Zinc Finger Proteins: From Atomic Contact to Cellular Function
Zinc Finger Proteins: 
From Atomic Contact to Cellular Function

Shiro Iuchi
Department of Cell Biology
Harvard Medical School
Boston, Massachusetts, U.S.A.

Natalie Kuldell
Biological Engineering Division
Massachusetts Institute of Technology
Cambridge, Massachusetts, U.S.A.
ZINC FINGER PROTEINS:
FROM ATOMIC CONTACT TO CELLULAR FUNCTION
Molecular Biology Intelligence Unit

Copyright ©2005 Landes Bioscience / Eurekah.com and Kluwer Academic / Plenum Publishers

All rights reserved.
No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Printed in the U.S.A.

http://www.wkap.nl

Please address all inquiries to the Publishers:
Landes Bioscience / Eurekah.com, 810 South Church Street, Georgetown, Texas, U.S.A. 78626
Phone: 512/ 863 7762; FAX: 512/ 863 0081
http://www.landesbioscience.com
http://www.eurekah.com

Zinc Finger Proteins: From Atomic Contact to Cellular Function, edited by Shiro Iuchi and Natalie Kuldell, Landes / Kluwer dual imprint / Landes series: Molecular Biology Intelligence Unit
ISBN: 0-306-48229-0

While the authors, editors and publisher believe that drug selection and dosage and the specifications and usage of equipment and devices, as set forth in this book, are in accord with current recommendations and practice at the time of publication, they make no warranty, expressed or implied, with respect to material described in this book. In view of the ongoing research, equipment development, changes in governmental regulations and the rapid accumulation of information relating to the biomedical sciences, the reader is urged to carefully review and evaluate the information provided herein.

Library of Congress Cataloging-in-Publication Data

Zinc finger proteins : from atomic contact to cellular function / [edited by] Shiro Iuchi, Natalie Kuldell. p. ; cm. -- (Molecular biology intelligence unit)
Includes bibliographical references and index.
ISBN 0-306-48229-0
QP552.Z55Z56 2005
572.8'645--dc22
2004023431
CONTENTS

Preface xi
Abbreviations xiv

1. The Discovery of Zinc Fingers and Their Practical Applications in Gene Regulation: A Personal Account 1
Aaron Klug

BINDING OF ZINC FINGERS TO DNA

2. C2H2 Zinc Fingers As DNA Binding Domains 7
Shiro Iuchi

3. TFIIIA: A Sophisticated Zinc Finger Protein 14
Raymond S. Brown and Jane Flint

4. GAGA: Structural Basis for Single Cys2His2 Zinc Finger-DNA Interaction 20
G. Marius Clore and James G. Omichinski

5. The DNA-Binding Domain of GATA Transcription Factors—A Prototypical Type IV Cys2-Cys2 Zinc Finger 26
Angela M. Gronenborn

Ryoji Masui, Noriko Nakagawa and Seiki Kuramitsu

Patrick Van Roey, Marlene Belfort and Victoria Derbyshire

8. Zinc Finger Interactions with Metals and Other Small Molecules 39
Jay S. Hanus, Jason L. Larabee and James R. Hocker

9. Synthetic Zinc Finger Transcription Factors 47
 Nicoletta Corbi, Valentina Libri and Claudio Passananti

BINDING OF ZINC FINGERS TO RNA

10. TFIIIA and p43: Binding to 5S Ribosomal RNA 56
Paul J. Romanik

11. RNA Binding by Single Zinc Fingers 66
Martyn K. Darby

12. Wig-1, a p53-Induced Zinc Finger Protein that Binds Double Stranded RNA 76
Cristina Mendez-Vidal, Fredrik Hellborg, Margareta T. Wilhelm, Magdalena Tarkowska and Klas G. Wiman

13. Tandem CCCH Zinc Finger Proteins in mRNA Binding 80
Perry J. Blackshear, Ruth S. Phillips and Wi S. Lai

14. Ribosomal Zinc Finger Proteins: The Structure and the Function of Yeast YL37a 91
John Dresios, Yuen-Ling Chan and Ira G. Wool

BINDING OF ZINC FINGERS TO PROTEINS

15. LIM Domain and Its Binding to Target Proteins 99
Algirdas Velyvis and Jun Qin

16. RING Finger-B Box-Coiled Coil (RBCC) Proteins As Ubiquitin Ligase in the Control of Protein Degradation and Gene Regulation 106
Kazuhito Ikeda, Satoshi Inoue and Masami Muramatsu

17. Structure and Function of the CBP/p300 TAZ Domains 114
Roberto N. De Guzman, Maria A. Martinez-Yamout, H. Jane Dyson and Peter E. Wright

18. A Zinc Ribbon Motif Is Essential for the Formation of Functional Tetrameric Protein Kinase CK2 121
Odile Filhol, Maria José Benitez and Claude Cochet

BINDING OF ZINC FINGERS TO SMALL MOLECULES

19. The FYVE Finger: A Phosphoinositide Binding Domain 128
Harald Stenmark

COMMON DOMAINS PRESENT IN ZINC FINGER PROTEINS

20. The BTB Domain Zinc Finger Proteins 134
Gilbert G. Privé, Ari Melnick, K. Farid Ahmad and Jonathan D. Licht
21. KRAB Zinc Finger Proteins:  
A Family of Repressors Mediating  
Heterochromatin-Associated  
Gene Silencing  
Shiro Iuchi  

22. The Superfamily of SCAN Domain  
Containing Zinc Finger  
Transcription Factors  
Tucker Collins and Tara L. Sander  

BIOLOGY OF ZINC FINGER PROTEINS  

23. Sp1 and Huntington's Disease  
Dimitri Krainc  

24. The Role of WT1 in Development  
and Disease  
Sean Bong Lee, Hongjie Li and Ho-Shik Kim  

25. Yin Yang 1  
Huifei Liu and Yang Shi  

26. The Multiple Cellular Functions of TFIIIA  
Natalie Kuldell  

27. The Role of the Ikaros Gene Family  
in Lymphocyte Development  
Pablo Gómez-del Arco, Taku Naito, John Seavitt,  
Toshimi Yoshida, Christine Williams  
and Katia Georgopoulos  

28. Basonuclin:  
A Zinc Finger Protein of Epithelial Cells  
and Reproductive Germ Cells  
Howard Green and Hung Tseng  

29. ZAS Zinc Finger Proteins:  
The Other KB-Binding Protein Family  
Carl E. Allen and Lai-Chu Wu  

30. Role of GATA Factors in Development  
Marc Haenlin and Lucas Waltzer  

31. The Androgen Receptor and Spinal  
and Bulbar Muscular Atrophy  
Federica Picioni, Charlotte J. Sumner  
and Kenneth H. Fischbeck  

32. The Role of XPA in DNA Repair  
Takahisa Ikegami and Masahiro Shirakawa  

33. MOF, an Acetyl Transferase Involved  
in Dosage Compensation in Drosophila,  
Uses a CCHC Finger  
for Substrate Recognition  
Asifa Akhtar and Peter B. Becker  

34. MDM2: RING Finger Protein and Regulator  
of p53  
Liqing Wu and Carl G. Maki  

BIOLOGY OF ZINC ION  

35. The Zip Family of Zinc Transporters  
David J. Eide  

36. Apoptosis by Zinc Deficiency  
Kirsteen H. Maclean  

Index  

213  
221  
232  
239  
247  
207  
261  
265  
273
<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Location</th>
<th>E-mail</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marc Haenlin</td>
<td>Centre de Biologie du Développement</td>
<td>CNRS UMR Toulouse, France</td>
<td>e-mail: <a href="mailto:haenlin@cict.fr">haenlin@cict.fr</a></td>
<td>30</td>
</tr>
<tr>
<td>Jay S. Hanas</td>
<td>University of Oklahoma College of Medicine</td>
<td>Oklahoma City, Oklahoma, U.S.A.</td>
<td>e-mail: <a href="mailto:jay-hanas@ouhsc.edu">jay-hanas@ouhsc.edu</a></td>
<td>8</td>
</tr>
<tr>
<td>Fredrik Hellborg</td>
<td>Department of Oncology-Pathology</td>
<td>Karolinska Institute</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>James R. Hocker</td>
<td>University of Oklahoma College of Medicine</td>
<td>Oklahoma City, Oklahoma, U.S.A.</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Kazuhiro Ikeda</td>
<td>Research Center for Genomic Medicine</td>
<td>Saitama Medical School</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Takahisa Ikegami</td>
<td>Laboratory of Structural Proteomics</td>
<td>Institute for Protein Research</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Satoshi Inoue</td>
<td>Research Center for Genomic Medicine</td>
<td>Saitama Medical School</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Ho-Shik Kim</td>
<td>Genetics of Development and Disease Branch</td>
<td>National Institute of Diabetes and Digestive and Kidney Diseases</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Aaron Klug</td>
<td>MRC Laboratory of Molecular Biology</td>
<td>Cambridge, United Kingdom</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Dimitri Krainc</td>
<td>MassGeneral Institute for Neurodegenerative Disease</td>
<td>Harvard Medical School</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Seiki Kuramitsu</td>
<td>Department of Biology</td>
<td>Osaka University</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Wi S. Lai</td>
<td>Office of Clinical Research and the Laboratory of Signal Transduction</td>
<td>National Institute of Environmental Health Sciences</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Jason L. Larabee</td>
<td>University of Oklahoma College of Medicine</td>
<td>Oklahoma City, Oklahoma, U.S.A.</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Sean Bong Lee</td>
<td>Genetics of Development and Disease Branch</td>
<td>National Institute of Diabetes and Digestive and Kidney Diseases</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Hongjie Li</td>
<td>Genetics of Development and Disease Branch</td>
<td>National Institute of Diabetes and Digestive and Kidney Diseases</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Valentina Libri</td>
<td>CNR</td>
<td>Istituto Biologia e Patologia Molecolari</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Jonathan D. Licht</td>
<td>Division of Hematology and Oncology</td>
<td>Mount Sinai School of Medicine</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Huifei Liu</td>
<td>Department of Pathology</td>
<td>Harvard Medical School</td>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>
Kirsteen H. Maclean  
Department of Biochemistry  
St. Jude Children’s Research Hospital  
Memphis, Tennessee, U.S.A.  
E-mail: kirsteen.maclean@stjude.org  
Chapter 36

Carl G. Maki  
The University of Chicago  
Department of Radiation and Cellular Oncology  
Chicago, Illinois, U.S.A.  
E-mail: cmaki@rover.uchicago.edu  
Chapter 34

Maria A. Martinez-Yamout  
Department of Molecular Biology  
The Scripps Research Institute  
La Jolla, California, U.S.A.  
Chapter 17

Ryoji Masui  
Department of Biology  
Osaka University  
Osaka, Japan  
Chapter 6

Ari Melnick  
Department of Developmental and Molecular Biology  
Albert Einstein College of Medicine  
Bronx, New York, U.S.A.  
Chapter 20

