: constituting centromeres or telomeres in some organisms and seeding heterochromatin in others. TE constitute a large fraction of transcription factor binding sites and provide an ongoing source of novel combinations which are responsive to stress signaling, MAP kinase activation and other developmental signals. Insertions of LINEs and Alu elements affect RNA processing because they encode cryptic splice sites, termination signals, and can target RNA editing.

Exapted mobile DNA coding sequences appear in novel contexts: transposases have evolved into the RAG endonuclease for V(D)J immunoglobulin gene diversification and into heterochromatic factor CENP-B; a reverse transcriptase evolved into telomerase; retrovirus envelope proteins became the trophoblast fusion protein syncytin. There are other examples of TE Open Reading Frames (ORFs) under selection, but with, as yet, unknown functions. Endogenous retroviruses forego prior allegiances and join strategies to resist new infections. For example, Fv-1, a retroviral Gag relic, thwarts incoming retroviruses of similar type. Repeated TE sequences are susceptible to DNA rearrangements via non-allelic recombination, aborted transposition, and generation of pseudo-genes—all of which might ultimately contribute to the resiliency of host genomes.

TE exploit their hosts as well. The bacterial XerCD tyrosine recombinases which function in bacteria to unlink multimeric chromosomes are exploited to integrate phage genomes or mediate invasion of the host chromosome on behalf of certain plasmid-borne mobile elements. Transposases, resolvases and integrases *in vivo* likely associate with host factors as they are joined with host genomes. TE are generally tightly controlled by host regulation so that some display opportunistic bursts of activity during specific windows of development. This is exemplified by yeast Ty transcription in response to MAP kinase signaling and activation of reverse transcription by DNA checkpoints. A common theme more generally is TE activation during stress. Diverse retroelements are derepressed during periods of germ cell development ensuring their vertical spreading in populations.

Despite these examples of cooperation, mobile DNAs are also in conflict with their hosts. RNA interference likely arose in part to combat mobilization of retroelements. Invaders of one sort or another engage in a dizzying unscored dance with their hosts. One result of this conflict is rapid evolution of genes encoding host innate immunity restriction factors, which for retroviruses include ones that prematurely uncoat incoming viruses, starve reverse transcriptases for nucleotides, and deaminate cytidines in replicating cDNA. Some of these same factors also suppress movement of endogenous retroelements.

Programmed variation is used by invaders and hosts alike for purposes of immune evasion or resistance, respectively. Examples include *Salmonella* Hin invertase flipping a promoter sequence to switch between expression of different antigenic flagellar proteins and DGR directing mutagenic reverse transcription of a template transcript coupled with directed conversion of a target expression site. *Neisseria gonorrhoea*, *Borrelia burgdorferi*, *Trypanosoma brucei*; and *Plasmodium falciparum*, agents of gonorrhoea, Lyme disease, sleeping sickness, and malaria, respectively, use amazingly diverse mechanisms to program variation of their antigenic surfaces for immune evasion. To counter this assault, there

is programmed variation of host immune proteins. In human immunoglobulin production, a DDE TE-derived RAG site-specific endonuclease initiates V(D)J switching, followed by transcription-activated somatic hyper-mutation (activation-induced cytidine deaminase), nuclease introduction of DSB, and final joint formation by redundant NHEJ pathways.

Next generation sequencing and development of methods for rapid TE mapping have greatly improved understanding of the distribution of TE as well as the utility of transposons for functional genomics. The bacterial Tn5 system has been exploited in particular for *in vitro* mutagenesis and next generation sequencing libraries by collapsing fragmentation and adapter ligation into a single step. P, Hermes, piggyBac, and Sleeping Beauty transposons have wide activity in eukaryotic systems and have been harnessed for genome-wide profiling, gene disruption and tagging, and genome modification. Retroviruses are additionally used in lineage tracing. The controlled, high-frequency mobilization of Mutator has made it indispensable for gene discovery in maize.

In medical research, understanding the impact of DNA mobilization is critical. In addition to individual TE, other mobile DNAs such as plasmids, Integrative Conjugative Elements (ICE) and both transposon-borne and chromosomal integrons are bacterial reservoirs of mobilizable antibiotic resistance. HIV, malaria, and sleeping sickness, and other pathogens, too numerous to mention here, remain threats to global health. Mobile element vectors transposons piggyBac, Sleeping Beauty, lenti-retroviruses and adenoviruses are being used as vectors to introduce exogenous DNAs in research, and in clinical trials. They differ with respect to targeting, excision footprints, payload size, and host activity profiles. Their mechanisms of DNA breakage and joining were among the systems first analyzed, now enabling them to be harnessed and used extensively for genome engineering including with developmentally-regulated expression, inactivation, and self-deletion strategies to enable probing essential or tissue-specific functions.

What challenges remain? One goal is to connect key findings from basic research, to clinical developments in drug resistance and genetic engineering. This volume is based on the considerable increase in understanding of molecular mechanisms of mobilization in the last decade. However, we have likely seen only the tip of the iceberg of how mobile DNAs affect the day to day biology of their hosts.

In the human genome alone, retroelements provide promoters for long non-coding and other RNAs of completely unknown function; Alu elements redirect RNA processing and delivery, and mobilization is occurring during neuronal development and in cancer with unknown consequences, just to mention a few. Finally, endogenization of a gamma retrovirus in Australian koalas is ongoing and those studies should provide insights into retroelement-host interaction. How have transposition events after separation from other great apes contributed to traits that make us human? What transposition events will provide key substrates for future evolution? And of course, perhaps the ultimate question, could we survive as a species were there no transposition?

We give our heartfelt thanks to all the authors who contributed diverse and fascinating chapters to Mobile DNA III. We express special thanks to Patti Kodeck, Administrative Assistant to Editor in Chief N. Craig, who mediated recruitment of and communications with authors and interactions between them and the publisher. Finally, our most sincere thanks to all of our supporters at ASM Press for their dedication in producing this volume, but especially to: Gregory Payne, Senior Editor; Larry Klein, Production Editor; Christine Charlip, Director of Administration; and Cathy Balogh, Administrative Assistant for Production.

Introduction



Nancy L. Craig¹

A Moveable Feast: An Introduction to Mobile DNA

1

INTRODUCTION

DNA has two critical functions: to provide the cell with the information necessary for macromolecular synthesis and to transmit that information to progeny cells. Genome sequence stability is important for both these functions. Indeed, cells devote significant resources to various DNA repair processes that maintain genome structure and repair alterations that can arise from DNA synthesis errors and assaults from both endogenous and exogenous sources. DNA sequence variation, however, provides the substrate for adaptation, selection, and evolution.

Genomes are, however, highly dynamic. Notably, they vary not only at the single or several base pair level (although such changes can be transformative and even deadly), but they also change by DNA rearrangements, that is, the movement of DNA segments that may be many kb (or even longer) in length. Such rearrangements can have enormous impacts on genome structure, function, and evolution.

The DNA rearrangements discussed here generally involve specific DNA sequences, or in some cases RNA sequences, that are recognized and acted on by specialized recombination proteins or recombinases that promote DNA breakage followed by joining of the broken DNAs to new sites. The involvement of a sequence-specific

recombinase is what distinguishes site-specific recombination from homologous recombination, which can occur between any two DNA segments as long as they are homologous to each other, as in RecA- and Rad51-dependent recombination. In some cases, the specialized recombinase is a sequence-specific nuclease that targets homologous recombination to a specific DNA site.

In some rearrangements, the recombinase alone breaks, exchanges, and joins DNA by using covalent protein-DNA intermediates. In other cases, DNA synthesis is also essential in these rearrangements. Notably, this DNA synthesis can involve not only conventional DNA synthesis in which a DNA polymerase uses DNA as a template, but also reverse transcription in which a novel DNA polymerase, reverse transcriptase, uses an RNA template to generate new DNA. A very wide number of other cellular processes can influence or be required for DNA rearrangements, including transcriptional activation of particular sites, DNA bending by bending proteins, DNA supercoiling, and many variations in chromatin structure, as well as DNA repair reactions including DNA end joining. Although a purified recombinase may execute DNA breakage and joining *in vitro*, it is critical to remember that this reaction and its consequences will be enormously influenced by its cellular environment.

¹Department of Molecular Biology and Genetics, Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

Although DNA rearrangements can provide very useful rapid and focused changes in genetic information, they are also very hazardous. Unrepaired DNA breaks can result in DNA mis-segregation and are often lethal. Not surprisingly then, DNA rearrangements often occur in elaborate nucleoprotein complexes that organize and juxtapose the participating DNAs and promote breakage and joining in carefully choreographed steps. The frequency of DNA rearrangements is usually highly regulated, often by restricting to low levels the recombinase that initiates or mediates the rearrangements.

Mobile DNAs also include a diverse variety of discrete mobile genetic elements, such as transposable elements, that move themselves or copies of themselves from place to place within and between genomes. Thus, in some cases, a copy of the element remains at its original site and there is a new copy at the new insertion site. This type of replicative mechanism leads to an enormously high element copy number, especially in some eukaryotic genomes. The majority of the maize genome, for example, is derived from a particular kind of transposon. High copy numbers of repetitive sequences result in increased susceptibility to nonallelic homologous recombination events that can lead to deletions, inversions, translocations, and other chromosomal rearrangements.

The movement of a transposable element within a single genome can have substantial genotypic and phenotypic consequences. The insertion of a transposable element into a gene can lead to gene disruption but even nearby insertions can effect gene expression as many elements carry regulatory signals, such as enhancers and promoters, as well as splice sites and transcription termination signals. Excision of an element also changes the donor site. Thus, the intracellular translocation of a mobile element results in genetic variation. The range of target sites used by the elements ranges from insertion into specific sites or regions that provide a "safe harbor" for the element with reduced negative consequence on the host, to preferences for actively transcribed regions to facilitate element expression to virtually random insertion, which can thus result in genetic variation anywhere within the host genome.

DNA rearrangements also play a crucial role in the interactions between viral chromosomes and their hosts, as well as the proper replication and segregation of host chromosomes. Many viruses integrate into and excise from host genomes, although in some cases integration is irreversible, such as with HIV-1. All of these reactions involve at least specific sites on the viral genome that are acted upon by site-specific recombinases and which sometimes involve specific sites on the host

genome. Recombination between specific sites to promote chromosome monomerization plays a key role in chromosome transmission in bacteria.

The translocation of mobile elements encoding a wide variety of determinants including genes encoding antibiotic resistances, virulence determinants, and diverse metabolic pathways from plasmids to chromosomes and from viruses and DNA fragments that are transduced or transformed into a cell, can also result in permanent chromosomal acquisition of these determinants. This sort of horizontal gene transfer involving mobile elements is rampant in bacteria and contributes greatly to genetic variation. There are also an increasing number of examples of horizontal gene transfer involving mobile elements in eukaryotes.

Perhaps the most profound example of the effect of mobile elements on eukaryotic genome evolution is the nuclear invasion of mobile group II introns into the nuclear genome from bacterial symbionts to form spliceosomal introns.

Cell type can also have substantial impact on DNA rearrangements. The elaborate DNA breakage and joining reactions that underlie immunoglobulin diversification are actually terminal differentiation events restricted to particular somatic cells. There is increasing interest in the somatic movement of transposable elements, which can also have profound organismal impact. The movement of human transposable elements in somatic tissue is associated with a variety of cancers, although it remains to be determined if such events can cause oncogenic transformation or are rather a consequence of transformation. The movement of transposable elements in neuronal tissue in several organisms raises the interesting possibility that such rearrangements are a deliberate strategy for neural plasticity.

Such terminal differentiation events involving DNA rearrangements are incompatible with the bacterial lifestyle, except in a few known cases such as spore formation by a subset of cells. By contrast, reversible DNA inversions that vary promoter or ORF orientation are well known in bacteria.

Thus, DNA rearrangements can contribute substantially to genetic variation. The frequency and potential advantage of the resulting variation must be carefully balanced with genome stability to avoid its potential for population-wide genomic catastrophe.

Although not exclusively so, the focus of this work is on the mechanism and regulation of DNA rearrangements. How do specific DNA (and sometimes RNA) sequences recognize each other and how do they assemble to form the machines in which DNA rearrangements occur? What are the mechanisms for DNA

strand breakage and joining? What processes determine when and where these reactions occur? How are actions at multiple DNA sites, for example, the two ends of a transposable element and its target DNA, coordinated? Importantly, how are nonproductive breakage and joining events avoided and how is intact duplex DNA regenerated?

Mobile DNAs are “natural” genome engineers. Although not a focus of this work, many of the mobile elements discussed here have been harnessed to facilitate researcher-directed rearrangements both *in vitro* and *in vivo*. Mobile elements are used for “random” insertional genome mutagenesis both *in vivo* and *in vitro*, as well as for “targeted mutagenesis.” Many mobile elements are used as vectors in both homologous and heterologous systems.

TARGETED DNA BREAKS LEAD TO GENE REPLACEMENT

DNA Double Strand Breaks Stimulate Homology-Dependent DNA Repair

Homologous recombination occurs without requirement for any particular sequence, depending only on base pairing between the participating DNA strands. However, the frequency of homologous recombination is stimulated by the presence of broken DNA, in particular double strand breaks. These breaks stimulate recombination because the action of nucleases and helicases at these breaks leads to the generation of DNA with single stranded 3′ trails that are the preferred substrate for DNA pairing mediated by RecA- and Rad51-like proteins. By interacting with a donor site, this pairing of 3′-OH ends can initiate homology-dependent DNA repair, which copies DNA sequence information from the donor site into the broken DNA target site. This repair leads to the replacement, or modification, of an existing gene or insertion of a new gene. The insertion of many mobile DNAs into a new site is targeted by double strand breaks by highly site-specific endonucleases.

There’s No Place Like Home: Homing Endonucleases

Homing endonucleases (HENs) are highly site-specific endonucleases (1). Although often associated with other genetic elements (see below), freestanding HEN genes can themselves be mobile DNA elements. If a HEN cleavage site lies within an “empty allele” of DNA that flanks the HEN ORF, cleavage of that target site can initiate homology-dependent DNA synthesis that will

transfer a copy of the HEN gene to that double strand break at the nuclease target site (Fig. 1).

HEN genes are also often found in other genetic elements such as self-splicing RNA introns, that is, group I introns, and self-splicing proteins, that is, inteins. Thus, if the HEN introduces a double strand break into the “empty” allele of a site occupied by the intron or intein, targeted DNA repair introduces a copy of the DNA encoding the intron or intein into that target site. Because the RNA intron can splice out of the RNA containing it and the protein intein can splice out of the protein containing it, the insertion of these elements is generally phenotypically silent.

Alternative Life Styles: Switching Mating Type in *Saccharomyces cerevisiae*

These yeasts have two different haploid cell types, mating type α and mating type a , which can mate to form diploids. During sporulation, meiotic recombination shuffles the two parental genomes, generating diverse haploid progeny. To facilitate diploid formation, haploids can switch mating type from mating type a

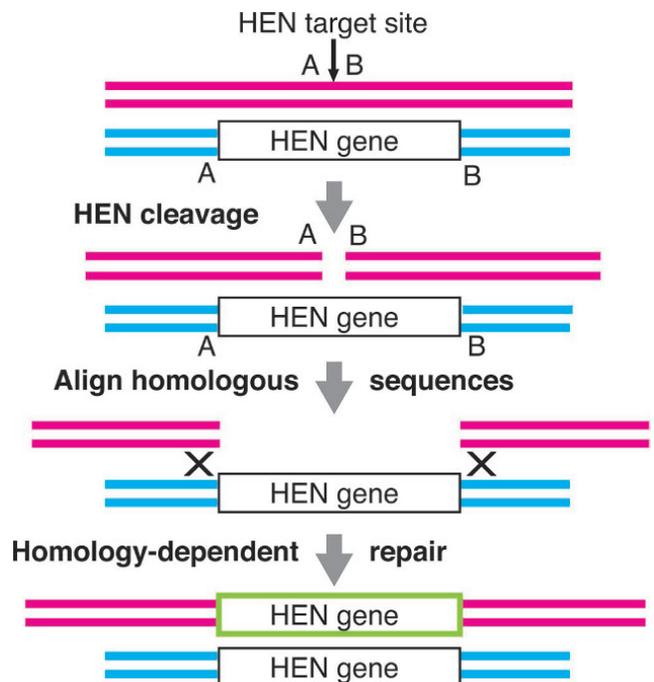


Figure 1 A targeted DNA double strand break can lead to gene insertion. Introduction of a site-specific double strand break by a homing endonuclease (HEN) in a homologous DNA duplex lacking the HEN gene targets homology-dependent DNA synthesis (green) that introduces a copy of the HEN gene to the broken DNA.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f1

to mating type α and from mating type α to mating type a (see Chapter 23).

MAT is the mating type expression site, which can express either of two different mating type regulators. One regulator set controls mating type a gene expression and the other regulator set controls mating type α gene expression. Mating type switching occurs when the highly site-specific HO nuclease, which is a member of the HEN family, introduces a double strand break into the *MAT* expression site (Fig. 2). This double strand break initiates homology-dependent DNA repair using either one of the two nonexpressed, silent storage copies of mating type information called *HML* α and *HMR* a , as a template to replace the mating type information at the *MAT* expression site. *HML* α has a silent copy of mating type α information and *HMR* a has a silent copy of the mating type a information. HO expression and the choice of *HML* α or *HMR* a as a donor site are highly regulated. When the *MAT* expression

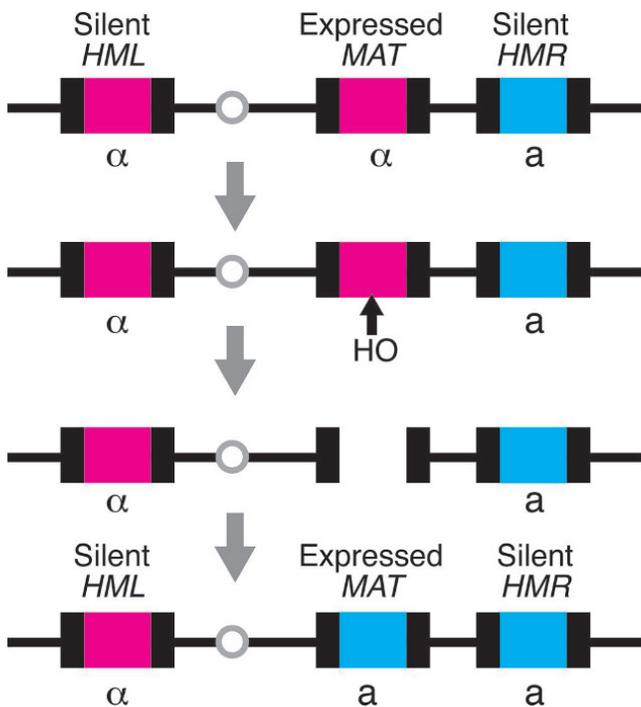


Figure 2 A targeted DNA double strand break causes mating type switching in *Saccharomyces cerevisiae*. Information that specifies either mating type a or mating type α is expressed at *MAT* and is also present at silent storage sites *HML* and *HMR*. Introduction of a DNA double strand break in *MAT* by the endonuclease HO targets homology-dependent DNA synthesis at *MAT* using either the silent *HML* and *HMR* storage sites as templates. DNA breaks in *MAT* a cells use silent *HML* α as the donor, thus switching *MAT* a to *MAT* α , resulting in a switch in mating type from a to α . Conversely, *MAT* a cleavage results in a switch to *MAT* α .

doi:10.1128/microbiolspec.MDNA3-0062-2014.f2

locus contains mating type information of one type, switching information from the silent locus of the opposite mating type is copied into the *MAT* expression site, thus switching a mating type α cell into a mating type a cell or a mating type a cell into a mating type α cell.

A different strategy is used for mating and cell type switching in several fission yeasts including *Schizosaccharomyces pombe*. Although these yeasts also have a *MAT* expression locus and two silent storage sites for mating type information, switching occurs, not by targeted DNA repair, but rather by another mechanism involving strand- and site-specific imprinting of one of the DNA strands at the *MAT* expression locus (see Chapter 24).

Taking Evasive Action: Changing Cell Surface Proteins to Elude the Host Immune Defense

A key step in successful pathogen invasion of a new host is evasion of the host's immune response directed against pathogen cell surface antigens. Many pathogens evade the immune system by changing their cell surface proteins, often by DNA rearrangements. The simplest variation system is to alternate expression between two different cell surface proteins. Switching between expression of two different surface protein types occurs in *Salmonella* via a DNA inversion. This DNA inversion flips an invertible segment containing a promoter, which in one orientation, drives expression of one surface flagellar protein, and in the other orientation, drives expression of the alternative flagellar protein (see below; see Chapter 9).

In other systems, homologous recombination between multiple silent gene variants and a single gene expression site underlies the alternate expression of surface protein variants (Fig. 3). The bacterial pathogens *Neisseria gonorrhoeae* and *N. meningitidis* contain about 20 silent variant copies of a surface pilin gene, *pilS*, and a single pilin gene expression site, *pilE* (see Chapter 21). *pilS* gene segments are transferred to *pilE* by RecA-dependent homologous recombination that appears to be stimulated by nicking of a DNA G4 quadruplex upstream of *pilE*. The surface lipoprotein VlsE of the bacterial spirochete *Borrelia burgdorferi*, which causes Lyme disease, also undergoes antigenic variation (see Chapter 22). As in *Neisseria*, there is a single *vlsE* expression site, which contains a required G4 quadruplex, and multiple variant silent *vlsS* genes whose sequences are transferred to *vlsE*. Interestingly, this reaction does not require RecA.

The *Borrelia* genome includes a linear chromosome and multiple linear plasmids with hairpin ends.

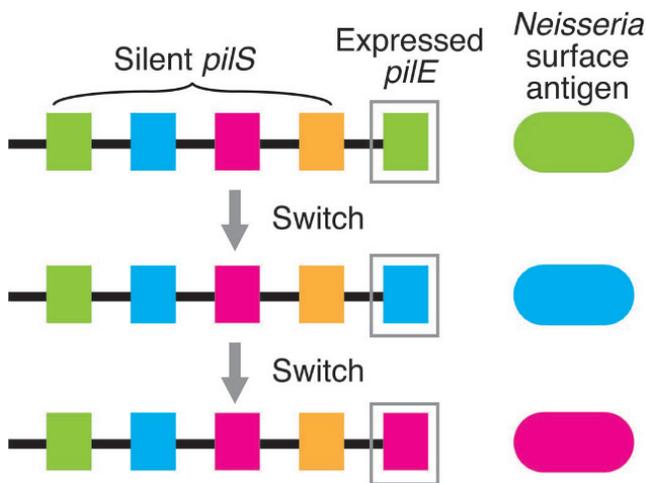


Figure 3 Gene replacement underlies antigenic variation in *Neisseria*. A pilin surface antigen is expressed from the *pilE* site and multiple variant pilin genes are stored in silent *pilS* sites. Homology-dependent repair using template information from a *pilS* gene changes the information in *pilE*, varying the surface antigen.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f3

To complete replication, these ends are processed by the proto-telomerase ResT, a phospho-Tyrosine recombinase (see below; see Chapter 12).

Antigenic variation of the variant surface glycoprotein (VSG) of *Trypanosoma brucei*, the protist that causes sleeping sickness, involves some 2,500 silent *vsg* gene variants and multiple *vsg* expression sites (see Chapter 19). The introduction of double strand breaks into repeats that flank the *vsg* expression sites results in gene switching but how such breaks might be generated remains to be determined.

Antigenic variation of several large multi-gene families of surface proteins also occurs in the malaria parasite *Plasmodium falciparum* (see Chapter 20). The best-studied system is the *var* family with about 60 members clustered in several different arrays. In this case differential gene expression regulated *in situ* mediates antigenic variation, but gene copies are also diversified by recombination.

LESS IS MORE: ACTIVE GENE ASSEMBLY BY DELETION

The Immune System Strikes Back: Immunoglobulin Gene Diversification Allows Detection of Millions of Antigens

In the vertebrate adaptive immune system, the B cells that make antibodies and T cells that make antigen receptors can produce millions of diverse immunoglobulins

that can recognize different antigens. Multiple processes underlie these gene diversifications. In both B and T cells, the combinatorial assembly of many different V, D, and J coding gene segments, which encode different protein segments to form the variable regions of immunoglobulin genes and proteins, is mediated by V(D)J recombination (see Chapter 14). The loci encoding these gene segments contain many, in some cases hundreds, of these coding segments separated by nonimmunoglobulin spacer DNA. In B cells, somatic hypermutation (SHM) then targets the assembled V(D)J gene segments to further diversify the variable regions (see Chapter 15). Class switch recombination (CSR) then adds one of several different classes of constant regions to the assembled, mutated V(D)J variable regions to yield antibodies for several different cellular activities (see Chapter 15).

V(D)J Recombination

An immunoglobulin locus contains arrays of multiple V (variable), D (diversity), and J (joining) gene coding segments upstream of a constant gene-coding segment. These V, D, and J gene-coding segments are assembled by the combinatorial joining of different V, D, and J coding gene segments, yielding many different $V_xD_yJ_z$ coding regions by V(D)J recombination (see Chapter 14). This combinatorial gene assembly occurs by the introduction of targeted DNA double strand breaks at the edges of the V, D, and J segments to be joined, resulting in excision of the DNA between the targeted V, D, and J segments. The coding segments are joined by nonhomologous end joining.

A highly specialized site-specific nuclease called RAG makes these initiating double strand breaks, acting at recombination signal sequences (RSSs), at the edges of the V, D, and J gene coding segments (Fig. 4). These RSSs may be either of two slightly different forms, RSS12 or RSS23, and RAG binding to and pairing of one RSS12 type and one RSS23 type site is required for RAG-dependent cleavage. RAG activity is also elaborately regulated *in vivo*. The RSSs are accessible to RAG only when activated by regulated transcription through them and by particular chromatin modifications. The interaction of two RSSs that may be separated by many tens of kb is facilitated by the highly structured 3D organization of the genome in each immunoglobulin locus (2).

The different positions and orientations of the RSS signals at the edges of the V, D, and J coding segments organize the breakage by RAG at particular positions, followed by the joining of V segments to D segments, and the joining of D segments to J segments. The joining

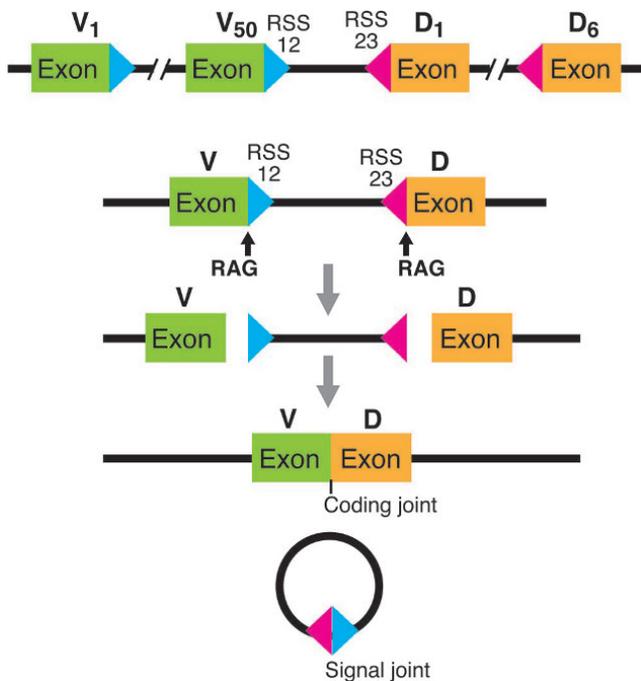


Figure 4 Targeted DNA double strand breaks by RAG mediate immunoglobulin gene assembly during V(D)J recombination. Diverse immunoglobulins that recognize many different antigens result from combinatorial assembly of different V, D, and J coding segments. Site-specific cleavage by RAG, at RSS12 and RSS23 sites that bound the multiple V, D and J segments, results in excision of intervening DNA between the gene segments to be joined and formation of a coding joint by NHEJ. The intervening DNA circularizes, forming a signal joint and is lost from the cell.
doi:10.1128/microbiolspec.MDNA3-0062-2014.f4

of V to D segments and D to J segments creates coding joints that link these protein-coding segments, which will then be transcribed into mRNA. The joining of the RSS12 to RSS23 sites creates the signal joint on the excised intervening DNA segment, which is lost from the cell.

RAG, whose structure has been recently determined (3), is closely related to DDE transposases (see below; see Chapter 25), that bind to and break at terminal inverted repeats (TIRs) at the ends of DDE transposons. Indeed, the DNAs between the coding segments, which are to be joined, are bounded by inverted RSSs just as the ends of DDE transposons are bounded by TIRs. Moreover, the chemical steps by which double strand breaks are made at RSSs by RAG and by which double strand breaks are made at TIRs at the ends of *hAT* transposons such as *Hermes* (see Chapter 35), occur by the same chemical mechanism (4, 5). This mechanistic similarity and the marked similarities between the

structures of RAG and *Hermes* (3,6) support the hypothesis that the RAG recombination system evolved from an ancient *Transib* transposon (7, 8).

Somatic Hypermutation

The V(D)J segment in an active immunoglobulin gene in a B cell is then subjected to local, targeted, high-frequency mutagenesis called somatic hypermutation (SHM) to further diversify the antigen binding region (see Chapter 15). This mutagenesis is targeted to particular regions by regulated transcription that recruits the activation-induced cytidine deaminase (AID). Deamination converts cytidine to uracil resulting in a U::G mismatch. Mutagenic repair, likely by error-prone polymerases, then occurs by base excision repair or mismatch repair, resulting in localized, high frequency mutagenesis that further diversifies the antigen-binding region of the antibody.

Class-Switch Recombination

Downstream of the V(D)J coding exon segments are multiple, different “C” (constant) exons that encode various immune system effectors. The V(D)J coding exon and a C exon are joined in the mRNA by RNA splicing. Class-switch recombination (CSR) positions particular, different isotype C exons immediately downstream of the V(D)J exon by DNA excision of the intervening C exons. This excision occurs by targeted double strand breaks at a switch site downstream of the V(D)J exon and the switch site upstream of a particular C exon (Fig. 5). As in SHM, transcription-directed AID changes cytidines to uracil in the switch sites, which are then further processed to introduce double strand breaks. The cleaved switch site downstream of the V(D)J coding segment then joins to the cleaved switch site upstream of a C exon by nonhomologous end joining. The intervening DNA is lost from the cell.

Massive Chromosome Destruction Leads to Active Ciliate Genes

The most spectacular examples of programmed DNA breakage and joining events to restructure chromosomes and genes (9) are those in the ciliates *Paramecium* (see Chapter 17), *Tetrahymena* (see Chapter 16), and *Oxytricha* (see Chapter 18). It should be noted, however, that large-scale chromosome diminution also occurs in some nematodes, insects, and vertebrates (see Chapter 18) (10).

Ciliates are single cell eukaryotes that have two nuclei, a germline micronucleus of ~50 to 100 Mbp, which is transmitted to progeny, and a somatic macronucleus

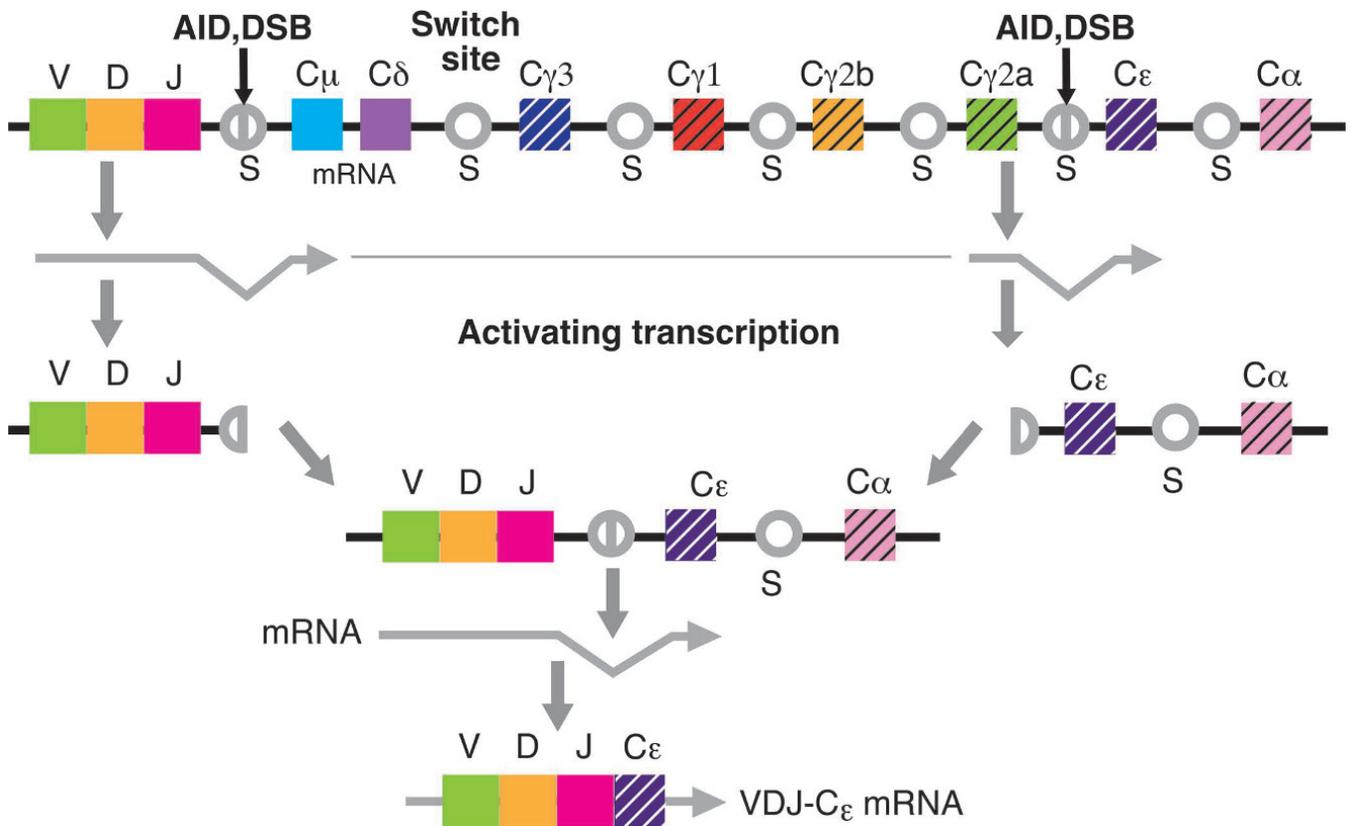


Figure 5 Targeted DNA double strand breaks at switch sites join V(D)J coding segments to different antibody class segments. DNA double strand breaks (DSB) are targeted by transcription-induced activation-induced cytidine deaminase (AID) modification of switch sites downstream of an assembled V(D)J coding region and upstream of different antibody class coding regions. The intervening DNA is excised and combinatorial joining of the V(D)J and antibody class segments by NHEJ results in different classes of antibodies.
doi:10.1128/microbiolspec.MDNA3-0062-2014.f5

from which all genes are expressed. In the macronucleus, the much larger micronuclear chromosome has been shattered by double strand breaks into smaller DNA fragments. Macronuclear development also involves excision and loss of a substantial fraction of the micronuclear genome, ranging from ~30% in *Tetrahymena* to almost 95% in *Oxytricha*. These internal eliminated sequences (IESs) lie in and between genes in the micronucleus, and thus, must be carefully excised and the genic material rejoined to maintain gene function. In *Paramecium* and *Tetrahymena*, many of these IESs are excised by a domesticated *piggyBac* transposase (see Chapter 39), which introduces targeted double strand breaks at the edges of the elements, following which the gene segments are rejoined by nonhomologous end joining. A smaller number of IESs are excised by other transposase-like proteins that make targeted DNA double strand breaks. An attractive view is that

the IESs began as transposon insertions and that their elimination is an extreme form of gene silencing.

RNA plays a prominent role in these programmed DNA rearrangements. Formation of the macronucleus is completed by the addition of new telomers to the ends of the new “minichromosomes.” Experiments in *Tetrahymena* lead to the proposal (11) that a new class of RNA, called scan RNA (scrRNA), actually directs the elimination of IES sequences. The scrRNAs specific for the sequences to be maintained in the next round of macronuclear development derive from pervasive transcription of the entire genome, which is then compared to the edited macronuclear DNA, thus, identifying the scrRNAs.

An even more remarkable role for RNA has been revealed in *Oxytricha* macronuclear development (see Chapter 18). Here, the micronuclear genes are not only interrupted by transposon-like sequences that are removed by *Tc1/mariner* transposases (see below), but

gene segment order is highly scrambled. For example, if the linear order of the segments that make up a gene in the macronuclear DNA is ABCDE, their order in the germline micronucleus may be BADCE! How can a sensible gene be assembled from this scrambled genome? The answer lies in gene-length RNAs with the correct gene segment order that are transcribed from the macronucleus and are then transported to the micronucleus where they provide a template for the correctly ordered assembly of genes in the next round of macronuclear development. The mechanism of this assembly process remains to be determined.

TRANSPOSABLE ELEMENTS

Transposable elements, which are present in virtually all genomes, are discrete DNA segments that can move themselves or a copy of themselves within and between genomes. Transposable elements underlie a wide variety of processes, such as the interaction of viral and host chromosomes, reactions that underlie the replication and accurate segregation of chromosomes and the regulation of gene expression. Transposable elements may also encode a wide variety of accessory determinants including antibiotic resistance genes, virulence determinants, and a wide variety of metabolic genes. Moreover, the movement of a transposable element can have a profound effect on host gene expression. Element insertion into a gene can result in gene inactivation. Element insertion near a gene can also alter gene expression as transposable elements can also encode regulatory signals such as promoters, enhancers, splice sites, polyadenylation signals, and transcription termination signals. These variations in genetic information are substrates for adaptation, selection, and evolution.

Other than viruses, it has long been thought that only mobile elements in bacteria carried accessory genes such as antibiotic resistance genes. Multiple examples of mobile elements containing genes or gene fragments have now been observed in eukaryotes, however, most notably with the DNA-only *Helitrons* (see Chapter 40) (12) and versions of *Mutator* transposons called *Pack-MULEs* (see Chapter 36) (13).

Elements that encode their own mobility functions, for example, a transposase, and its cognate recognition sequences at the transposon ends are called “autonomous” elements. Some elements, however, encode only the necessary *cis*-acting sequences at the transposon ends and are mobilized *in trans* by transposase from another autonomous version of the element elsewhere in the genome. These elements are called “nonautonomous” elements.

Some transposable elements move only via DNA intermediates. The DNA intermediates of some other transposable elements are generated by reverse transcription of RNA into double stranded DNA, which then interacts with and inserts into a target site. Transposition of another major class of elements occurs by the interaction of an RNA form of the element directly with a target DNA, followed by reverse transcription of the RNA *in situ* at the target site to generate a new DNA copy of the element. This ability to convert RNA copies of an element into DNA and the resulting amplification likely leads to the very high copy number of retroelements in some organisms.

There are many types of transposable elements that have different structures and move by different mechanisms. Confusingly, but not surprisingly, because many elements were first identified in different organisms and were named in the absence of molecular understanding of how they moved, what turn out to be mechanistically very related elements can have quite different types of names. Conversely, other elements that move by very different mechanisms can have similar names. For example, many bacterial insertion sequences move by the same mechanism, as do many eukaryotic transposable elements. The transposition mechanism of bacterial *IS4* family members, including *IS10* and *IS50* that form *Tn10* and *Tn5* (see Chapter 29), is related to that of the eukaryotic transposable element *piggyBac* (see Chapter 39). The transposable bacterial insertion sequence, *ISY100* (14), uses the same breakage and joining steps as do transposable *ITm* (*Tc1/mariner*) elements found in many eukaryotes (see Chapter 34). Conversely, the integrases of many bacterial viruses use a very different mechanism for DNA breakage and joining than do the integrases that mediate the integration of the DNA forms of retroviruses.

DNA-ONLY TRANSPOSONS

Transposases are sequence-specific DNA binding proteins that also contain a catalytic domain that mediates DNA breakage and joining. Some transposable elements move by breakage and joining mediated only by the transposase, whereas others also involve DNA synthesis and ligation by host proteins to regenerate intact duplex DNA. There are four major classes of DNA-only transposases: DDE transposases, tyrosine-histidine-hydrophobic-histidine (HUH) transposases, tyrosine-transposases, and serine-transposases. DDE transposases break and join DNA by direct transesterification. The other classes of transposases act via covalent-protein DNA intermediates. Eubacteria, archaea,

and eukaryotes all contain mobile elements with these four major classes of transposases.

A new class of transposons, casposons, and their novel transposases, was recently proposed based on bioinformatic analysis of bacterial mobile elements that also encode a DNA polymerase (15). These bacterial elements have been called casposons because the protein proposed to be their transposase (integrase) derives from Cas1, a protein component of bacterial adaptive immunity CRISPR-Cas systems (16). These immunity systems take DNA sequences from infecting nucleic acids, such as viruses, and incorporate them into “clustered regularly interspaced short palindromic repeats” (CRISPR) arrays. Transcription of these arrays generates RNA copies of these incorporated sequences, which are used as guide sequences for nucleolytic attack and destruction of re-invading foreign DNAs and their transcripts. Notably, it has recently been shown that a complex of Cas1-Cas2 proteins can integrate a new DNA segment into a target DNA *in vitro* by direct transesterification using a 3′-OH end as a nucleophile (17). This is the same chemical mechanism used for DNA joining as that used by DDE transposases (see below; see Chapter 25) and the closely related retroviral integrases (see Chapter 44). Notably, however, the structure of the Cas1-Cas2 complex (18), which has been determined by X-ray crystallography, has no structural homology with DDE transposases, thus, identifying *Cas1* as a new class of transposase.

Such apparently DNA self-synthesizing elements in eukaryotes called Polintons (Mavericks) that also encode a DDE transposase have also been described bioinformatically (19).

Double Strand DNA Transposons Move via DDE Transposases

DDE transposons are discrete DNA segments bounded by terminal inverted repeats (TIRs) that are the specific binding sites of the cognate DDE transposase (see Chapter 25). The TIRs position the transposase at each end of the element to carry out the DNA breakage and joining reactions. DDE transposases carry out two closely related reactions: (i) the cleavage of a DNA phosphodiester bond using water as the nucleophile to yield 3′-OH and 5′-P ends; and (ii) the joining or “strand transfer” of the 3′-OH transposon end to a target DNA in which the 3′-OH is the nucleophile in a direct transesterification reaction. The active site of a DDE transposase is formed by an RNase H-like fold that closely juxtaposes three conserved acidic amino acids - D, D, and E - to position the essential Mg²⁺ ion cofactors.

Thus, DDE transposases are sometime called RNase H-like transposases. As described below, retroviral integrases (see Chapter 44) are also DDE transposases.

There are many superfamilies of DDE transposable elements (20), which are defined by similarities in the transposase sequence, the sequence of the transposon ends, and in some cases their target sequence (Fig. 6). Members of some superfamilies are present in both bacteria and eukaryotes.

Different DDE transposases use different combinations of breakage and joining steps to disconnect at least their 3′ ends from the donor site (Fig. 6) (see Chapter 25 for an overview and element-specific chapters). The key events in transposition of all DDE elements are the release of the 3′-OH transposon ends from the donor site, which then attack and join to the target DNA at staggered positions on the top and bottom target strands. These joining reactions result in the covalent linkage of the 3′ transposon ends to the target DNA. Because of the staggered joining positions on the top and bottom strands, single strand gaps extend from 3′-OHs on the flanking target strands to the 5′ transposon ends. These 3′-OH target ends provide the primers for the DNA synthesis that will repair these gaps or, in some cases, copy the entire element. Repair of these gaps by host proteins can occur by several different mechanisms (see Chapter 31) (21), resulting in target site duplications, which are a hallmark of transposition.

Insertion of the eukaryotic DDE *Spy* element, however, occurs without target site duplication (22). Likely, the excised 3′-OH transposon ends join the target DNA at nonstaggered positions.

Transposition reactions occur within elaborate nucleoprotein complexes called transpososomes, whose assembly is a key control point in transposition, which bring the transposon ends and the target DNA together, such that uncoordinated unproductive events do not occur. In some systems, host DNA bending proteins are important for transpososome assembly. Known transpososomes contain at least a dimer of transposase, the active site of each transposase protomer mediating breakage and joining at one transposon end. Thus, protein-protein interactions for oligomerization are also important features of transposases.

In some systems, multiple transposon-encoded proteins are required for transposition. For example, the *Tn7* transposase contains two different polypeptides, each one cleaving a particular strand (see Chapter 30) (Fig. 7). Some systems, for example, *Mu* (see Chapter 31) and *Tn7* (see Chapter 30) encode an ATP-dependent DNA binding protein involved in target site selection. *Tn7* also encodes additional target site selection proteins.

DDE Transposons							
Cut & Paste							
Family	Element	Hairpins on transposon ends	Hairpins on donor flank	2-strand cleavage	Nick Copy-out Paste	Nick Paste Copy	Reference/Chapter
<i>IS4</i> <i>piggyBac</i>	<i>Tn5, Tn10</i>	■					29
	<i>piggyBac</i>	■					39
<i>hAT</i> <i>Transib</i>	<i>Hermes</i>		■				35
	RAG (RAG1)		■				14
<i>ITm</i> (<i>IS630-Tc1-mariner</i>) or <i>Tc1-mariner</i>	<i>Mariner</i>			■			34
	<i>Sleeping Beauty</i>			■			38
	<i>Mos1</i>			■			Richardson et al 2009
	<i>IS100</i>			■			Feng & Colloms 2007
<i>P</i>	<i>P</i> element			■			33
<i>Mutator</i>	<i>MuDR</i>						36
	<i>MULE</i> <i>Os3378</i>						Zhao et al 2015
	<i>IS256</i>						Hennig & Zeibuhr 2010
<i>IS3</i>	<i>IS911</i>				■		27
<i>Tn3</i>	<i>Tn3</i>					■	32
	<i>Tn4430</i>					■	

Figure 6 Different families of DDE transposases mediate the transposition of different elements. Different superfamilies of DDE transposases use different combinations of DNA breakage, replication, and joining reactions to move different DNA transposons (see Chapter 25 for details). Some elements move by excision and integration (cut and paste), whereas the movement of other elements involves copying of the element by DNA replication (nick-copy out-paste and nick-paste-copy). Transposases from different families can use related mechanisms. Richardson JM, Colloms DS, Finnegan DJ, Walkinshaw MD. 2009. Molecular architecture of the *Mos1* paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* 138:1096–1108; Feng X, Colloms SD. 2007. In vitro transposition of *ISY100*, a bacterial insertion sequence belonging to the *Tc1-mariner* family. *Mol Microbiol* 65:1432–1443; Zhao D, Ferguson A, Jiang N. 2015. Transposition of a Rice Mutator-Like Element in the Yeast *Saccharomyces cerevisiae*. *Plant Cell* 27:132–148; Hennig S, Zeibuhr W. 2010. Characterization of the transposase encoded by *IS256*, the prototype of a major family of bacterial insertion sequence elements. *J. Bacteriol* 192:4153–4163. doi:10.1128/microbiolspec.MDNA3-0062-2014.f6

Cut and Paste Transposons Move by Excision and Integration

Many elements move by a “cut and paste” mechanism in which the transposon is excised from the donor site by double strand breaks (Fig. 8), which can occur by several different pathways (see Chapter 25), all of which expose the 3'-OH transposon ends. Target sites range from nearly random to highly site specific. Note that the gapped donor backbone from which the element excised must also be repaired.

Nick, Copy-Out and Paste: Replicative Transposition of *IS911* and other *IS3* Family Elements

In DNA cut and paste reactions, DNA synthesis is limited to repair reactions at the ends of the newly inserted transposon. The movement of some other DDE transposons, however, involves replication of the entire element. DNA replication is essential to transposition of *IS911* and other members of the widespread *IS3* family (see Chapter 28). In this case, the transposase introduces a nick at only one 3' end of the transposon, that is, donor cleavage is asymmetric (Fig. 9). The released 3'-OH transposon end then joins intramolecularly to just outside its own 5' end, circularizing one strand of the transposon. DNA replication then initiates at the flanking target 3'-OH generated upon intramolecular

transposon strand joining. Replication of the transposon results in a free, double strand transposon circle in which the transposon ends are closely abutted. The transposase then breaks this junction and inserts the transposon by attack of its 3'-OH ends into a target site. Note that the noncircularized transposon strand at the donor site is also copied by DNA replication. Transposition is thus replicative, resulting in a transposon copy remaining at the donor site and a transposon copy at the new insertion site.

Nick, Paste and Copy: Replicative Transposition of *Tn3* Family Transposons

Members of the DDE *Tn3* transposon family, including the closely related transposon $\gamma\delta$ and *Tn4430* (see Chapter 32), are found in many types of bacterial plasmids, transposing from one plasmid to another. Upon transposition from a donor plasmid to a target plasmid, the primary product of their transposition is a cointegrate in which two copies of the transposon link a copy of the donor plasmid and a copy of the target plasmid (Fig. 10). Thus, transposition of these elements involves generating a copy of the transposon, that is, transposition is replicative. The *Tn3* and *Tn4430* transposases can make single-strand nicks at the 3' transposon end, exposing a 3'-OH. Assuming that *Tn3* replicative transposition proceeds as with *Mu* (see below and Chapter

Mu-like transposons

Transposon	Transposase		ATP-dependent targeting	Alternative target selectors		Cut & paste	Nick, paste and copy	Reference/Chapter
	5'	3'						
<i>Mu</i>	?	MuA	MuB			■	■	31
<i>Tn7</i>	TnsA	TnsB	TnsC	TnsD/ <i>attTn7</i>	TnsE/replicating DNA	■		30
<i>Tn522</i>	—	p480	p271	—			■	Rowland & Dyke 1990
<i>Tn5090/ Tn5053</i>	—	TniA	TniB	TniQ/ <i>res</i>			■	Minakhina et al 1999

Figure 7 Some transposons encode DDE Transposases and ATP-utilizing target choice regulators. Bacteriophage *Mu* uses cut and paste transposition to insert into the bacterial genome and replicative transposition to replicate its DNA during lytic growth. MuA is a DDE transposase that breaks and joins DNA and the ATP-dependent regulator MuB controls target DNA selection. Related transposons also encode a transposase, an ATP-dependent target regulator and, in some cases, an additional target type specification protein. Minakhina S, Kholodii G, Mindlin S, Yurieva O, Nikiforov V. 1999. *Tn5053* family transposons are *res* site hunters sensing plasmidal *res* sites occupied by cognate resolvases. *Mol Microbiol* 33:1059–1068; Rowland S-J, Dyke K. 1990. *Tn522*, a novel transposable element from *Staphylococcus aureus*. *Mol Microbiol* 4:961–975. doi:10.1128/microbiolspec.MDNA3-0062-2014.f7

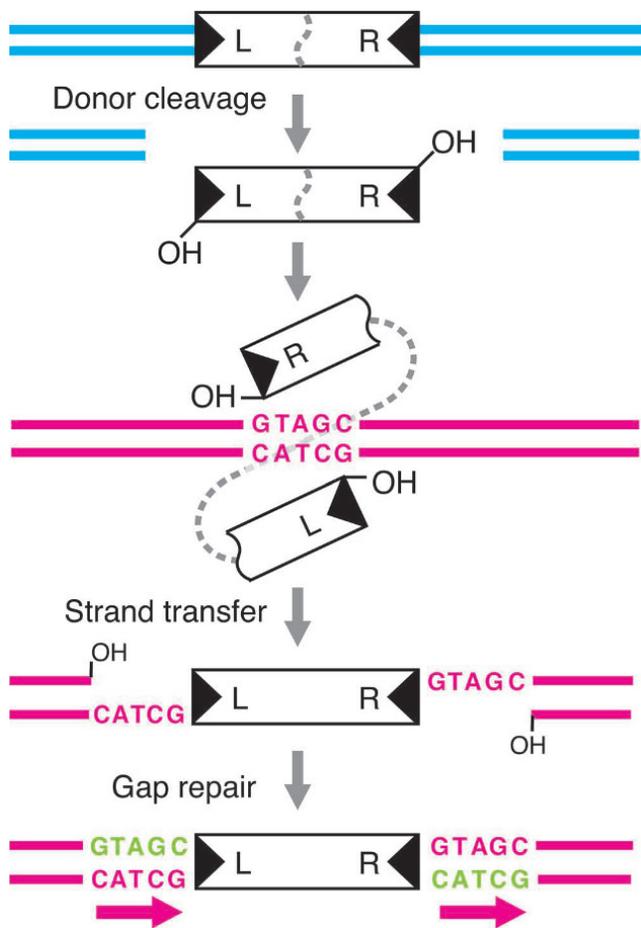


Figure 8 Mechanism of DNA cut and paste transposition by a DDE transposase. The transposase makes DNA double strand breaks at the transposon ends that excise the element from the donor site, exposing the 3'-OH transposon ends. These 3'-OH ends then attack the two target DNA strands at staggered positions by direct transesterification reactions that covalently link the 3' transposon ends to the target DNA. The staggered end joining positions result in single strand gaps at the 5' transposon ends that are repaired by host DNA synthesis (green) to generate flanking target site duplications. doi:10.1128/microbiolspec.MDNA3-0062-2014.f8

31), transposition begins with transposase nicking at both 3' ends of the element. The two released 3'-OH transposon ends then attack a target site on another plasmid, generating a strand transfer product in which the 3' transposon ends are covalently linked to the target plasmid and the 5' transposon ends remain covalently linked to the donor site. Target joining of the 3' transposon ends releases two 3'-OH target ends that flank the newly inserted transposon. DNA replication initiating from these 3'-OHs then copies both strands of the transposon, generating the two transposons that link the donor and target plasmids.

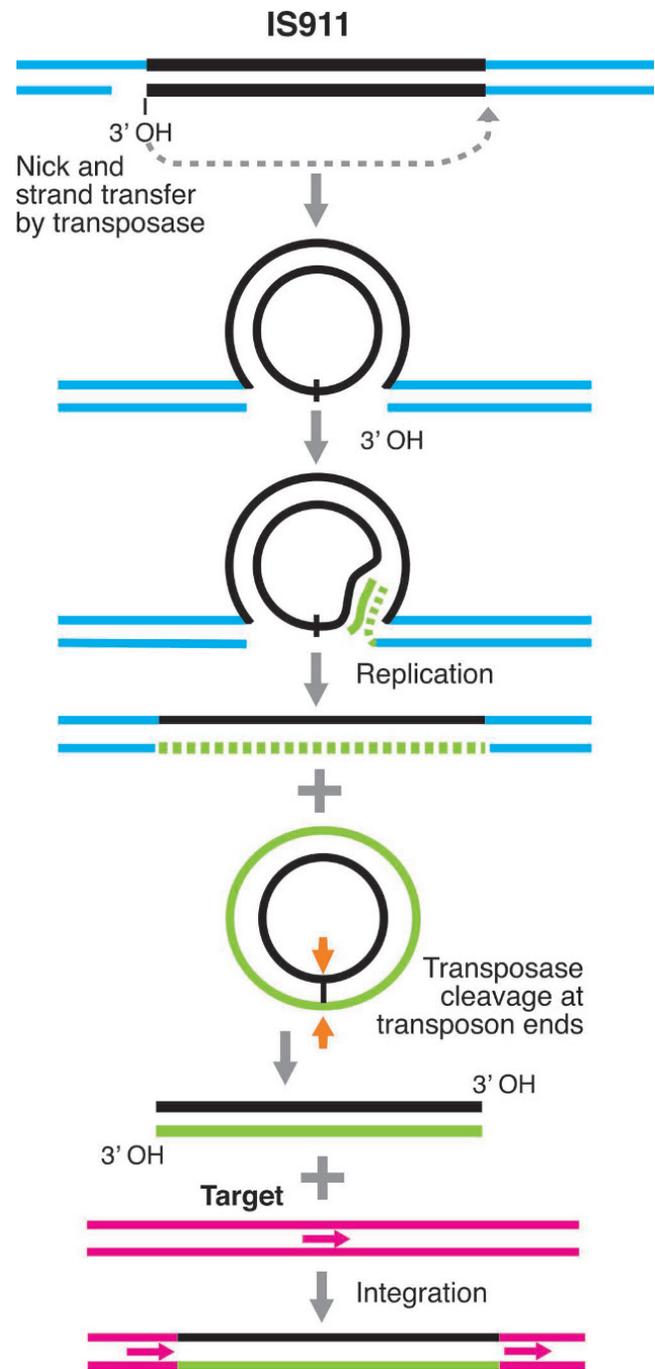


Figure 9 Mechanism of transposition of the nick-copy out and paste transposon *IS911* by a DDE transposase. Transposition begins by transposase nicking at one transposon end. The resulting 3'-OH then attacks its own 5' end, circularizing the transposon. DNA replication (green) initiated at a flanking target 3'-OH copies both transposon strands, releasing a circularized transposon and repairing the donor site. Transposase then cleaves the transposon ends in the transposon circle, releasing 3'-OH ends that attack the target DNA. doi:10.1128/microbiolspec.MDNA3-0062-2014.f9

Generation of the cointegrate is the only step of *Tn3* element-like transposition in which the transposase participates directly. Completion of *Tn3* transposition, however, involves another step that converts the cointegrate into two plasmids, one, the donor plasmid containing a copy of the transposon, and the other, the target plasmid now also containing a copy of the transposon. This monomerization reaction is called resolution, but note that the same reaction is called excision or deletion in other systems.

This cointegrate resolution requires another recombination system, which is also encoded in the transposon. This resolution system consists of a Serine recombinase called a resolvase that acts at the transposon-encoded *res* site to convert the dimeric cointegrate plasmid into separate donor and target plasmids. We consider the mechanism of such resolution reactions by resolvases below.

“Bacteriophage *Mu*: A Transposon” and “Transposon *Mu*: A Bacteriophage”

Bacteriophage *Mu* uses DDE transposition in two different steps of its life cycle (see Chapter 31) (23). *Mu* uses cut and paste transposition to integrate randomly into the bacterial chromosome upon lysogenization. *Mu* then uses replicative transposition that uses multiple chromosomal target sites to replicate its DNA during lytic growth. The *Mu* transposition machinery is elaborate: the ends of *Mu* contain multiple transposase binding sites, as well as internal binding sites that enhance transposition by facilitating correct assembly of the nucleoprotein machine that executes transposition. *Mu* also encodes two transposition proteins, the DDE transposase *MuA* and *MuB*, which facilitates the interaction of *MuA* bound to the transposon ends with target DNA. Perhaps the key event in *Mu* transposition is the formation of a *MuA* tetramer in which two subunits mediate *MuA* breakage and joining and the other two subunits play critical roles in transpososome assembly (Fig. 11). The central features of this machine have now been directly revealed in a crystal structure of a *MuA* tetramer bound to both transposon ends and to target DNA (24).

Single Strand DNA Transposons Move via Tyrosine-HUH Transposases

Hallmarks of DDE transposons are their termini inverted repeats and the absence of covalent transposase-DNA bonds during transposition. By contrast, the substrate of a tyrosine-HUH transposase is a single-stranded version of the element as would be transiently

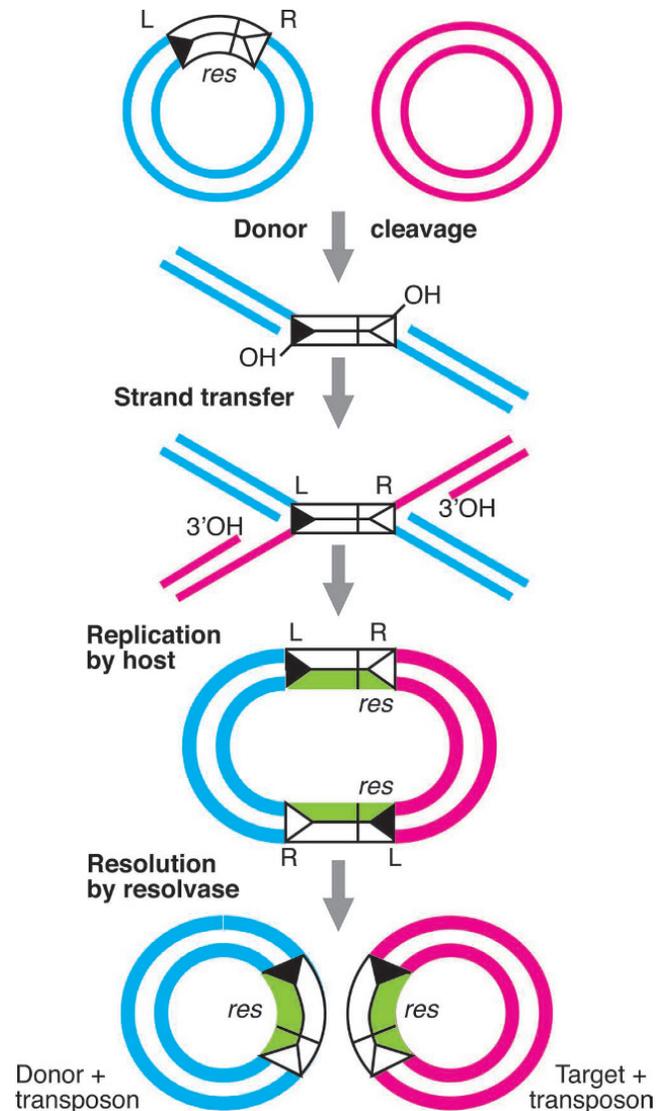


Figure 10 Mechanism of replicative transposition of *Tn3* by a DDE transposase. Transposition begins by transposase nicking to expose both 3'-OH transposon ends that attack and link to the target DNA while the 5' transposon ends remain linked to the donor plasmid. Replication (green) initiated at the flanking target 3'-OHs copies both strands of the transposon, generating a cointegrate in which two transposon copies link the donor and target plasmids. DNA breakage, strand exchange and rejoining at the directly oriented *res* sites by the Ser_{CSSR} recombinase resolvase that is also encoded by the transposon separates the donor and target plasmids, each containing a transposon copy.
doi:10.1128/microbiolspec.MDNA3-0062-2014.f10

present in replication forks. The element lacks terminal inverted repeats but rather contains two internal sets of palindromes that form hairpin structures that are specifically recognized by the transposase. The active sites of tyrosine-HUH transposases contain conserved

tyrosine (Tyr) and HUH motifs. To break DNA, the Tyr-OH acts as a nucleophile on a single strand DNA, releasing a free DNA 3'-OH and forming a 5' phosphotyrosine bond, conserving the high energy of the DNA phosphodiester bond. Attack of a DNA 3'-OH on a 5'-P-Tyr rejoins the DNA and releases the transposase.

