An Investigation into the DNA Binding Properties of the *Escherichia coli* Repressor Protein, IclR

by

David Hosfield

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science

Department of Chemistry
University of Manitoba
Winnipeg, Manitoba
R3T 2N2

ISBN 0-612-16159-5

Name _____

Dissertation Abstracts International and *Masters Abstracts International* are arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation or thesis. Enter the corresponding four-digit code in the spaces provided.

Biochemistry
SUBJECT TERM

0487
SUBJECT CODE

UMI

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Nom _____

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

_____ ☐☐☐ U·M·I

SUJET — CODE DE SUJET

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS
Architecture 0729
Beaux-arts 0357
Bibliothéconomie 0399
Cinéma 0900
Communication verbale 0459
Communications 0708
Danse 0378
Histoire de l'art 0377
Journalisme 0391
Musique 0413
Sciences de l'information 0723
Théâtre 0465

ÉDUCATION
Généralités 515
Administration 0514
Art 0273
Collèges communautaires 0275
Commerce 0688
Économie domestique 0278
Éducation permanente 0516
Éducation préscolaire 0518
Éducation sanitaire 0680
Enseignement agricole 0517
Enseignement bilingue et multiculturel 0282
Enseignement industriel 0521
Enseignement primaire. 0524
Enseignement professionnel 0747
Enseignement religieux 0527
Enseignement secondaire 0533
Enseignement spécial 0529
Enseignement supérieur 0745
Évaluation 0288
Finances 0277
Formation des enseignants 0530
Histoire de l'éducation 0520
Langues et littérature 0279
Lecture 0535
Mathématiques 0280
Musique 0522
Orientation et consultation 0519
Philosophie de l'éducation 0998
Physique 0523
Programmes d'études et enseignement 0727
Psychologie 0525
Sciences 0714
Sciences sociales 0534
Sociologie de l'éducation 0340
Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE
Langues
- Généralités 0679
- Anciennes 0289
- Linguistique 0290
- Modernes 0291

Littérature
- Généralités 0401
- Anciennes 0294
- Comparée 0295
- Médiévale 0297
- Moderne 0298
- Africaine 0316
- Américaine 0591
- Anglaise 0593
- Asiatique 0305
- Canadienne (Anglaise) 0352
- Canadienne (Française) 0355
- Germanique 0311
- Latino-américaine 0312
- Moyen-orientale 0315
- Romane 0313
- Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THÉOLOGIE
Philosophie 0422
Religion
- Généralités 0318
- Clergé 0319
- Études bibliques 0321
- Histoire des religions 0320
- Philosophie de la religion 0322

Théologie 0469

SCIENCES SOCIALES
Anthropologie
- Archéologie 0324
- Culturelle 0326
- Physique 0327

Droit 0398
Économie
- Généralités 0501
- Commerce-Affaires 0505
- Économie agricole 0503
- Économie du travail 0510
- Finances 0508
- Histoire 0509
- Théorie 0511

Études américaines 0323
Études canadiennes 0385
Études féministes 0453
Folklore 0358
Géographie 0366
Gérontologie 0351
Gestion des affaires
- Généralités 0310
- Administration 0454
- Banques 0770
- Comptabilité 0272
- Marketing 0338

Histoire
- Histoire générale 0578
- Ancienne 0579
- Médiévale 0581
- Moderne 0582
- Histoire des noirs 0328
- Africaine 0331
- Canadienne 0334
- États-Unis 0337
- Européenne 0335
- Moyen-orientale 0333
- Latino-américaine 0336
- Asie, Australie et Océanie 0332

Histoire des sciences 0585
Loisirs 0814
Planification urbaine et régionale 0999
Science politique
- Généralités 0615
- Administration publique 0617
- Droit et relations internationales 0616

Sociologie
- Généralités 0626
- Aide et bien-être social 0630
- Criminologie et établissements pénitentiaires 0627
- Démographie 0938
- Études de l' individu et de la famille 0628
- Études des relations interethniques et des relations raciales 0631
- Structure et développement social 0700
- Théorie et méthodes. 0344
- Travail et relations industrielles 0629

Transports 0709
Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES
Agriculture
- Généralités 0473
- Agronomie. 0285
- Alimentation et technologie alimentaire 0359
- Culture 0479
- Élevage et alimentation 0475
- Exploitation des pâturages 0777
- Pathologie animale 0476
- Pathologie végétale 0480
- Physiologie végétale 0817
- Sylviculture et faune 0478
- Technologie du bois 0746

Biologie
- Généralités 0306
- Anatomie 0287
- Biologie (Statistiques) 0308
- Biologie moléculaire 0307
- Botanique 0309
- Cellule 0379
- Écologie 0329
- Entomologie 0353
- Génétique 0369
- Limnologie 0793
- Microbiologie 0410
- Neurologie 0317
- Océanographie 0416
- Physiologie 0433
- Radiation 0821
- Science vétérinaire 0778
- Zoologie 0472

Biophysique
- Généralités 0786
- Medicale 0760

SCIENCES DE LA TERRE
Biogéochimie 0425
Géochimie 0996
Géodésie 0370
Géographie physique 0368
Géologie 0372
Géophysique 0373
Hydrologie 0388
Minéralogie 0411
Océanographie physique 0415
Paléobotanique 0345
Paléoécologie 0426
Paléontologie 0418
Paléozoologie 0985
Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT
Économie domestique 0386
Sciences de l'environnement 0768
Sciences de la santé
- Généralités 0566
- Administration des hôpitaux 0769
- Alimentation et nutrition 0570
- Audiologie 0300
- Chimiothérapie 0992
- Dentisterie 0567
- Développement humain 0758
- Enseignement 0350
- Immunologie 0982
- Loisirs 0575
- Médecine du travail et thérapie 0354
- Médecine et chirurgie 0564
- Obstétrique et gynécologie 0380
- Ophtalmologie 0381
- Orthophonie 0460
- Pathologie 0571
- Pharmacie 0572
- Pharmacologie 0419
- Physiothérapie 0382
- Radiologie 0574
- Santé mentale 0347
- Santé publique 0573
- Soins infirmiers 0569
- Toxicologie 0383

SCIENCES PHYSIQUES
Sciences Pures
Chimie
- Généralités 0485
- Biochimie 487
- Chimie agricole 0749
- Chimie analytique 0486
- Chimie minérale 0488
- Chimie nucléaire 0738
- Chimie organique 0490
- Chimie pharmaceutique 0491
- Physique 0494
- PolymÈres 0495
- Radiation 0754

Mathématiques 0405
Physique
- Généralités 0605
- Acoustique 0986
- Astronomie et astrophysique 0606
- Électronique et électricité 0607
- Fluides et plasma 0759
- Météorologie 0608
- Optique 0752
- Particules (Physique nucléaire) 0798
- Physique atomique 0748
- Physique de l'état solide 0611
- Physique moléculaire 0609
- Physique nucléaire 0610
- Radiation 0756

Statistiques 0463

Sciences Appliqués Et Technologie
Informatique 0984
Ingénierie
- Généralités 0537
- Agricole 0539
- Automobile 0540
- Biomédicale 0541
- Chaleur et thermodynamique 0348
- Conditionnement (Emballage) 0549
- Génie aérospatial 0538
- Génie chimique 0542
- Génie civil 0543
- Génie électronique et électrique 0544
- Génie industriel 0546
- Génie mécanique 0548
- Génie nucléaire 0552
- Ingénierie des systèmes 0790
- Mécanique navale 0547
- Métallurgie 0743
- Science des matériaux 0794
- Technique du pétrole 0765
- Technique minière 0551
- Techniques sanitaires et municipales 0554
- Technologie hydraulique 0545

Mécanique appliquée 0346
Géotechnologie 0428
Matières plastiques (Technologie) 0795
Recherche opérationnelle 0796
Textiles et tissus (Technologie) 0794

PSYCHOLOGIE
Généralités 0621
Personnalité 0625
Psychobiologie 0349
Psychologie clinique 0622
Psychologie du comportement 0384
Psychologie du développement 0620
Psychologie expérimentale 0623
Psychologie industrielle 0624
Psychologie physiologique 0989
Psychologie sociale 0451
Psychométrie 0632

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

AN INVESTIGATION INTO THE DNA BINDING PROPERTIES OF THE ESCHERICHIA COLI REPRESSOR PROTEIN, IclR

BY

DAVID HOSFIELD

A Thesis/Practicum submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Abstract

When the gram negative bacterium *Escherichia coli* is growing on acetate or fatty acids the enzymes of the glyoxylate cycle are required to allow the organism to incorporate the acetyl carbons of acetyl-CoA into C4 dicarboxylic acids and finally sugars. The enzymes, isocitrate lyase and malate synthase, are coded for by two structural genes *aceA* and *aceB,* which along with *aceK,* make up the *ace operon.* In glucose containing medium the *ace* enzymes are not present at any appreciable level, however, when the medium is changed to acetate or fatty acids the enzymes become induced. The molecular mechanisms controlling the regulation of the glyoxylate cycle are not well understood at present, however, it has been shown that the gene product of the *iclR* gene, IclR, binds to the promoter region of the *aceB* gene

The primary aim of this dissertation is to determine in greater detail the mechanisms by which IclR interacts with operator DNA. Using site directed mutagenesis several amino acids thought to be important for mediating DNA binding were mutated to alanine and there effects were measured by an electrophoretic mobility shift assay. In addition, the ability of IclR to bind to several different operators was tested using this method. To discover the identity of the DNA bases important for recognition a method termed SELEX, or Systematic Evolution of Ligands by Exponential enrichment, was employed. This method relies on the ability of a DNA binding protein to "fish out" tight binding sequences from a random pool of

oligonucleotides, and on PCR to amplify and enrich the selected molecules. The results from this experiment defined the IclR consensus sequence as being palindromic with the half site sequence AAWDTGGAAAV, where W= A or T, D = not C, and V = A or C. Using this consensus palindrome an additional site in the *E.coli* genome was discovered in the promoter region of the *gcl* gene, a gene encoding the protein glyoxylate carboligase. This region of DNA was amplified by PCR and shown to interact with IclR *in vitro*. The result suggests that IclR, aside from regulating the metabolism of acetate and fatty acids, may also regulate the metabolism of glycollate.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my gratitude to my supervisor, Dr. Harry W. Duckworth for the moral and intellectual support given to me in the years I have been lucky enough to work with him.

I would also like to thank the members of my advisory committee, Dr. Jim Jamieson, and Dr. Peter Loewen, for taking the time to read and critique this thesis. I am grateful to Dr. Loewen for the Taq polymerase used in many of the experiments described within

I express my sincere gratitude to the members of the laboratory, Dr. Lynda Donald and Ayeda Ayed, for the assistance and advise offered to me. Thanks also to the Department of Microbiology for the use of equipment.

I am also indebted to my parents, Bob and Dorrie, for supporting me through thick and thin. I extend my sincere thanks to Tanya for her unyielding support and commitment.

Finally I acknowledge the Natural Sciences and Engineering Research Council of Canada for financial support throughout my graduate career

Chapter 1

Introduction

1.0 The Role of DNA in Biology

The discovery of the structure of DNA some forty years ago (Watson and Crick, 1953) had a revolutionary impact on the science of molecular biology, as it suggested solutions to two fundamental problems. The double helical structure of the molecule immediately suggested that replication (the process in which the molecule could copy itself) could proceed through a process that involved complementary hydrogen bonds between base pairs such that the genetic information contained within could be passed from generation to generation in a semi-conservative manner. In addition, the structure suggested that the genetic information within a cell was encrypted by a four letter alphabet consisting of the DNA bases. This suggestion led to the discovery of mRNA and RNA polymerase, and the elucidation of the genetic code.

The mechanism by which a cell converts genetic information into useful biological products has become known as the central dogma of molecular biology (fig 1). In the most basic sense the dogma states that DNA directs its own replication, while also directing the synthesis of RNA (transcription), which, in turn, directs the synthesis of proteins (translation). In the years since the structure of DNA was discovered a great wealth of knowledge has emerged from studies attempting to discover the cellular machinery responsible for directing the above processes. One area that will be discussed in some detail in this dissertation, is the area of genetic regulation.

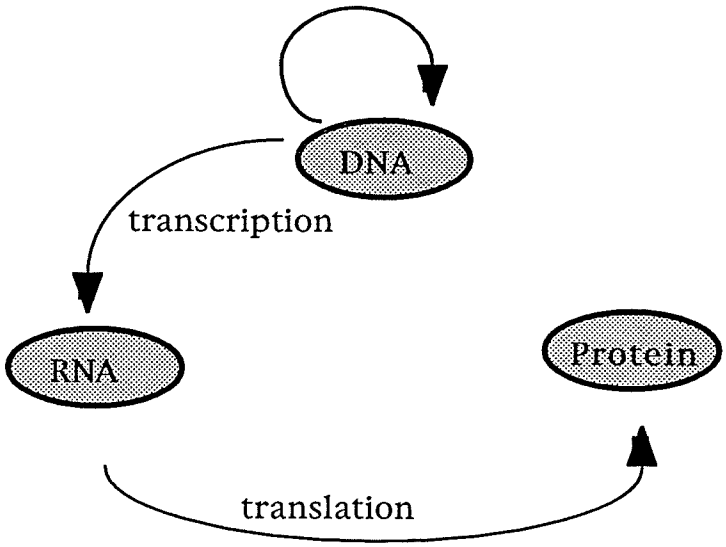


Fig.1 The central dogma of molecular biology: DNA directs its own replication, while also controlling the synthesis of proteins through the intermediate mRNA.

1.1 Genetic Regulation in Procaryotes

Before the discovery of the double helix it was well known that many organisms were able to effect the synthesis of certain enzymes in response to environmental conditions. For example, Stanier and colleagues (Stanier, 1951) showed that in a large number of procaryotic organisms the presence of a substrate would generally induce the formation of a series of enzymes that were sequentially involved in its metabolism. When the substrate was removed from the growth medium, the synthesis of the enzymes would desist. The logic of this system was indeed very rational: *To conserve energy, a cell would direct the synthesis of a protein only when that protein was needed.* In 1961 Jacob and Monod, working with the *lac* system of the gram negative bacterium *E. coli*, suggested a genetic model that could explain the above findings (Jacob and Monod,

1961). They termed this model the operon model of genetic regulation.

An operon can be described as a small unit of genomic DNA whose function is to ensure that proteins arising from the operon are transcribed only when the external environment dictates. As shown in Fig. 2 an operon generally contains three types of genes. The first type of gene, or the structural gene contains all the information necessary to dictate the primary sequence, and thus the

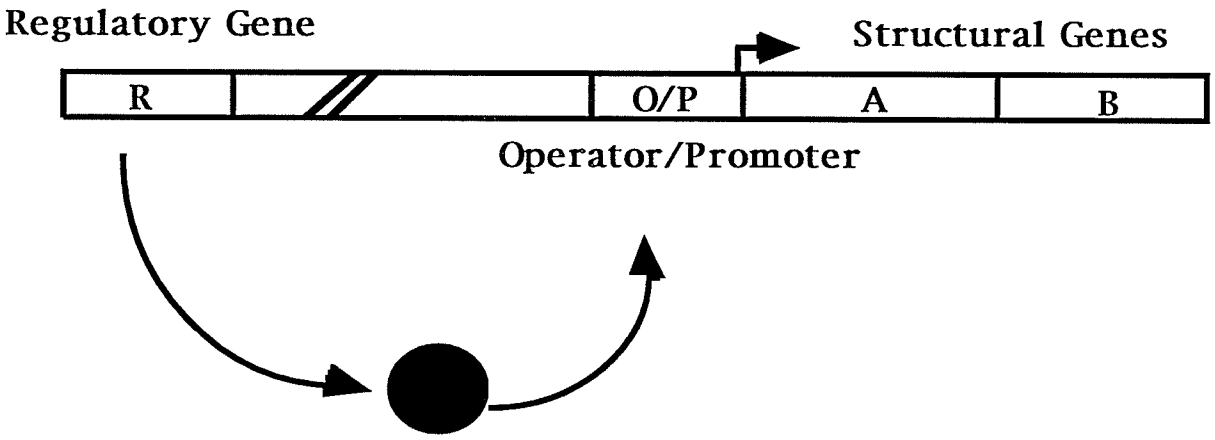


Fig. 2 Schematic illustration of an operon. Genes A and B, structural genes, are transcribed on one polycistronic mRNA. The O/P region houses both the operator, which binds the regulator, and the promoter, which recognizes RNA polymerase at the -10 and -35 regions. The regulatory gene, R, is often transcribed external to the operon. This gene encodes a protein that interacts with the operator. The arrow indicates the transcription start site.

structure and function, of the protein it encodes. The second type of gene, the promoter, contains initiation sites where RNA polymerase is able to bind and start transcription. These sites are found at positions -10 (Pribnow box) and -35(-35 region) relative to the

transcription start site. The third gene, or the operator, contains a DNA sequence that is recognized by a regulator protein. This gene is usually, but not always, located close to the promoter. A regulator protein is a molecule that is usually outside the operon but plays a role in controlling the rate of synthesis of the structural genes.

Regulator proteins can be grouped into two categories: those that mediate a positive effect on transcription, and those that mediate a negative effect. In a positively controlled system, the regulator increases the rate of transcription from a specified promoter, presumably by increasing the efficiency in which RNA polymerase initiates transcription. In general a small molecule can act as a co-activator by binding to the regulator, and increasing its affinity for DNA, thereby increasing the extent to which it controls transcription. Conversely, the promoter could be made inactive if a small ligand bound to the activator protein releasing it from DNA. The CAP-cAMP complex in its regulation of a number of operons (reviewed by deCrombrugghe, 1984), and the FadR protein in its regulation of *fabA* transcription (Henry and Cronan, Jr., 1992) are examples of each of these two models, respectively. In a negatively controlled system the levels of transcription are reduced by the presence of the regulator molecule which prevents RNA polymerase from initiating transcription. In a negatively regulated inducing system a small ligand binding to the regulator displaces it from DNA thus increasing the rate of transcription. In a negatively regulated repressing system the small ligand can act as a co-repressor by increasing the affinity of the protein for DNA. The LacI-allolactose

interaction (Gilbert and Müller-Hill, 1966; Riggs et al. 1968) and the TrpR-tryptophan interaction (Somerville, 1983) provide examples of each of these systems, respectively. In *E. coli* the *lac* operon provides the paradigm for procaryotic regulation. This system which is controlled both positively and negatively will be described in the following section to illustrate the complexities involved in genetic regulation in procaryotes.

1.1.2 The *lac* Operon: A History

When *E. coli* is growing on lactose as the sole source of carbon the enzymes of the *lac* operon are required to effect the efficient uptake and breakdown of the disaccharide. When *E. coli* is grown in glucose the cell only contains a few molecules of these proteins. However, when the glucose is replaced by lactose the cells increase the rate at which these enzymes are produced nearly 1000 fold, and they maintain this rate until the lactose is depleted (Monod et al. 1952), suggesting that lactose or one of its metabolic products is the inducer of the *lac* metabolizing proteins. Further work established that the actual physiological inducer was allolactose, a side product arising from the ability of β-galactosidase to occasionally trans-glycosylate lactose. The operon model suggested by Jacob and Monod was able to account for the above observations.

The structural genes of the *lac* operon, *lacZ*, *lacY*, and *lacA*, which code for β-galactosidase, lactose permease, and trans-acetylase, (an enzyme whose function is still largely unknown)

respectively, are located in that order, immediately distal to the *lac* operator/promoter (Jacob and Monod, 1961; Gilbert and Maxam, 1973) (fig. 3). As described above the promoter is recognized by RNA polymerase between positions -20 to +20 relative to the start of transcription. The operator, recognized by the product of the *lacI* gene, is located between positions -7 to +28 relative to the transcription start site. Thus the picture that emerged from the pioneering work of Jacob and Monod could be described as follows: When growing on a substrate other than lactose the *lac* operon would be "turned off" due to the fact that the repressor (the *lacI* gene product), which bound tightly to the DNA, blocked the ability of RNA polymerase to initiate transcription. When lactose was added to the growth medium allolactose formation was initiated by the minute quantities of permease and β-galactosidase present in the cell. The allolactose so formed would then allosterically bind to the Lac repressor, and in doing so would dramatically weaken the affinity of repressor for operator by inducing in the protein a conformational change that weakened DNA binding. This would then allow RNA polymerase to initiate transcription of the lac operon so that the cell could readily metabolize the substrate. This seemingly simple picture, like almost all things in nature, of course became much more complicated in the years that followed.

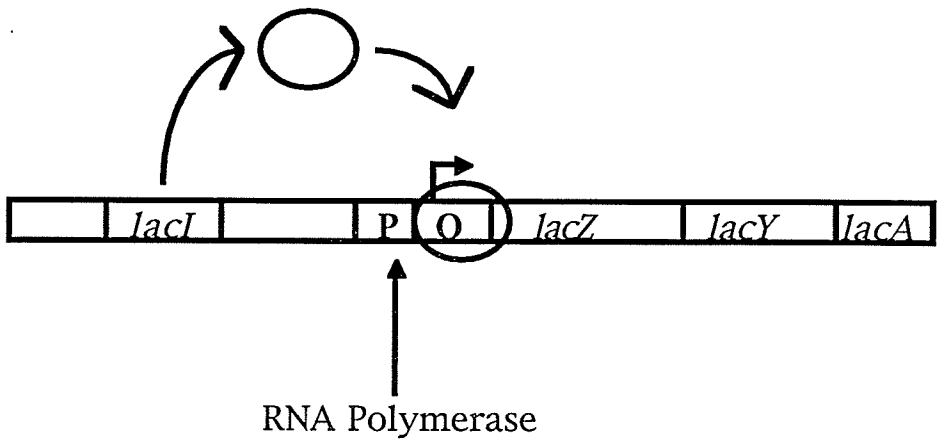


Fig. 3 Basic illustration of the *lac* operon. The *lacI* gene product, the Lac repressor (illustrated with an open circle) represses transcription by binding to the *lac* operator and preventing RNA polymerase from forming an open complex with its promoter. When lactose is present, a small amount of it is converted to allolactose which binds to the repressor and displaces it from its operator. This displacement allows RNA polymerase to form an open complex and initiate transcription.

1.1.3 The *lac* Operon: 1996

Today the structure and control mechanisms controlling the metabolism of lactose in *E. coli* are known in great detail. Since the time of Jacob and Monod, nuclease studies have demonstrated that three operator regions, termed O1, O2, and O3 (Oehler et al. 1990; Eismann and Muller-Hill, 1990) are located in the *lac* operon. While O1 conforms to the definition the classical operator described above, the other operators are located at regions not expected to interfere with polymerase binding (fig. 4) However, *in vivo* studies have demonstrated that indeed, all three operators are needed for maximal repression (Oehler et al. 1990). The cause of these effects is understood at the molecular level as originating from the ability of the Lac repressor to bind simultaneously to more than one operator. The Lac repressor tetramer can simultaeously bind to either O1 and

O2 or O1 and O3 in a cooperative manner to form a looped protein-DNA complex. Thus, binding at O2 for example, increases the affinity of the tetramer at O1 by 3 fold (Sasse-Dwight and Gralla, 1988). This increase in affinity leads to a more stable complex that is better able to repress the initiation process catalyzed by RNA polymerase. The recently determined crystal structure of the Lac repressor-DNA complex has confirmed that the Lac repressor tetramer can in fact bind to two operators simultaneously (Lewis et al. 1996)

Control of the *lac* operon has also been shown to be dependent on glucose concentration through the action of the cAMP receptor protein (CAP). In a non-glucose containing medium such as lactose the levels of cellular cAMP are high so that a complex between this ligand and CAP becomes favored. The ligand-receptor interaction greatly enhances the affinity of CAP for two specific regulatory sequences near the two *lac* promoters (fig. 4). The first CAP-DNA complex stimulates transcription at P1 by binding just upstream of this promoter and thus increasing the affinity of RNA polymerase for the lac promoter (Zubay et al. 1970). The mechanism of this control is unknown, but it probably involves cooperative protein-protein contacts that may be facilitated by DNA bending (Schultz et al. 1991). In addition, productive binding of CAP at this site prevents transcription from P2. The second CAP site, which overlaps the O1 site is thought to act as an anti-repressor by sterically preventing the Lac repressor from binding to O1 (Peterson and Reznikoff, 1985).

The *lac* operon thus provides a view into the complexities involved in genetic regulation: A large number of protein binding sites located within a relatively small distance from each other are able to communicate according to a number of environmental signals. These signals are propagated synergistically through a combination of protein-ligand, protein-DNA, and protein-protein contacts. Under different conditions the number and types of contacts will vary, thus allowing an organism to fine tune the amounts of a specific protein required for any particular environmental situation.

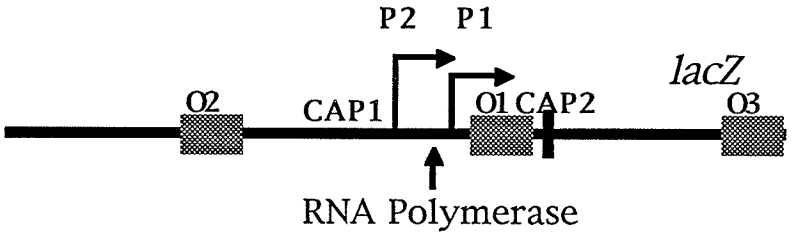


Fig. 4 Schematic illustration of the *lac* operator/promoter regions in *E. coli*. Shown are the three operators O1, O2, and O3, which communicate with each other through DNA looping which is mediated by bound Lac repressor. Shown also are the two promoters, P1 and P2, and the two CAP binding sites. Not shown exactly to scale

1.2 Metabolism in *E. coli*

In *E. coli* the central pathways of carbon metabolism are glycolysis and the tricarboxylic acid cycle. As shown in figure 5 the process can be thought of as starting with glucose (carbohydrate) and ending with the net oxidation of this molecule to six molecules of carbon dioxide and six molecules of water. Much of the net release of energy for this process is stored as high energy phosphate bonds

in the molecule ATP. In addition to providing energy to the cell these two central metabolic pathways also provide starting materials for the synthesis of cellular components such as amino acids, fatty acids, cell wall components, and nitrogen bases.

For organisms growing on compounds other than glucose these central metabolic pathways are almost always the final destination in their metabolism. For example, the amino acids valine and isoleucine, are converted through a series of steps to succinyl-CoA, an intermediate of the TCA cycle. If energy is required by the cell the succinyl-CoA can be oxidized through the TCA cycle. However, if cellular material is required some of the succinyl-CoA can be funneled out of the TCA cycle to provide for this. For example, the succinyl-CoA can be converted to oxaloacetate (OAA) by standard TCA reactions, whereupon the enzyme PEP carboxykinase can convert the OAA into the glycolytic intermediate phosphoenolpyruvate (PEP). The PEP so produced can be shuttled backwards through gluconeogenesis to produce glucose which can be used for creating cellular products such as cell wall polysaccharides.

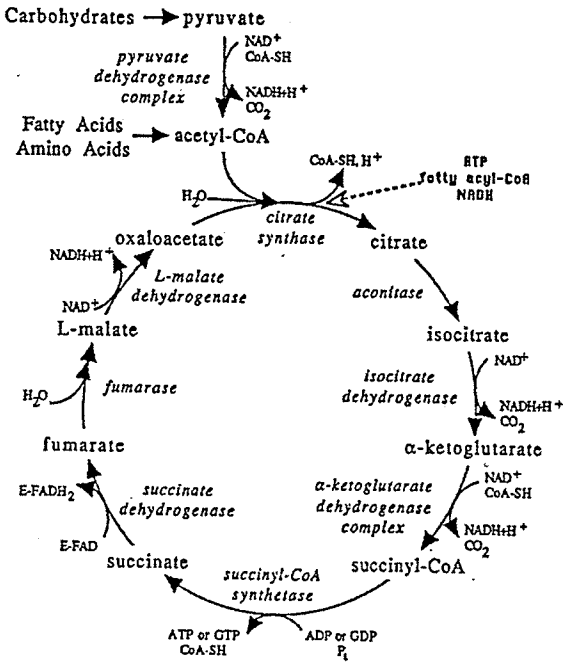


Fig. 5 Illustration of the cyclic TCA cycle in *E. coli*. Glucose is first converted by glycolysis into two molecules of pyruvate. The so-formed pyruvate is then either oxidized by the TCA cycle to provide ATP, or converted through other reactions into cellular material (Taken from Molgat, 1990).

1.2.1 Growth on Acetate

When *E. coli* is growing on acetate the situation is not so simple. Acetate is introduced into the central metabolic pathways as acetyl-CoA. This molecule is then condensed with OAA to form the six carbon molecule, citrate. As shown in figure 5, however, the acetyl-CoA is completely oxidized by the TCA enzymes by the time it reaches succinyl-CoA. Thus the TCA cycle is unable to provide the OAA that is required for gluconeogenesis. To adapt to this problem a number of organisms adapt by utilizing the enzymes of what has become known as the glyoxylate bypass (Kornberg and Sadler, 1961).

As shown in figure 6 the glyoxylate bypass consists of two enzymes: isocitrate lyase, whose function is the conversion of isocitrate to glyoxylate and succinate, and malate synthase, whose function is the condensation of acetyl-CoA and glyoxylate to form malate. The malate so formed by this cycle can then be converted to cellular material by the reactions of gluconeogenesis.