Cristina Mendez-Vidal  
Department of Oncology and Pathology  
Karolinska Institute  
Cancer Center Karolinska (CCK)  
Stockholm, Sweden  
Chapter 12

Masami Muramatsu  
Research Center for Genomic Medicine  
Saitama Medical School  
Saitama, Japan  
E-mail: muramasa@saitama-med.ac.jp  
Chapter 16

Taku Naito  
Cutaneous Biology Research Center  
Massachusetts General Hospital  
and Harvard Medical School  
Charlestown, Massachusetts, U.S.A.  
Chapter 27

Noriko Nakagawa  
RIKEN Harima Institute  
Hyogo, Japan  
Chapter 6

James G. Omichinski  
Département de Biochimie  
Université de Montréal  
Montréal, Quebec, Canada  
Chapter 4

Claudio Passananti  
CNR  
Istituto Biologia e Patologia Molecolari  
Sezione di Patologia Molecolare  
Rome, Italy  
E-mail: claudio.passananti@ibpm.cnr.it  
Chapter 9

Ruth S. Phillips  
Office of Clinical Research and the Laboratory  
of Signal Transduction  
National Institute of Environmental Health Sciences  
National Institutes of Health  
Research Triangle Park, North Carolina, U.S.A.  
Chapter 13

Federica Piccioni  
Neurogenetics Branch  
National Institute of Neurological Disorders and Stroke  
National Institutes of Health  
Bethesda, Maryland, U.S.A.  
Chapter 31

Gilbert G. Prive  
Department of Medical Biophysics  
University of Toronto and Division of Molecular  
and Structural Biology  
Ontario Cancer Institute  
Toronto, Ontario, Canada  
Chapter 20

Jun Qin  
Structural Biology Program  
The Lerner Research Institute  
The Cleveland Clinic Foundation  
Cleveland, Ohio, U.S.A.  
E-mail: qinj@ccf.org  
Chapter 15

Paul J. Romaniuk  
Department of Biochemistry and Microbiology  
University of Victoria  
Victoria, British Columbia, Canada  
E-mail: pjr@uvic.ca  
Chapter 10
In the early 1980s, a few scientists started working on a Xenopus transcription factor, TFIIIA. They soon discovered a novel domain associated with zinc, and named this domain "zinc finger." The number of proteins with similar zinc fingers grew quickly and these proteins are now called C\textsubscript{2}H\textsubscript{2}, Cys\textsubscript{2}His\textsubscript{2} or classical zinc finger proteins. To date, about 24,000 C\textsubscript{2}H\textsubscript{2} zinc finger proteins have been recognized. Approximately 700 human genes, or more than 2% of the genome, have been estimated to encode C\textsubscript{2}H\textsubscript{2} finger proteins. From the beginning these proteins were thought to be numerous, but no one could have predicted such a huge number. Perhaps thousands of scientists are now working on C\textsubscript{2}H\textsubscript{2} zinc finger proteins from various viewpoints. This field is a good example of how a new science begins with the insight of a few scientists and how it develops by efforts of numerous independent scientists, in contrast to a policy-driven scientific project, such as the Human Genome Project, with goals clearly set at its inception and with work performed by a huge collaboration throughout the world.

As more zinc finger proteins were discovered, several subfamilies, such as C\textsubscript{2}C\textsubscript{2}, CCHC, CCCH, LIM, RING, TAZ, and FYVE emerged, increasing our understanding of zinc fingers. The knowledge was overwhelming. Moreover, scientists began defining the term "zinc finger" differently and using various names for identical zinc fingers. These complications may explain why no single comprehensive resource of zinc finger proteins was available before this publication.

This book adopts a broad definition of zinc finger as a peptide domain with a special tertiary structure stabilized by Zn\textsuperscript{2+} coordination. These tertiary structured fingers confer specific binding activities to various molecules such as DNA, RNA, proteins or small molecules. Some groups of scientists exploited the binding specificity to develop tools that target zinc finger proteins to any DNA and RNA segments at will. Other groups studied additional domains that are required for executing functions of zinc finger protein molecules. Some of these domains, such as BTB, KRAB and SCAN, are often present in C\textsubscript{2}H\textsubscript{2} zinc finger proteins, while other associated domains are less common but may also play important roles. Functions of zinc finger proteins include gene expression, signal transduction, cell growth, differentiation and development. Given the broad role zinc finger proteins play in the cell, it is logical that mutations in some zinc finger proteins have been found to cause diseases such as cancer and neurological disorders. Likewise, zinc ion deficiency leads to cell death, probably because of the essential nature of zinc finger proteins.

This book systematically describes features of various zinc finger proteins, presenting specific activities of their domains in the earlier chapters and cellular activities of the protein molecules in the later chapters. The chapters are written by authors outstanding in the field. These chapters are grouped according to their binding function and typically include several illustrative examples of zinc finger proteins for each function. The arrangement reflects this book's intention to encompass the principles and the paradigms of zinc finger proteins at levels from atomic through cellular to organismal. This book should be a useful resource for experts in the field, but also valuable for scientists and graduate students of biological science.

We thank Drs. J.M. Baraban, D.A. Haber, A.G. Jochemsen, F.L. Rauscher III and B. Schulz for their suggestions regarding experts of research fields, Ms. K. Easley for her assistance, and Dr. H. Green for his support.