Out of and Into Single Strand DNA at Replication Forks

The bacterial IS elements, *IS200*, *IS608*, and *ISDra2*, move via Tyr-HUH transposition (see Chapter 28). The first step in DNA single strand transposition is mediated by a Tyr-HUH transposase dimer and it occurs by two, concerted breakage, exchange, and joining events that excise the element and rejoin the donor site (Fig. 12). Guided by base-pairing interactions between particular transposon sequences, one protomer acts at the upstream end of the element to generate "Donor flank-3'-OH" and "5'-P-Tyr-Transposon end." The other protomer acts at the downstream transposon end to give "Transposon end-3'-OH" and "5'-P-Tyr-Donor flank." The 3'-OH from one protomer then attacks the 5' phosphotyrosine link on the other protomer. Attack of the "Transposon end-3'-OH" on the "5'-P-Tyr-Transposon end" excises the element as a single strand circle and attack of the "Donor flank-3'-OH" on the "5'-P-Tyr-Donor flank" rejoins the single strand donor site.

Transposon integration into a single strand DNA target site again occurs within a transposase dimer. One protomer cuts the transposon circle, giving "Transposon end-3'-OH" and "5'-P-Tyr-Transposon end." The other protomer cuts the target site strand, giving "Target end-3'-OH" and "5'-P-Tyr-Target end." Rejoining, that is, integration, occurs by the attack of "Transposon end-3'-OH" on the "5'-P-Tyr-Target end" and attack of the "Target-3'-OH" on the "5'-P-Tyr-Transposon end."

These breakage and joining reactions occur at DNA single strands in replication forks. DNA replication is also needed, however, to regenerate intact duplex DNA at both the donor and target sites. Thus, when a replication fork passes through a donor duplex from which the transposon has excised, one daughter duplex lacks the element and replication of the other daughter duplex yields a daughter duplex containing the element. Similarly, when a replication passes through a target site into which the element has inserted, replication of the insertion strand gives rise to a daughter duplex containing the element and replication of the other strand yields a daughter duplex without the element. Whereas there was one copy of the element at the donor site before transposition, following replication, one copy

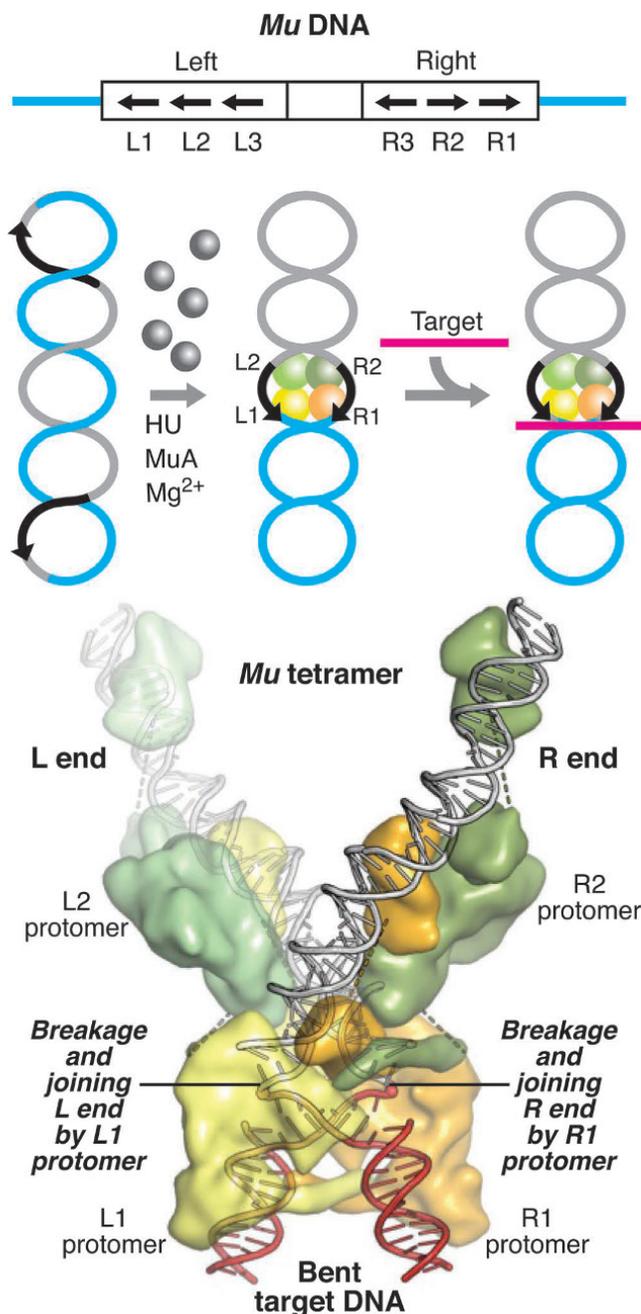
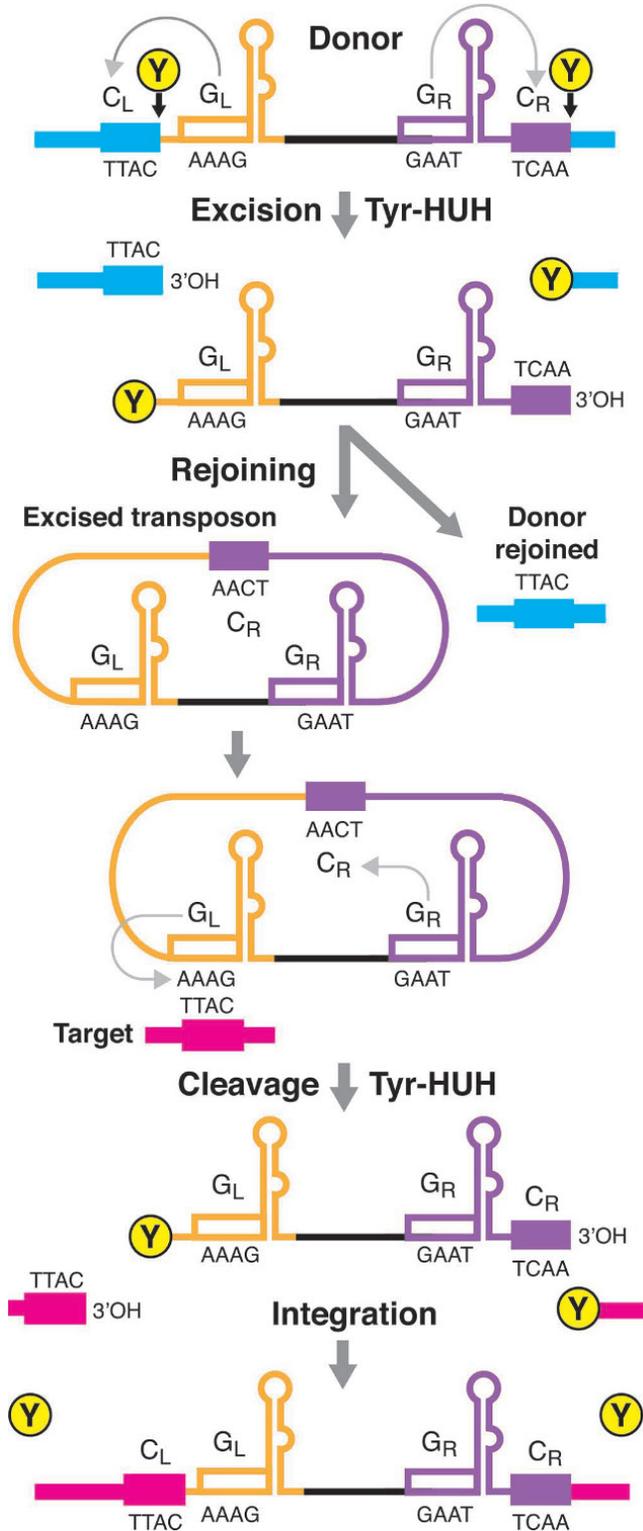


Figure 11 Assembly of an active tetramer of the MuA DDE transposase. The ends of *Mu* contain multiple binding sites for the MuA transposase. MuA interaction with the ends results in formation of a MuA tetramer that synapses the two ends and activates DNA breakage and joining by the two transposase protomers bound to the outermost *L1* and *R1* sites. Internal MuA binding sites in an enhancer (not shown) facilitate tetramer assembly. Usually recruited by MuB (not shown here), a sharply bent target DNA binds to the *L1* and *R1* promoters, which is attacked by the MuA 3'-OH ends exposed by MuA nicking.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f11



remains at the donor site and there is a new copy of the element at the target site.

Some Tyr-HUH Transposases are Rolling Circle Transposases

Tyr-HUH proteins also mediate rolling-circle replication reactions involved in bacterial plasmid replication and conjugation. In these reactions, after recognition of a DNA signal, such as a particular sequence or hairpin, the Tyr-HUH protein nicks one strand of a plasmid circle forming a DNA 3'-OH end, which serves as a primer for DNA replication that displaces the 5'-P-Tyr containing strand. Replication continues around the circle and then the displaced strand can be circularized by the attack of a 3'-OH on the 5'-P-Tyr containing strand. In transposition of bacterial *IS91*-like elements (25), it is thought that the displaced 5'-P-Tyr strand can join to a target DNA via attack of a target 3'-OH generated by the transposase on the incoming 5'-P-Tyr end. DNA replication at the new insertion site and donor site is required to regenerate intact duplex DNA at both sites.

Eukaryotic DNA transposons called *Helitrons* are thought to use a rolling circle mechanism (see Chapter 40). They encode an ORF thought to be the transposase that contains Tyr-HUH motifs and a helicase motif, which could provide a strand displacement activity (Fig. 13). Some elements also encode another nuclease motif. *Helitrons* are found in diverse eukaryotes but have been analyzed most intensively in the mammalian little brown bat and in maize where they make up several percent of the genome. Their presence in bats is

Figure 12 Mechanism of transposition of *IS200*-like transposable element by a tyrosine (Tyr)-histidine-hydrophobic-histidine (HUH) transposase acting on a single strand DNA substrate. A Tyr-HUH transposase breaks DNA by the attack of the hydroxyl of a higher conserved tyrosine, resulting in a free 3'-OH and a 5'-P-Tyr link. DNA rejoining occurs by attack of the 3'-OH on the 5'-P-Tyr link. Guided by base-pairing interactions between the guide and cleavage sequences at both transposon ends, one protomer of a transposase dimer acts at the upstream transposon end, generating "Donor flank-3'-OH" and "5'-P-Tyr-Transposon end." The other protomer acts at the downstream transposon end, generating "Transposon end-3'-OH" and "5'-P-Tyr-Donor flank." The 3'-OHs from one protomer then attack the 5' phosphotyrosine links on the other protomer. This excises the element as a single strand circle and rejoins the single strand donor site. A transposase dimer then integrates the transposon into a single strand DNA target site, guided by the guide and cleavage sequences, by another set of cleavage, strand exchange and rejoining reactions.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f12

particularly notable, as they appear to have been very recently active. Indeed, bats are the only known mammalian source of active endogenous DNA transposons, being host to *Helitrons* (12) and active *piggyBac* elements (26). The impact of *Helitrons* on a genome can be profound as they can acquire and shuffle a wide variety of host sequences when element replication extends into flanking DNA past the end of the element.

A Target Site-Specific Eukaryotic Tyr-HUH Transposon: Adeno-Associated Virus

Adeno-associated virus is a single strand DNA virus. A critical step in its replication requires site specific nicking of a folded region of its termini by a Tyr-HUH endonuclease (see Chapter 37). Although AAV can replicate extrachromosomally, regional specific integration does occur in a region of human chromosome 19.

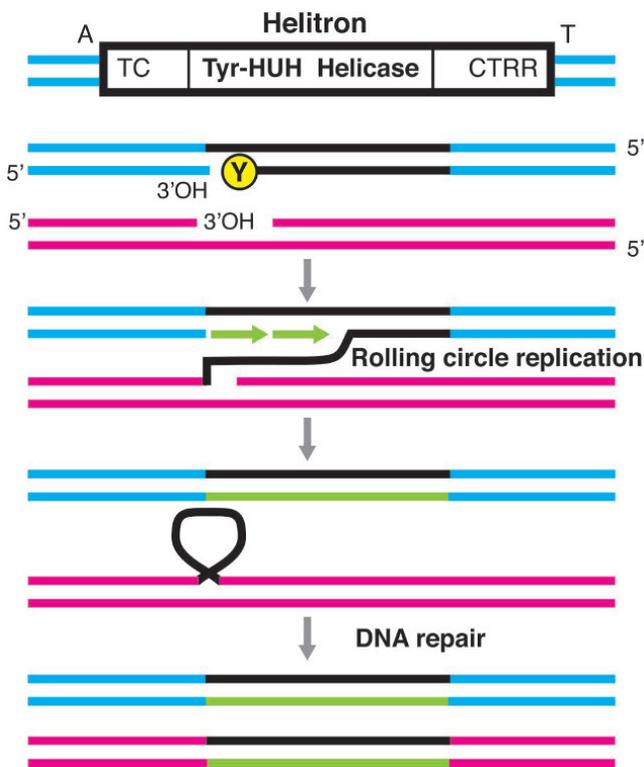


Figure 13 Mechanism of transposition of a *Helitron* by a tyrosine (Tyr)-histidine-hydrophobic-histidine (HUH)/helicase transposase. The Tyr-HUH transposase acts at the upstream end of the transposon, releasing a 3'-OH donor end and a 5'-P-Tyr end. The broken transposon strand is displaced from the donor DNA by rolling circle replication (green) likely assisted by the helicase and it is covalently linked to the target DNA by attack of a target 3'-OH on the 5'-P-Tyr transposon end. DNA synthesis (green) copies the single transposon strand in the target to generate intact duplex DNA.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f13

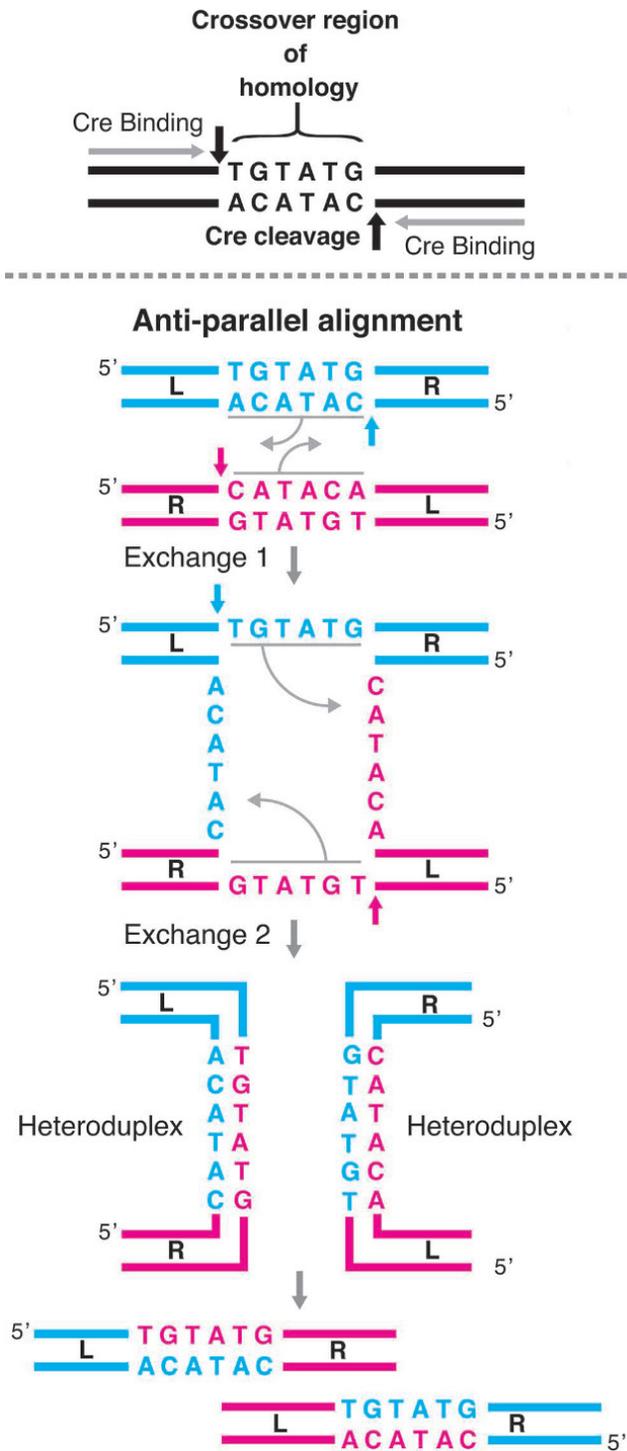
Tyrosine- and Serine-transposases: From Conservative to Libertarian

As with Tyr-HUH transposases, unrelated Tyr- and Ser-transposases also break DNA using a protein nucleophile to release an OH DNA end and a covalent protein-DNA link. Tyr^{TRANSP} use a conserved tyrosine as the nucleophile, yielding DNA 3'-P-Tyrosine and 5'-OH DNA products. Ser^{TRANSP} use a conserved Serine as the nucleophile, yielding DNA 3'-OH and 5'-P-Serine products. DNA rejoining occurs by attack of the DNA-OH on the protein-DNA link. Each transposase class has distinct DNA binding and catalytic domains, but these are not structurally related, thus they represent independent strategies for DNA breakage and joining.

These transposases can promote strand exchange between two recombination sites on two parental DNA duplexes, each site containing two specific transposase binding sites in inverted orientation flanking a short (2 to 8 bp) region of homology. Recombination occurs by DNA cleavage at the outside edges of the regions of homology, also called the crossover region, followed by strand exchange between the duplexes and DNA rejoining (Fig. 14). In these reactions, recombination is reciprocal, that is, no DNA is lost or synthesized, and conservative, that is, no high-energy cofactor is required to rejoin broken DNA because the phosphotyrosine and phosphoserine intermediates preserve the high energy of the phosphodiester bond. This type of recombination is called conservative site-specific recombination (CSSR) and the Tyr- and Ser-transposases that execute such homology-dependent exchange are called Tyr^{CSSR} and Ser^{CSSR} recombinases.

Conservative Site-Specific Recombination

In CSSR recombination sites, the inverted CSSR recombinase binding sites that flank the homology region position the recombinase to cleave at the edges of the homology such that each DNA duplex can bind two recombinase protomers. Recombinase dimers bound to each parental duplex interact with the bound dimers on the other duplex, thereby juxtaposing the substrate DNAs on a recombinase tetramer. The tetramers of Ser^{CSSR} recombinases break and exchange all four DNA strands simultaneously by two pairs of double strand breaks. By contrast, Tyr^{CSSR} recombinases first break, exchange, and rejoin one pair of strands from each duplex and then subsequently break, exchange, and rejoin the other pair of strands. Because the positions of strand exchange lie at the outer edges of the crossover regions, the DNA products are heteroduplex, that is, there is one strand from each parent in the recombination homology regions (Fig. 14). Note also



that the crossover region contains no internal repeats and is thus directional.

Some CSSR sites are far more elaborate than just a pair of inverted CSSR recombinase binding sites flanking a crossover region. There may be multiple additional recombinase binding sites flanking the binding sites at the crossover region that execute strand exchange, as well as binding sites for accessory proteins, which are often architectural DNA bending proteins that facilitate assembly of the elaborate nucleoprotein complexes in which many of these reactions occur. As we will see, assembly of these complexes is a key control point of recombination.

Different products for recombination sites in different orientations

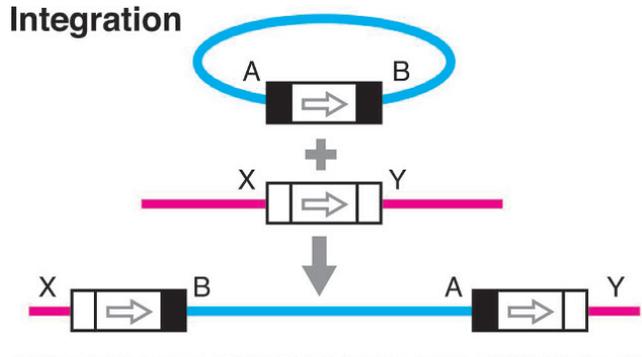
CSSR can mediate several types of DNA rearrangements, depending on the relative orientation of the recombination sites as defined by the direction of the sequence of the crossover region (Fig. 15). When the two recombination sites flank a DNA segment in direct orientation, recombination results in excision of the DNA segment. This reaction can also be called deletion or resolution, depending on the biological context. Recombination between inversely oriented sites results in inversion of the DNA segment between them. Recombination between sites on two different DNAs results in joining of the DNAs, that is, in integration.

Tyr_{CSSR} recombination: DNA nicking, strand swapping and joining

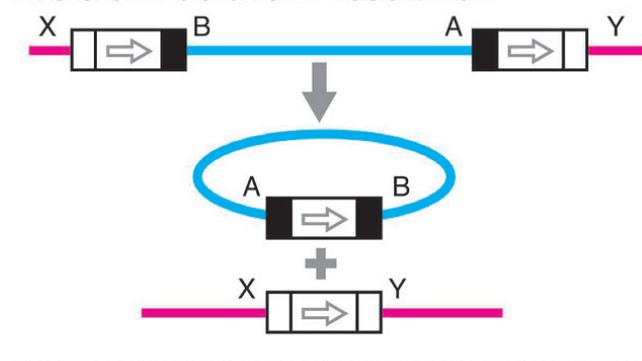
Anti-parallel is the way they go. Although it is often visually convenient to align the 6 to 8 bp cross-

Figure 14 Mechanism of DNA breakage, strand exchange and joining during conservative site-specific recombination by the Tyrosine_{CSSR} recombinase Cre. Cre is a Tyr_{CSSR} recombinase that acts at *lox* recombination sites consisting of two inverted Cre binding sites flanking a conserved central crossover region. Strand breakage, strand exchange and rejoining by Cre occur at the edges of the crossover region. Cre dimers bind to each *lox* site and pair to form the active Cre tetramer that pairs the *lox* sites in antiparallel alignment. Recombination begins by cleavage of the two *lox* sites on strands of the same polarity, making 3'-P-Tyr and 5'-OH ends. Strand exchange and rejoining occurs by the attack of each 5'-OH on the 3'-P-Tyr of the other strand. The second round of strand exchange occurs by Cre cleavage of the other pair of strands, making 3'-P-Tyr and 5'-OH ends. Strand exchange and rejoining occurs by attack of each 5'-OH on the 3'-P-Tyr of the partner *lox* site. Note that because of staggered positions of strand exchange, the recombinant duplexes are heteroduplex in the crossover region, which is bounded by the sites of strand exchange. doi:10.1128/microbiolspec.MDNA3-0062-2014.f14

Integration



Excision / deletion / resolution



Inversion

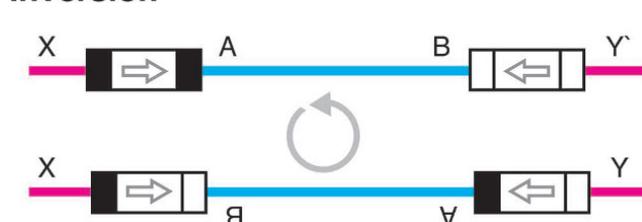


Figure 15 The relative orientation of the substrate recombination sites determines the structure of the recombination products. Sites for CSSR consist of two recombinase binding sites flanking a short crossover region of sequence homology sequence, which lacks repeats and is thus asymmetric. Although the local DNA strand breakage and joining reactions are the same in all cases, the overall structure of the recombination products is determined by the relative orientations of the substrate sites.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f15

over regions of CSSR sites in parallel, strand exchange actually occurs between DNA molecules that are aligned in antiparallel fashion in an active tetramer of the recombinase (Fig. 16). DNA strands are exchanged between duplexes by displacement of short segments of the crossover region such that strand swapping occurs between the duplexes (27). If the sites are nonhomologous, the exchange reactions

abort and the donor duplexes rejoin, leaving the duplexes in their parental configuration.

Going it alone: Cre and Flp can do it all. CSSR reactions promoted by Cre (see Chapter 5) and Flp (see Chapter 1) are the simplest of the CSSR reactions. Each recombination site on each parental duplex contains only the crossover homology region with its two flanking Tyr_{CSSR} recombinase binding sites. These systems can recombine sites in any orientation, making them powerful tools for genome engineering (28). Cre dimers bind to each parental duplex and synapse in antiparallel fashion to form the active tetramer.

Danger, dimers ahead: XerCD to the rescue. A hazard of having a circular genome is that if a circular chromosome dimerizes because of homologous recombination, then when segregation occurs one daughter does not receive a chromosome monomer. Thus, many bacteria encode a CSSR system that promotes chromosome monomerization in a reaction also called resolution. The chromosomally encoded Tyr_{CSSR} recombinases XerC and XerD collaborate to promote resolution at a specific chromosomal site called *dif* (see Chapter 7). This system is distinguished by its tight control of when and where recombination occurs by requiring the host protein FtsK, which acts at the septum, to pump DNA into daughter cells.

My way or no way: control, control, and more control. Some CSSR reactions are very tightly regulated, both in time, that is, only when particular proteins are synthesized, and in directionality, that is, once integration occurs cannot immediately occur without additional excision proteins because the proteins that mediate integration are not sufficient to mediate excision. The integration/excision cycle of bacteriophage *lambda* is elaborately regulated (see Chapter 4). Integration, that is, recombination between *attachment site Phage* (*attP*) on the phage chromosome and *attachment site Bacterial* (*attB*) on the bacterial chromosome, generates the hybrid sites *attL* and *attR* that flank the newly inserted element. Excision, that is, recombination between the *attL* and *attR* sequences, regenerates *attP* and *attB*. Integration and excision require both the phage-encoded Tyr_{CSSR} integrase and host-encoded integration host factor, a sequence-specific DNA bending protein. By contrast, excision also requires the phage-encoded protein excisionase (Xis), which

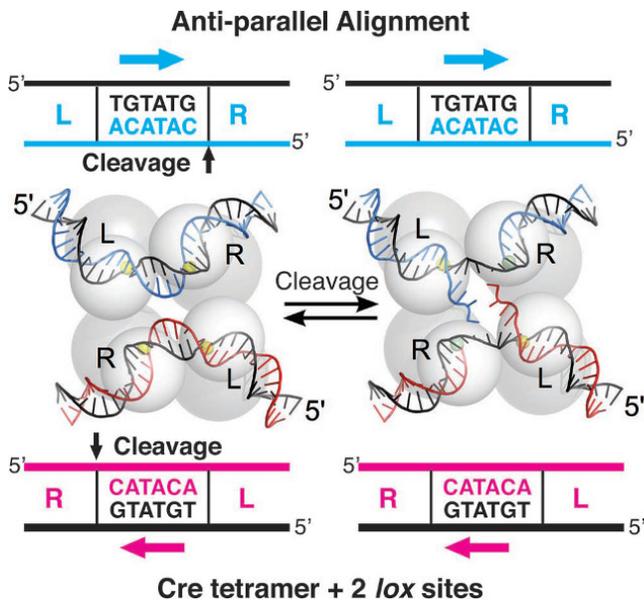


Figure 16 The *lox* sites of Tyrosine_{CSSR} recombinase Cre align in antiparallel orientation in the active tetramer. Dimers of Cre bind to each parental *lox* site and pair to form the active tetramer in which the *lox* sites are aligned in antiparallel fashion. Thus, once cleavage has occurred, strand exchange can occur by local melting and swapping of short, closely juxtaposed DNA segments. Structure graciously provided by Greg Van Dyne.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f16

is expressed only under excision conditions, and the host-encoded FIS protein.

An additional control element is that Int also has two different DNA binding activities (Fig. 17). One binding activity is encoded in the C-terminal region of Int along with the active site catalytic region and it specifically recognizes the sequences of Int binding sites that immediately flank the crossover region. The other Int binding activity is encoded in the Int N-terminal domain and it recognizes binding sites of a different sequence specificity in the arms of *attP* far outside the region of strand exchange (Fig. 17). IHF, Xis, and FIS are sequence-specific DNA binding proteins that bind to the arms of *attP* and promote the DNA bending reactions necessary to form active recombination complexes.

As with Cre and Flp, as well as other Ty_{CSSR} systems, the active form of *lambda* integrase is a tetramer of Int, each protomer of which acts as one of four sites of strand breakage and joining at the crossover regions in *attP* and *attB*. However, the affinity between the Int C-terminal domains is much lower than for Cre and Flp. Thus, the active Int tetramer is only assembled when Int binds to both its crossover and arm sites and when IHF binds to facilitate the DNA bending necessary

for formation of the “integrative intasome.” The active Int tetramer likely assembles on *attP* and then captures the much less complex *attB* site, followed by strand exchange.

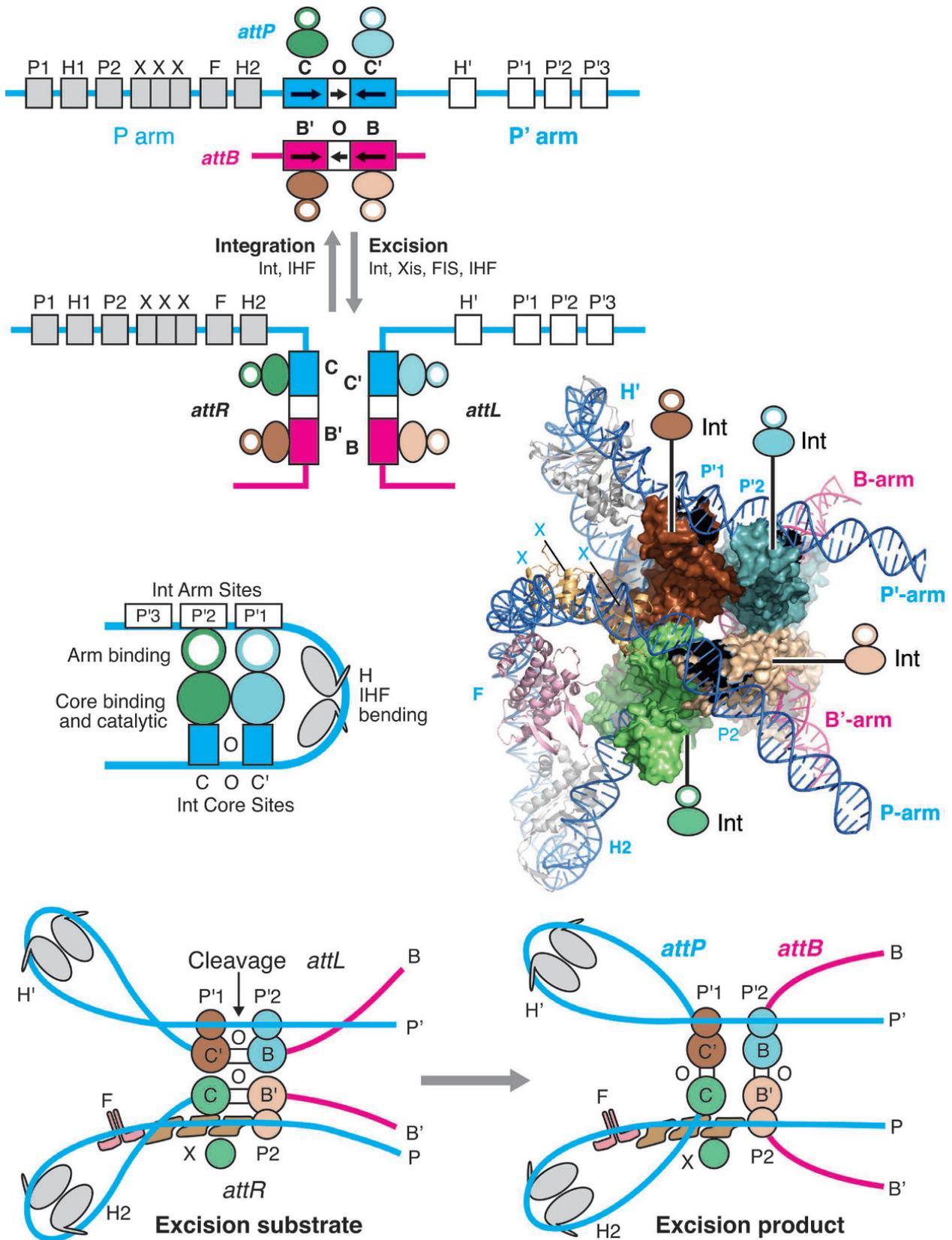
Once integration occurs generating the new hybrid *attL* and *attR* sites, the configurations of the Int and IHF protein-DNA binding sites in the arms of *attL* and *attR* are very different from those required to assemble the “integrative intasome.” Thus, a different “excisive intasome” must be formed to assemble an active Int tetramer at *attL* and *attR* (Fig. 17). Assembly of this “excisive intasome” additionally requires the DNA bending proteins Xis and FIS.

Therefore, although Int and IHF alone can assemble the “integrative intasome” with *attP* and *attB*, Int and IHF alone cannot assemble the “excisive intasome” with *attL* and *attR* so that integration cannot be reversed until Xis and FIS are expressed, which happens only when a particular excision developmental program occurs.

Such directionality control of recombination reactions by changes in the arrangements of flanking accessory sites is a common strategy, especially in bacteriophage integration and excision reactions, which need to be unidirectional. The fundamental difference between these regulated reactions with *lambda* Int and their flanking accessory sites that are necessary to promote formation of the catalytic recombinase tetramer, and the more permissive Cre and Flp recombinases, is that protein-protein interactions between the protomers of Cre and of Flp are sufficiently high that they can bind to the sites of strand exchange and make a reactive tetramer in the absence of accessory proteins or binding sites.

Ser_{CSSR} recombinases exchange broken DNAs by rotation on greasy protein swivels

As with Ty_{CSSR} recombinases, Ser_{CSSR} recombinases act as tetramers that bind specifically to two sites in inverted orientation that flank a region of homology on each parental duplex, exchange DNA strands between the duplex, followed by rejoining (see Chapter 3). However, in contrast to Ty_{CSSR} recombinases, which use two consecutive cycles of single strand exchange between two duplexes, Ser_{CSSR} recombinases make two concerted double strand breaks, one on each parental duplex. Once the DNA strands are broken, rotation of two of the recombinase subunits with their bound DNAs occurs which positions the broken ends of one duplex adjacent to the broken ends of the other duplex, thereby promoting strand exchange. Joining of the juxtaposed ends then occurs. This subunit rotation can occur because there is an interface between



the two sets of dimers that is flat and hydrophobic allowing rotation of the dimers. The flat surface of the interface within the dimers has been visualized by X-ray crystallography.

The Ser_{CSSR} recombinases fall into two categories. Both types contain related catalytic domains with the conserved serine but differ in their DNA binding domains and the complexity of their substrates. The small Ser_{CSSR} recombinases, which include the resolvases and invertases described below, have simple H-T-H DNA binding domains, whereas the large Ser_{CSSR} recombinases, which include phage integrases, have a much more elaborate DNA binding domain.

A small Ser_{CSSR} recombinase mediates inversion. The inversion of a promoter-bearing DNA segment in *Salmonella* directs alternative expression of two surface antigens (see Chapter 9). Inversion is carried out by the Ser_{CSSR} recombinase Hin, which mediates DNA breakage, strand exchange by subunit rotation, and rejoining at *hix* sites that lie in inverted orientation that bound the invertible segment (Fig. 18). Essential to inversion are two host proteins that bind and bend DNA, factor for inversion stimulation (FIS), a sequence-specific DNA binding protein, and HU, a sequence nonspecific DNA bending protein. Also important to inversion is a recombination enhancer sequence usually found between the *hix* sites, which contains multiple binding sites for FIS.

Activation of the DNA breakage and joining reactions that underlie Hin inversion requires the assembly of an Invertasome (Fig. 18). In the Invertasome, FIS bound to the enhancer provides an assembly platform for the two dimers of Hin bound on each *hix* site to form the Hin active tetramer. The conserved active site serines of Hin introduce DNA double strand breaks at each *hix* site and strand exchange occurs by the 180° rotation of one dimer with respect to the other. The broken DNAs rejoin by the attack of the DNA 3'-OH ends on the 5'-P-Ser links.

Notably, FIS-independent Hin mutants have been isolated that can assemble the active Hin tetramer without the aid of FIS and the Enhancer. These mutants, now without directionality control, can recombine *hix* sites in any orientation, as do Cre and Flp with their cognate recombination sites.

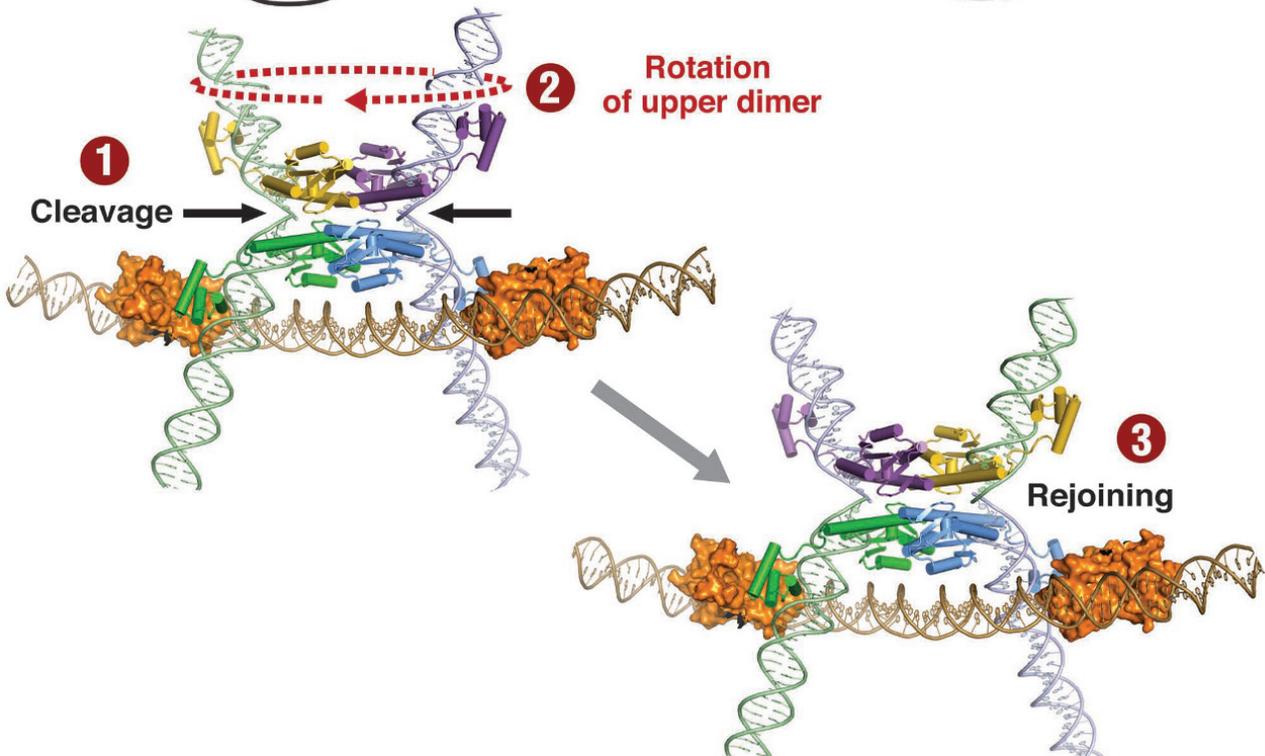
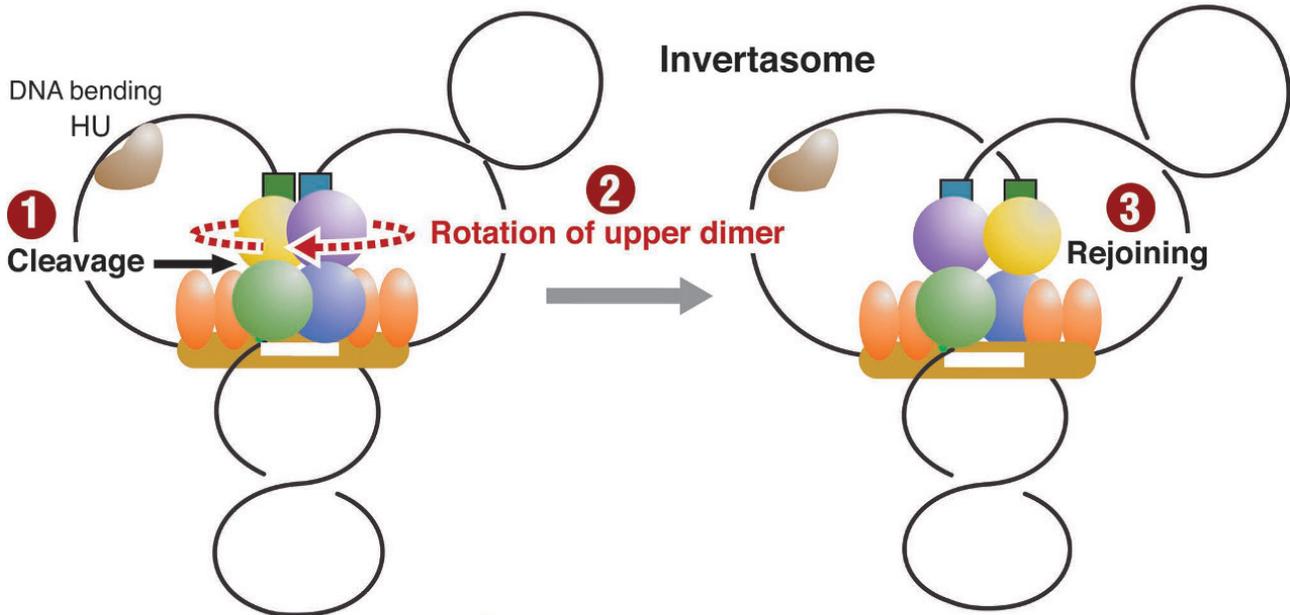
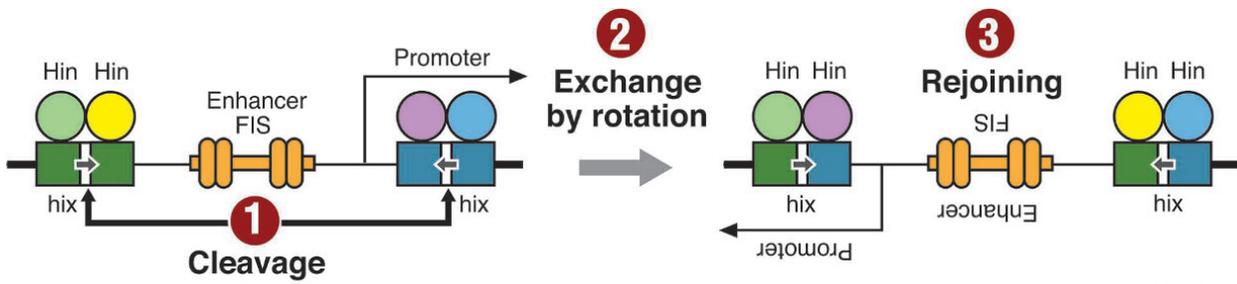
A small Ser_{CSSR} recombinase promotes resolution of plasmid dimers to monomers. As described above, a DDE *Tn3* transposase reaction generates *Tn3* cointegrates, which contain two transposon copies linking the donor and target plasmids. The cointegrate can undergo further recombination to separate donor and target plasmids which now each contain a copy of the transposon (Fig. 10). Ser_{CSSR} recombinases called resolvases promote CSSR between directly repeated copies of *res* recombination sites on a cointegrate plasmid to generate plasmid monomers.

In each parental *res* site, resolvase binding sites in inverted orientation flank the region of homology where strand exchange will occur. Resolvase dimers bind to each *res* site and the dimers pair to form the active tetramer that juxtaposes the parental *res* sites. As with the Ser_{CSSR} invertases, strand exchange occurs by subunit rotation after the *res* sites are broken.

Res sites also contain additional resolvase binding sites that regulate recombination so that resolution, not inversion, occurs (see Chapter 10). These additional resolvase binding sites assemble with the crossover site dimers to form a synaptosome that contains the active tetramer, facilitated by the interwrapping of DNAs at the other resolvase binding sites. Again, nucleoprotein complex assembly activates the resolvase and assures that only resolution occurs because the activating synaptic structure cannot be formed with *res* sites in inverted orientation.

Plasmid multimers can also result from several other reactions, for example, some plasmid multimers arise from homologous recombination between monomers. Instead of using the chromosomal encoded XerCD

Figure 17 The integration/ excision cycle of bacteriophage *lambda* is elaborately regulated. *Lambda* attachment (*att*) sites contain two Int core binding sites, COC' and BOB', in inverted orientation flanking a 7 bp crossover region of homology, "O," where strand exchange occurs. COC' is flanked by *P* and *P'* arms containing multiple protein binding sites: P1, P2, P'1, P'2, and P'3 sites = Int arm sites, which have a different recognition sequence than Int core C and B sites. Schematics of the excisive intasome paired substrate DNAs containing *attL* and *attR* DNAs and the product *attP* and *attB* DNAs following DNA breakage, strand exchange and rejoining at the "O" regions between the core Int binding sites are shown. H, IHF binding sites; X, Xis binding; F, FIS binding.
doi:10.1128/microbiolspec.MDNA3-0062-2014.f17



Tyr_{CSSR} system (see Chapter 7), some plasmids encode their own Ser_{CSSR} resolvase and *res* sites.

Large Ser_{CSSR} recombinases mediate phage integration and excision. The other class of Ser_{CSSR} recombinases is called large serine recombinases (see Chapter 11). The Large Ser_{CSSR} recombinases were discovered in two phage integration and excision systems, ϕ C31 from *Streptomyces* and *Bxb1* from *Mycobacteria*. In both systems the integration reaction, *attP* \times *attB*, can occur with only the large Ser_{CSSR} integrase in the absence of other phage or host proteins and is highly directional. Excisive recombination, *attL* \times *attR*, requires an additional phage-encoded protein, recombination directionality factor (RDF) (Fig. 19). These large Ser_{CSSR} integrases have the same catalytic domain as the small Ser_{CSSR} invertases and resolvases but have much more elaborate DNA binding domains (29, 30). These DNA binding domains, RZ and ZD, are contained in a several hundred amino acid domain at their C-terminal ends. These DNA binding domains bind to their cognate binding sites, *RZ* and *ZD*, which flank the *att* crossover regions in inverted orientation. Notably, the spacing between the RZ and ZD sites is different in *attP* and *attB* such that the conformations of Int bound to these sites, and hence to *attL* and *attR*, are distinct (see below). Each *att* site binds a dimer of integrase.

The integrase C-terminal domain also contains a protein-protein interaction domain, CC, which can interact with the CC domain from another Integrase. Notably, however, although the integrase can bind to all four *att* sites - *attP*, *attB*, *attL*, and *attR* - the conformation of the CC domain is different when bound to the different *att* sites because of spacing differences between multiple DNA binding motifs in *attP* and *attB*. Thus, *attP* CC interacts only with *attB* CC and *attL* CC interacts only with *attR* CC. Therefore, the integrase dimer bound to the *attP* site can pair only with the Integrase dimer bound to the *attB* site (Fig. 19). Similarly, the integrase dimer bound to the *attL* site can pair only with the integrase dimer bound to the *attR* site.

This defined synapsis pathway controls integration and excision. As with small Ser_{CSSR} invertases and resolvases, the active form of integrase is a tetramer assembled from two dimers and breakage and joining occurs only when dimers synapse to form a tetramer. Once the tetramer is assembled, double strand breaks occur at each *att* site, strands are exchanged between the parental duplexes by rotation of integrase dimers as with the resolvases and invertases, and then the exchanged ends are rejoined.

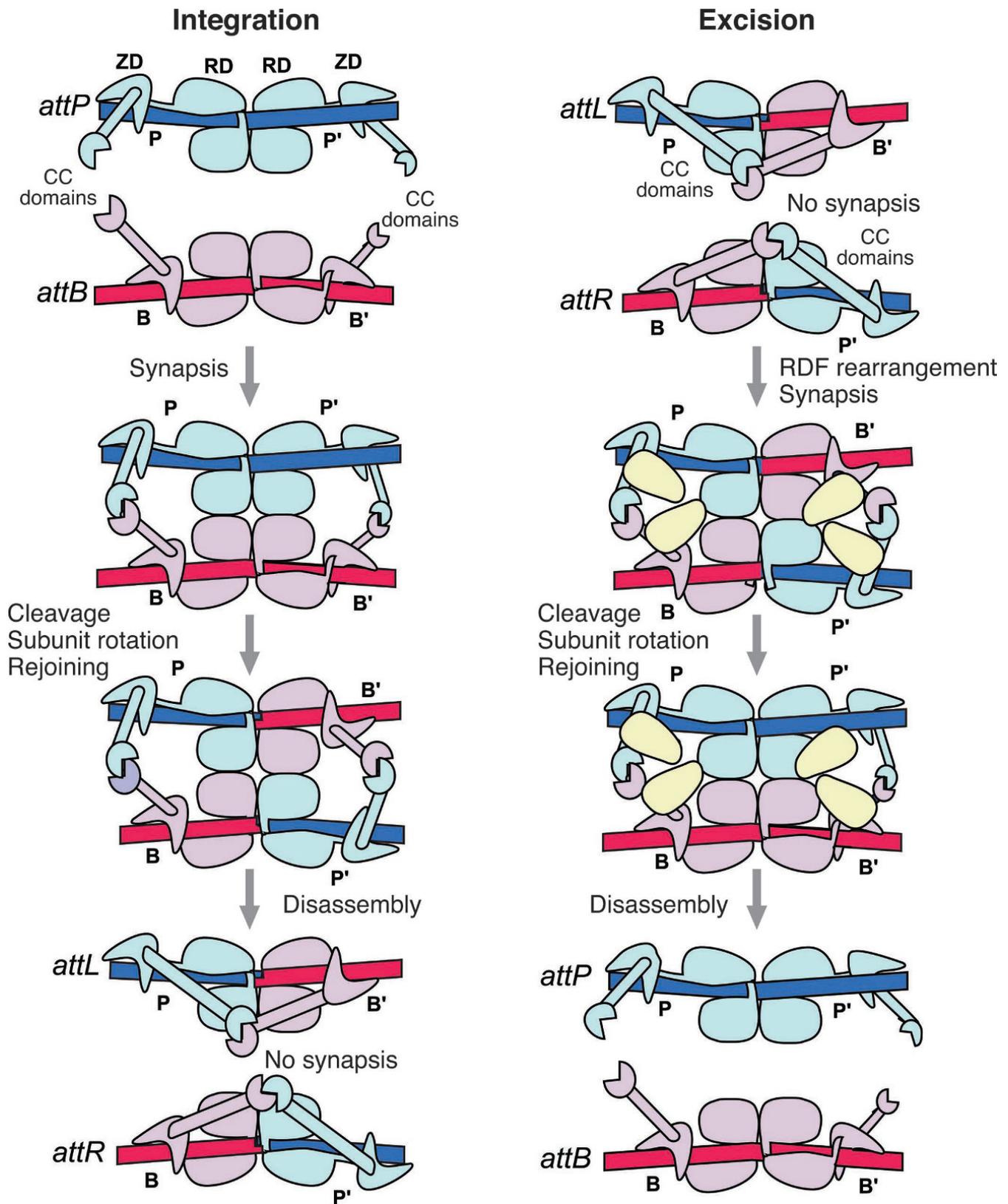
In the presence of the phage-encoded RDF that is expressed only as part of the developmental program for excision, synapsis between integrase dimers bound to *attL* and integrase dimers bound to *attR* can occur, leading to *attL* \times *attR* recombination, that is, excision. The mechanism by which RDF acts remains to be determined in molecular detail but an attractive view is that RDF interacts with the Integrase to change the conformation of the CC domain to allow Int-Int interactions when bound to *attL* and *attR*. This RDF strategy of changing integrase conformation is distinct from that in Tyr_{CSSR} systems such as bacteriophage *lambda* in which the proteins uniquely required for excision, Xis and FIS, are both DNA binding and bending proteins.

Tyr- and Ser-Transposases Mediate Transposition: DNA Breaking, Strand Exchange, and Joining in the Absence of Homology at the Region of Strand Exchange

The ICE-element cometh (and goeth)

Integrative conjugative elements (ICEs) are bacterial transposable elements that can excise from the chromosome in a donor cell, transfer between cells by conjugation, and then integrate into the chromosome in the recipient cell (see Chapters 8 and 13). ICEs encode conjugation functions but do not encode their own replication functions. In addition to encoding antibiotic resistance genes and other accessory determinants, ICEs also encode proteins that mediate their integration and excision. Different ICEs vary in their target site selectivity, which varies from integration into only a few sites to integration into many sites. ICE systems use

Figure 18 Inversion of a promoter-containing DNA segment by the Ser_{CSSR} recombinase Hin within an invertasome controls gene expression. Inversion of a DNA segment containing a promoter changes which surface antigen gene is expressed. *bix* recombination sites bound the invertible segment in inverted orientation. Hin dimers bind to each *bix* site and FIS binds to the enhancer segment. The Hin dimers interact with FIS on the Enhancer platform and with each other to form the active Hin tetramer. Cleavage of both *bix* sites occurs in the central region of homology and strands are exchanged by rotation of the upper dimer of Hin, followed by DNA rejoining. Drawing and structures adapted from material from Reid Johnson. doi:10.1128/microbiolspec.MDNA3-0062-2014.f18



either $\text{Tyr}_{\text{TRANSF}}$ or $\text{Ser}_{\text{TRANSF}}$ -transposases that use amino acid nucleophiles to make reversible Tyr-DNA and Ser-DNA covalent intermediates but have much looser requirements for homology at the crossover sites between the recombining sites than do Tyr_{CSSR} and Ser_{CSSR} recombinases.

Two well studied ICEs, *CTnDOT* (see Chapter 8) and *Tn916* (31), which are found in a wide range of bacteria, encode $\text{Tyr}_{\text{TRANSF}}$ integrases that mediate element integration and excision. In *CTnDOT*, multiple $\text{Tyr}_{\text{TRANSF}}$ integrase binding sites flank the crossover sites where strand exchange will occur. The bound proteins are thought to form an active tetramer assembled from two sets of dimers bound to the parental substrate DNAs. However, in contrast to Tyr_{CSSR} , recombination with $\text{Tyr}_{\text{TRANSF}}$ can proceed with nonhomologies in the crossover region (Fig. 20; see Chapter 8, Fig. 2; and Chapter 25, Fig. 3). As occurs in integration/excision systems mediated by Tyr_{CSSR} and Ser_{CSSR} recombinases, directionality is tightly controlled as excision also requires several *CTnDOT*-encoded excision proteins and a host-encoded factor.

Plug and Play: integrons capture and express multiple gene cassettes

Integrons are gene expression platforms that may be present on mobile elements such as transposons or in bacterial chromosomes (see Chapter 6). They encode a promoter upstream of an *attI* recombination site that uses an element-encoded $\text{Tyr}_{\text{TRANSF}}$ integrase to capture gene cassettes, which carry a recombination site *attC*, by recombination between *attI* and *attC* (Fig. 21). Excision of the cassettes allows their capture by other integrons. These gene cassettes can encode a wide variety

of determinants ranging from antibiotic resistance to metabolic functions. After recombination, the gene cassettes lie downstream of the *attI* promoter and they are expressed.

Recombination between *attC* \times *attI* does not occur by CSSR because there is no region of homology between these sites.

Phage exploitation of XerCD and *dif* sites

As described above, in the highly conserved Tyr_{CSSR} XerCD system (see Chapter 7), the XerCD recombinases act on chromosomal *dif* sites to convert hazardous chromosomal dimers generated by homologous recombination to monomers to facilitate chromosomal segregation. Many phages that do not encode their own recombinase hijack the XerCD system to promote their integration into bacterial chromosomes (32). For example, the integration of *CTX ψ* , a filamentous phage, which encodes the diphtheria toxin, into the *Vibrio cholerae* genome (33), converts nonpathogenic *V. cholerae* into a pathogen. Despite the fact that XerCD mediates CSSR between *dif* sites that have a region of homology, *CTX ψ* integration does not occur by CSSR. As in *attC* containing cassettes in integrons (see Chapter 6), the *CTX ψ* genome is single strand DNA that is folded by several palindromes into a double strand form to be able to recombine with *dif*.

$\text{Tyr}_{\text{TRANSF}}$ in eukaryotes: Cryptons and Tecs

Eukaryotic DNA elements, called *Cryptons* that contain $\text{Tyr}_{\text{TRANSF}}$ integrases have been identified in fungi (see Chapter 53) (34, 35). *Tec* elements containing $\text{Tyr}_{\text{TRANSF}}$ that undergo excision during macronuclear development in the ciliate *Euplotes* have also

Figure 19 A large Ser_{CSSR} integrase can mediate highly regulated cycles of bacteriophage integration and excision. The integration and excision cycle of phage ϕC31 is mediated by a large Serine_{CSSR} integrase. Integration between *attP* and *attB*, which generates the hybrid sites *attL* and *attR*, requires integrase. Excision between *attL* and *attR*, which generates *attP* and *attB*, requires integrase + recombination directionality factor (RDF). Integrase binds specifically to *attP* and *attB* using its RZ and ZD domains that recognize RZ and ZD DNA sequences, which are present on all *att* sites. Note the difference in RZ and ZD spacing in *attP* vs *attB* such that integrase binds to each in a slightly different conformation. CC domains interact with each other. A dimer of Integrase can bind to both *attP* and *attB*. Pairing between the Int dimers forms the active tetramer, which synapses *attP* and *attB*. However, an *attP* dimer cannot pair with another *attP* nor can *attB* pair with *attB* because of their different Int configurations. Once synapsis occurs, the integrase cleaves the crossover region of homology; the broken DNA ends exchange by rotation of one pair of Integrase subunits, followed by rejoining to generate the hybrid *attL* and *attR* sites. Note that the CC domains of the dimer integrase bound to *attL* and to *attR* interact intramolecularly and are thus unavailable for interdimer pairing. RDF is proposed to interact with integrase to change its conformation when bound to *attL* and *attR*, such that CC domains can interact intermolecularly and pair the dimers on *attL* and *attR* to form the active tetramer. Excision occurs between *attL* and *attR* by cleavage of *attL* and *attR* and subunit rotation, followed by rejoining to generate *attP* and *attB*. doi:10.1128/microbiolspec.MDNA3-0062-2014.f19

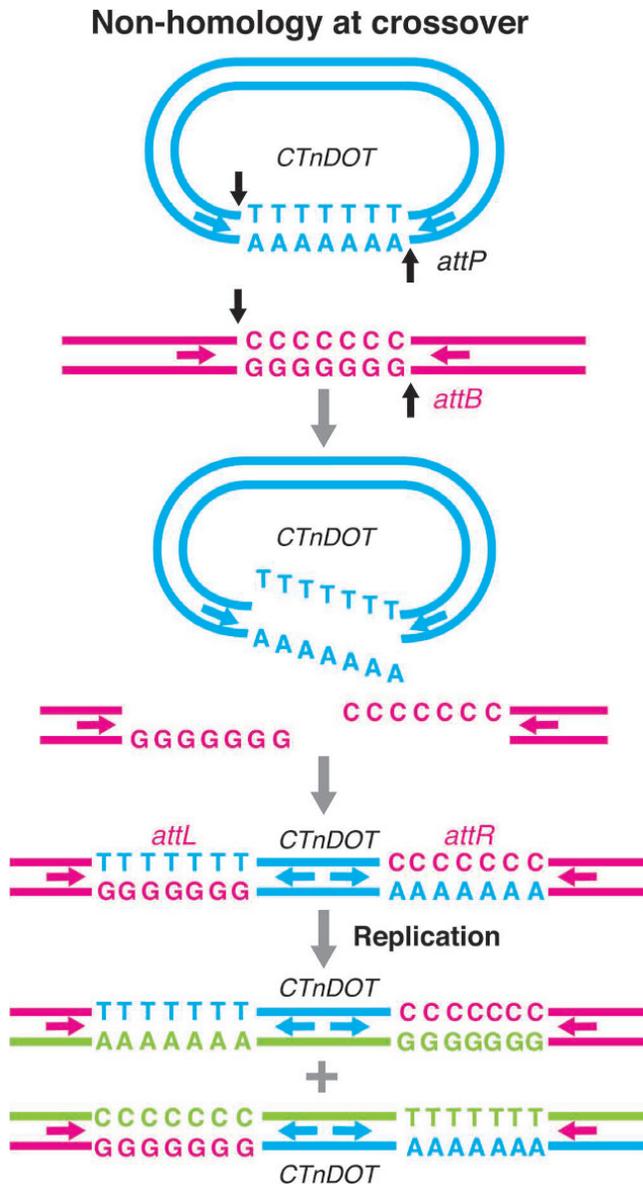


Figure 20 A $\text{Tyr}_{\text{TRANSF}}$ integrase mediates integration of the ICE *CTnDOT* in the absence of crossover homology between the recombination sites. A hallmark of CSSR crossover regions, which are flanked by inverted recombinase binding sites that promote breakage, exchange and joining at the outer edges of the crossover region, is that the crossover regions are identical. Some Ser and Tyr recombinases can, however, promote recombination between nonhomologous crossover regions. The crossover regions of *attP* and *attB* of *CTnDOT* are nonhomologous such that when strand exchange occurs, the heteroduplex regions contain base pair mismatches as shown in this extreme example. Replication of the recombination product yields two daughter chromosomes with different sequences in *attL* and *attR*. Recombinases that do not require absolute homology can be considered transposases. doi:10.1128/microbiolspec.MDNA3-0062-2014.f20

been identified bioinformatically (36,37). $\text{Tyr}_{\text{TRANSF}}$ have also been found in eukaryotic retroelements (see below; see Chapter 53).