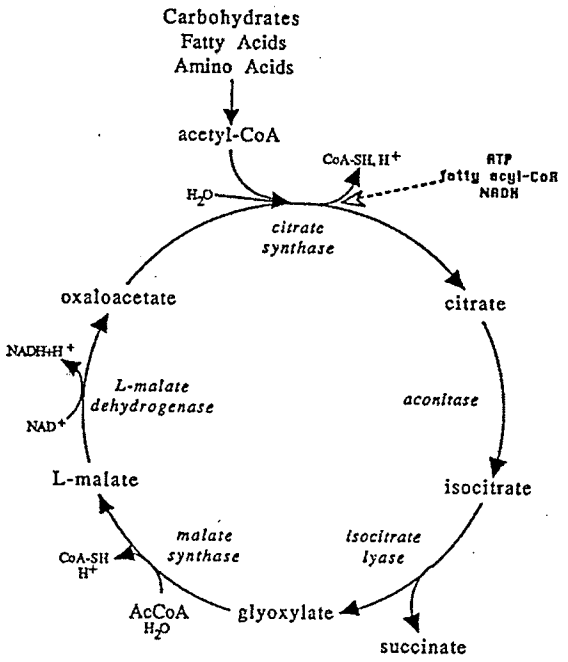


Fig. 6 The glyoxylate cycle in *E. coli* is a variation of the TCA cycle. Isocitrate lyase hydrolyzes isocitrate to succinate and glyoxylate, while malate synthase condenses the glyoxylate with acetyl-CoA to form malate. These reactions skip the two oxidative steps of the TCA (fig. 5) allowing for the carbon in acetate to be assimilated into cellular material rather than CO_2. (Taken from Molgat, 1990).

1.3 Regulation of the Glyoxylate Cycle

Cells growing on glucose or other sugars show only very low levels of the enzymes of the glyoxylate cycle, but, when the medium is changed to acetate the levels of the enzymes rise to much higher levels (Kornberg, 1963). Using the operon model of genetic regulation, this fact suggested that acetate, or one of its metabolites, was acting as an activator of the synthesis of the glyoxylate cycle enzymes. Several early observations by Kornberg (Kornberg, 1963), however, indicated that this was not the case. For example, it was shown that *E. coli* growing on glucose was not able to induce isocitrate lyase formation when acetate was added to the growth medium: for the enzymes to be induced the glucose first had to be removed, however, this removal was not required due to the common mechanism of metabolite repression as seen with CAP-cAMP and the *lac* operon (Duckworth, 1980). Secondly, Kornberg was able to demonstrate that when the growth medium was proline (a substrate that is somewhat effective in inducing isocitrate lyase synthesis) addition of acetate raised the levels of isocitrate lyase roughly 4 fold. This effect, however was not observed in the same medium if the organism was *gltA* (i.e. lacking citrate synthase). This result demonstrated that neither acetate, acetyl phosphate, or acetyl-CoA could be the inducer of the glyoxylate cycle, but that a product dependent on the action of citrate synthase was. Continued work with several mutants deficient in one or more steps of the central metabolic pathways allowed Kornberg to hypothesize that the

intracellular intermediate was phosphoenolpyruvate (Kornberg, 1966). The suggestion was made that this compound derepressed synthesis of the glyoxylate enzymes when growing on acetate due to a drop in its intracellular concentration. This suggested that PEP was acting as the co-repressor of a specific protein (like tryptophan in the TrpR-tryptophan complex) in glucose grown cells. Lowry and colleagues, however, refuted this claim by direct measurements of intracellular intermediate, which showed that the levels of PEP decline when glucose is added to acetate adapted *E. coli.* (Lowry et al. 1971).

1.3.1 The *ace* Operon

In 1968 Kornberg was able to provide evidence that the enzymes isocitrate lyase and malate synthase were coded for by an operon located at 91 minutes on the *E. coli* functional map (Brice and Kornberg, 1968). The genes encoding the enzymes, *aceB* (malate synthase) and *aceA* (isocitrate lyase) were later shown to exist in that order (Maloy and Nunn, 1982), along with a third gene, *aceK*, which encodes the multifunctional enzyme isocitrate-dehydrogenase kinase/phosphatase (IDH kinase/phosphatase) (LaPorte and Chung, 1985). During growth on acetate the kinase/phosphatase is able to disactivate the TCA enzyme isocitrate dehydrogenase by phosphorylating an active site serine residue. When the growth medium is changed to glucose the IDH kinase/phosphatase dephosphorylates IDH, returning it to its functional form. The action of this protein allows the isocitrate to be shuttled through the

glyoxylate cycle rather than the TCA cycle during growth on acetate. Aside from the above described system, the entire *aceBAK* operon has been successfully cloned and sequenced allowing a careful examination of the factors that control its regulation (Byrne et al. 1988; Chung et al. 1988). The structure of the operon, along with genes adjacent to it is depicted in figure 7.

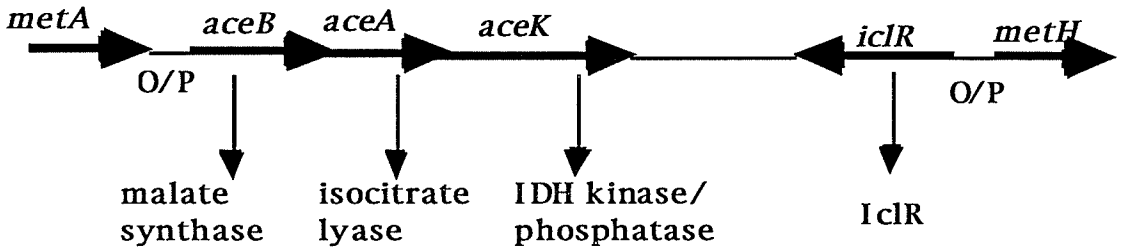


Fig. 7 Schematic of the *aceBAK* operon in *E. coli*. The *aceB*, *aceA* and *aceK* genes code for the enzymes malate synthase, isocitrate lyase, and IDH kinase/phosphatase, respectively. The *iclR* locus is located immediately downstream from the *aceK* gene.

Transcriptional regulation of the *aceBAK* operon has been shown to depend on the products of several genes. The first genes to be identified as important in this process were *iclR* (Brice and Kornberg, 1968; Maloy and Nunn, 1982) and *fadR* (Maloy et al. 1980). The *iclR* gene is located immediately adjacent to the *ace* operon at 91 minutes on the *E. coli* linkage map. The *fadR* gene on the other hand is located at 25 minutes. Using merodiploid studies Maloy and Nunn (1982) were able to demonstrate that the products of both these genes showed their effects in trans, that is, each coded for a diffusible repressor that could negatively effect the control of the *ace* operon. The effect of each of these gene products is described below

1.3.2 IclR

As discussed above, the protein product of the *iclR* gene, IclR, controls the synthesis of the glyoxylate cycle enzymes in a trans-dominant manner. As with the Lac repressor this suggests that IclR is able to interact with regulatory elements just upstream of the *ace* operon. To date the *iclR* gene has been cloned and sequenced by three groups; First by Cortay et al. (1991) and Nègre et al. (1991), and finally by Donald et al. (1996). The purified IclR protein resulting from these studies has been shown to interact specifically with DNA sequences in the *ace* promoter region thus verifying the earlier genetic results.

1.3.3 The IclR Binding Site

There exist several common methods for defining the site at which a protein binds to DNA. In the studies mentioned above several of these methods were used to examine IclR binding. Cortay et al. (1991), using a electrophoretic mobility shift assay, were first able to show that the purified repressor could specifically interact with a radiolabelled restriction fragment containing the *ace* operon. Moreover, the EMSA assay seemed to suggest that formation of the IclR-DNA complex was impeded by inclusion of PEP in the binding

reaction. This suggestion suggested that contrary to Kornberg's earlier suggestion, PEP could act as the inducer, rather than a co-repressor, of the *ace* operon. This finding, however, has never been mentioned again by the authors and evidence presented in this thesis suggest that this finding was in error. Cortay et al. (1992) later showed that the binding of IclR to the *ace* operator occurred with moderate to high affinity. The dissociation constant for the complex was reported to be 0.7nM.

The DNA binding site for this protein was also investigated by DNase footprinting (Donald et al. 1996; Cortay et al. 1991), dimethyl sulfate interference experiments (Nègre et al., 1992), and de-purination-depyrimidation interference experiments (Nègre et al., 1992). The DNase protection results of Cortay et al. (1991) suggested that IclR bound to a pseudo-palindromic region of DNA overlapping the -35 region of the *ace* promoter. The region protected was very large for a typical procaryotic repressor, spanning a region of approximately 35 base pairs. The protection data reported, however, only contained information for the top (sense) DNA strand. The results obtained by Donald et al. (1996), for both the top and bottom strands of the operator suggested that IclR protected the same region from digestion by DNase, however the number of bases protected were only 29 and 27 for the top and bottom strands, respectively.

The interference experiments of Nègre et al. (1992) verified the importance of the bases found by the DNase work. In an interference experiment bases in the DNA helix are first modified

with chemicals that generally alkylate or remove DNA bases. The modified DNA is then incubated with protein and the complex formed is separated from the unbound DNA. Modifications that affect binding are usually overrepresented in the "free DNA" fraction, and thus examination of this fraction following cleavage of the modified bases reveals DNA sites important for complex formation. In this way Cortay et al. (1992) were able to verify some of the bases identified to be important for IclR binding. In particular, they were able to show the importance of the two guanine residues located at position -43, and -44. A summary of the DNase protection, and the interference data is shown in figure 8.

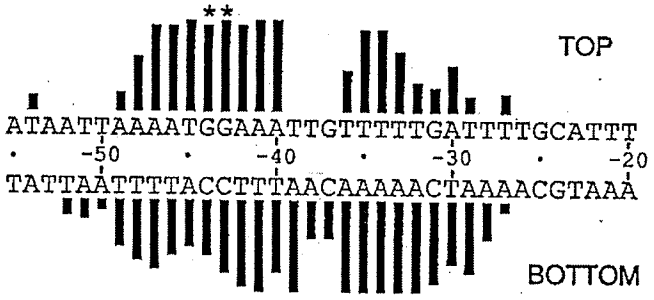


Fig. 8 Summary of the protection and interference data obtained to date for the binding of IclR to the *aceBAK* operator. The bases shown are those which were protected from DNase by IclR according to Cortay et al. (1991) The solid bars summarize the interference data obtained by Nègre et al. (1992). The magnitude of the effect on binding of removing a base is indicated by the size of the bar above the base. Asterisks above the two guanines refer to the bases shown to be important by a guanine interference assay (Nègre et al. 1992).

Although the work presented in the above studies has indicated the site to which IclR binds, it remains unclear as to the

identity of the DNA bases that are critical for a strong protein-DNA interaction. The guanine methylation interference results suggest that the bases at positions -43 and -44 are crucial determinants of IclR binding, however, the depurination-depyrimidation results must be viewed with caution. In this experiment, removing the entire base would be expected to effect dramatic changes in the structure of the molecule. Since protein-DNA interactions depend critically upon the formation of complementary interacting surfaces, the results obtained from this method can probably only be considered to reflect the general importance of a specific stretch of DNA rather than the importance of an individual base. In this thesis a different PCR-combinatorial based method was employed to examine the bases important for IclR binding to the *ace* promoter. The results obtained from these experiments allowed me to define a consensus sequence to which IclR binds. Using this consensus sequence I discovered that IclR specifically interacts with a region of DNA overlapping the promoter to the *gcl* gene in *E. coli*. This gene encodes the protein glyoxylate carboligase, an enzyme required for the efficient metabolism of the two carbon molecule glycollate (Kornberg, and Gotto, 1961; Chang, et al., 1993).

1.3.4 IclR is Autoregulated

Recently, it has also been shown that IclR regulates its own expression by binding to *cis*-acting elements on the *iclR* promoter (Gui et al. 1996). These elements, like those found in the *ace* operator, overlap the -35 region of the RNA polymerase start site.

The sequence of bases in this region slightly resembles that found in the *ace* promoter region. Gui et al. (1996) were able to show using an EMSA that IclR bound to this region with high affinity, though the dissociation constant was not reported. In this thesis, a quantitative EMSA experiment was performed with the *iclR* operator in order to determine the precise affinity of IclR for this binding site.

Unlike the *ace* operon, expression of the *iclR* gene was shown to be relatively independent of carbon source (Gui et al. 1996) In acetate containing medium the levels of IDH kinase/phosphatase (*aceK*) were approximately 40 fold higher than they were in glucoses. The level of *iclR* expression, however, was the same in both carbon sources. The fact that these differences are seen implies that additional protein factors may be responsible for regulating the *ace* operon. If IclR was the sole regulatory protein for both genes then *iclR* transcription would have been expected to be induced roughly at the same levels as the *ace* operon. As described below, additional proteins do in fact play an important role.

1.3.5 FadR

The primary role of the *fadR* gene is controlling the rate of transcription of the genes required for the synthesis and β-oxidation of fatty acids. However, several lines of evidence suggest that the protein also regulates the *ace* operon. Maloy and Nunn (1982), using operon fusions between the *ace* and *lac* operons showed that the synthesis of β-galactosidase increased in strains that were *iclR* and

fadR. In addition, the authors demonstrated that the two genes acted synergistically to repress the *ace* operon: the synthesis of β-galactosidase was higher in a strain that was *iclR fadR* as opposed to those that were either *iclR fadR*$^+$ or *iclR*$^+$ *fadR.* Recently the *fadR* gene product, FadR, has been cloned and purified, and shown to interact specifically with DNA sequences upstream of the enzymes required for fatty acid synthesis and degradation (DiRusso et al. 1993). These studies have identified a FadR binding consensus sequence of 5'-AGCTGGTCCGAYNTGTT-3', and have revealed that the formation of the FadR-DNA complex is specifically impeded by long-chain acyl-CoA esters. To date, however, no detailed molecular mechanism as to how the FadR protein exerts its control over the *ace* operon has been put forward. The lack of any DNA sequence resembling the FadR consensus in the immediate region of the *ace* promoter (either up or downstream) suggests, however, that the mechanism must be quite complex.

1.3.6 Other factors Controlling *ace* Expression

Recently, reports from several groups has indicated that the *ace* operon is also under the control of the FruR repressor and the Integration Host Factor. The FruR repressor has been shown to act as a pleiotropic regulator, regulating all the major pathways concerning carbon metabolism. Aside from the *ace* operon, FruR has been shown to exert its effect on the pathways of glycolysis, the TCA cycle, gluconeogenesis, electron transport, and the pentose phosphate pathway (Chin et al., 1989; Saier and Chin, 1990; Ramseier et al.,

1993). In mutants defective in FruR expression the levels of the enzymes involved in these pathways are dramatically affected. In general, *fruR* strains show elevated levels of glycolytic enzymes and depressed levels of the enzymes involved in gluconeogenesis, suggesting that FruR is an activator of gluceogenenesis and a repressor of the glycolytic pathway. With respect to the glyoxylate cycle enzymes a *fruR* strain growing on acetate had roughly one tenth the levels of malate synthase and one half the levels of isocitrate lyase, as compared to the isogeneic wild type strain. Interestingly these levels were roughly the same as those obtained by adding 0.5% glucose to the wild type strain (Chin, et al., 1989). These results thus suggested that FruR is a positive regulator of *ace* transcription.

In vitro, the cloned and purified FruR protein has been shown to interact with the operators of several genes in *E. coli*, verifying the observation that the protein regulates gene expression (Ramseier, et al., 1993). To the *ace* operon a variety of interference and protection experiments have indicated that the protein binds with moderate affinity (3nM) to a pseudo-palindromic region located 170 bp upstream of the transcriptional start site (Ramseier et al., 1993; Cortay et al., 1994). The binding was prevented by 50μM fructose-1-phosphate, and to lesser extent by fructose-1,6-bisphosphate, suggesting that one or both of these compounds is the *in vivo* effector molecule. The mode of action of this protein, however, with respect to its ability to activate transcription at the *ace* promoter, has yet to be investigated.

The Integration Host Factor (IHF), has also been shown to effect the levels of the glyoxylate cycle enzymes (Freundlich et al. 1992). This protein, which was first discovered as a host factor required for site-specific recombination of lambda bacteriophage, has been shown to control the expression of a number of *E. coli* genes. In many cases it has been found that IHF regulates transcription by facilitating the bending of DNA (Freundlich et al. 1992). DNase protection experiments have shown that this protein interacts with two regions upstream of the *aceB* transcription start site. The first site is located at -81 bp relative to the transcriptional start site of the *aceB* gene, while the second site is located at position -164. The exact mechanism by which this protein effects regulation of the *ace* operon is presently unknown, however, mutation of the two genes coding for the IHF protein results in reduced levels of *ace* expression. Similarily, disruption of the IHF binding sites 5' to the *aceB* gene show a similar effect *in vivo* (Freundlich et al. 1992; Gui et al. 1996; Resnik et al. 1996)

The structure of the *ace* operator/promoter region, with all its known regulatory sites is shown in figure 9.

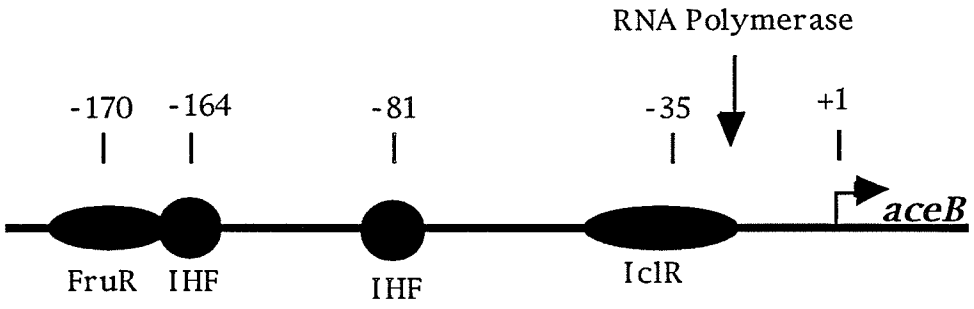


Fig.9 Schematic illustration of the O/P region of the *ace* operon. IclR functions by preventing RNA polymerase from accessing the 10 and -35 regions of the promoter. The molecular mechanisms by which FruR and IHF mediate their effects on *ace* expression are currently unknown.

1.4 Is IclR a Helix-Turn-Helix Protein?

Sequence alignments have suggested that IclR may contain a helix-turn-helix motif (Nègre et al. 1991; Sunnarborg et al., 1990)) This 20 amino acid motif, which is the most common DNA binding domain found in procaryotic repressors generally consists of two α-helices separated by a short turn consisting of three amino acids. The second helix, or the recognition helix, is able to insert itself into the major groove of the DNA and make base specific contacts. Proteins containing this domain have all been found to be multimeric (usually dimers) as this allows a recognition helix from each monomer to insert into adjacent turns of a palindromic B-DNA sequence (fig. 10). Common features of the helix-turn-helix motif suggest that residue 9 should be a glycine, residues 4 and 15 should not be charged, residues 3-8 and 15-20 should not be proline, and residue 5 should not be ß branched (as it is wedged between the 2 helices) As shown in figure 11, a comparison of the primary

sequence of IclR with that of 434 CRO, 434 repressor, λ-repressor, λ-CRO, Hin Recombinase, the cAMP receptor protein (CAP) and the tryptophan repressor suggests that amino acids 46-65 make up the bulk of this motif.

Fig. 10 Illustration of the structure of a helix-turn-helix protein complexed with B-DNA. In this example two CRO repressor monomers (Mondragon and Harrison, 1991) form a dimer and interact with adjacent major grooves of the DNA. As shown in the figure, an α-helix, or the recognition helix protrudes away from the protein and into the major groove.

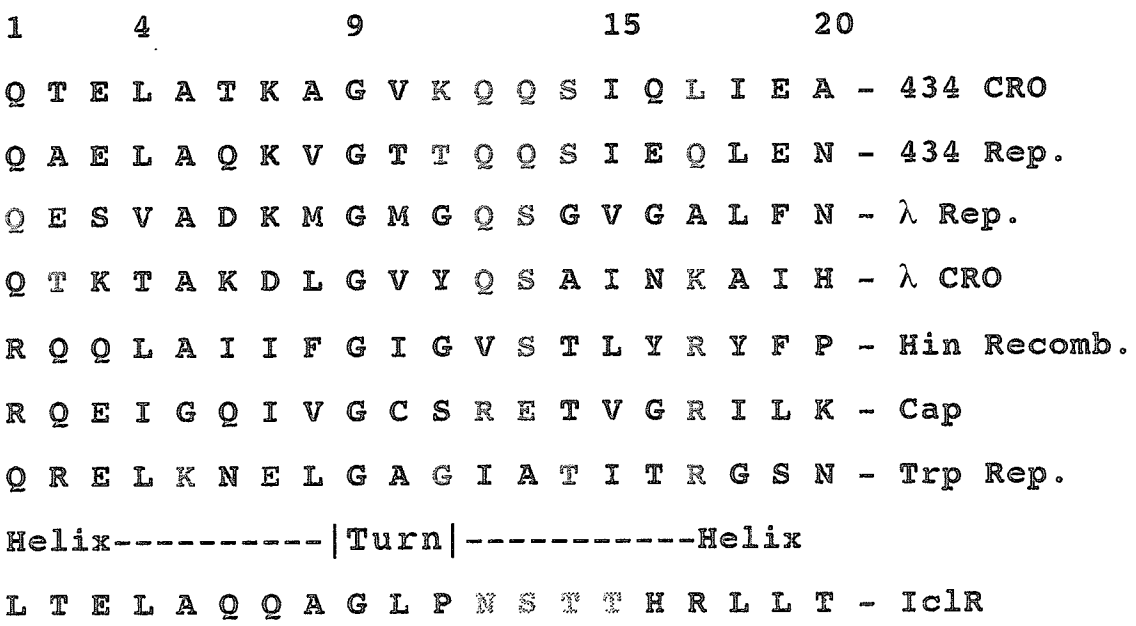


Fig. 11 Sequence alignments of IclR with known helix-turn-helix proteins. The residues highlighted in red have been shown to directly interact with DNA in protein-DNA co-crystal structures. The residues highlighted in green were targetted for site-directed mutagenesis in the experiments described in this thesis. The numbering in the figure represents the position of the amino acid in the helix-turn-helix motif.

In this thesis site-directed-mutagenesis was used to change the identity of the first four amino acids in the recognition helix, asparagine 57, serine 58, threonine 59, and threonine 60. The effects of the mutations on the DNA binding properties of the purified IclR were examined to test the hypothesis that this region was in fact important for mediating the interaction of IclR to its operator.

Chapter 2

Materials and Methods

2.0 Oligonucleotides

Synthetic DNA used in this study was obtained through standard phosphoramidite chemistry performed on an Applied Biosystems automated DNA synthesizer. In all cases small scale (0.2μM) syntheses were performed and the DNA was isolated from the column using standard procedures (Applied Biosystems Bulletin Issue #30). Briefly, the column was washed with concentrated ammonium hydroxide to release the DNA from the solid support. This sample was dried by evaporation in a SpeedVac, taken up in water, extracted with n-butanol, dried again, and resuspended in water. The concentration of each oligonucleotide was determined by measuring the absorbance at 260 nM and making the assumption that an optical density of 1.0 was equivalent to a DNA concentration of 40ug/ml. The molecular weights of the oligonucleotides were determined from the base sequences.

Purification of oligonucleotides SELEX-A, and SELEX-D required an additional step due to their large size. This was accomplished by electrophoresing the entire, n-butanol extracted mix into a preparative 12% denaturing acrylamide gel in 1X TBE buffer. After electrophoresis the oligonucleotides were visualized by UV shadowing and eluted by shaking the macerated gel in a buffer containing 0.5M NH_4OAc, 10mM $MgCl_2$, 1mM EDTA. The DNA was isolated by passing the filtered DNA solution through a Waters C-18 Sep-pak column that had been activated with acetonitrile, washed in water, and equilibrated with 0.5M NH_4OAc. The DNA was eluted

from the column by washing with 60% aqueous methanol and evaporated to dryness in a SpeedVac. This purified DNA was taken up in an appropriate volume of nanopure water.

The oligonucleotide used in the experiments described in this thesis are listed in table 1.

Mutagenic Oligonucleotide

Name	Sequence
DH(R9L)	CCCGCGAAACTCGGCAGAAAA
DH(N57A)	GTTACCCGCTTCCACGA
DH(S58A)	ACCCAATGCCACGACCC
DH(T59A)	CAATTCCGCGACCCACC
DH(T60A)	AATTCCACGGCCCACCG

Sequencing Oligonucleotides

Name	Sequence
DH(SP1)	TACTGGAGTGGATTGCC
DH(SP2)	TATGGTCGGCAGCAGCT
DH(SP3)	TGTCCGCGCCTATCGGC
DH(SP4)	CAAACGCGCAAACGGGG

Selex Oligonucleotides

Name	Sequence
SELEX-A	CCGAAGCGGAATTCAATTN$_{24}$GCATTGGATCCTGATCGC*
SELEX-D	CCGAAGCGGAATTCAATTAAAATGGAAATTGTTTTTGAT TTTGCTAAGGATCCACTAGCG
SELEX-B	CCGAAGCGGAATTCAATT
SELEX-C	GCGATCAGGATCCAATGC

* N is an equal mixture of A, C, G, and T

Consensus 30-mer

Name	Sequence
30-T	AACTAAAATGGAAATGATTTCCACTATACA
30-B	TTGTATAGTGGAAATCATTTCCATTTTAGT

gcl PCR Primers

Name	Sequence
gcl-A	GGCGTTGGGACTGAATTCACATCCAT
gcl-B	GGCGGTAGTGAAAGCTTCTTTCTCCA

Table 1 Oligonucleotides used in the experiments described in this thesis. All of the sequences are listed 5' - 3'.

2.1 Construction of an expression vector for IclR

The structural gene for the IclR protein was obtained from Dr. L.J. Donald in the M13 vector M13mp18Icl*. The gene was contained within an EcoRI site located 31 bp upstream of the transcription start site and a HindIII site located approximately 1.3 kb downstream of the 3' end of the gene. To generate a vector for protein expression this DNA was digested with EcoRI and HindIII and the resulting 1.1kb fragment containing the *iclR* gene was ligated into the Pharmacia plasmid pKK223-3 that had been cut with the same enzymes and dephosphorylated with alkaline phosphatase. This plasmid was chosen as it contains a strong inducible T7 promoter (ptac) and a gene that affords ampicillin resistance to transformants. The ligation mixture was used to transform the *E. coli* strain JM103 that had been made competent with $CaCl_2$, and the resulting transformants were isolated by their ability to grow on LB plates

supplemented with 20 μg/ml ampicillin. To check that the transformants had the correct insert, plasmid DNA was isolated from a number of colonies, digested with EcoRI and HindIII, and electrophoresed on a 0.8% agarose gel. Transformants containing an plasmid with a 1.1kB insert were selected for IclR protein production. The resulting recombinant plasmid, pKKICL8, thus contained the entire *iclR* coding region and 30 bp of the 5' region including a ribosome binding site and initiation codon. As already mentioned, transcriptional control was directed by the ptac promoter.

2.2 Site Directed Mutagenesis

In the experiments performed in this thesis, site directed mutagenesis was performed according to the method of Kunkel (1989). In short, this method utilizes an *E. coli* strain that allows for the *in vivo* synthesis of DNA which contains some uracil. Single stranded template M13 DNA is isolated from this strain and is used in the mutagenesis reaction. Upon transformation into a strain that does not allow for uracil incorporation the uracil containing M13 strand is degraded and theoretically only the mutant strand survives. In practice, however, the number of mutants is usually about 40%, and a mutant is usually detected by sequencing the DNA from the resulting transformants. The oligonucleotides used to generate the mutations are shown in table 1

In table one, the first letter in brackets is the one letter amino acid code of the residue that was changed. The number and the second letter within these brackets refer to the position of the amino acid that was mutated and the residue to which it was mutated. For example, R9L refers to the changing of arginine at position 9 to a leucine.

2.2.1 Construction of Uracil Template

The M13 construct M13mp18Icl* was used as the source for the DNA template for site directed mutagenesis. A single plaque was transferred to LB and incubated for 5 min at 60°C to kill any JM103 cells that may have been transferred. The cells were pelleted and 100 µl of the phage supernatant were transferred to 100 ml of prewarmed LB containing 5ml of log phase JM103 and 10µl of 2.5µg/ml uridine. After growing for 18 hours at 37°C the single stranded DNA template was isolated as described in Maniatis (1982). The purified uracil containing template had a phage titer 10^6 fold greater in CJ236 than it did in JM103.

2.2.2 Mutagenesis

The mutagenesis reaction was performed as described by Kunkel (1987) using the above described template and primers. After ligation, the mixture was taken to 15°C, left overnight, and the following day used to transform the *E. coli* strain JM103 made competent with $CaCl_2$.

2.2.3 Isolation and Identification of Mutants

For each mutagenesis reaction, several plaques arising from the transformation were sequenced with Sequenase v.2.0. The sequencing primer used to identify the R9L mutation was the M13 universal primer while the primer SP1 was used to identify the rest. Once a mutant was identified by this method a large scale preparation of both M13 RF and single stranded DNA was undertaken (Maniatis, 1982). To verify that no changes to the *iclR* gene (other than the desired mutation) had occurred the single stranded DNA was sequenced using the M13 universal primer and sequencing primers SP1, SP2, SP3, and SP4. The double stranded RF isolated from each mutant was digested with *Eco*R1 and *Hind*III and the resulting 1.1kb fragment containing the mutant *iclR* gene was subcloned into pKK223-3 as described above under 'Construction of an expression vector for IclR.' The resulting plasmids were named pKKICL8(R9L), pKKICL8(N57A), pKKICL8(S58A), pKKICL8(T59A), and pKKICL8(T60A). The primers used for sequencing the *iclR* gene are shown in table 1.