Shiro Iuchi
Natalie Kuldell
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aACRYBP1</td>
<td>aA-crystallin binding protein</td>
</tr>
<tr>
<td>AAV</td>
<td>adeno associated virus</td>
</tr>
<tr>
<td>ADPRT</td>
<td>poly(ADP-ribose) transferase</td>
</tr>
<tr>
<td>AF-1</td>
<td>activation function 1</td>
</tr>
<tr>
<td>AF-2</td>
<td>activation function 2</td>
</tr>
<tr>
<td>AFC</td>
<td>antibody forming cell</td>
</tr>
<tr>
<td>AGM</td>
<td>aorta gonad mesonephros</td>
</tr>
<tr>
<td>AIF</td>
<td>apoptosis-inducing factor</td>
</tr>
<tr>
<td>Aio</td>
<td>Aiolos</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>ARE</td>
<td>androgen responsive element</td>
</tr>
<tr>
<td>AT-BP2</td>
<td>a1-antitrypsin promoter binding protein</td>
</tr>
<tr>
<td>ATM</td>
<td>ataxia telangiectasia mutated</td>
</tr>
<tr>
<td>ATR</td>
<td>ATM and Rad3-related</td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>BAF</td>
<td>brg1 associated factors</td>
</tr>
<tr>
<td>BLAST</td>
<td>basic local alignment search tool</td>
</tr>
<tr>
<td>BM</td>
<td>bone marrow</td>
</tr>
<tr>
<td>BMI-1</td>
<td>B-cell-specific Moloney murine leukemia virus integration site 1</td>
</tr>
<tr>
<td>BMP</td>
<td>bone morphogenetic protein</td>
</tr>
<tr>
<td>Bn1</td>
<td>Basoncin 1</td>
</tr>
<tr>
<td>Bn2</td>
<td>Basoncin 2</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BRCA1</td>
<td>breast cancer predisposition gene 1</td>
</tr>
<tr>
<td>BTB</td>
<td>broad-complex, tramtrack, and bric-a-brac</td>
</tr>
<tr>
<td>CARD</td>
<td>caspase recruitment domain</td>
</tr>
<tr>
<td>Chbl</td>
<td>casitas B-lineage lymphoma protein</td>
</tr>
<tr>
<td>CBP</td>
<td>cAMP response element binding protein</td>
</tr>
<tr>
<td>CD</td>
<td>circular dihirosem; chromo domain; chromosome modifier domain</td>
</tr>
<tr>
<td>Cdk</td>
<td>cyclin-dependent kinase</td>
</tr>
<tr>
<td>CFTR</td>
<td>cystic fibrosis trans-membrane conductance regulator</td>
</tr>
<tr>
<td>CHD</td>
<td>congenital heart defect</td>
</tr>
<tr>
<td>Ch1p</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CHX</td>
<td>cyclohexamide</td>
</tr>
<tr>
<td>CK2</td>
<td>protein kinase CK2</td>
</tr>
<tr>
<td>(formerly known as casein kinase II)</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>caseine</td>
</tr>
<tr>
<td>CKII</td>
<td>casein kinase II</td>
</tr>
<tr>
<td>CLP</td>
<td>common lymphoid progenitor</td>
</tr>
<tr>
<td>CoaA</td>
<td>coenzyme A</td>
</tr>
<tr>
<td>Co124</td>
<td>type II collagen gene</td>
</tr>
<tr>
<td>CpG</td>
<td>cytidine-guanidine dinucleotide pairs</td>
</tr>
<tr>
<td>CREB</td>
<td>cAMP response element binding</td>
</tr>
<tr>
<td>CRIP</td>
<td>cysteine-rich intestinal protein</td>
</tr>
<tr>
<td>CRP</td>
<td>cysteine-rich protein</td>
</tr>
<tr>
<td>CS</td>
<td>cleavage site</td>
</tr>
<tr>
<td>CSD</td>
<td>chromo shadow domain</td>
</tr>
<tr>
<td>CtBP</td>
<td>c-terminal binding protein</td>
</tr>
<tr>
<td>SS DNA</td>
<td>5S ribosomal RNA gene</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA binding domain</td>
</tr>
<tr>
<td>DCC</td>
<td>dosis compensation complex</td>
</tr>
<tr>
<td>DDS</td>
<td>Denys-Drash syndrome</td>
</tr>
<tr>
<td>DHFR</td>
<td>dihydrololate reductase</td>
</tr>
<tr>
<td>DHS</td>
<td>Dnase I hypersensitivity site</td>
</tr>
<tr>
<td>DHT</td>
<td>5a-dihydrotestosterone</td>
</tr>
<tr>
<td>DMD</td>
<td>Duchenne muscular dystrophy</td>
</tr>
<tr>
<td>dn</td>
<td>dominant negative</td>
</tr>
<tr>
<td>DN</td>
<td>double negative</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>DP</td>
<td>double positive</td>
</tr>
<tr>
<td>Dpp</td>
<td>decapsenategic</td>
</tr>
<tr>
<td>ds</td>
<td>double-stranded</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>E6AP</td>
<td>viral E6-associated protein</td>
</tr>
<tr>
<td>ED</td>
<td>effector domain</td>
</tr>
<tr>
<td>Edn</td>
<td>endothelin</td>
</tr>
<tr>
<td>EED</td>
<td>embryonic ectoderm development</td>
</tr>
<tr>
<td>Efp</td>
<td>estrogen-responsive finger protein</td>
</tr>
<tr>
<td>EGR1</td>
<td>early growth response gene 1</td>
</tr>
<tr>
<td>ERCC</td>
<td>excision repair cross-complementing rodent repair deficiency</td>
</tr>
<tr>
<td>ES cells</td>
<td>embryonic stem cells</td>
</tr>
<tr>
<td>EWS</td>
<td>Ewings sarcoma gene</td>
</tr>
<tr>
<td>EXAFS</td>
<td>extended X-ray absorption fine structure</td>
</tr>
<tr>
<td>EZF-2</td>
<td>endothelial zinc finger protein-2</td>
</tr>
<tr>
<td>EZH</td>
<td>human homolog of enhancer of Zeste</td>
</tr>
<tr>
<td>FA</td>
<td>focal adhesion</td>
</tr>
<tr>
<td>FKB25</td>
<td>FK506-binding protein</td>
</tr>
<tr>
<td>FL(2)D</td>
<td>female-lethal(2)-d</td>
</tr>
<tr>
<td>FS</td>
<td>Frasier syndrome</td>
</tr>
<tr>
<td>FSHD</td>
<td>facioscapulo-humeral muscular dystrophy</td>
</tr>
<tr>
<td>FYVE</td>
<td>conserved in Fab1, YOTB Vacl and EEA1</td>
</tr>
<tr>
<td>GATA family</td>
<td>GATA sequence-binding zinc finger protein</td>
</tr>
<tr>
<td>GEF</td>
<td>GDP/GTP exchange factor</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylglucosamine</td>
</tr>
<tr>
<td>Gly</td>
<td>glycine</td>
</tr>
<tr>
<td>GO</td>
<td>8-oxoguanine</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione transferase</td>
</tr>
<tr>
<td>GTF</td>
<td>general transcription factor</td>
</tr>
<tr>
<td>H2TH</td>
<td>helix-two turns-helix</td>
</tr>
<tr>
<td>HAT</td>
<td>histone acetyl transferase</td>
</tr>
<tr>
<td>HD</td>
<td>Huntington's disease</td>
</tr>
<tr>
<td>HDAC</td>
<td>histone deacetylase</td>
</tr>
<tr>
<td>HIF-1a</td>
<td>hypoxia-inducible factor-1a</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HIV-1</td>
<td>human immunodeficiency virus type I</td>
</tr>
<tr>
<td>HIV-EP</td>
<td>human immunodeficiency virus type 1 enhancer binding protein</td>
</tr>
<tr>
<td>HMG</td>
<td>high-mobility group</td>
</tr>
<tr>
<td>HmL37ae</td>
<td>Haloarcula marismortui ribosomal protein L37ae</td>
</tr>
<tr>
<td>HMQC</td>
<td>heteronuclear multiple quantum coherence</td>
</tr>
<tr>
<td>HMTs</td>
<td>histone methyltransferases</td>
</tr>
<tr>
<td>HP1a</td>
<td>heterochromatin protein 1 alpha</td>
</tr>
<tr>
<td>HPC</td>
<td>human homolog of Polycamb</td>
</tr>
<tr>
<td>HPH</td>
<td>human homolog of Polyhomeotic</td>
</tr>
<tr>
<td>HPV</td>
<td>human papillomavirus</td>
</tr>
<tr>
<td>HSC</td>
<td>hemopoietic stem cell</td>
</tr>
<tr>
<td>HSED</td>
<td>Heteronuclear Spin-Echo Difference</td>
</tr>
<tr>
<td>HSP</td>
<td>heat shock protein</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum correlation</td>
</tr>
<tr>
<td>ICAD</td>
<td>inhibitor of calcium-activated DNAase</td>
</tr>
<tr>
<td>ICR</td>
<td>internal control region</td>
</tr>
<tr>
<td>IGR</td>
<td>immunoglobulin</td>
</tr>
<tr>
<td>IGF-1</td>
<td>insulin growth factor-1</td>
</tr>
</tbody>
</table>
IL-2 interleukin-2
ileu isoleucine
ILF3 interleukin enhancer binding factor 3
ILK integrin-linked kinase
IS insertion site
KAP-1 KRAB-associated protein 1
kb kilobase
kD (KDa) kilodalton
KLK Krüppel-like factor
LBD ligand binding domain
LRP-1 leader binding protein-1
Ldb LIM domain binding
LDD low density lipoprotein
Leu leucine
LFSE ligand field stabilization energy
Lhx LIM homeobox
LID LIM interaction domain
LIM-HD LIM homedomain
LMO LIM only protein
LSF late SV40 factor
LTR long terminal repeat
Lys lysine
MBP major histocompatibility complex enhancer binding protein; maltose binding protein
MIF middelastula transition
MDM murine double minute
Mdm2 murine double minute 2
MEFs mouse embryo fibroblasts
MGI mouse genome informatics
MIS Mullerian inhibiting substance
MMLV Moloney murine leukemia virus
MMPs matrix metalloproteinases
MOF males-absent-on-the-first
MOZ monocytic leukemia zinc finger protein
MTA2 metastasis-associated protein 2
MTF metalloregulatory transcription factor
MVE multivesicular endosome
MZF1 myeloid zinc finger gene 1
NB nuclear body
NCBI National Center for Biotechnology Information
NER nucleotide excision repair
NES nuclear export signal
NLS nuclear localization signal
NMR nuclear magnetic resonance
NO nitric oxide
NOE nuclear Overhauser effect
nt nucleotide
NuRD nucleosome remodeling deacetylase
OS Opitz syndrome
oTFIII A oocyte form of TFIII A
PAR4 prostate apoptosis response gene 4
Pc polycomb
PCAF p300-CBP associated factor
PCG polycomb group
Pcl polycomb-like
PCNA proliferating cell nuclear antigen
Peg3 paternally expressed gene 3
Pgc-2 PPARG coactivator-2
Ph polyhomeotic
PHA phytohemagglutinin
PHAX phosphorylated adaptor for RNA export
PHD plant homeobox domain protein; plant homeodomain
Pho pleiohomeotic
PI 3-kinase phosphatidylinositol 3-kinase
PI(3,5)P_{2} phosphatidylinositol 3,5-bisphosphate
PI3P phosphatidylinositol 3-phosphate
PMA phorbol myristate acetate
PMI promyelocytic leukemia protein
PML-Nbs promyelocytic leukemia protein-nuclear bodies
pol III RNA polymerase III
POZ poxvirus and zinc finger
PPARγ peroxisome proliferator-activated receptor gamma
ppm parts per million
PRDII-BFl positive regulatory domain II binding factor
PRE polycomb response element
Pro proline
R DNA degenerate alphabet A or G
Rb retinoblastoma protein
RBCC RING finger-B box-coiled coil
RCF replication factor C
RFP ret finger protein
RING really interesting new gene
RMSD root-mean-square-deviation
RNA pol II RNA polymerase II
RNAi RNA interference
RNP ribonucleoprotein particle
rp ribosomal protein
RPA replication protein A
RPB9 RNA polymerase II subunit 9
RRM RNA recognition motif
rRNA ribosomal RNA
RSS recombination signal sequences of V(D)J
Rang recombination
SAAl serum amyloid A1
SAS something about silencing
SBMA spinal and bulbar muscular atrophy
SCP1 SCAN-domain containing protein 1
SELEX systematic evolution of ligands by exponential enrichment
SCZ-2 SCAN-ZFP SCAN-domain-containing C2H2 zinc finger protein
Scm sex combs on midleg
SDF1 steroidalogenic factor 1
SDF1 steroidalogenic factor 1
SFT trimeter peptide
SJT solute carrier family 39
SNOC S-nitrosocysteine
SOX9 SRY-related HMG box
SP single positive
SREC steroid receptor coactivator-1
SRC-1 steroid receptor coactivator-1
SRE serum response element
SREBP serum response element binding protein
SRY sex-determining region Y gene
SS single-stranded
SSTR2 somatostatin receptor type II
sTFIII A somatic form of TFIII A
SUMO small ubiquitin-related modifier
TBP TATA-box binding protein;
TBP TATA-binding protein
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFIIA</td>
<td>transcription factor IIA</td>
<td>VEGF</td>
<td>vascular endothelial growth factor</td>
</tr>
<tr>
<td>TFIIH</td>
<td>transcription factor IIH</td>
<td>VHL</td>
<td>von Hippie Lindau</td>
</tr>
<tr>
<td>TFIIIA</td>
<td>transcription factor IIIA</td>
<td>W</td>
<td>DNA degenerate alphabet A or T</td>
</tr>
<tr>
<td>RRE</td>
<td>Rev response element</td>
<td>WAGR</td>
<td>Wilms tumor, aniridia, genitourinary malformations, and mental retardation</td>
</tr>
<tr>
<td>RRE-IIB</td>
<td>Rev response element stem loop IIB</td>
<td>wt</td>
<td>wild type</td>
</tr>
<tr>
<td>TTP</td>
<td>tristetraprolin</td>
<td>WT1</td>
<td>Wilms tumor suppressor gene 1</td>
</tr>
<tr>
<td>TGF-β</td>
<td>transforming growth factor beta</td>
<td>WTAP</td>
<td>WT1-associating protein</td>
</tr>
<tr>
<td>TIF1</td>
<td>transcriptional intermediary factor 1</td>
<td>X. laevis</td>
<td><em>Xenopus laevis</em></td>
</tr>
<tr>
<td>TIF1β</td>
<td>transcriptional intermediary factor 1 beta</td>
<td>Xi</td>
<td>X chromosome inactivation</td>
</tr>
<tr>
<td>Tip</td>
<td>TAT-interacting protein</td>
<td>Xlo</td>
<td>Xenopus oocyte type</td>
</tr>
<tr>
<td>tkv</td>
<td>thick veins</td>
<td>Xls</td>
<td>Xenopus somatic type</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane domain</td>
<td>XP</td>
<td>xeroderma pigmentosum</td>
</tr>
<tr>
<td>TNF</td>
<td>tumor necrosis factor</td>
<td>XPA</td>
<td>Xeroderma pigmentosum A</td>
</tr>
<tr>
<td>TPEN</td>
<td>N,N,N',N'-tetakis-(2-pyridylmethyl)ethylenediamine</td>
<td>YAF2</td>
<td>YY1-associated factor 2</td>
</tr>
<tr>
<td>TRAF</td>
<td>tumor necrosis factor receptor-associated factor</td>
<td>YL37a</td>
<td>yeast ribosomal protein L37a</td>
</tr>
<tr>
<td>TRIM</td>
<td>tripartite motif</td>
<td>YY1</td>
<td>Yin Yang 1</td>
</tr>
<tr>
<td>ts</td>
<td>temperature sensitive</td>
<td>ZBRK1</td>
<td>zinc finger and BRCA1-interacting proteins with KRAB domain 1</td>
</tr>
<tr>
<td>TSA</td>
<td>trichostatin A</td>
<td>ZF</td>
<td>zinc finger</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine</td>
<td>ZFP</td>
<td>zinc finger protein</td>
</tr>
<tr>
<td>URR</td>
<td>upstream regulatory region</td>
<td>ZIP</td>
<td>Zrt-, Irt-like protein</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
<td>ZNF</td>
<td>zinc finger</td>
</tr>
<tr>
<td>V(D)J</td>
<td>variable, diversity, and joining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>valine</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 1

The Discovery of Zinc Fingers and Their Practical Applications in Gene Regulation: A Personal Account

Aaron Klug

Abstract

A
r
t

count is given of the discovery of the classical Cys2His2 (C2H2) zinc finger, arising from biochemical studies on the protein transcription factor IIIA found in Xenopus oocytes, and of subsequent structural studies on its 3D structure and its interaction with DNA. Each finger is a self-contained domain stabilized by a zinc ion ligated to a pair of cysteines and a pair of histidines, and by an inner structural hydrophobic core. This work showed not only a novel protein fold but also a novel principle of DNA recognition. Whereas other DNA binding proteins generally make use of the symmetry of the double helix, zinc fingers can be linked linearly in tandem to recognize nucleic acid sequences of different lengths. This modular design offers a large number of combinatorial possibilities for the specific recognition of DNA (or RNA). It is therefore not surprising that this zinc finger is found widespread in nature, in 3% of the genes of the human genome.