RETROTRANSPOSONS: DNA→RNA→DNA

In contrast to the mobile elements described above that have DNA substrates, intermediates, and products, mobile elements called retrotransposons move via an RNA intermediate. The movement of a retrotransposon from a donor site to a new insertion site begins with the synthesis of an RNA copy of the element by the host pol II polymerase. This RNA copy is then converted into DNA by an element-encoded RNA-dependent DNA polymerase, reverse transcriptase (see Chapter 46). Notably, in contrast to DNA elements that excise from their donor sites, the retrotransposon donor site is unchanged, and thus, can generate more RNA copies, which are converted into DNA. This replicative capacity likely contributes to the high copy number of retrotransposons in some organisms.

Different types of retrotransposons use different mechanisms to convert the RNA copy of the element into a DNA copy at a new insertion site. The DNA form of retroviruses and retroviral-like retrotransposons is generated by reverse transcription in the cytoplasm far from the nuclear DNA where it is integrated by a DDE transposase, in this case called a retroviral integrase. The DNA cleavage reactions, which for some retrotransposon elements include several nucleotides from their 3' ends to expose their reactive 3'-OH termini, and the staggered attacks of these 3'-OH ends on both strands of the target DNA, occur by the same mechanism as with DDE DNA-only transposons (see below).

By contrast, other retrotransposons, including non-LTR retrotransposons and mobile group II introns, use a quite different strategy in which their RNA copies interact directly with their new DNA insertion site prior to reverse transcription. With these elements, a target DNA 3'-OH is used as the primer for reverse transcription of the template element RNA that generates the DNA form of the element *in situ* in a reaction called target primed reverse transcription (TPRT).

Retroviruses and Retroviral-like Retrotransposons

When integrated into the genome, the central gene-encoding region of retroviruses (see Chapter 48) and retroviral-like retrotransposons is flanked by directly repeated long (hundreds of bp) terminal repeats (LTRs), and thus, these elements are called LTR retrotransposons

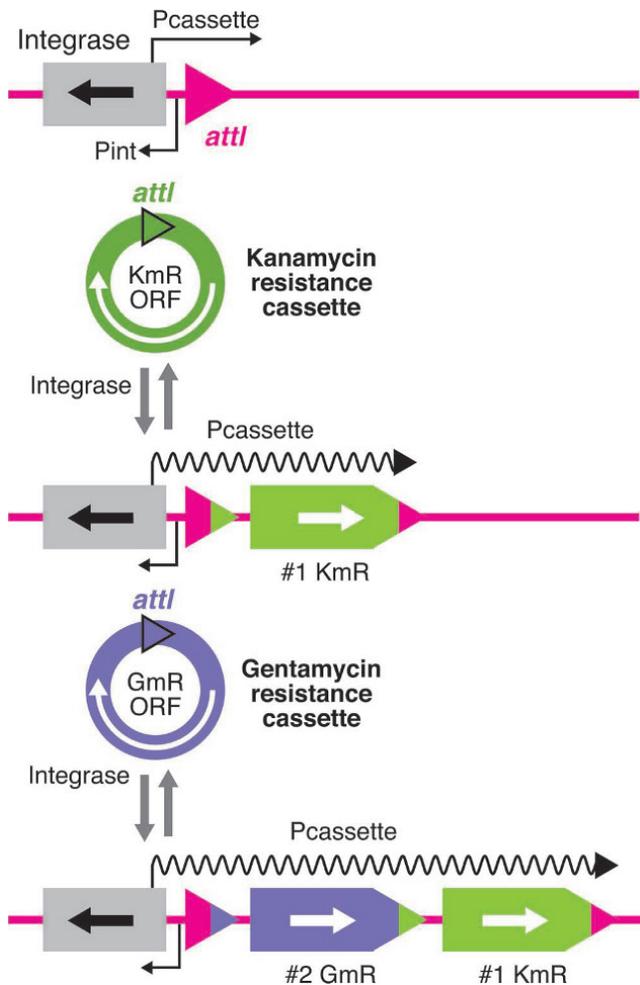


Figure 21 A Ty_T-TRANSP integrase mediates capture and expression of multiple different gene cassettes. Integrons are gene-expression platforms containing an attachment site (*attI*) and cognate integrase that are found on other mobile DNAs such as transposons and in bacterial chromosomes. The integron can capture, integrate, and express gene cassettes that contain related but often not identical *att* sites, *attC*, by site-specific recombination. Once integration occurs, the *P_{cassette}* promoter drives expression of the promoter-less cassettes. Some chromosomal integrons contain hundreds of gene cassettes. doi:10.1128/microbiolspec.MDNA3-0062-2014.f21

(Fig. 22). LTR retrotransposons are widespread in fungi, plants, and animals but retroviruses are found mostly in vertebrates. Although HIV and HTLV are the only known active human retroviruses, many copies of autonomous and nonautonomous retroviral elements are present in the human genome, reflecting past retroviral infections (Fig. 23). These endogenous retroviruses are called ERVs and are thought to be inactive for retrotransposition but provide important regulatory elements to the cell (see Chapter 47).

Transcription of the integrated virus, also called the provirus, by the host polymerase begins within the upstream LTR and terminates in the downstream LTR. The central region of all LTR elements encodes the polyprotein Gag, which includes the RNA binding proteins capsid and nucleocapsid, and Pol (Fig. 22). The Pol domain is typically expressed at lower levels than Gag as a Gag-Pol fusion precursor polyprotein. It can include Protease and does contain reverse transcriptase-RNase H (see Chapter 46) and integrase domains (see Chapters 44 and 45). Integrases of retrovirus and LTR retrotransposons are closely related to the DDE class of DNA transposases.

Retroviruses, which are distinct from retrotransposons, bud from host cells and infect new cells. This process is enabled by an Env, a membrane protein that mediates cellular exit and entry. Following expression of Gag and Gag-Pol, these proteins condense together with some host proteins around two copies of genomic retroviral RNA to form virus-like particles (VLPs). In the case of retroviruses, VLPs can form intracellularly or at the plasma membrane, but in any case, bud from the cell and in the process acquire a membrane containing Env. As budding occurs activation of protease results in processing of the precursor polyproteins into their mature forms. Maturation occurs intracellularly in the case of retrotransposons.

In multiple steps involving both copies of the viral RNA, the element-encoded reverse transcriptase makes a double-strand DNA copy, sometimes called a cDNA, containing the LTRs in the cytoplasm in a large complex containing viral and host proteins. In the case of fungi where LTR retrotransposition has been most extensively studied, the nuclear envelope does not break down and VLPs are likely to be significantly remodeled as the DNA is translocated together with integrase and other proteins into the nucleus.

Integrase then inserts the viral DNA into its new target site. Using the same steps as DNA cut and paste DNA-only transposons, integrase mediates attack by the viral DNA 3'-OH ends on staggered positions on the top and bottom strands of the target DNA (Fig. 24). Therefore, integrated retroviral elements are flanked by target site duplications and LTR elements from yeast to humans follow this same fundamental pathway. Retroviruses can also encode auxiliary determinants, for example, several proteins encoded by HIV limit the efficacy of host restriction factors.

The retroviral family is large and diverse (see Chapter 48) and the study of many different elements has contributed to our current understanding. Not surprisingly, however, much recent work has focused on

Retrovirus	Family
Avian Sarcoma Virus (contains Src) Rous Sarcoma Virus (contains Src)	Alpha
MLV Murine Leukemia Virus MoMLV Moloney Murine Lukemia Virus	Gamma
HIV-1 Human Immunodeficiency Virus	Lenti
PFV Prototype Foamy Virus	Spuma

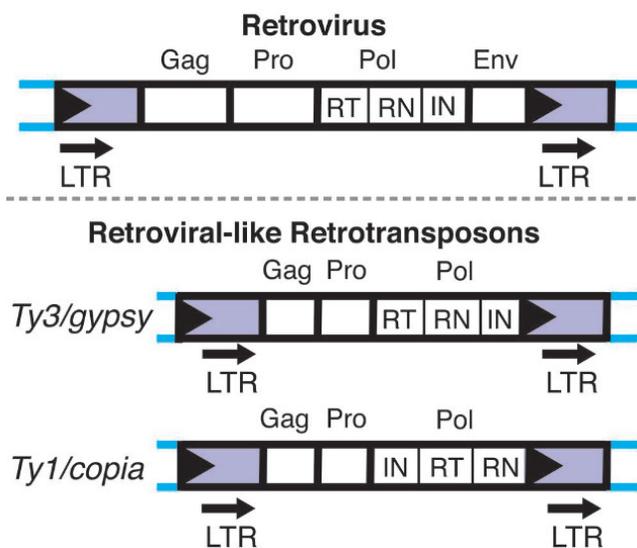


Figure 22 Structures of some well-studied retroviruses and retroviral-like retrotransposons. Retroviruses and the closely related retroviral-like retrotransposons contain related central protein-coding regions flanked by direct long terminal repeats. The central regions of both retroviruses and retroviral-like transposons encode Gag, which includes several nucleic acid domains, a protease that cleaves polyprotein precursors, and Pol, which encodes reverse transcriptase, RNase-H and integrase. Retroviruses also encode Env, a membrane protein that facilitates viral particle exit from host cells and entry into new host cells. In different families of retroviruses, different combinations of protein domains are expressed as fusion polyproteins. Some retroviruses also encode other genes, for example, the avian transforming viruses, ASV and RSV, encode the oncogene Src. Reverse transcriptase and RNase-H convert the two viral RNA copies in the viral particle into the DNA form of the virus, which terminates in direct long terminal repeats. Integrase, which is a DDE transposase, integrates the viral DNA into the host genome. In retroviral-like elements of the yeast *Ty3* and *Drosophila gypsy* family, the order of Gag, Pro and RT, RN and in Pol are the same order as in retroviruses, whereas in the yeast *Ty1* and *Drosophila copia* family, Integrase proceeds RT-RNase H.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f22

HIV-1, contributing greatly to human AIDS treatment including development of drugs that inhibit integration, replication, maturation, and cell fusion.

Much recent work on LTR elements has focused on dissecting the roles of host factors and on understanding target site selection, which has a major impact on the effect of insertions on host cells. Early-on genome-wide screens for budding yeast *Ty1* of the *Ty1/Copia* class (see Chapter 41), *Ty3* of the *Ty3/Gypsy* class (see Chapter 42), and fission yeast *Tf1* of the *Ty3/Gypsy* class (see Chapter 43) elements identified multiple host cofactors and restriction factors for retrotransposition. More recently, siRNA and dominant-negative screens have identified numerous host factors for retroviruses, particularly for HIV-1 (see Chapter 45). Perhaps not surprisingly, host factors for both LTR retrotransposons and retroviruses include RNA helicases, translation factors, replication factors, and nuclear porins.

Both retrotransposons and retroviruses have distinct targeting biases that influence where they insert into the host genome. These preferences are now explained, at least in part, by interactions with chromatin proteins including host transcription factors and proteins that mediate interactions with histone modifications.

HIV-1 inserts preferentially into transcribed regions, perhaps to facilitate transcription of the virus. An intriguing possibility is that one factor in this preference is a preferential association between HIV-1 and components of the nuclear pore complex that are, in turn, preferentially associated with active chromatin. Integrase interacts directly with the host factor LEDGF, which also interacts with histones associated with active genes (see Chapters 44 and 45).

The *S. cerevisiae Ty1*, *Ty3*, and *Ty5* elements and *S. pombe Tf1* element (see Chapter 43) display strong target site specificity (Fig. 25). *Ty1* (see Chapter 41) and *Ty3* (see Chapter 42) elements target tRNA genes. *Ty3* is the most selective, inserting site-specifically at *pol III* initiation sites by direct interaction with a *pol III* transcription factor. *Ty1* inserts preferentially within about 1 kb upstream of *pol III* transcription initiation sites, being guided by direct interaction with polymerase with a bias toward specific nucleosome surfaces. This preferential targeting to small genes discourages potentially harmful insertions at other sites. *Ty5* inserts preferentially into heterochromatic regions such as telomers and the silent *HML α* and *HMRa* storage cassettes. *Tf1* insertions are concentrated in *pol II* promoters and correlate with Sap1 binding sites with preference for stress response genes. This preference could be important for increased genetic diversity in response to stress.

Transposable Elements in the Human Genome

		Structure	Size	Copy number	% of genome	Copies active for transposition
Retrotransposons						
Active						
Autonomous		5'UTR ORF1 ORF2 3'UTR Poly A				
LINE	<i>L1</i>		~6 kb	~516,000	17.6%	50-100
Non-autonomous		L-7SL A rich R-7SL Poly A				
SINE	<i>Alu</i>		~300 bp	~1,100,000	11%	~900
		Alu-like SINE-like Poly A				
SVA		CCCTCT VNTR Poly A	~2 kb	~2,700	0.2%	20-50
Inactive						
Processed pseudogene/ Retrogene		Cellular mRNA Poly A	Variable	~12,000		None
ERV		LTR Gag Pro Pol Env LTR	6-11 kb	~54,000	3.0%	None
Solo LTR		LTR	0.5-3 kb	~490,000	5.3%	None
DNA transposons						
Inactive						
Autonomous		ORF Transposase	2-3 kb	~300,000	3.4%	None
<i>hAT, ITm</i>						
Non-autonomous			0.1-3 kb			

Figure 23 A large fraction of the human genome is comprised of transposable elements. The structures of examples of the major classes of transposable elements found in the human genome are shown. Their total copy numbers and the number of estimated active elements are also shown. The active elements, LINE element *L1* and the SINES *Alu* and *SVA*, whose movement depends on *L1* proteins, contribute to human genetic variation. Although processed pseudogenes do not retrotranspose, formation of new pseudogenes has occurred in humans. Note the amount of human genome derived from mobile elements, in some cases significantly (DNA elements=3.4%), or in other cases very dramatically (*L1*=17.6% and *Alu* =11%), exceeds the fraction of the human genome that encodes ORFs, about 1.5%.
doi:10.1128/microbiolspec.MDNA3-0062-2014.f23

Important insights into the mechanism of retrotransposition at the molecular level have come from structural characterization of individual steps of reverse transcription and with the first crystal structure of a retrotransposon reverse transcriptase, as well as RNase H complexed with the oligonucleotide substrates (see Chapter 46) and with achievement of a crystal structure of the prototypic foamy virus integrase tetramer com-

plexed with virus and host target DNA (Chapter 44). The latter structure has allowed modeling of the HIV-1 integrase and a deeper understanding of the mode of action of integrase inhibitors.

Retroviral integrases that are DDE transposases are not the only type of integrase used for the integration of a mobile element with a genome generated by reverse transcriptase. *DIRS*, *Ngaro* and *Viper* are retro-

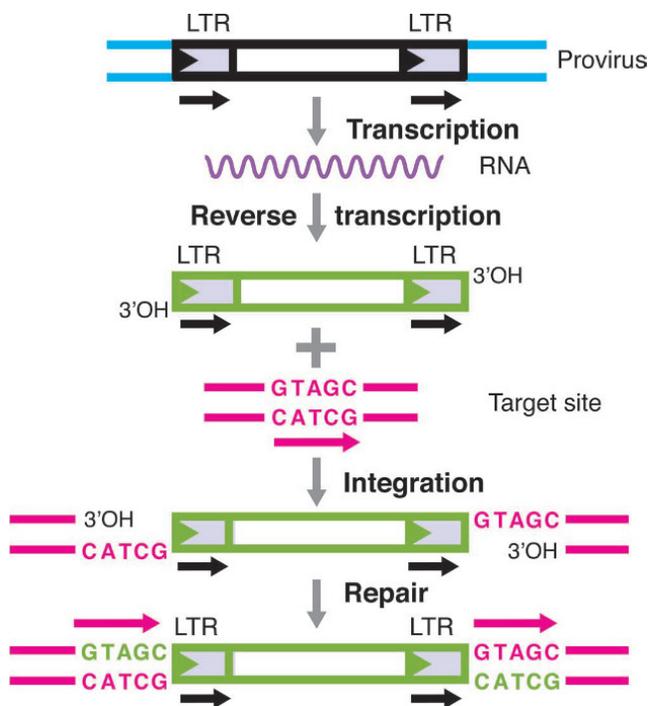


Figure 24 The mechanism of integration of the DNA forms of retroviruses and retroviral-like retrotransposons. The first step in retroviral replication and integration is transcription of the provirus by host RNA polymerase. Two RNA copies are packaged into each virus particle along with the proteins Reverse transcriptase-RNaseH and IN. Reverse transcriptase uses the two viral RNA copies to synthesize a cDNA (green) extending from the 5' end of one LTR to the 3' of the other with exposed 3-OHs. The retroviral integrase, a DDE transposase, inserts the cDNA into the target DNA by direct nucleophilic attack of the 3'-OH cDNA ends at staggered positions on the target. The 5' gaps are repaired (green) by host functions to give target site duplications. doi:10.1128/microbiolspec.MDNA3-0062-2014.f24

elements that appear to use a $Ty_{I-TRANSP}$ for integration (see Chapter 55).

Non-LTR Retrotransposons

As with the LTR-containing retroviruses and retroviral-like retrotransposons, the RNA copy of a non-LTR element is generated by a host polymerase and moves from the nucleus to the cytoplasm where it is translated. Non-LTR-encoded proteins always include reverse transcriptase but may also include other mobility functions including nucleases. Although the RNA and reverse transcriptase assemble to form a non-LTR element RNP in the cytoplasm, no reverse transcription occurs until the RNP interacts with the target nuclear DNA where the element will insert. A DNA 3'-OH at the target site serves as the primer for reverse transcrip-

tion using an RNA copy of the element as a template that generates the DNA copy of the element at the insertion site. This reaction is called target primed reverse transcription.

R2 is a Target Site-Specific Non-LTR Element

One of the best-understood non-LTR elements is R2 (see Chapter 49), originally found in *Drosophila melanogaster* but now known to be widespread in animals. The signature of this element is that it is highly target site-specific, inserting into a particular site in rDNA (Fig. 26). Other target site-specific elements have also been identified and are often associated with specific sites in repeated DNAs (see Chapter 50).

The R2 RNA is derived from the rRNA transcript and is excised by a self-cleaving ribozyme that generates the precise 5' end of the RNA. R2 encodes a single ORF that has both reverse transcriptase and target site-specific endonuclease, as well as binding sites for the 5' and 3' ends of the element (Fig. 26). R2 insertion involves two R2 RNPs. One subunit nicks the template DNA at the specific insertion site and then uses the released

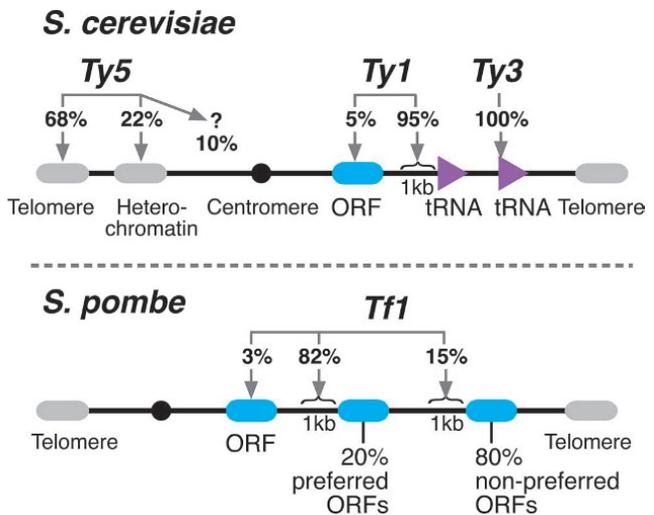


Figure 25 Retroviral-like retrotransposons in yeast are highly target site-selective. Next gene sequencing has been used to map hundreds of thousands of retroelement insertions in yeast. In *Saccharomyces cerevisiae*, Ty_1 and Ty_3 insert at tRNA genes (and other *poll III* genes). Ty_3 is highly site-specific, inserting within a nucleotide or two of the transcript start site and is positioned by interaction with tRNA transcription factors. Ty_1 inserts preferentially on nucleosomal DNA within about 1 kb upstream of transcription start sites. Ty_5 inserts preferentially into heterochromatic regions, including telomeres. In *S. pombe*, $Tf1$ inserts preferentially within about 1 kb upstream of *pol II* ORF promoters. The regions upstream of some genes are far more attractive to $Tf1$ than others. doi:10.1128/microbiolspec.MDNA3-0062-2014.f25

DNA 3'-OH on the target DNA as the primer for reverse transcription using the *R2* RNA as a template. As polymerization proceeds, this will remove the 5' end from the subunit. The second RNP then cleaves the top strand, using the released 3' end of the target DNA as the primer. TPRT-dependent *R2* insertion, like all TPRT events, results in variously sized target site duplications.

LINEs are Another Class of Non-LTR Retrotransposon

Long interspersed elements (LINE) elements are major components of mammalian genomes and have been studied intensively in humans and mice (see Chapter 51). A full-length of the *L1* LINE element is about 6 kb long, encodes two ORFs, ORF1 and ORF2, and ends with a polyA tail (Fig. 23). ORF1 encodes an RNA chaperone protein and ORF2 contains reverse transcriptase and APE-like endonuclease domains. Following translation, the *L1* RNA and the ORF1 and ORF2 proteins assemble into an RNP, which then returns to the nucleus. Formation of this RNP is thought to underlie the preferential *cis*-action of the LINE transposition proteins. Although they have yet to be defined in detail, the RNP also contains multiple host proteins.

The binding site preference of the APE-endonuclease for AT-rich DNA sequences mediates *L1* target site selection. Cleavage of the target DNA by the endonuclease generates a free 3'-OH DNA end that will be used as the primer for reverse transcriptase during TPRT (Fig. 27). Pairing of the *L1* polyA tail with the T-rich DNA strand released by cleavage at the target sites guides APE pairing of the template RNA to the target DNA (Fig. 25). The details of synthesis of the second DNA strand synthesis and the involvement of host proteins remain to be determined. *L1* insertion results in variable length target site duplications. Not surprisingly, multiple regulatory systems modulate the frequency of *L1* transposition.

Not all *L1* transposition events yield straightforward insertions of the full-length element. Many are 5' truncated or contain internal rearrangements. Notably *L1* transposition can also result in rearrangements of sequences outside the *L1* (see Chapter 51). Transduction of host sequences, both upstream and downstream of *L1*, have been observed as have alterations around the target site such as large deletions.

As with the movement of other mobile genetic elements, *L1* insertions can lead to gene inactivation by gene disruption and there are multiple examples of human disease resulting from *de novo* *L1* insertions. As with other mobile elements, regulatory signals within *L1* can also affect expression of adjacent genes.

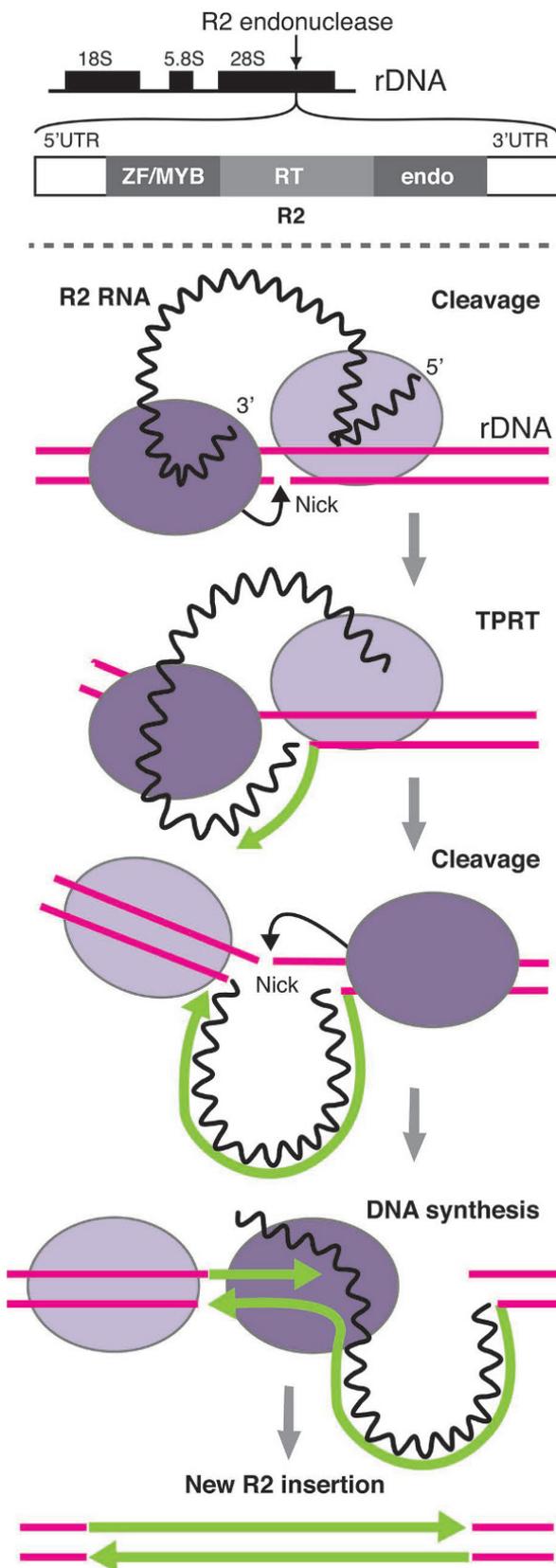
L1 has very successfully colonized the human genome and many other genomes (Fig. 23). Sequences derived from *L1* make up about 17% of the human genome. Although most of these sequences are inactive *L1* fossils, the human genome does contain about 80 to 100 *L1* active elements. Although the great majority of human *L1*s are fixed, that is, occurred before the emergence of modern humans, the ongoing activity of some *L1*s means that *L1* retrotransposition does contribute to human variation. Indeed, there may be millions of "private" *L1* elements that have recently transposed in different human lineages. Comparison of multiple human genomes suggests that a new germline *L1* insertion occurs about once every 100 births.

L1 transposition in the human germline (or in very early embryos) is required to pass new *L1* insertions to progeny. *L1* transposition also occurs, however, in somatic cells. Notably, *L1* transposition has been observed in the human brain and in multiple tumor types, raising interesting questions about the possible contributions of *L1* transposition to somatic phenotypic changes.

SINEs are Nonautonomous Non-LTR Elements

L1 is an autonomous transposable element, that is, it encodes its own mobility proteins as well as the mobile element itself. The *L1* retrotransposition proteins also mobilize other nonautonomous, non-LINE RNAs, a prominent class of which are short interspersed elements (SINEs) (as well as mRNAs, see below) (Fig. 23). RNA copies of SINE elements transpose via TPRT that requires only the *L1* ORF2 and, like *L1* insertions, are flanked by variable target site duplications. SINEs do not encode proteins and are derived from short (<400 bp) housekeeping RNAs, such as tRNAs, rRNAs, and 7SL RNA, the RNA component of the signal recognition particle, that transpose repetitively (Fig. 23). A very successful class of human SINE elements is derived from 7SL RNA, which is called *Alu* because their DNA copies contain an *AluI* restriction site and also have an internal *pol III* promoter. The human genome contains more than a million *Alu* elements, forming about 11% of the genome.

There are multiple families of *Alu* elements but only some are currently active. The frequency of *Alu* transposition is, however, higher than the frequency of *L1* transposition. *Alus* continue to contribute to variation within the human population again with likely millions of "private" insertions. As with *L1*s, there are multiple examples of *de novo* *Alu* insertions that have resulted in human disease. In addition to insertional mutagenesis, *Alu* elements also contribute to human variation



and cause disease by genome rearrangements that result from nonallelic homologous recombination between *Alu* elements that generate in chromosomal deletions, inversions, and translocations.

Another active SINE element in the human genome is SVA (Fig. 23), which is about 2 kbp long and is a composite of other SINE elements. These are very young elements and are present in only a few thousand copies.

Processed Pseudogenes

Pseudogenes are nonfunctional copies of active genes, which accumulate mutations as they are no longer under selection. One class of pseudogenes called processed pseudogenes (38), also called retrogenes (39), have arisen by *L1*-mediated TPRT of mRNAs that have had their introns removed by splicing (Fig. 23). Pseudogenes do have the *L1* TPRT hallmark of target site duplications. Pseudogenes generally lack active promoters, however, because the necessary regulatory sequences are located upstream of transcription start sites, and thus, are not present in mRNAs. Pseudogenes are common in mammalian genomes. For example, there are about 12,000 processed pseudogenes in the human genome compared to about 20,000 protein coding human genes. These elements are polymorphic in the human genome, indicating ongoing insertion.

Mobile Group II Introns: Back to the Future

Group II introns are both introns, which splice from mRNA, and mobile genetic elements, which reverse-splice into target DNA sites where they are converted to DNA via reverse transcription (see Chapter 52). They are widespread in bacteria and are also found in mitochondrial and chloroplast genomes. Notably, they are also the likely progenitors of eukaryotic spliceosomal introns, having invaded the nuclear genome from organelle genomes.

Figure 26 The non-LTR element R2 inserts site-specifically into rDNA by target-primed reverse transcription. R2 inserts into a specific site in the rDNA genes, which is determined by the target-site selectivity of its own endonuclease. Retrotransposition begins with the formation of a RNP containing R2 RNA and two R2 proteins. The R2 subunit bound to the 3' end of the RNA nicks the target DNA at a specific sequence. The resulting target 3'-OH is used as a primer for reverse transcription (green) by the element-encoded RT that uses R2 RNA as a template. Reverse transcription extends to the end of the RNA template, followed by cleavage of the top strand by the other subunit. DNA synthesis (green) of the other R2 DNA strand completes insertion of the element.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f26

Target-Primed Reverse Transcription

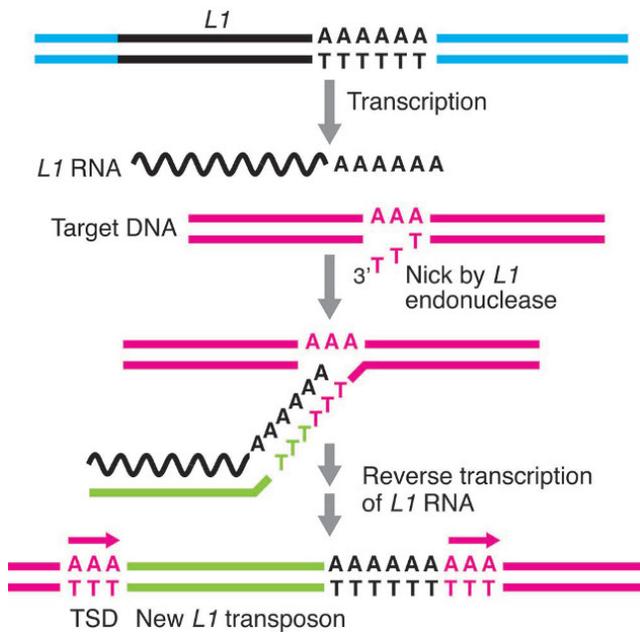


Figure 27 Long interspersed elements (LINE) elements insert by target-primed reverse transcription into target sites cleaved by ORF2 endonuclease. Transcription of the *L1* element initiates retrotransposition. Following synthesis of ORF1 and ORF2 proteins, which associate preferentially with *L1* RNA, the RNP enters the nucleus where the ORF2 APE endonuclease makes a nick in the AT-rich target site. Nicking releases a T-rich strand that pairs with the polyA tail of the *L1* RNA and the 3'-OH of the target DNA is used as the primer for reverse transcription (green) of the template *L1* RNA. Following multiple other processing steps, a new copy of *L1* occupies the target DNA.
doi:10.1128/microbiolspec.MDNA3-0062-2014.f27

Introns encode an RNA that can auto-catalytically excise (Fig. 28). They also encode a multifunctional ORF called an intron-encoded protein (IEP), or sometimes called a maturase, that assists with both RNA-splicing from mRNA and reverse splicing of the RNA into DNA. The IEP also encodes a reverse transcriptase and sometimes an endonuclease that facilitates TPRT by cleaving the target DNA to release a 3'-OH, which will be the primer for reverse transcription of the element RNA.

Following transcription of a parental mRNA, the intron folds into an elaborate structure, which is a self-splicing ribozyme in which intron sequences pair specifically with the exon-intron boundaries. The IEP interacts with the RNA to form a RNP. This ribozyme is highly similar in structure and enzymatic activity to the spliceosomal machinery. Facilitated by the IEP, RNA

splicing joins the exonic RNA and generates the spliced lariat RNA.

The mobile group II intron inserts into an intron-less allele of the donor gene.

Bound tightly to the IEP, intron boundary sequences in the RNA base pair specifically with exonic sequences on a single strand of the target DNA, leading to reverse splicing of the RNA into the target DNA. Conversion of this RNA strand into a DNA strand occurs by IEP-mediated TPRT. The 3'-OH DNA required as the primer for reverse transcription is generated by the endonuclease, which cuts the unpaired target DNA strand, or, in a clever variation, the 3'-OH of an Okazaki fragment if reverse splicing occurs into a target single strand adjacent to a replication fork. A wide variety of bacterial proteins, including RNases and DNases, are needed to generate intact duplex DNA at the target site (40).

Because of base pairing between the RNA and the DNA target during reverse transcription, group II retro-mobility is usually highly target site-specific, the intron inserting into the intron-less allele of its parent gene in a process called “retrohoming.” At lower frequency, group II introns insert into targets with relaxed sequence specificity in a process called “retrotransposition.” Deliberate manipulation of the intron sequence can result in targeted insertion into a chosen site.

Although the mechanism of group II intron mobility is well understood, the cellular regulation of mobility is relatively unexplored. Intriguing recent work suggests that they may be activated to move under conditions of stress (41).

PLEs are Another Class of Retroelement

Penelope-like elements (PLEs) are a widespread class of retroelement named after *Penelope*, an element isolated from *Drosophila virilis* (42), that have, to date, been observed only in eukaryotes. *Penelope* encodes a single ORF that has a reverse transcriptase domain and usually an endonuclease domain that is related to a homing endonuclease. Both activities have been shown to be functional in *Penelope* (43). *PLEs* also often encode spliceosomal introns, which is unusual for an element that presumably moves by a RNA intermediate. The ends of *PLEs* are usually direct repeats several hundred base pairs long that encode self-splicing hammerhead ribozymes (44), which may mediate the excision of the *Penelope* RNA from a larger transcript as occurs with *R2* (see Chapter 49).

An intriguing class of *PLEs*, which lack an obvious endonuclease domain, are associated with telomeres in diverse species (45) and have been postulated to move using the 3' chromosome ends as a priming site. The

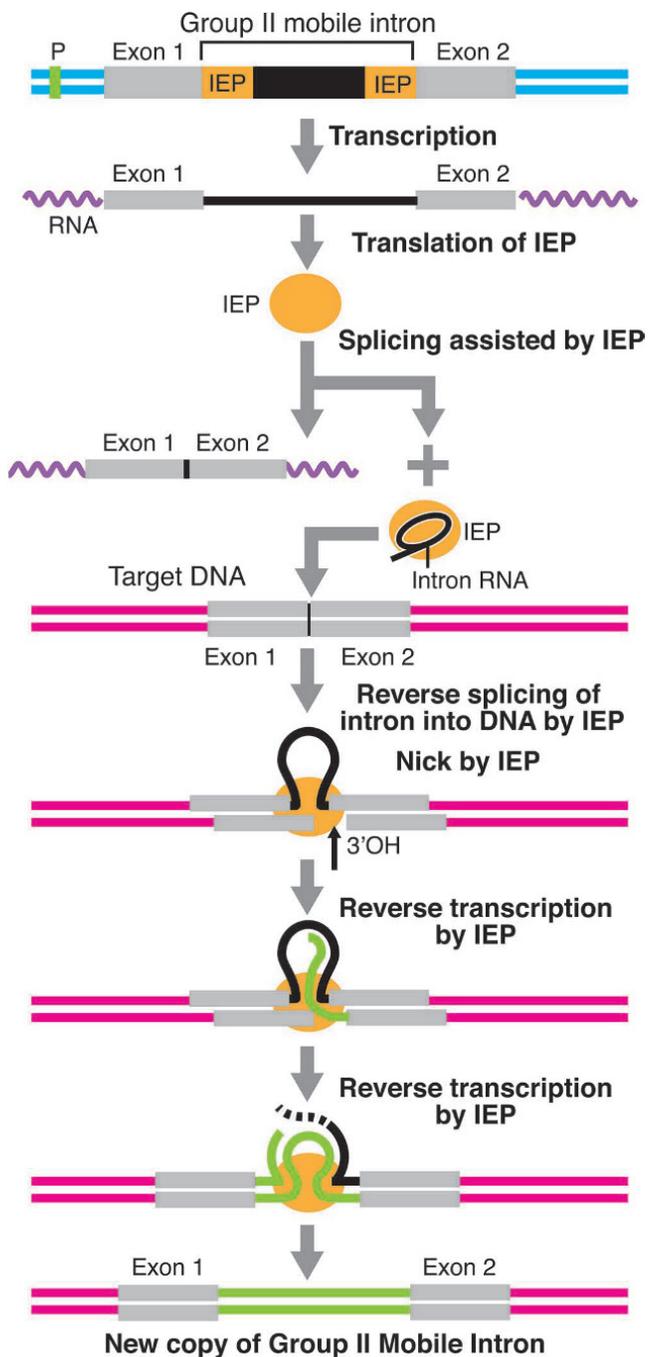


Figure 28 Mechanism of group II intron mobility. Transposition begins with transcription of the gene containing the intron, followed by synthesis of the multifunctional intron-encoded protein (IEP) protein that assists in RNA splicing and reverse splicing into the target DNA and has reverse transcriptase and endonuclease activity. After RNA splicing, the excised lariat RNA remains bound to IEP and then reverse splices into an allele of the gene that lacks the intron. This is highly site-specific, being mediated by base pairing between the intronic sequences and exonic sequences in the target

PLE reverse transcriptase is closely related to the reverse transcriptase of telomerase and it has been suggested that *PLEs* may be a missing link between telomerase and modern retroelements (46). Interestingly, the non-LTR elements *Het-A* and *TART* form the telomers of chromosomes in *Drosophila*.

DOMESTICATED REVERSE TRANSCRIPTASES

Bacterial Reverse Transcriptases

It is generally thought that reverse transcriptase originated in bacteria. group II introns are the only known mobile retroelements in bacteria, but they are not the only source of bacterial reverse transcriptase (see Chapter 54).

Retrons: A Retroelement Without a Function

The first reverse transcriptase to be identified in bacteria is encoded by an element called a “retron” (47). A retron is an unusual DNA-RNA molecule in which single strand DNA is covalently linked at its 5' end to single strand RNA in what is called multicopy single strand DNA (msDNA). msDNA has been found in a number of bacteria including *Escherichia coli* B and can be present in hundreds of copies per cell but its function remains undetermined.

Diversity-Generating Retroelements

While group II introns and retrons account for about 90% of bacterial reverse transcriptases, there are other reverse transcriptases. Some of these are found in diversity generating retroelements (DGRs) (see Chapter 53), which generate high frequency, targeted mutagenesis into particular gene segments to generate very high levels of sequence diversity. This mutagenesis results from error-prone reverse transcription that changes an A in the mutagenized region to any other nucleotide. Similar to TPRT in the movement of non-LTR elements, a region of the template RNA is used as the template for error-prone reverse transcription. It is not yet known, however, how the mutagenized DNA replaces the original coding region. DGRs have been characterized that diversify phage tail fibers, which mediate phage binding to variable surface proteins in *Bordatella*, and which diversify surface lipoproteins in

DNA. A 3'-OH on the target DNA provides the primer for reverse transcription (green) of the intron, which generates a target gene containing the intron.

doi:10.1128/microbiolspec.MDNA3-0062-2014.f28

Legionella. Genome sequencing has identified hundreds of DGRs in phage and bacterial genomes whose functions remain to be determined.

Domesticated Reverse Transcriptases in Eukaryotes

The most widespread domesticated reverse transcriptase in eukaryotes is telomerase, the ribonucleoprotein complex containing reverse transcriptase that uses an RNA template to add DNA sequences called telomers to the tips of linear chromosomes to complete their end-replication. As noted above, the reverse transcriptase of *Penelope* elements may be a link between ancient retroelements and telomerases (48). Other eukaryotic genes called RVT genes appear to be domesticated reverse transcriptases (45) but their functions are not yet known.

PERSPECTIVE: “ALL THAT WE KNOW IS STILL INFINITELY LESS THAN ALL THAT REMAINS”

William Harvey, 17th century scientist

Although *Mobile DNA III* includes much about what is known about many interesting aspects of mobile DNA, what is unknown, of course, remains most intriguing and exciting.

Although multiple types of DNA breakage and joining reactions have been analyzed *in vitro* at the biochemical and structural level, we have only a few snapshot views of a limited number of steps in what are active and dynamic processes. Structural analysis of multiple steps in these reactions, as well as single molecule approaches, will provide deeper insights. Biochemical and structural analysis, even of relatives of already known systems, will also be valuable. It is notable that an entirely new class of transposases has been discovered in the past year. Other systems, such as *Helitrons*, seem poised for *in vitro* analysis.

The conditions of *in vitro* reactions are, however, quite different from the cellular environment. Much remains to be learned about mobile DNA in the context of chromatin and three- and four-dimensional genomes. Other than retroviruses and retroviral-like transposons, few genome-wide screens for host factors have been done even in well-developed model organisms. New genetic, biochemical, and proteomic methods will be very useful for such investigations. Advances in microscopy and *in vivo* imaging should also allow more “real-time” analysis of mobile DNA rather than assays that depend on many generations of cell growth.

No doubt, genome sequencing, especially of single cells from different environments in multicellular organisms, as well as genomes of newly discovered organisms, in particular eubacteria and archaea from novel environments, will reveal more examples of mobile DNA. As viruses are the most numerous entities in the biosphere and they are intimately dependent on their hosts, these interactions, both in facilitating and blocking viral growth, are likely to be rich sources of information about DNA rearrangements. It will be especially important to look *de novo* for novel types of mobile DNA rather than just identifying new relatives of already known elements.

One already very well known mechanism for DNA rearrangement is DNA breakage and joining via covalent protein-DNA intermediates of the tyrosine and serine recombinases. These systems are wide spread in eubacteria and archaea. The very successful use of these recombinases for genome engineering in a variety of eukaryotes raises the question of why endogenous systems of this type are not more widespread in eukaryotes.

Another interesting issue concerns the role of RNA in DNA mobility. Certainly, RNA is a very useful template for reverse transcription and various types of retroelements have been extremely successful mobile DNAs. Reverse splicing of group II intron RNA into DNA reveals the awesome power of RNA-DNA chemistry and the accuracy and efficiency of RNA-templated CRISPR-Cas DNA cleavage is also notable. The apparent involvement of RNAs as guides for some aspect(s) of gene assembly in ciliates is also intriguing. Perhaps RNA can play an, as yet unknown, intimate role in templating and/or executing the DNA cleavage and even DNA joining reactions that underlie some DNA rearrangements.

Many adventures in mobile DNA await!

Acknowledgments. I thank Mick Chandler, Marty Gellert, Alan Lambowitz, Phoebe Rice, and Suzanne Sandmeyer for being such excellent editors. I also thank Patti Kodeck for so successfully managing all the communications between the editors, authors, and ASM, and for keeping us all on track. I am grateful to Mick Chandler, Fred Dyda, Tom Eickbush, Rasika Harshey, Reid Johnson, Haig Kazazian, Dixie Mager, Didier Mazel, John Moran, Phoebe Rice, Suzanne Sandmeyer, and Greg Van Duyne for their help with the figures and answering my questions. Special thanks to Reid, Greg, and Phoebe for providing files of their lovely structures, to Suzanne for her useful comments and to E. Hemingway (“A Moveable Feast”) and W. Harvey (“Everything we know...”) for being so quotable. I am also very grateful to Patti Kodeck for her skill and patience in editing the too many versions of this manuscript. It was a very special pleasure to work with Helen McComas on the figures – they are very much the better for her skill, suggestions, and willingness to keep revising them. I am an Investigator of the Howard Hughes Medical Institute.

References

1. Stoddard B. 2014. Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA* 5:7.
2. Chaumeil J, Skok J. 2013. A new take on v(d)j recombination: transcription driven nuclear and chromatin reorganization in RAG-mediated cleavage. *Front Immunol* 4:423.
3. Kim M, Lapkouski M, Yang W, Gellert M. 2015. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature* 518:507–511.
4. Jones J, Gellert M. 2004. The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev* 200:233–248.
5. Zhou L, Mitra R, Atkinson P, Hickman A, Dyda F, Craig NL. 2004. Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature* 432:995–1001.
6. Hickman AB, Ewis H, Li X, Knapp J, Laver T, Doss A-L, Atkinson P, Craig NL, Dyda F. 2014. A hAT DNA transposase with an unusual subunit organization to bind long asymmetric ends. *Cell* 158:353–367.
7. Hencken C, Li X, Craig NL. 2012. Functional characterization a transposon with a RAG connection. *Nat Struct Mol Biol* 19:834–836.
8. Kapitonov V, Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* 3:e181.
9. Bracht J, Fang W, Goldman A, Dolzhenko E, Stein E, Landweber L. 2013. Genomes on the edge: programmed genome instability in ciliates. *Cell* 152:406–416.
10. Streit A. 2012. Silencing by throwing away: a role for chromatin diminution. *Dev Cell* 23:918–919.
11. Madireddi M, Coyne R, Smothers J, Mickey K, Yao M, Allis C. 1996. Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. *Cell* 87:75–84.
12. Thomas J, Phillips C, Baker R, Pritham E. 2014. Rolling-circle transposons catalyze genomic innovation in a mammalian lineage. *Genome Biol Evol* 6:2595–2610.
13. Ferguson A, Jiang N. 2011. Pack-MULEs: recycling and reshaping genes through GC-biased acquisition. *Mob Genet Elements* 1:135–138.
14. Feng X, Colloms S. 2007. *In vitro* transposition of ISY100, a bacterial insertion sequence belonging to the Tc1/mariner family. *Mol Microbiol* 65:1432–1443.
15. Krupovic M, Makarova K, Forterre P, Prangishvili D, Koonin E. 2014. Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol* 12:36.
16. van der Oost J, Westra E, Jackson R, Wiedenheft B. 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* 12:479–492.
17. Nunez J, Lee A, Engelman A, Doudna J. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519:193–198.
18. Nunez JK, Kranzusch P, Noeske J, Wright A, Davies CW, Doudna J. 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* 21:528–534.
19. Krupovic M, Koonin EV. 2015. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat Rev Microbiol* 13:105–115.
20. Yuan Y, Wessler S. 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci U S A* 108:7884–7889.
21. Syvanen M, Hopkins JD, Clements M. 1982. A new class of mutants in DNA polymerase I that affects gene transposition. *J Mol Biol* 158:203–212.
22. Han M, Xu H, Zhang H, Feschotte C, Zhang Z. 2014. Spy: a new group of eukaryotic DNA transposons without target site duplications. *Genome Biol Evol* 6:1748–1757.
23. Harshey R. 2012. The Mu story: how a maverick phage moved the field forward. *Mobile DNA* 3:21.
24. Montano S, Pigli Y, Rice P. 2012. The mu transposome structure sheds light on DDE recombinase evolution. *Nature* 491:413–417.
25. Garcillan-Barcia M, Bernales I, Mendiola M, de la Cruz F. 2002. IS91 Rolling-Circle Transposition, p 891–904. In Craig NL, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA II*. ASM Press, Washington DC.
26. Mitra R, Li X, Kapusta A, Mayhew D, Mitra R, Feschotte C, Craig NL. 2013. Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils the first active mammalian DNA transposon. *Proc Natl Acad Sci U S A* 110:234–239.
27. Nunes-Duby S, Azaro M, Landy A. 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr Biol* 5:139–148.
28. Birling MC, Gofflot F, Warot X. 2009. Site-specific recombinases for manipulation of the mouse genome. *Methods Mol Biol* 561:245–263.
29. Rutherford K, Van Duyne G. 2014. The ins and outs of serine integrase site-specific recombination. *Curr Opin Struct Biol* 24:125–131.
30. Rutherford K, Yuan P, Perry K, Sharp R, Van Duyne G. 2013. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res* 41:8341–8356.
31. Roberts A, Mullany P. 2011. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev* 35:856–871.
32. Das B, Martinez E, Midonet C, Barre F. 2013. Integrative mobile elements exploiting Xer recombination. *Trends Microbiol* 21:23–30.
33. Val M, Bouvier M, Campos J, Sherratt D, Cornet F, Mazel D, Barre F. 2005. The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholerae*. *Mol Cell* 19:559–566.
34. Goodwin T, Butler M, Poulter R. 2003. Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology* 149:3099–3109.
35. Kojima K, Jurka J. 2011. Crypton transposons: identification of new diverse families and ancient domestication events. *Mobile DNA* 2:12.

36. Doak T, Witherspoon D, Jahn C, Herrick G. 2003. Selection on the genes of *Euplotes crassus* Tec1 and Tec2 transposons: evolutionary appearance of a programmed frameshift in a Tec2 gene encoding a tyrosine family site-specific recombinase. *Eukaryot Cell* 2:95–102.
37. Jacobs M, Sanchez-Blanco A, Katz L, Klobutcher L. 2003. Tec3, a new developmentally eliminated DNA element in *Euplotes crassus*. *Eukaryot Cell* 2:103–114.
38. Kazazian H Jr. 2014. Processed pseudogene insertions in somatic cells. *Mobile DNA* 5:20.
39. Richardson S, Salvador-Palomeque C, Faulkner G. 2014. Diversity through duplication: whole-genome sequencing reveals novel gene retrocopies in the human population. *Bioessays* 36:475–481.
40. Coros C, Piazza C, Chalamcharla V, Belfort M. 2008. A mutant screen reveals RNase E as a silencer of group II intron retromobility in *Escherichia coli*. *RNA* 14:2634–2644.
41. Coros C, Piazza C, Chalamcharla V, Smith D, Belfort M. 2009. Global regulators orchestrate group II intron retromobility. *Mol Cell* 34:250–256.
42. Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, Corces VG. 1997. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proc Natl Acad Sci U S A* 94:196–201.
43. Pyatkov K, Arkhipova I, Malkova N, Finnegan D, Evgen'ev M. 2004. Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. *Proc Natl Acad Sci U S A* 101:14719–14724.
44. Cervera A, De la Pena M. 2014. Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. *Mol Biol Evol* 31:2941–2947.
45. Gladyshev E, Arkhipova I. 2011. A widespread class of reverse transcriptase-related cellular genes. *Proc Natl Acad Sci U S A* 108:20311–20316.
46. Curcio M, Belfort M. 1996. Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell* 84:9–12.
47. Lampson B, Inouye M, Inouye S. 2005. Retrons, msDNA, and the bacterial genome. *Cytogenet Genome Res* 110:491–499.
48. Curcio MJ, Belfort M. 2007. The beginning of the end: links between ancient retroelements and modern telomerases. *Proc Natl Acad Sci U S A* 104:9107–9108.

*Conservative
Site-Specific
Recombination*

II

Makkuni Jayaram,¹ Chien-Hui Ma,¹ Aashiq H Kachroo,¹ Paul A Rowley,¹
Piotr Guga,² Hsui-Fang Fan,³ and Yuri Voziyanov⁴

An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective

2

INTRODUCTION

Tyrosine family site-specific recombinases (YRs), named after the active site tyrosine nucleophile they utilize for DNA strand breakage, are widely distributed among prokaryotes. They were thought to be nearly absent among eukaryotes, the budding yeast lineage (*Saccharomycetaceae*) being an exception in that a subset of its members houses nuclear plasmids that code for YRs (1, 2). However, YR-harboring DIRS and PAT families of retrotransposons and presumed DNA transposons classified as Cryptons have now been identified in a large number of eukaryotes (3, 4). The presence of functional YRs encoded in Archaeal genomes has been established by a combination of comparative genomics and modeling complemented by biochemical and structural analyses (5, 6). Over 1300 YR sequences mined from bacterial genome databases have been organized into families and subfamilies, providing a better understanding of the evolutionary relationships among them (7). These classifications also encourage investigations into the potential functional significance of YRs whose

genes are present as pairs or trios in bacterial and plasmid genomes.

YRs are remarkable enzymes that utilize a common chemical mechanism to bring about a wide array of biological consequences. They range from the choice of lysogenic or lytic developmental pathways in phage λ and related phage, equal segregation of phage, plasmid and bacterial chromosomes by resolving genome dimers or multimers formed by homologous recombination into monomers, to the resolution of hairpin telomeres that mark the termini of certain bacterial and phage genomes (8, 9, 10, 11, 12). In addition, YRs promote the transposition of conjugative mobile elements, the resolution of cointegrate intermediates formed by the Tn3-related toluene catabolic transposon Tn4651, the unidirectional insertion of the *Vibrio cholerae* phage CTX ϕ into the host chromosomes and the copy number control of budding yeast plasmids (13, 14, 15, 16, 17). A subset of YRs has been utilized as tools for directed genome manipulations with potentially important biotechnological and medical applications (18).

¹Department of Molecular Biosciences, UT Austin, Austin, TX 78712; ²Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Department of Bio-organic Chemistry, Lodz, Poland; ³Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei 112, Taiwan; ⁴School of Biosciences, Louisiana Tech University, Ruston, LA 71272.

In this overview of tyrosine site-specific recombination, we present our current understanding of the mechanism of the reaction from biochemical, chemical, structural and topological perspectives, and highlight the utility of this knowledge in addressing problems of fundamental importance in biology and in developing new technologies for biomedical engineering (see also chapters by M. Boocock, A. Landy, A. Segall, G. van Duynne, D. Mazel, F-X Barre, J. Gardner, and G. Chaconas).

THE RECOMBINATION REACTION: SYNAPTIC ORGANIZATION OF DNA PARTNERS AND STRAND EXCHANGE MECHANISM

The biochemically and structurally most well characterized YRs are phage λ integrase (λ Int), phage P1 coded Cre, Flp coded for by the *Saccharomyces cerevisiae* plasmid 2 micron circle and XerCD of *Escherichia coli* (10, 19, 20, 21, 22, 23). They have provided the templates for the chemical and conformational attributes of the strand cleavage and strand exchange steps during tyrosine recombination. The reaction is executed in the context of two core DNA target sites, each bound by two recombinase monomers, brought together in a synaptic complex by protein-protein interactions [Fig. 1(A)].

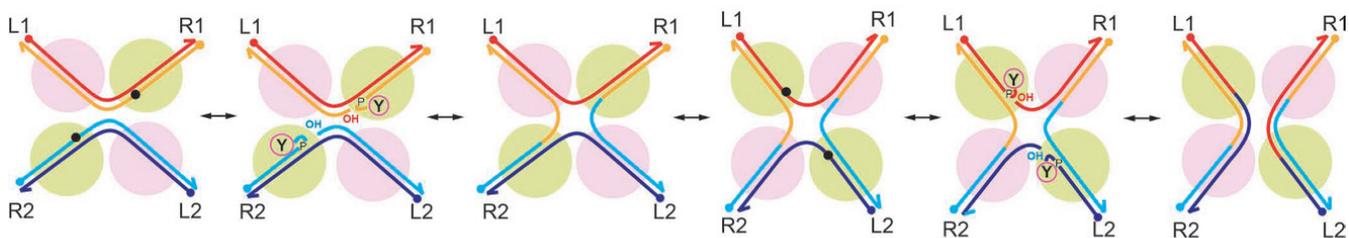
The association of a recombinase, a monomer in solution, with its binding element (a little over one turn of DNA) activates the scissile phosphate adjacent to it. The amino- and carboxyl-terminal domains of the recombinase cradle the DNA between them through a small number of base-specific and many more phosphate contacts. Recognition specificity is imparted to a significant degree through indirect readout. The 13 bp Flp binding element, for example, contains an A/T-rich segment with a characteristically narrow minor groove, and A to T changes within it are well tolerated with respect to binding (24). At the same time, C to G changes are detrimental to binding, and replacement of guanosine with inosine alleviates this negative effect by eliminating the obstructive 2-amino group from the minor groove. Within a DNA substrate, the scissile phosphates are positioned 6 to 8 bp apart (depending on individual systems) on opposite strands, specifying the extent of the strand exchange region. In general, two identical monomers of a recombinase occupy the two binding elements flanking the strand exchange region in a head-to-head (inverted) fashion. In rare instances, as with XerCD, a target site is bound by one monomer each of XerC and XerD.

Strand cleavage by the active site tyrosine nucleophile utilizes the type IB topoisomerase mechanism, yielding a 3'-phosphotyrosyl intermediate and an adjacent 5'-hydroxyl group (20). Strand exchange involves the nucleophilic attack by the 5'-hydroxyl group on the phosphotyrosyl bond across DNA partners to reseal the strand breaks in the recombinant configuration. The reaction is completed in two temporally separated cleavage-exchange steps, the first yielding a Holliday junction intermediate and the second resolving it into reciprocal recombinants.