2.3 Preparation and Purification of IclR Protein

Cells of the *E. coli* strain JM103 containing the plasmid pKKICl8, pKKICL8(R9L), pKKICL8(N57A), pKKICL8(S58A), pKKICL8(T59A), or pKKICL8(T60A) were grown at 37°C in LB medium that was supplemented with ampicillin at a final concentration of 100μg/ml. In a typical experiment 100 ml of an overnight culture were used to

inoculate one liter of the same medium and when this culture reached early log phase (i.e. A_{600} = 0.50) IPTG was added to a final concentration of 0.5mM to induce transcription from the ptac promoter. When truncated IclR protein was the desired product the cells were allowed to grow for an additional 18 hours before harvesting. When full length protein was the desired product the cells were grown for only an additional 2.5 hours. Harvesting was performed by collecting the cells by centrifugation at 5000g and suspending them in 3 volumes of buffer. When truncated protein was the desired product the buffer was 20mM TRIS pH 7.8, 0.2M KCl, 1mM EDTA, 20mM β-mercaptoethanol. It was found that addition of β-mercaptoethanol greatly improved the stability of the protein during purification. To minimize truncation the cells were harvested in a buffer that was similar to that described above, however, the EDTA was omitted and $ZnSO_4$ was added to a final concentration of 1mM. This step was added as it was postulated that the IclR protein was being cleaved by the *E. coli* protease OmpT (Sugimura and Nishihara, 1988). Since EDTA solubilizes this outer membrane protein and Zn^{2+} inhibits it the above described adjustment usually prevented truncation.

Cells were disrupted in a French Press Cell at 18000psi and the cell supernatant was separated from the cell debris by centrifugation at 27000g [12000 r.p.m. in Sorvall SS34 rotor]for one hour at 4°C. The supernatant was loaded onto a BioRex-70 (Bio-Rad) column (approximately 1.0 cm X 10.0 cm) that had been equilibrated in the same buffer as used for cell suspension, and protein that did not

stick to the column was eluted with 3 column volumes of the same buffer. To elute the IclR protein a linear gradient was set up using 500ml each of a buffer similar to the ones described above, except containing either 0.2 M KCl or 0.7M KCl. Fractions containing 150 drops (~12ml) were collected and protein content was determined using the Bradford (1976) protein assay (Bio-Rad) while the salt concentration was determined by comparing the conductivity of the fractions to a set of standards that were equilibrated at the same temperature as the protein fractions. Fractions identified as containing a large amount of protein were subsequently analyzed by SDS-PAGE (Laemmli, 1970) and IclR containing fractions were identified on the basis of a strong band that migrated just before carbonic anhydrase (~30K). It was found that full length IclR protein eluted from the column at a KCl concentration of 0.35M while truncated IclR generally eluted in a broad peak that was centered at 0.6M KCl. For the case of the R9L mutant the salt gradient was reduced from 0M KCl to 0.5M KCl as it was found that the full length protein eluted at a salt concentration of approximately 0.2M KCl. Fractions containing the IclR protein were pooled and concentrated by ultrafiltration using an Amicon YM10 membrane. This procedure usually afforded IclR protein that was already highly purified as determined by SDS-PAGE. It was estimated that IclR was at least 90% pure at this stage of purification.

To further purify the protein the BioRex fractions containing a high proportion of IclR were concentrated and then loaded onto a Sephadex G-100 column (~2cm X 100cm) that had been equilibrated

in fresh buffer (20mM TRIS pH7.8, 0.4M KCl, 1mM EDTA, 20mM β-mercaptoethanol). Fractions (12ml) were collected and those containing IclR were identified by SDS-PAGE as described above. The fraction containing the least amount of impurities was immediately dialyzed against two 500ml changes of a buffer containing 20mM TRIS pH7.8, 1mM EDTA, 0.4M KCl, 1mM DTT, 30%(v/v) glycerol (SB). This process not only changed the buffer, but effectively concentrated the protein due to the decrease in volume that accompanied dialysis. The final purified protein was stored at -20°C where it retained activity for periods of several months. The protein concentration was determined using the Bradford Protein assay using bovine gamma globulin as a standard. In all the steps described above the temperature was maintained at 4°C.

2.4 Oligomeric State of IclR

Sephadex G-100 chromatography (column dimensions; 2.0 X 120 cm) was used to examine the native molecular weight of the IclR protein. Samples (approximately 2-3 mg) were applied to the column in 1.0 ml of column buffer (20mM Tris (pH 7.8), 0.4M KCl, 50mM β-mercaptoethanol, 1mM EDTA) containing 5% glycerol. Fractions containing 100 drops (5.7 ml) were collected and protein was determined by the method of Bradford (1976). The column was standardized using lysozyme (14K), carbonic anhydrase (30K), BSA (66K), and alcohol dehydrogenase (150K), and the void volume was determined using dextran blue.

2.5 Preparation of Samples for Mass Spectrometry

For mass spectrometry experiments it proved necessary to remove all of the buffer components to achieve satisfactory spectra. This was accomplished by extensive dialysis of the purified protein against water. This procedure caused the protein to precipitate, however, it could be resolubilized by adding acetic acid to a final concentration of 5%. The acidified samples were usually diluted to a final protein concentration of approximately 1μM and methanol was added to a concentration of 40%. All the mass spectra data were obtained on a ESI-TOF mass spectrometer by Igor Chernushevich of the Department of Physics at the University of Manitoba (Donald et al. 1996).

2.6 Radiolabeled IclR binding probes

Labeled DNA fragments containing the IclR binding site were obtained from a variety of sources. The plasmid ptZop a construct obtained from Dr. L.J. Donald, houses the 5' region of the *aceB* gene (Donald et al. 1996). This fragment can be excised from the plasmid by digestion with *Bam*HI or by digestion with *Eco*RI and *Hind*III. Thus, to obtain a fragment 170 bp long the plasmid was first digested with *Bam*HI, and the 5' overhangs were filled in using the Klenow fragment of *E. coli* DNA polymerase and α-^{32}P dATP. To obtain a 227 bp fragment the DNA was cut with HindIII and EcoRI prior to labeling. The small DNA fragments were separated from the rest of the plasmid by electrophoresis on a non-denaturing 5%

polyacrylamide gel. The DNA of interest was located by autoradiography and purified as described in the SELEX procedure using either the shaking or freeze squeeze methods.

The plasmid pCHicl, a 6383bp plasmid containing all of the *iclR* gene including the 5' operator-promoter region was also obtained from Dr. L.J. Donald. Upon digestion of this plasmid with *Xmn*I and *Mlu*I, a 333 bp fragment containing the *iclR* operator-promoter region was liberated and subsequently 5' end labeled. The end labeling was achieved by treating the restriction digest with calf intestinal alkaline phosphatase and, following phenol-chloroform extraction and ethanol precipitation, phosphorylation with T4-polynucleotide kinase in the presence of [γ-^{32}P] ATP (Maniatis, 1982). The labeled fragment was purified by electrophoresis as described above.

Oligonucleotide SELEX-D, a synthetic 60-mer oligonucleotide containing base pairs -53 to -24 (Nègre et al. 1992) of the *aceB* promoter, was purified by 12% denaturing PAGE as described in "Oligonucleotides." To label this fragment, the oligonucleotide SELEX-B was first 5' end labeled with T4-polynucleotide kinase and [γ-^{32}P] ATP. This labeled probe was then used without purification to render an equimolar amount of SELEX-D double stranded in a standard primer extension reaction containing unlabeled dNTPs and the Klenow fragment of *E. coli* DNA polymerase (Maniatis, 1982). The labeled duplex was purified away from unincorporated primers and dNTPs by electrophoresis through a 12% polyacrylamide gel. The

DNA was visualized by autoradiography and eluted as described above.

To label the synthetic 30-mer containing the IclR consensus sequence, the oligonucleotide 30-T was first end labeled with [γ-^{32}P] ATP and T4-polynucleotide kinase. Duplex formation was initiated by adding a two-fold excess of the complementary oligonucleotide, 30-B, to the kinase reaction, heating the mixture at 80°C for 5 min, and letting it cool to 20°C over a period of ten min. The duplex was purified by chromatography through a Sephadex G-50 column.

Double stranded M13 RF was used to generate gel shift probes from the selected SELEX DNA. The M13 RF was isolated as described above (see subcloning of the iclR gene) and purified as described in Maniatis (1982) with omission of the CsCl purification step. Gel shift probes were labeled and purified as described above following digestion of the RF with *Eco*RI and *Hind*III.

The IclR site 5' to the *gclB* gene was obtained by a PCR reaction containing in 100µl, 100ng of *E. coli* genomic DNA, 1µM of primers *gcl*-A and *gcl*-B, 200µM of each dNTP, 10mM Tris (pH 8.8), 50mM KCl, 1.5 mM $MgCl_2$, and 2.5U of Taq polymerase. For purposes of cloning and labeling primer A was designed to contain an *Eco*RI site and primer gclB was designed to contain a *Hind*III site. Following phenol/ $CHCl_3$ extraction and ethanol precipitation, the PCR product was gel

purified and cloned into M13mp18 RF. DNA binding probes were generated from the M13 clone as described above.

2.7 DNA Binding Assays

Operator DNA binding assays were monitored by electrophoresis mobility shift assays (Carey, 1991) in which IclR protein was used to titrate a constant amount of one of the above described labeled operators. In a typical reaction (12μl) approximately 1500 cpm (0.07fmol) of DNA and varying amounts of freshly diluted IclR protein were incubated for 20 min at room temperature in a buffer (GSB) containing 20 mM TRIS (pH7.8), 50mM KCl, 8.3mM $MgCl_2$, 0.6mM EDTA, 0.4mM DTT, 42μg/ml BSA, and 5% glycerol. In some experiments more than one type of operator fragment was used. In these experiments the total counts were still 1500cpm. After incubation 10 μl of the reaction mixture was loaded onto a pre-run 5% polyacrylamide gel (19:1-acrylamide:bis-acrylamide) in 1X TBE buffer and electrophoresed at 100V until the bromophenol blue marker had migrated about 5 cm. After drying and autoradiography, the exposed film was used as a template to excise gel regions corresponding to bound and free DNA. The bands were quantitated by scintillation counting and the percentage of bound DNA was plotted against IclR concentration. Data were fit to a modified form of the Hill equation (Hill, 1910) that accounted for background using the Macintosh program KaleidaGraph Version 2.1.1 (Abelbeck Software) The equation fitted was:

$$\%bound = A+95(K_D^n)/(1+K_D^n)$$

The parameter A, which ideally should have been zero, was allowed to be adjusted, while the upper limit was set at 95 as this was commonly found to be the percentage of DNA bound at saturating IclR concentrations. The term K_D was taken to be the IclR concentration at which half of the operator was bound. The Hill number, n, was also used as an adjustable parameter in the fits. This equation was valid as the concentration of DNA used in the assays (approximately 5-10 pM) was well below that of the protein.

To monitor the effect of KCl and $MgCl_2$ on complex formation the binding reactions were performed exactly as described above in GSB buffer containing the indicated amounts of KCl or $MgCl_2$. The KCl concentrations tested were 2.5, 10, 25, 50, 100, 200, 300, and 400mM while the $MgCl_2$ was tested at 0, 5, 10, and 20mM. In all cases the obtained data were fit to the modified Hill equation described above.

2.7.1 Mixed Oligomer Experiments

Mixed oligomer experiments were performed in an attempt to determine the stoichiometry of binding of IclR to the *ace* operator. This method took advantage of the fact that the complex formed between the double stranded 30 mer (30-T and 30-B that had been

annealed) and full length IclR had a slower mobility then a complex formed by truncated IclR and the 30-mer. Full length and truncated wild type IclR, at a concentration of 23μM in SB were mixed together and urea was added to a final concentration of 5.6M. The solution was incubated for one hour at room temperature and then dialyzed against two changes of 500 ml of SB at 4°C to remove the urea and refold the protein. Preliminary fluorescence experiments suggested that roughly half of the individual protein molecules refolded under these conditions. The mixed sample was then serially diluted in GSB and mixed with 3000cpm of labeled, double stranded 30-mer in a final volume of 15μl. After a 20 min incubation at room temperature the sample was loaded onto a pre-run, 37cm long, 1X TBE/5% polyacrylamide gel running at 250V. The sample was electrophoresed until the xylene cyanol marker had migrated approximately 35cm (typically about 12hrs), at which time the gel was dried and autoradiographed as described above.

2.7.2 Association Kinetics

Rates of association of IclR protein with operator DNA were monitored by PAGE in a similar manner to the equilibrium binding measurements described above. In a typical experiment 100μl of a solution containing approximately 5000 cpm (5 pM) of the *ace* operator in GSB was gently mixed with 20μl of IclR that had also

been diluted in GSB. At different times 10µl aliquots were removed from the reaction and immediately loaded onto a 1X TBE/5% acrylamide gel running at 300V. When all the samples were loaded onto the gel the voltage was kept at 300V for 1 min and then reduced to 150V until the bromophenol blue had migrated approximately 5cm. The gels were dried and exposed to X-ray film. As the concentration of IclR was much greater than that of operator, the rate constant for association was obtained using standard first order treatments.

2.7.3 Dissociation Kinetics

Dissociation rates were monitored by following the rate of loss of labelled DNA from the bound complex upon addition of unlabeled *ace* operator. In a typical experiment 20µl of IclR (diluted in GSB to a concentration such that approximately 90% of labeled *ace* operator would be bound at equilibrium) was added to 100µl of GSB containing 15000cpm (15 pM) of *ace* operator. The mixture was incubated for 10 min at room temperature to allow the reaction to reach equilibrium, and then 5µl of unlabeled *ace* operator was added to a final concentration of 10-50 nM. At various times 10µl aliquots were removed from the reaction and loaded onto a 1X TBE/5% polyacrylamide gel running at 150V. When the bromophenol blue marker had migrated approximately 5 cm the gels were dried, autoradiographed, and quantitated as described above. It should be noted that in all cases the rate of dissociation was not dependent on the concentration of unlabeled competitor in the range 10-50nM. In

addition, when the competitor was added to the reaction at the same time as the labeled DNA, no binding of the labeled operator was observed. First order rate costants were determined by plotting the natural logarithm of the percentage of bound DNA as a function of time.

2.8 Dimethyl Sulfate Protection Experiments

Dimethyl sulfate (DMS) protection experiments were performed as described by Vershon et al. (1987) with the following modifications. IclR was diluted into a reaction buffer containing 50 mM sodium cacodylate (pH7.5) 50mM KCl, 8.3mM $MgCl_2$, 0.6mM EDTA, 0.4mM DTT, 42μg/ml BSA, 5% glycerol, and approximately 25000 cpm of 3' end labeled operator fragment (from ptZop). After a 20 min incubation at room temperature, 5μl of a fresh 1:10 dilution (in water) of DMS was added and the reaction was allowed to proceed for two minutes before the addition of 50 ul of G-Stop buffer (1.5M-NaOAc (pH6.0), 1.0M β-mercaptoethanol, 100μg/ml tRNA) and 750 μl of 100% ethanol. The DNA was precipitated in a dry ice ethanol bath for 10 min, resuspended in 250 μl of 0.3M-NaOAc (pH6.0), precipitated in ethanol, and allowed to dry. Cleavage reactions were performed in one of two ways: Cleavage at G residues was accomplished by suspending the sample in 100μl of a solution containing 1M piperidine, 5mM EDTA, and heating for 30 min at 90°C. The samples were dried with a SpeedVac concentrator, taken up in 100μl of nanopure water and dried again in the SpeedVac

concentrator. The water wash was repeated two more times and the dried sample was taken up in sequencing sample buffer (Maxam and Gilbert, 1980). These samples were then heated to 90°C for 2 min before being loaded onto a 8% (w/v) polyacrylamide/urea sequencing gel. The gel was run at 1200V in 1X TBE buffer until the xylene cyanol marker had migrated to the bottom of the gel. The gels were then fixed in 10% methanol, 10% acetic acid, blotted on to Whatman 3MM paper, dried under vacuum, and exposed to X-ray film at -70°C with two intensifying screens. Cleavage at guanine and adenine residues was performed as described in Maxam and Gilbert (1980). The dried, methylated samples were depurinated by resuspended them in 20 µl of 20mM KPO_4 (pH7.0), 1mM EDTA, and heating to 90°C for 30 min. After a 5 min incubation on ice strand cleavage was initiated by the addition of 2 µl of 1M NaOH and heating at 90°C for 30 min. The DNA was precipitated at -70°C by first diluting the sample with 180µl of 0.3M NaOAc (pH6.0) 10mM $MgCl_2$, and then adding 600µl of 100%-ethanol. After washing the sample with 70% ethanol, the DNA was allowed to dry before being taken up in sequencing sample buffer and electrophoresed as described above.

2.9 Selection of DNA that binds to IclR: Selex

Initial experiments were performed using 5' end labeled double stranded Selex-D (see table 1) to determine the mobility of DNA that was bound to IclR under standard gel shift conditions. These experiments demonstrated that the IclR-Selex-D complex

migrated to a position roughly 0.5 cm in front of the xylene cyanol marker when the bromophenol blue had migrated 5 cm into the gel. In the first round of binding approximately 1.0μg (1800nM) of double stranded random oligonucleotide (SELEX-A) was incubated under standard gel shift conditions (see DNA binding) with IclR protein at a concentration of 160nM. After a 30 minute incubation at room temperature the entire mixture was electrophoresed on a pre-run 5% polyacrylamide gel until a bromophenol blue marker had migrated approximately 5 cm. To recover the DNA bound to IclR a gel slice measuring approximately 0.5cm X 1cm was excised just above the xylene cyanol marker, placed in a 1.5 ml microfuge tube crushed with a sterile 1 ml syringe plunger, and mixed with 500 μl of SDS containing elution buffer (0.5M NH_4OAc, 10mM $MgCl_2$, 1mM EDTA, 0.1%(w/v) SDS). After shaking for 5-6 hours at room temperature, the resulting slurry was filtered through a 0.2 micron filter (Nalgene), 5μg of tRNA was added, and the resulting mixture was precipitated with 1.0 ml of 100% ethanol. In some experiments the DNA was isolated from the gel using the squeeze freeze method (Beutel and Gold, 1992). Briefly, the DNA containing gel slice was frozen in a dry-ice ethanol bath for 5 minutes at which time 600μl of 0.5M NaOAC, 2mM EDTA was added. The gel was crushed with a one ml plunger and the resulting slurry forced through a 0.2 micron filter. As above, 5μg of tRNA was added and the DNA was precipitated with ethanol. After washing the precipitate twice with 70% ethanol, the dried sample was taken up in 25 μl of TE buffer. Fifteen μl of the sample was used in a 100 μl PCR reaction containing 0.24 mM of the primers SELEX-B and SELEX-C, 500μM of each dNTP,

10mM Tris (pH 8.8), 50mM KCl, 2.5 mM $MgCl_2$, and 2.5U of Taq polymerase. In all PCR experiments the reaction mix containing everything but template and Taq polymerase were heated to 75°C for 5 minutes and then to 94°C for 3 minutes. While the samples were at 75°C either template or water was added to the appropriate tube. Taq polymerase was added when the temperature reached 94°C. The PCR was performed for 10 cycles with denaturation occurring at 94°C for 1 min., annealing at 55°C for 1 min., and extension at 72°C for 1 min. The temperature was held at 72°C for 5 min. after the last cycle to ensure synthesis was complete on each template. In all cases 10% of the reaction product were analyzed by 12% PAGE and in most cases only samples containing template gave detectable amounts of product (as described later, the control reaction sometimes gave a small amount of product).

To obtain additional product a second PCR reaction was performed on the product obtained above. In this reaction 1 μl of the first PCR reaction was used as a source of template for each of three separate 100 μl PCR reactions. After PCR all three reaction mixes were pooled, cleaned up with phenol and chloroform, and precipitated with ethanol. The precipitated product was washed twice with 70% ethanol, dried, and purified by PAGE on a 12% gel. The product band was visualized by UV shadowing and the DNA was eluted from the gel as described above with the exception that tRNA was omitted from the precipitation stage. This DNA was used as the starting material for the next round of the SELEX procedure. It should be noted that all the solutions used in the PCR, with the

exception of the primers and the Taq polymerase, were treated in microfuge tubes (with the lids open) with long wave UV light for 30 minutes prior to use. It was found that this precaution reduced the amount of non-specific amplification product.

The second, third, and fourth SELEX cycles were performed as described above except that the IclR concentrations were 150nM, 35nM, and 8nM, respectively. In addition only 250 ng (450 nM) of DNA were used for these later rounds.

The conditions described above were worked out in a number of preliminary experiments. First, the $MgCl_2$ concentration had to be optimized. This was accomplished by using single stranded Selex-A as a template for a PCR reaction containing variable amounts of $MgCl_2$. The hot start (adding Taq polymerase only when the reaction reached 94°C) was required as early experiments showed evidence of a band smaller than 60 bp (presumably a primer dimer, Saiki, 1989) when the template and Taq polymerase were added to the mix at room temperature. This band was never present in reactions that used the hot start procedure. Finally, the number of cycles had to optimized in order to obtain the maximal yield of product. As observed by Pollock and Treisman (1990) this is best accomplished by taking care that all PCR products are double stranded. This can be achieved by using a large excess of primers and dNTPs, and by using the minimum number of PCR cycles to avoid primer or dNTP depletion. As pointed out by Cui et al. (1995) primer depletion could result in the annealing of two non-complementary strands that are

double stranded at the ends but single stranded in the middle. Consistent with this fact was that preliminary experiments using either 15 or 20 PCR cycles showed evidence of a slow moving band that was heterogeneous in size. This band was observed with a concomitant decrease in the amount of amplified 60-mer.

To further improve the affinity of the above selected oligonucleotides for IclR a nitrocellulose filter was used to partition the bound and free DNA. This method works because of the strong affinity nitrocellulose has for protein but not for double stranded DNA. Thus, only DNA that is bound to the protein remains on the filter. When the free DNA is washed away, the bound portion can be eluted from the protein and amplified.

The concentrations of IclR used for the first three rounds of the SELEX procedure (using nitrocellulose) were chosen to be relatively high, in order to maximize the proportion of selected selected DNA over that which might be non-specifically retained by the filter. This was required as control experiments using a radiolabelled probe demonstrated that a small amount of DNA bound to the filter even in the absence of IclR. Thus in the first three rounds of the SELEX procedure the IclR concentrations were 175nM, 100nM, and 75nM, respectively. For the next three rounds of SELEX, the IclR concentration was kept steady at 22nM. In these experiments IclR was incubated with double stranded random oligonucleotide as described above. The resulting mixture was filtered on a 2cm X 2cm nitrocellulose filter that had been equilibrated in the same buffer,

before being washed with 6 ml of GSB. The DNA was eluted from the filter by a 30 min incubation in 50µl of 0.5M KCl at 65°C. The eluted DNA was amplified as described above with the exception that the KCl component of the reaction came from the solution containing the eluted DNA. Thus 15µl of the KCl solution was added to a standard PCR reaction at 75°C and the material was amplified for 10 cycles before being used to start 3 secondary PCR reactions that provided additional product. The amplified product was pooled and purified as described above. In the first two rounds of the procedure approximately 1.0 µg of DNA was used, while only 0.2µg of DNA was used for the next 4.

Two more rounds of SELEX, using 5nM IclR and 0.2µg of DNA were attempted, however, binding assays determined that the DNA selected in these rounds had diminished affinity for the protein. Consequently, these samples were discarded and the DNA selected in the previous rounds were used for the experiments described below.

2.9.1 Preparation of gel shift probes from SELEX selected DNA pools

Approximately one nanogram of the selected, gel purified, random oligonucleotide was labeled by incorporation of [α-^{32}P] dATP during PCR. The PCR reaction was performed as described above

with the following modifications: The unlabeled dNTP's (dCTP, dGTP, and dTTP) were added at a final concentration of 1μM, and 30μCi of [α-^{32}P] dATP was used as a source of label. The low concentrations of unlabeled dNTPs were chosen in order to prevent the Taq polymerase from incorporating the wrong nucleotide at positions that should have been occupied by dATP. The PCR was continued for 3 cycles and the labeled probe was purified by electrophoresis as described above. In this case, however, the position of the probe was determined by autoradiography rather than UV shadowing.

2.9.2 Cloning and Sequencing of SELEX selected Oligonucleotides

Gel purified DNA selected in the last round of the SELEX procedure were digested with both BamHI and EcoRI before being ligated into M13mp18 RF that had been cut with the same enzymes. Transformation of this ligation mixture into the *E. coli* strain JM103 was carried out in the presence of X-Gal and IPTG and plaques containing an insert (i.e. those that were not blue) were used to generate single stranded DNA for sequencing (Maniatis, 1982). The DNA was sequenced using Sequenase version 2.0 (United States Biochemical) according to the manufacturer's protocol. The M13 universal sequencing primer was used in all cases. Sequences of the selected DNAs were aligned by eye by placing them around the conserved GGAAANNNTTTCC found in most of the clones.

2.10 Database Searching

E. coli sequences that contained the IclR binding sequence were discovered using the *E. coli* database collection and the on-line Fasta search found at the world wide web site **http://susi.bio.uni-giessen.de/usr/local/www/html/ecdc.html.** In addition to the sites upstream of the *aceBAK* and *iclR* genes, the search engine found two high scoring sequences in known *E. coli* genes. The first site was located at nucleotides 93-107 of the *gclB* gene (Chang et al., 1993) and the second site was located at nucleotides 6310-6323 of the *hisF* gene (Carlomagno, et al. 1988).

2.11 Sequence Comparisons and Secondary Structure Predictions

Protein Sequences that were similar to IclR were downloaded from the Swiss-Prot database found on-line at **http://expasy.hcuge.ch/cgi-bin/sprot-search-ful.** The downloaded sequences included IclR from *E. coli* and *S.typhimurium* (P16528 and P17430 respectively), GylR from *S.coelicolor* and *S.griseus* (P15360, and P22866 respectively), and KdgR from *E.chrysanthemi* (P37728). The sequences of YcsO from *B.subtilus* (P42968), YiaJ from *H.influenzae* (P44996) and YjhI from *E. coli* (P39360) were also downloaded, however, these proteins are only hypothetical as they are open reading frames discovered by genome sequencing projects which have yet to be studied. The numbers in

brackets are the Swiss-Prot identification numbers. Sequence alignments were performed using the Pam250 scoring matrix (Feng et al. 1985) and secondary structure predictions were performed using both the Chou-Fasman (Chou and Fasman, 1974) and Robson-Garnier (Robson and Suzuki, 1976) methods. Regions of secondary structure similarity were only considered significant when both the Chou-Fasman, and Robson-Garnier methods predicted similar regions of secondary structure in six of the eight above protein sequences. Both the sequence alignments and secondary structure predictions were performed using the Macintosh program MacVector 4.0$^{T.M.}$ (Kodak Life Science Products).

Chapter 3

Results

3.0 IclR Production, Purification, and Physical Properties

IclR protein was obtained from the *E. coli* cell line JM103 using the techniques described in the previous section. As described in previous work (Donald, et al., 1996) care had to be taken to prevent truncation of the protein during the course of its purification. This truncation, which removes the first eight (and sometimes nine) amino acids, is thought to be caused by prolonging the exposure of the protein to the *E. coli* protease OmpT, a protein that is solubilized in the purification procedure by EDTA, but inhibited by Zn^{2+} ions. In this study Zn^{2+} was usually present during IclR purification to prevent truncation.

The first step of the purification procedure utilized the affinity of the IclR protein for the cation exchange resin BioRex-70. This step was extremely efficient and usually gave IclR that was at least 90% pure as judged by an SDS gel that had been stained with Coomassie Blue. Fractions identified as containing a high proportion of IclR were concentrated and further purified by chromatography through a Sephadex G-100 column. This last chromatographic step generally gave protein that was at least 95% pure and was of suitable quality for physical characterization and binding studies. Figure 12 shows the elution profile for the IclR mutant N57A from both the BioRex-70 and Sephadex G-100 columns.

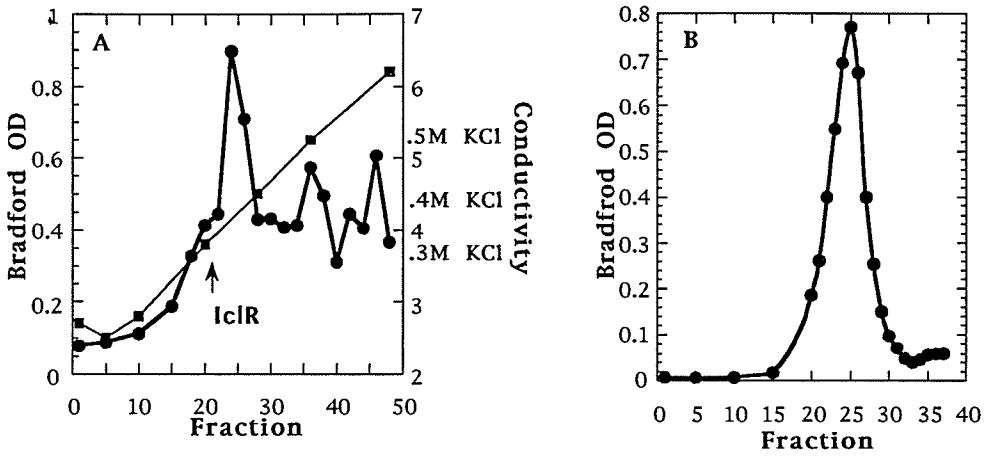


Fig. 12 Elution profiles of full length, N57A IclR from A)BioRex-70 and B) Sephadex G-100

The protein was eluted from the Bio-Rex column at a salt concentration of approximately 0.35 M KCl. This peak, when concentrated and run through the Sephadex column elutes as a large single peak. The elution profiles of the wild type, S58A, T59A, and T60A proteins were all similar to that of the N57A mutant. The elution profile of the R9L protein was similar to that obtained for the other mutants; however, it eluted from the BioRex column at a KCl concentration 0.15M less than the others, suggesting that this residue is important for anchoring the protein to the exchanger. Truncated protein generally eluted from the column at a salt concentration of 0.5-0.6M. The G-100 elution profile of the R9L mutant and the truncated IclR was identical to the others.

3.1 Mass Spectrometry

In all the proteins studied it proved imperative to characterize them by mass spectrometry. The mass data so obtained were used to characterize the extent of truncation of each batch of protein , and also to verify mutations. Knowing whether truncation had occurred proved to be very important as it was later shown that DNA binding was dramatically effected by this modification (see DNA Binding) The extent of truncation proved to be variable from preparation to preparation, however, when the proper precautions were taken, it was usually found to affect less than 5% of the sample. An example of the ESI-TOF mass spectrum obtained for the N57A mutant is depicted in figure 13.