It had long been the goal of molecular biologists to design DNA binding proteins for specific control of gene expression. It has been demonstrated that the zinc finger design is ideally suited for such purposes, discriminating between closely related sequences both in vitro and in vivo. The first example of the potential of the method was in 1994 when a three-finger protein was constructed to block the expression of an oncogene transformed into a mouse cell line. By fusing zinc finger peptides to repression or activation domains, genes can be selectively switched off and on. Several recent applications are described.

After the initial discovery, other types of zinc-binding domains which fold and interact with DNA or RNA in a different way were found, and these have become loosely grouped under the name of zinc finger proteins.

Introduction

After ten years of research on the structure of chromatin which had led to the discovery of the nucleosome and an outline of its structure, as well as the next level of folding of DNA in the 300Å chromatin fiber,1,2,3 I became interested in the then so-called “active chromatin,” the chromatin which is involved in transcription or poised to do so. I looked for a system which was tractable, that is, offered the possibility of extracting relatively large amounts of material for biochemical and structural studies.

I became intrigued by the work of Robert Roeder, then at Washington University, and Donald Brown, of the Carnegie Institution of Washington in Baltimore on the 5S RNA genes of Xenopus laevis, which are transcribed by RNA polymerase III (reviewed in Brown).3 They discovered that the correct initiation of transcription requires the binding of a 40 kD protein factor, variously called factor A or transcription factor IIIA (TFIIIA), which had been purified from oocyte extracts. By deletion mapping it was found that this factor interacts with a region about 50 nucleotides long within the gene, called the internal control region.

Immature oocytes store 5S RNA molecules in the form of 7S ribonucleoprotein particles,4 each containing a single 40 kD protein which was later shown5 to be identical with transcription factor IIIA. TFIIIA therefore binds both 5S RNA and its cognate DNA and it was therefore suggested that it may mediate auto-regulation of 5S gene transcription.5 Whether this autoregulation occurred in vivo or not, the dual interaction provided an interesting structural problem with could be approached because of the presence of large quantities of the protein TFIIIA in immature Xenopus oocytes.

In the autumn of 1982, I therefore proposed to a new graduate student, Jonathan Miller, that he begin studies on TFIIIA. This led to the discovery of a remarkable repeating motif within the protein, which we later, in laboratory jargon, called zinc fingers, because they contained zinc and gripped or grasped the DNA. The full story of the experiments is told in our first paper6 and I will only summarize it here. I should however emphasize that the repeating structure was discovered through biochemistry, not as some reviews have stated, by computer sequence analysis. When the sequence was published I looked for, and found by eye, a repeating pattern, which was then confirmed and aligned as a motif of 30 amino acids by Andrew McLachlan’s computer analysis.6

The 7S RNP Particle

When Miller repeated the published protocols for purifying the 7S particle, he obtained very low yields, which we attributed to dissociation. Brown and Roeder had used buffers which contained variously dithiothreitol (DTT), used because the protein had a high cysteine content, and/or EDTA to remove any contamination by metals which hydrolyse nucleic acids. We observed that gel filtration of the complex in 0.1 mM DTT resulted in separate elution of protein and 5S RNA. However, when we found that the strong reducing agent sodium borohydride did not disrupt the complex, we realized that the protein was not being held
When the particle was incubated with a variety of chelating agents, particle dissociation could be prevented only by prior addition of Zn. We therefore repeated the analysis with pure and undisassociated particle preparations, taking great care to ensure no contamination. We concluded that the native 7S particle contains between 7 and 11 zinc ions. This result was consistent with the fact that the protein contains large numbers of histidine and cysteine residues, the commonest ligands for zinc in enzymes and other proteins. This hinted at some kind of internal substructure.

A natural step was therefore to see if any such substructure could be revealed by proteolytic digestion, and Miller had already begun such studies. He found two products, an intermediate 33 kD fragment, and a limit 23 kD. At about that time Brown's group also showed that, on treatment with proteolytic enzymes, the 40 kD TFIIIA protein breaks down to a 30 kD product, which is then converted to a 20 kD product. They proposed that TFIIIA consists of three structure domains which they identified as binding to different parts of the 50 base-pair internal control region of the 5S RNA gene.

Carrying on these proteolytic studies, we found that on prolonged proteolysis the TFIIIA protein breaks down further, finally to a limit digest of about 3 kD. In the course of this breakdown, periodic intermediates differing in size by about 3 kD could be seen. The correspondence in size between these last two values suggested that the 30 kD domain of TFIIIA might contain a periodic arrangement of small, compact domains of kD. If each such domain contained one Zn atom, then the observed high Zn content would be accounted for.

This novel idea of small Zn-stabilized domains was strengthened by the timely publication by Roeder's group of the sequence of TFIIIA derived from a cDNA clone. By inspection, it could be seen that the large number of cysteines and histidines present in the protein appeared to occur in more or less regular patterns. A rigorous computer analysis showed that, of the 344 amino acids of the TFIIIA sequence residues, numbers 13-276 form a continuous run of nine tandemly repeated, similar units of about 30 amino acids, each containing two invariant pairs of histidines and cysteines. Repeating patterns in the sequence were also noticed by R S Brown, Sander and Argos who concluded, however, that the whole protein was divided into twelve repeats, indexed on a 39 amino acid unit (although their abstract states "about 30").

**A Repeating Structure for TFIIIA**

From the three different lines of evidence described above, namely (1) a 30 amino acid repeat in the sequence, which (2) corresponds in size to the observed periodic intermediates and the limit-digest product of 3 kD, and (3) the measured Zn content of 7-11 atoms, we proposed that most of the TFIIIA protein has a repeating structure in which each of the nine 30 amino acid units folds around a Zn ion to form a small independent structural domain, the "finger." 25 of the 30 amino acid residues form a loop around the central Zn ion and the five intervening amino acids provide the linkers between consecutive fingers. The Zn ion forms the basis of the folding by being tetrahedrally coordinated to the two invariant pairs of cysteines and histidines. Each repeat also contained besides this unique conserved pattern of Cys-Cys .... His-His, three other conserved amino acids, namely Tyr6 (or Phe6), Phe17 and Leu23, all of which are large hydrophobic residues. The whole of the 30 amino acid repeat is rich in basic and polar residues but the largest number are found concentrated in the region between the second cysteine and the first histidine, implicating this region in particular in nucleic acid binding.

Formally, when indexed on a 30 amino acid repeat, the repeating structure could be written at

```
1 5  9 13 17 21 25 29 30
```

where h represents a conserved (large) hydrophobic residue. The proposal that each 30 amino acid unit formed an independently folded, Zn-stabilized domain soon gained support from two lines of research. First, we carried out a study using EXAFS (extended X-ray absorption fine structure) confirming that the Zn ligands are two cysteines and two histidines. Secondly it was found by Tso et al that, in the DNA sequence of the gene for TFIIIA, the position of the intron-exon boundaries mark most of the 9 proposed finger domains.

In evolutionary terms, the multi-fingered TFIIIA may have arisen by gene duplication of an ancestral domain comprising about 30 amino acids. Because one such self-contained small domain would have had the ability to bind to nucleic acids, and could be passed on by exon shuffling, we suggested that these domains might occur more widely in gene control proteins than in just this case of TFIIIA. The extent to which this prediction has been borne out (3% of the genes of the human genome, at the latest count) still, on occasion, astonishes me. Indeed, within months of the publication of our paper, I received word of sequences homologous to the zinc finger motif of TFIIIA. The first two were from *Drosophila*, the serendipity gene from Roshbash's group and Kruppel from Jackie's group.

The key point that emerged from our first paper was that, not only had there emerged a novel protein fold for nucleic acid binding, but also a novel principle of DNA recognition. The overall design for specific DNA recognition was distinctly different from that of the helix-turn-helix motif found in the first DNA binding proteins to be described. The latter binds to DNA as a symmetric dimer to a palindromic sequence on the DNA, thus making use of both the 2-fold symmetry of the DNA helix backbones and also the nucleotide sequence. (Later heterodimeric variations of this and related designs were found, but they still make use of the helix symmetry).

In contrast, the zinc finger is a module that can be used singly or linked tandemly in a linear fashion to recognize DNA (or RNA) sequences of different lengths. Each finger domain has a similar structural framework but can achieve chemical distinctiveness through variations in a number of key amino acid residues. This modular design offers a large number of combinational possibilities for the specific recognition of DNA (or RNA). It was not surprising that it is found widespread throughout so many different types of organisms.
The Structure of the Zinc Finger and Its Interaction with DNA

We had noted,⁶ that in addition to the characteristic arrangements of conserved cysteines and histidines which are fundamental in the folding of the finger by the coordinating Zn, there are several other conserved amino acids, notably Tyr6, Phe17 and Leu23, and that were likely to form a hydrophobic structural core of the folded structure. In other words, the seven conserved amino acids in each unit provide the framework of tertiary folding, whereas some of the variable residues determine the specificity of each domain. Jeremy Berg¹⁵ built on these original observations by fitting known structural motifs from other metallo-proteins to the consensus sequence of the TFIIIA finger motifs. His proposed model consisted of an antiparallel β-sheet, which contains the loop formed by the two cysteines and an α-helix containing the His-His loop. The two structural units are held together by the Zn atom. In analogy with the way in which the bacterial helix-turn-helix motif binds DNA, DNA recognition was postulated to reside mostly in the helical region of the protein structure.

Berg’s model was soon confirmed by the NMR studies of Peter Wright’s group¹⁶ on a single zinc finger in solution, and subsequently by David Neuhaus in our laboratory¹⁷,¹⁸ on a two-finger peptide. Our work took longer to solve but had the merit of showing that adjacent zinc fingers are structurally independent in solution, being joined by flexible linkers.