INHIBITION OF TYROSINE RECOMBINATION BY AGENTS THAT TARGET HOLLIDAY JUNCTIONS

Short synthetic hexapeptides rich in aromatic amino acids inhibit tyrosine recombination by trapping the Holliday junction intermediate (25, 26, 27) (also chapter by A. Segall). The current model for peptide action, based on gel mobility shift and fluorescence quenching results, together with crystal structure data for a Cre recombinase-Holliday junction-peptide ternary complex, posits that the binding of a peptide dimer across the junction core stabilizes the junction in a nearly square-planar (but nonfunctional) conformation (27, 28, 29). More recent analyses of peptide-junction interactions by a combination of single molecule FRET (fluorescence resonance energy transfer), SAXS (small angle X-ray scattering) and gel mobility shifts suggest that peptide binding yields an ensemble of highly dynamic junction conformations that do not fit the canonical square-planar and stacked X-conformations (unpublished observation). The induced conformational heterogeneity likely results from multiple stacking arrangements of aromatic amino acids with the bases surrounding the junction core, perhaps reflecting an intrinsic property of positively charged hydrophobic peptides. Peptide association with a protein-Holliday junction complex may inhibit subsequent reaction steps by inducing global changes in the junction conformation or local changes in the active site environment. Alternatively, peptide binding could accelerate protein dissociation from the junction, and then inhibit further reaction by inducing unfavorable junction conformations. The concept of inhibiting biologically important nucleic acid transactions by enhancing, rather than constraining, conformational freedom may be broadly applicable to peptide and nonpeptide ligands that recognize specific nucleic acid structures. In addition to inhibiting tyrosine recombination, hexapeptides also impede the unwinding of branched DNA structures by

A



B

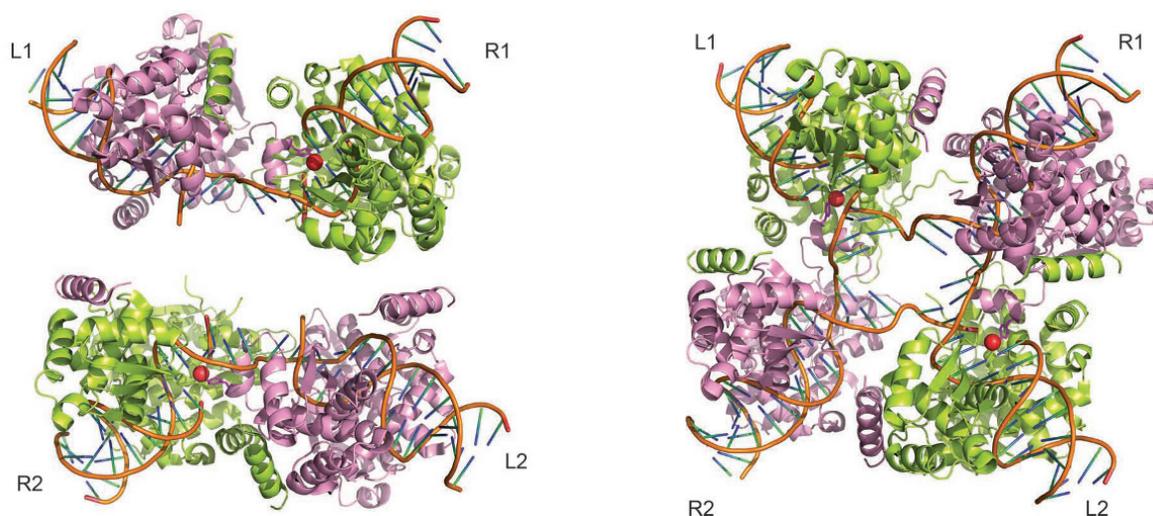


Figure 1 Tyrosine family site-specific recombination. (A) The two target sites, each bound by two recombinase monomers across the strand exchange region, are arranged within the recombination synapse in an almost perfectly planar, antiparallel fashion. The left and right arms of the sites are marked as L1, L2 and R1, R2, respectively. The reaction proceeds by the cleavage/exchange of one pair of strands to form a Holliday junction intermediate, isomerization of the junction, and exchange of the second pair of strands to give the recombinant products ($L1R1 + L2R2 \rightarrow L1R2 + L2R1$). The scissile phosphates engaged by the “active” active sites at distinct stages of the reaction are indicated by the filled circles. (B) The “half-of-the-sites” activity, responsible for the two-step strand exchange mechanism, is revealed by the crystal structure of the Flp-DNA complex (34, 36). Within each recombination partner (left), the green Flp monomer (bound at R1 or R2) is poised to promote the cleavage of the scissile phosphate adjacent to it (red circle). The tyrosine nucleophile for cleavage is donated in *trans* by the neighboring Flp monomer (bound at L1 or L2; magenta). Following isomerization of the Holliday junction intermediate (right), there is a switch between the active and inactive Flp pairs, signifying the imminent cleavage of the scissile phosphates adjacent to Flp monomers bound at L1 and L2. The tyrosine nucleophiles are donated across DNA partners, in the R1 to L2 and R2 to L1 configuration.
doi:10.1128/microbiolspec.MDNA3-0021-2014.f1

the RuvG helicase of *E. coli*, and interfere with Holliday junction resolution by the RuvABC complex (28). Consistent with these properties, the inhibitory peptides appear to hold promise as potential antimicrobial agents (30).

SIMPLE AND COMPLEX YRS: CONTROLLING THE DIRECTIONALITY OF RECOMBINATION

Simple YRs such as Cre and Flp are not particular about DNA topology or target site orientation. They

can act on supercoiled or nicked circles as well as linear molecules, and promote intra- and intermolecular reactions. They bring about DNA inversion between a pair of sites in head-to-head (inverted) orientation and DNA deletion between sites in head-to-tail (direct) orientation. More complex YRs (λ Int and XerCD, for example), depending on the reaction context, may require DNA supercoiling, and may utilize the interaction between accessory factors and their cognate sites to regulate the reaction and/or impart directionality to it. The crystal structures of λ Int tetramers bound to synapsed DNA partners and the Holliday junction intermediate, together with biochemical data, suggest how interactions of the amino-terminal domains of Int with the ‘arm-type’ sequences (which also include multiple binding sites for the accessory proteins: IHF, Xis and Fis) can stabilize 2-fold symmetric configurations of the recombination complex (21, 31). The cumulative DNA-protein and protein-protein interactions thus coordinate Int activity at the core recombination sites as well as bias strand exchange towards a particular outcome. Consistent with this model, when the amino-terminal domain of λ Int is fused to the normally unregulated and bias-free Cre, the latter acquires the regulatory features and directionality of Int (32, 33). The action of the Int-Cre chimera on attenuated core target sites containing appropriate embellishments with the arm-type sequences is controlled by the accessory factors as if recombination were being performed by native Int. The conversion of a simple recombinase, whose origin likely traces back to an ancestral type IB topoisomerase, into a complex one by just the addition of a peptide domain suggests a possible “self-promoting” evolutionary scheme for the emergence of the latter class of recombinases. The relevant gene fusion may be performed by the recombinase itself via low frequency crossover events between suitably positioned secondary target sites, which are fortuitously scattered within a genome and may be harbored by the recombinase gene. Alternatively, the “complexity” domain may be acquired via the action of the host’s recombination machinery. λ Int may thus be a representative of the evolutionary trajectory from topoisomerase to simple and then complex recombinases (see chapter by A. Landy).

HALF-OF-THE-SITES ACTIVITY OF THE YR ACTIVE SITE

Within the recombination synapse, which has a 2-fold symmetry, only two of the four potential active sites are active at any one time (34, 35) [Fig. 1(B)]. The two active sites responsible for the first cleavage-exchange

step become inactive following the isomerization of the Holliday junction intermediate. The other two active sites, which now become activated, resolve the junction into recombinant products. This “half-of-the-sites” activity accounts for the two-step, single-strand exchange mechanism of recombination. Consistent with this mechanism, three Cre or Flp recombinase monomers bound to a three-armed DNA substrate (Y-junction) can yield two functional active sites capable of resolving the junction into a linear and a hairpin product (36, 37, 38).

THE LOCAL GEOMETRY OF PARTNER SITES WITHIN THE RECOMBINATION SYNAPSE

The DNA target sites are almost entirely planar in their paired state, and are arranged in an antiparallel fashion [Fig. 1(A)]. Topological analyses and crystal structure data support this synapse geometry, with strand cleavage and exchange occurring in a diagonal fashion (31, 34, 39, 40, 41, 42, 43, 44). However, a few experiments based on electron microscopy, proximity of DNA ends reported by ligation, and a combination of atomic force microscopy, tangle analysis (see below under “Topological and chiral features of tyrosine recombination”) and modeling suggest parallel arrangement of sites or their nonplanar configuration with a potential tetrahedral geometry for the Holliday junction intermediate (45, 46, 47). These could represent transient or intermediate states that precede the functional synapse or isomerization of the Holliday junction. Or, they could be comprised of aberrant complexes. The reactive orientation of key catalytic residues with respect to the scissile phosphates in the crystal structures strongly imply that the antiparallel disposition of the partner sites revealed by them represents the functional geometry of the recombination synapse (19, 31, 39).

THE YR ACTIVE SITE: KEY CATALYTIC RESIDUES

The signature active site motif of YRs consists of a tyrosine nucleophile assisted by an invariant or highly conserved catalytic pentad: Arg, Lys, His, Arg, and His/Trp (23). In addition, a sixth conserved residue, Asp/Glu, appears to contribute indirectly to transition state stabilization by hydrogen bonding to catalytic residues and promoting the integrity of the active site (39). In Flp, the catalytic hexad is comprised of Arg-191, Asp-194, Lys-223, His-305, Arg-308, and Trp-330 with Tyr-343 providing the cleavage nucleophile (Fig. 2). In Cre (see chapter by G. van Duyne), the corresponding

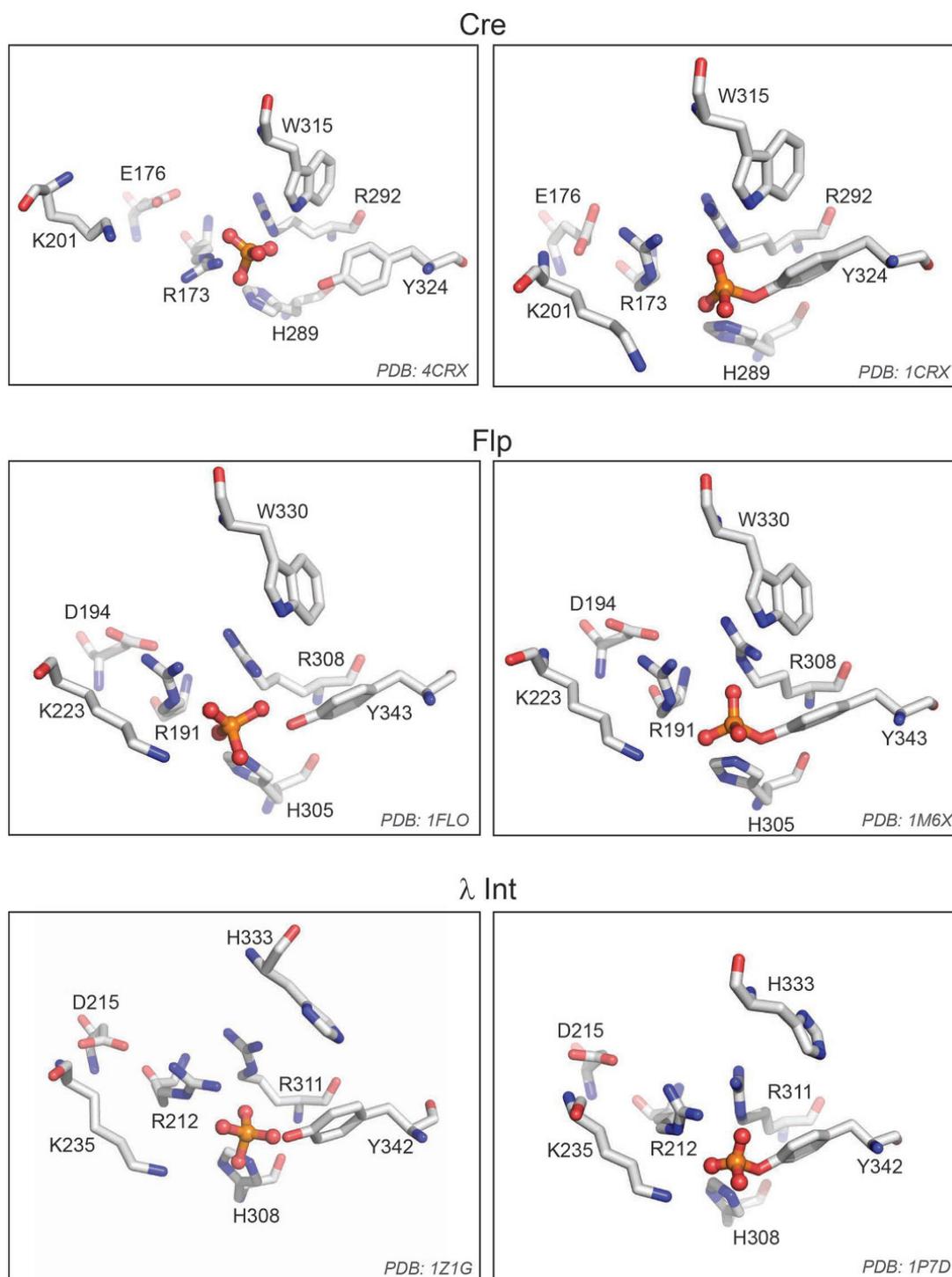


Figure 2 Organization of conserved catalytic residues within the recombinase active site. The arrangements of the catalytic hexad (Arg-Asp/Glu-Lys-His-Arg-His/Trp) and the tyrosine nucleophile in Cre, Flp and λ Int active sites are shown (31, 34, 36, 52, 94, 147). The states of the active site with the scissile phosphate uncleaved and cleaved are shown at the left and right, respectively. The role of the conserved Asp/Glu of the hexad in transition state stabilization is likely indirect, by promoting the structural integrity of the active site.
doi:10.1128/microbiolspec.MDNA3-0021-2014.f2

residues are Arg-173, Glu-176, Lys-201, His-289, Arg-292, Trp-315, and Tyr-324. In λ Int, the second and sixth conserved positions are aspartic acid (Asp-215) and histidine (His-333), respectively. Mutational analyses combined with structural data have provided insights into the mechanistic roles of several of these residues.

As would be expected from chemical principles, the two invariant arginine residues balance the negative charges on the nonbridging oxygen atoms of the scissile phosphate group (31, 34, 39). In Cre, Lys-201 appears to function as the general acid that stabilizes the leaving 5'-hydroxyl group during strand cleavage. The absence of this residue can be rescued when the leaving group pK_a is decreased by substituting the 5'-oxygen by sulfur (5'-thiolate) (48). The potential general acid function of Lys-223 of Flp has not been similarly tested. In the type IB vaccinia topoisomerase, Lys-167 (which corresponds to Lys-201 of Cre) collaborates with Arg-130 (corresponding to Arg-173 of Cre and Arg-191 of Flp) to facilitate leaving group departure, perhaps by a proton relay mechanism (49, 50). His-305 of Flp may serve as the general base that abstracts the proton from Tyr-343 to activate it as a nucleophile (51). A tyrosine mimic with reduced pK_a (3-fluorotyrosine, $pK_a = 8.2$ as opposed to tyrosine, $pK_a = 10.0$), when supplied in the context of a short native peptide, can restore the cleavage potential of Flp

(H305Q) to a large extent. The predominant majority of YRs contain a histidine rather than tryptophan at the final hexad position. The hydrogen bonding between the indole-nitrogen of Trp-315 and the scissile phosphate observed in the Cre crystal structure (52) seemed to suggest that either histidine or tryptophan at this position helps catalysis through their hydrogen bonding potential. However, this is not the case for Cre or Flp, as replacement of Trp-315 or Trp-330, respectively, by histidine results in lower recombination activity compared to replacement by phenylalanine or tyrosine in Cre and phenylalanine in Flp (39, 53). Consistent with these biochemical results, structural data for Flp suggest that hydrophobic/van der Waals contacts made by Trp-330, over a hydrophobic surface of $\sim 380 \text{ \AA}^2$, help position Tyr-343 in its active orientation (53) (Fig. 3). In addition, Trp-330 appears to play a secondary role in stabilizing the 5'-hydroxyl leaving group, as suggested by the stimulation in the cleavage activities of Flp mutants lacking this residue on 5'-thiolate substrates (54). The vanadate transition state mimic structure of Cre-DNA reveals Trp-315 to be located on a turn between two helices (αL and αM), providing a sizable hydrophobic surface on to which the αM helix carrying the Tyr-324 nucleophile docks through van der Waals contacts (39) (see chapter by G. van Duyne).

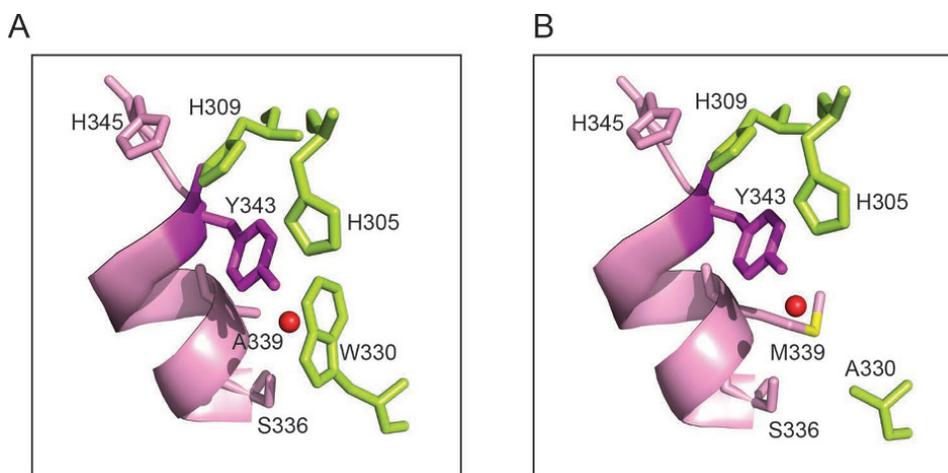


Figure 3 The assembly of the Flp active site in *trans*. (A) In the shared active site of Flp, the tyrosine nucleophile (Tyr-343) is delivered by the donor Flp as part of the M helix (shown in magenta) to the proactive site, whose residues are shown in green, of the recipient Flp. The van der Waals' contacts made by Trp-330 (recipient Flp) with Ser-336 and Ala-339 (donor Flp) are important for the positioning of Tyr-343 (donor) (53). The stacking of His-309 (recipient) over Tyr-343 is stabilized by His-305 (recipient) and His-345 (donor). (B) Consistent with the importance of Trp-330–Ala-339 interaction, the loss of active site function resulting from the W330A substitution can be rescued by the second site A339M mutation, which increases the side chain length at this position (54). The red circle in A and B denotes the scissile phosphate. doi:10.1128/microbiolspec.MDNA3-0021-2014.f3

HIDDEN RNA CLEAVAGE ACTIVITIES OF THE YR ACTIVE SITE

When a ribonucleotide is substituted at the cleavage position or immediately 3' to it within the Flp target (*FRT*) site, two latent RNase activities of Flp can be unveiled (55, 56, 57). The 2'-hydroxyl, when present as part of the cleavage position nucleotide, can attack the phosphotyrosyl intermediate formed by strand cleavage to give a 2',3' cyclic phosphate. Subsequent attack of the cyclic bond by a water nucleophile gives the 3'-phosphate as the end product. This activity, which closely follows the recombination mechanism, has been termed type I RNase. When the 2'-OH is placed on the nucleotide adjacent to the normal cleavage position, it directly attacks the correspondingly shifted scissile phosphate to yield the cyclic phosphate intermediate, which is then hydrolyzed to the 3'-phosphate. This activity, which resembles the classical pancreatic RNase mechanism, has been termed type II RNase. Perhaps the latent RNase activities are the relics of the evolutionary progression of Flp from an elementary nuclease to a recombinase, likely via a topoisomerase. When the interaction between two Flp monomers bound to an *FRT* site is weakened by increasing the length of the strand exchange region, a weak topoisomerase activity of Flp can be unmasked. The type I, but not the type II, RNase activity has been detected in Cre as well (58). The ability of 2'-hydroxyl groups to compete effectively with the tyrosine nucleophile (in the case of Flp) and with the 5'-hydroxyl group (in the case of Cre and Flp), as suggested by their latent RNase activities, speaks to the considerable catalytic flexibility of the tyrosine recombinase active site. These activities also expose the potential threat to recombination by errant nucleophiles that might gain entry into the active site.

ASSEMBLY OF THE YR ACTIVE SITE IN CIS OR IN TRANS

In general, the active site of a tyrosine recombinase is assembled entirely within a monomer, although its strand cleavage activity may be stimulated by allosteric contact with an adjacent monomer. Flp and the related subfamily of YRs coded for by 2 micron-like plasmids of budding yeasts are unusual in that they assemble an active site at the interface of two neighboring monomers (34, 59, 60) [Fig. 1(B); Fig. 3]. Biochemical and structural evidence suggests that the integrase of SSV1, a virus that infects the extremely thermophilic archaeon *Sulfolobus shibate*, may also harbor a shared active site (61, 62). A *cis* active site (the Cre active site, for

example) is responsible for the activation and cleavage of the scissile phosphate engaged by it. A *trans* active site, exemplified by that of Flp, activates the scissile phosphate but relies on the tyrosine nucleophile donated to it for strand cleavage. There are two *trans* modes of DNA cleavage. For the first strand exchange step and formation of the Holliday junction, the tyrosine nucleophile performs cleavage across the strand exchange region within a DNA substrate [Fig. 1(B), left]. For the resolution of the Holliday junction and formation of the recombinant products, the cleavage by tyrosine occurs across partner substrates [Fig. 1(B), right].

Comparison of the crystal structures of Cre and Flp synaptic structures suggests how a simple switch in the connectivity of two helices can switch a *cis* active site into a *trans* active site or vice versa. (23, 34). In the shared active site, the Tyr-343 from an Flp monomer gains entry into the proactive site of the second monomer as part of the protruding "M" helix. Trp-330 of the hexad in the proactive site of Flp may play a particularly important role in helping to dock the M helix by packing against it through contacts with Ser-336, Ala-339, and Tyr-343 (53) [Fig. 3(A)]. These interactions are further augmented by the nearly perfect stacking of His-309 from the recipient Flp over Tyr-343, with likely assistance from His-345 (donor) and His305 (recipient) (34, 63). The extremely weak recombination activity of Flp(W330A) can be restored to nearly wild type level by a second mutation that changes Ala-339 to methionine (54). The longer side-chain of Met-339 located in the M helix can compensate for the lack of Trp-330 in the proactive site, further highlighting the importance of interprotomer hydrophobic interactions in the assembly of the *trans* active site [Fig. 3(B)].

From a purely mechanistic point of view, there is apparently little advantage of a *cis* active site over a *trans* active site or vice versa. By the "Cheshire cat" paradigm, if one were to erase all the amino acid residues from Cre and Flp structures, except for the tyrosine nucleophile and its principal catalytic cohort, the 'catalytic grins' of the two recombinases would look almost identical (64, 65) (Fig. 2). However, the *trans* active site offers considerable advantages in the analysis of recombination mechanism. For example, strand cleavage can be performed by providing exogenous nucleophiles such as hydrogen peroxide or tyramine or by supplying chemically modified tyrosines embedded in a short native Flp peptide (51, 66, 67). As already noted, the potential general base/acid role of His305 of Flp has been inferred from the ability of tyrosine analogs with lowered pK_a to confer cleavage competence on Flp(H305Q).

From a functional perspective, each type of active site may have its strengths and weaknesses. The *trans* active site will delay DNA cleavage until a target site has been occupied by two monomers of the recombinase. However, the time delay between the binding of the two monomers may allow rogue nucleophiles, the abundant water nucleophile for example, to attack the activated phosphodiester bond (68). Since the *cis* active site positions all the catalytic residues, including the tyrosine nucleophile, in concert around the scissile phosphate, the chances for aberrant strand cleavage are minimized. However, cleavage could potentially occur even before the full occupancy of a DNA site by the recombinase. Premature strand breaks may be minimized if the cleavage potential of the *cis* active site is activated by allosteric interactions between recombinase monomers within a DNA site or between partner sites within a recombination synapse. For Flp, whose physiological function is to trigger plasmid DNA amplification by a replication-coupled recombination event (14, 17), DNA cleavage in *trans* may hold special significance. Should an Flp monomer covalently linked to the cleaved phosphate be dislodged from its binding element by the replication machinery, and be unfolded or partially degraded by a protease as a consequence, the Flp monomer bound to the other binding element will be able to promote healing of the DNA break via ligation.

REQUIREMENT OF HOMOLOGY BETWEEN THE STRAND EXCHANGE REGIONS OF RECOMBINATION PARTNERS

According to the generally accepted paradigm, based on evidence from the λ Int, Cre and Flp systems, successful recombination requires perfect homology in the strand exchange regions of the DNA partners, even though the sequence per se of the exchanged segment can be altered in a number of ways without affecting reaction efficiency. The original notion that homology promotes end-to-end branch migration of the Holliday junction through the strand exchange region (69, 70) has been discounted by biochemical and structural evidence (31, 34, 52, 71, 72, 73, 74). The cleaved strands are swapped in a segmental fashion, perhaps as triplets during the first and strand exchange steps (74). Non-homology would disfavor stable strand exchange, as mismatches in DNA hinder the ligation reaction (73). In this model, recombination is blocked by non-homology either because the Holliday junction is not formed or because the junction with mismatched DNA is quickly resolved in the parental mode. For Flp, which

mediates the exchange of 8 bp rather than 6 bp (the extent of exchange in the Cre system), the strand swaps at the initiation and termination steps of recombination may be separated by an intermediate step of limited branch migration through the central base pairs (75).

The strict requirement of homology in strand exchange has been called into question as the mechanisms of more YRs have been brought to light, in particular, the integrases of integrative conjugative elements (ICEs), also referred to as conjugative transposons (16) (see chapter by J. Gardner). These enzymes mediate strand exchange across overlap regions that include 5 to 6 bp nonhomologies. IntDOT, the integrase of the *Bacteroides* conjugative transposon CTnDOT, utilizes a 2 bp homology within a 7 bp overlap region for the first exchange step, and carries out the second exchange in a homology-independent manner [Fig. 4 (A)]. The resulting heteroduplex DNA is resolved by replication.

The CTX ϕ phage of *Vibrio cholera* manipulates the XerCD recombinase of its host for its integration in a rather unusual reaction that also challenges the homology rule in its conventional sense (13) [Fig. 4(B)] (see chapter by F-X Barre). The + strand of the phage DNA folds itself into a forked hairpin structure to generate a pseudo XerCD target site within it. The first strand exchange between the phage and chromosome target sites is mediated by XerC, and utilizes 3 bp of homology adjacent to its cleavage site. The nonhomology adjacent to XerD stops the reaction at this pseudo-Holliday junction stage. Resolution of this structure by replication generates a chromosome with an integrated copy of the phage. As the lysogen is not flanked by functional recombination sites, the integration reaction is irreversible, proscribing the excision of the replicative form of the phage. However, tandem copies of the lysogen generated by successive integration events permits the production of + single-stranded phage genome by a rolling circle type of replication. An analogous strategy of integration via single-strand exchange by a tyrosine recombinase, followed by replication-mediated resolution of the resulting Holliday junction, is also devised during the capture of exogenous gene cassettes by integrons (76), which are important not only for their role in the spread of antibiotic resistance but also for their potential relevance to bacterial genome evolution. As in the case of the *V. cholerae* phage integration, the recombination target site on the cassette is assembled by the folding of a single-stranded DNA region. The phage and integron systems, rather than breaching the homology rule outright, seem to bend it by cleverly manipulating the strand exchange reaction in their favor.

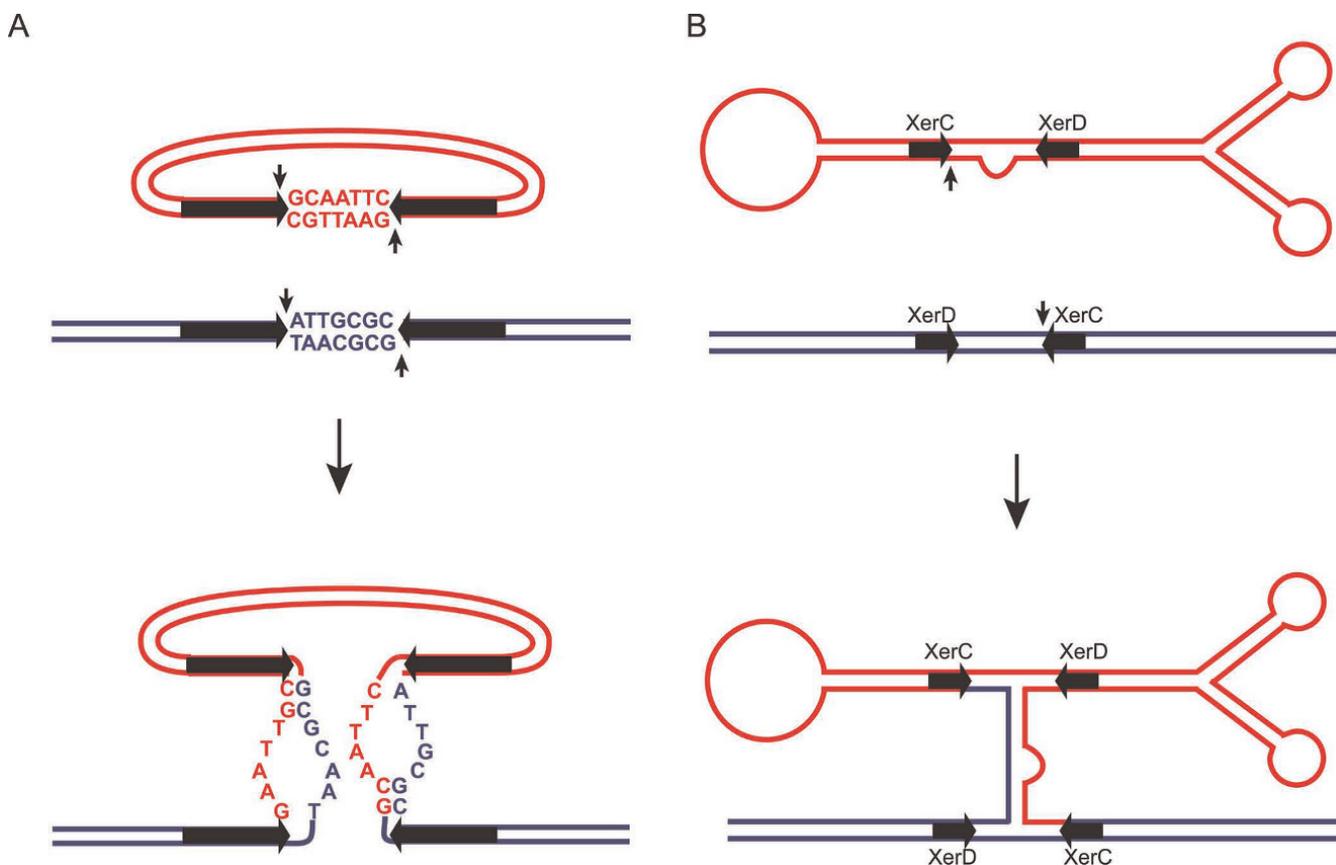


Figure 4 Challenges to the homology rule during tyrosine site-specific recombination. (A) The integrase (IntDOT) of the conjugative transposon CTnDOT catalyzes exchange of both strands between target sites that contain five consecutive nonhomologous positions within the 7 bp segments swapped between them (16). (B) The folded form of the “+” strand of the CTXφ phage contains an imposter target site for XerCD recombinase of its host bacterium, *V. cholera*. Single-strand exchange mediated by the XerC active site between the phage DNA and the bacterial chromosome results in phage integration (13). The heteroduplex integrant in (A) and the pseudo-Holliday junction in (B) are likely resolved via replication. The flat horizontal arrows indicate recombinase binding sites. The short vertical arrows denote points of strand cleavage. doi:10.1128/microbiolspec.MDNA3-0021-2014.f4

Even the archetypal YRs may violate the homology rule in its strictest sense. Analysis of the Cre reaction between a wild type LoxP (the Cre target site) and mutant LoxPs containing single bp substitutions of the 6 bp strand exchange region reveals several types of outcomes: significant amounts of the Holliday junction intermediate without detectable recombinant product, small amounts of the product with higher amounts of the Holliday junction, small amounts of the Holliday junction with higher amounts of the product and small amounts of the product with no detectable Holliday junction (77). The overall strand exchange, the sum of the Holliday junction and recombinant yields, is most diminished for substitutions adjacent to the scissile phosphatase,

whose cleavage initiates the first strand exchange step, with one exception. The T to A transversion at position 2 from the initiation end gives a modest amount of the product with much smaller amounts of the Holliday junction. The accumulation of the Holliday junction as nonhomology shifts from the initiation-proximal positions to the central and termination-proximal positions of the strand exchange region is consistent with the normal execution of the first strand exchange step, while the second exchange step is blocked or severely impeded by nonhomology. Stable single-strand exchange by Cre, dictated by the location of nonhomology, is thus somewhat analogous to the formation of the pseudo-Holliday junction by XerCD during CTXφ integration.

NONHOMOLOGY INDUCED KNOTTING OF SUPERCOILED PLASMIDS BY A YR

An even more flagrant violation of the homology rule is brought to light by Flp reactions between two *FRT* sites, nonhomologous at the central two positions of their 8 bp strand exchange regions and located within negatively supercoiled plasmids (44, 78). While no stable recombinant products result, the reaction ties the plasmids into knots of wide ranging complexity, but all in their parental configuration. When the two sites are in head-to-head orientation (with respect to the six homologous bp), the knots are even numbered; when the sites are in head-to-tail orientation, the knots are odd numbered. The knotting may be explained by two or repeated even rounds of recombination, giving rise initially to unstable recombinants containing mismatched base pairs that then rapidly recombine back to the parental form.

An obvious question is whether the observed knot complexity is due to DNA crossings added during an iterated series of recombination events occurring within a given synapse. A similar increase in topological complexity has been described for the serine recombinases when they attempt to recombine sites that harbor nonhomology in the overlap region (79, 80). However, members of this recombinase family follow a completely different mechanism (see chapter by W. M. Stark). They make concerted double-strand cuts within target sites arranged in a parallel fashion, promote right-handed rotation of the broken DNA through 180 degrees, and reseal the strands in the recombinant configuration. Mismatches between partner strands impede or reverse joining and encourage a second 180 degree rotation to restore complementarity, which favors joining. A repetition of these dual half-rotation steps will progressively increase the complexity of products. Can nonhomology alter the normal synapse geometry of tyrosine recombination and processively generate products of increasing complexity?

An alternative explanation for the knotting reaction by Flp is that the complexity of the knots reflects the topological complexity of the plasmid substrate from which they are generated. Since the pairing of *FRT* sites occurs by random collision, a range of supercoils (crossings between the two DNA domains bordered by the sites; blue \times red in Fig. 5) can be trapped by the synapse. The antiparallel geometry of the *FRT* sites requires an odd number of such crossings between head-to-head sites and even number of crossings between head-to-tail sites (Fig. 5). Depending on the number of trapped interdomainal crossings, the first recombination event will generate an unknotted inversion circle

plus 3-, 5-, 7- etc. crossing knots from the head-to-head sites [Fig. 5(A)]. Similarly, the products from the head-to-tail sites will be a pair of unlinked deletion circles plus 2-, 4-, 6- etc. crossing catenanes (linked circles) [Fig. 5(B)]. When the parental *FRT* sites are nonhomologous, these products contain mismatches, and are prone to a second round of recombination after dissociation of the original synapse. The addition of one more crossing during this step will convert the knots with odd number crossings from the inversion reaction to knots with even number crossings (4, 6, 8 etc.) [Fig. 5(A)]. Similarly, the catenanes from the deletion reaction will be converted to fusion knots with odd number crossings (3, 5, 7 etc.) [Fig. 5(B)]. The prediction then is that when the synapse topology is simplified and made unique, the product topology must be correspondingly simple and unique. This indeed is the case (44). When Flp reaction is carried out after assembling the Tn3 resolvase synapse [which traps precisely three interdomainal negative supercoils, as in Fig. 5(A)], and taking care to minimize random entrapment of supercoils, the product yielded by the head-to-head sites is predominantly a 4-noded knot; that yielded by the head-to-tail sites is predominantly a 5-noded knot (44). The topologies of the products from corresponding reactions between two native (homologous) *FRT* sites are a 3-noded knot and a 4-noded catenane. The difference of one in the crossing numbers between the knot and the catenane is consistent with the need to arrange the sites in the same functional geometry, antiparallel, for them to recombine [Fig. 5(B)]. In addition to the three crossings anchored by resolvase, a fourth crossing must be trapped from the supercoiled plasmid substrate to keep the head-to-tail sites antiparallel (see also the section on "Difference topology"). Thus, nonhomology does not block recombination by Flp; nor does it induce processive recombination by altering the normal reaction mechanism. The unstable (mismatched) recombinants resulting from the first recombination event are restored to the more stable parental state by a second dissociative recombination event.

PROBING ACTIVE SITE MECHANISM USING CHEMICAL MODIFICATIONS OF THE SCISSILE PHOSPHATE GROUP

Mechanistic analysis of strand breakage and joining reactions in nucleic acids has greatly benefited from chemical modifications of the phosphate group in the nonbridging oxygen atoms to alter its electronegativity and/or stereochemistry, and in the bridging oxygen atoms to manipulate leaving group properties. The po-

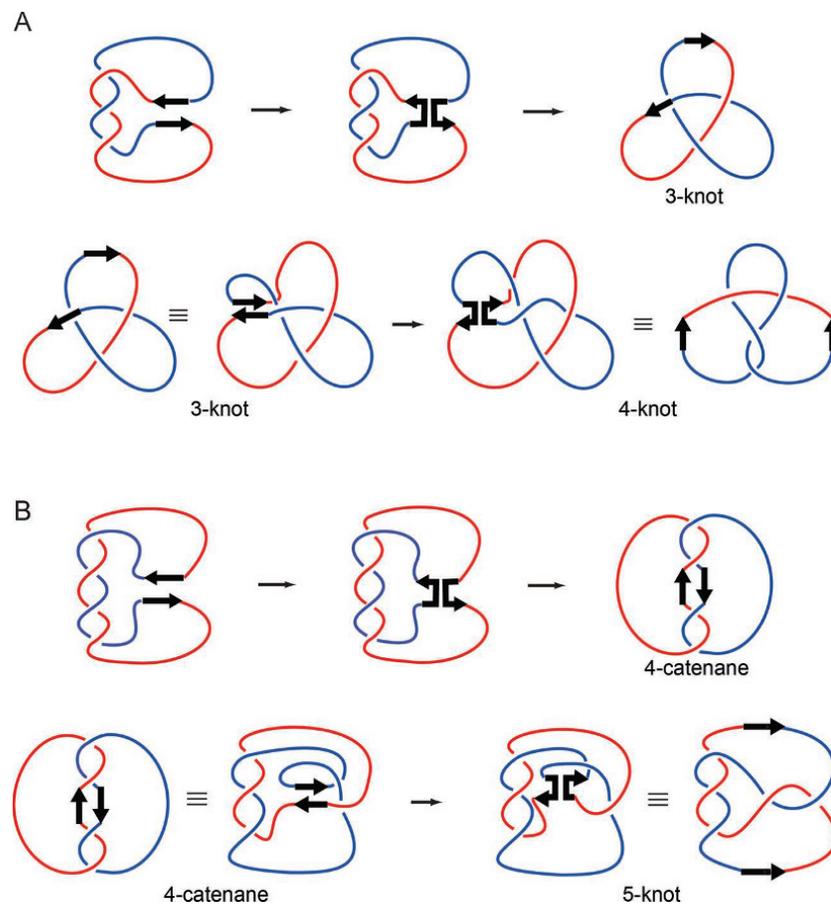


Figure 5 Flp-mediated knotting of supercoiled plasmids by recombination between two *FRT* sites harboring nonhomology within the strand exchange region. (A) The first recombination event between two head-to-head (inverted) *FRT* sites from a synapse containing an odd number of interdomainal (blue \times red) supercoil crossings will generate a torus knot with the same number of crossings. The product from a synapse with one blue \times red crossing will be an unknotted inversion circle, as it takes a minimum of three crossings to form the simplest knot. In the example shown, a 3-noded knot is formed from a 3-crossing synapse. A second recombination event after dissociation of the first synapse, and the assembly of a *de novo* synapse, can give rise to a twist knot with four crossings. (B) For *FRT* sites in head-to-tail (direct) orientation, the first recombination event from a synapse with an even number of interdomainal crossings yields a catenane with the same number of crossings. The product from a synapse with no crossings will be two unlinked deletion circles. The diagram illustrates the formation of a 4-noded catenane from a 4-crossing synapse. A second round of dissociative recombination can convert the 4-noded catenane into a 5-noded knot. In the reactions shown in (A) and (B), intradomainal supercoils (blue \times blue or red \times red crossings) are omitted for clarity, as they do not contribute to knot or catenane crossings. The products from the second rounds of recombination revert to the parental configuration. The non-complementarity in the product formed by recombination between *FRT* sites containing nonhomology in their strand exchange regions encourages a second recombination event that restores base pairing and parental DNA configuration (44). doi:10.1128/microbiolspec.MDNA3-0021-2014.f5

tential general acid role for Lys-201 and a subsidiary role for Trp-330 in leaving group stabilization during strand cleavage by Cre and Flp, respectively, have been revealed with the help of 5'-thiolate substrates (48, 54).

Shuman and colleagues have successfully exploited phosphorothioate (replacement of a nonbridging oxygen by sulfur), methyl phosphonate (MeP; replacement of a nonbridging oxygen by the methyl group) and 5'-

thiolate substrates to investigate the active site mechanisms of vaccinia topoisomerase (81, 82, 83, 84).

Recent studies employing MeP-substrates [Fig. 6(A)] have revealed active site attributes of Cre and Flp that could not have been deduced from reactions of native phosphate containing substrates. These analyses have been performed predominantly using half-site substrates containing a single scissile phosphate or a modified scissile phosphate [Fig. 6(B)], together with recombinase variants harboring specific active site mutations. The chemical synthesis of MeP-half-sites is considerably easier than that of full-sites. Although the half-site reaction involves the breakage of one scissile phosphate within a substrate molecule, it faithfully preserves the chemical mechanism of the normal reaction. Associations of a recombinase-bound half-site can give rise to dimers, trimers and tetramers (85), so that the shared active site assembly and the *trans* mode of DNA cleavage are obeyed during Flp half-site reactions.

The reactivity of Flp variants on MeP-substrates demonstrates that neutralization of the phosphate negative charge in its ground state permits transition state stabilization in the absence of one of the two conserved arginines (either Arg-191 or Arg-308) (68, 86). Flp (R191A) and Flp(R308A) are active in the MeP reaction [Fig. 7(A)], while both these variants are almost completely inactive on phosphate containing DNA substrates. The electrostatic suppression of the lack of a positive charge in the recombinase active site by a compensatory charge substitution in the scissile phosphate of the DNA substrate has been demonstrated for the Cre recombinase as well (87, 88). Not only do Cre(R173A) and Cre(R292A) yield strand cleavage in an MeP-half-site, the double mutant Cre(R173A, R292A) also mediates this reaction. Presumably, the overall electrophilic character of the Cre active site is sufficient to neutralize the diminished negative charge present in the MeP, compared to the phosphate, transition state.

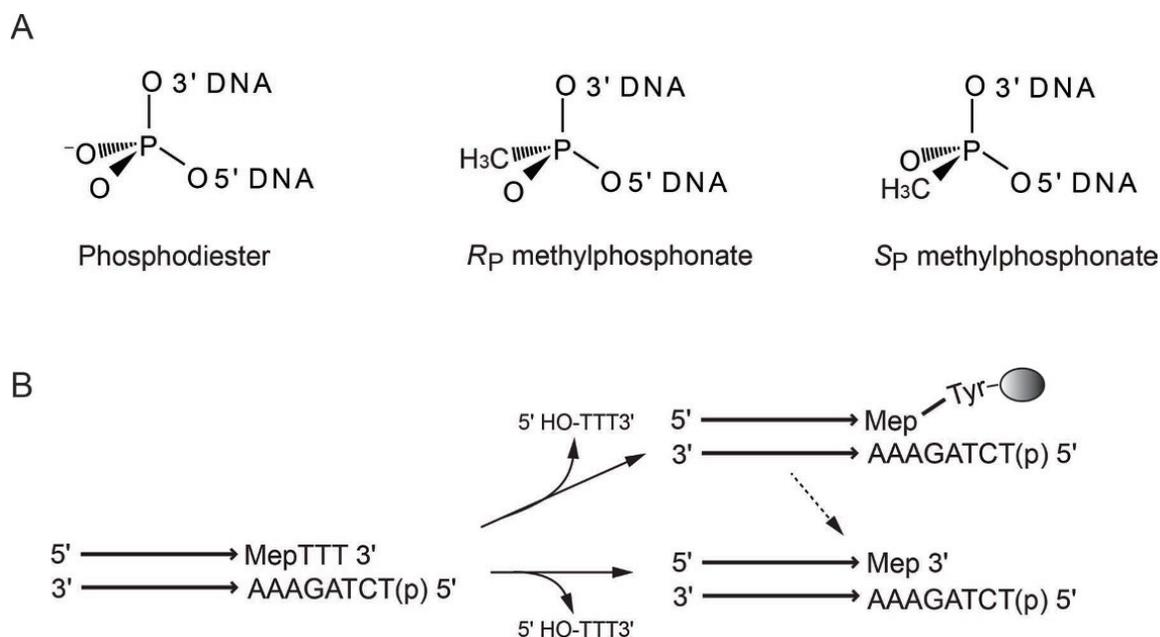


Figure 6 Reactions of half-sites containing methylphosphonate substitution at the scissile phosphate position. (A) The structures of methylphosphonate (MeP) are compared to that of the native phosphate in DNA. There are two possible stereoisomers of MeP (R_P or S_P). (B) The possible reactions of a half-site containing MeP at the scissile phosphate position are illustrated. The 5'-hydroxyl group on the bottom strand of the half-site is blocked by phosphorylation to prevent it from taking part in a pseudo-joining reaction. Attack of the MeP bond by the active site tyrosine will give the MeP-tyrosyl intermediate, which may undergo slow hydrolysis. The hydrolysis product may also be formed by direct water attack on the MeP bond. The two-step (type I) and single-step (type II) reaction pathways are mechanistically analogous to the type I and type II RNA cleavage activities of Flp (see text).
doi:10.1128/microbiolspec.MDNA3-0021-2014.f6

As the methyl substitution of one of the nonbridging oxygen atoms turns the normally symmetric phosphate group into an asymmetric center [Fig. 6(A)], an additional utility of the MeP substrates is in unveiling the stereochemical course of the recombination reaction. Stereochemically pure R_P and S_P forms of the MeP substrates are currently being used to dissect the individual stereochemical contributions of Arg-191 and Arg-308 in Flp, and to probe how other members of the catalytic hexad might influence these contributions.

DISTINCT ACTIVITIES OF Flp(R191A) AND Flp(R308A) IN THE MeP REACTION

The absence of Arg-191 or Arg-308 has strikingly different effects on the activity of Flp on an MeP-half-site [Fig. 7(A)] (68, 86). Flp(R308A) does not utilize the Tyr-343 nucleophile, but promotes direct hydrolysis of the MeP bond. Consistent with this mechanism, the double mutant Flp(R308A,Y343F) also yields the hydrolysis product with similar kinetics and V_{max} (maximal velocity) as Flp(R308A). Apparently, the lack of

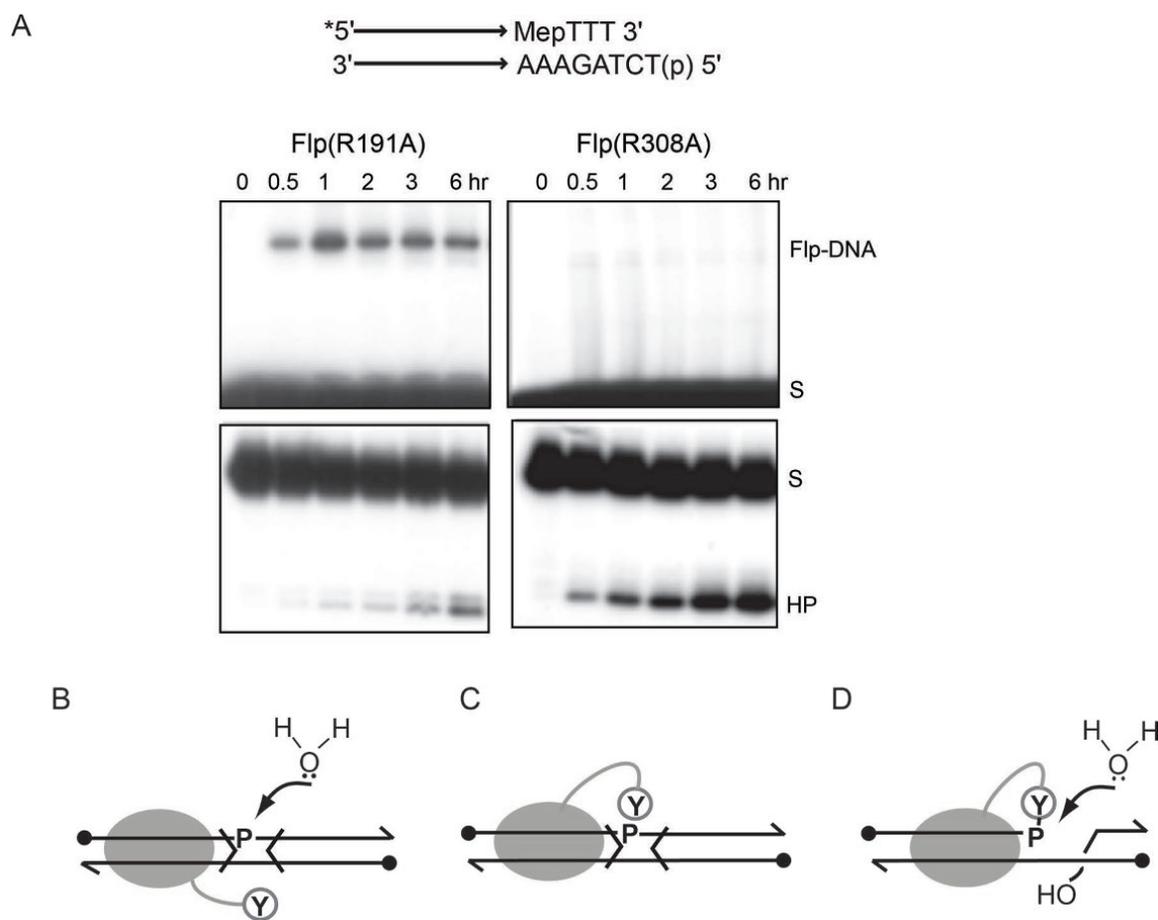


Figure 7 Distinct activities of Flp(R191A) and Flp(R308A) on an MeP-half-site. (A) Flp (R191A) cleaves the MeP-half-site (S) using Tyr-343 to form the protein–DNA adduct (revealed by SDS-PAGE; top) (86). This intermediate is converted to the hydrolysis product (HP) (revealed by denaturing PAGE; bottom) in a subsequent slow reaction. Flp(R308A), by contrast, yields the hydrolysis product directly, without going through the MeP-tyrosyl intermediate (68). (B) The binding of an Flp monomer to *FRT* activates the scissile phosphate, leaving it exposed until the binding of a second Flp monomer delivers Tyr-343 to the active site in *trans*. (C) Concomitant with the binding of a Cre monomer to the LoxP site, Tyr-324 engages the scissile phosphate in *cis*, thus protecting it against direct water attack. (D) As vaccinia topoisomerase, like Cre, assembles its active site in *cis*, the scissile phosphate is protected at the strand cleavage step during DNA relaxation. However, the protein's grip on DNA is loosened during the strand rotation step, leaving the phosphotyrosyl bond vulnerable to attack by water. doi:10.1128/microbiolspec.MDNA3-0021-2014.f7

Arg-308 permits the abundant water nucleophile to access the reaction center, where it out competes Tyr-343 to give a dead-end product. However, the possibility that Arg-308 is required for the positioning of Tyr-343 or its nucleophilic activation cannot be ruled out. The corresponding arginines, Arg-292 of Cre and Arg-410 of *Leishmania* topoisomerase I, are hydrogen bonded to the catalytic tyrosine in their respective vanadate-transition state structures (39, 89). Flp(R191A), by contrast, utilizes Tyr-343 as the cleavage nucleophile to yield the tyrosyl intermediate. Direct hydrolysis in the Flp(R191A) reaction is only a minor component. Cre (R173A) and Cre(R292A) are mechanistically similar to Flp(R191A) in that they promote Tyr-324-mediated cleavage of MeP (87, 88).

POTENTIAL ROLES FOR ACTIVE SITE AND PHOSPHATE ELECTROSTATICS IN PREVENTING FUTILE PHOSPHORYL TRANSFER

As suggested by the MeP reactions, in addition to balancing the phosphate negative charge, Arg-308 of Flp appears to protect the normal reaction course from abortive hydrolysis, perhaps by electrostatically misorienting water nucleophile (which is a dipole) from the activated phosphate. The phosphotyrosyl bond formed by vaccinia topoisomerase during DNA relaxation is apparently protected from hydrolysis by an analogous mechanism, utilizing the negative charge on the scissile phosphate (84). Furthermore, the potential role of the Arg-308 side-chain in orienting or activating Tyr-343 (see above under “Distinct activities of Flp(R191A) and Flp(R308A) in the MeP reaction”) suggests an alternative or collaborative mechanism for preventing futile breakage of the DNA backbone by increasing the local concentration of the tyrosine nucleophile. As noted earlier, the need to shield the scissile phosphate from extraneous nucleophiles would be more critical for Flp because of its *trans* active site. The scissile phosphate, activated by the proactive site of a bound Flp monomer, stays exposed until Tyr-343 is provided in *trans* [Fig. 7(B)]. Binding by a Cre or topoisomerase monomer to DNA and the alignment of the tyrosine nucleophile with respect to the scissile phosphate would be nearly concomitant events because of their *cis* active sites [Fig. 7(C)]. In the case of the topoisomerase, which acts as a monomer, the strand rotation step may open the phosphotyrosyl bond to attack by water [Fig. 7(D)], which is prevented by phosphate electrostatics. Such a protective mechanism is likely unnecessary for the recombinases, as the tight organization of

the recombinase tetramer-DNA complex (Fig. 1) and the dynamics of strand exchange within it preclude water from accessing the phosphotyrosyl bond. The extrusion of the cleaved strand into the center of the “strand exchange cavity” seen in the structure of the Cre-recombination synapse (52) would be consistent with strand swap being nearly concomitant with strand cleavage.

TYROSINE RECOMBINATION STEP-BY-STEP FROM START TO FINISH: SINGLE MOLECULE ANALYSIS

Single molecule analysis of tyrosine recombination using real-time tethered particle motion (TPM), tethered fluorophore motion (TFM) and fluorescence energy transfer (FRET) have provided deeper insights into the prechemical, chemical, and conformational steps of the reaction pathway by revealing transient states as well as long- and short-range movements of DNA (90, 91, 92, 93). The results of these studies reveal interesting similarities and contrasts among Cre, Flp and λ Int. The kinetics of recombinase binding to target sites and the pairing of bound sites are quite fast in all three cases, ruling out intrinsic barriers to synapsis, at least *in vitro*. There is a strong commitment to recombination following the association of Flp with the *FRT* sites. The formation of nonproductive complexes (those that do not synapse) and wayward complexes (those that do not form the Holliday junction intermediate or complete recombination after synapsis) constitute only minor detractors from the productive pathway (91). The stability of the synapse is enhanced by strand cleavage in the case of Flp and λ Int. However, Cre forms stable synapse even in the absence of strand cleavage (90). Recombination by Flp is efficient, and the frequency of occurrence of the Holliday junction intermediate is quite low (91). λ Int exhibits a strong and early commitment to a directed reaction path, likely assisted by its accessory factors bound to their cognate sites (92). Unidirectionality of an initiated recombination event would be a desirable attribute *in vivo* in bringing about the desired DNA rearrangement, without reversing course midway through a reaction. The Holliday junction formed during Cre recombination, however, is long lived, thanks to a rate-limiting step that follows its isomerization (93). This kinetic barrier affords the opportunity for the reaction to be interrupted and to go backwards, at least *in vitro*. It is possible that the *in vitro* Cre reaction fails to recapitulate the native regulatory features of recombination occurring within the P1 phage genome organized into a nucleoprotein complex.

SINGLE MOLECULE TPM AND FRET ANALYSES AS PROBES FOR THE GEOMETRY OF SITE-PAIRING AND ORDER OF STRAND EXCHANGE

TPM analysis is based on the rationale that the Brownian motion (BM) amplitude of a small polystyrene bead (~200 nm in diameter) attached to one end of a DNA molecule, whose other end is held in place, will be determined by the length of the DNA (Fig. 8). Since individual steps of recombination (binding of recombinase to target sites and their bending, synapsis of two bound sites by DNA looping, formation of the Holliday junction, and excision of DNA between two head-to-tail sites) are accompanied by characteristic changes in DNA length, TPM is well suited for the stepwise analysis of recombination (90, 91, 92). The pre- and post-

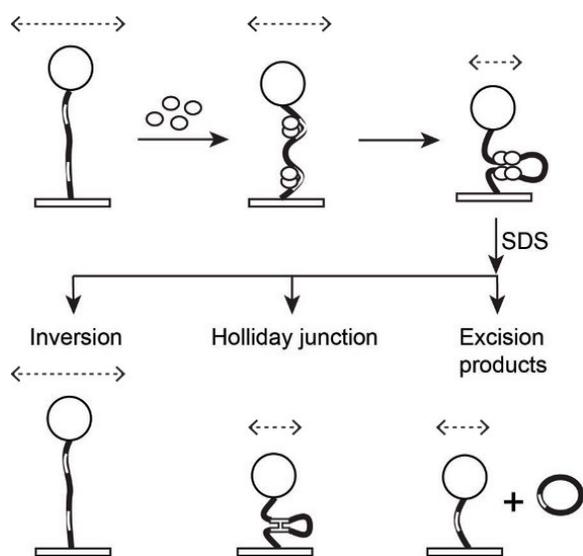


Figure 8 Stepwise analysis of recombination by TPM. The DNA molecule containing two recombination target sites (open boxes) in head-to-head or head-to-tail orientation is attached to a glass slide at one end and tethered to a polystyrene bead at the other. The change in DNA length occurring at individual steps of recombination is reported by the corresponding changes in the BM amplitude of the bead (schematically indicated by the dashed lines with arrowheads at either end). The bending of the sites bound by the recombinase (shown as globules) causes a shortening of DNA, which is magnified upon synapsis. Chemical steps of recombination within the synapse can result in Holliday junction formation or completion of recombination (DNA excision in the case of head-to-tail sites and DNA inversion in the case of head-to-head sites). Upon dissociation of the recombinase from DNA by SDS treatment, the Holliday junction intermediate and the linear excision product will retain their low BM amplitude. The inversion product has the same length, and hence the same BM amplitude, as the starting DNA molecule. doi:10.1128/microbiolspec.MDNA3-0021-2014.f8

chemical changes in DNA length can be distinguished by challenging the reaction with SDS. Upon protein dissociation, the prechemical states and completed inversion reactions (from head-to-head sites) will return to the length of the substrate DNA molecules. The Holliday intermediates from the inversion or the excision reaction and the linear product from the latter will retain their reduced “tether” lengths. By performing the excision and inversion reactions in parallel using DNA substrates identical in length and in the location and spacing of the recombination sites, a complete analysis of the reaction path is possible. TPM is also useful for verifying the geometry of a pair of target sites within the recombination synapse, as described below.

For sites in head-to-head orientation, the entry and exit points of the DNA will be at the same end of the synapse if the sites are aligned in a parallel fashion [Fig. 9(A)]. If the sites are in antiparallel alignment, the DNA will enter and exit the synapse from opposite ends. These situations will be reversed for a pair of sites in the head-to-tail orientation [Fig. 9(B)]. The proximal disposition of the entry and exit points imposes a stronger constraint on the DNA than their distal configuration, and makes it effectively shorter by a small amount. A significant difference in the BM amplitudes of two DNA molecules of identical length harboring a pair of equally spaced recombination sites, head-to-head in one case and head-to-tail in the other, would indicate preferential synapsis in one geometry. If the BM amplitude of the synapsed state is larger for the head-to-head sites, the preferred geometry is antiparallel. This is indeed the observed result for *FRT* sites synapsed by Flp (44). This conclusion is supported by single molecule FRET measurements in two *FRT* sites whose synapse geometry is restricted to being either parallel or antiparallel by a short single-stranded tether joining them (44) (Fig. 10). A change in the FRET state, in the expected direction, upon binding of Flp(Y343F) is observed only for the pair of *FRT* sites constrained in the antiparallel sense.