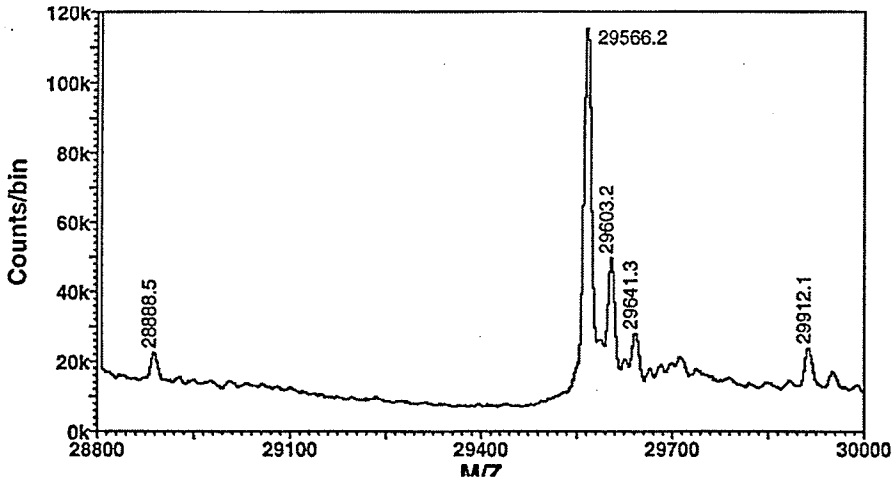


Fig. 13 ESI-TOF mass spectrum for the IclR mutant N57A. The full length protein corresponds to the peak at 29566.2u, while the truncated peak has a mass of 28888.5u.

The full length protein corresponds to the peak shown at 29566.2u while the truncated corresponds to the peak at 28888.5u. The peaks

that are multiples of 37u bigger than the above described peaks are potassium adducts. The masses found for the truncated and mutant IclR proteins used in this study are listed in table 2.

A)

	Full Length		Truncated	
Protein	Theoretical	Experimental	Theoretical	Experimental
W.T.	29608	29608	28931	28775
R9L	29564	29561	28882	n.d.
N57A	29566	29566	28886	28889
S58A	29592	29591	28913	28913
T59A	29578	29579	28899	28901
T60A	29578	29578	28899	28902

B)

Val-Ala-Pro-Ile-Pro-Ala-Lys-Arg-Gly-Arg-Lys-Pro-Ala*

↑

OmpT site

Table 2. A) Masses obtained by ESI-TOF Mass Spectrometry for the IclR Proteins described in this thesis. B) Illustration of the N-terminal sequence of wild type IclR and the peptide bond thought to be cleaved by the OmpT protease.

3.2 Sizing IclR by Exclusion Chromatography

The subunit composition of the IclR protein was studied by monitoring the elution of the protein from a Sephadex G-100 column that had been standardized with lysozyme (14K), carbonic anhydrase (30K), bovine serum albumin (66K), and alcohol dehydrogenase (150K); and the void volume was determined using Dextran Blue. As shown in Figure 14 IclR elutes from the column with a calculated mass of 117 000 Da, suggesting that the native protein exists as a tetramer at micromolar concentrations.

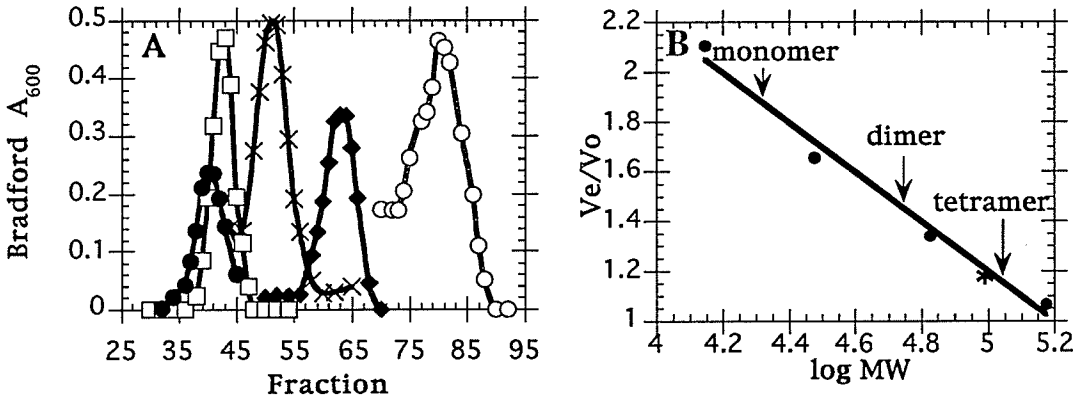


Fig. 14 A) Elution profiles of IclR(❐) and the standards lysozyme(O), carbonic anhydrase(♦), BSA(×), and alcohol dehydrogenase(●). from Sephadex G-100. B) Plot showing the normalized elution volume (V_e/V_O; where V_e is the elution volume of the standard, and V_O is the void volume) of the standards and IclR as a function of log M.W. The arrows indicate the expected position for species of molecular mass 30kDa (monomer), 60 kDa (dimer), and 120kDa (tetramer). The starindicates the elution volume of IclR from the column, suggesting that the native molecular mass is approximately 117kDa.

On the basis of ultracentrifugation experiments, Nègre et al. (1991) found that IclR was a dimer with an average molecular weight of 60kDa. The discrepancy between their result and mine suggests that IclR may exist as a tetramer at high concentrations, but perhaps dissociates to dimers or even monomers, as the concentration falls to levels that are used in DNA binding experiments (10^{-8}-10^{-12}M). The Sephadex elution profiles of all the mutants were identical to that obtained from the wild type preparation indicating that the mutations did not affect the subunit-subunit interactions. The truncated protein also eluted in the same position as full length indicating that the N-terminal arm of the protein doesn't contribute significantly to oligomerization.

3.3 Prevention of Truncation with the Mutation R9L

As described above, the purification scheme employed to obtain full length IclR protein required that Zn^{2+} be included in the purification buffer. This precaution is thought to inhibit the activity of the protease OmpT (see above). Unfortunately this method does not always yield a protein that is 100% full length. It was surmised that mutation of the OmpT recognition site in IclR might prevent the modification while not affecting the structure and function of the molecule in a dramatic fashion. To test this hypothesis the R9L mutant was prepared in a manner that always led to truncation when using wild type protein; the cells were allowed to grow for 18 hours after the addition of IPTG and in addition, they were harvested in a buffer containing EDTA, and devoid of Zn^{2+}. The first indication

that this modification was successful in preventing truncation was observed in the elution profile of the protein from the cation exchanger (see above). Normally truncated protein eluted from the column in a broad peak centered around 0.6M KCl, while full length eluted at 0.35 M. The R9L, as described above, eluted at approximately 0.2M KCl, a value closer to that expected for a full length protein missing a positively charged arginine residue. The ESI-TOF mass spectrum of the protein gave a mass of 29561, the mass expected for the full length protein with an arginine changed to a leucine, within experimental error. The lack of a peak at a mass of 28882 in the same ESI-TOF spectrum indicated that no truncation of the sample had occurred using this procedure. Unfortunately the DNA binding properties of this protein were different from those found with the full length wild type protein thus preventing it from being used as a "wild type mimic" (see below).

To date, the exact nature of the agent(s) responsible for truncation of the IclR protein is not exactly known. The fact that Zn^{2+} and mutation of the OmpT recognition site somewhat inhibits the process suggests that OmpT is in fact the agent responsible. However, the fact that wild type IclR obtained from cells grown overnight is still truncated even when Zn^{2+} is added suggests that some other proteases may be responsible, perhaps acting synergistically with OmpT. This possibility is supported by the observation that an *E. coli* strain defective in *ompT* will still give truncated protein if the cells are grown overnight (L.J. Donald, personal communication).

3.4 Prediction of IclR Secondary Structure

To date there does not exist a structure for the IclR repressor protein either alone or in complex with DNA. Until a structure becomes a available regions of secondary structure can only be predicted. In the case of IclR, the primary sequence was used at the Swiss-Prot protein database to search for amino acid sequences that were similar to that of IclR from *E. coli*. In total, and as described in Materials and Methods, eight proteins were found by this search. The proteins included the KdgR repressor from the organism *E. chrysanthemi*, GylR from the organisms *S. coelicolor* and *S. griseus* , IclR from the organism *S. typhimurium* and three other proteins whose function is unknown but whose sequences have been discovered via genome sequencing projects in various organisms. All of the proteins possess a putative helix-turn-helix motif and all are thought to interact with DNA. Of the known examples, the KdgR repressor has been shown to negatively regulate expression of the genes involved in pectin metabolism (Reverchon, et al. 1991; Nasser et al. 1994) while GylR has been shown to act as an activator of the glycerol operon (Smith and Chater, 1988).

Alignment of the IclR sequence from *E. coli* with the other eight was performed using the program MacVector which uses a Pam250 scoring matrix. This matrix allows for the optimal alignment of amino acids and allows the introduction of breaks in the sequence to improve the fit. The results of the alignment are shown in figure 15. Solid lines indicate regions of high homology, while hollow lines

indicate regions where breaks in the sequence had to be added to improve the fit.

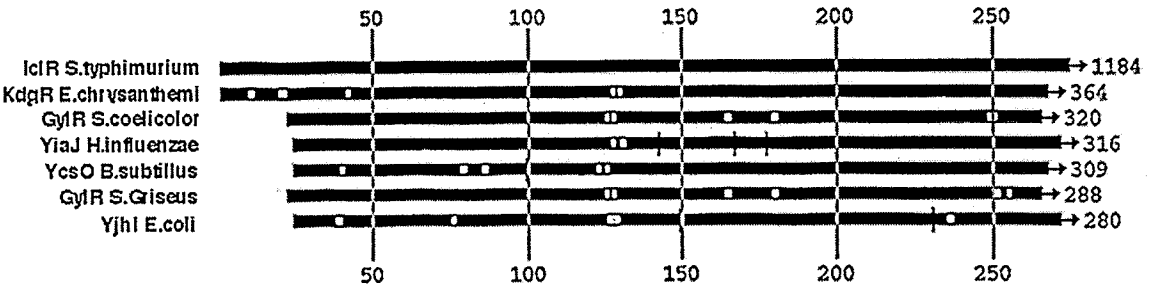


Fig. 15 Results of the sequence alignments using IclR from *E. coli* against the proteins listed in the text. Solid bars indicate regions that contain similar amino acid sequences, while open bars indicate regions where gaps had to be introduced to improve the fit. The numbers to the right of each line indicate the Pam250 score, with high numbers resembling the sequence of IclR from *E. coli*

In order to predict regions of secondary structure in the IclR protein from *E. coli,* secondary structure predictions were carried out for all of the aligned proteins described above. Although secondary structure predictions alone are not very reliable for non-membrane proteins, it was assumed that if regions of predicted secondary structure were common to six of the eight proteins these regions had a greater chance of really existing. The regions that meet the criteria described above are summarized in figure 16, using IclR as the representative sequence.

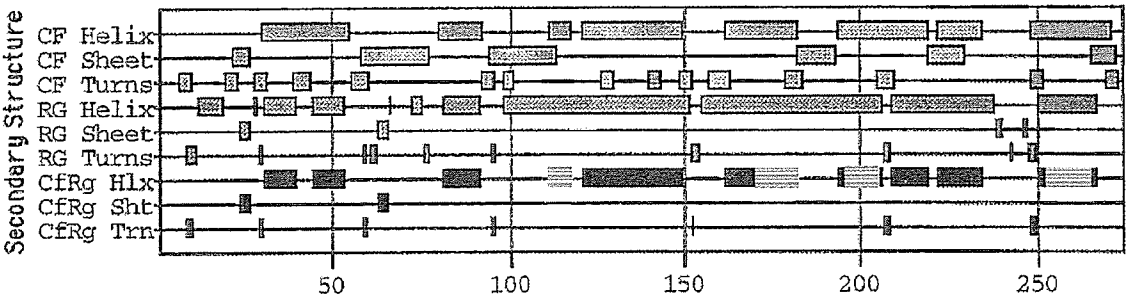


Fig. 16 Secondary structure predictions for the IclR from *E. coli* using both the Robson-Garnier(RG) and Chou-Fasman(CF) methods. CfRg is the consensus derived from both methods. The regions of secondary structure that are predicted in six or more of the above sequences (fig. 15) are highlighted in green. These regions correspond to a-helices running from amino acids 110-118, 170-180, 192-201, and 253-266. The C-terminal helix was predicted in all eight structures.

One result that is of interest to come from the alignments is the fact that the proposed helix-turn-helix motif is not predicted by either the Chou-Fasman nor the Robson Garnier methods. In all eight proteins studied by these methods in fact not one of them was predicted to have a helix-turn-helix, suggesting that this motif is probably undetectable by traditional secondary structure predictions (data not shown). The helix found at the C-terminal end of the protein, however, was conserved in all eight of the proteins suggesting that this helix plays a significant role in the structure and function of these proteins. As shown in figure 17, a typical helix wheel analysis of this proposed region in IclR indicates that this helix would have one face that was extremely hydrophobic. In the Lac repressor a similar situation is observed for the C-terminal amino acids (Chakerian et al. 1991). In this case the helix forms a leucine zipper type structure that allows the molecule to tetramerize. It is tempting to speculate that this helix, if it in fact does exist, may also be responsible for mediating subunit-subunit contacts in the protein.

The fact that an apparent IclR C-terminal deletion mutant is unable to repress the *ace* operon (Sunnarborg et al. 1990) lends support to this proposal.

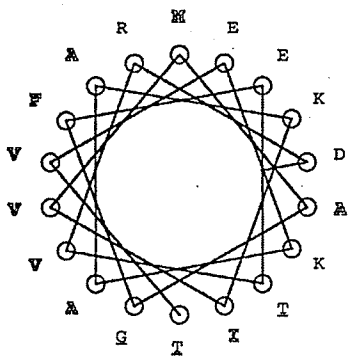


Fig. 17 Helical representation of the predicted C-terminal helix in the IclR repressor from *E. coli*. Notice that one face (left side) of the helix is extremely hydrophobic. This helix was predicted in all eight of the proteins whose primary sequence is similar to that of IclR.

3.5 Equilibrium DNA Binding

Binding of the IclR repressor to its natural binding sites was monitored by an electrophoretic mobility shift assay (EMSA). In this assay various amounts of protein and DNA are mixed together in a pre-defined buffer and allowed to come to equilibrium. Once this equilibrium has been reached the solution is loaded onto a polyacrylamide gel and electrophoresed to resolve the DNA that is bound to the protein from that which is not. As the DNA is radiolabelled with ^{32}P, the position of the DNA in the gel can be determined by autoradiography and the amount of each species can be quantitated using a number of different methods.

At the outset of the work reported here only one naturally occurring IclR binding site was known. This site is located in the promoter region of the *aceBAK* operon (Cortay et al., 1991,1992; Donald et al., 1996) spanning a region of approximately 31bp as defined by DNase protection experiments (fig. 8). The binding of full length, truncated, and mutant proteins was first assayed using this segment of DNA which was excised as a 227 bp restriction fragment upon digestion of the plasmid ptZop.

Shown in fig. 18 are typical autoradiograms depicting the binding of truncated and full length IclR to the *aceBAK* O/P region. To quantitate the binding isotherms the bands corresponding to "bound" and "free" DNA were excised from the gel and quantitated by scintillation counting; then the percentage of DNA that was bound in each lane was plotted against the total concentration of IclR. The data were fitted to the Hill equation (Hill, 1910) as it was observed that the binding did not follow a simple 1:1 Langmuir type isotherm. Shown in figure 19 are the binding data for both the full length and truncated (-8 and -9) IclR proteins. The protein concentration required for half-maximal binding was significantly lower for the full length protein than it was for either truncated protein: for the full length molecule the value was 0.36±0.04 nM while for both truncated proteins it was 40 fold greater, or 14±0.6 nM.

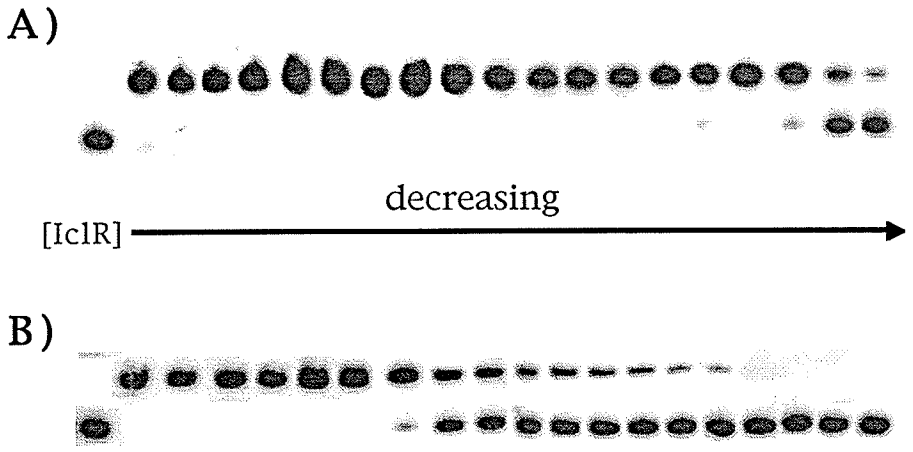


Fig. 18 Autoradiograms showing the binding of A) full length and B) truncated IclR to the *ace* O/P region. The highest and lowest concentration of full length IclR were 1700 and 0.06nM respectively, while for truncated IclR the same values were 1700 and 1.1 nM. Lane 1 contained no protein.

To determine the effect of mutating the first four amino acids in the proposed recognition helix of IclR, DNA binding assays were performed with each mutant protein as described above. Shown in Figure 20 are the Hill plots for each of the five mutants, R9L, N57A, S58A, T59A, and T60A. The curves obtained from these mutants showed half-maximal binding at protein concentrations of 6.0±0.3nM, 2.3±0.2nM, 4.7±0.3nM, and 0.9±0.5 nM for the R9L, N57A, S58A, and T60A mutants, respectively. Thus compared to the wild type protein the mutations weakened the binding by a factor of 16 for the R9L mutant, 13 for the S58A mutation, 6.5 for the N57A mutation, and only a factor of 2.5 for the T60A mutation. The T59A mutant gave the most dramatic effect. As shown in figure 20, even at concentrations >100nM, T59A was not able to bind all of the *aceB* O/P

region. The binding was so weak that the EMSA technique was not suitable to measure the protein concentration required to bind half of the operator.

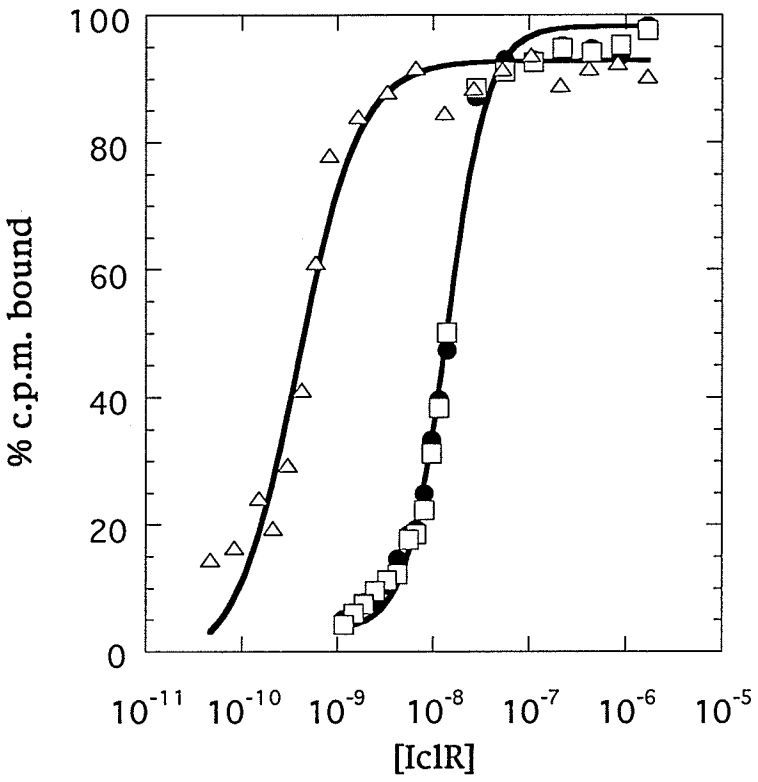


Fig. 19 Hill Plots showing the binding of wild type full length (Δ), truncated-8 (●), and truncated-9(□) IclR proteins to the *ace* O/P region

An interesting feature of the binding isotherms for all of the proteins is the steepness of the transition region, indicating that the binding of IclR to the *aceBAK* promoter is cooperative. Cooperative binding is usually thought to originate from the binding of multiple ligands to a substrate. Binding of the first ligand makes binding of the second either easier (positive cooperativity) or more difficult

(negative cooperativity). In a typical Hill analysis the fractional saturation of the substrate (DNA) is plotted against the free ligand concentration (IclR) and the value of the Hill coefficient (n) indicates the degree of cooperativity. A Hill coefficient of 1 indicates non cooperative ligand binding whereas a value equal to the stoichiometry of binding indicates infinite cooperativity.

The Hill coefficients for the various IclR proteins studied here are shown in table 3 along with the K_Ds reported above and the difference in binding free energies, $\Delta\Delta G$.

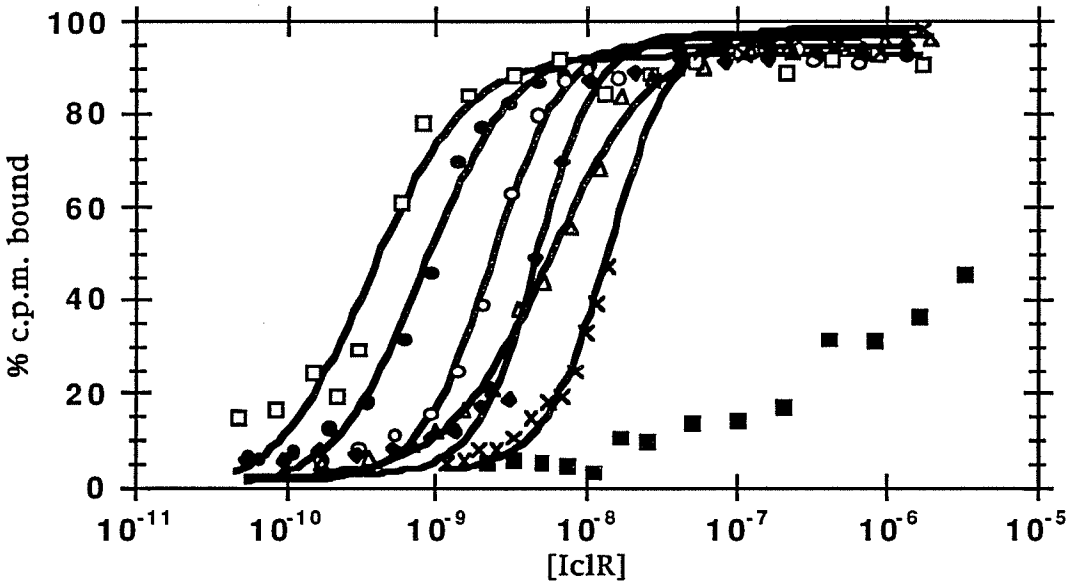


Fig. 20 Hill plots for the IclR mutants R9L (Δ), N57A (O), S58A (♦), T59A (■), and T60A (●). Shown for comparison is the wild type full length (❐) and truncated (×) proteins. The binding data for the T59A protein could not be fit by the Hill equation

Protein	K_D (M)	$\Delta\Delta G$ kcal/mol	n
W.T. (full length)	$(3.6\pm0.4) \times 10^{-10}$	---	1.6±0.2
W.T. (truncated-8)	$(1.32\pm0.05) \times 10^{-8}$	2.14	2.1±0.2
W.T (truncated -9)	$(1.41\pm0.06) \times 10^{-8}$	2.17	2.1±0.2
R9L (full length)	$(6.0\pm0.3) \times 10^{-9}$	1.68	1.5±0.2
N57A (full length)	$(2.3\pm0.2) \times 10^{-9}$	1.09	1.8±0.2
S58A (full length)	$(4.7\pm0.3) \times 10^{-9}$	1.52	1.9±0.3
T59A (full length)	n.d.	n.d.	n.d.
T60A (full length)	$(8.5\pm0.5) \times 10^{-10}$	0.51	1.6±0.1

Table 3 Summary of the binding data for the wild type, truncated, and mutant IclR proteins to the *aceBAK* promoter region. The data were obtained from the Hill plots shown in figures 19 and 20. n.d. - not determined

3.5.1 Effect of Salt

It is commonly found in the study of protein-DNA interactions that the concentration of salt has a large effect on the equilibrium parameters obtained from EMSA experiments. To examine the effect of the potassium chloride concentration on the binding affinity and the Hill coefficient (n), full length IclR was titrated against a constant amount of labeled DNA at KCl concentrations ranging from 2.5mM to 0.4M. Shown in figure 21 are the binding isotherms generated at some of the KCl concentrations performed in this experiment. It can clearly be seen in this plot that the apparent binding affinity decreases as the concentration of KCl increases. In addition, the sharpness of the transition seems to decrease as the salt

concentration is increased, suggesting that the dissociation reaction becomes less cooperative at higher salt concentrations.

Figure 22 summarizes the effect of [KCl] in the form of a log-log plot. It can be seen from this plot that the binding affinity seems to increase slowly as the salt concentration is increased to 50mM, at which time it begins to decrease dramatically. In the range 50mM to 400mM the slope of the plot is 1.7±0.2. According to the method of Record (Record, et al. 1977; Winter and vonHippel, 1981), this result suggests that in this range of salt concentrations about two ions are released when the protein binds to DNA. This value is significantly lower than the value of eight reported for the Lac repressor (Whitson et al. 1989) but it is nearly identical to that of the Gal repressor which also releases two ions upon operator binding (Brenowitz et al. 1990). As ions are generally associated with the phosphate backbone of a DNA molecule, IclR probably interacts strongly with two such groups when it forms a complex with the *ace* promoter region

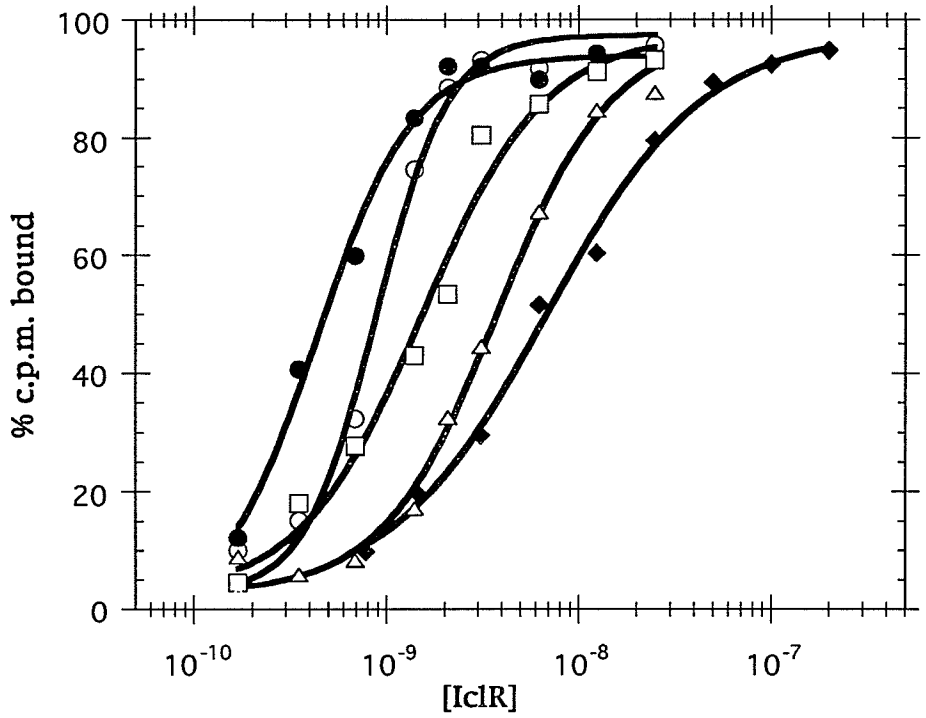


Fig. 21 Illustration of the effect of [KCl] on the binding affinity of IclR for the *aceBAK* operator. The KCl concentrations shown are 2.5mM(O), 50mM(●), 100mM(❐), 200mM(Δ), and 300mM(♦).