The question remained of the precise pattern of amino acid interactions of zinc fingers with DNA. The breakthrough came in 1991 when Nikola Pavletich and Carl Pabo,¹⁹ both then at Johns Hopkins, solved the crystal structure of a complex of a DNA oligonucleotide specifically bound to the three-finger DNA binding domain of the mouse transcription factor Zif268, an early response gene. The primary contacts are made by the α-helix which binds in the DNA major groove through primary hydrogen-bond interactions from helical positions 2, 3, and 6 to three successive bases (a triplet) on one strand of the DNA, and through a secondary interaction from helical position 2 on the other strand. This is the canonical docking arrangement, but there

In our laboratory, my colleagues and I adopted a different approach. The reason was that the “rules” did not take into account the fact that real DNA structures are not fixed in the canonical B form but there are wide departures, depending on the DNA sequence.²⁰,²¹ This was further brought home to us by the structure of the second zinc finger DNA complex to be solved, that with the Drosophila tramtrack protein, by Fairall et al²² in Daniela Rhodes’s group our laboratory. Here the helical position used for the primary contact with the 3’-most base of one of the triplets (thymine) is not -1, but 2. The reason was that the DNA structure was much distorted from the B form, with the thymine followed by the adenine at a helical rotation angle of 39° rather than the canonical 36°. The reason is that the T-A step in DNA is unstable as I had noted long ago.²⁰

Affinity Selection from a Library of Zinc Fingers by Phage Display

The alternative to this rational but biased design of proteins with new specificities is the isolation of desirable variants from a large pool or library. A powerful method of selecting such proteins is the cloning of peptides or protein domains as fusions to the minor coat protein (pIII) of bacteriophage fd, which leads to their expression on the tip of the capsid. Phage displaying the peptides of interest can then be affinity purified by binding to the target and then amplified for use in further rounds of selection and for DNA sequencing of the cloned gene. We applied this technology to the study of zinc finger-DNA interactions, after my colleague, Yen Choo, demonstrated that functional zinc finger proteins could be displayed on the surface of fd phage, and that such engineered phage could be captured on a solid support coated with the specific DNA.²⁵,²⁶ The phage display method was also adopted by other groups working on zinc fingers, including those of Carl Pabo and Carlos Barbas.

We created phage display libraries comprising about 10⁷ variants of the middle finger from the DNA-binding domain of Zif268. A DNA oligonucleotide of fixed sequence was used to bind and hence purify phage from this library over several rounds
such application,27 in which we built a protein which recognized a specific DNA sequence both in vitro and in vivo. This was a crucial test of our understanding of the mechanism of zinc-finger DNA recognition. The proof of principle stimulated ourselves, and later others, to devote our future studies to potential applications in gene regulation for research purposes or for therapeutic correction.

In summary, we created a three finger peptide able to bind site-specifically to a unique nine base-pair region of the p190 bcr-abl cDNA: this is a transforming oncogene which arises by translocation between the tips of chromosomes 9 and 22, of which one product is the Philadelphia chromosome. The latter contains a novel DNA sequence at the junction of two exons, one each from the two genomic parent bcr and abl genes. Our engineered peptide discriminated in vitro against like regions of the parent bcr and c-abl genes, differing in only a single base, by factors greater than one order of magnitude.

Our peptide also contained a nuclear localization signal fused to the zinc finger domain so that the peptide could accumulate in the nucleus. Consequently, stably transformed mouse cells, made interleukin-3 independent by the action of the oncogene, were found to revert to IL-3 dependence on transient transfection with a vector expressing the peptide. Our construct was also engineered to contain a c-myc epitope, which enabled us to follow by immunofluorescence the localization of the peptide to the nuclei of the transfected cells. When IL-3 is subsequently withdrawn from cell culture, over 90% of the transfected p190 cells become apoptotic (that is, showing chromosome degradation) within 24 hours. Our experiments were repeated on cells transformed by another related oncogene p210 bcr-abl, which served as a control. All transfected p210 cells maintained their IL-3 dependence, and remained intact on entry of the engineered peptide.

Measurements of the levels of p190 bcr-abl mRNA extracted from cells treated with the peptide showed that the repression of oncogenic expression by the zinc finger peptide was due to a transcriptional block imposed by the sequence-specific binding of the peptide, which presumably obstructed the path of the RNA polymerase.

**Promoter-Specific Activation by Zinc Finger**

These experiments showed that a zinc finger peptide could be engineered to switch off gene expression in vivo. In the same paper27 we described other experiments on a different cell system (cultured mouse fibroblasts) to show that a gene could also be switched on in a similar way. We used the same nine base pair sequence, but this time as a promoter for a CAT reporter gene contained in a plasmid. The peptide, which recognized the promoter, was fused to a VP16 activation domain and, on transient transfection, stimulated expression of the reporter gene by a factor of 30-fold above controls.

**Improving Zinc Finger Specificity**

(i) Our more recent work has focused on improving the specificity of recognition by zinc fingers of the DNA target. While the main source of specificity lies in the amino acids at positions -1, 3 and 6 of the recognition α-helix of a zinc finger for successive bases lying on one strand of a DNA triplet, we found that the “cross-stand” interaction described above from helical position 2 to the neighboring base pair on the adjacent triplet can significantly influence the specificity.28 Therefore it has been necessary to revise the simple model that zinc fingers are essentially independent modules that bind three base-pair subites to a model that considers functional synergy between adjacent independently folded zinc fingers. In this revised model, Zif268-like zinc fingers potentially bind four base-pair overlapping subites. We therefore redesigned our method of phage library construction to take account of this refinement,29 which also has the merit of being more widely applicable and rapid.

(ii) An important step forward has been to increase the length of the DNA sequence targeted and hence its degree of rarity. Three zinc fingers recognize nine base pairs, a sequence which would occur randomly several times in a large genome. However six fingers linked together would recognize a DNA sequence 18 base pairs in length, sufficiently long to constitute a rare address in the human genome. One cannot simply go on adding fingers, because the periodicity of packed fingers does not quite match the DNA periodicity, so that they get out of register. We have learned how to engineer longer runs of zinc fingers which can target longer DNA sequences.30,31 By fusing functional groups to the engineered DNA binding domains, for example silencing or activation domains, highly specific transcription factors can be generated to up- or down-regulate expression of a target gene.

**Some Applications of Engineered Zinc Fingers**

Examples of some recent applications by ourselves and others, using either three finger or six finger peptides are:

(i) inhibition of HIV-1 expression;32

(ii) the disruption of the infective cycle of infection by herpes simplex virus;33

(iii) activating the expression of VEGF-A in a monkey kidney cell line;34

(iv) activating the expression of vascular endothelial growth factor (VEGF) in a human cell line, and in an animal model; and35

(v) regulation of zinc finger expression by small molecules.36

**Other Classes of Zinc Fingers and Zinc-Binding Domains**

Shortly after the classical C2H2 zinc finger was discovered in TFIIIA, sequence motifs that appeared to be related were found in several other protein or cDNA sequences of molecules which bound DNA. It was therefore at first thought that these might have a rather similar structure to the TFIIIA type finger domains. The most important and widespread examples are those from members of the superfamily of hormone-activated nuclear receptors which play a central role in the control of eukaryotic gene expression, and are indeed transcription factors.37,38

The DNA-binding domains (DBDs) of such receptors all include two motifs in tandem, each about 30 amino acids long, but each motif contains two pairs of cysteines rather than a pair of cysteines and a pair of histidines as in the first class. They do indeed bind Zn2+, but the three-dimensional structure of two such DBDs, determined in solution using 2-D NMR spectroscopy, showed that the receptor DBD is structurally distinct from the TFIIIA type of zinc finger.39,40 The two motifs in each domain each fold up into an irregular loop followed by an α-helix, but the two together form a single structural unit with their helices crossing at right
angles, so that the DNA recognition helix (from the first motif) is supported by the helix from the second motif.

Hormone receptors bind to palindromic sites (response elements, RE) on the DNA as dimers, and the DBDs alone also form dimers, the dimer interface arising from a region of the loop of the second motif of each receptor. These different roles for the two motifs within on structural unit, namely helix recognition and dimerization, were deduced by mapping onto the three-dimensional structures the site-directed mutagenesis data from a number of laboratories, particularly those of P. Chambon, R. Evans and G. Ringold. This combination of structural analysis and biochemical and genetic experiments pointed toward a mechanism of interaction with DNA and a general model was proposed by Hard et al.\(^3\) and Schwabe et al.\(^4\)

The detailed chemistry of the interactions at the protein-DNA interface, however, had to await a crystal structure of a complex. The first to be determined was that at 2.9 Å resolution by Luisi and Sigler\(^41\) of the DBD of the glucocorticoid receptor (GR) complexed with a DNA segment 18 bp long. However, the DNA segment used was composed of two half sites (each of 6 bp) separated by a non-native spacing, with four (rather than three) intervening base pairs. As a consequence of this, the two DBDs did not bind equivalently to DNA.

The crystal structure of a second hormone-receptor-DNA complex with a correct cognate binding site was solved (at 2.4 Å resolution) by Schwabe et al.\(^42\) in our laboratory. This was of the DBD of the estrogen receptor which recognizes a different DNA half site from GR but with the same native separation of 3 base pairs between half sites (ERE\(_3\)). The protein binds as a symmetrical dimer in an equivalent manner to both half sites. The interactions seen in this complex are characteristic in number and type of those later found in specific interactions in other families of protein-DNA complexes.

Clearly the second class of zinc finger DNA-binding domains is not a simple variant of the first TFIIIA class. They differ both in their structure and the way in which they interact with DNA. Above all the GR and ER receptors operate as dimers which bind to palindromic DNA sites, whereas the binding of zinc fingers of the first C\(_2\)H\(_3\)2 class makes no use of the symmetry of the DNA structure nor of base sequence. The latter function as independent dimers which can be strung together in a directly repeating (tandem) fashion with no restriction on their number.

It should however be added that some members of the nuclear receptor family (for example thyroid, vitamin D, retinoic acid) also bind as dimers, but as nonsymmetrical dimers to a DNA binding site made of two directly repeated identical "half-sites." Here clearly another interface is brought into play, but again the discrimination depends entirely on the separation between the repeats in the DNA sequence.\(^43\)

The structure of a member of a third class of zinc binding domain of a distinct structural type emerged soon afterwards.\(^44\) This was the GAL4 transcriptional activator, representative of a small family found only in yeast. Here two zinc ions and six cysteines in the DBD form a binuclear cluster with each Zn\(^2+\) coordinated by four cysteines, so that two of the cysteines are shared. This cluster holds together two short helices, related by a quasi dyad, one of which inserts into the major groove of the DNA as in the hormone receptors; we thus see the recognition helix supported by a second helix. Again, the molecule binds as a dimer to a palindromic DNA binding site, with two short half-sites separated by approximately 1.5 turns of the DNA helix, so that the DBDs bind on opposite faces of the DNA.

The fourth class of zinc fingers discovered was constituted by the Cys-X\(_2\)-Cys-X\(_4\)-His-X\(_4\)-Cys sequences found in the nucleocapsid proteins of retroviruses, which form 'stubby' fingers. Thereafter several more types were found, which are described in later chapters in this book.