Given the approximate 2-fold symmetry of the core recombination sites, one might have imagined that they would synapse in a parallel or antiparallel fashion, even if only one of the two arrangements was productive for recombination. Topological and FRET results argue for preferred antiparallel synapsis of *FRT* sites even in the absence of the chemical steps of recombination (41, 44). Perhaps an asymmetric DNA bend within the strand exchange region of an Flp bound *FRT* site may preclude two similarly bent sites from occupying the synapse in a parallel fashion. A sharp bend located at a single bp step at one end of the strand exchange region

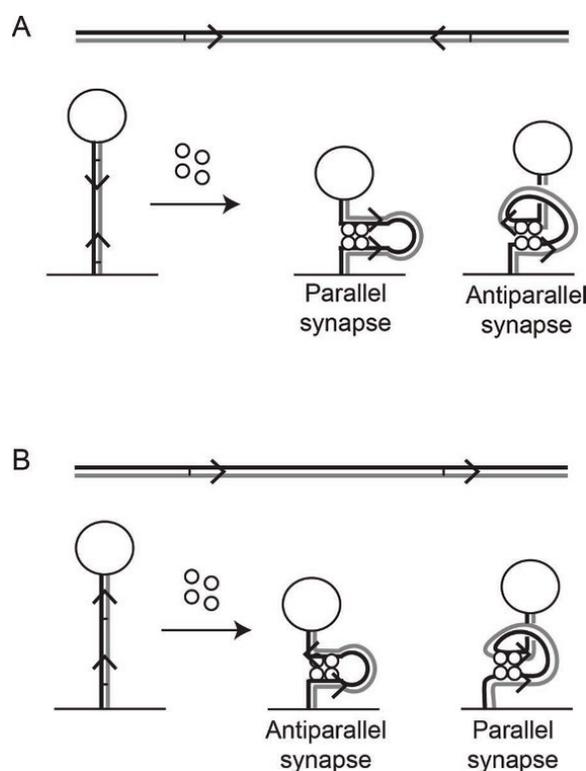


Figure 9 Effect of synapse geometry on the BM amplitude of DNA. (A) The DNA contours for a pair of synapsed head-to-head sites are outlined for their alignment in parallel (left) or antiparallel (right) geometry. (B) Similar diagrams as in (A) represent the antiparallel (left) and parallel (right) synaptic configurations for head-to-tail sites. The effective length of DNA is slightly larger when its entry and exit points are at opposite ends of the synapse than when they are at the same end. For two DNA substrates that differ only in the relative orientation of the recombination sites, a difference in the BM amplitudes of synapsed head-to-head versus head-to-tail sites signifies a preferred geometry of the synapse.
doi:10.1128/microbiolspec.MDNA3-0021-2014.f9

has been observed in the structures of *LoxP* complexed with cleavage-incompetent mutants of *Cre* (94).

The synapse geometry raises the question of “order of strand exchange” during recombination. Depending on the location of the asymmetric bend with respect to the scissile phosphate, there are two geometrically equivalent and chemically competent configurations of the antiparallel synapse. One would correspond to “top strand” cleavage, and the other to “bottom strand” cleavage [Fig. 11(A)]. If one of the two synaptic configurations is preferred, the order of strand cleavage/exchange will reflect this preference. Current evidence suggests that *Flp* performs strand exchange without obvious bias (95, 96), indicating that the two modes of antiparallel synapsis are equally likely. *Cre*, by contrast,

performs ordered strand exchange. FRET analysis with donor–acceptor dye pairs suitably positioned with respect to the strand exchange region demonstrates a preferred synapse, the DNA bend within which is consistent with the biochemically mapped preference in strand cleavage [Fig. 11(B), (C)] (48).

Ordered strand exchange is the norm in the λ *Int* and *XerCD* systems as well. The constraints imposed by high-order protein assemblies and DNA topology on the synapsis of the *Int*-bound core sites can dictate which pair of scissile phosphates is primed for initial cleavage (31, 70, 97). In the case of *XerCD*, depending on the reaction context, cleavage susceptibility may be determined by the synapse topology organized by accessory factors or may be altered by the presence or absence of an interacting regulatory protein (98, 99).

TOPOLOGICAL AND CHIRAL FEATURES OF TYROSINE RECOMBINATION

Tyrosine recombinases (*Cre*, *Flp* and λ *Int*) in general assemble the synapse by random collision of their target sites (see chapter by M. Boocock). In the case of *Int*, this randomness appears to be superposed over an intrinsic topological specificity (see below). As noted in discussing the role of homology in *Flp* recombination, it is the interdomainal crossings trapped during synapsis (blue \times red in Fig. 5) that appear as knot or catenane crossings in the recombination products. As pointed out earlier, the inversion reaction results in an unknotted circle (with the blue domain inverted with respect to the red) together with a range of increasingly complex knots; the deletion reaction produces unlinked circles as well as a range of increasingly complex catenanes.

The topology of *Cre* recombination is sensitive to reaction conditions. Relatively high pH tends to increase the complexity of the products, while lower pH has the opposite effect (43). Computer simulations, combined with DNA cyclization assays, suggest that the topological difference between *Cre* and *Flp* can be accounted for by the difference in the presynaptic bends that they induce in their target sites ($\sim 35^\circ$ for *LoxP* and $\sim 78^\circ$ for *FRT*) (100). The larger bend tends to localize two presynaptic *FRT* sites within separate branches of a plectonemically supercoiled circle [Fig. 12(A)], while the smaller bend tends to place two presynaptic *LoxP* sites in the same branch [Fig. 12(B)]. Interbranch recombination results in topologically complex products; intrabranched recombination gives simple products. Thus, protein-induced local changes in the statistical properties of large DNA molecules can strongly influence

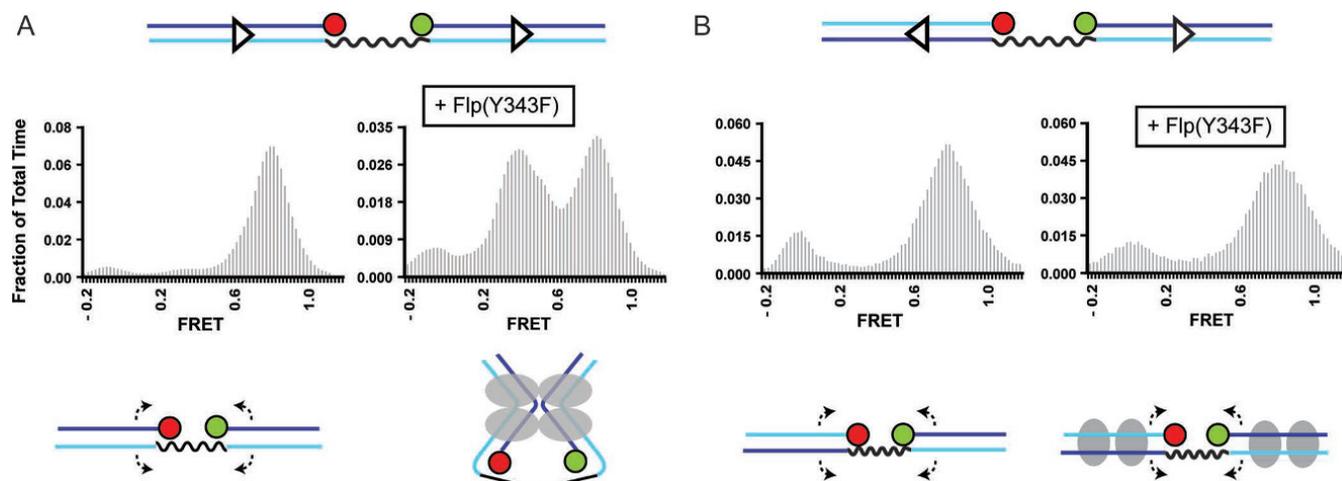


Figure 10 Preferred antiparallel synopsis of a pair of tethered *FRT* sites. (A) The two *FRT* sites, whose orientation is indicated by the arrowheads, are constrained by a single-stranded tether (wavy line) to align only in the antiparallel geometry. The positions of the donor (Cy3) and acceptor (Cy5) fluorophores are indicated by the green and red circles, respectively. The shift towards lower FRET upon Flp(Y343F) binding is consistent with the synopsis of the *FRT* sites as schematically diagrammed (44). (B) In the tethered DNA substrate, analogous to that diagrammed in (A), the *FRT* sites are constrained to pair only in the parallel geometry. Flp(Y343F) binding produces no change in FRET, suggesting the absence of parallel synopsis. doi:10.1128/microbiolspec.MDNA3-0021-2014.f10

their global topology, and dictate the outcomes of the reactions they partake in.

The topological outcomes of XerCD recombination are dictated by the contexts in which the reaction occurs. Recombination between *dif* sites (utilized for the resolution of *E. coli* chromosome dimers) requires the ATP driven DNA translocase FtsK (99). The reaction yields topologically simple products from negatively supercoiled substrates. When *dif*-recombination is activated by the carboxyl-terminal γ -domain of FtsK, which lacks the ATPase function, the products are topologically complex. The topology of recombination between two *cer* sites or two *psi* sites (utilized for the resolution of plasmid dimers) is dictated by accessory protein factors (101). The reaction normally occurs between sites in head-to-tail orientation, and requires negative supercoiling. The unique right-handed 4-crossing catenane produced from *psi* \times *psi* recombination, as well as the structure of the Holliday junction formed by *cer* \times *cer* strand exchange, conforms to a three-crossing synapse topology. According to tangle analysis, the recombination synapse fits a unique three-dimensional model, with three solutions that correspond to three distinct views obtained by rigid body movements of the synapse and projection on to a planar surface (102).

The FtsK-dependent topology simplification by XerCD recombination is also manifested in the unlinking of cate-

nanas harboring *dif* sites, either in the parallel or antiparallel sense (103). Catenanes with parallel *dif* sites are topologically analogous to catenanes resulting from the replication of circular plasmids and chromosomes, which are unlinked by the type II topoisomerase Topo IV (104). Rather surprisingly, XerCD-FtsK can support the resolution of chromosome catenanes *in vivo* when Topo IV activity is compromised (105). The recombination mechanism would suggest that unlinking by XerCD-FtsK proceeds by removing one crossing at a time (Fig. 13). This intuitive model, based on product distributions observed in *in vitro* reactions with plasmid substrates, has been validated mathematically by a combination of tangle analysis and knot theory under the assumption that each recombination event reduces the topological complexity of the substrate (106, 107).

An intriguing aspect of tyrosine recombination, brought to light primarily from the analysis of λ Int reactions is the apparent chirality of the reaction (108). The chirality of knots and catenanes formed from inversion and deletion reactions, respectively, in negatively supercoiled substrates follows from the right-handed chirality of plectonemic negative supercoils. However, quite unexpectedly, even reactions of nicked substrates turn out to be chiral. In the reactions between attP and attB sites in nicked substrates, two inter-domainal right-handed crossings are trapped by the DNA Inter-

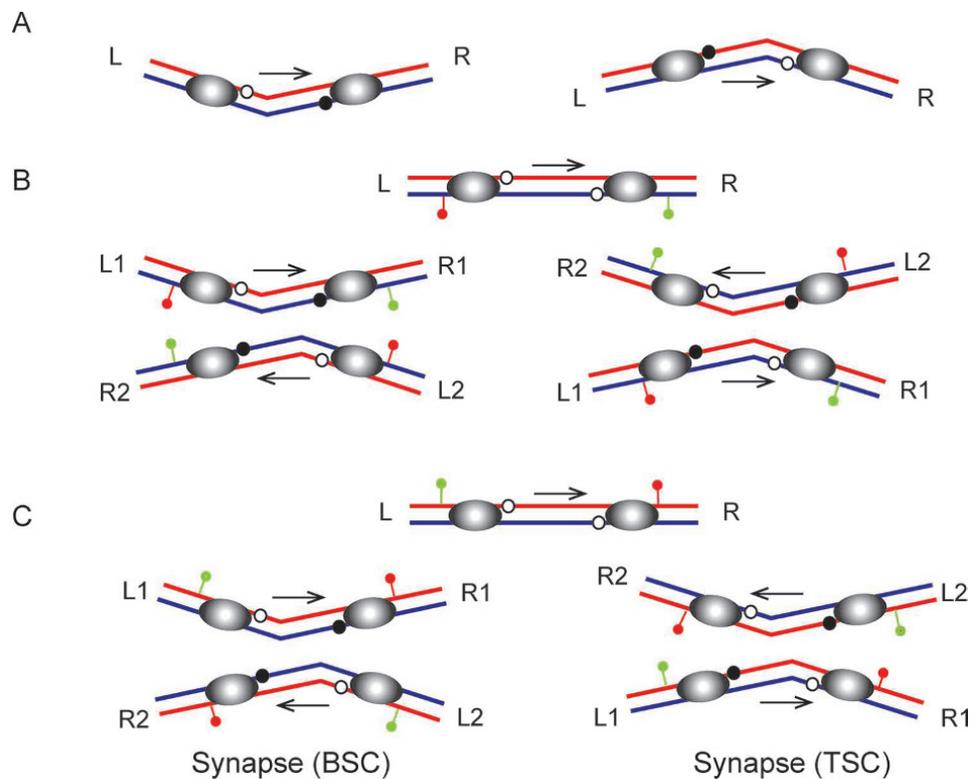


Figure 11 The preferred assembly of one of two possible types of antiparallel synapse can specify the order in which strands are cleaved and exchanged during recombination. (A) A LoxP site bound by Cre is bent asymmetrically, the bend center being located close to one end of the strand exchange region. The two possible asymmetric bends would specify the cleavage of the bottom (blue) or the top strand (red). The scissile phosphates primed for cleavage are indicated by the filled circles; the quiescent ones are shown as open circles. For convenience of orienting the sites, the DNA arms are labeled as L (left) and R (right) as in Fig. 1. (B) Based on the structure of the Cre-LoxP complex, fluorophores can be so positioned as to minimize donor (green)–acceptor (red) distance, and induce efficient FRET when the synapse favoring bottom strand cleavage (shown at the left) is assembled by Cre. In this fluorophore configuration, the FRET efficiency will be low for the synapse favoring top strand cleavage (shown at the right). (C) By reversing the left–right orientation of the fluorophores with respect to the strand exchange region, while maintaining their relative positioning, the synapsis favoring top strand cleavage (right) can be made to acquire the high FRET state. Experimental results indicate a clear preference for the synapse shown at the left in (B) suggesting that recombination is initiated by bottom strand cleavage and exchange (48). doi:10.1128/microbiolspec.MDNA3-0021-2014.f11

actions of Int and its accessory factors; in the reactions between attL and attR sites, the corresponding number is one. It would take only a single additional right handed DNA crossing randomly trapped in the substrate to generate a chiral three-noded knot as the product of an attP–attB inversion reaction. During attP–attB reactions in negatively supercoiled molecules, Int (in conjunction with accessory factors) traps three right handed DNA crossings, one more than that deduced for the same reactions in nicked circular molecules. However, for attL–attR reactions, this number is still

one, unchanged between nicked and negatively supercoiled substrates.

The topological and chiral features of a recombination reaction are conveniently and succinctly summarized by tangle diagrams such as those illustrated in Fig. 14. A tangle may be perceived as a three-dimensional ball within which strings representing double stranded DNA may cross in a variety of ways. In the two dimensional projection of a tangle, the entry and exit points of DNA are placed at the NE, NW, SE and SW corners (in a geographical sense). The O_b and

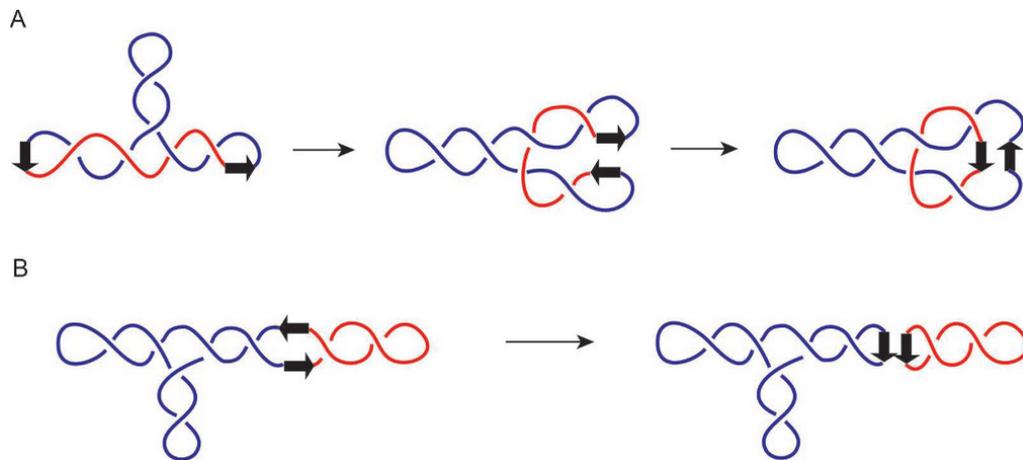


Figure 12 The magnitude of the DNA bend at the recombination target sites influences their localization within the branches of pleconemically supercoiled DNA. (A) The large DNA bend induced by Flp tends to localize presynaptic *FRT* sites in separate pleconemic branches. Recombination between such sites yields topologically complex products. In the example shown, the excision reaction yields a 4-noded catenane. (B) The relatively small DNA bend induced by Cre tends to place presynaptic *LoxP* sites within the same pleconemic branch, thus simplifying the topology of recombination products. The excision reaction shown here yields two unlinked deletion circles.
doi:10.1128/microbiolspec.MDNA3-0021-2014.f12

O_f tangles harbor DNA constrained by protein binding and ‘free’ DNA, respectively. The P and R tangles contain the DNA segments that engage in crossover in their parental and recombined states, respectively. O_b for attL-attR reactions, containing one right-handed crossing, is a +1 tangle for nicked as well as negatively supercoiled substrates (as shown in Fig. 14 A, B). For attP-attB reactions in nicked substrates, the O_b tangle is +2; for the same reactions in negatively supercoiled substrates, the O_b tangle is +3.

Most surprisingly, recombination reactions from nicked substrates by Cre and Flp also appear to be chiral, trapping one right-handed inter-domainal crossing in the synapse. This crossing is proposed to predispose the reaction towards a chiral product via a right-handed Holliday junction intermediate. The near perfect planarity of the DNA arms in the crystal structures of Cre and Flp (34, 52) challenges this postulate. Nevertheless, the slight out-of-plane disposition of the Holliday junction arms in the Flp crystal structure is consistent with the proposed right-handed chirality.

An irksome aspect of chirality is the difficulty in accommodating the experimental observation that the linking number change (ΔLk) associated with Flp- or λ Int-mediated inversion reactions between *FRT* sites and attL-attR sites, respectively, is either +2 or -2 (41, 109), and the two outcomes are equally likely for a

nearly perfectly relaxed substrate. The right-handed chirality would predict a ΔLk of exclusively +2. For example, the right-handed crossing (a - node) trapped by Int would change its sign (a + node) as a result of DNA inversion ($\Delta Lk = +1 - (-1) = +2$). The tangle diagram depicting this change in the node sign in O_b is shown in Fig. 14A. The two suggested tangle solutions to resolve this paradox are shown in Fig. 14B, C. In Fig. 14B, the substrate DNA enclosed by the O_f tangle harbors two + crossings, one to compensate for the - crossing trapped by Int (in the O_b tangle) and an additional one to arrange the recombination sites with the antiparallel geometry in the P (parental) tangle. The inversion of each of these crossings would give $\Delta Lk = -2$, $[(+1 - (-1)) + [(-2) - (+2)]$. The problem, though, is that the energetic cost of introducing additional O_f crossings should make $\Delta Lk = -2$ less likely than $\Delta Lk = +2$, in violation of the experimental result. In Fig. 14C, the P tangle is switched from an ∞ tangle to a 0 tangle, so as to preserve the tangle notation in O_b , and still produce a $\Delta Lk = -2$, $[-1 - (+1)]$. This is also unsatisfactory, as it accommodates $\Delta Lk = -2$ by a sleight of hand(edness). If one follows the contour of the DNA circle, it is obvious that the crossing in O_b is left-handed, not right-handed. Chirality of tyrosine recombination and the ΔLk paradox arising from it remain an enigmatic curiosity that calls for further exploration.

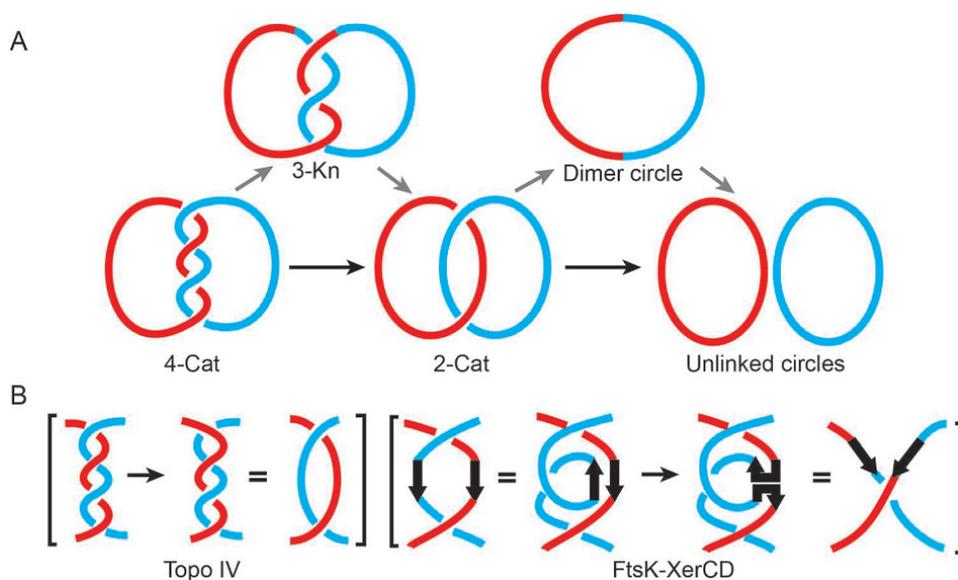


Figure 13 Unlinking of replication catenanes by XerCD-FtsK. (A) The unlinking of replication catenanes in *E. coli* is normally carried out by the type II topoisomerase Topo IV. For a 4-noded replication catenane containing parallel *dif* sites, unlinking by Topo IV will be completed in two steps (the straight path), removing two crossings at each step. Unlinking of the same catenane by FtsK-XerCD-mediated recombination at the *dif* sites requires four steps (the zigzag path), by removal of one crossing at a time. (B) The mechanisms for topology simplification by Topo IV and FtsK-XerCD are illustrated. doi:10.1128/microbiolspec.MDNA3-0021-2014.f13

CONTRIBUTIONS OF YRs TO BASIC BIOLOGY AND BIOENGINEERING APPLICATIONS

Biochemical, biophysical, and structural studies of YRs have been seminal to unveiling the mechanisms of an important class of phosphoryl transfer reactions in nucleic acids and to understanding conformational dynamics associated with strand exchange between two DNA partners (20, 21, 23, 90, 91, 92, 93). The simple requirements of Cre and Flp have been exploited to carry out specific genetic rearrangements in bacteria, fungi, plants, and animals. By combining the DNA delivery properties of mobile group II introns in bacteria and the DNA exchange potential of tyrosine recombination, a new platform for genome editing via targetrons and recombinases (GETR) has been developed (110). In general, prokaryotic and eukaryotic cells engineered to express a recombinase and harboring its target sites in the genome or housed in an extrachromosomal vector carry out the expected reaction with high efficiency. Directed insertion of a desired foreign DNA into a genome as well as inversions, deletions, or translocations of selected genomic segments can thus be accomplished with reasonable ease. These manipulations have been

particularly helpful in addressing fundamental problems in cell and developmental biology. The utilization of controlled and efficient site-specific recombination between homologous chromosomes to generate mosaic flies has provided a technical breakthrough for tracking cell lineages in *Drosophila* (111). Analogous strategies, in conjunction with multicolored reporter genes and live-cell imaging, have expanded the power and range of lineage tracking to higher organisms and facilitated its integration with the monitoring of intracellular signaling pathways (112). Methodologies for tissue-, cell type- and stage-specific induction of recombination activity make it possible to analyze spatial and temporal controls of developmental programs in intricate detail (113, 114). Another, perhaps less widely publicized, utility of Cre and Flp in basic biology is exemplified by “difference topology,” an analytical method for tracing the topological path of DNA within high-order DNA-protein complexes (115, 116). Finally, Cre, Flp, and to a limited extent, λ Int have been put to practical use in a number of biotechnology-related applications. A brief description of the principles and practice of difference topology and of the potential impact of site-specific recombination on biotechnology is given below.

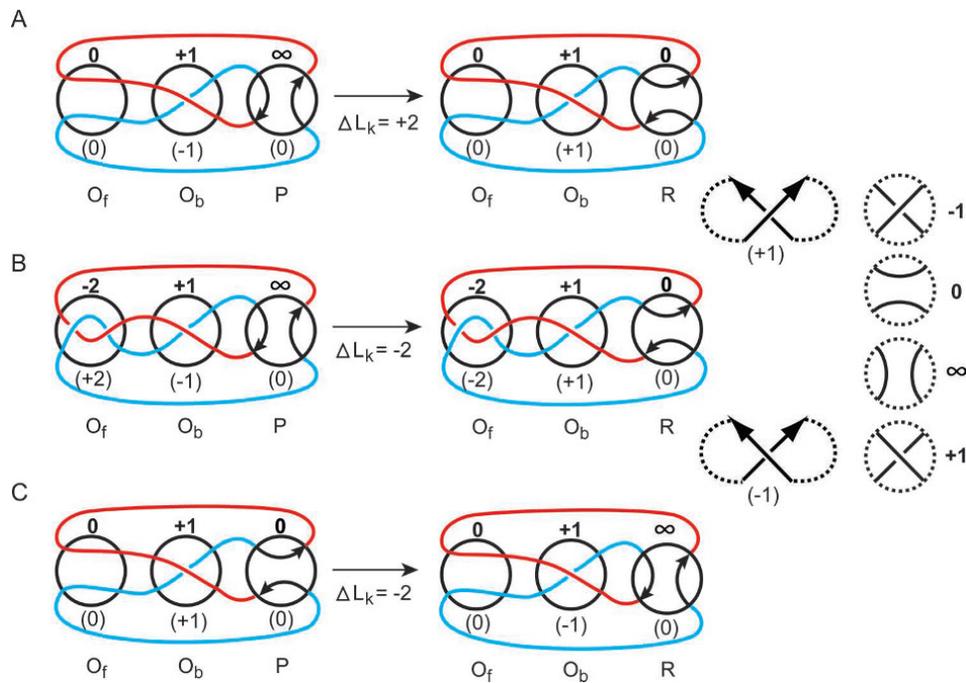


Figure 14 Tangle diagrams of attL- attR recombination performed by λ Int; ΔLk associated with DNA inversion. The λ Int mediated inversion reaction between attL and attR sites in three relaxed circular substrate molecules is represented by tangle diagrams (A-C). The O_b tangle contains inter-domainal DNA crossings trapped by Int (likely assisted by the accessory factors). O_f contains randomly trapped crossings in the ‘free’ DNA. The core recombination sites reside in the P tangle in anti-parallel geometry. The R tangle represents the post-recombination state of the sites. The tangle notations are shown at the top in bold; the corresponding DNA crossing (node) signs are given at the bottom in parentheses. The convention for the crossing signs (+1 or -1) is illustrated at the right, with the arrow heads denoting the direction (arbitrarily assigned) for the circular DNA axis. The simplest tangles (0, +1, -1, ∞) are diagrammed at the far right. **A.** In the DNA molecule shown here, one right-handed crossing is trapped in O_b , and none are contained in O_f . In tangle parlance, recombination changes $P(\infty)$ tangle to the $R(0)$ tangle, yielding an unknotted inversion circle. Note that a right-handed crossing in O_b in the substrate is +1 by the tangle convention, but -1 by the sign convention. In the recombinant product, the crossing sign in O_b becomes +1 because of DNA inversion. The linking number change (ΔLk) accompanying the attL-attR reaction is +2. **B.** The same reaction as in **B** is shown for a molecule with two left-handed crossings present in O_f . The ΔLk for the reaction is -2. **C.** A molecule performing the same reaction as in **A** and **B** is represented with $P(\infty)$ and $R(0)$ switched to $P(0)$ and $R(\infty)$, respectively. The ΔLk associated with recombination is -2 in this case as well. The ΔLk changes are explained in more detail in the text. doi:10.1128/microbiolspec.MDNA3-0021-2014.f14

DIFFERENCE TOPOLOGY: DECIPHERING DNA TOPOLOGY WITHIN DNA-PROTEIN MACHINES

The elimination of topological randomness from Cre and Flp reactions by assembling a unique synapse with the assistance of protein factors (40, 41, 42, 43) is the basis for the analytical tool called difference topology. The method is useful for determining the number of supercoils sequestered by two DNA sites when they functionally interact with each other. As we have seen already, when three negative supercoils are trapped

adjacent to Cre or Flp synapse, say by utilizing the resolvase synapse, the inversion and deletion reactions yield a 3-noded knot and a 4-noded catenane, respectively (Fig. 5). The inversion knot faithfully preserves the number of DNA crossings in the external synapse, as three (or an odd number) crossings would bring head-to-head *LoxP* or *FRT* sites in the antiparallel geometry that promotes recombination. A fourth crossing, easily provided by the negatively supercoiled substrate, is necessary for the deletion reaction, as it takes an even number of crossings to confer antiparallel

geometry on head-to-tail sites. By similar arguments, two external negative supercoil crossings would be revealed in the difference topology analysis as a 2-noded catenane for the deletion reaction and a 3-noded knot for the inversion reaction. Random entrapment of supercoils in the hybrid synapse can be avoided by suitable placement of the recombination sites with respect to the external synapse. The crossings in the inversion knot and the deletion catenane, analyzed by gel electrophoresis and electron microscopy, would thus accurately report on the DNA topology of the external synapse. A simplified description of the concepts and experimental applications of difference topology can be found in a recent review (2). Using this analysis, the topology of the interactions among the left and right ends of phage Mu and its transposition enhancer element within the transposition complex organized by the MuA protein has been mapped as a three-branched, five-crossing plectoneme (117) (see chapter by R. M. Harshey).

ENGINEERING OF EUKARYOTIC GENOMES USING YRS

Before the practical utilities of site-specific recombination came into prominence, genome manipulations in higher eukaryotes relied either on nonhomologous or homologous recombination. Nonhomologous recombination promotes the efficient integration of linear DNA molecules into most genomes, but does so randomly. Homologous recombination permits modification of genomic loci with high specificity, but suffers from very low efficiency. Site-specific recombinases circumvent the drawbacks of both inefficiency and promiscuity, but require the prior integration of their cognate target sites into the genome to be modified. Furthermore, targeting multiple loci within a genome is limited by the number of available recombinases with the desired properties. This problem can at least be partly circumvented by taking advantage of the homology rule that dictates successful recombination. Mutually incompatible, but individually functional, target sites may be designed by introducing nonhomologies within their strand exchange regions. An even better solution, at least in principle, is the directed evolution of recombinases with altered target specificities (118, 119, 120, 121). Among YRs, Cre and Flp have been, by far, the enzymes of choice for applications in biotechnology. Variants of λ Int that are not functionally limited by cofactor requirements (122) have so far been running a rather distant third. The integrase of the *Streptomyces* phage ϕ C31 and chimeras derived from an activated form of the Tn3 resolvase represent serine recombinases that have shown promise

as tools in applied genetics (123, 124). Zinc finger nucleases (ZFNs), transcription activator effector-like nucleases (TALENs) and clustered regulatory interspersed short repeats (CRISPR)-Cas based RNA guided nucleases have complemented and augmented site-specific recombinases in the bioengineer's arsenal for analyzing and reshaping genomes (125, 126, 127).

The optimal performance of a recombinase and the tight regulation of its activity in a given biological context often require amino acid substitutions in the native protein sequence and/or the introduction of allosteric control regions. For example, the preferred growth temperature of budding yeast ($\sim 30^\circ\text{C}$), in which Flp normally functions, differs from that of mammalian cells (37°C). A thermo-tolerant variant of Flp (Flpe), with higher activity at 37°C than Flp, was obtained by mutagenesis coupled to selection (128). Flpe harbors four amino acid changes from Flp: P2S, L33S, Y108N, and S294P. Flpe, which still underperforms Cre in mammalian cells, has been further improved by adding a nuclear localization signal and by 'humanizing' its codons (129). The new variant Flpo is comparable to Cre in its activity in mammalian cells. Continuous expression of a recombinase from a constitutively active promoter could be counterproductive because of the likelihood of the intended reaction being reversed and potential toxic effects arising from rare off-target recombination events. These impediments can be overcome or minimized by conditional recombinase expression from a regulatable promoter (Tet-on or Tet-off, for example) (130). Small molecule effectors are also useful for controlling recombination activity. The recombination potential of Flp or Cre fused to the ligand binding domain of the steroid hormone receptor is activated only in the presence of natural estrogens or synthetic estrogen receptor antagonists (131, 132, 133, 134).

Attempts to increase the target repertoire of a recombinase by generating variants with nonoverlapping specificities have been moderately successful. The basic strategy involves the screening of a large pool of the mutagenized recombinase gene for those that code for "shifted" or "switched" specificities. Simple bacterial genetic assays or more rapid high-throughput cell sorting screens, based on chromogenic reporter genes, have been effectively employed for identifying recombinases with the desired recombination potential (119, 121). Substrate designs that place the target sites and the recombinase genes in *cis* so as to link them by the act of recombination (substrate linked protein evolution; SLiPE) can accelerate screening by simple PCR-based protocols (135). In a distinct cell-free approach, *in vitro* compartmentalization (IVC) has been used to

obtain altered specificity variants of λ Int (120). The IVC method relies on compartments of an oil-in-water emulsion in which *in vitro* expressed Int variants and the target sites are encapsulated.

In the Flp system, the strategy of mutagenesis followed by a bacterial dual reporter screen yielded recombination capability initially towards *FRT* sites containing single point mutations in the Flp binding element and subsequently towards sites containing combinations of these mutations (121, 136). It was further demonstrated that hybrid *FRT* sites, harboring distinct specificities in their two binding elements, can be recombined by a binary combination of Flp variants, each with the appropriate monospecific recognition potential (118). A step-wise directed evolution scheme with intermediate DNA shuffling steps is necessary to progressively coax Flp into accepting multiple changes within the *FRT* site. Consistent with the array of DNA contacts made by Flp, and the rather complex mode of substrate recognition, mutations involved in the acquisition of new specificities are distributed among amino acids that directly contact DNA as well as those that are located at monomer–monomer interfaces or in the proximity of catalytic residues.

Altered specificity variants of Cre have been evolved by structure-based substitution of base pairs recognized by Cre and randomization of selected amino acid positions in close proximity to them (119). Structural analysis of a subset of these variants suggests that two target sequences can be functionally recognized by a Cre variant through similar backbone contacts in conjunction with distinct base-specific contacts (137). These alternative modes of recognition are facilitated by a network of water-mediated contacts and an unexpected shift in the DNA backbone configuration. The contributions of water networks and macromolecular plasticity to DNA–protein interactions may thus complicate efforts to evolve new target specificities based on predictive schemes.

Directed evolution of recombinases that can act on naturally occurring sequences in their native biological context would signify a giant step forward in site-specific genome remodeling. Search algorithms such as Target Finder and TargetSiteAnalyzer have been developed to identify genomic sequences that match the size of a given recombination target site, and rank them according to the degree of their resemblance in organization and sequence to the chosen site (138, 139). There are >600,000 potential *FRT*-like sequences in the human genome (roughly one such sequence per 5,000 bp). Their distributions are inversely correlated to the average G/C content of individual chromosomes,

in agreement with the A/T richness of the *FRT* site. The highest density (one *FRT*-like sequence per ~4 kb) is in chromosomes 4 and 13 with an average G/C content of 38%, and the lowest (one *FRT*-like sequence per ~8 kb) is in chromosomes 19 and 22 with an average G/C content of 48%. The majority of duplicate *FRT*-like sequences are located in the copies of LINE1, while others form part of the LTRs (long terminal repeats) of endogenous retroviruses, Alu repeats and other repetitive DNA sequences. The potential genomic target sites located by search algorithms not only facilitate the manipulation of genetic loci of interest but also promote stringent specificity by providing sequences for counterselection during the steps of directed evolution of novel specificity recombinases.

Once an appropriately placed “high-ranking” site has been identified, the procedures of progressive directed evolution of the recombinase, aided by structural information in at least some of the cases, can be employed to turn it into an authentic recombination target. This strategy has produced Flp variants that utilize an *FRT*-like sequence located upstream of the human IL-10 gene and analogous sequences found in the human β globin locus (138, 139). Members of the latter set of sites perform well in mammalian cells when they are present on episomal vectors (139). However, further optimization of specificity, recombinase expression, and activity will be required before the system operates efficiently in the native chromosomal context. The step-wise evolutionary approach has proven to be powerful enough to yield a Cre variant capable of deleting a proviral DNA of HIV-1 pseudotype by recombination between LoxP-like sequences located in the LTRs (140).

An important and frequently employed genome engineering reaction is recombinase-mediated cassette exchange (RMCE), a replacement reaction that exchanges two DNA fragments by a double recombination event between sites flanking them at either end (141, 142) (Fig. 15). The two sites harbored by each DNA partner are designed to be incompatible (heterotypic) for intramolecular recombination but compatible (homotypic) for intermolecular recombination in one configuration of the partners. In its early formats, RMCE reactions utilized a single recombinase such as Cre or Flp to perform exchange at both DNA ends (141, 143) [Fig. 15 (A)]. In more recent versions of RMCE, referred to as dual RMCE (144), recombination at each end is mediated by a separate recombinase, Flp and Cre or Flp and the integrase of the λ related HK022 phage [Fig. 15(B)]. To obtain the best replacement results by dual RMCE, the relative expressions of the two recombinases have

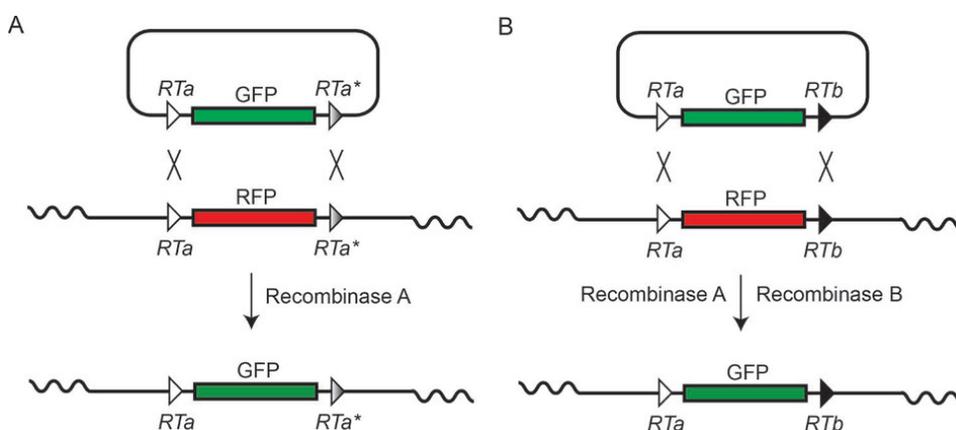


Figure 15 Recombination-mediated cassette exchange. (A) In the classical RMCE, the replacement of a native locus by a donor DNA fragment is mediated by the same recombinase acting on two pairs of target sites (RTa and RTa^*) that are compatible only in one configuration of the DNA partners. The reaction may be followed by the replacement of one fluorescent reporter (RFP; red fluorescent protein) by another (GFP; green fluorescent protein). (B) In dual RMCE, the same reaction as in (A) is mediated by two separate recombinases acting on their respective cognate sites (RTa and RTb).
 doi:10.1128/microbiolspec.MDNA3-0021-2014.f15

to be carefully optimized (145, 146). The Cre-Flp pair can yield RMCE in up to 35 to 45% of the transfected cells, while the corresponding yield for the Flp-HK022 Int pair is ~12%.

The Cre-Flp based dual RMCE system is being successfully employed by several mouse genome engineering programs for systematically knocking out protein coding regions, expressing reporter cassettes from cellular promoters, and for amino-terminal protein tagging of gene trap clones *in situ*. The organizations leading these efforts are the International Mouse Phenotyping Consortium (IMPC; www.mousephenotype.org), the European Conditional Mouse Mutagenesis Program (EUCOMM; www.eucomm.org), the KnockOut Mouse Project (KOMP; www.knockoutmouse.org) and the German Gene Trap Consortium (www.genetrapp.de). The beneficiaries from these endeavors will be high-throughput genomics and proteomics related to molecular medicine.

EPILOGUE

The intellectual seed for the advances in our understanding of site-specific recombination was sown more than fifty years ago by a simple and elegant model proposed by Allan Campbell for the integration of the phage λ genome into the *E. coli* chromosome. Over these five decades, the study of recombination has been transformed from a geneticists' sanctuary to the playing

fields of biochemists and to the roaming grounds of crystallographers and biophysicists. Their collective contributions have unveiled the chemical simplicity, mechanistic elegance, and structural sophistication of the reaction. Genome engineers, biotechnologists, and system biologists have now almost completely taken over the field and seem poised to lead it in new directions.

Acknowledgments. I thank Phoebe Rice, Ian Grainge, F.-X. Barre and Jeff Gardner for helpful criticisms and suggestions. The work in the Jayaram laboratory on the mechanism of site-specific recombination has been supported by the National Institutes of Health, the National Science Foundation, the Human Frontiers in Science Program, and the Texas Higher Education Coordinating Board. Funding for our recent work on the DNA topology and single molecule analysis of recombination has been provided by the National Science Foundation (MCB-1049925) and the Robert F Welch Foundation (F-1274).

Citation. Jayaram M, MA C-H, Kachroo AH, Rowley PA, Guga P, Fan H-F, Voznyanov Y. 2014. An overview of tyrosine site-specific recombination: from an flp perspective. *Microbiol Spectrum* 3(1):MDNA3-0021-2014.

References

1. Blaisonneau J, Sor F, Cheret G, Yarrow D, Fukuhara H. 1997. A circular plasmid from the yeast *Torulaspora delbrueckii*. *Plasmid* 38:202–209.
2. Chang KM, Liu YT, Ma CH, Jayaram M, Sau S. 2013. The 2 micron plasmid of *Saccharomyces cerevisiae*: a miniaturized selfish genome with optimized functional competence. *Plasmid* 70:2–17.

3. Piednoel M, Goncalves IR, Higuete D, Bonnivard E. 2011. Eukaryote DIRS1-like retrotransposons: an overview. *BMC Genomics* 12:621.
4. Poulter RT, Goodwin TJ. 2005. DIRS-1 and the other tyrosine recombinase retrotransposons. *Cytogenet Genome Res* 110:575–588.
5. Cortez D, Quevillon-Cheruel S, Gribaldo S, Desnoves N, Sezonov G, Forterre P, Serre MC. 2010. Evidence for a Xer/dif system for chromosome resolution in archaea. *PLoS Genet* 6:e1001166.
6. Serre MC, El Arnaout T, Brooks MA, Durand D, Lisboa J, Lazar N, Raynal B, van Tilbeurgh H, Quevillon-Cheruel S. 2013. The carboxy-terminal α N helix of the archaeal XerA tyrosine recombinase is a molecular switch to control site-specific recombination. *PLoS One* 8:e63010.
7. Van Houdt R, Leplae R, Mergeay M, Toussaint A, Lima-Mendez G. 2012. Towards a more accurate annotation of tyrosine-based site-specific recombinases in bacterial genomes. *Mobile DNA* 3:6.
8. Austin S, Ziese M, Sternberg N. 1981. A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell* 25:729–736.
9. Azaro MA, Landy A. 2002. λ integrase and the λ int family, p 118–148. In Craig NL, Craigie R, Gellert M, Lambowitz AM (ed), *Mobile DNA II*, ASM Press, Washington, DC.
10. Barre F-X, Sherratt DJ. 2002. Xer site-specific recombination: promoting chromosome segregation, p 149–161. In Craig NL, Craigie R, Gellert M, Lambowitz AM (ed), *Mobile DNA II*, ASM Press, Washington DC.
11. Chaconas G, Kobryn K. 2010. Structure, function, and evolution of linear replicons in *Borrelia*. *Annu Rev Microbiol* 64:185–202.
12. Ravin NV. 2011. N15: the linear phage-plasmid. *Plasmid* 65:102–109.
13. Das B, Martinez E, Midonet C, Barre FX. 2013. Integrative mobile elements exploiting Xer recombination. *Trends Microbiol* 21:23–30.
14. Fitcher AB. 1986. Copy number amplification of the 2 micron circle plasmid of *Saccharomyces cerevisiae*. *J Theor Biol* 119:197–204.
15. Genka H, Nagata Y, Tsuda M. 2002. Site-specific recombination system encoded by toluene catabolic transposon Tn4651. *J Bacteriol* 184:4757–4766.
16. Rajeev L, Malanowska K, Gardner JF. 2009. Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol Mol Biol Rev* 73:300–309.
17. Volkert FC, Broach JR. 1986. Site-specific recombination promotes plasmid amplification in yeast. *Cell* 46:541–550.
18. Turan S, Zehe C, Kuehle J, Qiao J, Bode J. 2013. Recombinase-mediated cassette exchange (RMCE)- a rapidly-expanding toolbox for targeted genomic modifications. *Gene* 515:1–27.
19. Chen Y, Rice PA. 2003. New insight into site-specific recombination from Flp recombinase-DNA structures. *Annu Rev Biophys Biomol Struct* 32:135–159.
20. Grindley ND, Whiteson KL, Rice PA. 2006. Mechanisms of site-specific recombination. *Annu Rev Biochem* 75:567–605.
21. Radman-Livaja M, Biswas T, Ellenberger T, Landy A, Aihara H. 2006. DNA arms do the legwork to ensure the directionality of λ site-specific recombination. *Curr Opin Struct Biol* 16:42–50.
22. Van Duyne GD. 2005. λ integrase: armed for recombination. *Curr Biol* 15:R658–660.
23. Van Duyne GD. 2002. A structural view of tyrosine recombinase site-specific recombination, p 93–117. In Craig NL, Craigie R, Gellert M, Lambowitz AM (ed), *Mobile DNA II*, ASM Press, Washington, DC.
24. Whiteson KL, Rice PA. 2008. Binding and catalytic contributions to site recognition by Flp recombinase. *J Biol Chem* 283:11414–11423.
25. Boldt JL, Pinilla C, Segall AM. 2004. Reversible inhibitors of λ integrase-mediated recombination efficiently trap Holliday junction intermediates and form the basis of a novel assay for junction resolution. *J Biol Chem* 279:3472–3483.
26. Cassell G, Klemm M, Pinilla C, Segall A. 2000. Dissection of bacteriophage λ site-specific recombination using synthetic peptide combinatorial libraries. *J Mol Biol* 299:1193–1202.
27. Ghosh K, Lau CK, Guo F, Segall AM, Van Duyne GD. 2005. Peptide trapping of the Holliday junction intermediate in Cre-loxP site-specific recombination. *J Biol Chem* 280:8290–8299.
28. Kepple KV, Boldt JL, Segall AM. 2005. Holliday junction-binding peptides inhibit distinct junction-processing enzymes. *Proc Natl Acad Sci U S A* 102:6867–6872.
29. Kepple KV, Patel N, Salamon P, Segall AM. 2008. Interactions between branched DNAs and peptide inhibitors of DNA repair. *Nucleic Acids Res* 36:5319–5334.
30. Su LY, Willner DL, Segall AM. 2010. An antimicrobial peptide that targets DNA repair intermediates in vitro inhibits *Salmonella* growth within murine macrophages. *Antimicrob Agents Chemother* 54:1888–1899.
31. Biswas T, Aihara H, Radman-Livaja M, Filman D, Landy A, Ellenberger T. 2005. A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* 435:1059–1066.
32. Van Duyne GD. 2009. Teaching Cre to follow directions. *Proc Natl Acad Sci U S A* 106:4–5.
33. Warren D, Laxmikanthan G, Landy A. 2008. A chimeric Cre recombinase with regulated directionality. *Proc Natl Acad Sci U S A* 105:18278–18283.
34. Chen Y, Narendra U, Iype LE, Cox MM, Rice PA. 2000. Crystal structure of a Flp recombinase-Holliday junction complex. Assembly of an active oligomer by helix swapping. *Mol Cell* 6:885–897.
35. Lee J, Tonzuka T, Jayaram M. 1997. Mechanism of active site exclusion in a site-specific recombinase: role of the DNA substrate in conferring half-of-the-sites activity. *Genes Dev* 11:3061–3071.
36. Conway AB, Chen Y, Rice PA. 2003. Structural plasticity of the Flp-Holliday junction complex. *J Mol Biol* 326:425–434.

37. Lee J, Whang I, Jayaram M. 1996. Assembly and orientation of Flp recombinase active sites on two-, three- and four-armed DNA substrates: implications for a recombination mechanism. *J Mol Biol* 257:532–549.
38. Woods KC, Martin SS, Chu VC, Baldwin EP. 2001. Quasi-equivalence in site-specific recombinase structure and function: crystal structure and activity of trimeric Cre recombinase bound to a three-way Lox DNA junction. *J Mol Biol* 313:49–69.
39. Gibb B, Gupta K, Ghosh K, Sharp R, Chen J, Van Duyne GD. 2010. Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Res* 38:5817–5832.
40. Gourlay SC, Colloms SD. 2004. Control of Cre recombination by regulatory elements from Xer recombination systems. *Mol Microbiol* 52:53–65.
41. Grainge I, Buck D, Jayaram M. 2000. Geometry of site alignment during Int family recombination: antiparallel synapsis by the Flp recombinase. *J Mol Biol* 298:749–764.
42. Grainge I, Pathania S, Vologodskii A, Harshey RM, Jayaram M. 2002. Symmetric DNA sites are functionally asymmetric within Flp and Cre site-specific DNA recombination synapses. *J Mol Biol* 320:515–527.
43. Kilbride E, Boocock MR, Stark WM. 1999. Topological selectivity of a hybrid site-specific recombination system with elements from Tn3 res/resolvase and bacteriophage P1 loxP/Cre. *J Mol Biol* 289:1219–1230.
44. Ma CH, Liu YT, Savva CG, Rowley PA, Cannon B, Fan HF, Russell R, Holzenburg A, Jayaram M. 2014. Organization of DNA partners and strand exchange mechanisms during Flp site-specific recombination analyzed by difference topology, single molecule FRET and single molecule TPM. *J Mol Biol* 426:793–815.
45. Cassell G, Moision R, Rabani E, Segall A. 1999. The geometry of a synaptic intermediate in a pathway of bacteriophage λ site-specific recombination. *Nucleic Acids Res* 27:1145–1151.
46. Huffman KE, Levene SD. 1999. DNA-sequence asymmetry directs the alignment of recombination sites in the FLP synaptic complex. *J Mol Biol* 286:1–13.
47. Vetcher AA, Lushnikov AY, Navarra-Madsen J, Scharein RG, Lyubchenko YL, Darcy IK, Levene SD. 2006. DNA topology and geometry in Flp and Cre recombination. *J Mol Biol* 357:1089–1104.
48. Ghosh K, Lau CK, Gupta K, Van Duyne GD. 2005. Preferential synapsis of loxP sites drives ordered strand exchange in Cre-loxP site-specific recombination. *Nat Chem Biol* 1:275–282.
49. Krogh BO, Shuman S. 2002. Proton relay mechanism of general acid catalysis by DNA topoisomerase IB. *J Biol Chem* 277:5711–5714.
50. Nagarajan R, Kwon K, Nawrot B, Stec WJ, Stivers JT. 2005. Catalytic phosphoryl interactions of topoisomerase IB. *Biochemistry* 44:11476–11485.
51. Whiteson KL, Chen Y, Chopra N, Raymond AC, Rice PA. 2007. Identification of a potential general acid/base in the reversible phosphoryl transfer reactions catalyzed by tyrosine recombinases: Flp H305. *Chem Biol* 14:121–129.
52. Guo F, Gopaul DN, Van Duyne GD. 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombinase synapse. *Nature* 389:40–46.
53. Chen Y, Rice PA. 2003. The role of the conserved Trp330 in Flp-mediated recombination. Functional and structural analysis. *J Biol Chem* 278:24800–24807.
54. Ma CH, Kwiatek A, Bolusani S, Voziyarov Y, Jayaram M. 2007. Unveiling hidden catalytic contributions of the conserved His/Trp-III in tyrosine recombinases: assembly of a novel active site in Flp recombinase harboring alanine at this position. *J Mol Biol* 368:183–196.
55. Jayaram M, Grainge I, Tribble G. 2002. Site-specific DNA recombination mediated by the Flp protein of *Saccharomyces cerevisiae*, p 192–218. In Craig NL, Craigie R, Gellert M, Lambowitz AM (ed), *Mobile DNA II*, ASM Press, Washington, DC.
56. Xu CJ, Ahn YT, Pathania S, Jayaram M. 1998. Flp ribonuclease activities. Mechanistic similarities and contrasts to site-specific DNA recombination. *J Biol Chem* 273:30591–30598.
57. Xu CJ, Grainge I, Lee J, Harshey RM, Jayaram M. 1998. Unveiling two distinct ribonuclease activities and a topoisomerase activity in a site-specific DNA recombinase. *Mol Cell* 1:729–739.
58. Sau AK, Tribble G, Grainge I, Frohlich RF, Knudsen BR, Jayaram M. 2001. Biochemical and kinetic analysis of the RNase active sites of the integrase/tyrosine family site-specific DNA recombinases. *J Biol Chem* 276:46612–46623.
59. Chen JW, Lee J, Jayaram M. 1992. DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell* 69:647–658.
60. Yang SH, Jayaram M. 1994. Generality of the shared active site among yeast family site-specific recombinases. The R site-specific recombinase follows the Flp paradigm. *J Biol Chem* 269:12789–12796.
61. Letzelter C, Duguet M, Serre MC. 2004. Mutational analysis of the archaeal tyrosine recombinase SSV1 integrase suggests a mechanism of DNA cleavage in trans. *J Biol Chem* 279:28936–28944.
62. Zhan Z, Ouyang S, Liang W, Zhang Z, Liu ZJ, Huang L. 2012. Structural and functional characterization of the C-terminal catalytic domain of SSV1 integrase. *Acta Crystallogr, Sect D: Biol Crystallogr* 68:659–670.
63. Grainge I, Lee J, Xu CJ, Jayaram M. 2001. DNA recombination and RNA cleavage activities of the Flp protein: roles of two histidine residues in the orientation and activation of the nucleophile for strand cleavage. *J Mol Biol* 314:717–733.
64. Jayaram M. 1997. The *cis-trans* paradox of integrase. *Science* 276:49–51.
65. Yarus M. 1993. How many catalytic RNAs? Ions and the Cheshire cat conjecture. *FASEB J* 7:31–39.
66. Kimball AS, Lee J, Jayaram M, Tullius TD. 1993. Sequence-specific cleavage of DNA via nucleophilic attack of hydrogen peroxide, assisted by Flp recombinase. *Biochemistry* 32:4698–4701.

67. Lee J, Jayaram M. 1995. Functional roles of individual recombinase monomers in strand breakage and strand union during site-specific DNA recombination. *J Biol Chem* 270:23203–23211.
68. Ma CH, Rowley PA, Maciaszek A, Guga P, Jayaram M. 2009. Active site electrostatics protect genome integrity by blocking abortive hydrolysis during DNA recombination. *EMBO J* 28:1745–1756.
69. Kitts PA, Nash HA. 1987. Homology-dependent interactions in phage lambda site-specific recombination. *Nature* 329:346–348.
70. Nunes-Duby SE, Matsumoto L, Landy A. 1987. Site-specific recombination intermediates trapped with suicide substrates. *Cell* 50:779–788.
71. Arciszewska L, Grainge I, Sherratt D. 1995. Effects of Holliday junction position on Xer-mediated recombination in vitro. *EMBO J* 14:2651–2660.
72. Dixon JE, Sadowski PD. 1994. Resolution of immobile chi structures by the FLP recombinase of 2 micron plasmid. *J Mol Biol* 243:199–207.
73. Lee J, Jayaram M. 1995. Role of partner homology in DNA recombination. Complementary base pairing orients the 5'-hydroxyl for strand joining during Flp site-specific recombination. *J Biol Chem* 270:4042–4052.
74. Nunes-Duby SE, Azaro MA, Landy A. 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr Biol* 5:139–148.
75. Voziyanov Y, Pathania S, Jayaram M. 1999. A general model for site-specific recombination by the integrase family recombinases. *Nucleic Acids Res* 27:930–941.
76. Cambray G, Guerout AM, Mazel D. 2010. Integrons. *Annu Rev Genet* 44:141–166.
77. Lee G, Saito I. 1998. Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* 216:55–65.
78. Azam N, Dixon JE, Sadowski PD. 1997. Topological analysis of the role of homology in Flp-mediated recombination. *J Biol Chem* 272:8731–8738.
79. Kanaar R, Klippel A, Shekhtman E, Dungan JM, Kahmann R, Cozzarelli NR. 1990. Processive recombination by the phage Mu Gin system: implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action. *Cell* 62:353–366.
80. Stark WM, Grindley ND, Hatfull GF, Boocock MR. 1991. Resolvase-catalysed reactions between res sites differing in the central dinucleotide of subsite I. *EMBO J* 10:3541–3548.
81. Krogh BO, Shuman S. 2000. Catalytic mechanism of DNA topoisomerase IB. *Mol Cell* 5:1035–1041.
82. Stivers JT, Jagadeesh GJ, Nawrot B, Stec WJ, Shuman S. 2000. Stereochemical outcome and kinetic effects of Rp- and Sp-phosphorothioate substitutions at the cleavage site of vaccinia type I DNA topoisomerase. *Biochemistry* 39:5561–5572.
83. Tian L, Claeboe CD, Hecht SM, Shuman S. 2005. Mechanistic plasticity of DNA topoisomerase IB: phosphate electrostatics dictate the need for a catalytic arginine. *Structure* 13:513–520.
84. Tian L, Claeboe CD, Hecht SM, Shuman S. 2003. Guarding the genome: electrostatic repulsion of water by DNA suppresses a potent nuclease activity of topoisomerase IB. *Mol Cell* 12:199–208.
85. Qian XH, Inman RB, Cox MM. 1990. Protein-based asymmetry and protein-protein interactions in FLP recombinase-mediated site-specific recombination. *J Biol Chem* 265:21779–21788.
86. Rowley PA, Kachroo AH, Ma CH, Maciaszek AD, Guga P, Jayaram M. 2010. Electrostatic suppression allows tyrosine site-specific recombination in the absence of a conserved catalytic arginine. *J Biol Chem* 285:22976–22985.
87. Kachroo AH, Ma CH, Rowley PA, Maciaszek AD, Guga P, Jayaram M. 2010. Restoration of catalytic functions in Cre recombinase mutants by electrostatic compensation between active site and DNA substrate. *Nucleic Acids Res* 38:6589–6601.
88. Ma CH, Kachroo AH, Maciaszek A, Chen TY, Guga P, Jayaram M. 2009. Reactions of Cre with methylphosphonate DNA: similarities and contrasts with Flp and vaccinia topoisomerase. *PLoS One* 4:e7248.
89. Davies DR, Mushtaq A, Interthal H, Champoux JJ, Hol WG. 2006. The structure of the transition state of the heterodimeric topoisomerase I of *Leishmania donovani* as a vanadate complex with nicked DNA. *J Mol Biol* 357:1202–1210.
90. Fan HF. 2012. Real-time single-molecule tethered particle motion experiments reveal the kinetics and mechanisms of Cre-mediated site-specific recombination. *Nucleic Acids Res* 40:6208–6222.
91. Fan HF, Ma CH, Jayaram M. 2013. Real-time single-molecule tethered particle motion analysis reveals mechanistic similarities and contrasts of Flp site-specific recombinase with Cre and lambda Int. *Nucleic Acids Res* 41:7031–7047.
92. Mumm JP, Landy A, Gelles J. 2006. Viewing single lambda site-specific recombination events from start to finish. *EMBO J* 25:4586–4595.
93. Uphoff S, Reyes-Lamothe R, Garza de Leon F, Sherratt DJ, Kapanidis AN. 2013. Single-molecule DNA repair in live bacteria. *Proc Natl Acad Sci U S A* 110:8063–8068.
94. Guo F, Gopaul DN, Van Duyne GD. 1999. Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc Natl Acad Sci U S A* 96:7143–7148.
95. Dixon JE, Sadowski PD. 1993. Resolution of synthetic chi structures by the FLP site-specific recombinase. *J Mol Biol* 234:522–533.
96. Lee J, Tribble G, Jayaram M. 2000. Resolution of tethered antiparallel and parallel Holliday junctions by the Flp site-specific recombinase. *J Mol Biol* 296:403–419.
97. Kitts PA, Nash HA. 1988. Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *J Mol Biol* 204:95–107.
98. Bregu M, Sherratt DJ, Colloms SD. 2002. Accessory factors determine the order of strand exchange in Xer recombination at psi. *EMBO J* 21:3888–3897.