Figure 23 depicts the change in the Hill coefficient with an increase in potassium chloride concentration. It is clear from this plot that the cooperativity drops off dramatically as the salt concentration increases. One explanation for this finding is that at the concentrations used for DNA binding IclR exists as an equilibrium mixture of different multimeric forms. As the cooperativity observed in the low salt region may reflect the sequential addition of IclR subunits onto the DNA, the fact that the cooperativity decreases

as the salt is raised may imply that the DNA binding subunit composition of IclR is stabilized by the high salt. It is also possible,

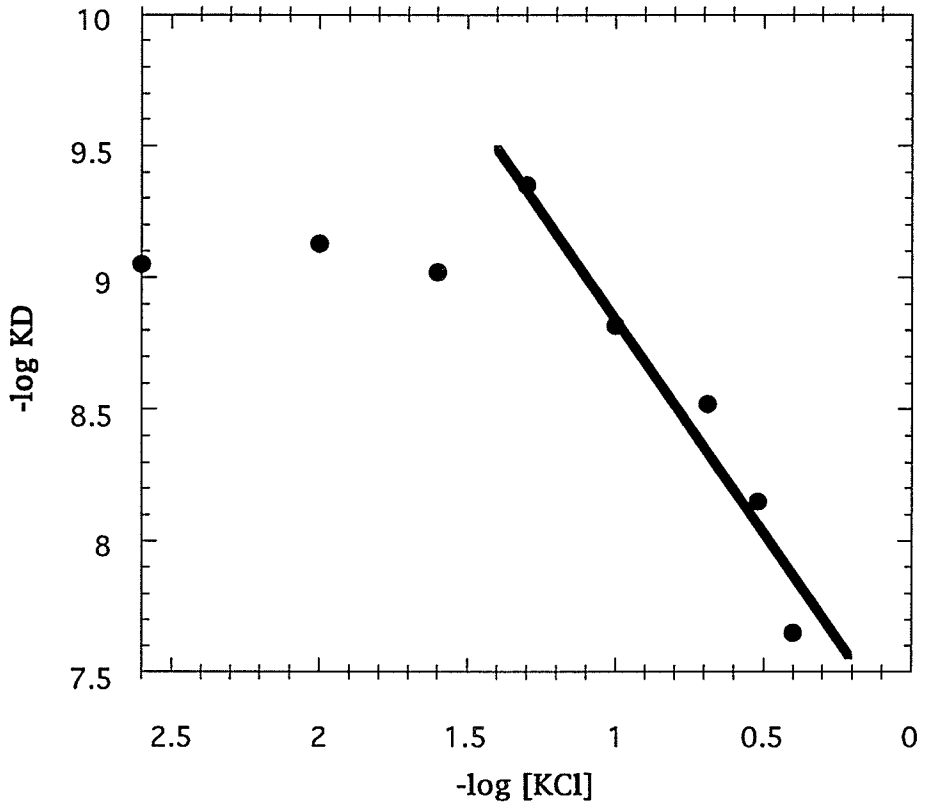


Fig. 22. Log-log plot summarizing the effect of KCl on the binding affinity of IclR for the aceBAK operator. The line shown has a slope of 1.7±0.2 from 50 to 300mM KCl suggesting that about two ions are released upon complex formation.

however, that the cooperativity arises primarily from conformational changes in the DNA. In this model, the high salt concentration might alter the DNA conformation, converting it to an IclR binding form. This is touched upon further in the discussion.

At a constant KCl concentration of 50mM, the effect of varying the $MgCl_2$ was investigated using the EMSA. The binding isotherms generated at $MgCl_2$ concentrations of 0 mM, 5mM, 10mM, and 20mM, are shown in figure 24. It can be seen in this figure that the binding affinity is lowest when $MgCl_2$ is absent (approximately five fold lower), and greatest when $MgCl_2$ is present at concentrations greater than 10mM. This effect should be contrasted to the situation observed with the P22 Mnt repressor (Vershon, et al., 1987). In this

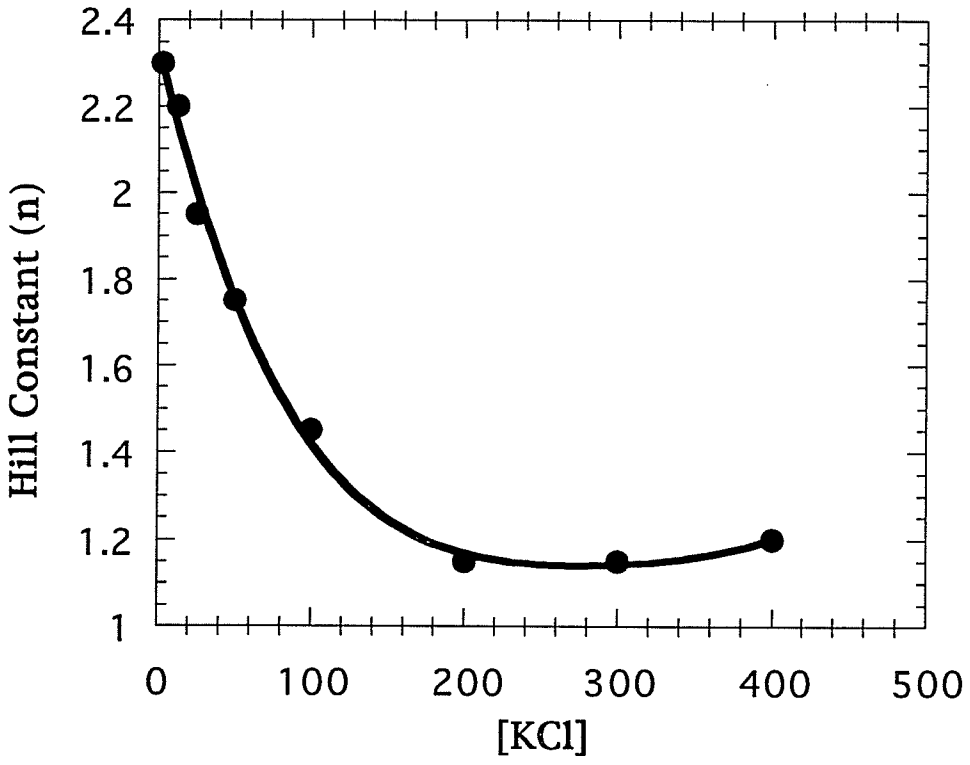


Fig. 23 Effect of [KCl] on the apparent cooperativity of IclR binding to the *aceBAK* operator. The curve has no theoretical significance, it is simply drawn to illustrate the decrease in n that was observed.

system, $MgCl_2$ decreases the affinity of the repressor 25 fold as it increases from 0 to 10mM. The two different situations suggest that in the IclR-DNA complex, some of the specific interactions are mediated by Mg^{2+} ions, while in the P22 Mnt-DNA complex, Mg^{2+} ions compete with the protein for binding sites on the DNA. At all the $MgCl_2$ concentrations tested, the Hill constant, within experimental error, was the same as reported above for full length IclR.

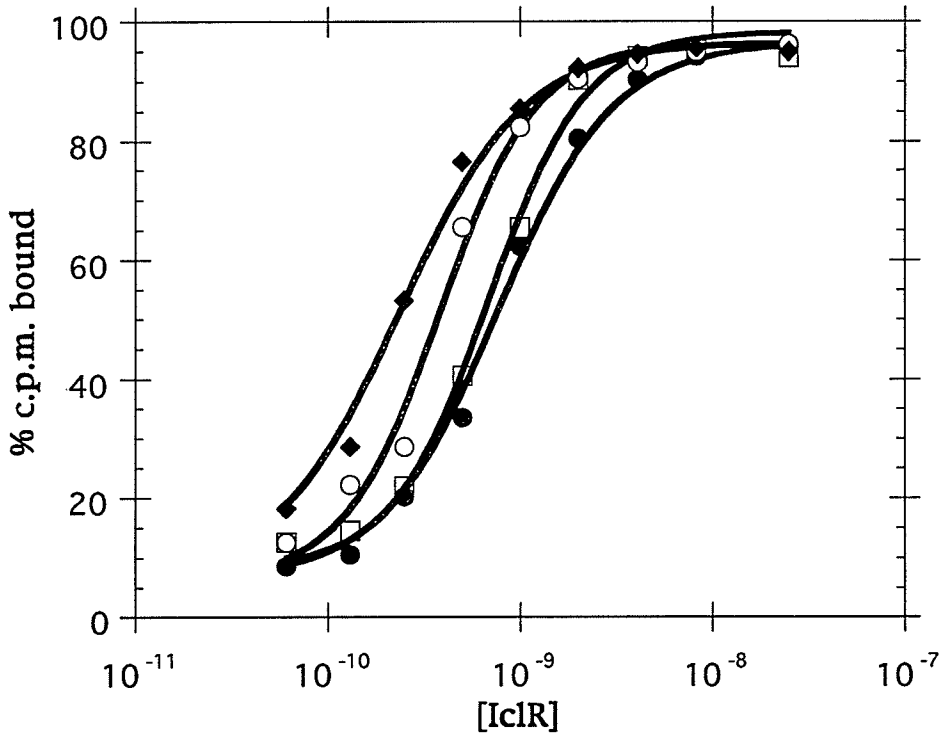


Fig. 24 Binding isotherms generated by IclR binding to the aceBAK operator as a function of [$MgCl_2$]. The $MgCl_2$ concentrations tested were 0(●), 5mM(❐), 10mM(○), and 20mM(♦) In each reaction the KCl concentration was kept constant at 50mM.

3.5.2 Effect of Metabolites on the IclR-DNA Complex

To test the possibility that a small molecule may in some way either activate or repress the DNA binding properties of IclR, a gel shift experiment was performed in the presence of a variety of freshly prepared metabolites from glycolysis, the TCA cycle, and other pathways known to exist in *E. coli*. In the experiment, IclR was added to a constant amount of DNA at a concentration that bound approximately 50% of the operator when no effector molecule was present. A change in affinity, either positive or negative would have indicated that the metabolite was having an effect on IclR binding. None of the compounds tested showed any effect on IclR binding *in vitro*, suggesting that none of these molecules can be an *in vivo* effector molecule of *ace* operon expression. The compounds tested, and the concentrations at which they were added are shown in table 4. The concentrations used were similar to the *in vivo* concentration measured by Lowry et al. (1971). It should be noted that the concentration of the buffer was 20mM in these experiments, thus the metabolites should not have affected the pH of the solution at the concentrations used here.

Of interest is the result that PEP has no effect on the binding affinity of IclR for the *aceBAK* operon *in vitro*. This result should be contrasted with that obtained by Cortay et al. (1991). In that work it was reported that PEP, at a concentration of 4μM was able to inhibit the formation of the IclR-DNA complex. This experiment was

repeated several times in our laboratory and at a concentration of 0.6mM no effect was ever noticed with PEP. As the concentration we used was 150 fold greater than that used by Cortay et al. (1991) we are confident that PEP has no effect on IclR complex formation with the *aceBAK* operon.

Compound Tested	Concentration (mM)
Glucose	5.0
Fructose-1,6-Bisphosphate	7.0
3-Phosphoglyceric Acid	1.0
Phosphoenolpyruvate	0.6
Pyruvate	1.0
Acetate	1.0
Acetyl-CoA	1.3
CoA	1.3
Citrate	3.4
Isocitrate	3.4
α-Ketoglutarate	1.2
Succinate	6.7
Malate	6.7
Oxaloacetate	6.7
Glyoxylate	1.0
NADH	1.0
NAD	1.0
ATP	1.5
ADP	1.0
AMP	1.0

Table 4 Metabolites used in conjunction with IclR in an EMSA with the ace*BAK* operator. None of these compounds had any effect on the formation of the IclR-DNA complex suggesting that none of them were the *in vivo* effector molecule.

3.5.3 Stoichiometry of the IclR-DNA complex

To determine the stoichiometry of the IclR DNA complex several experiments were performed that took advantage of the sensitivity of the EMSA technique in resolving complexes of different sizes. On the plasmid ptZop the *aceBAK* operator can be excised on a fragment of two different sizes depending on the restriction enzymes used to cut it. Digestion with *Eco*RI and *Hind*III yields a fragment 227 bp in size while digestion with *Bam*HI gives a fragment 170 bp in size. In addition, digestion of the plasmid pCHicl with *Xmn*I and *Mlu*I yields a fragment containing the *iclR* operator (to which IclR binds with high affinity, see below), 333 bp in length. In the first experiment full length IclR was incubated with operator DNA of two different sizes at a concentration that bound approximately 95% of the DNA. It was expected that if only one DNA molecule was bound to IclR in the protein-DNA complex, then two discrete bands would be observed: one band for IclR bound to the larger of the two fragments, and one band for IclR with the smaller of two fragments. If IclR was able to bind to two pieces of DNA simultaneously, the pattern of bands expected would be the same as above, but an additional band of intermediate mobility, corresponding to IclR bound to both pieces of DNA, would be observed. As shown in figure 25, when the experiment was performed, the former situation was observed, indicating that only one molecule of DNA was present in the IclR-DNA complex. In light of the fact that IclR is a tetramer at the high micromolar concentrations, this result was somewhat surprising. If IclR is in fact a helix-turn-helix protein, then it might

be expected that the tetramer (or a dimer of dimers) could bind to two fragments of DNA, one with each dimer. This situation would be expected in analogy to that observed for the Lac repressor molecule which can bind to two DNA sites and mediate DNA looping (Mossing and Record, 1986; Borowiec et al., 1987; Kramer et al. 1988). The result obtained thus suggests that at the concentrations used for DNA binding the IclR tetramer dissociates into dimers that can only bind to one operator. Another possibility, however, is that the IclR tetramer interacts with DNA in a novel fashion. The experiments described below favor the former interpretation.

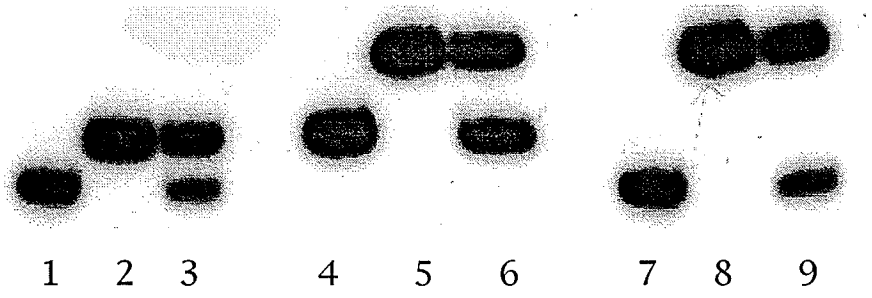


Fig.25 Autoradiogram showing the results from the EMSA using IclR and operator DNA of different lengths. Lanes 1,2, and 3 contain BamHI digested *ace* operator, EcoRI/ HindIII digested *ace* operator, and BamHI and EcoRI/HindIII digested *ace* operator, respectively. Lanes 4, 5, and 6, contain EcoRI/HindIII digested *ace* operator, XmnI/MluI digested *iclR* operator, and EcoRI/HindIII *ace* operator and XmnI/MluI digested *iclR* operator, respectively. Lanes 7, 8, and 9, contain BamHI digested *ace* operator, XmnI/MluI digested iclR operator, and BamHI digested *ace* operator and XmnI/MluI digested iclR operator, respectively. The lack of an intermediate band in the lanes with two operators (3, 6, and 9) indicates that only one DNA molecule is involved in the specific IclR-DNA complex. The free DNA is not shown in this figure as it was run off the end of the gel. The IclR concentration was 1nM in each lane.

To investigate the subunit composition of IclR in the protein-DNA complex full length and truncated (-8) IclR protein were mixed together and denatured in the presence of 5.6M urea. The proteins were then allowed to refold by slowly removing the urea with dialysis. Shown in figure 26 are the fluorescence spectra of the mixture before the addition of urea, after the addition of urea, and after dialysis. The recovery of protein fluorescence suggests that under these conditions a small proportion of the molecules refold. The refolded mixture was then used in gel retardation assays with a synthetic 30bp duplex containing the IclR binding site.

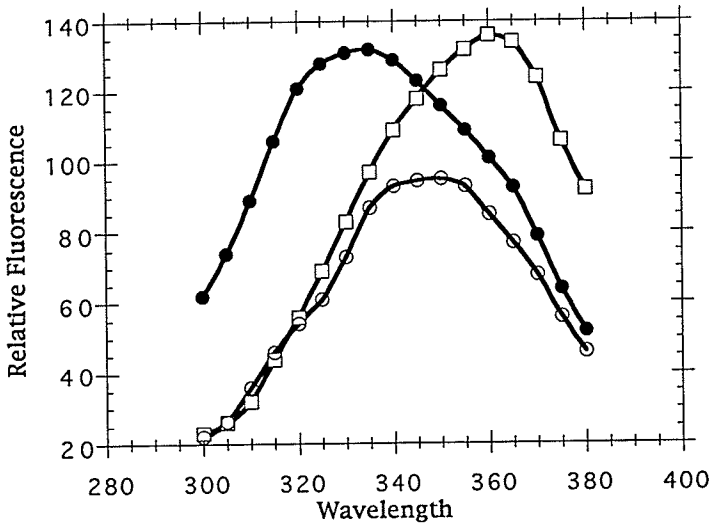


Fig. 26 Fluorescence emission spectra (λ_{ex}=285 nm) of fully folded IclR(●), IclR unfolded in the presence of 5.6M urea(□), and IclR that had been unfolded in urea, and refolded by removing the urea by dialysis(O). The relative fluorescence of the refolded sample is lower than that of the other two as dialysis diluted the protein approximately 1.5 fold.

After electrophoresis the gel was dried and autoradiographed. As shown in lanes 1-4 of figure 27, the mixed sample gave rise to three differentially retarded complexes, with the slowest and fastest

migrating bands comigrating with the bands corresponding to full length and truncated protein, respectively. Although the bands were not quantitated it can be seen that the intermediate band is twice as intense as the outer two bands. The simplest interpretation of this result is that IclR is a dimer when it binds to DNA, and the intermediate band represents the complex formed between a heterodimer containing a full length and truncated subunit. If the protein bound DNA as a tetramer one would have expected the mixed species to show five bands, those corresponding to full length:truncated subunit ratios of 0:4, 1:3, 2:2, 3:1, and 4:0.

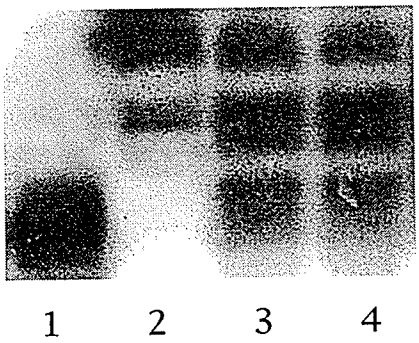


Fig. 27 Autoradiogram showing the pattern of bands observed when full length and truncated IclR subunits are allowed to exchange. Lanes 1, and 2, contained refolded truncated and full length proteins, respectively, at a concentration of 10nM. Lanes 3, and 4 contain equal molar mixtures of refolded truncated and full length IclR at a concentration of 10 nM and 5 nM, respectively. The intermediate band in the last two lanes is presumed to arise from a hetero-oligomeric species consisting of one full length subunit and one truncated subunit. The faint intermediate band seen when just the full length protein is present arises from the small amount of truncated protein (approximately 5%) that is always present in full length IclR preparations.

3.6 Dissociation Kinetics

Dissociation kinetics for the DNA complexes with full length IclR, truncated IclR, and four of the five mutants were performed using EMSA. In a typical dissociation experiment IclR was added to a level such that approximately 90% of the labeled *aceBAK* promoter region was bound to the protein. At time t=0, an excess of unlabeled *aceBAK* promoter (which in preliminary experiments prevented any binding if the labeled and unlabeled DNA were added at the same time) was added to the mixture and at various times a portion of the sample was added to a running polyacrylamide gel. The fact that the gel was running while the samples were loaded was important as it minimized the dead time that the complexes had to sit in the wells of the gel before being resolved. After electrophoresis, the gels were dried, autoradiographed, and quantitated as a function of time. The percentage of bound DNA for the initial part of the dissociation were fitted to a first order rate equation as shown in figure 28. The kinetic parameters derived from these plots are listed in table 5.

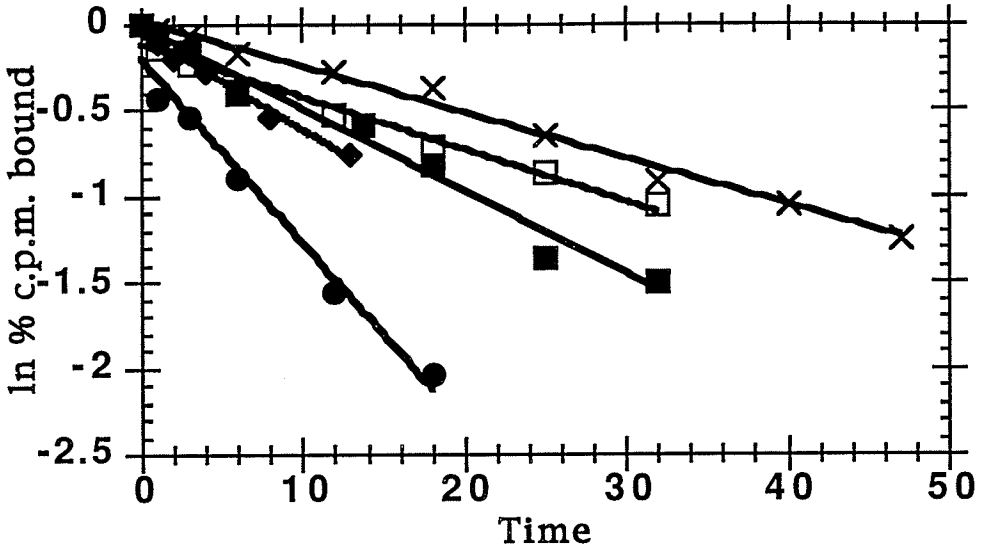


Fig. 28 Dissociation kinetics of full length(×), truncated(♦), N57A(■), S58A(●), and T60A(❐) IclR proteins from the 227 b.p. *ace* operator containing DNA fragment. The T59A mutant did not form stable complexes thus preventing the determination of it's kinetic parameters.

Protein	k_d (min-1)	$t_{1/2}$ (min)
W.T. (full length)	0.027±0.001	26±1
W.T. (truncated)	0.056±0.001	12±1
N57A (full length)	0.048±0.003	14±1
S58A (full length)	0.106±0.009	6.5±0.5
T60A (full length)	0.030±0.002	23±2

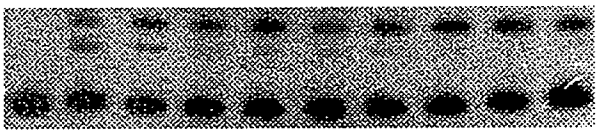
Table 5 First order dissociation parameters extracted from the graph shown in fig. 28. The truncated protein was missing the first eight amino acids

The kinetic data almost exactly parallel the equilibrium data presented above (see table 3). The full length protein, which has the highest affinity for the *aceBAK* promoter in equilibrium experiments also forms the longest lived complex, with a half life for dissociation of 26±1 minutes. The T60A protein, whose equilibrium binding affinity is closest to that of wild type shows the next longest half life of 23±2 minutes. The mutant with the lowest affinity, S58A, shows the shortest half life, 6.5±0.5 minutes. The only exception to the pattern is the observation that the truncated protein, which had an equilibrium affinity two to three times less than S58A, formed a complex with a half life of 12±1 minutes, a value roughly twice that obtained for S58A. This result suggests that although the truncation weakens the stability of the complex (roughly two fold according to the half lives) the N-terminal arm probably functions in DNA binding primarily by increasing the kinetics of association. This is commented on further in the next section.

3.7 Association Kinetics

The association kinetics of the IclR-DNA complex could not be measured precisely using the EMSA method described above. The reaction kinetics were so fast that at best an estimate of the lower limit was available. Shown in figure 29 is an autoradiogram of a typical association experiment. In this figure the first six lanes were loaded onto the running gel at five second intervals and it can be seen that the amount of complex does not increase with an increase in time after this point. Thus it is probably a reasonable estimate to

assume that equilibrium is reached in 30 seconds upon mixing full length IclR with its operator. At the concentrations used in this experiment (IclR = 0.5nM, *ace* operator = 5pM), and assuming that 30% of the DNA was bound when equilibrium was reached (30 seconds), the first order rate constant, k_a, is calculated to be 2.4 X 10^7 s^{-1} M^{-1}. It should be noted that the ratio of the association constant to dissociation constant (i.e. the equilibrium constant) is worked out to be approximately 2.0 X 10^{-11} M^{-1}. Although the association rate constant was only estimated, it could not be any slower, thus the disagreement of the two results verify the equilibrium results; that is the binding reaction is not simply described by a Langmuir type isotherm.



1 2 3 4 5 6 7 8 9 10

Fig 29. Autoradiogram of a typical association kinetic experiment using IclR at a concentration of 0.3 nM and *ace* operator at a concentration of approximately 5.0 pM. Lane 1 had no protein, while complexes observed in lanes 2-6 were loaded onto the running gel at time points corresponding to 5, 10, 15, 20, 25, and, respectively. Complexes observed in lanes 7, 8, 9, and 10 were loaded onto the gel at 60, 90, 120, and 180s, respectively. The amount of bound complex does not increase after 25 s indicating that the reaction has probably come to completion by this time. The faint intermediate band is only observed in the early time points.

In addition to providing a lower limit estimate of k_a, the association experiments revealed the transitory presence of a complex that was smaller than that observed in equilibrium

experiments (fig 29 lanes, 1-7). This band disappeared over time suggesting that it was a kinetic intermediate along the reaction pathway. Since the equilibrium complex probably contains the IclR dimer it is tempting to assume that the kinetic intermediate observed in these experiments is an IclR monomer bound to a half site on the DNA molecule of interest. This in turn implies that at the concentrations used for DNA binding IclR probably exists as a mixture of monomers and dimers.

The association kinetics for the truncated protein proved impossible to acquire using the EMSA technique, even though the dissociation kinetics suggested that the rate would be slower. To get the protein to bind to DNA a concentration of approximately 15nM IclR was required. It was found that at this concentration the rate of association was too fast to be even estimated, as the amount of bound DNA seemed to reach equilibrium even at the first time point (5 seconds).

3.8 IclR binding to the *iclR* O/P region

Very recently it has been shown that the *iclR* gene is autoregulated: that is the IclR protein controls the rate at which it is synthesized (Gui et al. 1996). DNase analysis of this region indicated that the IclR protein can specifically interact with a region of DNA 5' of the *iclR* gene that is similar, but not identical to the IclR target region 5' of the *aceBAK* operon. To try to gain some insight into the importance of the bases involved in IclR binding, this region of DNA

was isolated by a *MluI/XmnI* digestion of the plasmid pCHicl. The sequence of the region protected, compared with the region protected in the *aceB* operator region, is shown below (fig.30)

```
          1         10        20        30
aceB      TAATTAAAATGGAAATTGTTTTTGATTTTGCATT
iclR      CAATAAAAATGAAAATGATTTCCACGATACAGAA
```

Fig. 30 Comparison of the IclR binding sites at the *aceB* and *iclR* O/P regions. Bases that are conserved among the two sequences are highlighted. No clear indications of a palindromic structure are seen when looking at the highlighted bases.

It can be seen from the figure that several bases are conserved in the two sequences, however, no clear indication of a palindromic sequence is obviously apparent (as would be expected for a helix-turn-helix protein). The binding isotherm generated for the binding of full length IclR to the *iclR* O/P region is shown in figure 31 The dissociation constant of 0.9±0.1 nM is approximately 2.6 fold weaker than that obtained for the *aceBAK* operator. Interestingly, the Hill coefficient for this site, n=1.6±0.2 is identical to that obtained for the *aceBAK* promoter.

3.9 Defining an IclR Target Sequence

To define the mechanism of action of a DNA binding protein it is important to understand at the molecular level both the protein side chains and the DNA bases important for mediating the

interaction. The experiments described above have demonstrated the importance of some of the protein groups involved in this process, however, the identities of the individual DNA bases important for recognition remain elusive.

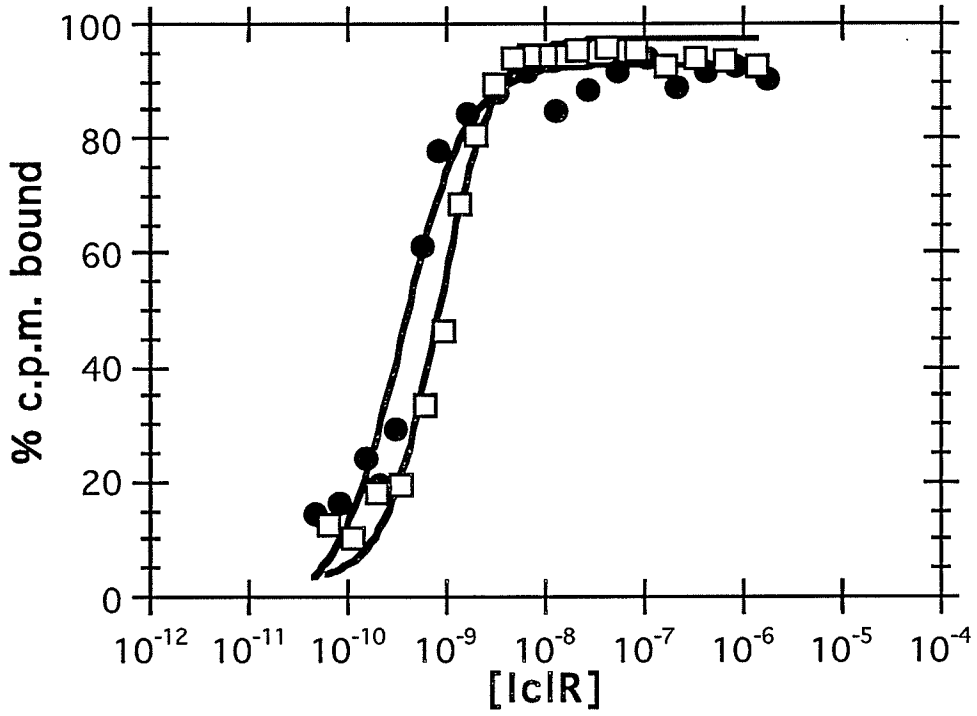


Fig. 31 Hill plot showing the binding curve of full length IclR to the *iclR* promoter region(□). Shown for comparison is the curve obtained with the same protein and the *ace* operator region(●) which binds with approximately 2.6 fold greater affinity.

In figure 30 the sequences protected from DNase digestion by IclR at the *aceBAK* and *iclR* operons are compared. An interesting feature of these regions is the high content of AT base pairs. Classically, base specific protein-DNA interactions have been probed

by studying the protection and interference patterns that a protein affords to the guanine residues within its binding site. However, as stated above, the small number of guanine residues in the IclR binding site is only able to provide limited amounts of information. For example Cortay et al. (1992) were able to demonstrate using interference techniques that the two guanines at positions eleven and twelve (as numbered in fig 30) were vital for allowing the formation of a stable complex. In these experiments, pre-methylation of the N7 position of the guanine bases prevented formation of the IclR-DNA complex. Interestingly, however, only one of these two guanines is present in the *iclR* binding site, indicating that the protein must somehow be able to tolerate this change. In the process of this work further characterization of the IclR binding site at the *aceBAK* operon was attempted by employing methylation protection experiments. IclR was first incubated with the *ace* promoter prior to the addition of the methylating agent, dimethyl sulfate. Dimethyl sulfate is known to methylate double stranded DNA at the N7 position of guanine, and also at the N3 position of adenine, albeit at a rate 10 fold slower. Thus a protection experiment operates under the premise that the N7 position of guanine (in the major groove), and the N3 position of adenine (in the minor groove) will not be methylated if they are interacting with the protein. Upon cleavage of the DNA at the methylated residues, a typical sequencing gel analysis reveals positions protected by the protein.