This general use of zinc became increasingly clear, even though the structural information was often still limited. A diverse set of families of proteins which bind Zn\(^2+\) and interact with nucleic acids were uncovered and became loosely grouped under the name of zinc finger proteins. The term was not inappropriate since the zinc-binding domains in all these cases do grip or grasp the double helix. Nor would it be inappropriate if the domain does not bind DNA or RNA but another protein, as was foreshadowed,\(^46\) and as has been increasingly found.

However in more recent years the term zinc finger has begun to be used even when the zinc-binding motif or amino acid sequence does not form an independent, self-folding domain or minidomain. I personally would prefer the term to be restricted to the latter cases. The cases where a Zn ion, say, is merely used to link together two regions of a protein, belong to the field of the bioinorganic chemistry of zinc. Its widespread use in proteins for such structural purposes must be attributed to the fact that zinc is what I would call a "safe" metal, in that it has only one main oxidation state and hence has no redox chemistry, as for instance copper and iron do. As a structural element in a protein it has the advantage over a disulfide bridge in bringing together two parts of a protein, because it cannot be reduced in the reducing atmosphere inside a cell.

Postscript

While, as described in the Introduction, it has long been known that TFIIIA binds RNA as well as DNA, there is now increasing evidence that zinc fingers are widely used to recognize RNA. However the molecular basis of the recognition has remained elusive. In this laboratory, we had not forgotten that TFIIIA binds 5S RNA. After our failure to crystallize the native 7S RNP particle itself (or a reconstituted version) we set out\(^47\) to find relevant subcomplexes which could show the main interactions between TFIIIA and the RNA. We succeeded and have recently determined the X-ray structure of a zinc finger RNA complex which reveals two modes of zinc finger binding, both different from that for DNA.\(^48\)

References

Zinc Finger Proteins: From Atomic Contact to Cellular Function


47. Searles MA, Lu D, Klug A. The role of the central zinc fingers transcription factor IIIA in binding to SS RNA. J Mol Biol 2000; 301:47-60.

C$_2$H$_2$ Zinc Fingers As DNA Binding Domains

Shiro Iuchi

Abstract

A great number of C$_2$H$_2$ zinc finger proteins selectively bind to specific DNA sequences and play a critical role in controlling transcription of genes. The specific binding is achieved by zinc finger domains with $\beta\beta\alpha$ structure that is formed by tetrahedral binding of Zn$^{2+}$ ion to the canonical cysteine and histidine residues. Two to three tandem zinc fingers are necessary and sufficient for the specific binding without participation of any other domains. Zinc fingers bind in the major groove of the DNA, wrapping around the strands, with specificity conferred by side chains of several amino acid on the $\alpha$ helices. Some zinc finger proteins undergo homodimerization by hydrophobic interactions or by finger-finger binding and reinforce the specific binding to DNA. Conserved linkers between tandem fingers are necessary for stabilizing the DNA complex. Regulatory mechanisms of zinc finger binding to DNA are emerging. Some cellular factors are found to acetylate and phosphorylate zinc fingers and the linkers of a few proteins. These modifications alter the binding activity of the zinc finger proteins and hence control expression of their target genes. Other factors can methylate promoter regions of genes. This modification alters affinity of zinc finger proteins for the DNA segments and hence controls expression of their target genes.

Introduction

The C$_2$H$_2$ zinc finger consists of twenty to thirty amino acid residues that have a special secondary structure stabilized by zinc tetrahedral binding to two cysteine and two histidine residues. Proteins with these zinc fingers are called C$_2$H$_2$ zinc finger proteins. The C$_2$H$_2$ zinc finger protein family is the largest group of all zinc finger protein families (Table 1) and the second largest group of all protein classes after the envelope glycoprotein GP120 family. The proteins are present in prokaryotes as well as eukaryotes and are abundant in mammals. More than 700 human genes, or greater than 2% of the total human genes, encode C$_2$H$_2$ zinc finger proteins. Not surprisingly, C$_2$H$_2$ zinc finger proteins participate in a variety of cellular activities including development, differentiation, and tumor suppression. Among C$_2$H$_2$ zinc finger proteins, many bind to DNA duplexes in a finger-mediated specific manner and participate in controlling expression of the target genes. C$_2$H$_2$ zinc fingers, which are often described as $\text{X}_2\text{CX}_2\text{CX}_2\text{HX}_2\text{SH}$ to show the intervals between the zinc binding residues, contain two $\beta$ strands and one $\alpha$ helix. All the primary, secondary and tertiary structures are important for binding to DNA duplexes. In this chapter the general features of C$_2$H$_2$ zinc fingers are described.

Folding of C$_2$H$_2$ Zinc Fingers

Requirement of zinc ion for transcription factors to bind their cognate DNA was found in Xenopus TFIIIA (Transcription Factor IIIA) first. The requirement was due to the ion coordination with a small peptide domain, named zinc finger, that contains two canonical cysteine and histidine residues. Now, thousands of proteins are known to have C$_2$H$_2$ zinc fingers and the majority of the fingers are thought to bind to DNA. Zinc finger proteins can take three states: unfolded, folded and DNA-bound forms (Fig. 1 and for details of folded zinc finger domain see ref. 1 and 3 as well as Fig. 1 of Chapter 8). Unfolded zinc fingers do not bind to target DNA, but folded fingers bind to the cognate DNA duplexes. The protein molecules of the DNA complex are usually associated with other transcription factors that bind to different domains on the zinc finger protein, or to different parts of the zinc fingers. The C$_2$H$_2$ zinc finger motif contains all the information necessary for its folding but folds properly only when Zn$^{2+}$ binds to the canonical residues, two cysteines and two histidines. The change in Gibbs free energy ($\Delta G$) of the folding is enthalpy-driven. The amount, about -8.8 kcal per mole (Table 2), indicates that the folded fingers are very stable. C$_2$H$_2$ zinc fingers contain three conserved hydrophobic amino acids at position -12, -3 and +4 in addition to the two canonical cysteines and histidines (see bolded amino acids in the sequence below where the first amino acid residue of the $\alpha$ helix is designated as position 1). It has been shown that these seven amino acid residues are necessary and sufficient to fold peptides properly by using a designed-synthetic peptide, K(-13)-YACAACTAFAKAKAAHAAA-K13. This peptide binds Co$^{2+}$, a substitute for Zn$^{2+}$, to the cysteines with a higher affinity than to the histidines, and the proper folding occurs only when the ratio of the ion to the peptide is one or higher. Similar binding of Zn$^{2+}$ to the zinc finger motif has also been observed in the synthetic peptide of the zif268 third finger [F(-12)-ACDICGRKSRDSERKHTKHLRQ-K15]. The authors of this work have proposed that the zinc finger folding begins with binding of Zn$^{2+}$ to the canonical cysteines and then establishes the tetrahedral structure involving the histidines. They have also suggested that the $\alpha$ helix emerges at the S(1)-DERKRTIK-H11 sequence as the metal ion binding proceeds.

Tandem C$_2$H$_2$ Zinc Fingers

C$_2$H$_2$ zinc finger proteins often contain the fingers as tandem repeats connected by short oligopeptides, called linkers. Based on the number and repeat pattern of the fingers, C$_2$H$_2$ zinc
Type
C2H2
CCHC (C2HC)
RING
LIM
C2C2 (2x)*
CCCH (C2CH)
TAZ

Table 1. Number of zinc finger proteins reported as of May 2003

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of Proteins</th>
<th>Average Domain's Length (AA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2H2</td>
<td>23,989</td>
<td>23</td>
</tr>
<tr>
<td>CCHC (C2HC)</td>
<td>5,215</td>
<td>17</td>
</tr>
<tr>
<td>RING</td>
<td>2,010</td>
<td>41</td>
</tr>
<tr>
<td>LIM</td>
<td>1,246</td>
<td>58</td>
</tr>
<tr>
<td>C2C2 (2x)*</td>
<td>1,243</td>
<td>65</td>
</tr>
<tr>
<td>CCCH (C2CH)</td>
<td>1,019</td>
<td>26</td>
</tr>
<tr>
<td>TAZ</td>
<td>51</td>
<td>81</td>
</tr>
</tbody>
</table>

Data are taken from Pfam 9.0 at Washington University in St. Louis. *The number is for proteins containing two C2C2 fingers.

finger proteins can be divided into four classes (Fig. 2). (A) single C2H2, (B) triple C2H2, (C) multiple-adjacent C2H2, and (D) separated-paired C2H2 zinc finger proteins. This classification is useful to predict how the zinc fingers of the proteins exert their binding activity. The single zinc finger differs from the other zinc fingers in that it requires an additional, non-zinc finger domain to establish the binding to the target DNA. The other classes, that is the triple, multiple-adjacent and separated-paired C2H2 fingers, bind to the specific DNA sequence without the aid of other domains. Both the triple C2H2 and the multiple-adjacent C2H2 zinc fingers bind to the cognate DNA at the three consecutive fingers. Another experiment has shown that a finger peptide with four-tandem identical repeats binds to the target DNA sequence at the three consecutive fingers only. Furthermore, the separated-paired C2H2 finger can also specifically bind to the cognate DNA, often at the one pair finger. Taking all these results into account, it may be concluded that two to three successive C2H2 zinc fingers are the most suitable unit to specifically bind to the cognate DNA.

Multiple-adjacent C2H2 zinc fingers bind to the DNA, based on the rule that two to three successive fingers are responsible for the specific DNA binding, but the fingers have additional DNA contacts. For example, TFIIIA with nine zinc fingers establishes the DNA binding at fingers 1-3, but briefly touches the DNA at finger 5 and weakly binds to the DNA at fingers 7-9. Another example is Zac. This has seven zinc fingers and binds to a GC rich DNA duplex. Biochemical and genetic analysis have shown that Zac binds to the DNA at the two consecutive fingers (finger 2-3) and also at the three consecutive fingers (finger 5-7), without involving finger 4. Zac contacts the DNA at more than three fingers, keeping the two to three finger-DNA binding rule. It is curious that extra fingers are present in multiple-adjacent zinc finger and separated-paired zinc finger proteins, but it is becoming increasingly clear that the extra fingers are engaged in other interactions not only with a secondary locus of DNA but also with distinct molecules such as RNA and proteins.