99. Grainge I, Lesterlin C, Sherratt DJ. 2011. Activation of XerCD-dif recombination by the FtsK DNA translocase. *Nucleic Acids Res* 39:5140–5148.
100. Du Q, Livshits A, Kwiatek A, Jayaram M, Vologodskii A. 2007. Protein-induced local DNA bends regulate global topology of recombination products. *J Mol Biol* 368:170–182.
101. Colloms SD. 2013. The topology of plasmid-monomerizing Xer site-specific recombination. *Biochem Soc Trans* 41:589–594.
102. Vazquez M, Colloms SD, Sumners D. 2005. Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single three-dimensional topological pathway. *J Mol Biol* 346:493–504.
103. Ip SC, Bregu M, Barre FX, Sherratt DJ. 2003. Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination. *EMBO J* 22:6399–6407.
104. Zechiedrich EL, Khodursky AB, Cozzarelli NR. 1997. Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in *Escherichia coli*. *Genes Dev* 11:2580–2592.
105. Grainge I, Bregu M, Vazquez M, Sivanathan V, Ip SC, Sherratt DJ. 2007. Unlinking chromosome catenanes in vivo by site-specific recombination. *EMBO J* 26:4228–4238.
106. Jayaram M. 2013. Mathematical validation of a biological model for unlinking replication catenanes by recombination. *Proc Natl Acad Sci U S A* 110:20854–20855.
107. Shimokawa K, Ishihara K, Grainge I, Sherratt DJ, Vazquez M. 2013. FtsK-dependent XerCD-dif recombination unlinks replication catenanes in a stepwise manner. *Proc Natl Acad Sci U S A* 110:20906–20911.
108. Crisona NJ, Weinberg RL, Peter BJ, Sumners DW, Cozzarelli NR. 1999. The topological mechanism of phage lambda integrase. *J Mol Biol* 289:747–775.
109. Nash HA, Pollock TJ. 1983. Site-specific recombination of bacteriophage lambda. The change in topological linking number associated with exchange of DNA strands. *J Mol Biol* 170:19–38.
110. Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD. 2013. Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* 9:685–700.
111. Lee T. 2013. Generating mosaics for lineage analysis in flies. *Wiley Interdiscip Rev: Dev Biol* 3:69–81.
112. Kretschmar K, Watt FM. 2012. Lineage tracing. *Cell* 148:33–45.
113. Smedley D, Salimova E, Rosenthal N. 2011. Cre recombinase resources for conditional mouse mutagenesis. *Methods* 53:411–416.
114. Zhang DJ, Wang Q, Wei J, Baimukanova G, Buchholz F, Stewart AF, Mao X, Killeen N. 2005. Selective expression of the Cre recombinase in late-stage thymocytes using the distal promoter of the Lck gene. *J Immunol* 174:6725–6731.
115. Harshey RM, Jayaram M. 2006. The Mu transpososome through a topological lens. *Crit Rev Biochem Mol Biol* 41:387–405.
116. Jayaram M, Harshey RM. 2009. Difference topology: analysis of high-order DNA-protein assemblies, p 139–158. In Santosa F, Keel M (ed), *The IMA volumes in mathematics and its applications, Mathematics of DNA structure, function and interactions*. Springer.
117. Pathania S, Jayaram M, Harshey RM. 2002. Path of DNA within the Mu transpososome. Transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils. *Cell* 109:425–436.
118. Konieczka JH, Paek A, Jayaram M, Voziyanov Y. 2004. Recombination of hybrid target sites by binary combinations of Flp variants: mutations that foster interprotomer collaboration and enlarge substrate tolerance. *J Mol Biol* 339:365–378.
119. Santoro SW, Schultz PG. 2002. Directed evolution of the site specificity of Cre recombinase. *Proc Natl Acad Sci U S A* 99:4185–4190.
120. Tay Y, Ho C, Droge P, Ghadessy FJ. 2010. Selection of bacteriophage lambda integrases with altered recombination specificity by in vitro compartmentalization. *Nucleic Acids Res* 38:e25.
121. Voziyanov Y, Konieczka JH, Stewart AF, Jayaram M. 2003. Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. *J Mol Biol* 326:65–76.
122. Christ N, Droge P. 2002. Genetic manipulation of mouse embryonic stem cells by mutant lambda integrase. *Genesis* 32:203–208.
123. Akopian A, He J, Boocock MR, Stark WM. 2003. Chimeric recombinases with designed DNA sequence recognition. *Proc Natl Acad Sci U S A* 100:8688–8691.
124. Brown WR, Lee NC, Xu Z, Smith MC. 2011. Serine recombinases as tools for genome engineering. *Methods* 53:372–379.
125. Carroll D. 2011. Genome engineering with zinc-finger nucleases. *Genetics* 188:773–782.
126. Gaj T, Gersbach CA, Barbas CF 3rd. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31:397–405.
127. Proudfoot C, McPherson AL, Kolb AF, Stark WM. 2011. Zinc finger recombinases with adaptable DNA sequence specificity. *PLoS One* 6:e19537.
128. Buchholz F, Angrand PO, Stewart AF. 1998. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 16:657–662.
129. Raymond CS, Soriano P. 2007. High-efficiency FLP and phiC31 site-specific recombination in mammalian cells. *PLoS One* 2:e162.
130. Garcia-Otin AL, Guillou F. 2006. Mammalian genome targeting using site-specific recombinases. *Frontiers Biosci* 11:1108–1136.
131. Feil R, Brocard J, Mascrez B, LeMeur M, Metzger D, Chambon P. 1996. Ligand-activated site-specific recombination in mice. *Proc Natl Acad Sci U S A* 93:10887–10890.
132. Feil R, Wagner J, Metzger D, Chambon P. 1997. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* 237:752–757.

133. Logie C, Stewart AF. 1995. Ligand-regulated site-specific recombination. *Proc Natl Acad Sci U S A* 92:5940–5944.
134. Metzger D, Clifford J, Chiba H, Chambon P. 1995. Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc Natl Acad Sci U S A* 92:6991–6995.
135. Buchholz F, Stewart AF. 2001. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat Biotechnol* 19:1047–1052.
136. Voziyanov Y, Stewart AF, Jayaram M. 2002. A dual reporter screening system identifies the amino acid at position 82 in Flp site-specific recombinase as a determinant for target specificity. *Nucleic Acids Res* 30:1656–1663.
137. Baldwin EP, Martin SS, Abel J, Gelato KA, Kim H, Schultz PG, Santoro SW. 2003. A specificity switch in selected Cre recombinase variants is mediated by macromolecular plasticity and water. *Chem Biol* 10:1085–1094.
138. Bolusani S, Ma CH, Paek A, Konieczka JH, Jayaram M, Voziyanov Y. 2006. Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites. *Nucleic Acids Res* 34:5259–5269.
139. Shultz JL, Voziyanova E, Konieczka JH, Voziyanov Y. 2011. A genome-wide analysis of FRT-like sequences in the human genome. *PLoS One* 6:e18077.
140. Sarkar I, Hauber I, Hauber J, Buchholz F. 2007. HIV-1 proviral DNA excision using an evolved recombinase. *Science* 316:1912–1915.
141. Schlake T, Bode J. 1994. Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* 33:12746–12751.
142. Turan S, Galla M, Ernst E, Qiao J, Voelkel C, Schiedlmeier B, Zehe C, Bode J. 2011. Recombinase mediated cassette exchange (RMCE): traditional concepts and current challenges. *J Mol Biol* 407:193–221.
143. Bethke B, Sauer B. 1997. Segmental genomic replacement by Cre-mediated recombination: genotoxic stress activation of the p53 promoter in single-copy transfectants. *Nucleic Acids Res* 25:2828–2834.
144. Osterwalder M, Galli A, Rosen B, Skarnes WC, Zeller R, Lopez-Rios J. 2010. Dual RMCE for efficient re-engineering of mouse mutant alleles. *Nat Methods* 7:893–895.
145. Anderson RP, Voziyanova E, Voziyanov Y. 2012. Flp and Cre expressed from Flp-2A-Cre and Flp-IRES-Cre transcription units mediate the highest level of dual recombinase-mediated cassette exchange. *Nucleic Acids Res* 40:e62.
146. Voziyanova E, Malchin N, Anderson RP, Yagil E, Kolot M, Voziyanov Y. 2013. Efficient Flp-Int HK022 dual RMCE in mammalian cells. *Nucleic Acids Res* 41:e125.
147. Aihara H, Kwon HJ, Nunes-Duby SE, Landy A, Ellenberger T. 2003. A conformational switch controls the DNA cleavage activity of λ integrase. *Mol Cell* 12:187–198.

W. Marshall Stark¹

The Serine Recombinases

3

INTRODUCTION

Site-Specific Recombination: A Brief Primer

The term site-specific recombination encompasses a group of biological processes that, unlike homologous recombination, promote rearrangements of DNA by breaking and rejoining strands at precisely defined sequence positions. In a canonical site-specific recombination event, two discrete sites (sequences of DNA, typically a few tens of base pairs long) are broken, and the ends are reciprocally exchanged and rejoined, resulting in recombinant products (Fig. 1). Site-specific recombination does not require extensive sequence homology; the sites are identified and brought together by protein–DNA and protein–protein interactions involving specialized recombinase proteins, unlike homologous recombination where DNA–DNA interactions define the loci of strand exchange. “Conservative” site-specific recombination systems form recombinants without any requirement for DNA synthesis or high-energy cofactors. Some other biological processes such as transposition are sometimes categorized with site-specific recombination because of common features including cleavage and rejoining of DNA strands at precise positions defined by protein–DNA interactions, but these processes may require DNA synthesis and/or ligase-mediated rejoining of DNA strands. The systems discussed in this chapter conform to the strict “conservative”

definition. General aspects of site-specific recombination have been reviewed elsewhere (1, 2, 3).

Site-specific recombination can have different outcomes depending on the nature of the DNA substrate(s) (Fig. 2). Recombination between two sites, each on a separate linear DNA molecule, results in linear recombinants. Two outcomes are possible, depending on which “half-site” is joined to which, as shown in Fig. 2a. However, typical sites have a polarity, such that the “left half” of one site is joined to the “right half” of the other, and vice versa; thus, only one of these possibilities normally occurs. The origin of the site polarity is discussed below. If the two sites are on separate molecules but one or both molecules are circular (Fig. 2b), recombination will join the two molecules together (this is called integration or fusion). The product molecule contains two sites, oriented in a direct repeat (head-to-tail) relationship. Conversely, recombination of this two-site molecule splits it into two products (this is called excision or resolution). If two sites within a single DNA molecule are in an inverted relationship (Fig. 2c), recombination inverts the orientation of one DNA segment bounded by the sites, relative to the other. In most real site-specific recombination systems, restrictions imposed by the mechanism of recombinase-mediated catalysis allow only some of these possibilities (see below). Site-specific recombination is seemingly isoenergetic; the products, like the substrates, are normal double-stranded DNA molecules. Reactions might

¹Institute of Molecular, Cell and Systems Biology, University of Glasgow, Bower Building, Glasgow G12 8QQ, Scotland, United Kingdom.

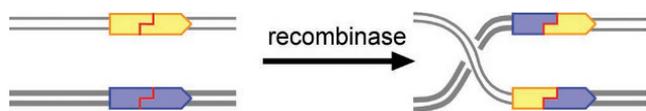


Figure 1 Site-specific recombination. Two sites (pointed boxes) in double-helical DNA (shown as double lines) are recognized by a recombinase protein (not shown), and then cut and rejoined to form recombinants.

doi:10.1128/microbiolspec.MDNA3-0046-2014.f1

therefore be expected to reach a 1:1 equilibrium of substrates and recombinants. However, natural systems have evolved strategies to bias the reaction toward the desired products; some examples are described in the sections that follow.

Conservative site-specific recombination has been adopted widely for diverse programmed DNA rearrangements essential to the biology of bacteria, archaea and the mobile DNA elements that infest them (bacteriophages, plasmids and transposons) (2, 4, 5, 6). Curiously, however, there are only a few known conservative

site-specific recombination systems in eukaryotes, and some of these may be associated with bacterial symbionts or bacterial-derived organelles, or may be recent acquisitions from horizontal transfer of mobile DNA (1, 5, 6, 7, 8). Roles of site-specific recombination systems include temperate bacteriophage DNA integration and excision from the host bacterial genomic DNA, transposon cointegrate resolution, monomerization of plasmid multimers, switching of gene expression by inversion of regulatory sequences relative to coding sequences and developmentally programmed excision of intervening genomic sequences. There is no clear distinction of the biological functions of systems based on serine recombinases, the subject of this chapter, from those based on the other large family, the tyrosine recombinases (see Chapter 2, this volume). It seems that Nature has evolved two quite different ways of doing site-specific recombination, both of which are sufficiently “fit for purpose” to survive and prosper in present-day organisms.

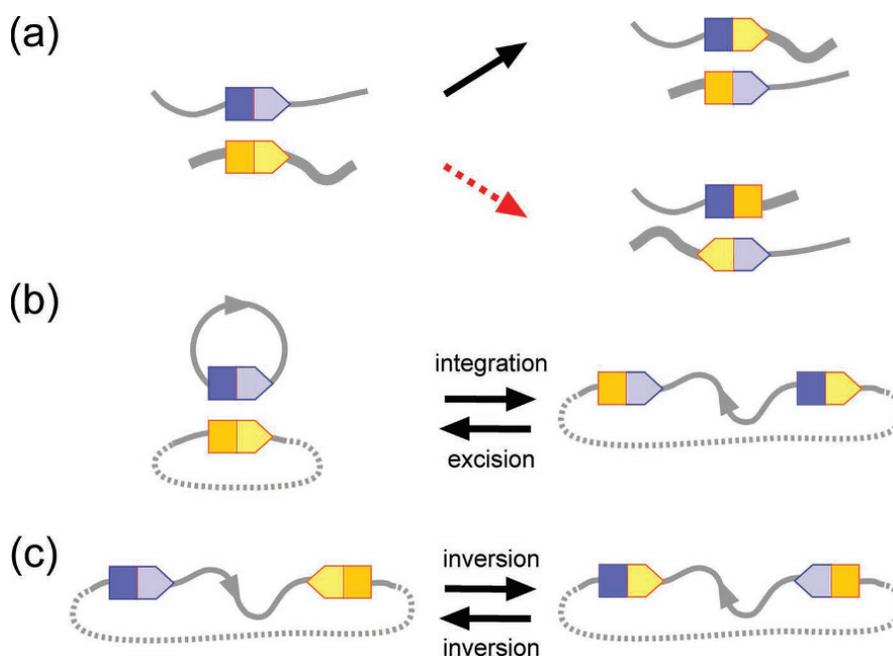


Figure 2 Site-specific recombination outcomes. (a) Recombination between two sites in separate linear DNA molecules results in two linear recombinant products. Usually, the sites have a polarity (indicated by the pointed boxes) such that the lower pathway (red arrow) is forbidden. (b) Recombination between two sites in separate DNA molecules, when one or both of the molecules is circular, results in a single product molecule containing two sites in direct repeat. This is called integration or fusion. The “reverse” reaction splits a molecule containing two sites into two product molecules, one or both of which are circular. This is called resolution, excision, or deletion (depending on the biological context). (c) Recombination between two sites in inverted repeat in a DNA molecule inverts the orientation of one segment of DNA relative to the other. This is called inversion.

doi:10.1128/microbiolspec.MDNA3-0046-2014.f2

Mechanistic Nuts and Bolts

In this section, I will give a brief overview of the molecular mechanisms of conservative site-specific recombination.

In many systems, recombination takes place between two identical sites, and two identical sites are reformed in the recombinants. However, there are examples (notably bacteriophage integrase systems; see “Regulation of recombination by large serine recombinases” below) where recombination is between two different sites. Sites range in length from about 25 up to several hundreds of base pairs. The shortest sites typically have imperfect 2-fold (dyad/palindrome) DNA symmetry, consistent with their observed or inferred property of binding a symmetric dimer of recombinase. The specific phosphodiester linkages that are cut and rejoined during recombination are located close to the center of the site (Fig. 3). Longer sites comprise a “crossover site” conforming to the above description, which binds a recombinase dimer and within which are the points of strand exchange, as well as adjacent “accessory sequences” on one or both sides of the crossover site, which may include binding sites for additional recombinase subunits or other “accessory proteins,” or for looping interactions with recombinase subunits bound at the crossover site (Fig. 3a). The roles of the accessory sequences are in regulation of recombinase activity; initiation of catalysis typically depends on their presence and their correct interactions with other components of the system (1, 2, 3).

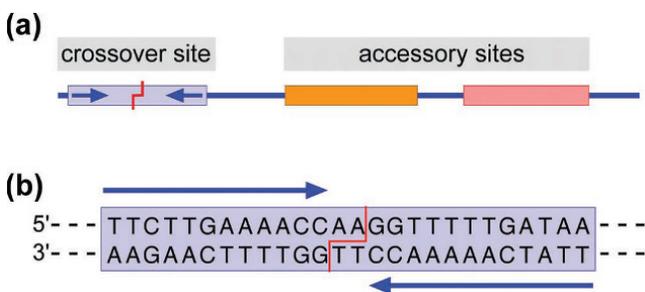


Figure 3 Recombination sites. (a) A typical recombination site. The crossover site, where strand exchange takes place (at the position marked by the staggered red line), binds a recombinase dimer and typically has partial dyad symmetry (indicated by the blue arrows). “Accessory sites,” which may be adjacent on one side of the crossover site (as shown), on both sides or more distant, may bind additional recombinase subunits or other proteins, or may make looping interactions with recombinase bound at the crossover site. (b) Example of a real crossover site (*hixL*, a site acted upon by *Hin* recombinase). The colors and symbols are as in part (a). *Hin*, like all serine recombinases characterized to date, cuts the DNA at the center of the crossover site with a 2 bp “stagger” as shown. doi:10.1128/microbiolspec.MDNA3-0046-2014.f3

Recombinases do not cut the two DNA strands at the precise center of the site. Instead, the break points are symmetrically positioned off-center, so that there are a few base pairs between the top strand and bottom strand break points. These base pairs are often referred to as the “overlap sequence” because the top and bottom strands of this sequence in the recombinant sites originate from different parent sites. All serine recombinase systems examined in this respect have 2 bp overlap sequences with the strand breaks staggered as shown in Fig. 3b; in contrast, the overlap sequences for tyrosine recombinases vary in length (typically 6 to 8 bp), and the stagger is in the opposite direction. If the “half-sites” that are to be joined to form recombinants do not have complementary overlap sequences, the products would have mismatched base pairs. This scenario can arise if two identical crossover sites are misaligned in the catalytic intermediate such that strand exchange pairs two identical, noncomplementary ends. Serine recombinases do not normally form mismatched recombinants; this is one origin of the site polarity discussed above. However, reactions of “mismatched” sites can have other consequences (see “Subunit rotation” below).

Each crossover site binds a recombinase dimer. A critical subsequent step is when two crossover sites come together; this is called synapsis. The “synapse” or “synaptic complex” that is thus formed comprises the two crossover sites bridged by a recombinase tetramer, and it is in this intermediate that the chemical steps of strand cleavage, exchange and ligation will take place. In regulated systems, crossover-site synapsis is typically a control point that depends on interactions with accessory factors.

In any conservative site-specific recombination event, there are eight chemical steps: four strand cleavages and four ligations. Cleavage occurs when a nucleophilic amino acid functional group at the recombinase active site attacks the scissile phosphodiester bond of a DNA strand; for the serine recombinases, this is the hydroxyl group of a serine residue. The immediate product of cleavage has a broken DNA strand, with a covalent phosphodiester linkage between one DNA end and the recombinase at the break point. Serine recombinases become linked to the 5′ end of the DNA, leaving a 3′-hydroxyl group on the other end at the break. Serine recombinases cleave all four DNA strands in the synaptic complex, creating double-strand breaks at the center of each crossover site. Each half-site thus formed has a recombinase subunit covalently attached to its 5′ end, and 2-nt single-stranded protrusions terminated by a 3′-OH group (Fig. 4). The half-sites are then exchanged

and re-ligated, creating recombinants. This mechanism contrasts with that of the tyrosine recombinases, which become linked to the 3' end of the DNA at the strand break and do not make double-strand break intermediates. Instead, they cleave, exchange and re-ligate pairs of single DNA strands; thus, strand exchange proceeds via a “four-way junction” intermediate with two recombinant and two non-recombinant strands (see Chapter 2 of this volume).

SERINE RECOMBINASES

Some History

Following the discovery of the first site-specific recombinase, λ Int, in the 1970s, it was realized that the product of the *tnpR* gene encoded by the bacterial penicillin resistance transposon Tn3 has a similar function (9, 10). Detailed characterization of the *tnpR* gene product (resolvase) and the recombination site (*res*) soon followed (11, 12, 13). It quickly became apparent that there was a group of enzymes related to Tn3 resolvase, encoded by other bacterial transposons and DNA inversion systems. The group came to be known as the resolvase or resolvase/invertase family (14, 15). Pioneering *in vitro* studies of the $\gamma\delta$ transposon resolvase (closely related to Tn3 resolvase) by Reed and Grindley revealed basic mechanistic differences from λ Int and its relatives (16, 17, 18). It was later shown that the resolvase–DNA linkage is via a serine residue, unlike the tyrosine that is used by λ Int and its brethren (19, 20, 21, 22). In the 1990s, the two families came to be referred to as the “serine” and “tyrosine” recombinases (23, 24).

Serine Recombinase Proteins

All serine recombinases possess a characteristic catalytic domain, which implements the chemical steps of strand exchange. I will call it the “SR” (serine recombinase) domain throughout this review. The size of the

SR domain is remarkably constant (usually about 150 amino acid residues). Several of its amino acid residues are highly conserved and are now known to contribute to the structure of the active site (3). All known serine recombinases have “attachments” to the SR domain, usually at the C terminus; these vary substantially and their specific properties have roles in definition of the recombinase function (25, 26) (Fig. 5). The recombinases studied in the early days (transposon resolvases and invertases) have a simple configuration with the SR domain at the N-terminus linked to a small C-terminal helix–turn–helix (HTH) DNA-binding domain, giving a total length of ~180–200 residues. These have come to be known as the “small serine recombinases”. However, as identification of putative serine recombinases by sequence analysis gathered pace, the diversity of the family became apparent (25, 26) (Fig. 5). Many sequences could be aligned with the entire length of the small serine recombinases but have extensions at the C-terminal end, such as the ISXc5 resolvase. Others have large C-terminal extensions immediately after the SR domain, in place of the HTH domain. An important subgroup of these “large serine recombinases” includes the bacteriophage serine integrases, the first to be identified being that of the *Streptomyces* phage ϕ C31 (28). These proteins (~400 to 700 amino acids) have an N-terminal SR domain followed by a complex, variable multidomain region with DNA-binding and regulatory functions, which are still only partially characterized (27). At first, it seemed that a “rule” was that the SR domain should be at the extreme N terminus of the protein, but proteins with a HTH domain preceding the SR domain were then identified and are now known to be transposases (29, 30).

Biological Roles of Serine Recombinases

The role of the patriarchs of the serine recombinase family, the transposon resolvases, is to divide (“resolve”) a large circular DNA molecule into two smaller

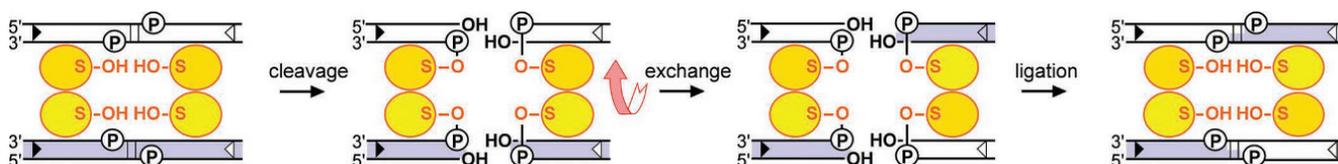


Figure 4 The serine recombinase strand-exchange mechanism. A synaptic complex of two crossover sites bridged by a recombinase tetramer (yellow ovals) is shown. The four subunits are spaced out, so that the catalytic steps can be seen clearly. The catalytic serine residues are indicated by S-OH. The scissile phosphodiesteres are represented as circled Ps, and in the first and last panels the 2-bp overlap sequence is indicated by vertical lines. For further details, see text. doi:10.1128/microbiolspec.MDNA3-0046-2014.f4

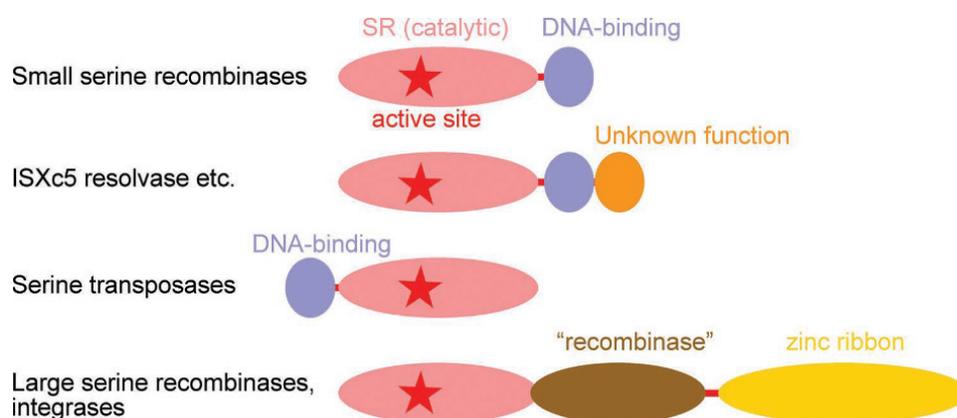


Figure 5 Domain structures of serine recombinases (26). The SR (catalytic) domain (shown in pink; typically ~150 amino acids) is common to all serine recombinases and contains the active site (red star). Small serine recombinases (including resolvases and invertases) have a helix–turn–helix DNA-binding domain (blue; ~40 amino acids) at the C terminus. Some related recombinases such as ISXc5 resolvase have additional C-terminal domains (orange) of unknown function. Serine transposases have a similar helix–turn–helix domain at the N terminus. Large serine recombinases have multiple domains at the C terminus of the SR domain (27). doi:10.1128/microbiolspec.MDNA3-0046-2014.f5

circles. The natural substrate is a “cointegrate” molecule formed by replicative transposition, which contains two transposons, each with a *res* recombination site (31, 32). Closely related resolution systems are encoded by some bacterial plasmids and act to reduce plasmid multimers to monomers. DNA invertases promote flipping of the DNA sequence between two sites, thereby switching between different modes of gene expression (often to evade host defenses against infection). The activity of one invertase, Hin, from *Salmonella typhimurium*, is responsible for the phenomenon of flagellar phase variation studied since the 1920s. Other invertases are encoded in bacterial, bacteriophage and plasmid genomes (33). Bacteriophage serine integrases integrate and excise the DNA genomes of “temperate” or “lysogenic” phages to/from the bacterial host chromosomal DNA, like the famous tyrosine recombinase-based phage λ system (34, 35). Recently, some small serine recombinases with similarity to the DNA invertase group have also been shown to be phage integrases (36). As noted above, a group of transposases have an SR domain with a HTH domain at the N terminus (30). In addition, substantial numbers of proteins in the databases have homology to the SR domain but have unknown functions (M. R. Boocock, personal communication); there may still be surprises in store. More details on the functions of particular groups of serine recombinases are given in Chapters 9, 10 and 11, this volume.

Serine Recombinase Structures

There is now a substantial bank of structural data at atomic resolution on members of the serine recombinase family. In particular, a series of groundbreaking crystal structures of $\gamma\delta$ resolvase obtained by the Steitz laboratory in Yale has gone hand in hand with our developing understanding of serine recombinase mechanisms (37, 38, 39, 40, 41, 42). Recently, crystallographic studies of a distantly related serine resolvase, Sin, have transformed our understanding of the regulatory mechanisms and catalytic active site (43, 44). Chapter 10 (this volume) gives an in-depth review of these data. Further insights have come from the structures of the “attachments” to the SR domain. Table 1 summarizes the crystallographic and nuclear magnetic resonance (NMR) structural data on serine recombinases available at the time of submission of this review (45, 46, 47, 48, 49, 50, 51, 52). Small-angle scattering-based methods have also provided important structural information (53).

As an example, Fig. 6a shows the structure of a $\gamma\delta$ resolvase dimer bound to the crossover site of the $\gamma\delta$ *res* recombination site (40). Each subunit (183 amino acids) comprises an SR domain connected via a short linker peptide to a C-terminal HTH domain. The HTH domains recognize sequence motifs at the ends of the crossover site. Each SR domain comprises a core β -sheet decorated with α -helical and irregular regions, and ends in a long α -helix whose C-terminal region

Table 1 Structural data for serine recombinases

PDB accession no.	Description	Reference(s)
2RSL	$\gamma\delta$ resolvase; supersedes 1RSL	37, 39
1GDR	$\gamma\delta$ resolvase	38
1GDT	$\gamma\delta$ resolvase dimer bound to <i>res</i> site I DNA	40
1GHT, 1HX7	$\gamma\delta$ resolvase catalytic domain (NMR structures)	45
1RES, 1RET	$\gamma\delta$ resolvase DNA-binding domain (NMR structures)	46
1ZR2, 1ZR4	$\gamma\delta$ resolvase activated mutant tetramer in cleaved-DNA synaptic intermediate	41
2GM4	$\gamma\delta$ resolvase activated mutant tetramer in cleaved-DNA synaptic intermediate	42
2GM5	$\gamma\delta$ resolvase mutant tetramer	42
1HCR	Hin invertase C-terminal domain bound to DNA motif	47
1IJ6+	Hin C-terminal domain bound to wild-type and mutant DNA motifs (also 1IJW, 1JJ8, 1JKO, 1JKP, 1JKQ, and 1JKR)	48
3UJ3	Gin activated mutant tetramer. Supersedes 3PLO	49
4M6F	Gin dimer bound to <i>gix</i> site DNA	50
2R0Q	Sin tetramer in synaptic complex with <i>res</i> site II DNA	43
3PKZ	Sin activated mutant tetramer	44
4KIS	Bacteriophage A118 integrase (C-terminal part bound to <i>att</i> site DNA)	51
4BQQ	Bacteriophage Φ C31 integrase (N-terminal part)	McMahon SA, McEwan AR, Smith MCM, and Naismith, JH, unpublished data
3GUV	Large serine recombinase from <i>Streptococcus pneumoniae</i>	Bonanno JB, Freeman J, Bain KT, Do J, Sampathkumar P, Wasserman S, Sauder JM, Burley SK, and Almo SC, unpublished data
3BVP	Bacteriophage TP901-1 integrase catalytic domain	52
3G13	Transposase from CTn7 of <i>Clostridium difficile</i>	Bagaria A, Burley SK, and Swaminathan S, unpublished data
3ILX	TnpA transposase from <i>Sulfolobus solfataricus</i> ISC1904	Chang C, Bigelow L, Bearden J, and Joachimiak A, unpublished data
3LHF	TnpA transposase from <i>Sulfolobus solfataricus</i> IS1921	Stein AJ, Osipiuk J, Marshall N, Bearden J, Davidoff J, and Joachimiak A, unpublished data
3LHK	TnpA transposase from IS607-family element in <i>Methanocaldococcus jannaschii</i>	Chang C, Chhor G, Cobb G, and Joachimiak A, unpublished data

contacts the DNA minor groove near the center of the crossover site. The linker between this helix and the HTH domain lies in the minor groove and bears a structural resemblance to the “AT-hook” motif found in other DNA-binding proteins (54). The two SR domains make a complex network of interactions to form a dimer with imperfect 2-fold symmetry. The crossover site DNA is significantly bent but essentially “B-form.” The positions of the catalytic serine residues are shown in Fig. 6a; like other putative active-site residues, they are not in contact with the DNA in this structure. The resolvase is therefore considered to be bound to the DNA in a precatalytic conformation. However, subsequent structures, also solved by the Steitz group using “activated” $\gamma\delta$ resolvase variants (see below), revealed a catalytic intermediate containing a resolvase tetramer

bridging two crossover sites that have each been cleaved in both strands (Fig. 6b) (41, 42).

Taken together, the structural data (Table 1) show that the SR-domain fold is very well conserved throughout the serine recombinase family, despite substantial amino acid sequence divergence.

Serine Recombinase Mechanism

The early studies of Reed and Grindley demonstrated that the resolvase catalytic mechanism was significantly different from that of λ Int, Cre and other members of the “integrase family” (now tyrosine recombinases) (17, 18). DNA cleavage and rejoining by $\gamma\delta$ resolvase occur at precise positions within the 28-bp crossover site of *res*. Alteration of the reaction conditions allowed isolation of products with double-strand breaks at the

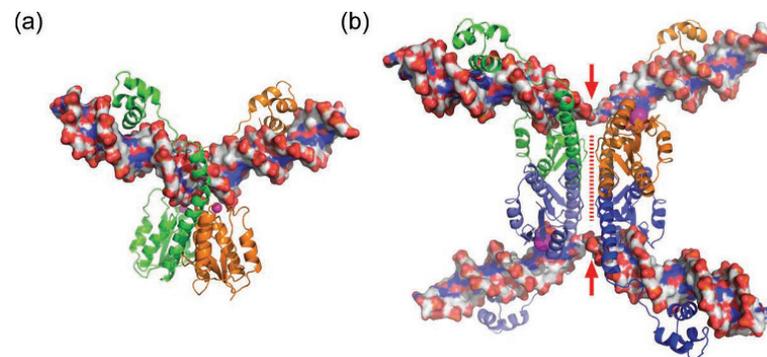


Figure 6 Crystal structures of $\gamma\delta$ resolvase–NA complexes. (a) Wild-type $\gamma\delta$ resolvase dimer bound to crossover-site DNA (PDB 1GDT; 40). The subunits are in cartoon representation (green and orange). The active site serine residues (α carbons) are indicated by magenta spheres. (b) Activated mutant $\gamma\delta$ resolvase tetramer in a cleaved-DNA synaptic intermediate (PDB 1ZR4; 41). The resolvase is rendered as in (a). The active-site serines are covalently linked to DNA ends (see Fig. 4); only two are visible. This view emphasizes the flat interface (marked by a dashed red line) between “rotating pairs” of resolvase subunits. The red arrows indicate positions of double-strand breaks in the DNA. The structure corresponds to the intermediates cartooned in the two central panels of Fig. 4. doi:10.1128/microbiolspec.MDNA3-0046-2014.f6

center of the crossover site. A resolvase subunit is covalently linked to each 5' end of the linearized DNA (18) (Fig. 4). The protein–DNA linkage was shown later to be a phosphodiester bond with the resolvase Ser10 (19, 21). The *in vitro* reaction is very efficient; under standard conditions, nearly all the substrate is converted into recombinant products within a few minutes. No cofactors or metal ions such as Mg^{2+} are required for activity. The analysis of Reed and Grindley also revealed that the product of resolution of a supercoiled plasmid substrate *in vitro* was a specific simple catenane in which the two product circles are linked as in a chain, an intriguing observation that led to many further studies and insights (see “Topological studies” below) (16).

Studies on related systems, including the resolution systems of Tn3 and Tn21, and the DNA invertases Gin, Hin and Cin, confirmed the generality of the mechanistic insights from the $\gamma\delta$ resolvase system (32, 33). However, the products of the inversion systems, their site structures and their regulation are substantially different, as will be discussed below.

The products with DNA double-strand breaks were presumed to be derived from a recombination intermediate, and suggested a simple “cut-and-paste” mechanism of strand exchange (Fig. 4). Together with the specific, simple catenane or unknotted circle product topologies of resolvases and invertases, respectively, the data suggested that exchange of DNA ends by serine recombinases is a well-ordered process, taking place within a synaptic complex of two crossover sites and a

recombinase tetramer, after double-strand cleavage of both sites (55, 56).

Topological Studies

In the absence of protein structural information, most early analysis of the mechanism focused on the DNA reaction products. Analysis of the product topologies from supercoiled circular (plasmid) substrates was especially significant (57). Studies with the tyrosine recombinase λ Int (and later FLP and Cre) had revealed that a supercoiled two-site substrate could give products with a wide range of knot/catenane topologies. These results were interpreted by a “random collision” mechanism of synapsis; that is, the sites collide due to natural random motions of the supercoiled substrate molecule. Various numbers of coils/tangles are trapped as the two sites synapse. A subsequent simple strand exchange mechanism results in products with a range of topologies (Fig. 7). Consistent with a random collision synapsis mechanism, these tyrosine recombinase systems did not distinguish between substrates with sites in different relative orientations: both “head-to-tail” (direct repeat) and “head-to-head” (inverted repeat) arrangements of sites were recombined equally well (57). The serine resolvases and invertases were clearly different. Resolvases yield almost exclusively simple catenane recombination products (Fig. 7a), and invertase recombination products are almost exclusively unknotted circles. Furthermore, resolvases only recombine substrates with sites in direct repeat, and

invertases only recombine sites in inverted repeat. These selectivities are very strong (for example, a $>10^4$ -fold rate difference for Tn3 resolvase), and persist even when the sites are separated by several kilobase pairs of DNA. Neither resolvases nor invertases recombine sites on separate supercoiled plasmids. The question therefore arose: how and why do these systems avoid the formation of random collision products? The phenomenon, which became known as “topological selectivity,” has been reviewed (57), this volume. To summarize very briefly, the catalytic activity of these serine

recombinases is strictly regulated so as to take place only when an elaborate synaptic complex is properly formed. This structure includes the accessory DNA sequences and protein subunits, and involves intertwining of the sites (as shown for resolvase in Fig. 7). The twisting/writhing of the DNA involved in synaptic complex formation is energetically favorable only when the sites come together in a specific way, in a substrate with the correct relationship between the two sites. The regulatory properties of synaptic complexes are discussed further below.

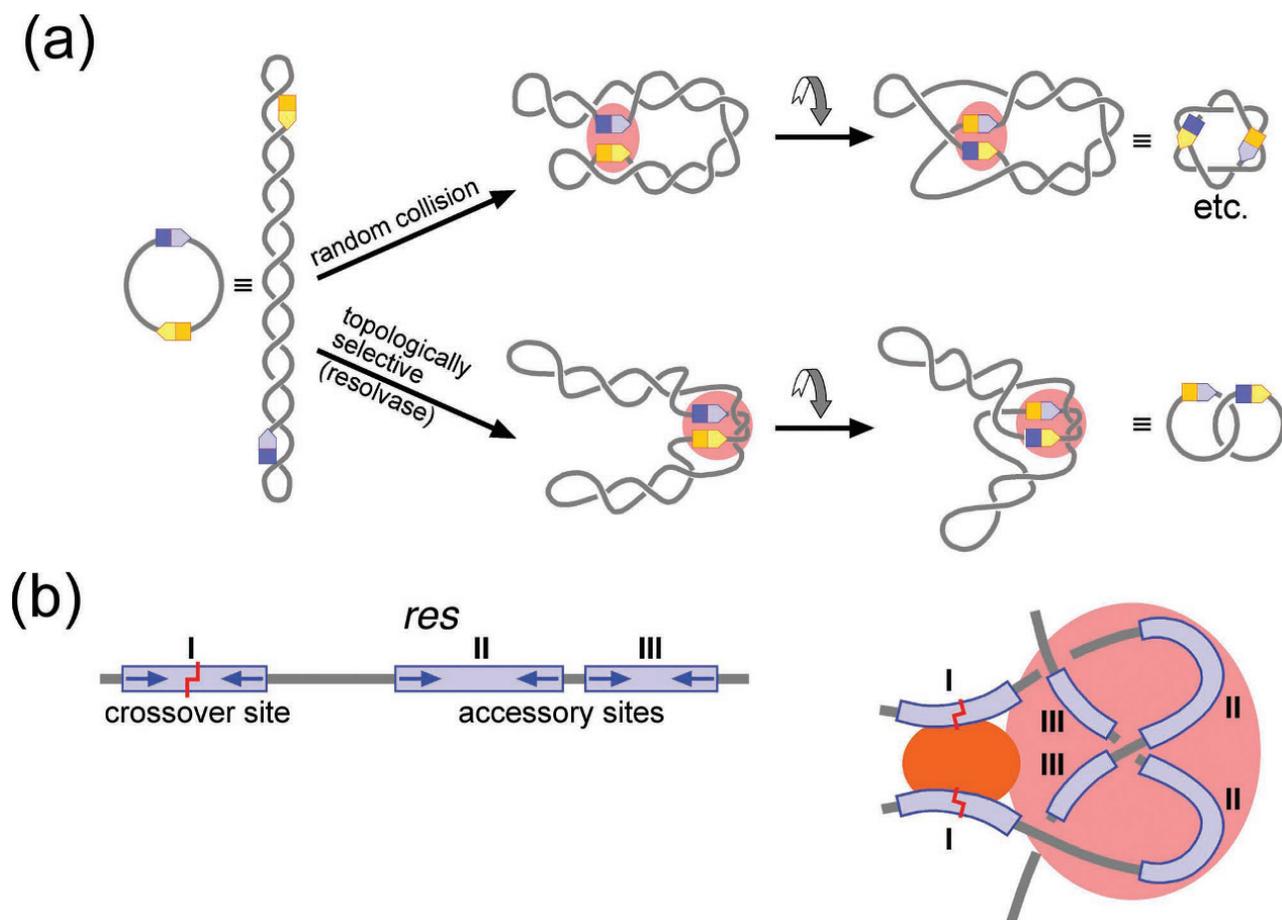


Figure 7 Topologically selective recombination by Tn3/γδ resolvase. (a) The reaction pathway of resolvase (lower row) is contrasted with that of a non-selective recombinase (upper row). Random collision of sites results in products with a variety of topologies (a 6-noded catenane is shown as an example here). Selective synapsis by resolvase results in a product with a specific topology (2-noded catenane). (b) Architecture of the synapse. The Tn3/γδ *res* site is diagrammed on the left. On the right, the arrangement of DNA in the synapse is shown. The catalytic tetramer bound to the crossover sites (the “catalytic module”) is represented as an orange oval, and the eight resolvase subunits bound at the accessory sites (the “regulatory module”) are collectively represented by the pink oval. Chapter 10, this volume, gives more details on the structures of this and other synaptic complexes.
doi:10.1128/microbiolspec.MDNA3-0046-2014.f7

Subunit rotation

Pioneering electron microscopy studies by Cozzarelli's group revealed the precise topologies of a series of minor resolvase reaction products (Fig. 8a). These were proposed to be made by repeated rounds of strand exchange equivalent to half-turns of one pair of DNA ends relative to the other, in an intermediate with double-strand breaks in both crossover sites (58, 59, 60). The changes in DNA linkage that accompany the first round of the series (the standard resolution reaction) and its reverse reaction (catenane fusion) were determined and are consistent with this "simple rotation" mechanism (61). However, the simple hypothesis that the recombinase subunits attached to the half-sites rotate along with the DNA ends ("subunit rotation"; Fig. 8b) was difficult for many to accept, because of its radical biochemical implication; one half of the recombinase tetramer must rotate through 180° relative to the other half, but somehow disastrous dissociation of the two halves must be avoided. There is no biochemical precedent for this model; it was a "unicorn in

the garden," which would require extraordinarily rigorous testing.

A synapse with the recombining crossover sites on the outside of a recombinase tetramer core ("DNA-out") was argued to be most consistent with the subunit rotation model (61). Later, alternative models that retained a fixed tetramer structure (and thus avoid the dissociation issue) were proposed. Some models placed the two recombining DNA double helices close to each other near the center of the tetramer ("DNA-in"). However, it was very difficult then to account for the observed topological changes after DNA strand exchange. Another model proposed that part of the DNA-out tetramer remains fixed, while the N-terminal parts of two subunits rotate with their attached DNA ends (38, 31).

A strange property of serine recombinase-mediated recombination, first discovered in the Gin DNA invertase system, led to strong experimental support for subunit rotation. If two recombining sites have different 2-bp sequences at their central "overlap," the recombinants that would have mismatched base pairs are not

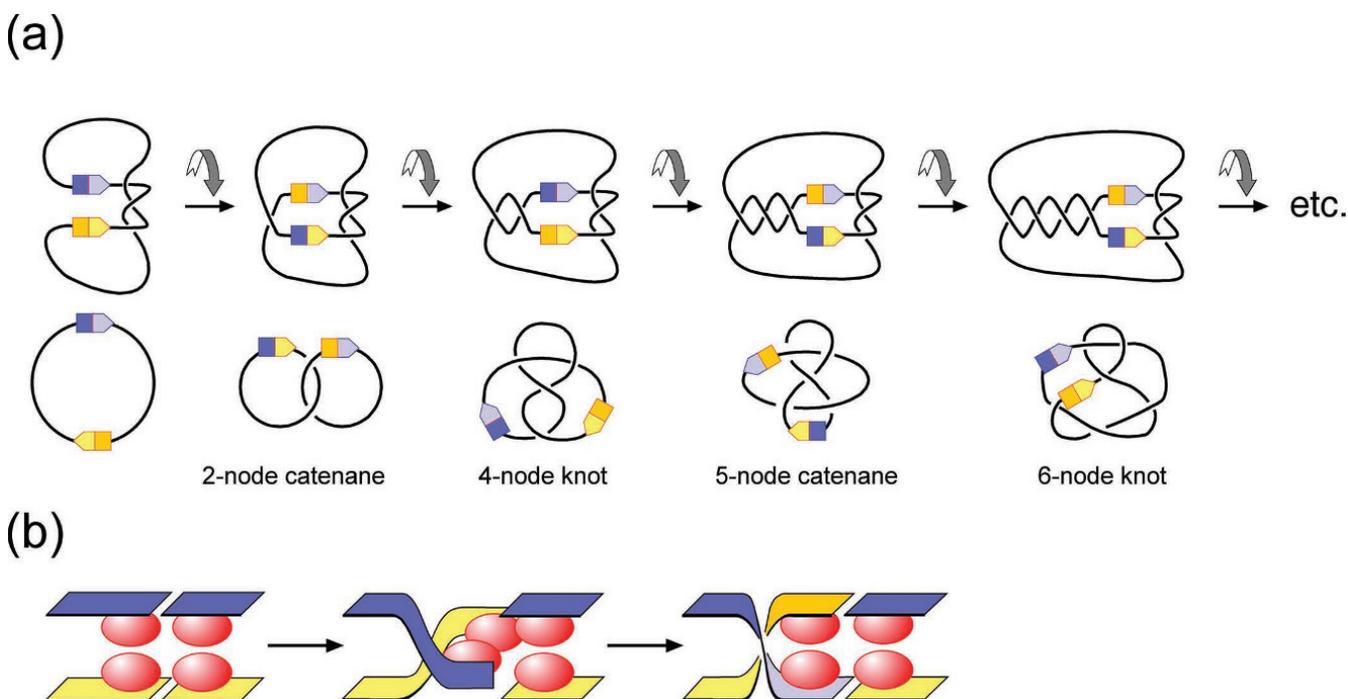


Figure 8 Subunit rotation mechanism of resolvase. (a) Topologies of first round and "iteration products" observed by Cozzarelli's group (58, 59, 60). The upper part shows the products predicted by a rotation mechanism in a resolvase synapse with topology as shown in Fig. 7. The lower panels show the simplified topologies of these products. "Mismatched" substrates (see text) form only the nonrecombinant knot products, starting with the 4-noded knot. (b). Cartoon illustrating the proposed subunit rotation mechanism. DNA is represented as ribbons and recombinase subunits as ovals. The crystal structure of a proposed intermediate in subunit rotation is shown in Fig. 6b.
doi:10.1128/microbiolspec.MDNA3-0046-2014.f8

formed (see above). Instead, a second round of strand exchange ensues, restoring the ends to a nonrecombinant configuration but leaving a record of the transaction as a change in the DNA topology (knotting) (62, 63, 64, 65). This behavior has since been shown to be general to many if not all serine recombinases, and *per se* suggests subunit rotation. In the case of resolvase, reaction of a “mismatched” substrate leads to a 4-noded knot product, consistent with a 360° rotation, and further double rounds of strand exchange give more complex knotted products (65; Fig. 8a). Further analysis of the products supported subunit rotation, but not alternative mechanisms (66, 67, 68, 69, 70). It was shown that knotting of a mismatched substrate proceeds with the DNA linkage changes predicted for subunit rotation, that the recombinase subunits move in concert with the DNA ends to which they are bound and that the knotting reaction of a mismatched substrate proceeds without any intermediate rejoining of the DNA ends that would allow a “reset” of the protein subunits, as would be necessary for all nonrotary mechanisms.

The dimer interface seen in the early $\gamma\delta$ resolvase crystal structures (37, 40) is quite rugged and apparently incompatible with subunit rotation. The structural breakthrough came from further crystallographic studies using “activated” $\gamma\delta$ resolvase variants. Activated serine recombinase mutants (first identified in the Gin and Cin invertase systems) have lost their dependence on regulation by accessory factors (71, 72, 73, 74). Activated resolvase variants were shown to form synapses *in vitro*, comprising two crossover sites bridged by a resolvase tetramer (75, 76), and low-resolution structural studies confirmed that the sites were bound on the outside of the tetramer (53). The resolvase–DNA cocrystal structures of Li *et al.* and Kamtekar *et al.* (41, 42) revealed a synaptic intermediate with both crossover sites cleaved at their centers and resolvase subunits covalently attached to each 5′ end via the active-site serines, in line with the earlier biochemical experiments (Figs 4 and 6b). The conformation of the resolvase SR domains is dramatically different from that in the dimer structures; the tetramer has a remarkably flat, hydrophobic surface between the two “halves” that are predicted to rotate with respect to each other (Fig. 6b). It is proposed that a flat, greasy interface is maintained throughout rotation, and structure-based modelling has demonstrated that there would be no major energy barriers to this process (41).

Biochemical studies on invertases (69, 70, 77) and serine integrases (78, 79) have provided further support for a subunit rotation mechanism, and recent crystal structures of activated Sin resolvase and Gin invertase

variants show recombinase tetramers with flat hydrophobic interfaces but different rotational relationships of the “rotating dimers” compared with the $\gamma\delta$ resolvase structures (44, 49). It looks like all serine recombinases work this way.

The Active Site

Each of the four active sites in a serine recombinase tetramer has to perform two chemical steps during a round of recombination (a strand cleavage and then a ligation). We would like to understand the mechanism of catalysis, the contributions of individual subunits, and the choreography and reversibility of the reaction steps.

Alignments of serine recombinases reveal about a dozen well-conserved polar or charged residues that might contribute to catalysis at the active site, and studies involving mutagenesis and *in vitro* biochemical analysis have identified six key residues including the nucleophilic serine (3, 80, 81). Proposed roles for these residues include those typical of phosphoryl transfer enzymes: generation of a strong base to increase the nucleophilicity of the very weakly acidic serine hydroxyl (cleavage reaction) and deoxyribose 3′-hydroxyl (re-ligation reaction) groups, stabilization of the transition state geometry and/or charge, and provision of an acid for protonation of the leaving group during cleavage and re-ligation (82). Other active-site features that must be present include interactions to guide the incoming nucleotide bearing the 3′-OH to the active site for ligation and contacts that detect base pairing (or the lack of it) in the product overlap sequence.

REGULATION OF RECOMBINATION ACTIVITY

Introduction

The “programmed” DNA rearrangements promoted by natural site-specific recombinases typically involve sophisticated regulation to ensure that strand cleavages and subsequent events happen only at the right times and places. Serine recombinase-based systems adapted for resolution, inversion and integration have evolved distinct regulatory strategies, as will be discussed in the following sections.

All site-specific recombinases must have high fidelity for their target sites; off-target reactions are very likely to be deleterious. The C-terminal HTH domains of small serine recombinases recognize sequence motifs at the ends of the crossover site, but their sequence specificity is limited (83) and DNA contacts by the SR

(catalytic) domains make a substantial additional contribution to specificity (84). Even so, some variation of the crossover site sequence is tolerated (85, 86). The observed high site specificity of the complete systems is presumably due to tight dependence of catalytic activity on cooperative assembly of all the components including accessory factors.

Formation of the “catalytic module” by bringing together two recombinase dimer-bound crossover sites may be a key regulatory step for most systems. For example, the crossover-site DNA-bound $\gamma\delta$ resolvase dimers in the crystals studied by Yang and Steitz (40) do not make a synaptic interaction despite their extremely high concentration, whereas “activated” mutants that are defective in regulation readily form tetramer-containing synaptic complexes (75, 87). This checkpoint apparently prevents wild-type resolvase catalysis until the full synaptic complex including the accessory sites and their bound subunits is correctly assembled (88) (see below).

Regulation of Resolvase Recombination

The serine resolvase systems are described in detail in Chapter 10, this volume. Recombination by resolvases takes place following formation of a specific synaptic complex involving intertwining of the *res* recombination site accessory sequences. As noted above, this complex forms only when the two *res* sites are in direct repeat in a negatively supercoiled DNA molecule. The synaptic complex also plays an important but as yet mysterious role in restricting strand exchange to a single half-turn, so that the first-round simple catenane resolution product is released and inert to further reaction (see Figs 7 and 8). The resolvase-bound accessory sites of Tn3/ $\gamma\delta$ *res* can pair and intertwine to form a “regulatory module” even in the absence of the crossover sites (“site I”) (Fig. 7b); this property has been used to impose topological selectivity on normally non-selective recombinases (such as Cre) by putting their crossover sites in place of *res* site I (89, 90). The detailed molecular architecture of the Tn3/ $\gamma\delta$ regulatory module is still unclear. However, it has been shown that a specific protein interface between resolvase subunits plays a key role in coupling accessory site synapsis to synapsis of the crossover sites and activation of catalysis (43, 88, 91, 92). One hypothesis is that crossover-site synapsis and the dramatic concomitant protein conformational changes (see “Subunit rotation” above; Fig. 6b) are brought about simply by forcing the recombinase dimers into close proximity by their interactions with accessory subunits (“extreme mass action”). Alternatively, the interactions with accessory

protein subunits might play an essential role in promoting the required conformational changes.

The arrangements of protein-binding sites within *res*-type recombination sites are quite diverse (93), as exemplified by Sin *res*, which contains just a single accessory binding site for Sin resolvase and a site for an “architectural” DNA-bending protein (HU/IHF). A crystal structure of Sin in a synaptic interaction with accessory DNA has led to a model for the complete synaptic complex formed by that system (43). The intertwining of the DNA and the contacts between Sin subunits bound at the crossover and accessory sites are strikingly similar to the corresponding features in current models of the Tn3/ $\gamma\delta$ resolvase complex. It seems plausible that many resolution systems adopt a similar strategy for activation of catalysis, despite significantly different regulatory module architectures.

Regulation of Invertase Recombination

The serine invertase systems are described in detail in Chapter 9, this volume. Most research on invertase mechanism has been on the Hin, Gin and Cin systems (33). Like resolvases, the invertases recombine at precise positions within dimer-binding crossover sites, but unlike the resolvases there are no adjacent accessory sequences. However, it was discovered that a sequence quite far from the crossover sites (called the enhancer, or *sis*) which binds an *Escherichia coli* protein FIS (factor for inversion stimulation) was essential for efficient recombination in each system (94, 95, 96, 97, 98). The lengths of DNA between the crossover sites and the enhancer could be varied without loss of activity (99). It is proposed that the invertase-bound crossover sites and FIS-bound enhancer come together to form a synaptic complex, as shown diagrammatically in Fig. 9. The molecular architecture of this complex is still incompletely understood, but structure-based models have been built following characterization of specific invertase–FIS interfaces, and it has been shown recently that Hin subunits make direct contacts with the enhancer DNA (100).

Like resolvases, invertases selectively recombine sites within the same supercoiled molecule, in a specific relative orientation, in this case inverted repeat. However, there is a notable difference. Resolvase has no activity at all on substrates with two *res* sites in an inverted repeat, whereas Hin (or Gin) invertase substrates with *hix* (or *gix*) sites in a direct repeat do not give recombinants but do undergo efficient double rounds of rotational strand exchange, giving knotted nonrecombinant products (62, 63, 64). It was concluded that the synaptic complex (Fig. 9) is formed regardless of relative

site orientation, but sites in a direct repeat are thereby misaligned in “antiparallel” such that strand exchange would give recombinants with mismatched base pairs; instead, double rounds of strand exchange give knotted nonrecombinant products (62) (see “Subunit rotation” above).

Regulation of Recombination by Large Serine Recombinases

Serine integrases and other large serine recombinases are the subject of Chapter 11 of this book. The integrases do not display topological selectivity; they adopt a quite different strategy to ensure correct product formation. The phage (*attP*) and bacterial genome (*attB*) crossover sites are not identical, and recombination results in two further nonidentical sites *attL* and *attR*, flanking the integrated prophage DNA (Fig. 10). Unlike phage tyrosine integrases related to λ Int, the large serine integrases (and other large serine recombinases such as the *Clostridium* transposase TnpX)

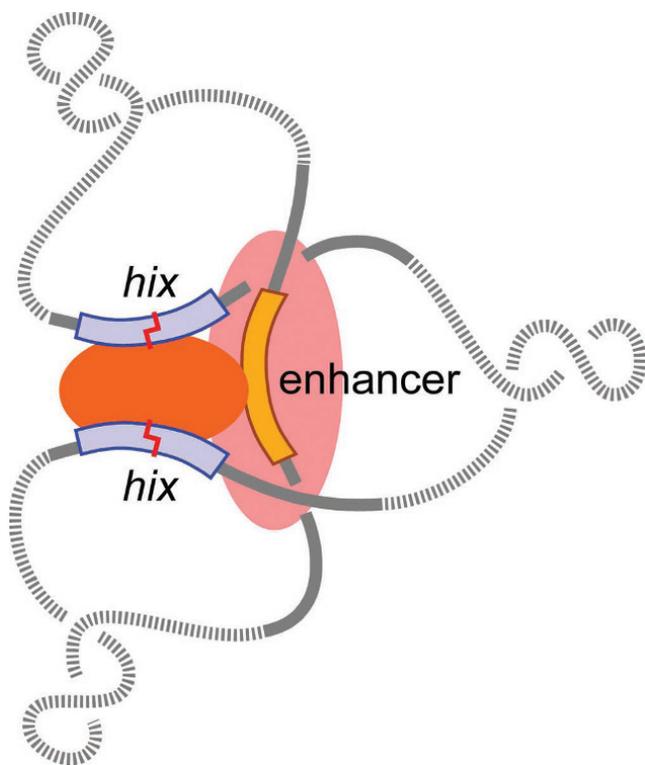


Figure 9 Cartoon of proposed inversion synaptic complex. An invertase tetramer bridging the two crossover sites is shown as an orange oval, and contacts the enhancer DNA (brown) and a FIS dimer bound there (pink). The DNA in the complex is intertwined as shown. Supercoiled loops of DNA outside the complex are shown as dashed lines.
doi:10.1128/microbiolspec.MDNA3-0046-2014.f9

apparently do not have accessory DNA sequences (35, 101, 102). The details that follow derive from *in vitro* studies on the two best-characterized serine integrases, ϕ C31 Int and Bxb1 Int (35). The *att* sites (~40 bp) each bind an integrase dimer and have an asymmetric central 2-bp overlap sequence, which has been shown to be the sole determinant of site polarity in *attP* \times *attB* recombination (103, 104). An *attP* site only recombines with an *attB* site, not with another *attP*, *attL* or *attR* site; likewise, *attB* only recombines with *attP*. Synapsis is a key selective step; only the “correct” pair of sites (*attP* and *attB*) forms a stable complex (105, 106). The integrase alone does not recombine the “lysogen” sites *attL* and *attR* at all. However, a phage-encoded recombination directionality factor (RDF) protein binds to and transforms integrase so that it efficiently and specifically recombines *attL* \times *attR*, whereas *attP* \times *attB* recombination is inhibited (107, 108, 109) (Fig. 10). Recent crystallography of the C-terminal part of A118 integrase bound to DNA (51) has led to a structure-based hypothesis for integrase *att* site selectivity (27). Unlike the small serine recombinases, integrases do not require specific connectivities between sites or DNA supercoiling, making them attractive for applications in biotechnology and synthetic biology (see “Serine recombinases in biotechnology and synthetic biology”).

PROTEINS RELATED TO SERINE RECOMBINASES

There are no other families of proteins with known functions that can be unambiguously shown to be related to the serine recombinases. The SR fold has similarity to the structures of a group of 5′-to-3′ exonucleases and to the TOPRIM domain of type IA and type II topoisomerases (110, 111), but the active site residues (and thus presumably the catalytic mechanisms) are quite different, so it is not clear that there is any relationship by descent.