For the IclR protein-DNA complex, in my hands, using IclR at a concentrations up to 100nM, no protection was observed at any guanine or adenine residues on either strand (data not shown). This result was surprising as the result obtained by Cortay et al. (1992) suggested that the two guanine residues at positions 10 and 11 (as labelled in fig. 30) were critical for allowing complex formation. A possible explanation for this observation is that these two residues function primarily in assisting the association of the complex, rather than in stabilizing a complex already pre-formed. The fact that no protection of adenine residues was observed in these experiments suggests that IclR does not make strong contacts to the minor groove of the operator. The lack of sequence specific information obtained from these protection experiments required that a different approach be taken to determine the bases important for the IclR-DNA interaction.

Theoretically an excellent approach for studying the importance of the bases in a DNA binding site is to synthesize chemically a number of oligonucleotides containing a number of base replacements at the sites thought to be important for function. This approach has been successfully employed in studying the interaction of the Lac repressor with its operators (Lehming et al. 1987; Sasmor and Betz 1990). In that system, however, a good understanding as to the importance of each base had already been obtained from footprinting and interference experiments. In the case of the IclR binding site very little information was known and a study of this type would have involved the synthesis of a number of

oligonucleotides spanning the 31 or so base pair region protected by DNase. This would have not only been a time consuming, laborious process, it would have also been prohibitively expensive. Instead, a method was devised that took advantage of the ability of IclR to select tight binding sequences from a random pool of oligonucleotides.

3.10 SELEX

SELEX, or Systematic Evolution of Ligands by Exponential enrichment, was initially developed as a method for screening huge combinatorial libraries, consisting of $>10^{15}$ RNA molecules, for molecules of predefined catalytic activity (Tuerk and Gold, 1990; Robertson and Joyce, 1990). Since then the method has evolved to screening large pools of double stranded DNA for the ability to bind to a protein (Thiesen and Bach, 1990; Norby et al., 1992; Toledano et al., 1994; Ades and Sauer, 1994; Cui et al., 1995; and He et al., 1996). In its simplest sense the method is conceptually straightforward: in a DNA synthesizer a completely random sequence consisting of equal amounts of all four bases in all positions is generated and flanked by two defined primer binding sites. This generates an enormous pool of sequences (4^n, where n is the number of randomized positions) which are then made double stranded by the action of a DNA polymerase. DNA molecules that bind to a particular protein are selected by a method such as affinity chromatography, filter binding, or electrophoretic mobility. As the starting pool is often very complex, several purification steps are usually required to "fish

out" functional molecules. Therefore, the small percentage of molecules selected at each step is amplified in a polymerase chain reaction. The process is then repeated with these amplified molecules and continued until the high affinity ligands dominate the population. A schematic illustration of this procedure is shown in figure 32.

For IclR this protocol was followed using both EMSA and nitrocellulose binding to partition the bound and free DNA molecules. The random oligonucleotide, SELEX-A, which is 60 base pairs in length, contains two primer binding sites which flank 24 randomized bases in the middle. To facilitate cloning of the selected sequences, the primer binding sites were designed to contain an EcoRI site at the 5' end and a BamHI site at the 3' end. For the first round of selection SELEX-A was made double stranded in a standard primer extension reaction using the primer SELEX-B. This generated a possible 4^{24} (3×10^{14}) different sequences. Approximately 1μg of this DNA (approximately 1.7×10^{13} molecules) was then incubated with full length IclR protein in gel shift buffer, and molecules which bound to the protein were selected by EMSA. To determine the position at which any protein-DNA complex migrated a 60 base pair oligonucleotide (SELEX-D) containing the natural *aceBAK* binding site was synthesized, made double stranded, and labeled with ^{32}P. Thus, the region of the gel in which the SELEX-D-DNA complex migrated

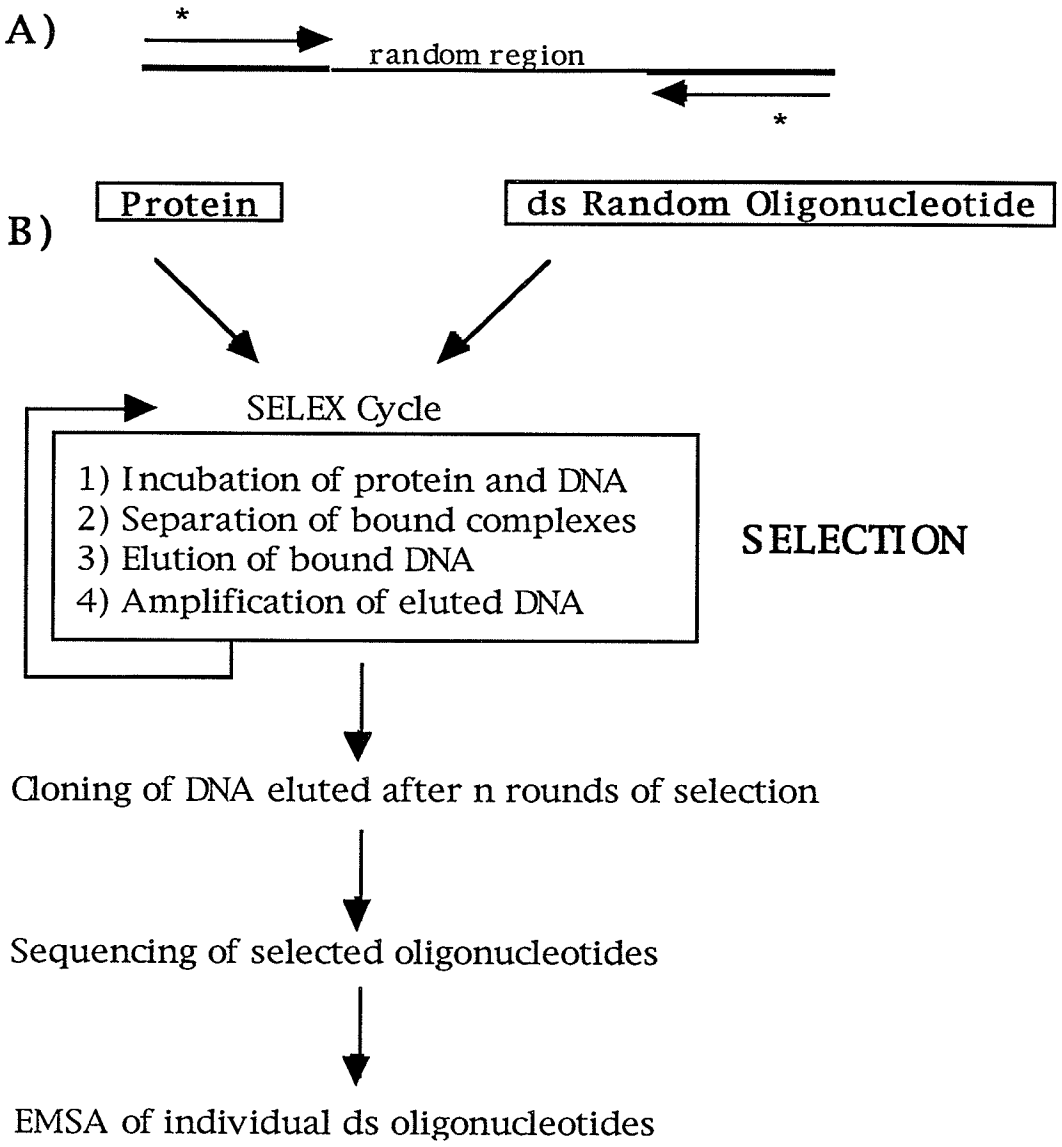


Fig. 32 Schematic illustration of the SELEX procedure. In A, the random oligonucleotide is first made double stranded by the action of DNA polymerase and one of the PCR primers (arrows). The PCR primers and oligonucleotides are designed to carry restriction enzyme recognition sites (*) in order to facilitate cloning at a later stage In B, the selection stage, protein is mixed with randomized double stranded oligonucleotides and DNA which binds with high affinity is selected. After elution of the bound DNA from the protein the DNA is amplified by PCR and the process is repeated for n cycles. At the completion of n cycles the selected DNA is cloned, sequenced and tested for its ability to interact with the protein.

was the region containing SELEX-A-DNA complexes, which was excised from the gel. After elution and PCR of the DNA that bound to IclR, the process was repeated for three more cycles, each time using less DNA and less protein.

At the conclusion of the first four cycles the selected DNA was labeled internally and used as a binding probe to check its ability to bind to IclR. Unfortunately, the DNA selected did not bind with even moderate affinity to the IclR protein at this stage. A mathematical analysis of the SELEX procedure (Irvine et al., 1991) revealed a possible reason for this poor result. When performing SELEX one has to be very careful to limit the amount of non-specific "background" that partitions along with the molecules binding with high affinity to the protein. This background can be acquired from a number of sources, including non-specific partitioning of the DNA with the protein, and contamination of PCR reactions by randomized DNA that can be introduced throughout the many purification steps. As the PCR will amplify all the molecules introduced into the reaction, this non-specific background must be carefully controlled.

In an attempt to alleviate the above described problem three modifications to the original process were employed. The first modification involved irradiating all of the solutions (except for the primers and Taq polymerase) used in the PCR reaction with long wave UV light. The second modification involved partitioning the bound and free DNA with nitrocellulose filters. These changes were expected to reduce the amount of contamination as the bound DNA

was simply eluted from the filters in high salt and directly introduced into a PCR solution that had just been irradiated with UV light. Using the EMSA method the DNA had a much greater chance of becoming contaminated as the DNA had to be extracted from the gel, ethanol precipitated twice, and gel purified before it could be used in the next round of SELEX The third modification simply involved raising the IclR concentrations used in each round of selection. Keeping the IclR at elevated concentrations allowed for the introduction of a greater amount of selected DNA into the PCR reaction. This should have had the effect of reducing the amount of contamination, as the starting concentrations of specific molecules would vastly outnumber the contaminating background. The combination of these modifications improved the results dramatically.

After eight rounds of SELEX the selected DNA was assayed for its ability not only to bind to IclR, but also to compete with the naturally occurring *aceBAK* promoter region. Shown in figure 32 is the result obtained from the competition experiments using the selected DNA from rounds 5, and 10. (round 5 was the first round that employed the nitrocellulose filters). The DNA selected in round 5 was only mildly effective in competing with the natural promoter, and only at the highest concentrations used. In contrast, the DNA selected from round 10 was completely effective as a competitor at the highest concentration, and was still an effective competitor when it had been diluted up to 500 fold.

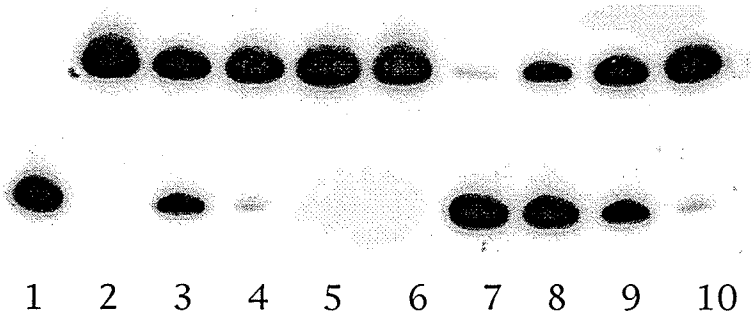
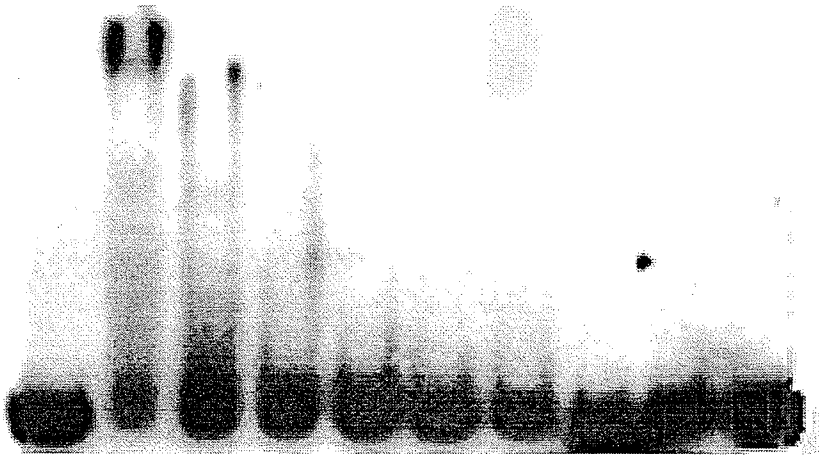


Fig. 32B The ability of the SELEX selected DNA to compete with the naturally occurring *aceBAK* operator was monitored by a competition experiment. Lanes 3-6 contained SELEX DNA from round five at a concentration of approximately 1μM, 200nM, 50nM, and 10nM, respectively. Lanes 7-10 contained SELEX selected DNA from round 10 at concentrations of 500nM, 100nM, 10nM, and 1nM, respectively. Lane 1 contained no IclR, and lane 2 contained no competitor. In each lane the IclR concentration was 3nM.

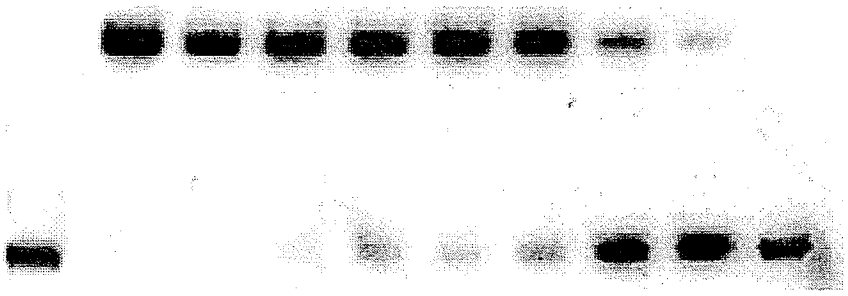
To verify this result, the DNA from a number of rounds was labeled internally and its affinity for IclR was monitored by an EMSA. The results shown in figure 33 clearly demonstrate that the affinity of the selected DNA had dramatically increased. The DNA used in round 1 was unable to form a stable 1:1 complex even at protein concentrations of 5.6 μM. In fact, only a smear was observed at this concentration, suggesting that most of the binding was non-specific. The DNA selected in round 10, however, showed binding even at an IclR concentration of 1.4 nM, indicating that the binding affinity had improved at least 5000 fold. Interestingly, analysis of the DNA selected from rounds 11 and 12 showed a decrease in binding affinity by a factor of approximately ten fold. As described above, it is possible that the lower concentrations of IclR used in these rounds were not sufficient to provide enough DNA to the PCR reactions, allowing for these samples to become contaminated.

A)



[IclR]nM 0 5600 1800 620 210 69 35 17 9 4

B)



[IclR]nM 0 280 175 44 22 11 5.5 2.8 1.4 0.7

Fig. 32C Illustration of the increase in binding affinity for the DNA selected after 10 rounds of SELEX (fig. B) as opposed to DNA that was used in round 1 (fig.A). The IclR concentrations used in each lane are listed in the figure.

3.10.1 Sequence analysis of IclR Selected Binding Sites

To analyze the sequences to which IclR bound the DNA selected from round 10 of the SELEX procedure was cut with BamHI and EcoRI and ligated into the replicative form of phage M13mp18 DNA that had been cut with the same enzymes. After transformation and plating onto IPTG/XGal LB plates, white plaques were picked and used to prepare both RF and single stranded DNA for sequencing. The nucleotide sequences of 32 such isolates are shown in figure 34.

The sequences can be divided into three groups. In the first group the selected sequences almost all contain the palindromic sequence **GGAAANNNTTTCC.** When the selected oligonucleotides are aligned around this motif a consensus half site for IclR can be defined as **(AAWDTGGAAAV)N**, where W= A or T, D = not C, and V = A or C (fig. 34). It should be noted that it was not possible to distinguish the orientation in which IclR bound, therefore both strands had to be included in the above analysis. Unfortunately, this precluded any observation of possible asymmetries in DNA sequence that may have been selected for by IclR binding. In the second group, only one half of the consensus palindrome is found suggesting that IclR can bind to these sequences with at least moderate affinity. In the third group the isolates do not seem to contain any of the sequence elements required for IclR binding. As suggested by Irvine et al. (1991) these sequences probably represent molecules that are carried over from the contaminating background or from the huge number of weak binding sequences present in the original pool.

A)

FULL SITES:

```
2-B                    TGCACAGAATTGGAAACAGTTTCCATTCTGTGCATTGATCCCGGG
2-T          CCCGGGATCAATGCACAGAATGGAAACAGTTTCCATTCTGTGCA
3-B                      ATCCAATGCGGAAAGTATTTCCAAACTTAGCATAATTGAATTC
3-T           AATTCAATTATGCTAAGTTTGGAAATACTTTCCGCATTGGAT
5-B                      ATCCAATGCGGAAATGGTTTCCATTTATATACCAATTGAATTC
5-T           AATTCAATTGGTATATAAATGGAAACCATTTCCGCATTGGAT
4-B             ATCCAATGCACCACCTCTGGAAACGATTTCCATTATGTACCAATTGAATTC
4-T            AATTCAATTGGTACATAATGGAAATCGTTTCCAGAGGTGGTGCATTGGAT
6-B             ATCCAATGCGCCAAGACTGGANACNNTTTCCATATTGNN
6-T                      NNCAATATGGAAANNGTNTCCAGTCTTGGCGCATTGGAT
8-B                 ATCCAATGCAAATTGGAAACCGTTTCCAAAGCCATCCCAATTGAATTC
8-T          GAATTCAATTGGGATGGCTTTGGAAACGGTTTCCAATTTGCATTGGAT
11-B                     ATCCAATGCGGAAGACTTTTCCCCATGGATAGGAATTGAATTC
11-T         GAATTCAATTCCTATCCATGGGGAAAAGTCTTCCGCATTGGAT
12-B                     ATCCAATGCGGAAATGGTTTCCATTTATATACCAATTGAATTC
12-T         GAATTCAATTGGTATATAAATGGAAACCATTTCCGCATTGGAT
16-T                  CGGAATTCAATTGGAAANNGTTTCCATTGTCTGGTTGCA
16-B                TGAAACCAGACAATGGAAACNNTTTCCAATTGAATTCCG
22-B              ATCCAATGCACAGAAAGGAAACAGTTTCCNCCCAATTGAATTC
22-T                GAATTCAATTGGGNGGAAACTGTTTCCTTTCTGTGCATTGGAT
25-B               GGATCCAATGCCAATGGAAATTCTTTCCATTCCCCAATTGAATTC
25-T             GAATTCAATTGGGGAATGGAAAGAATTTCCATTGGCATTGGATCC
26-B       GGATCCAATGCCACAACAAAATGGAAATCCTTTCCAATTGAATTC
26-T                    GAATTCAATTGGAAAGGATTTCCATTTTGTTGTGGCATTGGATCC
27-B       GGATCCAATGCCAACACAGAATGGAACCTATTTCCAATTGAATT
27-T                    GAATTCAATTGGAAATAGGTTCCATTCTGTTTGGCATTGGATCC
28-B                   GGATCCAATGCGGAAATGGTTTCCATTTATATACCAATTGAATTC
28-T         GAATTCAATTGGTATATAAATGGAAACCATTTCCGCATTGGATCC
29-B                   GGATCCAATGCGAAAACCGTTTCCGTATCCATGGAATTGAATTC
29-T          GAATTCAATTCCATGGATACGGAAACGGTTTTCGCATTGGATCC
30-B                   GGATCCAATGCGGAAATGGTTTCCATTTATATACCAATTGAATTC
30-T         GAATTCAATTGGTATATAAATGGAAACCATTTCCGCATTGGATCC
31-B                   GGATCCAATGCGGAAAACGTTTCCAAAGCCATCCCAATTGAATTC
31-T         GAATTCAATTGGGATGGCTTTGGAAACGTTTTCCGCATTGGATCC
32-B       GGATCCAATGCACCTAGCAAATGGAAACGATTTCCACTTTCCCTCGAATTGAATTC
32-T        GAATTCAATTCGAGGGAAAGTGGAAATCGTTTCCATTTGCTAGGTGCATTGGATCC
33-B       GGATCCAATGCACCTAGCAAATGGAAACGCTTTCCAATTGAATTC
33-T                    GAATTCAATTGGAAAGCGTTTCCATTTGCTAGGTGCATTGGATCC
34-B     GGATCCAATGCGCGGCCAGCCGTGGAAATAATTTCCGATACTAATTGAATTC
34-T              GAATTCAATTAGTATCGGAAATTATTTCCACGGCTGGCCGCGCATTGGATCC
35-B         GGATCCAATGCGCCAAGACTGGAAACTGTTTCCATAATTGAATTC
35-T                  GAATTCAATTATGGAAACAGTTTCCAGTCTTGGCGCATTGGATCC
```

SINGLE SITES:

```
7-T                 CAATTCCGCCTGGTGGAAACTAGGTATGTGCATTGGAT
10-T          GAATTCAATTCCGTATCGTGGAAATACTCGGTGCGCATTGG
14-T          GAATTCAATTCCTAGATGNGGAAATTGGTGTGGTGCATTGG
20-T          GAATTCAATTCCGTATNGTGGAAATACTCGGTGNGCATTGG
21-T                NNAGGACTTTGGAAACGGGGTTGGCATTGG
23-T          GAATTCAATTCCATAGGGTGGAAATTGTTATTCAGCATTGG
18-T         GAATTCCAATTCCATCAGGTGGAAAATAAGCCTCGGCATTGG
```

NO CONSENSUS:

```
19-T        GAATTCAATTGGCCTATATATGATGCTGTGCCTAGCATTGG
1-B         CCAATGCTGCCGATACTGGGGTGCCCTCTAAAATTGAATTC
9-B         CCAATGCGACCCGTTAACCCCAATCTGATGCAATT
36-B        GGATCCAATGCGGCCTCCCACTCCACCGCACTCGGAATTGAATTC
```

B)

	1	2	3	4	5	6	7	8	9	10	11	12	13
A	26	51	58	56	42	3	0	2	100	100	95	7	19
C	40	21	14	2	7	23	0	0	0	0	2.5	53	29
G	21	15	20	2	29	3	100	98	0	0	2.5	10	29
T	13	12	7	40	22	69	0	0	0	0	0	30	14
	N	A	A	A T	A G T	T	G	G	A	A	A	C T	N

Fig. 33 Nucleotide sequences of DNAs selected and cloned after 10 rounds of the SELEX procedure. The sequences can be grouped into three groups: The first group is the largest and contains two palindromic half sites with the sequence TGGAAA. The second group contains DNA with only one half TGGAAA site and the third group bears no resemblance to first two groups. In the first two groups the sequences were aligned around the conserved half sites. In some sequences it was found that more than 24 bases existed in the random region. This was presumably due to errors made by the Taq polymerase throughout the procedure. Nucleotides designated N could not be unambiguously determined from the sequencing reactions. B) The percentage of times that each base was found at positions through the half site is shown with the bases that are most strongly conserved shown below. The absolutely conserved GGAAA sequence is underlined.

The results of this experiment thus demonstrate that IclR binds preferentially to a consensus palindrome that extends over 23 bp. In

contrast to other experiments that only show the importance of a few nucleotides, this experiment was able to define a binding site signature for the IclR protein. As IclR was only known to bind to two naturally occurring sites at the completion of this experiment this consensus sequence could not have been determined by a direct comparison of the bases in both operators. The SELEX method thus proved to be an extremely powerful tool in our study of IclR-DNA interactions.

3.10.2 Determination of the Affinity of IclR for the selected DNAs

The affinity of IclR for representatives of the selected DNAs was measured using an EMSA. The DNA samples for the assays were obtained by digestion of the M13mp18 RFs of the selected clones with *Eco*RI and *Hind*III, and were labeled with ^{32}P. Selected for analysis were clones 7 and 23, corresponding to half sites; clones 8, 25, and 26; corresponding to full sites; and clone 36, corresponding to a site with no homology to the consensus. The results of these experiments are presented in figure 34. The data for the full sites are similar to that obtained previously for the natural *aceBAK* operator. Clone 25 seems to have the highest affinity for the protein as the protein concentration required for half-maximal binding was 0.5±0.1nM. Clones 8 and 26 bind roughly four fold more weakly with affinities of 1.8±0.2nM, and 1.9±0.2nM, respectively. In all three cases the Hill coefficient 1.4±0.2, is somewhat lower than that for the natural operator. For the half site clones, 7 and 23, the

binding is clearly weaker as evidenced not only by the drop in affinity (6.7±0.3, 10.3±0.4nM, respectively) but also by the lack of cooperativity. The Hill constants for these sites are actually less than unity with a value

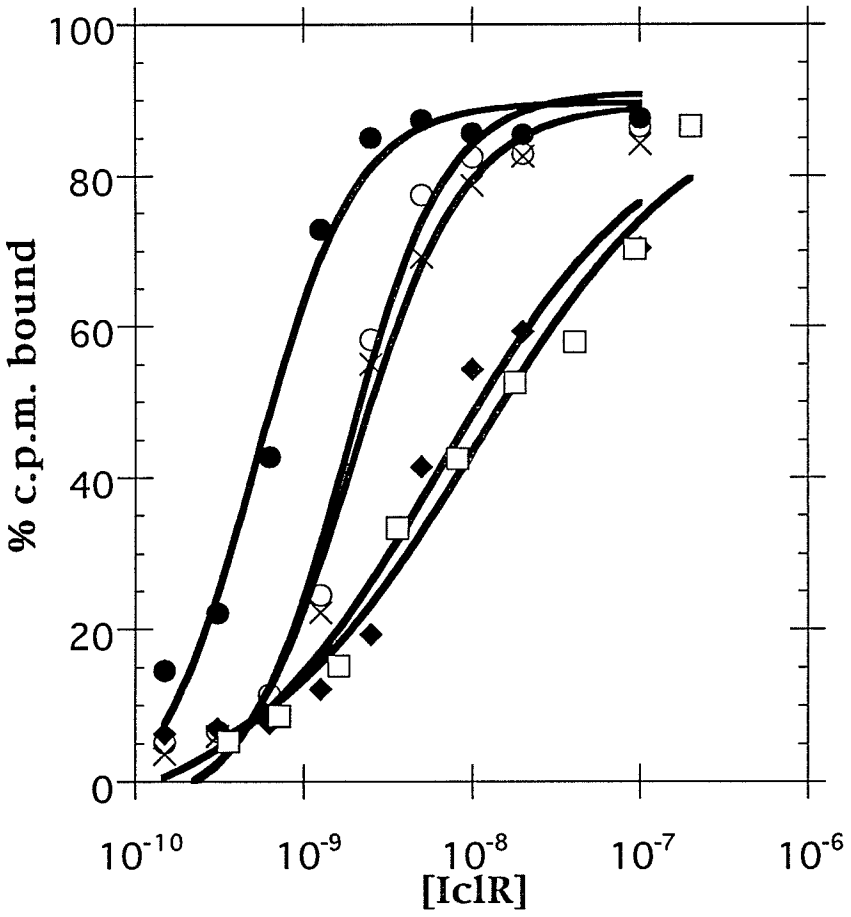


Fig. 34 Hill plots showing the binding of the SELEX clones 7(♦), 8(O), 23(□), 25(●), and 26(×). Clones 7, and 23 correspond to half sites while 8, 25, and 26 correspond to full sites.

of 0.7±0.1 for both clones. The loss of cooperativity observed when binding to half sites has also been observed for the P22 Arc repressor (Brown and Sauer, 1993). Isolate 36, the clone bearing no

homology to the consensus sequence showed almost no affinity to IclR, and quantitation by an EMSA was not possible. A summary of the data is presented in table 6

Clone	Sequence	K_D (nM)	Hill Constant
8	AAATT**GGAAA**CCG**TTTCC**AAAGC	1.8±0.2	1.4±0.2.
25	CCAAT**GGAAA**TTC**TTTCC**ATTCC	0.5±0.1	1.4±0.2.
26	AAAAT**GGAAA**TCC**TTTCC**AATTG	1.9±0.2	1.4±0.2.
7	CCTGGT**GGAAA**CTAGGTATGTGC	6.7±0.3	0.7±0.1
23	TAGGGT**GGAAA**TTGTTATTCAGC	10.3±0.4	0.7±0.1
36	AATGCGGCCTCCCACTCCACCGCA	n.d.	n.d.

Table 6 Summary of the binding properties of the SELEX clones studied in this work. The length of sequence shown roughly corresponds to the bases protected from DNase when IclR is bound to either the *ace* or *iclR* operators

3.11 Discovery of a Cryptic IclR Binding Site

Using the consensus IclR binding sequence as a template the *E. coli* genomic database was searched for additional binding sites using the on-line Fasta search engine. Two possible binding sites were discovered. The first site was located within the coding region of the *hisF* gene, a gene required for the biosynthesis of the amino acid histidine. The second site was located in the promoter (overlapping the -35 region) of the *E. coli gcl* gene (fig. 35). The location of this site, and the function of the gene itself, immediately suggested a functional significance with respect to regulation by IclR. The *gcl*

gene of *E. coli* encodes the protein, glyoxylate carboligase, whose function is the condensation of two molecules of glyoxylate to form tartronate semialdehyde.