It is interesting to speculate how these tandem zinc fingers have evolved from a single finger. The Echerichia coli gene, argE, encodes a helix-turn-helix DNA binding repressor protein for the genes directing aerobic respiration enzymes. This gene has an identical eleven-nucleotide sequence flanking a short region, and this genetic organization prompts the region to duplicate spontaneously (Luchi and Lin, unpublished). Similarly, the identical linker sequences flanking the C2H2 zinc finger DNA may allow a zinc finger sequence to duplicate together with the linker itself. This duplication would in turn stimulate the gene duplication further by using the linkers again or the zinc finger itself. The multiple zinc fingers can gain a better affinity for a DNA sequence and give the zinc finger protein molecules a selective advantage over the original finger proteins to function as transcription factors. This speculation is consistent with the fact that many of the separated-paired C2H2 and triple C2H2 zinc fingers conserve amino acid residues between tandem zinc fingers. After the duplication, point mutations would further improve its affinity for the target. Alternative splicing within the gene or recombination with another zinc finger gene would allow the finger peptide to acquire tandem hybrid zinc fingers with a very different specificity. It should not be difficult to show that gene duplication indeed causes the evolution in living cells using well-established bacterial and yeast genetic systems.
Table 2. Thermodynamic parameters for zinc finger reactions

<table>
<thead>
<tr>
<th>Reaction</th>
<th>ZF</th>
<th>$K_a$ (M$^{-1}$)</th>
<th>$\Delta G$</th>
<th>$\Delta H$</th>
<th>$TAS$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ZFP + Zn}^{2+} \leftrightarrow \text{ZFP-Zn}^{2+}$</td>
<td>1</td>
<td>$3.5 \times 10^6$</td>
<td>-8.8</td>
<td>-9.3</td>
<td>-0.5</td>
<td>11</td>
</tr>
<tr>
<td>$\text{ZFP-Zn}^{2+} + \text{DNA} \leftrightarrow \text{ZFP-Zn}^{2+}\text{DNA}$</td>
<td>2</td>
<td>$2.2 \times 10^{10}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>33</td>
</tr>
<tr>
<td>$\text{ZFP-Zn}^{2+} + \text{DNA} \leftrightarrow \text{ZFP-Zn}^{2+}\text{DNA}$</td>
<td>3</td>
<td>$2.8 \times 10^8$</td>
<td>-11</td>
<td>-6.9</td>
<td>4.5</td>
<td>29</td>
</tr>
<tr>
<td>$\text{ZFP-Zn}^{2+} + \text{DNA} \leftrightarrow \text{ZFP-Zn}^{2+}\text{DNA}$</td>
<td>4</td>
<td>$1.3 \times 10^7$</td>
<td>-9</td>
<td>-6.9</td>
<td>2.6</td>
<td>31</td>
</tr>
<tr>
<td>$\text{ZFP-Zn}^{2+} + \text{DNA} \leftrightarrow \text{ZFP-Zn}^{2+}\text{DNA}$</td>
<td>5</td>
<td>$2.6 \times 10^7$</td>
<td>-10</td>
<td>NA</td>
<td>NA</td>
<td>32</td>
</tr>
<tr>
<td>$\text{ZFP-Zn}^{2+} + \text{DNA} \leftrightarrow \text{ZFP-Zn}^{2+}\text{DNA}$</td>
<td>6</td>
<td>$8.8 \times 10^8$</td>
<td>-12</td>
<td>7</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>$\text{ZFP-Zn}^{2+} + \text{DNA} \leftrightarrow \text{ZFP-Zn}^{2+}\text{DNA}$</td>
<td>7</td>
<td>$1.8 \times 10^8$</td>
<td>-8</td>
<td>-11</td>
<td>-2.8</td>
<td>31</td>
</tr>
</tbody>
</table>

*ZFP denotes zinc finger peptide. ZF shows number of zinc fingers per finger peptide. The relation between Gibbs free energy, enthalpy and entropy is given by $\Delta G = \Delta H - TAS$, where $T$ is absolute temperature. Negative $\Delta G$ value suggests that each above reaction favors the association-over the dissociation-reaction. The unit of energy is kcal/mol.*

Overall Features of the Triple-C$_2$H$_2$ Zinc Fingers Binding

The change in Gibbs free energy of zinc finger-DNA binding is similar to or higher than that of the zinc finger folding (Table 2) and the binding requires no enzyme action. $\Delta G$ is constant over a biologically-meaningful temperature range, 5 to 45 °C. The reactions are mostly enthalpy- and entropy-driven reactions, but some are only enthalpy- or only entropy-driven. The $\Delta G$ values suggest that the C$_2$H$_2$ zinc fingers bind to DNA as strongly as some antibodies, whose $K_a$ for their antigens is $10^9$ to $10^{12}$ M$^{-1}$. These zinc finger-DNA complexes are of such high affinity that they routinely display an electrophoretic mobility shift on native polyacrylamide gels.

When an amino acid and guanine are mixed in water, the amino acid starts to associate with the nucleotide base and the reaction soon reaches equilibrium. Tendency of the reaction is described by the affinity constant, $K_a = 1/K_d = [\text{aa-guanine}] / [\text{aa}] [\text{guanine}]$. The $K_a$ of amino acids for guanine is in the order arginine > lysine > glutamine > glutamate > glycine. Amino acids also have an inherent $K_a$ for the three other bases. Accordingly, it is predictable that the amino acid-base associations are key for zinc finger peptides to specifically bind to the DNA duplexes. Indeed, the arginine-guanine contact, whose association is the greatest of all the combinations, is quite often present in zinc finger-DNA complexes (Fig. 4B and C). However, each amino acid residue of the zinc finger peptides does not have free mobility to access to the favorite bases due to the rigid zinc finger structure. How the zinc finger peptide recognizes and binds to the cognate DNA duplex is a big question for the zinc finger-DNA interactions and the answer has come mostly from structure-oriented and genetic oriented analyses of zinc finger-DNA complexes. In particular, the structure of Zif268-DNA and TFIIIA-DNA complexes contributed enormously to the understanding of how the triple fingers align with the DNA duplex and how each finger interacts with the nucleotides of the DNA. Three $\alpha$ helices of the Zif268 triple zinc fingers bind in the major groove of the target DNA duplex antiparallel to the primary strand (defined as the strand to which zinc finger contacts most, Fig. 4A and for details of the binding see ref. 2), making hydrogen bonds and forming hydrophobic interactions with nucleotide bases and wrapping around the DNA for almost one turn. In addition to hydrogen bonds and hydrophobic interactions, phosphate contacts also participate in the zinc finger-DNA complex formation. Phosphate contacts, linking to the DNA backbone, may not be significant in determining the specificity but appear to be important for strengthening the binding. Through further analysis of various zinc finger–DNA complexes, it was found that the overall DNA binding mode of Zif268 is shared with other tandem C$_2$H$_2$ zinc fingers. These fingers include TFIIIA, SP1, GL1, WT1 and Tramtrack as well as the single finger GAGA. In addition to these GC-rich DNA binding transcription factors, an AT-rich DNA recognizing transcription factor, CF2II, is also thought to take the same binding mode. However, one AT-rich DNA binding transcription factor, Nmp4, is proposed to associate with minor groove of the DNA.

Figure 2. Schematic representation of four classes of C$_2$H$_2$ zinc finger proteins. Only one example is shown for each class. Some multiple-adjacent C$_2$H$_2$ zinc finger proteins contain more than thirty zinc fingers. The number and the pattern of C$_2$H$_2$ zinc fingers indicate how the zinc fingers are involved in the DNA binding (see text).
Figure 3. A model of intramolecular duplication of the zinc finger region. ZF1 flanked by two identical linker sequences for (L) would form a loop overlapping at the identical sites, or form a hairpin structure complementing within the single strand at the identical sites. When DNA polymerase reads the loop or the hairpin structure twice, then the zinc finger region is duplicated together with the linker sequence.

DNA of the Zif268-DNA complex takes a slightly unwound B form so that the major groove is still wide and can be deep. This form contains 11.3 bp per turn, which is slightly more bases than the B-form itself (10.5 bp per turn).

Side Chain-Base Contacts in C2H2 Zinc Finger-DNA Complexes

Amino acid side chains on the surface of α helices in Zif268 are exposed to the cognate DNA duplex in the major groove, and the side chains at position 6, 3, 2 and -1 contact selectively with four successive bases (subsite) (Fig. 4B). Residues at position 6, 3 and -1 bind to three successive bases of the primary strand, the strand contacted most by the side chains, and the residue at position 2 binds to the fourth base present in the complementary strand. In this way, Zif268 finger 3, 2 and 1 recognize the primary strand’s subsite, 5’-GGGT-3’, 5’-TGGG-3’ and 5’-GGGT-3’, respectively. Of the four base pairs in the subsites, the base pairs at the end is shared by adjacent fingers. Consequently, the Zif268 triple zinc fingers bind to the ten nucleotide base pair, 5’-GGGTGGGGG-3’. Based on statistics, one can predict with a high accuracy which side chain-base contacts can happen at the key positions of the α helices in the Zif268 finger context. Stereochemistry between amino acids and bases can also predict the contacts with a similar accuracy or somewhat less accuracy.

The rule that side chains positioned at 6, 3, 2 and -1 contact bases is well preserved throughout the family of C2H2 zinc fingers, but additional contacts can occur in other fingers. Such an example has been observed in the TFIIIA-DNA complex (Fig. 4C). There is an additional contact of the side chain at position 10. Furthermore, some side chains at the regular positions 6, 3, 2 and -1 reach bases out of the subsites. When Zif268 was engineered to bind to AT-rich DNA duplexes, many irregular contacts occurred. In fact, many of these irregularities occurred in the complementary strand but not in the primary strand. These facts, together with results obtained by mutational analyses, have led to the conclusion that there is no simple code for side chain-base contacts. Superimposition of several finger-DNA complex images showed that the irregular contacts coincide with the slight difference in docking angle of the α helix to DNA. The difference may reveal the influences of all of the factors involved in specific zinc finger-DNA binding, including amino acid residues within the subsite, linkers, and adjacent fingers.

Absence of a strict rule in the side chain-base contacts makes it impossible to predict the side chain-base contact with 100% accuracy.
accuracy, but it rather indicates that C2H2 zinc fingers are able to bind to almost any DNA duplexes. The versatility of the zinc finger binding relies on hydrogen bonds that can (i) make long distance contact to bases with and without participation of a water molecule (>2.75 angstroms), (ii) make contacts to more than one base, and (iii) make contacts to bases with flexible angles. Zinc fingers’ binding also relies on nonspecific hydrophobic interactions. Although it is difficult to predict exactly which zinc finger peptide sequence specifically binds to a DNA sequence, one can obtain desired zinc fingers by the phage display methods and manipulate expression of the target genes with the obtained finger proteins.