SERINE RECOMBINASES IN BIOTECHNOLOGY AND SYNTHETIC BIOLOGY

Many recombinase systems have been investigated as possible tools for mediation of precise, inducible DNA rearrangements in the fields of experimental genetics, biotechnology and gene therapy (112). However, with one notable exception (ϕ C31 integrase), the utilization of serine recombinases has been relatively limited. ϕ C31 integrase has been adopted for targeted transgene integration in a number of organisms including

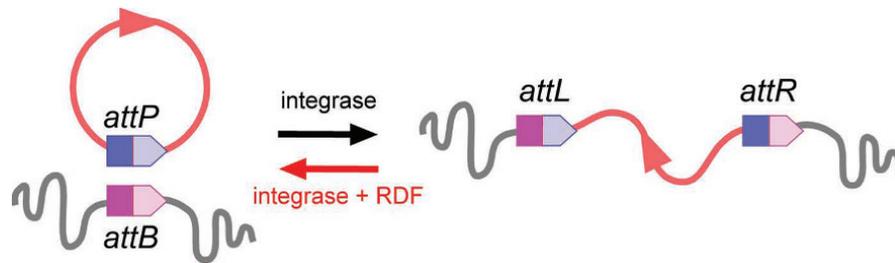


Figure 10 Recombination by serine integrases. In nature, integrase promotes recombination between a crossover site, *attP*, in the circular bacteriophage DNA and a different crossover site, *attB*, in the host bacterial genome (indicated by a squiggly line). Integrase alone does not promote any reaction between the product sites *attL* and *attR*. However, the presence of a bacteriophage-encoded recombination directionality factor (RDF) protein alters the properties of integrase so that it preferentially promotes *attL*×*attR* recombination (red arrow). See text for further details. doi:10.1128/microbiolspec.MDNA3-0046-2014.f10

humans, and it is now in widespread use in experimental research, notably in the *Drosophila* field (112, 113).

Serine recombinases are currently being exploited in the field of synthetic biology, for construction of artificial genetic switches and circuits. Recent studies have shown how all the standard Boolean logic operations can be implemented on gene expression in *E. coli* by the combined action of two orthogonal serine integrases (ϕ C31 Int and Bxb1 Int) (114). Other applications are in the assembly and manipulation of metabolic pathway genetic components (113, 115, 116).

ENGINEERING SERINE RECOMBINASES

Site-specific recombination has obvious potential as a tool for “genomic surgery” in organisms of interest to humans, but to realize this potential it will be necessary to engineer recombinases so that they recognize and act on sequences occurring in these organisms. Natural recombinases often require long complex sites, accessory factors, and DNA supercoiling, making this task seem quite daunting. However, the characterization of activated variants of small serine recombinases, which have simplified substrate requirements, has opened up engineering possibilities (71, 74, 117). The small serine recombinases are modular proteins with spatially distinct SR and HTH (DNA-binding) domains (see Fig. 6). Some sequence specificity changes were made by mutating the HTH domain or replacing it with a domain from a related serine recombinase (118, 119, 120). However, much more dramatic retargeting was achieved by linking the SR domain to a zinc-finger DNA-binding domain (121). These “zinc finger recombinases” can be

adapted to use a wide range of new “crossover sites” including natural genomic sequences, by engineering zinc-finger-domain specificity, reducing or altering the residual sequence specificity of the SR domain, or using SR domains from different recombinases (86, 122, 123, 124, 125). Recently, transcription activator-like effector (TALE) DNA-binding domains have been used instead of zinc-finger domains to retarget SR domain activity; the modularity of TALE domains and thus the ease of creating new specificities may greatly enhance the applicability of these “designer recombinases” (126).

Acknowledgments. I apologize to readers that, in this overview chapter on the serine recombinases, it has been painfully necessary for me to cover important aspects of the subject only sketchily or even not at all, and to curtail the reference list. However, much more detail on serine recombinases can be found in the three chapters in this book by Phoebe Rice, Maggie Smith and Reid Johnson. I also take this opportunity to acknowledge the many contributions of colleagues past and present to the advancement of this field, and to look forward to many more exciting developments and insights in the future.

Citation. Stark WM. 2014. The serine recombinases. *Microbiol Spectrum* 2(6):MDNA3-0046-2014.

References

1. Jayaram M, Grainge I. 2005. Introduction to site-specific recombination, p 33–81. In Mullany P (ed), *The dynamic bacterial genome*. Cambridge University Press, New York, NY.
2. Nash HA. 1996. Site-specific recombination: integration, excision, resolution and inversion of defined DNA segments, p 2363–2376. In Neidhardt FC (ed), *Escherichia coli and Salmonella typhimurium*. ASM Press, Washington, DC.

3. Grindley ND, Whiteson KL, Rice PA. 2006. Mechanism of site-specific recombination. *Ann Rev Biochem* 75: 567–605.
4. Hallet B, Sherratt DJ. 1997. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev* 21:157–178.
5. Berg DE, Howe MM (ed). 1989. *Mobile DNA*. ASM Press, Washington DC [see Chapters 1, 5, 28, 29, and 33].
6. Craig N, Craigie R, Gellert M, Lambowitz A (ed). 2004. *Mobile DNA II*. ASM Press, Washington DC [see Chapters 6 to 14].
7. Brembu T, Winge P, Tooming-Klunderud A, Nederbragt AJ, Jakobsen KS, Bones AM. 2013. The chloroplast genome of the diatom *Seminais robusta*: new features introduced through multiple mechanisms of horizontal gene transfer. *Mar Genomics* 16:17–7.
8. Piednoël M, Gonçalves IR, Higuier D, Bonniard E. 2011. Eukaryote DIRS1-like retrotransposons: an overview. *BMC Genomics* 12:621.
9. Heffron F, McCarthy BJ. 1979. DNA sequence analysis of the transposon Tn3: three genes and three sites involved in transposition of Tn3. *Cell* 18:1153–1163.
10. Arthur A, Sherratt DJ. 1979. Dissection of the transposition process: a transposon-encoded site-specific recombination system. *Mol Gen Genet* 175:267–274.
11. Kostriken R, Morita C, Heffron F. 1981. Transposon Tn3 encodes a site-specific recombination system: identification of essential sequences, genes and actual site of recombination. *Proc Natl Acad Sci U S A* 78: 4041–4045.
12. Grindley NDF, Lauth MR, Wells RG, Wityk RJ, Salvo JJ, Reed RR. 1982. Transposon-mediated site-specific recombination: identification of three binding sites for resolvase at the *res* site of $\gamma\delta$ and Tn3. *Cell* 30:19–27.
13. Kitts PA, Symington LS, Dyson P, Sherratt DJ. 1983. Transposon-encoded site-specific recombination: nature of the Tn3 DNA sequences which constitute the recombination site *res*. *EMBO J* 2:1055–1060.
14. Plasterk RHA, Brinkman A, van de Putte P. 1983. DNA inversions in the chromosome of *Escherichia coli* and in bacteriophage Mu: relationship to other site-specific recombination systems. *Proc Natl Acad Sci U S A* 80: 5355–5358.
15. Hatfull GF, Grindley NDF. 1988. Resolvases and DNA-invertases: a family of enzymes active in site-specific recombination, p 357–396. In Kucherlapati R, Smith GR (ed), *Genetic recombination*. ASM Press, Washington DC.
16. Reed RR. 1981. Transposon-mediated site-specific recombination: a defined *in vitro* system. *Cell* 25:713–719.
17. Reed RR. 1981. Resolution of cointegrates between transposons $\gamma\delta$ and Tn3 defines the recombination site. *Proc Natl Acad Sci U S A* 78:3428–3432.
18. Reed RR, Grindley NDF. 1981. Transposon-mediated site-specific recombination *in vitro*: DNA cleavage and protein-DNA linkage at the recombination site. *Cell* 25: 721–728.
19. Reed RR, Moser CD. 1984. Resolvase-mediated recombination intermediates contain a serine residue covalently linked to DNA. *Cold Spring Harbor Symp Quant Biol* 49:245–249.
20. Klippel A, Mertens G, Patschinsky T, Kahmann R. 1988. The DNA invertase Gin of phage Mu: formation of a covalent complex with DNA via a phosphoserine at amino acid position 9. *EMBO J* 7:1229–1237.
21. Hatfull GF, Grindley NDF. 1986. Analysis of $\gamma\delta$ resolvase mutants *in vitro*: evidence for an interaction between serine-10 of resolvase and site I of *res*. *Proc Natl Acad Sci U S A* 83:5429–5433.
22. Leschziner A, Boocock MR, Grindley NDF. 1995. The tyrosine-6 hydroxyl of $\gamma\delta$ resolvase is not required for the DNA cleavage and rejoining reactions. *Mol Microbiol* 15:865–870.
23. Esposito D, Scocca JJ. 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res* 25:3605–3614.
24. Kamali-Moghaddam M, Sundstrom L. 2000. Transposon targeting determined by resolvase. *FEMS Microbiol Lett* 186:55–59.
25. Smith MCM, Thorpe HM. 2002. Diversity in the serine recombinases. *Mol Microbiol* 44:299–307.
26. Rowland SJ, Stark WM. 2005. Site-specific recombination by the serine recombinases, p 121–150. In Mullany P (ed), *The dynamic bacterial genome*. Cambridge University Press, New York, NY.
27. Van Duyne GD, Rutherford K. 2013. Large serine recombinase domain structure and attachment site binding. *Crit Rev Biochem Mol Biol* 48:476–491.
28. Thorpe HM, Smith MCM. 1998. *In vitro* site-specific integration of bacteriophage DNA catalysed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A*, 95:5505–5510.
29. Kersulyte D, Mukhopadhyay AK, Shirai M, Nakazawa T, Berg DE. 2000. Functional organization and insertion specificity of IS607: a chimeric element of *Helicobacter pylori*. *J Bacteriol* 182:5300–5308.
30. Boocock MR, Rice PA. 2013. A proposed mechanism for IS607-family transposases. *Mobile DNA* 4:24.
31. Grindley NDF. 2002. The movement of Tn3-like elements: transposition and cointegrate resolution, p 272–302. In Craig N, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA II*. ASM Press, Washington DC.
32. Sherratt D. 1989. Tn3 and related transposable elements: site-specific recombination and transposition, p 163–184. In Berg DE, Howe MM (ed), *Mobile DNA*. ASM Press, Washington DC.
33. Johnson RC. 2002. Bacterial site-specific DNA inversion systems, p 230–271. In Craig N, Craigie R, Gellert M, Lambowitz A (ed) *Mobile DNA II*. ASM Press, Washington DC.
34. Groth AC, Calos MP. 2004. Phage integrases: biology and applications. *J Mol Biol* 335:667–678.
35. Smith MCM, Brown WRA, McEwan AR, Rowley PA. 2010. Site-specific recombination by ϕ C31 integrase and other large serine recombinases. *Biochem Soc Trans* 38:388–394.

36. Askora A, Kawasaki T, Fujie M, Yamada T. 2011. Resolvase-like serine recombinase mediates integration/excision in the bacteriophage ϕ RSM. *J Biosci Bioeng* 111:109–116.
37. Sanderson MR, Freemont PS, Rice PA, Goldman A, Hatfull GF, Grindley NDF, Steitz TA. 1990. The crystal structure of the catalytic domain of the site-specific recombination enzyme $\gamma\delta$ resolvase at 2.7 Å resolution. *Cell* 63:1323–1329.
38. Rice PA, Steitz TA. 1994. Model for a DNA-mediated synaptic complex suggested by crystal packing of $\gamma\delta$ resolvase subunits. *EMBO J* 13:1514–1524.
39. Rice PA, Steitz TA. 1994. Refinement of $\gamma\delta$ resolvase reveals a strikingly flexible molecule. *Structure* 2: 371–384.
40. Yang W, Steitz TA. 1995. Crystal structure of the site-specific recombinase $\gamma\delta$ resolvase complexed with a 34 bp cleavage site. *Cell* 82:193–207.
41. Li W, Kamtekar S, Xiong Y, Sarkis GJ, Grindley NDF, Steitz TA. 2005. Structure of a synaptic $\gamma\delta$ resolvase tetramer covalently linked to two cleaved DNAs. *Science* 309:1210–1215.
42. Kamtekar S, Ho RS, Cocco MJ, Li W, Wenwieser SVCT, Boocock MR, Grindley NDF, Steitz TA. 2006. Implications of structures of synaptic tetramers of $\gamma\delta$ resolvase for the mechanism of recombination. *Proc Natl Acad Sci U S A* 103:10642–10647.
43. Mouw KW, Rowland SJ, Gajjar MM, Boocock MR, Stark WM, Rice PA. 2008. Architecture of a serine recombinase-DNA regulatory complex. *Mol Cell* 30: 145–155.
44. Keenholtz RA, Rowland SJ, Boocock MR, Stark WM, Rice PA. 2011. Structural basis for catalytic activation of a serine recombinase. *Structure* 19:799–809.
45. Pan B, Maciejewski MW, Marintchev A, Mullen GP. 2001. Solution structure of the catalytic domain of $\gamma\delta$ resolvase. Implications for the mechanism of catalysis. *J Mol Biol* 310:1089–1107.
46. Liu T, DeRose EF, Mullen GP. 1994. Determination of the structure of the DNA binding domain of $\gamma\delta$ resolvase in solution. *Protein Sci* 3:1286–1295.
47. Feng JA, Johnson RC, Dickerson RE. 1994. Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science* 263:348–355.
48. Chiu TK, Sohn C, Dickerson RE, Johnson RC. 2002. Testing water-mediated DNA recognition by the Hin recombinase. *EMBO J* 21:801–814.
49. Ritacco CJ, Kamtekar S, Wang J, Steitz TA. 2012. Crystal structure of an intermediate of rotating dimers within the synaptic tetramer of the G-segment invertase. *Nucleic Acids Res* 41:2673–2682.
50. Ritacco CJ, Steitz TA, Wang J. 2014. Exploiting large non-isomorphous differences for phase determination of a G-segment invertase–DNA complex. *Acta Cryst D* 70: 685–693.
51. Rutherford K, Yuan P, Perry K, Sharp R, Van Duyne GD. 2013. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res* 41:8341–8356.
52. Yuan P, Gupta K, Van Duyne GD. 2008. Tetrameric structure of a serine integrase catalytic domain. *Structure* 16:1275–1286.
53. Nöllmann M, He J, Byron O, Stark WM. 2004. Solution structure of the Tn3 resolvase-crossover site synaptic complex. *Mol Cell* 16:127–137.
54. Aravind L, Landsman D. 1998. AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* 26:4413–4421.
55. Krasnow MA, Cozzarelli NR. 1983. Site-specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented *res* sites. *Cell* 32:1313–1324.
56. Stark WM, Boocock MR, Sherratt DJ. 1989. Site-specific recombination by Tn3 resolvase. *Trends Genet* 5: 304–309.
57. Stark WM, Boocock MR. 1995. Topological selectivity in site-specific recombination, p 101–129. In Sherratt DJ (ed), *Mobile genetic elements*. Frontiers in Molecular Biology series. Oxford University Press, Oxford, UK.
58. Wasserman SA, Cozzarelli NR. 1985. Determination of the stereostructure of the product of Tn3 resolvase by a general method. *Proc Natl Acad Sci U S A* 82: 1079–1083.
59. Wasserman SA, Dungan JM, Cozzarelli NR. 1985. Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science* 229:171–174.
60. Wasserman SA, Cozzarelli NR. 1986. Biochemical topology: applications to DNA recombination and replication. *Science* 232:951–960.
61. Stark WM, Sherratt DJ, Boocock MR. 1989. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell* 58:779–790.
62. Kanaar R, Klippel A, Shekhtman E, Dungan JM, Kahmann R, Cozzarelli NR. 1990. Processive recombination by the phage Mu Gin system: implications for the mechanism of DNA exchange DNA site alignment, and enhancer action. *Cell* 62:353–366.
63. Heichman KA, Moskowitz IPG, Johnson RC. 1991. Configuration of DNA strands and mechanism of strand exchange in the Hin invertasome as revealed by analysis of recombinant knots. *Genes Dev* 5:1622–1634.
64. Moskowitz IP, Heichman KA, Johnson RC. 1991. Alignment of recombination sites in Hin-mediated site-specific recombination. *Genes Dev* 5:1635–1645.
65. Stark WM, Grindley NDF, Hatfull GF, Boocock MR. 1991. Resolvase-catalysed reactions between *res* sites differing in the central dinucleotide of subsite I. *EMBO J* 10:3541–3548.
66. Stark WM, Boocock MR. 1994. The linkage change of a knotting reaction catalysed by Tn3 resolvase. *J Mol Biol* 239:25–36.
67. McIlwraith MJ, Boocock MR, Stark WM. 1996. Site-specific recombination by Tn3 resolvase, photocrosslinked to its supercoiled DNA substrate. *J Mol Biol* 260:299–303.
68. McIlwraith MJ, Boocock MR, Stark WM. 1997. Tn3 resolvase catalyses multiple recombination events

- without intermediate rejoining of DNA ends. *J Mol Biol* 266:108–121.
69. Dhar G, Sanders ER, Johnson RC. 2004. Architecture of the *Hin* synaptic complex during recombination: the recombinase subunits translocate with the DNA strands. *Cell* 119:33–45.
 70. Dhar G, McLean MM, Heiss JK, Johnson RC. 2009. The *Hin* recombinase assembles a tetrameric protein swivel that exchanges DNA strands. *Nucleic Acids Res* 37:4743–4756.
 71. Klippel A, Cloppenborg K, Kahmann R. 1988. Isolation and characterization of unusual *gin* mutants. *EMBO J* 7:3983–3989.
 72. Haffter P, Bickle TA. 1988. Enhancer-independent mutants of the *Cin* recombinase have a relaxed topological specificity. *EMBO J* 7:3991–3996.
 73. Klippel A, Kanaar R, Kahmann R, Cozzarelli NR. 1993. Analysis of strand exchange and DNA binding of enhancer-independent *Gin* recombinase mutants. *EMBO J* 12:1047–1057.
 74. Arnold PH, Blake DG, Grindley NDF, Boocock MR, Stark WM. 1999. Mutants of *Tn3* resolvase which do not require accessory binding sites for recombination activity. *EMBO J* 18:1407–1414.
 75. Sarkis GJ, Murley LL, Leschziner AE, Boocock MR, Stark WM, Grindley NDF. 2001. A model for the $\gamma\delta$ resolvase synaptic complex. *Mol Cell* 8:623–631.
 76. Leschziner AE, Grindley NDF. 2003. The architecture of the $\gamma\delta$ resolvase crossover site synaptic complex revealed by using constrained DNA substrates. *Mol Cell* 12:775–781.
 77. Dhar G, Heiss JK, Johnson RC. 2009. Mechanical constraints on *Hin* subunit rotation imposed by the *Fis*/enhancer system and DNA supercoiling during site-specific recombination. *Mol Cell* 34:746–759.
 78. Bai H, Sun M, Ghosh P, Hatfull GF, Grindley NDF, Marko JF. 2011. Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase. *Proc Natl Acad Sci U S A* 108:7419–7424.
 79. Olorunniji FJ, Buck DE, Colloms SD, McEwan AR, Smith MCM, Stark WM, Rosser SJ. 2012. Gated rotation mechanism of site-specific recombination by ϕ C31 integrase. *Proc Natl Acad Sci U S A* 109:19661–19666.
 80. Olorunniji FJ, Stark WM. 2009. The catalytic residues of *Tn3* resolvase. *Nucleic Acids Res* 37:7590–7602.
 81. Keenholtz RA, Mouw KW, Boocock MR, Li NS, Piccirilli JA, Rice PA. 2013. Arginine as a general acid catalyst in serine recombinase-mediated DNA cleavage. *J Biol Chem* 288:29206–29214.
 82. Mizuuchi K, Baker TA. 2002. Chemical mechanisms for mobilizing DNA, 12–23. In Craig N, Craigie R, Gellert M, Lambowitz A (ed) *Mobile DNA II*. ASM Press, Washington DC.
 83. Abdel-Meguid SS, Grindley NDF, Templeton NS, Steitz TA. 1984. Cleavage of the site-specific recombination protein $\gamma\delta$ resolvase: the smaller of the two fragments binds DNA specifically. *Proc Natl Acad Sci U S A* 81:2001–2005.
 84. Rimphanitchayakit V, Grindley NDF. 1990. Saturation mutagenesis of the DNA site bound by the small carboxy-terminal domain of $\gamma\delta$ resolvase. *EMBO J* 9:719–725.
 85. Wells RG, Grindley NDF. 1984. Analysis of the $\gamma\delta$ *res* site: sites required for site-specific recombination and gene expression. *J Mol Biol* 179:667–687.
 86. Proudfoot CM, McPherson AL, Kolb AF, Stark WM. 2011. Zinc finger recombinases with adaptable DNA sequence specificity. *PLoS ONE* 6:e19537.
 87. Olorunniji FJ, He J, Wenwieser SVCT, Boocock MR, Stark WM. 2008. Synapsis and catalysis by activated *Tn3* resolvase mutants. *Nucleic Acids Res* 36:7181–7191.
 88. Rowland SJ, Boocock MR, McPherson AL, Mouw KW, Rice PA, Stark WM. 2009. Regulatory mutations in *Sin* recombinase support a structure-based model of the synaptosome. *Mol Microbiol* 74:282–298.
 89. Watson MA, Boocock MR, Stark WM. 1996. Rate and selectivity of synapsis of *res* recombination sites by *Tn3* resolvase. *J Mol Biol* 257:317–329.
 90. Kilbride E, Boocock MR, Stark WM. 1999. Topological selectivity of a hybrid site-specific recombination system with elements from *Tn3 res/resolvase* and bacteriophage P1 *loxP/Cre*. *J Mol Biol* 289:1219–1230.
 91. Hughes RE, Hatfull GF, Rice PA, Steitz TA, Grindley NDF. 1990. Cooperativity mutants of the $\gamma\delta$ resolvase identify an essential interdimer interaction. *Cell* 63:1331–1338.
 92. Murley LL, Grindley NDF. 1998. Architecture of the $\gamma\delta$ resolvase synaptosome: oriented heterodimers identify interactions essential for synapsis and recombination. *Cell* 95:553–562.
 93. Rowland SJ, Stark WM, Boocock MR. 2002. *Sin* recombinase from *Staphylococcus aureus*: synaptic complex architecture and transposon targeting. *Mol Microbiol* 44:607–619.
 94. Huber HE, Iida S, Arber W, Bickle TA. 1985. Site-specific DNA inversion is enhanced by a DNA sequence element in *cis*. *Proc Natl Acad Sci U S A* 82:3776–3780.
 95. Hubner P, Arber W. 1989. Mutational analysis of a prokaryotic recombinational enhancer with two functions. *EMBO J* 8:577–585.
 96. Johnson RC, Simon MI. 1985. *Hin*-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp enhancer. *Cell* 41:781–791.
 97. Kahmann R, Rudt F, Koch C, Mertens G. 1985. G inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. *Cell* 41:771–780.
 98. Bruist MF, Glasgow AC, Johnson RC, Simon MI. 1987. *Fis* binding to the recombinational enhancer of the *Hin* DNA inversion system. *Genes Dev* 1:762–772.
 99. Haykinson MJ, Johnson RC. 1993. DNA looping and the helical repeat *in vitro* and *in vivo*: effect of HU protein and enhancer location on *Hin* invertasome assembly. *EMBO J* 12:2503–2512.
 100. McLean MM, Chang Y, Dhar G, Heiss JK, Johnson RC. 2013. Multiple interfaces between a serine

- recombinase and an enhancer control site-specific DNA inversion. *eLife* 2:e01211.
101. Adams V, Lucet IS, Lyras D, Rood JI. 2004. DNA binding properties of TnpX indicate that different synapses are formed in the excision and integration of the Tn4451 family. *Mol Microbiol* 53:1195–1207.
 102. Misiura A, Pigli YZ, Boyle-Vavra S, Daum RS, Boocock MR, Rice PA. 2013. Roles of two large serine recombinases in mobilizing the methicillin-resistance cassette SCCmec. *Mol Microbiol* 88:1218–1229.
 103. Ghosh P, Kim AI, Hatfull GF. 2003. The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol Cell* 12:1101–1111.
 104. Smith MCA, Till R, Smith MCM. 2004. Switching the polarity of a bacteriophage integration system. *Mol Microbiol* 51:1719–1728.
 105. Ghosh P, Pannunzio NR, Hatfull GF. 2005. Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites. *J Mol Biol* 349:331–348.
 106. Smith MCA, Till R, Brady K, Soutanas P, Thorpe H, Smith MCM. 2004. Synapsis and DNA cleavage in ϕ C31 integrase-mediated site-specific recombination. *Nucleic Acids Res* 32:2607–2617.
 107. Stark WM. 2011. Cutting out the ϕ C31 prophage. *Mol Microbiol* 80:1417–1419.
 108. Ghosh P, Wasil LR, Hatfull GF. 2006. Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol* 4:e186.
 109. Khaleel T, Younger E, McEwan AR, Varghese AS, Smith MCM. 2011. A phage protein that binds ϕ C31 integrase to switch its directionality. *Mol Microbiol* 80:1450–1463.
 110. Artymiuk P, Ceska TA, Suck D, Sayers JR. 1997. Prokaryotic 5'-3' exonucleases share a common core structure with gamma-delta resolvase. *Nucleic Acids Res* 25:4224–4229.
 111. Yang W. 2010. Topoisomerases and site-specific recombinases: similarities in structure and mechanism. *Crit Rev Biochem Mol Biol* 45:520–534.
 112. Bischof J, Basler K. 2008. Recombinases and their uses in gene activation, gene inactivation, and transgenesis. *Methods Mol Biol* 420:175–195.
 113. Fogg PCM, Colloms SD, Rosser SJ, Stark WM, Smith MCM. 2014. New applications for phage integrases. *J Mol Biol* 426:2703–2716.
 114. Maranhao AC, Ellington AD. 2013. Endowing cells with logic and memory. *Nat Biotechnol* 31:413–415.
 115. Zhang L, Zhao G, Ding X. 2011. Tandem assembly of the epothilone biosynthetic gene cluster by *in vitro* site-specific recombination. *Sci Rep* 1:141.
 116. Colloms SD, Merrick CA, Olorunniji FJ, Stark WM, Smith MCM, Osbourn A, Keasling JD, Rosser SJ. 2014. Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res* 42:e23.
 117. Burke ME, Arnold PH, He J, Wenwieser SVCT, Rowland SJ, Boocock MR, Stark WM. 2004. Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol Microbiol* 51:937–948.
 118. Grindley NDF. 1993. Analysis of a nucleoprotein complex: the synaptosome of $\gamma\delta$ resolvase. *Science* 262:738–740.
 119. Ackroyd AJ, Avila P, Parker CN, Halford SE. 1990. Site-specific recombination by mutants of Tn21 resolvase with DNA recognition functions from Tn3 resolvase. *J Mol Biol* 216:633–643.
 120. Schneider F, Schwikardi M, Muskhelishvili G, Dröge P. 2000. A DNA-binding domain swap converts the invertase Gin into a resolvase. *J Mol Biol* 295:767–775.
 121. Akopian A, He J, Boocock MR, Stark WM. 2003. Chimeric site-specific recombinases with designed DNA sequence recognition. *Proc Natl Acad Sci U S A* 100:8688–8691.
 122. Gordley RM, Gersbach CA, Barbas CF III. 2009. Synthesis of programmable integrases. *Proc Natl Acad Sci U S A* 106:5053–5058.
 123. Sirk SJ, Gaj T, Jonsson A, Mercer AC, Barbas CF III. 2014. Expanding the zinc-finger recombinase repertoire: directed evolution and mutational analysis of serine recombinase specificity determinants. *Nucleic Acids Res* 42:4755–4766.
 124. Akopian A, Stark WM. 2005. Site-specific recombinases as instruments for genomic surgery. *Adv Genet* 55:1–23.
 125. Gaj T, Sirk SJ, Barbas CF III. 2014. Expanding the scope of site-specific recombinases for genetic and metabolic engineering. *Biotechnology and Bioengineering* 111:1–15.
 126. Mercer AC, Gaj T, Fuller RP, Barbas CF III. 2012. Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Res* 40:11163–11172.

Arthur Landy¹

The λ Integrase Site-specific Recombination Pathway

4

INTRODUCTION

The λ site-specific recombination pathway has enjoyed the sequential attentions of geneticists, biochemists, and structural biologists for more than 50 years. It has proven to be a rewarding model system of sufficient simplicity to yield a gratifying level of understanding within a single (fortuitously timed) professional career, and of sufficient complexity to engage a small cadre of scientists motivated to peel this onion. The initiating highlight of the genetics phase was the insightful proposal by Allan Campbell for the pathway by which the λ chromosome integrates into, and excises from, the *Escherichia coli* host chromosome (1). The breakthrough for the biochemical phase was the purification of λ integrase (Int) and the integration host factor (IHF) by Howard Nash (2, 3). The first major step in the structural phase was the cocrystal structure of IHF bound to its DNA target site by Phoebe Rice and Howard Nash (4). Although the crystal structure of naked Fis protein had been determined earlier (5, 6), the full impact of Fis on understanding the fundamentals of the Int reaction did not come until much later (7, 8).

λ Integrase is generally regarded as the founding member of what is now called the tyrosine recombinase family, even though many family members are not strictly recombinases. Family membership is defined by

the creation of novel DNA junctions via an active site tyrosine that cleaves and reseals DNA through the formation of a covalent 3'-phospho-tyrosine high-energy intermediate without the requirement for any high-energy cofactors. Other important, well studied, and highly exploited family members each have their own chapter in this volume of *Mobile DNA III*. Limitations on space prevent the inclusion in this chapter of the many other interesting family members, which comprise a wide range of biological functions and interesting variations on the themes discussed here, including other well-studied members of the heterobivalent subfamily, such as Tn916 (9), HP1 (10), and L5 (11). For previous reviews that include sections on the tyrosine recombinase family and λ Int see references 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, and 27. In this review, I will try to emphasize as much as possible those features of the λ Int pathway that have been reported since, or were not the focus of, earlier reviews, an intention that will consequently highlight recent advances in structural aspects of the pathway.

OVERVIEW OF THE REACTION

The λ Int recombination pathway has evolved to provide a conditional, effectively irreversible, DNA switch

¹Division of Biology and Medicine, Brown University, Providence, RI.

in the life cycle of the virus. The “cost” (complexity) associated with regulated directionality in the λ Int pathway is what distinguishes it from its Cre and Flp siblings (Fig. 1). As in most of the family members, each recombining partner DNA contains a pair of inverted repeat recombinase binding sites (called core-type sites) that flank a 7 bp overlap region (O) (6 to

8 bp in other systems) that is identical in both DNAs. (Evolution of new core-type and overlap DNA sequences has been proposed to proceed by low frequency λ phage insertions at sites other than the canonical *attB* [28].) DNA cleavage and exchange of the top strands on one side of the overlap region by two active Ints creates a four-way DNA junction

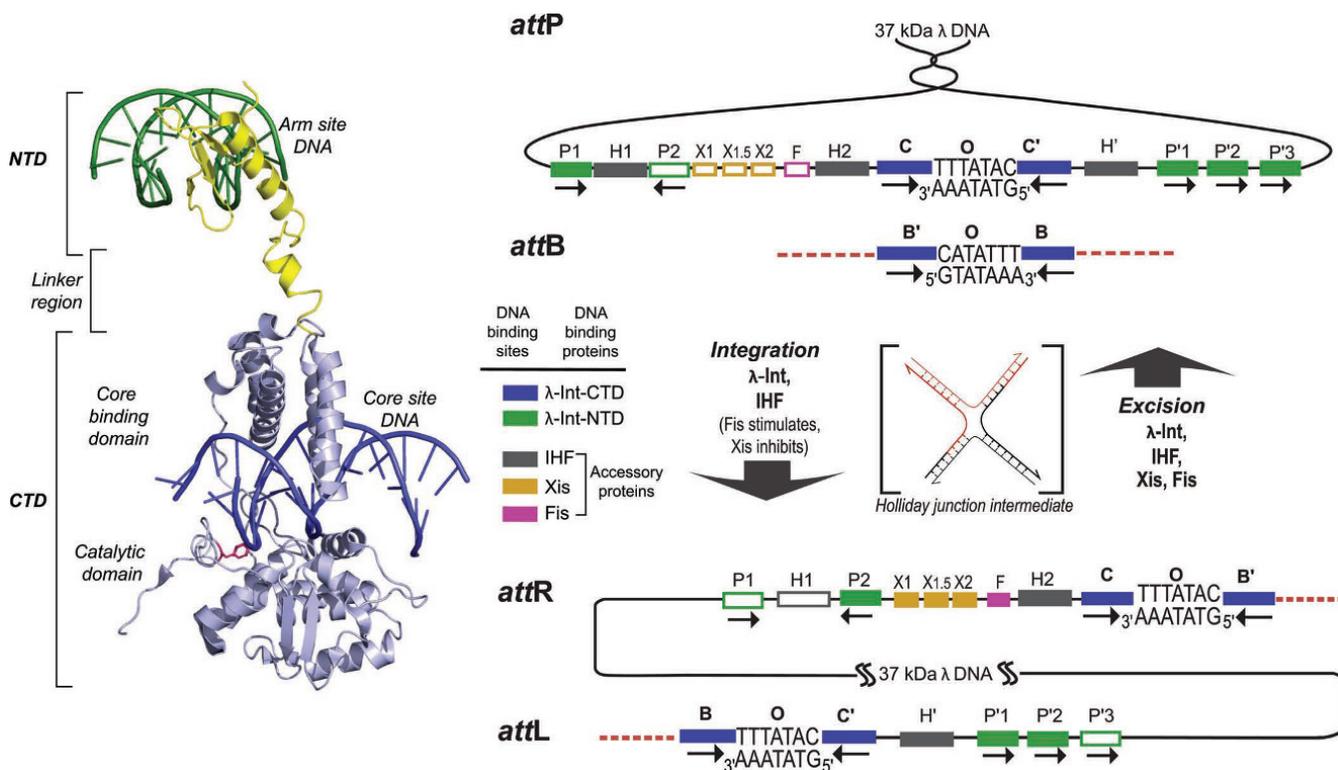


Figure 1 λ Integrase and the overlapping ensembles of protein binding sites that comprise *att* site DNA. The left panel shows the structure of a single λ Int protomer bound via its NTD to an arm site DNA and via its CTD to a core site DNA (adapted from the Int tetrameric structure determined by Biswas *et al.* [44], PDB code 1Z1G). The right panel shows the recombination reactions. Integrative recombination between supercoiled *attP* and linear *attB* requires the virally encoded integrase (Int) (2) and the host-encoded accessory DNA binding protein integration host factor (IHF) (4, 177) and gives rise to an integrated phage chromosome bounded by *attL* and *attR*. Excisive recombination between *attL* and *attR* to regenerate *attP* and *attB* additionally requires the phage-encoded Xis protein (which inhibits integrative recombination) (140) and is stimulated by the host-encoded Fis protein (8). Both reactions proceed through a Holliday junction intermediate that is first generated and then resolved by single strand exchanges on the left and right side of the 7 bp overlap region, respectively. The two reactions proceed with the same order of sequential strand exchanges (not the reverse order) and use different subsets of protein binding sites in the P and P' arms, as indicated by the filled boxes: Int arm-type P1, P2, P'1, P'2, and P'3 (green); integration host factor (IHF), H1, H2, and H' (gray); Xis, X1, X1.5, and X2 (gold); and Fis (pink). The four core-type Int binding sites, C, C', B, and B' (blue boxes) are each bound in a C-clamp fashion by the CB and CAT domains, referred to here as the CTD. This is where Int executes isoenergetic DNA strand cleavages and ligations via a high-energy covalent 3'-phosphotyrosine intermediate. The CTD of Int and the tetrameric Int complex surrounding the two overlap regions are functionally and structurally similar to the Cre, Flp, and XerC/D proteins. Reprinted with permission from reference 36. doi:10.1128/microbiolspec.MDNA3-0051-2014.f1

[Holliday junction (HJ)] that is then resolved to recombinant products by the remaining pair of Ints cleaving and exchanging the bottom strands on the other side of the overlap region (Fig. 2). Appended to two of the four core-type sites are additional DNA sequences that encode binding sites for the second (NTD) DNA binding domain of Int and the accessory DNA bending proteins, IHF, Xis, and Fis. As indicated by color coding in Fig. 1, some sites are required only for integrative recombination between *attP* (on the phage chromosome) and *attB* (on the bacterial chromosome), some are required only for excisive recombination between the *attL* and *attR* sites (flanking the integrated prophage), and some sites are required for both reactions. For more detail, see reference 27. It has been suggested that the additional complexity of the λ pathway evolved to

regulate the directionality of recombination in response to the physiological state of the host cell (29), a notion that is now well documented in latent human viruses, such as the ubiquitous herpes virus, cf., "... the [herpes] viral genome evolved to sense the infection status of the host... through highly evolved pathogen genomes with the capacity to sense host cytokines..." (30).

HOLLIDAY JUNCTION INTERMEDIATES

A hallmark of the tyrosine recombinase family, discussed here in terms of the λ pathway, is the formation of a four-way DNA junction (HJ) intermediate. For a long time, it was thought that this was a very unstable intermediate because it was difficult to identify without designing elaborate substrates (31, 32, 33). Only many

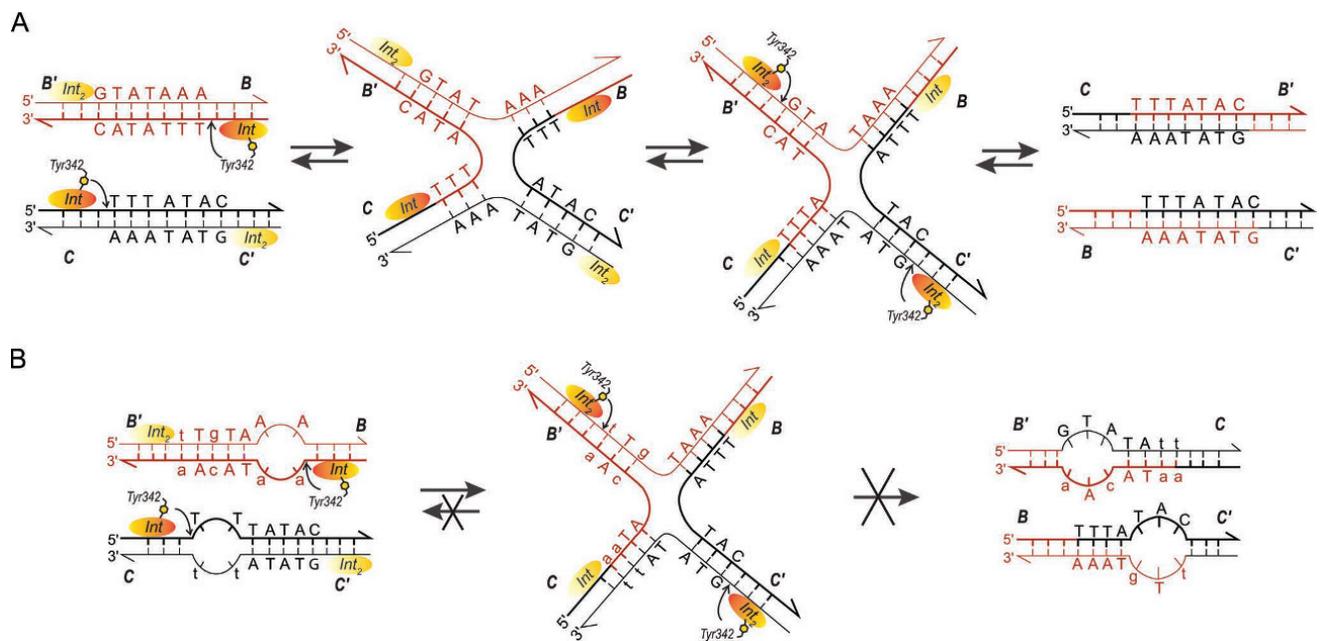


Figure 2 Formation, resolution, and trapping of Holliday junctions (HJ). (A) The top strand of each *att* site is cleaved via formation of a high-energy phosphotyrosine intermediate and the strands are exchanged (three bases are “swapped”) to form the HJ, thus, creating a branch point close to the center of the overlap regions. A conformational change of the complex that slightly repositions the branch point and more extremely repositions the Int protomers leads to the second swap of DNA strands and resolution of the HJ to helical products (44, 178). These features of the reaction suggested the mechanism-based method of trapping HJ complexes shown in (B). The left panel shows the DNA sequence changes made in the 7 bp overlap regions to trap HJ intermediates (lower case letters). Following the first pair of Int cleavages (via the active site Tyr) on one side of the overlap regions (arranged here in antiparallel orientation), the “top” strands are swapped to form the HJ; this simultaneously converts the unpaired (bubble) bases to duplex DNA. On the other side, the sequence differences between the two overlap regions strongly disfavor the second (“bottom”) strand swap that would resolve the HJ, because this would generate unpaired bubbles in the product complex (36, 37, 38, 39). This diagram applies to both integrative and excisive recombination (even though the labels refer to integrative recombination). Adapted in part, with permission, from reference 36. doi:10.1128/microbiolspec.MDNA3-0051-2014.f2

years later was it discovered that the standard sodium dodecyl sulfate (SDS) treatment employed for visualizing naked HJ DNA gave misleading results because SDS fails to quench Int ligation activity fast enough to prevent reformation of the initial phosphodiester bonds (see also below) (34).

A striking feature of HJ formation in the Int reaction is that it is always initiated by cleavage and exchange of the same (“top”) strands in both integrative and excisive recombination (one of the facts indicating that the two reactions are not the reverse of one another) (Fig. 2) (31, 32, 33, 35). These features of the reaction suggested a mechanism-based method of trapping HJ complexes (outlined in Fig. 2) that would prove useful in studies of the complete higher-order complexes (described below) (36, 37, 38, 39).

Based upon structural snapshots from X-ray crystallographic models, patterns of amino acid residues in the active sites, mutational studies, and biochemical analyses, it is supposed there are only small differences between the pathways of λ Int, Cre, and Flp at the level of HJ formation, structure, and resolution, and likely only minor differences in their respective chemistries of DNA cleavage and ligation. One exception to this generalization is the manner in which the active site tyrosine nucleophile is delivered to the active site. In the case of Flp it is delivered in *trans*, that is, the tyrosine of one protomer in the tetramer is activated as a nucleophile in the active site of its adjacent neighbor (40, 41). While in both Int and Cre the tyrosine nucleophile is in *cis*, its proper positioning within the active site depends upon the nature of an interprotomer interaction between adjacent protomers within the tetrameric complex (42, 43, 44, 45), as discussed further below. In the absence of the NTD DNA binding domain Int can efficiently resolve HJs but it cannot carry out a recombination reaction (discussed further below) (46, 47, 48, 49). It is also clear from mutational analyses that there are Int residues that are specifically critical for HJ resolution but not DNA cleavage (50).

HEXAPEPTIDE INHIBITORS

In a bold and formidable effort to find recombination inhibitors that would trap the HJ intermediate, Anca Segall and her collaborators used deconvolution of synthetic hexapeptide libraries to search for hexapeptides that would block recombination subsequent to the first HJ-forming strand exchange (51, 52). Their most potent peptide inhibitor, WRWYCR, whose active form is a dimer assembled via a disulfide bridge between two peptide monomers, stably traps HJ complexes in all

pathways mediated by Int as well as Cre (53, 54). Using this inhibitor, they were able to study the kinetics of HJ resolution under several different conditions and in several different Int-mediated pathways (55, 56). One of their conclusions from these studies was that spermidine stabilizes the “second” HJ isomeric form (the precursor to product formation) (57). Application of a hexapeptide inhibitor to studying the *Bacteriodes* NBU1 recombination pathway revealed that IntN1 recombinase is surprisingly more efficient when it forms HJs in the presence of mismatches, although their resolution to products does require homology (58).

In vitro, the hexapeptides inhibit a range of enzymes involving tyrosine-mediated transesterification, such as vaccinia virus topoisomerase and *E. coli* topoisomerase I (59). Subsequently, they were shown to be bactericidal to both Gram positive and Gram negative bacteria, presumably because they can interfere with DNA repair and chromosome dimer resolution by XerC/D. They were also shown to inhibit the excision of several different prophages *in vivo* (60). The *in vivo* successes of the hexapeptide inhibitors motivated the Segall group to search for therapeutically more useful small molecules with similar activities. Indeed, a search of over nine million compounds yielded one potentially interesting compound with properties that suggested the possible value of further searches for functional analogs of the hexapeptide inhibitors (61).

KINETICS

To overcome the difficulty of distinguishing kinetically relevant intermediates from off-pathway species, single molecule experiments were used to determine how binding energy from the multiple protein-DNA interactions is used to achieve efficiency and directionality in the overall Int recombination pathway (34). Protein binding (i.e., associated DNA bending), synapsis between *attL* and *attR*, HJ formation, and recombination were all monitored by changes in the length of a 1353 bp DNA that served as a diffusion-limiting tether of a microscopic bead to the flow chamber bed of a video-enhanced light microscope. In these experiments it was found that stable bent-DNA complexes containing Int, IHF, and Xis form rapidly (<20 s) and independently on *attL* and *attR*, and synapsis under these conditions is extremely rapid (1.0 min^{-1}). These single molecule experiments strongly suggest there are no intrinsic mechanistic features of the pathway that make synapsis slow. While Int-mediated DNA cleavage, before or immediately after synapsis, is required to stabilize the synaptic complexes, those complexes that synapsed ($\sim 50\%$ of

the total) yield recombinant with an impressive ~100% efficiency. The rate-limiting step of excision occurs after synapsis, but closely precedes or is concomitant with the appearance of a stable HJ. This single molecule result is consistent with the observation that in solution rates of stable HJ formation are similar to the rates of excisive recombination (62).

Given the reversibility of the underlying chemistry of recombination, the apparent irreversibility observed in these experiments of each step of the reaction (except for synapsis) is notable. This result indicates that the overall directionality of excisive recombination is a direct consequence of the sequence of protein–protein and protein–DNA interactions that efficiently drive the reaction forward through nearly every step. It was proposed that the slow step in the reaction is some conformational change that stabilizes the HJ (34). Candidates for this rate-limiting step, such as the scissoring movement of the HJ arms, the shift in the localized bend of the HJ, or the reorientation of the active and inactive pairs of Int protomers, are suggested by comparison of the different X-ray crystal structures of tyrosine family recombinases complexed with their respective four-armed DNAs (15, 22, 43, 44, 63).

A totally different aspect of the kinetics of recombination concerns the process by which λ DNA, once inside the cell, finds its cognate *attB* site. Surprisingly, λ DNA does not carry out an active search but rather remains confined to the point where it entered the cell; it is the directed motion of the bacterial DNA during chromosome replication that delivers *attB* to a waiting, relatively stationary, *attP* (64).

STRUCTURE OF THE Int CTD

Among the most significant recent advances in our understanding of λ Int recombination were those emanating from the X-ray crystallographic studies by the Ellenberger laboratory (44, 45, 65). The second Int fragment to be used by the Ellenberger laboratory for X-ray crystallography, lacked the NTD (arm binding domain) and consisted of residues 75 to 356 (45). Referred to as C75 in the literature and here called the CTD, it corresponds to the two domains comprising the well-studied monovalent family members such as Cre, Flp, and XerC/D. The λ Int CTD is not competent for recombination but it is an efficient topoisomerase, binds weakly to single core-type DNA sites, and resolves λ *att* site HJs (48, 66, 67). The weak binding of the λ Int CTD to single core-type sites was circumvented by trapping covalent Int-*att* site complexes with a “flapped” suicide substrate containing a nick within

the overlap region, three bases from the scissile phosphate (Fig. 3A).

As shown in Fig. 3, the λ Int CTD consists of a catalytic domain that is joined to the central binding (CB) domain by a flexible, interdomain linker, residues I160–R176, that is extremely sensitive to proteolytic degradation (45, 48). The CB and catalytic domains of Int both contribute to recognition of the core site, although the former, whose structure has also been determined (68), confers most of the sequence specificity (69, 70). Only two residues from each domain (K95 and N99 in the CB domain and K235 and R287 in the catalytic domain) directly form hydrogen bonds with DNA bases. Interestingly, one of them, K95, interacts with a base, Gua30, that is absent in the B' site, the weakest of the four core sites (71). The base-specific interactions are consistent with the effects of mutations of these and nearby residues that affect DNA binding specificity (72, 73, 74).

In comparison to the monovalent family members, Cre and Flp (41, 42, 43), the λ Int CTD displays fewer hydrogen bonds and total direct contacts to DNA bases in both its amino- and carboxy-terminal domains. Additionally, the extended unstructured interdomain linker of λ Int appears to be more flexible than the Cre linker (43), suggesting an increase in entropic cost of binding to DNA. Indeed, the helpful and informative *int-h* mutant (E174K), which substitutes a lysine in the middle of the interdomain linker adjacent to the site of DNA cleavage, increases the DNA binding affinity of λ Int and relaxes or eliminates the requirement for IHF during recombination (75, 76). It was proposed that the substituted lysine might enhance DNA binding affinity by contributing a stabilizing interaction with DNA, and/or by constraining the movement of the interdomain linker (45).

A comparison of the structure of the CTD Int covalently bound to DNA (45) with that of the unliganded catalytic domain (65), revealed that the tyrosine342 nucleophile had moved approximately 20 Å into the active site where it forms a 3'-phosphotyrosine linkage with the cleaved DNA (Fig. 4). Additionally, in the tetrameric complex, the eight carboxy-terminal residues (349 to 356) of a protomer extend away from the protein and pack against a neighboring protomer, contributing in *trans* an additional strand (β 7) to the sheet formed by strands β 1, β 2, and β 3 of the catalytic domain. This *trans* packing arrangement of β 7 is required for appropriate placement of the Tyr342 nucleophile into the active site. This fact, in conjunction with the phenotypes of a number of Int mutants, suggests a dual role for the alternative stacking arrangements of β 7.

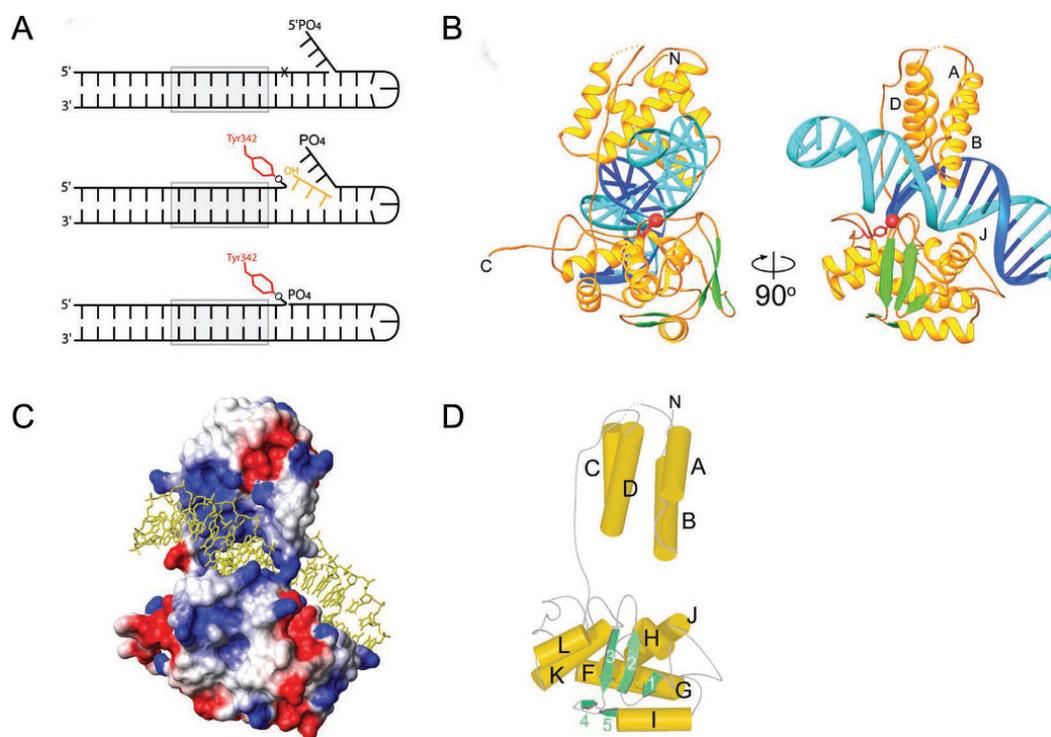


Figure 3 X-ray crystal structure of the Int CTD. (A) With this modified version of previously designed suicide recombination substrates (35, 47) covalently trapped CTD-DNA complexes were stable for weeks. Formation of the phosphotyrosine bond and diffusion of the three base oligonucleotide is followed by annealing of the three base flap to the three nucleotide gap, thus, positioning the 5'-phosphate such that it repels water and shields the phosphotyrosine linkage from hydrolysis. (B) Ribbon diagrams showing the central domain (residues 75 to 160; above the DNA) and the catalytic domain (residues 170 to 356; below the DNA) of λ Int, and their interactions with the major and minor grooves on the opposite sides of the DNA. A long, extended linker (residues I160 to R176) connects these domains. The scissile phosphate that is covalently linked to Y342 is shown as a red sphere. The central domain inserts into the major groove adjacent to the site of DNA cleavage. The catalytic domain makes interactions with the major and minor groove on the opposite side of the DNA, straddling the site of DNA cleavage. (C) The solvent accessible surface of the Int protein is shown, colored according to electrostatic potential. The DNA binding surface is highly positive (blue) and makes numerous interactions with the phosphates of the DNA (cf. Figure 3B). The polypeptide linker between domains joins the central and catalytic domains on one side of the DNA. A salt bridge between the N ζ of K93 and the carbonyl oxygen of S234 bridges between domains on the other side of the DNA, completing the ring-shaped structure that encircles the DNA. (D) The architecture of the λ Int C-75 protein is shown with cylinders and arrows representing helices and β strands, respectively. This view is oriented similarly to that in (A) (right side). The central domain of λ Int lacks helix E, corresponding to the fifth helix of Cre's N-terminal domain, which is involved in subunit interactions. Reprinted with permission from reference 45.
doi:10.1128/microbiolspec.MDNA3-0051-2014.f3

This also suggested an attractive explanation for the findings that a carboxy-terminal deletion of seven residues (commencing with Trp350), and mutations involving residues in or around $\beta 7$ (all of which were expected to untether the Tyr342) abolished recombination but enhanced the topoisomerase activity of monomers (77, 78, 79). Because these same mutations

decrease recombinase activity, the C-terminal tail could also be important in coordinating the catalytic activities of adjacent protomers, as seen in the X-ray crystal structure of the tetrameric higher order recombination complex (Fig. 4C) (44, 80).

In contrast to the large movement of the Tyr342 nucleophile in transitioning from the unliganded to the

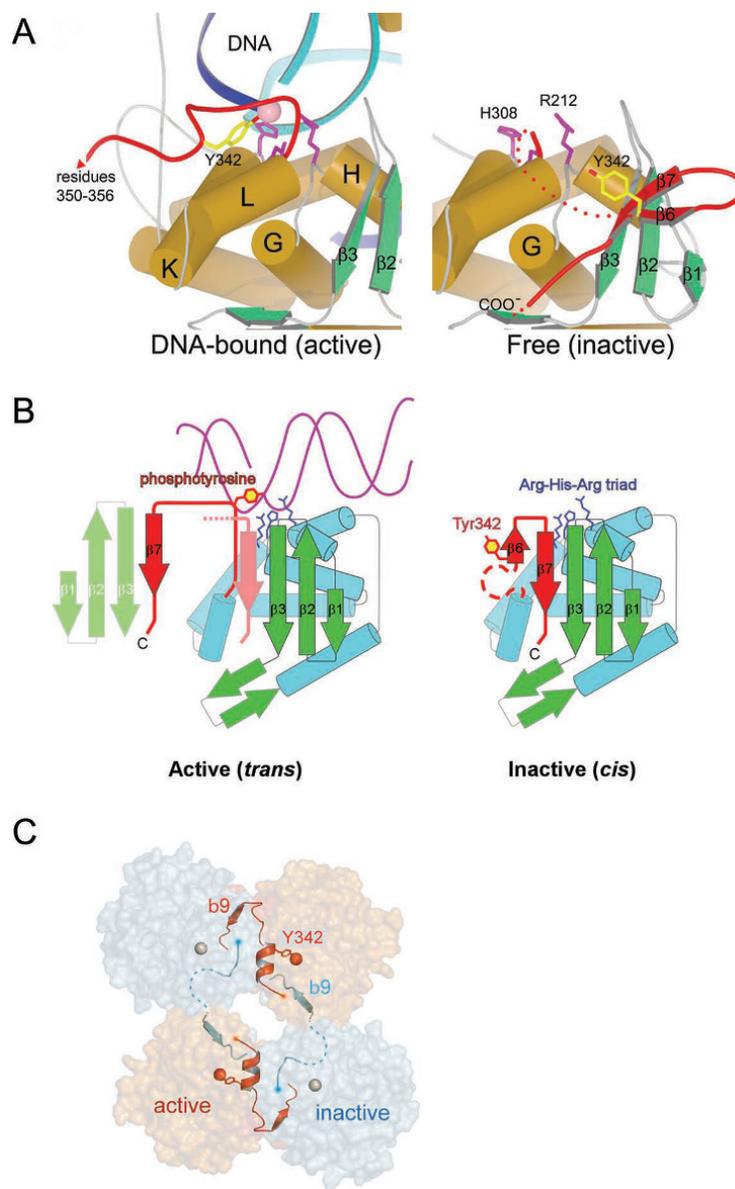


Figure 4 A Remodeling of Int's active site switches DNA cleavage activity on and off. (A) A comparison of the DNA-bound (left), and unbound (right), structures of λ Int shows a dramatic reorganization of the C-terminal region spanning residues 331 to 356 (red). In the absence of DNA, Y342 (yellow) is far from the catalytic triad of R212, H308, and R311 (magenta side chains). In the DNA complex (left panel), Y342 has moved into the active site. Another consequence of the DNA-bound conformation is that the extreme C-terminal residues 349 to 356 extend away from the parent Int molecule and pack against another molecule in *trans*. (B) A cartoon illustrating how the DNA-bound conformation of Int positions the Y342 for cleavage of DNA. The isomerization from the inactive form, in which Y342 is held some distance from the catalytically important Arg212-His308-Arg311 triad (65), to the active conformation seen in complex with DNA, is accompanied by the release of strand β 7 and its repacking in *trans* against a neighboring molecule. (C) The assembly of active (orange) and inactive (gray) catalytic sites results from a skewed packing arrangement of λ Int subunits (residues 75 to 356) in the tetramer. The scissile phosphates bound by active and inactive subunits are shown as red and gray spheres, respectively. Reprinted with permission from references 44 and 45. doi:10.1128/microbiolspec.MDNA3-0051-2014.f4

liganded Int, the other four catalytically important residues, R212, K235, H308, and R311, show less than 1 Å movement on average between the two structures, as is also true for most of the other residues in the catalytic domain. The role of these residues in catalysis was established by mutational analyses of several tyrosine recombinase family members, biochemical analyses (especially of topoisomerase I), sequence comparisons of other family members, and shortly thereafter, comparisons with the X-ray crystal structures of other DNA-bound family members (41, 43, 81, 82, 83, 84, 85, 86, 87, 88).

ROLE OF THE Int NTD

The following experiments were carried out to prove it was possible to “de-tune” a monovalent recombinase, for example, Cre, and convert it to a regulated unidirectional recombinase by appending an NTD (89). Cre recombinase is bidirectional, unregulated, does not require accessory proteins, and has a minimal symmetric DNA target. Rather than de-tuning the Cre recombinase its DNA target was attenuated: a single base pair change, previously shown to weaken the interaction between Cre and its DNA binding site (90) was introduced into each of the inverted repeat Cre binding sites and the DNA sequence and spacing between the DNA cleavage sites (the “overlap” region) was changed to the canonical seven base pair sequence of the λ *att* sites. λ P and P' arms were appended to the modified Cre target sites to generate analogs of the four λ *att* sites.

To complete the recombination pathway, a gene fusion encoding the first 74 residues of λ Int was fused to Cre. The resulting chimeric Cre protein product carried out recombinations between the analogs of the four λ *att* sites with all of the properties of canonical λ Int-dependent pathways: reactions were dependent upon IHF, Xis was required for the excision reaction but inhibited the integration reaction, integrative recombination required the P1 but not the P2 sites, and the excisive reaction required P2 but not P1 (cf. Fig. 1).

It appears from these experiments that the regulated directionality of the λ Int pathway has been conferred on Cre by the appended 74 N-terminal residues of λ Int coupled with the reduction in DNA binding efficiency between Cre and its DNA target sites. These experiments suggest that two simple steps, in no specified order, are all that is required for the evolution of the heterobivalent recombinases from their monovalent siblings. However, they do not rule out an alternative evolutionary trajectory in which the monovalent and heterobivalent site-specific recombinases evolved in parallel from a common, less efficient, precursor.