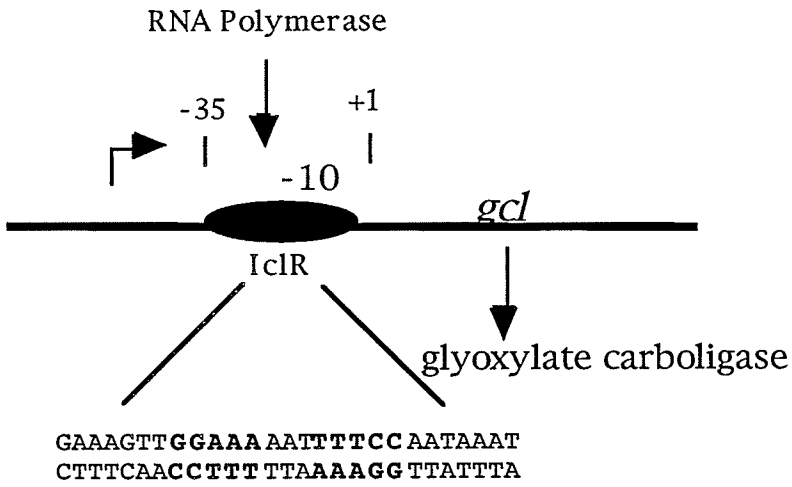


Fig. 35 Schematic illustration of the proposed location of the IclR binding site found 5' to the *gcl* gene. The location of the site is analogous to the situation observed for the *ace* and *iclR* operators. The DNA sequence of the operator contains the consensus sequence (in bold face) discovered in the experiments described above.

When *E. coli* is growing on the two carbon molecule glycollate, the first step in its metabolism is its oxidation to glyoxylate. The glyoxylate formed can be converted to tartronate semialdehyde as described above, and through a series of steps, introduced into glycolysis as 3-phosphoglyceric acid (Kornberg, and Gotto, 1961; Chang, et al., 1993). The fact that a potential IclR binding site has now been located in the promoter of the *gcl* gene suggests that IclR may also play a role in controlling the metabolism of glycollate

and/or glyoxylate. The fact that glyoxylate carboligase is an inducible enzyme was firmly established by the work of Chang et al. (1993). When *E. coli* was grown on succinate the levels of the carboligase were barely detectable. However, when either glycollate or glyoxylate was added to the medium the level of the enzyme rose dramatically. When the cells were grown on glycollate, however, induction was only observed in strains carrying a functional glycollate oxidase gene (which converts glycollate into glyoxylate), indicating that glycollate was not the *in vivo* inducer. The authors suggested that glyoxylate was the direct inducer of *gcl* expression. However, the results presented in this thesis have indicated that glyoxylate itself has no effect on the ability of IclR to bind to the *aceBAK* operon. Thus if IclR is regulating the *gcl* gene the glyoxylate must first be metabolized before transcription is initiated at this gene. In order to further characterize the ability of IclR to regulate the *gcl* gene, the operator/promoter region of this gene was generated by PCR, and the ability of IclR to interact with it *in vitro* was analyzed by an EMSA.

The sequence of the proposed IclR binding site in the *gcl* promoter region (fig. 35) shares a high degree of homology to the SELEX derived binding site (see fig. 34). It was thus expected that this region would interact strongly with IclR. After cloning and sequencing the PCR derived *gcl* operator the EMSA confirmed this prediction. Shown in figure 36 is the binding curve generated when IclR binds to the *gcl* promoter. The measured affinity for IclR binding to this site is 0.5±0.1nM while the Hill constant is 1.2±0.1.

These results clearly indicate that IclR is able to bind specifically to the *gcl* promoter region. The *in vivo* consequence of this interaction is not yet understood, however, operon fusion experiments currently underway in our laboratory will hopefully resolve this question.

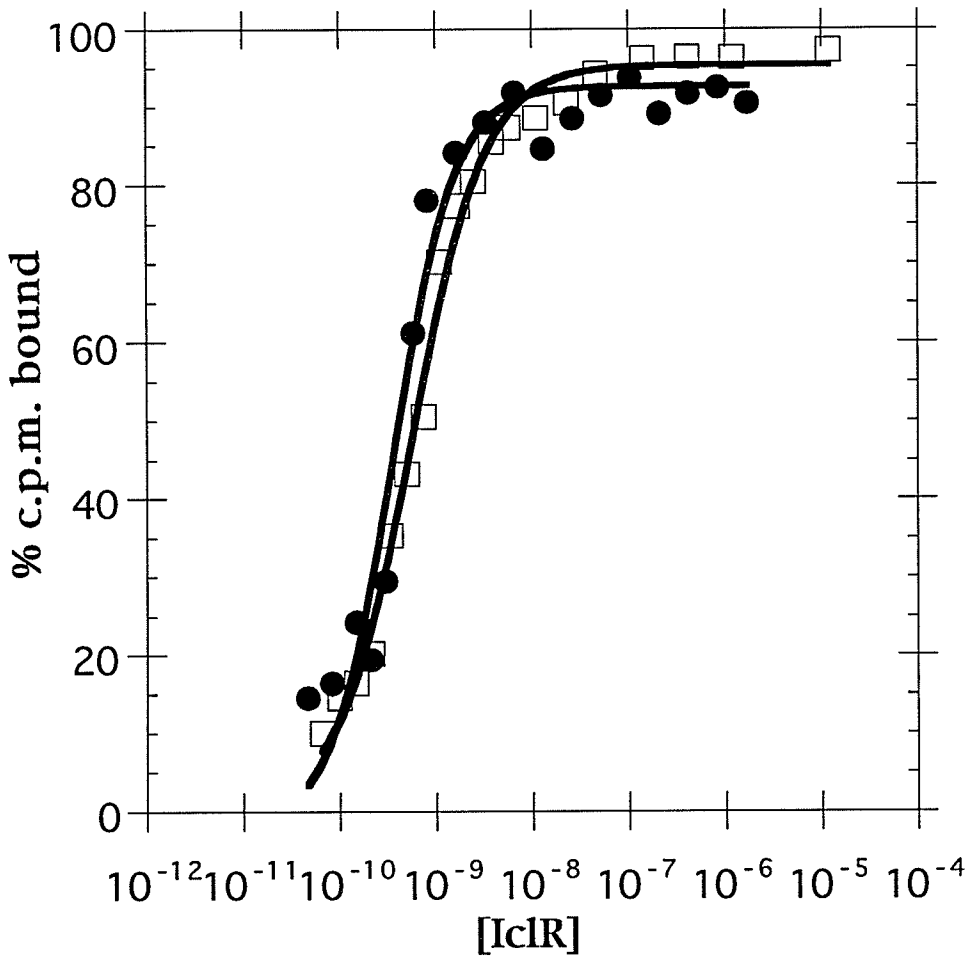


Fig. 36 Hill plot showing the binding affinity of full length IclR for the gcl operator(●). The binding curve generated using the *aceBAK* binding region(□) is also shown.

3.12 Design of a Synthetic IclR Binding Site

With the results of the SELEX experiment and the knowledge of three naturally occurring sites to which IclR bound I was able to design a 30 base pair oligonucleotide containing the IclR binding consensus (fig. 37). This oligonucleotide was tested for its ability to interact with IclR in a standard gel retardation assay (fig. 37) and it was found to have the highest affinity of any of the binding sites tested. The cooperativity, however, as measured by the Hill constant, was markedly decreased. The affinity was measured to be 0.07±0.03nM while the Hill constant was found to be 0.9±0.1..

3.13 FadR has no effect on the IclR-*aceBAK* interaction

As described in the introduction, the levels of the glyoxylate cycle enzymes are negatively regulated by the product of the *fadR* gene. The purified FadR protein was obtained from Dr. L.J. Donald and tested to determine whether this protein could interact with the *aceBAK* operator either by itself, or cooperatively with IclR. An EMSA was performed with both IclR and FadR present (fig 38), and the results clearly indicate that *in vitro*, FadR does not bind to controlling sequences close to the *aceB* promoter. This was not surprising, however, as examination of the sequence did not reveal any sequences that were obviously similar to the FadR consensus. In addition, when FadR and IclR were incubated simultaneously with *ace* operator no change in DNA mobility was observed beyond that

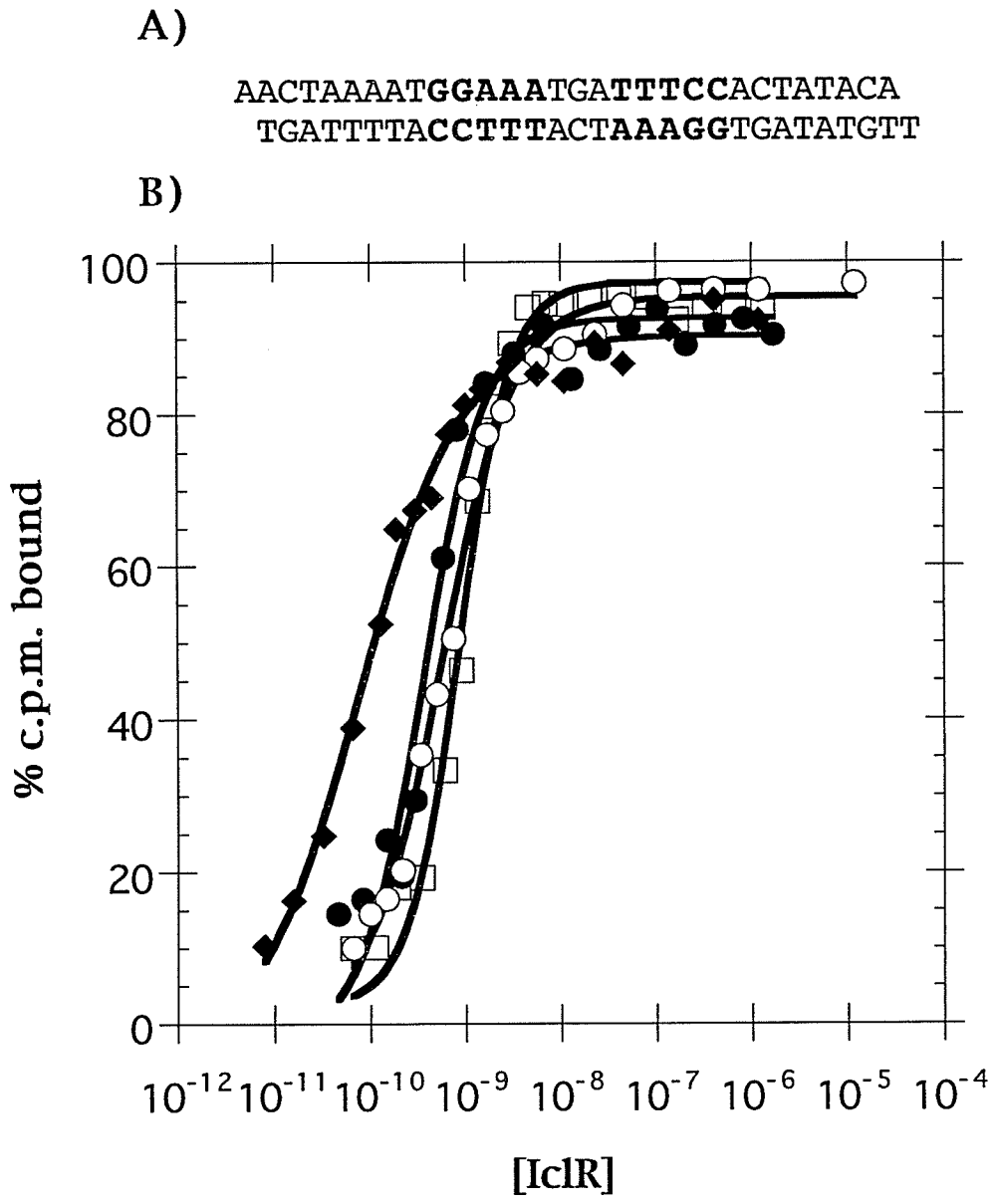


Fig. 37 A) Sequence of the double stranded 30-mer that contained the consensus IclR binding site. The ends were constructed to have an overhang to facilitate radio-labeling and crystal growth. B) Hill plot showing IclR binding to the 30-mer(♦) along with the plots from the three natural operators, *aceBAK*(●), *iclR*(□), and *gcl*(○).

induced by IclR alone. Thus, protein-protein interactions between FadR and IclR cannot explain the role of FadR in the regulation of the

ace operon. The *fadR* gene must therefore affect the production of the *ace* enzymes by a more complicated mechanism.

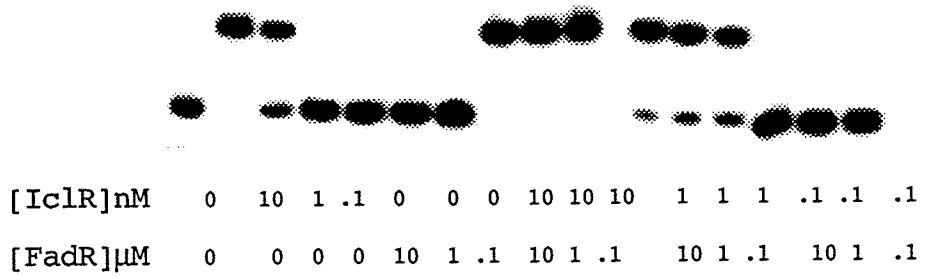


Fig. 38 An EMSA using both IclR and FadR with the *aceBAK* promoter was performed to determine if FadR either interacted alone with the DNA or perhaps affected the affinity of IclR for the DNA. The IclR and FadR concentrations present in each lane are shown in the figure. As seen in the figure, the amount of complex decreases with decreasing IclR concentration. However, the amount of complex is unaffected by the amount of FadR present, with and without IclR.

Chapter 4

Discussion

4.0 IclR Mutants

In this thesis the DNA binding properties of the IclR protein was examined by site-directed mutagenesis. The first four amino acids of the proposed recognition helix (N57, S58, T59, and T60) and a residue found in the N-terminus (R9) were mutated to alanine, and leucine respectively. The effect of the mutations on the DNA binding properties were measured by an EMSA. The recognition helix mutants N57A, S58A, and T60A each only showed a modest decrease in the binding affinity of the protein for DNA of 6.5, 13, and 2.5 fold, respectively. In several well characterized protein-DNA complexes, it has been shown that mutating side chains that directly contact the DNA results in a much more dramatic loss in binding affinity. For example, mutation of functionally important DNA-binding residues in the Mnt repressor reduces the affinity of the protein for DNA by 60-12000 fold (Knight and Sauer, 1989). The same situation is also observed for a variety of helix-turn-helix proteins including the 434 repressor (Huang et al. 1994), the LexA protein (Thliveris and Mount, 1992), and the Tet repressor (Baumeister et al. 1992). The fact that the mutations listed above did not lead to the same kind of decrease in binding affinity suggests that, perhaps with an exception of serine-58, these residues are probably not making critical contacts with DNA.

These results are not particularly surprising when one compares the amino acid alignments shown in fig. 39.

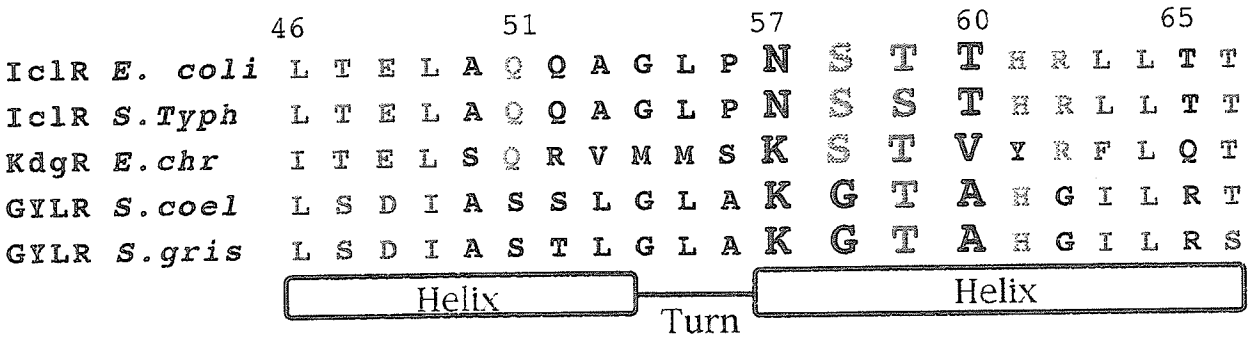


Fig. 39 Illustration of the proposed Helix-turn-helix motifs from IclR and known (not hypothetical open reading frames) proteins sharing a high degree of sequence homology (see fig. 15). The residues in red are those that are absolutely conserved among all five proteins, while those shown in green are conserved between IclR and either KdgR or GylR. The enlarged residues correspond to the IclR amino acids that were studied in this thesis. The numbering is from the *E. coli* IclR sequence (Nègre et al. 1991).

Of the N57, S58, and T60 residues, the only one that shows some conservation is that whose mutation shows the greatest effect on DNA binding, Ser58. This residue is found not only in the IclR sequences, but also in the KdgR sequence. At position 57 asparigine is only found in the two IclR proteins. In KdgR and GylR this residue is a lysine. At position 60 an even smaller degree of conservation exists, as this position is a valine in KdgR, and alanine in the two GylR proteins. Correspondingly, the T60A mutation has the smallest effect on DNA binding.

The T59A mutation decreases the operator binding affinity of IclR substantially (see fig. 20 and table 3), suggesting that this residue is making an important contribution to the DNA binding properties of the protein. When IclR is aligned against the other

proteins sharing a high degree of sequence homology (fig 39) it can be seen that this residue is highly conserved. Only in IclR from *Salmonella* is a difference observed; in this protein a serine replaces the conserved threonine. As it seems unlikely that the wild type *Salmonella* protein would have a dramatically lower DNA binding affinity than the *E. coli* protein, this substitution suggests that in the *E. coli* complex it is the hydroxyl function of threonine, rather than the extra methyl group that is responsible for DNA binding. Mutation of this residue to serine will be required to answer this question decisively.

The question of whether or not IclR is in fact a helix-turn-helix protein has not yet been verified from the experiments described in this thesis. Although the T59A mutation has a severe effect on DNA binding, the other three mutants show only a very small to moderate effect suggesting that of the first four N-terminal amino acids in the "recognition" helix only T59 makes base specifying contacts. This is unusual in prokaryotic helix-turn-helix proteins as a substantial number of these contacts usually are seen to arise from these residues (Harrison and Aggarwal, 1990). Exceptions, however, have been observed. In the co-crystal structure of the Trp-repressor-DNA complex the first two amino acids in the recognition helix only make contacts to the DNA through water mediated hydrogen bonds with their main chain amide nitrogens indicating the nature of the side chains at these two positions is relatively unimportant (Otwinowski et al. 1988; Lawson and Carey, 1993). Site directed mutagenesis have verified this result (Pfau et al. 1994). Site directed mutagenesis

experiments have also shown that the first four residues of the recognition helix in the Lex repressor (Thliveris and Mount, 1992), are probably not critical for recognition. Rather, amino acids further along the helix seem to be functionally important. Finally, the homeodomain proteins found in eucaryotic organisms also contain a helix-turn-helix motif. In these proteins, however, most of the base specifying interactions are found at the C-terminal rather than N-terminal portion of the "recognition" helix (Kissinger et al. 1990; Wolberger et al. 1991; Wilson et al. 1995). It is thus possible that IclR also utilizes a helix-turn-helix protein in a manner that is atypical for procaryotic repressors. Perhaps in the IclR-DNA complex, as in the homeodomain complexes, it is the C-terminal portion of the recognition helix that make base specifying contacts with the major groove of the DNA.

The sequence alignments presented above support the above suggestion as the degree of conservation seems to be much higher at the C-terminal portion of the recognition helix. For example, histidine 61 is conserved in both the IclR and GylR sequences. At positions 62 IclR and KdgR both possess an arginine side chain. The next two positions, 63 and 64, both show absolute conservation of a hydrophobic side chain, with position 64 showing absolute conservation of a leucine residue. Finally, conservation of a hydroxyl containing amino acid is evident at position 66 as all these residues are either threonine or serine. Site directed mutagenesis experiments targeting these amino acids will have to be performed to validate the importance of these conserved residues.

In addition to a marked degree of conservation in the recognition helix, a high homology is also present in the first helix of the proposed motif. At amino acid 46, 47, 48, and 49 conservation of hydrophobic, hydroxyl containing, acidic, and hydrophobic side chains, respectively, is obvious. Position 50 is also highly conserved (alanine), however, this is expected as the fifth amino acid of a helix-turn-helix- domain is usually alanine. These conservations suggest that IclR may utilize these residues, in addition to the C-terminal residues of the "recognition helix" in contacting DNA. Once again, site-directed mutagenesis will be required to answer this question.

The N-terminal arm of the IclR protein also plays a role in DNA binding. Upon removal of either the first eight or nine amino acids the affinity of the repressor falls approximately 40 fold suggesting that this region of the protein contributes significantly to DNA binding. In several examples reported in the literature the N-terminal arm is able to form an extended chain that can make base specific contacts with the minor groove, e.g. lambda repressor (Pabo et al. 1982; Jordan and Pabo, 1988); engrailed homeodomain (Kissinger et al. 1990); Matα2 homeodomain (Wolberger et al. 1991); and Hin recombinase (Feng et al. 1994). It is possible that IclR also contacts its cognate DNA by a similar mechanism.

In the N-terminal arm of IclR, mutation of arginine-9 to leucine reduces repressor binding 16.5 fold. This finding was unanticipated as removal of this arginine when the first eight amino acids were

already missing did not change the binding affinity of the protein at all (i.e. IclR had a similar affinity for binding regardless of whether it was missing the first eight or nine amino acids). As the R9L change corresponds to roughly half the binding energy associated with the N-terminal arm a main function of the eight N-terminal amino acids might involve stabilizing arginine-9 in a DNA binding conformation. Removal of the first eight residues results in a loss of this stabilization, consequently both truncated proteins bind to DNA with similar affinities.

4.1 IclR Binding to Various DNA Sites

The identification of a consensus sequence for IclR was a significant step in the characterization of the molecular mechanism by which IclR binds to DNA. The binding sites were selected by two methods throughout the course of the experiments. The first method, an EMSA, was unsuccessful in selecting high affinity oligonucleotides from a random pool. Although several groups have successfully used EMSAs in the past (Cui et al. 1996; He et al. 1995; and Toledano et al. 1994) I was unable to adapt this method for IclR, presumably because the background levels of contamination were too high. The situation was improved greatly when the partitioning of bound and free complexes was performed with nitrocellulose filters. The filters seemingly reduced the amount of contamination introduced into the PCR reaction, thus reducing the number of cycles required for enrichment (Irvine et al. 1991). In addition, the filter binding method was much quicker than the EMSA technique. While

the latter method generally required a minimum of two days to complete one cycle of selection, the former method was able to accommodate two rounds of selection per day, thus increasing the productiveness of the experiments.

The DNA consensus sequence selected by IclR was shown to be palindromic with a half site spacing of three base pairs. The consensus half site, or IclR-box, was found to be AAWDTGGAAA where W = A or T, and D = not C. While the first five bases in the consensus are highly conserved, the last five (GGAAA) are absolutely conserved. DNase analysis, however, has shown that all of the bases in the consensus half site are protected from digestion (Donald et al. 1996; Cortay et al. 1991). Taken together, these results suggest that the outer, not so well conserved bases, may contact IclR in a less specific manner, while the inner five nucleotides, GGAAA, may be responsible for providing the majority of the base specific interactions. The less specific interactions provided by the outer bases could involve phosphate or sugar moieties that interact with specific protein side chains. The observation that this region does, however, tend to prefer adenine and thymine bases suggests that IclR may, like a large number of well characterized DNA binding proteins such as CAP (Schultz et al. 1991), LacR (Lewis et al. 1996), and PurR (Schumacher et al. 1994) bend its operator upon binding.

In all of the sequenced clones that contained two IclR half-sites, three bases were always found in the intervening region. The fact that this half site spacing is so conserved is probably a

direct result of the distance between the recognition helices in the IclR dimer. If IclR is in fact a helix-turn-helix protein then the recognition helices would fit into consecutive turns of the major groove. Increasing or decreasing the spacing between the two half sites would alter the relative positions (by a 3.4Å translation, and 36° rotation around the helical axis of the molecule, per base pair) of the two major grooves thus preventing IclR binding. Thus, in addition to the chemical identity of the bases in the IclR binding site, the spacing between these critical residues affords a second source of sequence discrimination. Experiments that study the effect of changing the half site spacing between the two IclR boxes will probably prove fruitful in further dissecting the molecular mechanism by which IclR binds to DNA. Other proteins that only recognize half-sites separated by the proper spacing include, to name a few, the P22 Arc repressor (Smith and Sauer, 1995), the MetJ repressor (Phillips et al., 1989; He et al., 1996), the glucocorticoid receptor (Luisi et al., 1991), and the LexA repressor (Oertel-Bucheit et al. 1993)

The DNA binding properties of a number of isolated clones were examined for their ability to bind IclR. As expected the molecules containing a full site were able to interact with the protein with high affinity. In the three clones examined the inner five bases of the IclR-box were GGAAA (fig. 34), but the outer five bases, and the bases comprising the inner three were variant. This lack of sequence specificity, and the fact that IclR is still able to interact

with high affinity with these molecules again suggests that IclR may make a greater number of nonspecific interactions with the outer bases in the IclR binding sites, while making more specific interactions with the inner five bases. Interestingly, however, the clone with the lowest AT content, and thus the lowest similarity to the consensus sequence, in the outer five bases (clone 25) had the highest affinity for IclR. The reason for this is uncertain but perhaps it is due to the identity of the three nucleotides separating the two half sites. The isolation, and examination of additional selected sequences will be required to clarify this issue.

An interesting finding to emerge from the SELEX experiments was the fact that IclR was able to select a relatively large proportion of DNA sequences containing only one half site (fig. 33). When two of these sequences were isolated, EMSA experiments demonstrated that IclR was still able to bind rather tightly, with dissociation constants of approximately 10nM. The *in vivo* consequence of this result is not understood, but it suggests that IclR might interact weakly with additional sites in the genome. That this result is not simply a non-specific interaction of IclR with DNA is evidenced by the fact that the K_Ds for two different half-sites are similar, and also the fact that IclR showed almost no affinity for the non-consensus containing DNA obtained from clone 36 (fig. 33).

4.1.1 Comparison of the IclR Consensus to the in vivo Operators

The consensus sequence described above could not have been defined solely from a comparison of the *iclR* and *aceBAK* operators, the two naturally occurring IclR binding sequences known prior to the SELEX experiments. For example, neither of the naturally occurring sites contains two perfect half sites; the *aceBAK* site replaces the two cytosines found in the second (3') half site with thymines, while the *iclR* site replaces the second guanine found in the first half site (5') with an adenine. Only when the *gcl* O/P region is included in the alignment does the naturally occurring consensus more resemble that obtained from the SELEX experiments (fig. ?)

ace*B*	TAATTAAAAT**GGAAA**TTG**TTTTT**GATTTTGCATT
iclR	CAATAAAAAT**GAAAA**TGA**TTTCC**ACGATACAGAA
gcl	TTTGAAAGTT**GGAAA**AAT**TTTCC**AATAAATAGAG
SELEX	AAAAT**GGAAA**NNN**TTTCC**ATTTT
	TT AA
	G C
30-mer	AACTAAAA**TGGAAA**TGA**TTTCCA**CTATACA

Fig. 40 Alignment of the three naturally occuring IclR binding sites from E. coli. along with the SELEX selected sequence and the synthetic 30 mer. The sequences are aligned around the conserved GGAAA half sites. Note that neither the ace*BAK* or *iclR* sequences have two perfect IclR boxes.

It is of interest that neither the *aceBAK* or *iclR* O/P regions contain two perfect IclR boxes; each only contains one complete half site. In contrast, the *gcl* site contains two half sites where the inner five base pairs are identical in to the consensus defined by SELEX. A reasonable assumption from this result would be that the *gcl* O/P region might bind more tightly to IclR than the other two operators. This is not borne out in experiment, however. The naturally occurring *gcl* promoter binds with an affinity intermediate to the *aceBAK* (which binds the tightest) and *iclR* regions (fig. 36). Why then did IclR almost always select sequences that contained two complete half sites? One explanation is that the bases outside of the GGAAA motif play an important role in supplying binding free energy to the complex. In the naturally occurring *aceBAK* and *iclR* sites the bases 5' to the conserved GGAAA motif all agree perfectly with the bases selected with highest proportion in the SELEX experiment. In the *gcl* O/P region two of these bases do not agree with the SELEX consensus. Thus, it is possible that although some binding energy might be lost by not conforming to the inner consensus, fine tuning of the bases outside the GGAAA motif may afford extra binding energy. The fact that a synthetic oligonucleotide (fig. 40) which was constructed to contain the consensus of both the natural sites and the SELEX selected sites bound to IclR more tightly than any of the naturally occurring sites confirms this suggestion.

Why then did nature not select IclR binding sequences that more resemble the 30 base pair oligonucleotide? Perhaps evolution has modified the sequences not to maximize the binding affinity, but

rather to optimize the *in vivo* metabolic requirements of the organism. For example, a tighter binding sequence may dissociate from the protein much more slowly than the natural sequences do. This could be deleterious in the case of IclR if induction of the genes it controls depends on the ability of a small molecule, or inducer, to first bind to the protein preventing it from binding DNA. If the small molecule had to wait for the protein first to dissociate before it showed its effect, a tighter binding complex, would take a longer time to induce. It is also possible, of course, that the operators do not perfectly resemble the consensus sequence as a result of the fact that they share a binding site with RNA polymerase. Evolution, perhaps, has modified both sequences in such a way that they are better able to control the regulation of the glyoxylate cycle.

4.1.2 Why is IclR Binding Cooperative?

Cooperativity is common to a great number of biological systems, including protein-DNA interaction, protein-protein interactions, and protein-ligand interactions The classical definition of a cooperative system states that the binding of one ligand, say to a protein, affects the ability of a second ligand to bind to the same protein. In hemoglobin, for example, the paradigm of cooperative systems, binding of one oxygen molecule makes the binding of the second easier. When two molecules of oxygen are bound the third binds even more easily and finally when three molecules are bound, the fourth binds with the greatest ease of all. The molecular explanation for this observation is that the binding of oxygen to one

of the hemoglobin subunits effects a conformational (allosteric) change in the quaternary structure of the protein allowing for the next ligand to bind to another subunit with greater ease.