**Linkers**

About half of zinc finger proteins have a well-conserved linker, TGEKP, between adjacent fingers. The importance of the linker in zinc finger-DNA binding has been revealed by analyzing the effect of mutation on DNA affinity after making substitutions for the conserved residues, and by analyzing structures of finger-DNA complexes. The linker is flexible in solution without DNA. Upon forming the finger-DNA complex, however, the conserved lysine residue of the TGEKP sequence (Fig. 1) interacts with the phosphate backbone. Moreover, linkers contact the C terminus of the preceding α helix involving threonine and glycine residues so that the zinc finger-DNA complex becomes more stable (C capping). Alternative splicing of the WT1 message disrupts the conserved linker between finger 3 and 4 by inserting the sequence KTS. The change from TGEKP to sequence TGKTSEP is accompanied by a severe decrease in DNA binding. An NMR study of the finger peptide-DNA complex has shown that the insertion increases the flexibility between finger 3 and 4 and abrogates binding of finger 4 to its cognate site. These findings have demonstrated that the conserved linker is not only necessary to promote fingers to fit completely into the DNA major groove, but also necessary to strengthen the DNA binding. A separated-paired zinc finger class, ZAS family, also contains the TGEKP linker, but fingers belonging to a subtype of the same class, basonuclin-type fingers, have a distinct linker, LR(K)MHK. Tramtrack also has a distinct linker sequence, KRNVKYP.

**Dimerization of Zinc Finger Protein**

DNA binding proteins often bind to the target DNA duplex as dimers in order to increase their binding affinity and modulate their regulatory activities. This is true for C2H2 zinc finger proteins. Many multiple-adjacent C2H2 zinc finger proteins, such as Ikaros, Roaz, GL1, SW15, TRPS-1 and Zac, form homodimers on the target DNA duplex using their zinc fingers. GL1 forms a homodimer through the hydrophobic surface of zinc finger 1 that is not involved in the DNA binding. Similarly, SW15 dimerizes through finger 1 at the hydrophobic surface of both the β strand and the α helix’s C-terminal half. This hydrophobic dimerization was applied to make the homodimer of an engineered two-finger peptide on the DNA. However, the hydrophobic binding is not the only dimerization mechanism for zinc fingers. Ikaros contains six zinc fingers. The N-terminal four fingers participate in the specific binding to the DNA, and the C-terminal two fingers, separated from the four and making a pair of fingers, are responsible for the homodimerization.

**Regulation of Zinc Finger Binding to DNA**

In vitro zinc finger-DNA binding proceeds as long as both substances are present in a suitable buffer, but the in vivo binding does not happen automatically and proceeds only under certain conditions. The regulation is mediated by modification of either the zinc finger protein itself or the target DNA (Fig. 5). The former can be executed by acetylation and also by phosphorylation, whereas the latter is mediated by methylation. All these modifications increase or decrease the zinc finger protein-DNA complex formation. The direction in which the regulation proceeds, repression or activation, depends on the fingers and the DNA, as well as the regulatory signal involved.

**Acetylation of Zinc Fingers**

YY1 (Yin Yang 1), with four tandem C2H2 zinc fingers, binds to various genes and represses or activates their target gene expression. PCAF (p300/ p300/CBP associated factor) acetylates YY1 at the zinc finger domain and inhibits DNA binding. PCAF also acetylates members of the KLF (Kruppel-like factor) family belonging to the triple C2H2 zinc finger protein class. It acetylates KLF13 at two lysines of the TGEKK linker between finger 2 and 3 and results in stimulating the fingers to bind to DNA. Another coactivation factor CBP/p300 acetylates a total of six lysines including the linker’s two lysines, but it prevents the fingers from binding to DNA. Of the six residues, the lysine of finger 1 directly contacts bases. However, acetylation of this lysine is not enough to interrupt the DNA binding and acetylation of all six lysines is required. Moreover, in vitro experiments showed that KLF2 zinc finger cotransfected with PCAF increases the target γ globulin gene expression in vivo while the zinc finger cotransfected with p300 decreases gene expression. Therefore, it can be concluded that acetylation of KLF2 regulates the target gene expression both positively and negatively. Because lysine residues are abundant in a variety of zinc fingers, acetylation of C2H2 zinc fingers is likely to be a common mechanism to regulate gene expression by modulating the zinc fingers’ DNA binding activity.

Another domain of YY1, the central glycine-lysine-rich domain, is acetylated by p300 and PCAF. This acetylation has nothing to do with the DNA binding but is required for fully repressing the target gene transcription.
**Phosphorylation of Zinc Fingers**

During mitosis, Ikars is phosphorylated on threonine/serine residues of the three linkers and interrupted for its DNA binding. Phosphorylation of linkers also occurs to Sp1. This suggests that phosphorylation is a global inhibitory mechanism to keep C2H2 zinc fingers out of DNA during mitosis. Interestingly, phosphorylation modifies the linkers as the favored sites. A few interpretations can be given. First, modification of the conserved linker is the most effective way to modulate zinc finger binding activity because the linker is essential for the high affinity zinc finger-DNA binding. Second, the modification sites, theonines/serines, are always available in the conserved linkers. Finally, linkers may be accessible even in the finger-DNA complexes so that zinc finger peptides can be separated from the DNA as soon as the regulatory signal is delivered to the fingers. The same arguments can be applied to regulation by acetylation since it also modifies linkers.

A serine residue of the C terminus immediately after a zinc finger can be phosphorylated in the case of Cre1 (catabolite repression) of *Hypocrea jecorina*, perhaps by caseokinase-II-like enzyme. Surprisingly, this phosphorylation increases the fingers' DNA binding activity, suggesting that the unphosphorylated serine residue may mask the finger's DNA binding activity directly by interacting with the finger or indirectly by interacting with other intramolecular domains.

**Methylation of DNA**

Engineered zinc fingers have been shown to distinguish methylated DNA duplex from unmethylated. It is becoming increasingly clear that natural zinc fingers differentially recognize methylated and unmethylated CpG (cytosine-guanidene dinucleotide pairs) and play important regulatory roles in the expression of target genes in vivo. CTCF (CCCTC-binding factor) has eleven C2H2 zinc fingers and binds to an element present between a promoter and an enhancer in order to block the enhancer activity. The CTCF binding sites are widely distributed in vertebrates and several similar CpG-rich sequences are present in ICR (imprinted-control region) of the i^2///i5?as well. Electrophoretic mobility shift experiments have demonstrated that CTCF binds to the CpG-rich elements of ICR but only when these are unmethylated. Although more experiments have to be done to obtain the CTCF function in vivo, the zinc finger proteins are likely to participate in the genomic imprinting process by differentially binding to the methylated- and unmethylated CpG-rich elements.

Kaiso, with triple C2H2 zinc fingers, is another example of this type of DNA binding. However, in this case, the zinc fingers recognize the methylated CpG-rich sequence but not the unmethylated sequence. In vivo transfection experiments have shown that kaiso represses expression of genes with methylated binding sites in the promoter.

**Acknowledgements**

I thank Dr. R. Freeman for information on the Pfam database.

**References**


TFIIIA: A Sophisticated Zinc Finger Protein

Raymond S. Brown* and Jane Flint

Abstract

Transcription factor IIIA (TFIIIA) is widely regarded as the archetypal zinc finger protein. It is a member of a very large multigene family of eukaryotic DNA-binding proteins. More than two decades of research have been dedicated to understanding its interaction with the 5S ribosomal RNA gene (5S DNA). TFIIIA has nine tandem C2H2 zinc fingers along the peptide sequence. The three-dimensional structure of the N-terminal 6 zinc fingers bound to 31 base pairs of DNA shows that not all fingers are equal. Four of them make contacts located on both DNA strands while two fingers act as spacers. Individual fingers can recognize overlapping and interlocking base pair quartets. Side chains in the short alpha helices of fingers contact bases in the major groove. It is likely that the linker sequences connecting adjacent zinc fingers evolved to dictate which fingers bind to DNA. Signals essential for transcription initiation of 5S DNA, nuclear localization, and nucleocytoplasmic transport of 5S ribosomal RNA (5S rRNA) are located in the C-terminal part of TFIIIA.

Introduction

An RNA-binding protein accumulates in the oocytes of the ovaries of the African clawed frog (Xenopus laevis). This protein has a molecular mass of 40 kilodaltons (kD). About 10^10 molecules of the protein are estimated to be present per oocyte and using fluorescent antibodies it can be seen that this protein is localized mainly outside of the nucleus. Its partner in the cytoplasm was identified as 5S rRNA. About half of the 5S rRNA in the oocyte is associated with the 40 kD protein in a stable complex called the 7S storage particle (7S RNP). Similar 7S RNPs are also present in the oocytes of other amphibians, teleost fish and maize embryos.

In 1980 a transcription factor that is essential for RNA polymerase III (pol III) to transcribe Xenopus oocyte 5S rRNA genes was isolated. It was shown to be identical to the 40 kD RNA-binding protein. The protein isolated from the 7S RNP binds specifically to radioactively labeled 5S DNA. This transcription factor was named TFIIIA. In fact this was the first eukaryotic transcription factor to be purified and used to transcribe an isolated gene. Surprisingly the promoter region that TFIIIA recognizes was found to be located inside the gene itself. Even more impressive was the discovery that a single TFIIIA protein binds without interruption to a long sequence of about 55 base pairs. This region of the 5S DNA was named the internal control region (ICR).

Transcription of 5S DNA

Expression of a fluorescent green protein/TFIIIA fusion protein shows that TFIIIA is localized as expected in the nucleolus where rRNA genes are transcribed and ribosomes are assembled. Two other pol III transcription factors are also required for initiation of transcription of 5S DNA, namely TFIIIB and TFIIIC. In contrast to TFIIIA these factors contain a number of subunits. TFIIIB, which binds upstream of TFIIIA to 5S DNA, consists of three subunits, the TATA box-binding protein TBP, a single zinc finger protein Brf and a third subunit B. TFIIIC contains at least six subunits, some of which bind to sequences essential for transcription within the ICR (for a review see ref. 7).

5S rRNA is the product of transcription from 5S DNA. It is a component of the ribosome 60S subunit. The 120 nucleotide sequence of 5S rRNA is highly conserved among vertebrates. There are two major types of 5S rRNA genes in eukaryotes. Xenopus oocyte type (Xlo) and the somatic type (Xls) genes have sequences which differ at eight nucleotide positions between them. Six of these sequence positions lie within the ICR region. The differences in sequence do not appear to significantly affect TFIIIA binding to 5S DNA in vitro.

TFIIIA Structure and Function

Internal Domains

In addition to the oocyte form of TFIIIA (oTFIIIA) there is a larger somatic protein, sTFIIIA, which has 22 additional amino acids at the N-terminus originating from translation at an upstream AUG codon. The existence of sTFIIIA has been confirmed by precipitation with antibodies. Cellular levels of the two different forms of TFIIIA are probably related to switching from the production of oocyte 5S rRNA entirely to that of somatic 5S rRNA in the developing embryo.

The internal organization of oTFIIIA has been probed by limited digestion of 7S RNP with proteases such as elastase, trypsin or papain. These enzymes can readily cut in exposed regions of a protein that would be expected to lie between the relatively resistant folded domains. In each case, proteolytic fragments are released from the C-terminus of the 40 kD protein in a step-wise fashion to give a 35 kD product first and then a 30 kD core fragment. The same oTFIIIA digestion pattern is obtained from either the 7S RNP or a TFIIIA-5S DNA complex. The 30 kD core

* Corresponding author. See list of “Contributors”.