While the NTD of λ Int was able to confer regulated directionality on the Cre recombinase, it is possible, and even likely, that not all of the λ NTD functions were revealed in these experiments. For example, effects resulting from any interactions between the NTD and the CTD were not studied in those experiments and they would not likely even be manifest in the hybrid protein. One example of such interactions came from studies on the context-dependent effects of the NTD. These studies were prompted by the unexpected finding that the Int CTD (residues 65 to 356, called C65) is more active as a topoisomerase, in binding to core-type sites, cleaving DNA, and resolving synthetic Holliday junctions, than the full length Int. In other words, the NTD is an inhibitor of the primary Int functions (49). Equally surprising was the fact that when the cloned and purified NTD (residues 1 to 65) was added to the cloned and purified CTD, it stimulated all of the primary Int functions, well beyond the levels observed for either CTD or full length Int. In other words, when present in *cis* (i.e., in full length Int), the NTD is an inhibitor of Int functions, but when present in *trans*, it is a stimulator. Resolution of the apparent paradox came with the finding that addition of an oligonucleotide encoding the arm-type DNA sites (P'1–P'2) to full length Int abolished the *cis* NTD inhibition and resulted in the formation of a ternary complex between Int and core and arm-type DNAs.

These results led to the hypothesis of an enhanced dual role for the DNA bending accessory proteins. In addition to their structural function in facilitating the Int-mediated arm-core bridges that comprise the higher-order structure of recombinogenic complexes, they should also be viewed as a requirement to overcome the N-domain inhibition of recombinase functions (49). These data and the resultant hypothesis are consistent with the finding of mutants in one domain that effect the activity of the other (72, 77), and the important observation of Richet *et al.* that Int does not bind well to *attB* unless it part of a higher-order *attP* complex (91).

Residues Met1 to Leu64 comprise the minimal Int fragment that binds to arm-type sites and it does so with almost the same efficiency as full length Int (66, 92). However, an additional six residues are required (Met1 to Ser70) for cooperative binding to the adjacent arm-type sites P'1, P'2, and P'3. The greatest cooperativity in binding, which is between sites P'2 and P'3, depends upon the single bp between them and is resistant to an unopposed three base bulge in the top strand but not in the bottom strand. The asymmetric effect of the unopposed bulge is consistent with DNA bending

upon Int binding to the P' arm sites. Int's affinity for the single sites P'1 or P1 exceeds its net affinity for P'2–P'3 (44, 93). It is interesting that the two lowest affinity arm-type sites, P2 and P'3 are each required for only one of the two recombination reactions, excision and integration, respectively, and are also the outermost sites in their respective pathways. Int binding at P2 is greatly enhanced by its cooperativity with Xis binding at X1, and Int binding at P'3 is enhanced (to a lesser extent) by its cooperativity with Int binding at P'2, thus, rendering the excisive reaction very sensitive to Xis concentration and the integrative reaction more sensitive to Int concentration (66). The latter fit nicely with a very early observation by Enquist *et al.* that integrative recombination is more sensitive than excisive recombination to decreased intracellular levels of Int (94). The Int 1-70 NTD is also equally as competent as full length Int for cooperative interactions with Xis when the two are bound at P2 and X1, respectively (66).

STRUCTURE OF THE NTD

The first view of the NTD structure came from a nuclear magnetic resonance (NMR) analysis of the Met1-Leu64 peptide, which revealed a fold structurally related to the three-stranded β -sheet family of DNA-binding domains. However, it was supplemented with a disordered 10 residue amino-terminal basic tail, that was shown to be important for arm binding by its loss of function upon removing a single positive charge (G2K Δ 2R) (92). The importance and role of the amino-terminal basic tail was clearly shown in the subsequent NMR structure of the NTD in complex with its DNA target site (95). Only two other proteins containing this fold have been visualized in complex with their DNA targets: the N-terminal domains from the Tn916 Int protein (96) and from the ethylene responsive factor from *Arabidopsis thaliana* (AtERF1) (97). All three proteins recognize DNA via their unique three-stranded antiparallel β -sheet that is inserted into the major groove of their respective DNA targets. The smaller size of the β -sheet-DNA interface in the λ NTD, relative to the other two proteins, is presumably compensated by the additional contacts of the 11 residue amino-terminal tail that projects deep into the minor groove (95).

STRUCTURE OF A FULL Int TETRAMER COMPLEX

A structural view of the full λ Int did not come until it was cocrystallized with DNA bound at the NTD and

CTD, recognition domains for the arm- and core-type DNA sites, respectively. These studies by the Ellenberger laboratory were particularly informative because they represented Int-DNA complexes at three different steps along the recombination pathway (44). One of the structures, a synaptic complex between two COC' core-type sites bound by four CTDs (residues 75 to 356), represented an early step after the first DNA cleavage but before strand exchange. A second structure, with full length Ints in which the cleaved strands had exchanged but ligation was prevented by a modified DNA substrate, represented a post strand-exchange complex. And the third structure was a synthetic Holliday junction intermediate bound by four full length Ints, carrying the Tyr342Phe mutation, that were thus unable to cleave the DNA into products. In the last two structures, the NTDs of the full length Ints were bound to short oligonucleotides containing tandem P'1–P'2 arm-type DNA binding sites. It is likely that the presence of this arm-type DNA occupying the NTD domains was a critical factor in the successful crystallization of the full length Int, and additionally imposed a facilitating (albeit unnatural) 2-fold symmetry. The other factor critical for crystallization was the stable tetrameric arrangement of protomers within each complex.

The tetrameric complexes with full length Int assemble into three distinct layers. The NTD (residues 1 to 63) that binds to arm-type sites is joined to the core-binding domain (CB domain; residues 75 to 175) by a short α -helical segment (residues 64 to 74), and this, in turn, is connected to the C-terminal catalytic domain (residues 176 to 356) through another linker (residues 160 to 176). Together, the three domains of each Int form an ensemble that engages the core and arm DNA targets to form a tightly knit but flexible tetrameric complex (Fig. 5) (44).

The four NTDs are bound by two antiparallel arm DNAs that slightly bend towards each other, with each pair binding the adjacent P'1–P'2 binding sites. The basic N-terminal segment (residues 2 to 10), that was disordered in the NMR structure but shown to be required for recombination activity (92), tucks into the minor groove adjacent to the 3' side of the arm-type consensus sequence (44).

As noted above, the CB and catalytic domains (which are referred to together as the CTD in this review) are structurally analogous to the full-length monovalent tyrosine recombinases, Cre (42, 43), Flp (41), and XerC/D (98). Thus, it is not surprising that λ Int has a catalytic pocket that resembles the other family members with nearly identical conserved residues (Arg212, Lys235, His308, Arg311, and His333) that

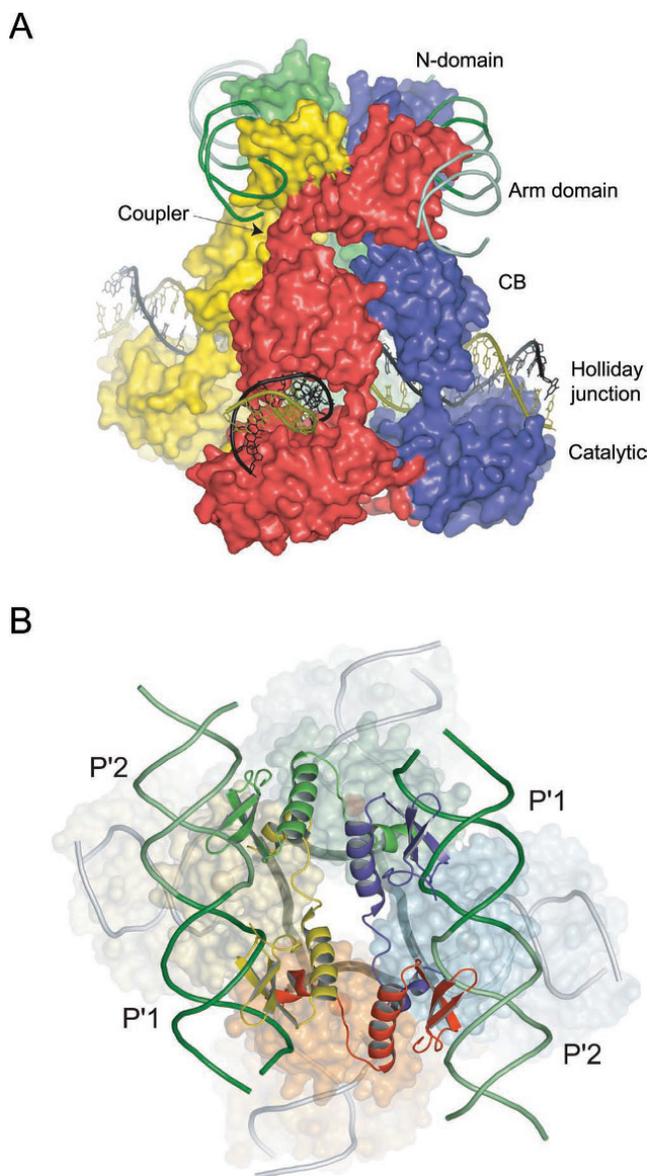


Figure 5 Structure of the λ Int tetramer bound to a Holliday junction and arm DNAs. (A) The domains of Int pack together as three stacked layers, with the NTDs cyclically swapped onto neighboring subunits. The NTD layer embraced by two antiparallel arm DNAs is linked through short α -helical couplers to the CTD, which encircles the branches of the Holliday junction. The active subunits are colored red/green and the inactive subunits are blue/yellow. (B) The 2-fold symmetry of the NTD layer is reflected in the skewed arrangement of the CTDs and the shape of the four-way junction (thick dark gray lines) in the bottom strands reactive isomer. Reprinted with permission from reference 44.

doi:10.1128/microbiolspec.MDNA3-0051-2014.f5

engage the scissile phosphate and Tyr342 nucleophile (44, 45, 65).

Among the factors likely to contribute to λ CTD's lack of recombination function, is the linker (residues 160 to 176) between the CB and catalytic domains. In contrast to the analogous linker in Cre, it lacks the α E helix that contributes many intersubunit interactions that stabilize the Cre tetramer (42, 43). Consequently, the loosely packed CB domains of the λ Int tetramer are able to rotate against each other by as much as 30° in the different isomers that were crystallized (44).

It was particularly interesting that each of the three independent crystal structures determined by Biswas *et al.* (44) illustrates a different conformation of the core DNAs and different subunit packing interactions (Fig. 6). The skewed packing of protomers generates two very different subunit interfaces comprising active versus inactive catalytic sites. In the former, the Tyr342 helix is well ordered and stabilized by electrostatic interactions with two catalytically essential residues. In the latter, the β 9 is incompatible with these stabilizing interactions and the region around Tyr342 is disordered (see also Fig. 4C). It should be noted that an α -helical conformation around Tyr342, that was not seen in the active conformation of the earlier crystal structure of the λ CTD (residues 75 to 356) (45), was confirmed by additional crystal structures (in the presence of orthovanadate) to likely be the true active conformation (44).

In the crystal structure of the synaptic, prestrand-exchange, complex, the tetramer deviates strongly from 4-fold symmetry: the scissile phosphates (which can be visualized as the corners of a parallelogram) of the cleaved DNA strands are 39 \AA apart while those of the uncleaved strands are 50 \AA apart (Fig. 6A, D). This translational offset brings the cleaved $5'$ ends closer to the phosphotyrosine of the synapsing partner, thus, facilitating strand exchange and ligation. In the post strand-exchange complex, the core DNAs resemble a HJ intermediate with approximate four-fold symmetry. Here the kink has moved to a more central position, 4 bp away from the cleaved site, bringing the cleavage sites of the bottom strands closer together, and possibly disfavoring reversal of the top strand cleavage (Fig. 6B, E). In the complex with a synthetic preformed HJ, the crossover point was fixed three nucleotides from one pair of cleavage sites, and consequently, these sites are used preferentially for resolution (67, 99). This complex is also highly skewed such that the scissile phosphates, bound by the active protomers, are brought close together (Fig. 6C, F). Although not apparent in the crystal structures, mutational analyses also reveal

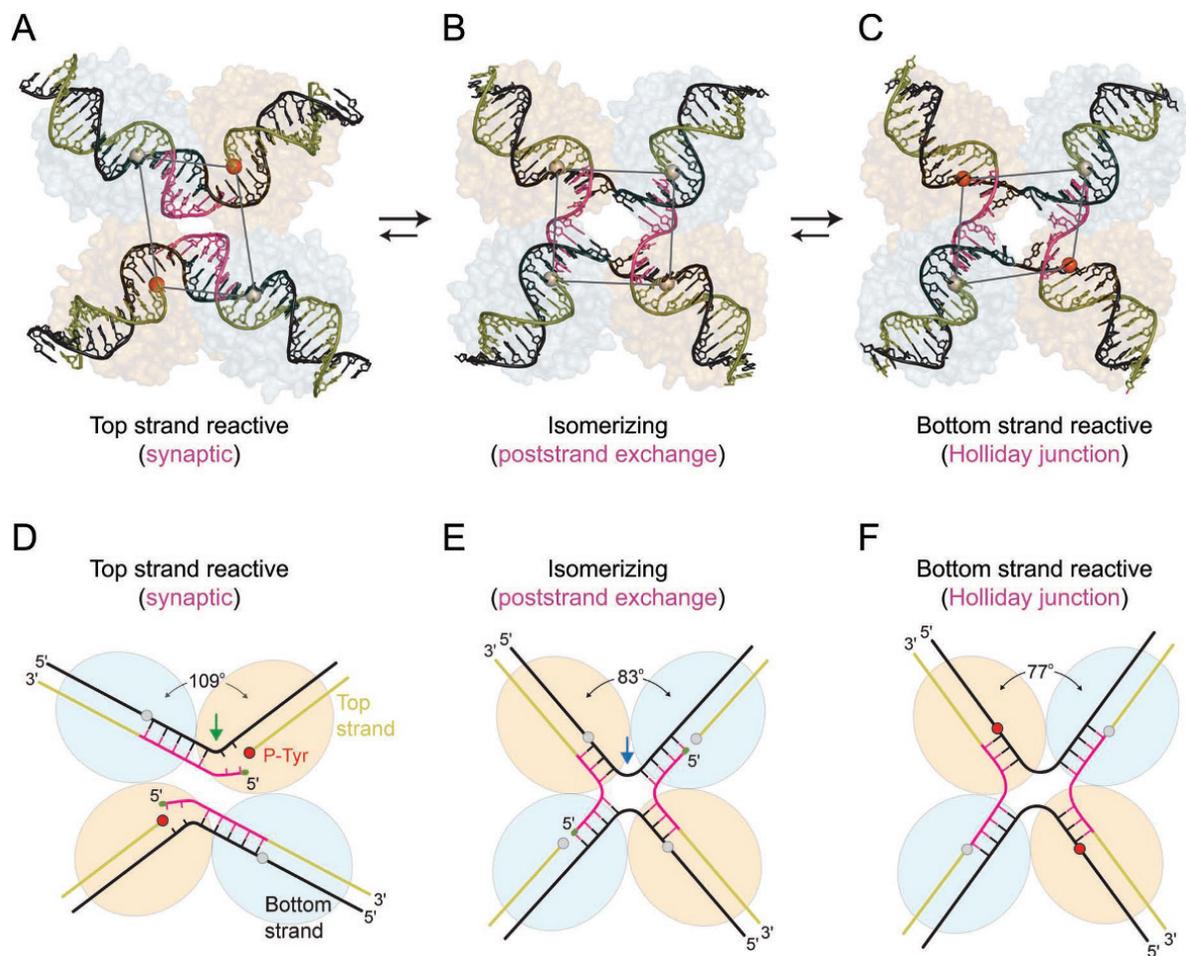


Figure 6 Three different conformations of λ Int tetramers representing distinct steps of the recombination reaction. The core DNAs within the λ -Int⁽⁷⁵⁻³⁵⁶⁾ synaptic complex (A, D), the λ -Int post-strand exchange complex (B, E), and the λ -Int Holliday junction complex (C, F) are shown along with schematic diagrams illustrating the interbranch angles and position of branch points. The pair of Int subunits in the active conformation (orange/red) is positioned closer to the center of each complex, whereas the inactive pair of subunits (gray) is further apart. Scissile phosphates (spheres) activated for cleavage are colored in red. Reprinted with permission from reference 44. doi:10.1128/microbiolspec.MDNA3-0051-2014.f6

nonequivalent interactions between the NTDs of neighboring Int protomers during HJ resolution (100).

One of the features of the tetrameric Int crystal structures, which was also inferred from solution studies of these small complexes (101), is a cyclically permuted topology, in which each NTD packs on top of the neighboring CB domain. It is now thought that this 2-fold symmetric NTD arrangement does not reflect that of a bona fide (integrative or excisive) recombinogenic complex, but rather is a consequence of the symmetric arm-type sites that are not connected by DNA and bending proteins to the core region, as discussed further below.

INTEGRATION HOST FACTOR

Integration host factor (IHF) was discovered in the very early studies of λ site-specific recombination by virtue of its role as a host-encoded protein that was essential both *in vitro* and *in vivo* for integrative and excisive recombination (102, 103). Its specific architectural role was demonstrated by the observation that IHF bending at the H' site could stimulate Int binding and cleavage at the low-affinity C' core site (104). As has not infrequently been the case, an *E. coli* protein discovered for its role in a phage life cycle, turned out to be an important player in the physiology of the cell. IHF is involved in regulation of gene transcription, especially

σ^{54} promoters (105), initiation of DNA replication (106), transposition (107), and phage packaging (108) (for reviews see references 109, 110, 111, 112, 113, 114, 115, 116).

To a large extent, the role and mechanisms of IHF in λ site-specific recombination, the crystal structure of IHF in complex with its DNA target, and the ways in which the physiology and biology of IHF in the host cell can impact recombination *in vivo*, have been reported and discussed prior to, and within, a previous review of λ site-specific recombination (27). More recent studies of IHF have centered on details of its interaction with DNA (117, 118, 119) and the mechanics and features of DNA bending by IHF (120, 121, 122, 123).

IHF is a hetero-dimeric protein consisting of two highly basic polypeptides, α and β , with molecular masses of 11,200 and 10,580 Da, respectively. These two subunits share approximately 30% homology to each other and also to the family of type II DNA-binding proteins that includes major histone-like proteins of *E. coli* such as HU.

The IHF structure, which is very similar to that of HU, is a compact, globular domain, consisting of symmetrically intertwined α and β subunits, from which two long β ribbon arms extend. The arms curl around the DNA and interact exclusively with the minor groove; most of the DNA bending ($>160^\circ$) occurs at two large kinks, 9 bp apart (Fig. 7) (4). It has also been possible to construct a functional IHF in which the two chains have been fused into one (124).

The sequence preference displayed by IHF does not come from specific side chain contacts: it makes no contacts at all within the major groove and only a few hydrogen bonds to positions in the minor groove. Rather its specificity comes from “indirect readout,” based on the sequence dependent structural parameters of its target DNA. Indeed, biochemical and structural studies of a relaxed-specificity mutant of IHF revealed how specificity is determined within the TTR portion of the consensus sequence (117). Within certain constraints, the structure of the DNA was driven by its own sequence and the protein side chains had readjusted to accommodate the different DNA structures.

Evidence that formation of a distinct DNA path was indeed the primary role of IHF came from “bend swap” experiments, where one or more IHF sites were replaced by unrelated DNA bending modules, either intrinsically bent DNA or different DNA bending proteins, such as HU (56, 125, 126). Although able to complement the lack of IHF-induced bending, none of the chimeric constructions performed as efficiently as the

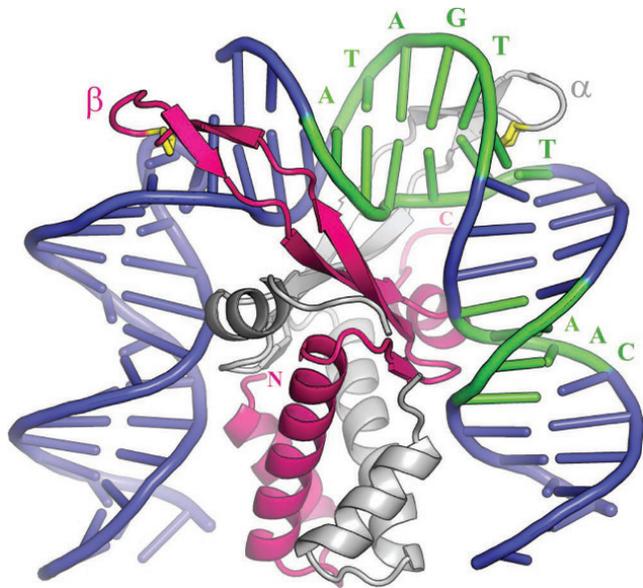


Figure 7 Complex of integration host factor with H1N. The α and β subunits are shown in white and pink, respectively. The consensus sequence is highlighted in green and interacts mainly with the arm of α and the body of β . The yellow proline at the tip of each arm (P65 α /P64 β) is intercalated between bp 28 and 29 on the left side and 37 and 38 on the right. Reprinted with permission from reference 4. doi:10.1128/microbiolspec.MDNA3-0051-2014.f7

wild-type arrangement. The inability of the bend swap chimeras to achieve wild-type efficiency was evidently due to a requirement for considerable precision, as evidenced by the observation that an IHF bend wrongly positioned by just 1 bp, in a loop of constant length between *C* and *P*1 of *attL*, could severely reduce excisive recombination efficiency (127).

It is attractive to propose that the evolution of Int's dependence on host-encoded accessory proteins derives, at least in part, from the benefits of linking the regulation and direction of recombination to the physiology of the host cell (see also the discussion of Fis protein below). In this regard, the changing levels of intracellular IHF are potentially interesting. The relative abundance of IHF increases approximately 5- to 7-fold over a 6 h span after entry into stationary phase (128, 129), and decreases when stationary-phase cells are diluted into fresh medium and cell mass begins to double (130). It is interesting that *in vitro* high IHF concentrations tend to inhibit the excisive reaction (131). The *in vivo* downshift in IHF concentration is probably not due to increased protein degradation, as IHF is not unstable in its dimeric form (132, 133), but instead appears to be a consequence of arrested transcription upon entry into exponential phase and increased

transcription of the individual subunits upon entry into stationary phase (134). Additionally, there is evidence that IHF may play an essential role in survival from cell starvation: not only is IHF critical for induction of 14 proteins from the glucose starvation stimulon but mutants lacking IHF appear to be severely compromised in their ability to survive glucose starvation (135).

Xis

As noted above, the small phage-encoded Xis protein is the key determinant of directionality in the λ pathway. Essential for the excisive reaction and stimulating more than 10^6 -fold *in vivo*, it is also inhibitory for the integrative reaction (136, 137). The NMR structure of $^{1-55}\text{Xis}^{\text{C28S}}$ revealed an unusual “winged”-helix structure formed by two α -helices that are packed against two extended strands. While this structure itself did not afford critical insights into how Xis plays such a critical role in the λ pathway, it did herald the start of a steady progression towards this goal by the Johnson and Clubb laboratories (138).

A 1.7 Å resolution cocrystal structure of $^{1-55}\text{Xis}^{\text{C28S}}$ complexed with a 15 bp DNA fragment containing its cognate X2 binding site comprised the second step of the Johnson/Clubb progression and provided a detailed view of the complex, which was largely in accord with their proposals based on the NMR structure of the free protein (Fig. 8A) (139). Although, the Xis-X2 complex is bent only modestly (approximately 25°), and hardly enough to account for the strikingly large curvature observed for a larger Xis-*attR* complex (93), it did suggest a molecular model for the Xis stimulation of Int binding to the adjacent P2 arm-type site (Fig. 8B).

A precursor to a larger and more informative cocrystal structure was the finding that, counter to the previous long-standing notion of two Xis binding sites (X1 and X2), a third Xis was bound at a site between them, called X1.5 (140, 141). The initial EMSA data were supplemented with protein-protein crosslinking experiments to further confirm the trimeric nature of the complex (140). More useful insights for understanding the role of Xis in directing recombination came from the 2.6 Å cocrystal structure of Xis bound to a larger DNA target comprising the entire 33 bp Xis binding region (Fig. 8C) (140). The three Xis proteins bind to this DNA in a head-to-tail orientation that generates a micronucleoprotein filament having approximately 72° of curvature and a slight positive writhe (Fig. 8D).

The differences in the specific interactions at X1 versus X2, combined with the observed nonspecific binding of Xis at the X1.5 site and the range of different

interactions at ostensibly similar protein-protein interfaces, foreshadowed experiments showing that the flexible recognition surfaces of Xis result in a relatively promiscuous binder of DNA. The propensity for non-specific DNA binding was further characterized in an Xis-DNA cocrystal with an 18 bp fragment of DNA (8). While this flexibility of DNA recognition is important for binding at the X1.5 site, where protein-protein interactions with the X1- and X2-bound Xis protomers provide additional stability, it also means that Xis is easily distracted from its *attR* target *in vivo*, where there is a huge excess of nonspecific DNA. Indeed, this latter point explains why excisive recombination *in vivo* is 50 to 200-fold lower in the absence of Fis (see also below) (8). Correspondingly, the Fis dependence of excisive recombination *in vitro* is only seen at limiting concentrations of Xis (142).

Fis

It is ironic, but understandable with hindsight, that although Fis protein was the first component of the λ recombination pathway to be crystallized (5, 6), it was the last component whose biological and molecular role was elucidated (8). Throughout this 16-year period (and beyond), the Johnson lab has played the leading role in studying the many roles and mechanisms of the Fis protein (112, 143, 144, 145). Fis was initially identified as a factor in promoting site-specific recombination by DNA invertases (146, 147) and was shortly thereafter shown to bind cooperatively with Xis at *attR* and to stimulate excisive recombination up to 20-fold when Xis is limiting (142). *In vivo*, the absence of Fis reduced *attP* formation from an induced lysogen by 100 to 1,000-fold (138, 148); it was also shown to be required along with Xis for binding to the *attR* region in the P22 challenge phage system (149).

Fis, like the other host-encoded accessory protein in the λ Int pathway, IHF, is a nucleoid associated protein of global, structural, and regulatory importance. Its role in determining overall chromosome structure is exerted by contributing to the looped-domain architecture of the nucleoid, and by influencing the regulation of genes encoding topoisomerases (150, 151, 152). Fis plays a role in the initiation of DNA replication, in several transposition reactions, and in the regulation of transcription at many different genes by several different mechanisms (for reviews, see references 112, 143, 151, and 153). The large number of critical sites of action for Fis becomes even more significant when considering how dramatically its intracellular levels vary as a function of cellular physiology.

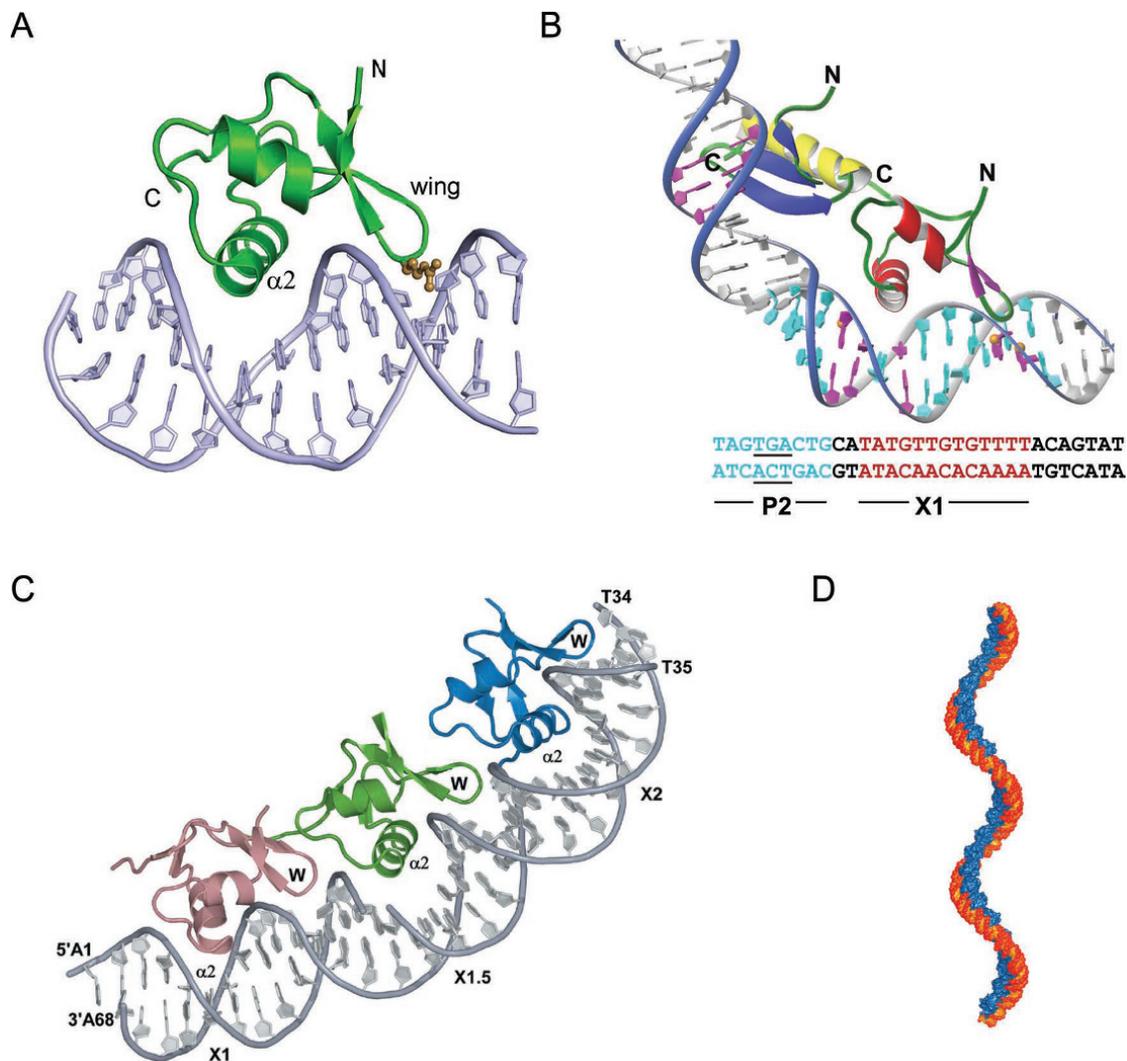


Figure 8 Complex of Xis with DNA. (A) The structure of $^{1-55}\text{Xis}^{\text{C28S}}$ specifically bound to X2 DNA penetrates adjacent grooves of the duplex by fastening on the phosphodiester backbone. The major groove is filled primarily with helix $\alpha 2$ with the side chains of Glu19, Arg23, and Arg26 playing a major role in specific DNA recognition. The adjacent minor groove is contacted by the “wing” which does not contribute significantly to the specificity of complex formation but does contribute to binding affinity, although to a smaller extent than helix $\alpha 2$. The side-chain of Arg39 (brown) extends along the floor of the minor groove where it makes direct and water-mediated hydrogen bonds. (B) A model for the Int (NTD)-Xis-DNA ternary complex. The Int (NTD) is modeled to interact with the TGA trinucleotide (underlined) of the P2 site (blue) in the DNA major groove. Xis is modeled on the X1 site (magenta) in the same manner as observed in the complex with the X2 site. The C-terminal tail of Xis, which is disordered in solution (not shown), is located adjacent to the C-terminal helix of Int to make a protein-protein interaction as shown by mutagenesis and NMR titration data (179). (C) X-ray crystal structure of Xis bound to the Xis binding region reveals the structural basis of cooperative binding. Xis monomers bound to the X1, X1.5, and X2 sites are colored dark salmon, green, and blue, respectively. (D) Structure-based model of an extended Xis-DNA filament. Units of the Xis-DNA^{X1-X2} crystal structure were stacked end-to-end by superimposing site X1 over X1.5 to assemble a pseudocontinuous helix with a pitch of ~22 nm. Proteins are blue; DNA is orange. Reprinted with permission from reference 139 (A and B) and reference 140 (C and D).
doi:10.1128/microbiolspec.MDNA3-0051-2014.f8

Thompson *et al.* (142) showed that Fis levels drop dramatically when cells entered stationary phase and, more significantly, that occupation of the Fis binding site on *att* site DNA also drops in stationary phase cells. More detailed studies revealed that from these extremely low levels in stationary phase, Fis levels increase 500-fold during the initial lag phase when cells are diluted into fresh medium, and reach a peak of 50,000 to 100,000 copies per cell as the culture enters exponential phase. The control of Fis protein synthesis is at the level of mRNA where it is repressed by Fis protein (154, 155, 156) and stimulated by IHF (157). Transcription from the *fis* promoter, *Pfis*, is critically influenced by DksA, a component of the transcription initiation complex that is also required for negative regulation of rRNA promoters (158, 159). DksA, which acts in part by reducing the half-life of (unstable) RNA polymerase-*Pfis* promoter complexes, elevates the required concentration of the initiating NTP (CTP) and amplifies the inhibitory effect of ppGpp on *Pfis* (154, 155, 158, 160). In so doing, it constrains *fis* expression primarily to early log phase at high growth rates, and it inhibits expression at low growth rates or following amino acid starvation (154, 155). However, as normal growth phase-dependent regulation of *fis* is observed in a $\Delta relA \Delta spoT$ strain, other mechanisms can evidently compensate for the role of ppGpp in the pathway (154).

The crystal structures of Fis revealed a globular dimer composed of four tightly intertwined α -helices with two helix-turn-helix (HTH) motifs in each monomer (5, 6). One of the most striking features of the Fis structure was that the D helices, which were proposed to fit into adjacent major grooves of the DNA helix, are only 25 Å apart, approximately 10 Å shorter than the pitch of normal B DNA.

The long-standing hurdle to obtaining Fis-DNA co-crystals was the weak similarity among the many 15 bp sequences capable of forming stable complexes with Fis, thus, making it extremely difficult to derive an optimal consensus sequence for Fis binding. This obstacle was finally overcome by Stella *et al.* (7), who compiled the results of many analyses of Fis binding affinities into an informative hierarchy of DNA sequences, culminating in two 27 bp oligonucleotides whose Fis binding affinities were sufficiently optimized for crystallography (see Fig. 9A). Having established that compression of the central AT-rich minor groove is a critical feature of Fis binding, the authors went on to show that intrinsic DNA bends are unlikely to contribute significantly to Fis binding. Rather, they proposed that Fis initially searches for DNA with an intrinsically narrow minor

groove, where AT composition, not sequence, is the critical determinant. Most recently, the Johnson lab has shown that the primary molecular determinant modulating minor groove widths is the 2-amino group on guanine (145).

While intrinsic DNA bends are not very important for targeting Fis binding, the bends induced by bound Fis are critical for its many functions, including DNA compaction, assembly of invertasomes, regulating transcription, and, most importantly for this article, directing the

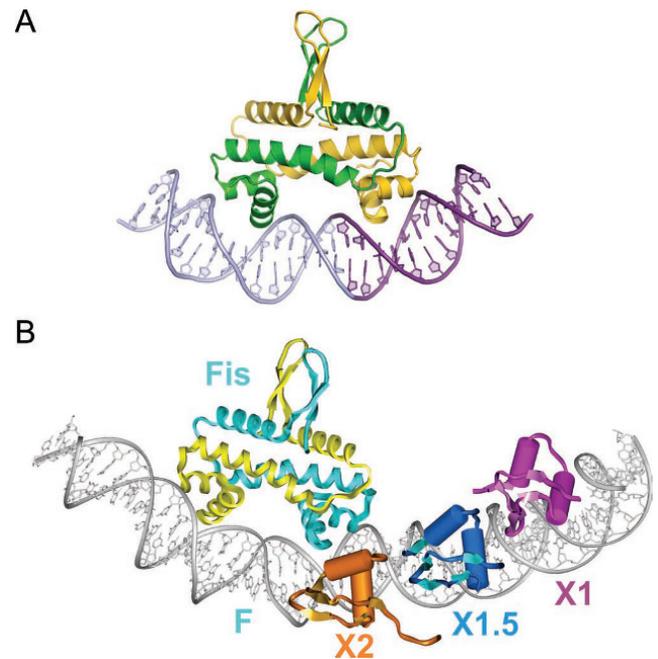


Figure 9 X-ray crystal structure of a Fis dimer complexed with DNA (A) and its relation to Xis binding (B). (A) The C-terminal helix representing the recognition helix of the HTH unit of each subunit is inserted into adjacent major grooves on the concave side of the 21 bp curved DNA. Only base contacts with a single residue, Arg85, are important for binding. The DNA undergoes substantial conformational adjustments, including adoption of $\sim 65^\circ$ overall curvature, to fit onto the Fis binding surface. The central 5 bp of the DNA interface are not contacted by Fis, but compression of the central minor groove to almost half the width of canonical DNA at the center enables the α -helices to insert into the adjacent major grooves, which do not show any appreciable change in width. (B) Model of the Fis-Xis cooperative complex. The X-ray crystal structure of three Δ^{55} Xis monomers bound to the X1 (magenta), X1.5 (blue), and X2 (gold) binding sites was superimposed onto the model of the Fis K36E X-ray structure docked to DNA representing the F site. Fis subunits are cyan and yellow. The DNA recognition helices of Xis bound at X2 and the proximal Fis subunit nearly form a continuous protein surface within the major groove. Reprinted with permission from reference 7 (A) and reference 8 (B).

doi:10.1128/microbiolspec.MDNA3-0051-2014.f9

curvature of the *attR* complex. The Johnson lab proposed that cooperative DNA binding between Fis and its partners, which bind immediately adjacent to Fis but generate only a small number of interfacial amino acid residues, is likely facilitated by mutually compatible changes in DNA shape.

Early experiments indicating that the F and X2 sites on *attR* overlap one another had been interpreted to suggest that both sites could not be bound simultaneously by their cognate proteins. However, subsequent experiments using quantitative gel shifts, stoichiometry determinations, nuclease footprinting and protein-protein crosslinking clearly established that Fis and Xis bind to the F and X2 sites simultaneously and cooperatively (8, 141). Most interestingly, Papagiannis *et al.* showed that Fis binds to the *attR* site *in vitro* with approximately 100-fold greater affinity than Xis alone, and *in vivo*, the rate of excision is reduced approximately 100-fold when Fis is absent (8). They proposed that *in vivo* Fis targets the otherwise peripatetic Xis to the X2 site, which then recruits Xis to X1.5 and X1. Based on their Xis-DNA microfilament cocrystal structures and their Fis crystal structures they built a model of the Fis-Xis complex on *attR* DNA, in which their observed protein induced DNA distortions are proposed to favor the cooperative binding of Fis and Xis (Fig. 9B).

PATTERNS OF λ NTD BINDING AND BRIDGING

Prior to considering the patterns of λ NTD binding and bridging in recombination reactions between canonical pairs of *att* sites, it should be noted the λ Int is also capable of efficiently carrying out an IHF-dependent recombination between two identical *attL* sites lacking a P1 arm site (55). The existence of such a bidirectional pathway lacking the usual complement of components raises interesting questions about the kinds of recombinogenic complexes Int is capable of forming (161, 162, 163) and also underscores the caveat of off-pathway reactions.

The caveat of off-pathway reactions was echoed by a caveat about the artificially imposed symmetry of the NTD domains of the complexes used for X-ray crystallography of Int tetramers bound to Holliday junctions (discussed above). Earlier genetic and nuclease protection experiments had suggested that the patterns of NTD binding to arm-type sites were asymmetric (see Fig. 1) (reviewed in reference 27) and these results were subsequently reinforced by nuclease protection studies on Holliday junction intermediates (trapped with a

hexapeptide inhibitor [51]) and biotin interference assays (BIA) (164). The latter, which probe the requirements for protein binding at a particular DNA locus by obstructing the major groove with a biotin bound to the C5 position of designated thymines, was particularly compelling because it monitored a complete integrative or excisive recombination reaction. From these experiments, it became clear that any attempt to understand the architecture and function of canonical recombinogenic complexes would require an analysis of the full ensemble of proteins and DNAs.

A requisite step in moving towards a panoptic investigation of the recombinogenic complexes was the deciphering of which “core-type” and “arm-type” binding sites are joined to one another by Int-mediated bridges. Towards this end, a disulfide trapping technology (165, 166) was used, in conjunction with trapped Holliday junction complexes (see Fig. 2), to introduce disulfide crosslinks at the protein-DNA interfaces between an Int NTD and its cognate arm-type site, and between an Int CTD and its cognate core-type site (36). Trapped nucleo-protein HJ complexes doubly cross-linked to Int were only observed with those *att* sites in which cystamine-labeled arm site and the cystamine-labeled core site are “bridged” by the same Int molecule.

From such analyses, it was concluded that the Int bridges between arm- and core-type sites in the integrative HJ recombination intermediate are: P'1-C'; P'2-C; P'3-B'; and P1-B. The Int bridges in the excisive HJ intermediate are: P'1-C'; P'2-B; and P2-B'. This leaves the C core site as the one that does not form an Int bridge with one of the three arm-type sites required for excisive recombination.

The Int bridges determined by site-directed cross-linking in HJ complexes were confirmed and complemented in full recombination reactions by a genetic approach using two chimeric recombinases. The first, called Crn1, consists of a Cre recombinase fused to the NTD of λ -integrase; it has all the properties of λ Int (described above) (89). This was complemented by construction of a second chimeric recombinase, Crn2, in which the NTD and CTD domains recognized different arm- and core-type DNA target sequences (36). A collection of hybrid *att* sites was constructed in which one of the bridged arm-core pairs (identified by the chemical crosslinking experiments) had the arm and core sequences recognized by Crn2, while the remaining arm-core bridges had the arm and core sequences recognized by Crn1. Using these substrates, it was shown that Crn1 could not carry out recombination unless Crn2 was also present (and vice versa). The results of the chimeric recombination reactions confirmed, and

also provided information complementary to, the results from chemical crosslinking (as discussed below).

These results argue strongly against models in which regulated directionality of λ Int recombination depends upon some degree of Int bridge remodeling during the course of the reaction. Furthermore, the monogamous relationship of each arm-core bridged pair throughout the course of the recombination reaction makes it possible to extrapolate from the patterns observed in the HJ recombination intermediate to those predicted for the presynaptic recombination partners and the post HJ recombination products. Inspection of Fig. 10 reveals that for excisive recombination, the presynaptic partners have only intramolecular bridges, suggesting that Int bridging is not a driving force in synapsis of *attL* and *attR*. This is also likely to be the case for integrative recombination, even though the capture of a naked *attB* by a fully assembled (supercoiled) *attP* complex requires two intermolecular bridges (91). It was postulated that the reason *attB* cannot bind Int protomers unless they are part of a higher-order complex stems from the need to overcome the NTD inhibition of CTD function, described above (49), and not from any driving force by intermolecular bridges (see also discussion below).

ARCHITECTURES OF RECOMBINOGENIC COMPLEXES

In an attempt to derive architectural models for the HJ recombination intermediates, the Int bridging results were augmented with in-gel fluorescence resonance energy transfer (FRET) experiments (101, 167, 168, 169) to determine the apparent distances between selected positions within the excisive recombination HJ intermediate (37).

Using the Int bridging data, the apparent FRET distances for the HJ recombination complex, and the 3D structures for all of the protein components in their DNA-bound forms (4, 7, 8, 44, 45, 139, 140), it was possible to computationally build a model for the architecture of the λ excision complex (Fig. 11A, B, C). Insights gained from the excisive complex along with the integrative Int bridging data and 3D structures were used to generate a corresponding model of the integrative complex (Fig. 11D, E, F). Considered individually and together, the two architectures afford a number of interesting insights, as discussed below and in the figure legends (37).

In the excision complex, the P' and B arms form a left-handed crossing, while the overall path of *attR* DNA indicates a left-handed, nucleosome-like, wrapping by

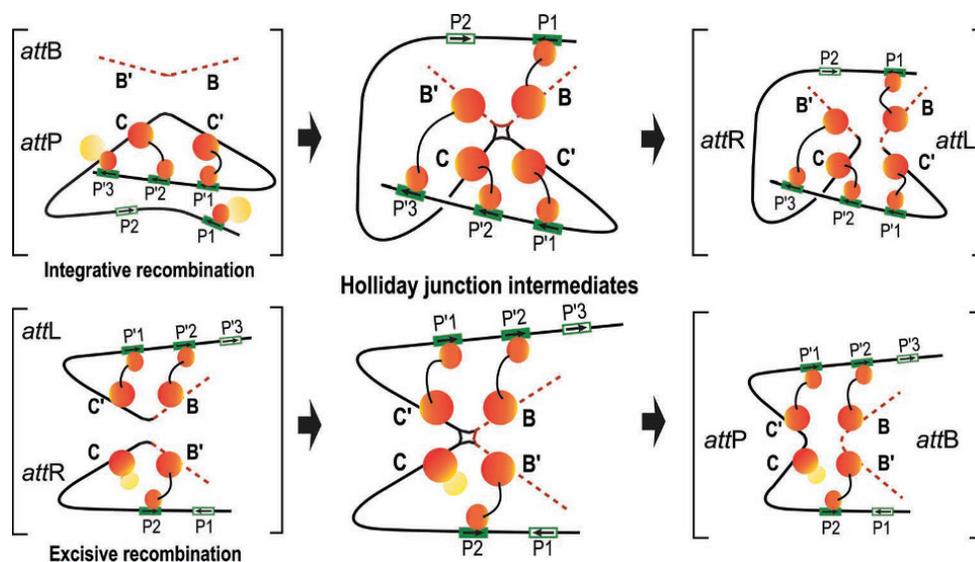


Figure 10 Schematic summary of the Int bridges in integrative and excisive recombination. The middle panel diagrams the Int bridges of the Holliday junction (HJ) recombination intermediates determined by Tong *et al.* (36). In the integrative complex, all four core sites and four of the five arm sites enjoy an Int bridge while the excisive complex engages three of the four core sites and three of the five arm sites. The flanking panels (brackets) depict extrapolations from the HJ complexes to the respective *att* site recombination partners (substrates) and recombinants (products) based on the deduction that Int bridges are not broken and reformed during recombination. doi:10.1128/microbiolspec.MDNA3-0051-2014.f10

IHF, Fis, Xis, and Int. The model thus predicts a negative DNA crossing node in *attL* and left-handed solenoidal wrapping in *attR*, both of which are consistent with negative supercoiling in the normal substrates. In the integrative complex the asymmetric mode of binding to C' , C , and B' by the P' arm requires considerable flexibility in the linker segments between the CTDs and NTDs, and this model explicitly predicts the formation of a negative DNA-crossing node in the recombination complex, where the P-arm crosses over P' . The model also features an unusual and flexible P-arm tether that positions Int-B for *attB* binding.

ASYMMETRY AND FLEXIBILITY

The architectures proposed for the recombinogenic complexes differ in several ways from the crystal structures of the HJ-bound Int tetramers bound to arm site DNA duplexes (44). While this would not necessarily have been predicted, it is not surprising, as the crystal structures did not include accessory DNA bending proteins or their cognate DNA sites, which join the core- and arm-type sites. An additional compromise required to form crystals was the substitution of a pair of $P'1$ - $P'2$ -containing oligonucleotides for the canonical asymmetric arrangement of arm-type binding sites. Indeed, subsequent experiments involving biotin-interference mapping of complete recombination reactions (described above) are more consistent with the asymmetric architectures than the symmetric arrangement in the smaller complexes designed for crystallization (164). In contrast to the symmetric and tightly packed NTD organization observed in crystal structures, the models for the architectures of the complexes feature highly asymmetric arrangements of the NTDs. In the former, the domains are swapped, with the NTD of one Int subunit located above the CB domain of an adjacent Int. The latter is incompatible with domain-swapped NTDs and implies considerable flexibility in the CB-NTD linkers.

TOPOLOGY

Excisive recombination between directly repeated *attL* and *attR* sites results in a large fraction of free circles when supercoiling levels are low, similar to that observed for the Cre and F1p recombinases (170). Integrative recombination between directly repeated *attP* and *attB* sites results in catenated circles for supercoiled substrates, implying that the recombination process itself imposes a strand crossing (170, 171, 172). Seah *et al.* (37) argue that the proposed architectures are

consistent with these results and explain many of the other topological findings of Crisona *et al.* (170). Additionally, the tightly wrapped nature of the integrative complex model and the inclusion of a negative DNA crossing node are consistent with, and may explain, in part, the requirement for negative supercoiling for efficient integration (91).

CAPTURING THE HOST *attB* SITE

From the time Richet and Nash (91) first showed that *attB* comes naked to a recombination with its fully decorated *attP* partner there has been considerable speculation about the details of this synaptic event. Because of the pseudodyad symmetry of the core-type sites the openings of the bound integrase C-clamps must face in opposite directions (45). While this is not a problem for the monovalent family members it implies that for the fully assembled *attP* complex one of the Int subunits (the one destined to bind the B core site of *attB*) must have the flexibility to wrap around the host chromosome from the opposite face.

Indeed, the architecture proposed for the integrative complex does contain an inherently flexible P-arm that tethers the Int-B subunit and allows for the dynamic binding required to engage the bacterial chromosome and ultimately lock onto the *attB* sequence. The model is also consistent with, and explains, a difference between the two kinds of Int bridging experiments reported by Tong *et al.* (36). Whereas chemical crosslinking of the $P1$ -B Int bridge was the most robust of all the Int bridges, in the genetic analyses, the $P1$ -B Int bridge was the weakest, precisely the difference expected for a flexible arm.

ARCHITECTURAL BASIS FOR DIRECTIONALITY

The source of the strong bias towards the top strands being exchanged first in formation of the HJ (33, 35) is evident from the models in Fig. 11 and Fig. 12. During excisive recombination, both *attL* and *attR* are bent at their core sites in order to promote the bridging interactions that form between core and arm binding sites. The core site bend directions that lead to stable complexes are coupled to IHF-induced bends and commit both *attL* and *attR* to top strand cleavage upon synapsis of the sites. Similarly, only one bend direction of the *attP* core site will lead to stable bridging interactions between C/C' core sites and $P'1/P'2$ arm sites. This direction commits *attP* to top strand cleavage in the synaptic complex with *attB*. Thus, the order of strand

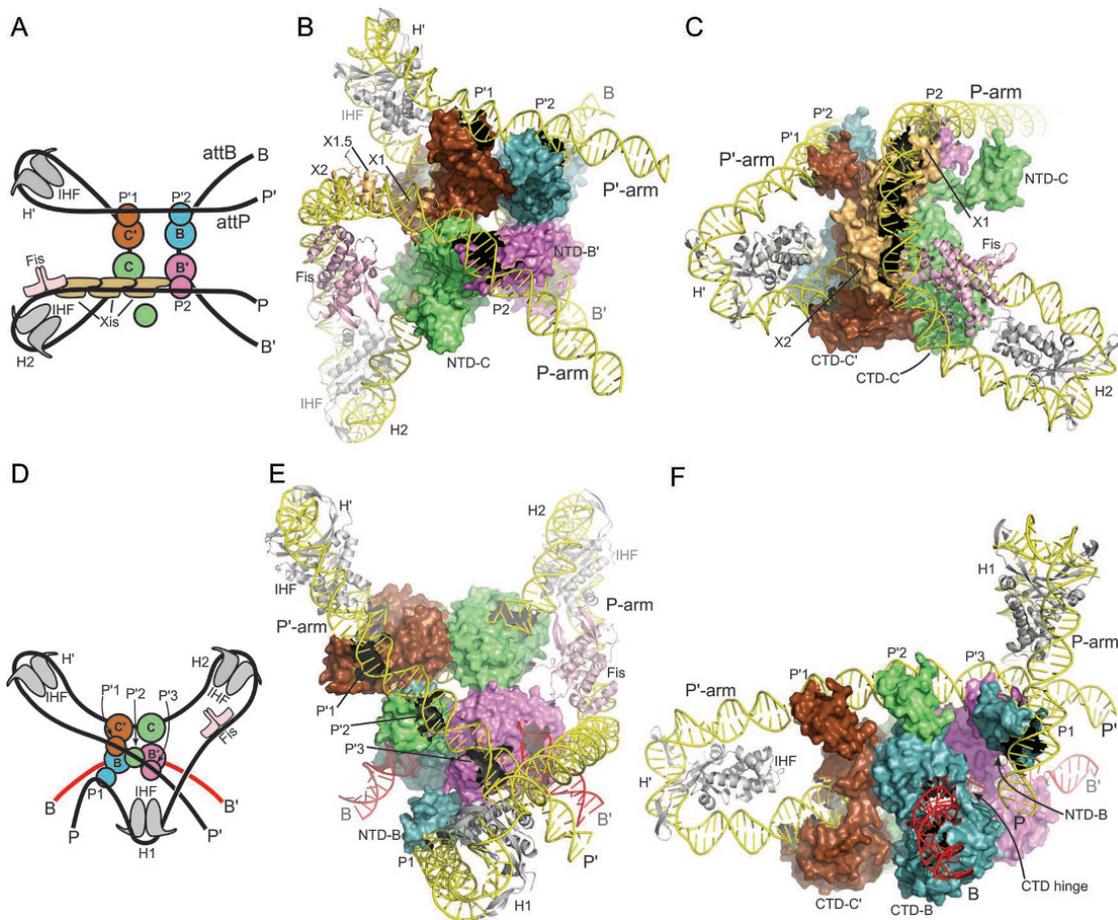


Figure 11 Models of the λ excisive and integrative recombination complexes. (A) Schematic representation of the excisive complex architecture. The excision reaction product resulting from Holliday junction (HJ) resolution is shown. Int subunits (blue, green, magenta, brown) are represented by a small circle (NTD) and a large circle (CTD). Integration host factor (IHF) heterodimers (gray) are shown bound to the H' and H2 sites. Fis dimer (pink) and Xis (tan) subunits are indicated. (B) Model of the excisive complex in the same "top view" orientation as the schematic drawing in panel A. The NTD of the Int subunit bound at the C core site (NTD-C) is shown separated from the rest of the complex to improve clarity of the P-arm trajectory. (C) Side view of the excisive complex, highlighting the trajectory of the P-arm. IHF bending of the P' arm at H' directs the DNA over the CTD domains of the Int tetramer, facilitating engagement of the P'1 and P'2 arm sites by the Int subunits bound at the C' and B' core sites, respectively. In the P-arm of *attR* the phasing of the IHF-induced bend at H2 is different from that at H'; at H2, the P-arm is directed along the plane of the catalytic domain tetramer. An A-tract sequence that is stabilized by Fis binding (7, 8) directs the P-arm upwards, towards the Int CB domains. The cooperative Xis filament (8, 140) then redirects the P-arm across the top of the Int CTD domains, where the P2 site is bound by the Int subunit bound at the B' core site. The Xis subunit bound at X1 resides close to the position where the NTD of the Int subunit bound at the C core site (Int-C) would be expected. The NTD of Int-C was not docked in a specific location of the excisive complex model, but it seems plausible, even attractive, that this domain could bind nonspecifically to the P-arm near the X1 site, perhaps interacting with Xis. (D) Schematic of the integrative complex architecture. The arm-type binding sites engaged by the four Int subunits are indicated. (E) Model of the integrative complex in the same "top view" as illustrated in panel B. In this orientation, the P-arm rises towards the viewer, crosses over the P' arm, and is directed back towards the Int tetramer by the IHF bend at the H1 site. (F) Side view of the integrative model, looking approximately down the B core site. The NTD of the Int subunit bound at the B core site (NTD-B) is shown bound at the P1 site, on the flexible P-arm. The CB and catalytic domains of the Int subunit bound at the B site can be seen wrapped around the opposing face of *attB*, with the interdomain hinge indicated. The CTD-NTD linkers were not modeled and are not shown. IHF bending at H' directs the P' arm over the CTD domains of the Int tetramer, but in this case the P'1, P'2, and P'3 binding sites are engaged by the Int subunits bound to C', C, and B', respectively. As Xis is not present in the integrative complex, the P-arm is directed upwards, parallel to the Int tetramer, and as Fis stimulation of integration has been reported (180, 181), it was included in the model. IHF bound to the H1 site redirects the P-arm back towards the Int tetramer, crossing over the P' arm in the process. The P1 arm-type site is thereby brought to a position where it can bind the NTD of the Int subunit poised for capture of the B core half-site (Int-B). Reprinted with permission from reference 37. doi:10.1128/microbiolspec.MDNA3-0051-2014.f11

exchange in both pathways is determined prior to synthesis by formation of specific *attL*, *attR*, and *attP* complexes (34).

The architecture of the excisive complex provides a bird's eye view of how Xis mediates its critical role as the regulator of directionality (137, 173) (Fig. 13A). In the absence of Xis, the P-arm would not be directed across the top of the Int CTDs to make the required P2-B' bridge and the P-arm would not be properly positioned to stabilize a functional *attR*. An additional critical role for Xis is to promote the cooperative binding of the Int NTD at P2 (66, 130, 174).

The architecture of the excisive complex also explains the long-standing question of why the excision reaction does not run efficiently in reverse once *attB* is released (Fig. 13B). After dissociation of *attB*, the *attP* complex is expected to be less stable because it now only contains a single intramolecular bridge (P'1-C'). Furthermore, this complex has the potential to rearrange, such that the *attP* core bends in the opposite direction and facilitates the formation two intramolecular bridges (P'1-C' and P'2-C). While this complex resembles a portion of the *attP* substrate complex, it is prevented from proceeding to a competent complex by

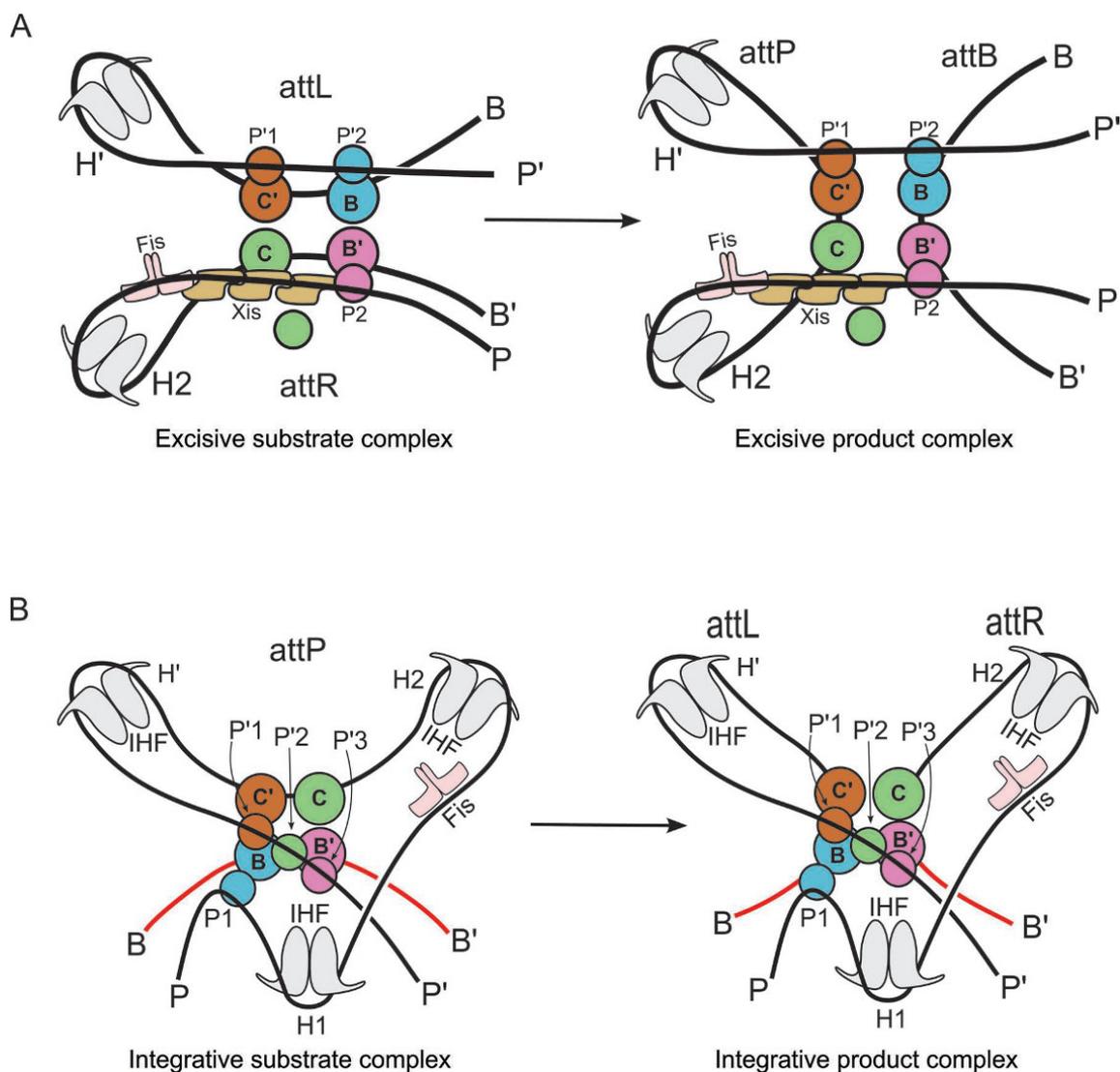


Figure 12 Schematic representation of the excisive and integrative reactions, based on the structural models shown in Fig. 13. Coloring of the protein subunits matches that shown in Fig. 11. Reprinted with permission from reference 37. doi:10.1128/microbiolspec.MDNA3-0051-2014.f12