In the case of protein-DNA interactions, cooperativity is thought to arise by the assembly of protein oligomers on the DNA helix. Interactions between the bound oligomers are then thought to account for the cooperativity. In the simplest systems the assembly generally involves the sequential addition of homo-oligomers to different DNA sites. A good illustration of this phenomenon is seen with the P22 Arc repressor. This protein, at the concentrations used for DNA binding exists as an equilibrium mixture of dimers and unfolded monomer (Bowie and Sauer, 1989). The DNA binding reaction involves the sequential addition of Arc dimer to individual half sites within its operator. When the first dimer is bound to one half site the affinity of a second dimer for the other half is enhanced 5900 fold relative to the non-cooperative case (Brown and Sauer, 1993). The crystal structure of the Arc-DNA complex (Raumann et al., 1994) has shown that this cooperativity arises primarily from protein-protein interactions across the dimer-dimer interface.

Cooperativity in protein-DNA interactions, however, can also arise through conformational changes in the DNA. For example, binding of a protein subunit at a particular DNA binding site can alter the conformation of a second binding site. This alteration would be expected to change the affinity of the second site for protein and cooperativity could theoretically be observed even when the two

protein subunits are not contacting each other. In the literature there is evidence that this is the case for binding of the λ-repressor to its operators (Strahs, and Brenowitz, 1994), binding of the paired class (Pax) homeodomain dimer to its operator (Wilson et al. 1995), and activation of transcription by integration host factor (Parekh and Hatfield, 1996). In the paired class homeodomain, the cooperative interactions are also mediated, to a small extent, by protein-protein contacts between adjacently bound monomers.

Evidence presented in this thesis suggests that the cooperative interactions observed for IclR binding arise primarily through DNA conformational changes. The association kinetics, and stoichiometry experiments, suggest that at the concentrations used for DNA binding IclR exists as a mixture of monomers and dimers. In the equilibrium DNA binding experiments it is observed that the cooperativity of binding (as measured by the Hill number) tends to decrease as the DNA binding site more conforms to the consensus. For example, the Hill number for binding to the *aceBAK* and *iclR* operators is 1.6, while it is close to unity for the consensus 30 mer. In addition, the Hill number for IclR binding to half sites is somewhat less than unity. These data thus suggest that the exact sequence of DNA bases in the IclR binding site are critical determinants for cooperative interactions. If protein-protein contacts formed when two monomers form a dimer on the DNA were responsible for the cooperativity, the sequence of DNA bases found in the binding site would not be expected to alter the cooperativity dramatically. This was clearly not the case in our hands.

When only one perfect half site is present (*aceBAK* and *iclR*) in the IclR binding site it is possible that the kinetic route of assembly is as follows: one subunit of IclR, either monomer or dimer, first binds to the perfect half site present in each operator. Conformational changes in the DNA, then might have the effect of altering the structure of the non-consensus half site to one that more resembles the consensus site. Following this conformational change, a second IclR subunit binds to the vacant and altered half-site to form the stable protein-DNA complex.

This model for IclR binding suggests that if both half sites resembled the consensus sequence less cooperativity would be observed. This was in fact borne out in the experiment with the 30 base pair oligonucleotide described above. In addition, cooperative interactions were never observed when IclR bound to sites containing only one IclR-box (fig. 34). Presumably the binding of one IclR subunit can only alter the second half site, and prepare it for binding, when that site closely resembles the consensus. Thus, the cooperative interactions observed for IclR binding to DNA depend critically on the local DNA sequence and not on protein-protein contacts between bound monomers. The reason that truncated IclR binds with higher cooperativity to the *aceBAK* operator is not yet understood in the context of the above described mechanism. DNA binding experiments performed with the other operators described in this work will hopefully clarify the issue somewhat.

4.2 How Does IclR Regulate the Glyoxylate Cycle?

To date no satisfactory explanations have been put forward that explain the molecular processes involved in the expression of the *ace* operon. In the absence of any ligand, IclR has been shown to bind to the operon with moderate to high affinity, at sites that overlap the recognition sequence for RNA polymerase. Thus, in glucose containing medium it would be expected that IclR would prevent transcription by sterically preventing RNA polymerase from forming an initiation complex. When the medium is changed to acetate, the *ace* operon becomes deregulated, suggesting that the concentration of a specific inducer molecule may rise and prevent the formation of the IclR-DNA complex, thus permitting transcription. An alternative hypothesis, however, is that the *ace* operon is deregulated by a drop in concentration of a specific co-repressor molecule. The fact that the affinity of IclR for DNA is quite strong in the absence of any ligand argues that the former situation is most likely to occur *in vivo*.

The nature of the effector molecule for the *ace* operon still remains to be elucidated despite intensive study over the past 40 years. However, several lines of evidence have demonstrated that for induction to occur the following criteria have to be met: 1) The organism must contain a functional citrate synthase (Kornberg, 1966) suggesting that acetyl-CoA cannot be the inducer. Proline, and α–ketoglutarate, however, can still induce the glyoxylate cycle without a functional citrate synthase indicating that induction of the

ace operon when growing on acetate first requires some of the acetate to be metabolized (Duckworth, 1980). 2) A functional malate synthase is also required for induction on acetate again indicating that acetate must first be metabolized before induction can occur. Finally, induction is independent of cyclic-AMP, as *crp* and *cya* mutants are still able to induce isocitrate lyase (Duckworth, 1980).

Experiments performed in this thesis have demonstrated that a number of compounds from glycolysis, the TCA cycle, and other pathways do not show any effect on IclR binding. For example, it was suggested that on the basis of the experiments described above that neither acetate or acetyl-CoA could be the inducer of the *ace* operon. This was confirmed here and has been confirmed previously (Cortay et al. 1991). In addition I have shown that the TCA cycle intermediates α-ketoglutarate, oxaloacetate, malate, succinate, citrate, and isocitrate have no effect on IclR binding suggesting that none of these compounds is the intracellular inducer. The fact that α-ketoglutarate had no effect on IclR binding to the *ace* operon suggests that the ability of this compound (and proline, which is metabolized through α-ketoglutarate) to induce the *ace* operon must arise from its metabolism to another, probably non-TCA compound. As fumarate and succinyl-CoA were not tested it is not possible to comment on the role played by these TCA cycle intermediates. These results thus indicate that for acetate to induce the *ace* operon it must first be metabolized to a compound outside of the TCA cycle. If this compound resides in the gluceogenic pathway, it cannot be pyruvate, phosphoenolpyruvate, fructose-1,6-bisphosphate,

3-phosphoglycerate, or glucose as none of these compounds had an effect on DNA binding.

An alternative hypothesis as to the nature of the inducing molecule comes from the observation that IclR also interacts with the *gcl* O/P region. This gene codes for the enzyme glyoxylate carboligase, a key enzyme in the metabolism of glycollate. It is possible that IclR responds to the levels of the intermediates resulting from the action of this enzyme. Thus it is possible that tartronate semialdehyde (or its interconvertible isomer hydroxypyruvate), or other products resulting from the metabolism of glycollate, induce IclR expression. However, as isocitrate lyase is only induced to one tenth its maximal level in glycollate containing medium (Kornberg, 1966; Duckworth, 1980) other factors must play a role in *ace* expression if this hypothesis is in fact correct. As described below, other factors clearly are important.

As it stands today the expression of *ace* operon is also under control of three gene products completely unrelated to IclR. The first such molecule is the FadR protein, a protein which also regulates the levels of the enzymes controlling the synthesis and degradation of fatty acids. The oxidative fatty acid degradative enzymes are repressed by this molecule, while those responsible for fatty acid synthesis are activated. The allosteric effectors of this protein are long chain acyl CoAs which bind to FadR, preventing it from interacting with DNA. Thus when *E. coli* is growing on long chain fatty acids, FadR is in its non-DNA binding form and the degradative

enzymes are able to break down the fatty acids into acetyl-CoA. The acetyl-CoA is then metabolized via the glyoxylate cycle, allowing for the generation of gluconeogenic substrates.

It is believed that FadR regulates the *ace* operon at the level of transcription as the levels of isocitrate lyase and malate synthase in a *fadR* strain are greatly increased under non-inducing conditions (Maloy et al. 1980). In this study it was shown that this increase was a direct result of increased transcription rather than from the accumulation of metabolic products resulting from the *fadR* mutation. In the work presented here I have demonstrated that FadR is unable to interact with any *cis*-acting regulatory elements within the local proximity of the *ace* operon. In addition I have shown that FadR does not regulate *ace* expression by interacting with IclR itself.

It is possible that FadR regulates the levels of IclR by binding to cis-acting elements at the *iclR* gene. This mechanism, however, could not account for the fact that strains carrying both *iclR* and *fadR* mutations have higher levels of isocitrate lyase than strains carrying only an *iclR* mutation (Maloy and Nunn, 1982). If *iclR* was regulated by FadR, the strain carrying both mutations should have had the same levels of isocitrate lyase as the strain that carried only the *iclR* mutation. It is also possible that FadR may affect the *ace* operon indirectly by interacting with the *gcl* O/P region controlling transcription of glyoxylate carboligase. If this enzyme leads to the accumulation of a positive effector for the *ace* operon, a *fadR* mutation would increase the levels of the carboligase and the

accumulation of a positive effector resulting from this enzyme would then de-repress *aceBAK*. A direct observation of a protein-DNA complex formed by FadR and either the *iclR* and *gcl* O/P regions will hopefully clarify this issue somewhat.

In addition to IclR and FadR, the *ace* operon is also regulated by the FruR protein and the integration host factor. As both of these molecules have been shown to directly interact with the *ace* operon the mechanism by which these proteins regulate it is less of a mystery than the FadR protein. Both FruR and IHF regulate the *ace* operon in a positive manner, that is mutation of the protein results in decreased levels of *ace* transcription. Thus when growing on acetate these proteins must somehow activate transcription at the *ace* promoter. FruR binds to a region approximately 170bp upstream of the transcription start site suggesting that it cannot activate RNA polymerase by simply interacting with it at an adjacent site (as is often observed with CAP-cAMP). For activation to occur, it is possible that the DNA forms a loop complex such that FruR is able to interact with RNA polymerase and activate transcription in a manner similar to that observed for other well characterized procaryotic and eucaryotic activators (Vignais and Sentenac, 1989). This looping may be mediated by the IHF, a protein whose known function is exactly this (Freundlich et al. 1992). Mutation of either IHF or FruR would prevent these contacts from forming, thus explaining why both FruR and IHF are positive activators of *ace* expression.

This mechanism of transcriptional activation by FruR and IHF can be readily explained by the fact that FruR is easily displaced from its binding site by fructose-1-phosphate, and fructose-1,6-phosphate. When cells are growing on acetate, glucose addition raises the levels of fructose-1,6-bisphosphate approximately 1500 fold (Lowry, 1971). Thus glucose addition could cause cessation of the glyoxylate cycle enzymes for two reasons: 1) it could change the levels of the hypothetical IclR inducer molecule, and 2) it could displace FruR from its binding site at the *ace* operon preventing it from activating transcription. Alternatively, the mechanism by which repression on glucose arises may not depend on the action of any small effector molecule. If the FruR-IHF mediated DNA loop in some way prevented IclR from binding to its operator, it could effectively act as the small molecule. In glucose grown cells, the looping would not occur as FruR could not bind DNA. Thus in glucose, IclR would be bound to its operator and transcription would be repressed. In acetate the levels of fructose-1,6-bisphosphate (and perhaps fructose-1-phosphate) would dramatically decrease allowing for FruR to bind to the *ace* operon and form the IHF mediated loop. It is then possible that this looped complex instead of activating transcription by interacting with RNA polymerase, activates transcription by displacing IclR from its binding site. As visual inspection of the region 5' to the *iclR* gene has revealed no IHF or FruR sites this latter mechanism would explain why *iclR* expression is

insensitive to carbon source. Figure 41 illustrates the mechanisms discussed above.

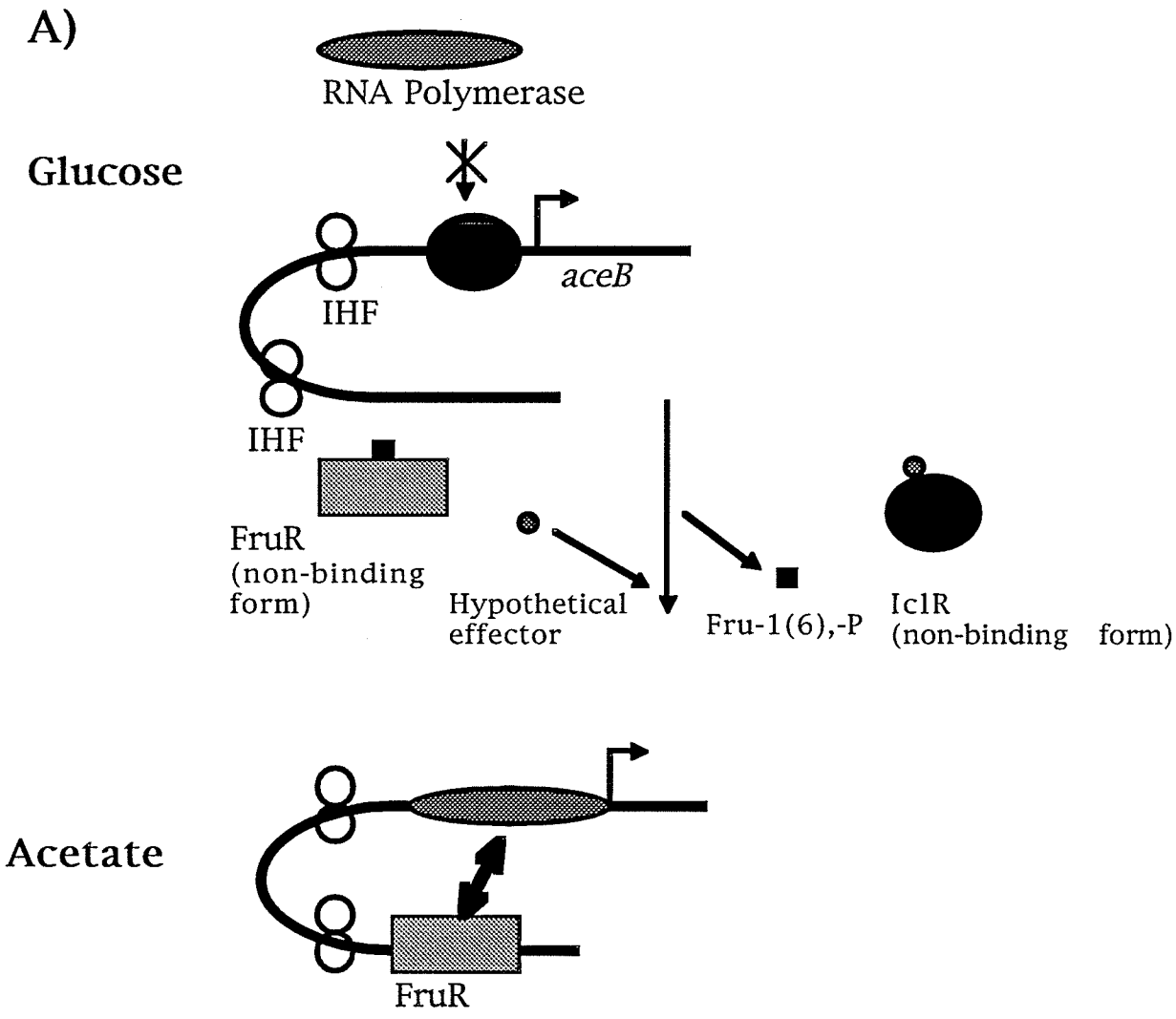


Fig. 41A Possible mechanism describing how the *ace* operon becomes deregulated. When the medium is changed from glucose to acetate the levels of fructose-1,6 (and possibly fructose-1) phosphate decrease. In addition the levels of a hypothetical inducer that effects IclR binding may also rise. These physiological changes would then displace IclR from DNA while allowing FruR to activate transcription by interacting with RNA polymerase. In this model, the integration host factor would simply serve as an agent that bends the DNA allowing for the productive contacts to be formed. It is possible, of course, that IHF may also play additional roles. Whether or not the bending would occur in glucose medium is unknown.

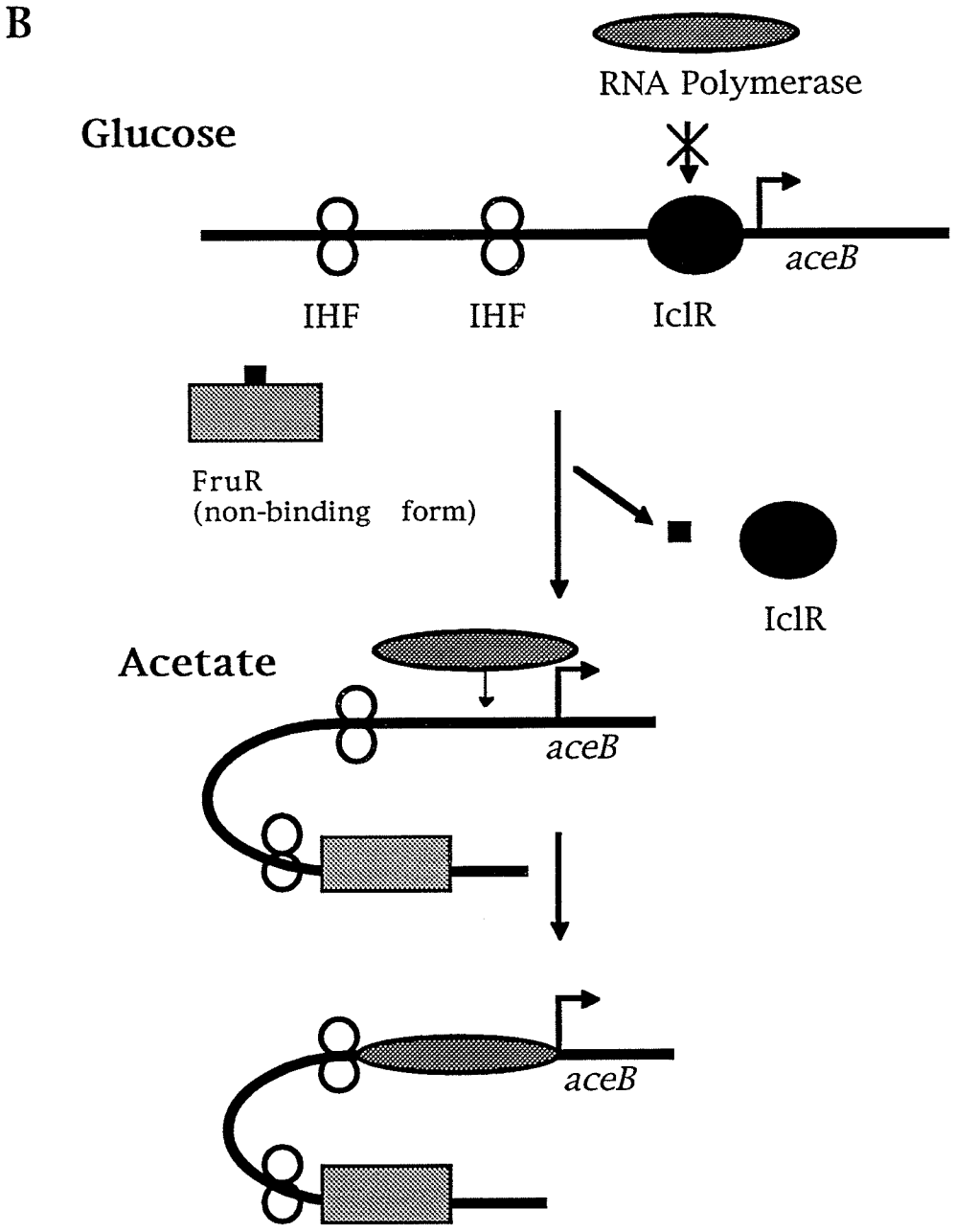


Fig. 41B Alternative mechanism explaining how changing the medium from glucose to acetate induces *ace* expression. In this model, like the first model, IclR is tightly bound to the *ace* operator in glucose. However, when the medium is changed to acetate there is no requirement for an inducer to bind IclR. Instead the binding of IHF (which would also be bound in glucose-not shown) and more specifically FruR would cause IclR to fall off the operator. This could be caused by perturbations in the DNA molecule or by FruR making the binding of RNA polymerase at the *ace* promoter much more favorable than the binding of IclR. This model would explain the fact that expression of the *iclR* gene, which does not posess any IHF sites (and probably no FruR sites) is insensitive to carbon source.

The molecular mechanisms controlling the regulation of the *ace* operon have now been the focus of intense research for over 40 years. Although we now have a better idea as to the identity and function of the protein products involved in the metabolism of acetate the picture is not nearly complete. Experiments in the future will have to absolutely identify the mechanisms by which FruR and IHF mediate transcription. In addition the search for a small effector molecule, perhaps using a method that utilizes the affinity of IclR for this small molecule, should be continued. Additionally, characterization of the mechanisms by which IclR regulates the *gcl* gene, if it in fact does, will able us to better connect the pathways of acetate and glycollate metabolism in *E. coli*. Finally, the role played by FadR in regulating the *ace* operon will have to be established.

References

Ades, S.E., and Sauer, R.T., *Biochemistry*, **33**, 9187-9194, (1994)

Baumeister, R, Helbl, V., and Hillen, W,. *J. Mol. Biol*, **226**, 1257-1270 (1992)

Beutel, B.A., and Gold, L., *J. Mol. Biol.*, **228**, 803-812 (1992)

Borowiec, J.A., Zhang, L., Sasse-Dwight, S., and Gralla J.D., *J. Mol. Biol.* **196,** 101-111 (1987)

Bowie, J.U., and Sauer, R.T., *Biochemistry*, **28**, 7139-7143 (1989)

Bradford, M.M., *Anal Biochem.* **72**, 248-254 (1976)

Brenowitz, M., Jamison, E., Majumdar, A., and Adhya, S., *Biochemistry*, **29**, 3374-3383 (1990)

Brice, C.D., and Kornberg, H.L., *J. Bacteriol.* **96,** 2185-2186 (1968)

Brown, B.M., and Sauer, R.T., *Biochemistry* **32**, 1354-1363 (1993)

Byrne, C, Stokes, H.W., and Ward, K.A., *Nucl. Acids Res.* 16, 9342. (1988)

Carey, J., *Methods Enzymol.*. **208**, 103-117 (1991)

Carlomagno, M.S., Chiariotti, L., Alifano, P., Nappo, A.G., and Bruni C.B. *J. Mol. Biol.* **203**, 585-606, (1988)

Chang, Y-Y.,, Wang, A-Y., and Cronan, Jr., J.E., *J. Biol. Chem.*, **268,** 3911-3919 (1993)

Chakerian, A.E., Tesmer, V.M., Manly, S.P., Brackett J.K., Lynch, M.J., Hoh, J.T., and Matthews, K.S., *J. Biol. Chem.*, **266**, 1371-1374 (1991)

Chin, A.M., Feldheim, D.A., and Saier, M.H., Jr., *J. Bacteriol.*, **171,** 2424-2434 (1989)

Chou, P.Y., and Fasman, G.D., *Biochemistry*, **13**, 222-245, (1974)

Chung. T., Klumpp, D.J., and Laporte, D.C. *J. Bacteriol.* **170:** 386-392

Cortay, J-C., Negre, D., Galnier, A., Duclos, B., Perriere, G., and Cozzone, A.J. *EMBO J.,* 10, 675-679 (1991)

Cortay, J-C., Negre, D., Scarabel, M., Ramseier, T.M., Vartak, N.B., Reizer, J., Sauier, M.H., Jr., and Cozzone, A.J., *J. Biol. Chem*, 269, 14885-14891, (1994)

Cui, Y., Wang, Q., Stormo, G., and Calvo J.M., *J. Bacteriol.*, **177**, 4872-4880 (1995)

deCrombrugghe, B., Busby, S., and Buc, H., *Science*, **224**, 831-838 (1984)

DiRusso, C.C., Metzger, A.K., and Heimert, T.L., *Mol. Microbiol* **7**, 311-322 (1993)

Donald L.D., Chernushevich, I.V., Zhou, J., Verentchikov, A., Poppe-Schriemer, N., Hosfield, D.J., Westmore, J.B., Ens, W., Duckworth, H.W., and Standing, K.G., *Protein Science* (in press) 1996

Duckworth, M.L. PhD thesis (Regulation of the Glyoxylate cycle in *Escherichia coli* K12), University of Manitoba, 1980

Eismann, E.R., and Muller-Hill, B., *J. Mol. Biol.*, **213**, 763-775 (1990)

Feng, D.F., Johnson, M.S., and Doolitle R., *J. Mol. Evol.*, **21**, 112-125, (1985)

Feng, J.A., Johnson, R.C., and Dickerson, R.E., *Science*, **263**, 348-354 (1994)

Freundlich M., Ramani, N., Mathew, A., Sirko, A., and Tsui, P., *Mol. Microbiol* **6**, 2557-2563 (1992)

Gilbert, W. and Maxam, A., *Proc. Natl. Acad. Sci. U.S.A.* 70, 3581 (1973)

Gilbert, W., and Muller-Hill, B., *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3581-3584 (1966)

Gui, L., Sunnaborg A., Pan, B., and LaPorte D., *J. Bacteriol.* **178,** 321-324 (1996)

Hansen, R.W., and Hayashi, J.A., *J. Bacteriol.* **83**: 679-687 (1962)

Harrison, C.S., and Aggarwal, A.K., *Annu. Rev. Biochem.*, **59**:933-969 (1990)

He, Y-y., Stockley, P.G, and Gold, L. *J. Mol. Biol.*, **255**, 55-66 (1996)

Henry, M.F., and Cronan, Jr., J. E., *Cell*, **70**, 671-679 (1992)

Hill, A. V., *J. Physiol. (London)*, **40**, 190-224 (1910)

Huang, L., Sera, T., and Schultz P.G., *Proc. Natl. Acad. Sci.*, **91**, 3969-3973 (1994)

Irvine, D., Tuerk, C., and Gold, L. J. Mol. Biol., 222, 739-761 (1991)

Jacob, F, and Monod, J. *J. Mol. Biol.* 3, 318-356 (1961)

Jordan, S.R., and Pabo, C.O., *Science*, 242, 893-899 (1988)

Kissinger, C.R., Liu, B., Martin-Bianco, E, Kornberg, T.B., and Pabo, C.O., *Cell*, 63, 579-590, (1990)

Knight, K.L., and Sauer, R.T., *J. Biol. Chem.,* 264, 13706-13710 (1989)

Kornberg, H.L., and Gotto, A.M., *Biochem. J.,* 78, 69-82, (1961)

Kornberg, H.L. and Sadler, J.R. *Biochem J.* 81, 503-513 (1961)

Kornberg H.L. *Biochim. Biophys. Acta* 73, 517-519 (1963)

Kunkel, T., *Methods Enzymol*, **154**, 367-382 (1987)

Kramer, H., Amouyal, M., Nordheim, A., and Muller-Hill, B., *EMBO J.*, **7**, 547-556 (1988)

Laemmli, U.K., *Nature*, **227**, 280-285 (1970)

LaPorte, D.C., and Chung, T. *J. Biol. Chem.* **260**: 15291-15297 (1985)

Lawson, C.L., and Carey, J., *Nature*, **366**, 178-182 (1993)

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P., *Science*, **271**, 1247-1254 (1996)

Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B., and Muller-Hill, B., *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 7947-7951 (1987)

Lowry, O.H., Carter, J., Ward, J.B., and Glaser, L. *J. Biol. Chem.*, **246** 6511-6521 (1971)

Luisi, B.F., Xu, W.X., Otiwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B., *Nature*, **352**, 497-505, (1991)

Maniatis, T., Fritsch, E.F., and Sambrook, J., *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Springs Harbor; New York, (1982)

Maloy, S.R., Bohlander, M., and Nunn, W.D., *J.Bacteriol,* **143**, 720-725 (1980)

Maloy, S.R., and Nunn, W.D. *J. Bacteriol*, **149**, 173-180 (1982)

Maxam, A., and Gilbert, W., *Methods Enzymol*, **65**, 499-560 (1980)

Molina, I., Pellicer, M-T., Badia, J., Aguilar, J., and Baldoma, L., *Eur. J. Biochem*, **224**, 541-548, (1994)

Mondragon, A., and Harrison, S.C. *J. Mol. Biol.* **219**, 321-3334, (1991)

Mossing, M.C., and Record, Jr., M.T., *Science,* **223**, 889-892 (1986)

Molgat, G.F. M.Sc. Thesis, University of Manitoba, (1990)

Monod, J., Pappenheimer, A.M., and Cohen-Bazire, G. *Biochim. biophys. Acta.* **9**, 648 (1952)

Nasser, W., Reverchon, S., Condemine, G., and Robert-Baudouy, J., *J. Mol. Biol.* **244**, 427-440 (1994)

Negre, D., Cortay, J-C., Old, I.G., Galnier, A., Richaud, C., Saint Girons, I. and Cozzone, A.J., *Gene* **97**, 29-37, (1991)

Negre, D., Cortay, J-C., Galinier, A., Sauve, P., and Cozzone, A.J., *J. Mol. Biol.*, **228**, 23-29, (1992)

Norby, P.L., Pallisgaard, N., Pedersen, F.S., and Jorgensen P., *Nucleic Acids Res.*, **20**, 6317-6321, (1992)

OehlerS., Eismann, E.R., Kramer, H., and Muller-Hill, B., *EMBO J.* **9**, 973-979 (1990)

Oertel-Buchheit, P., Schmidt-DÖrr, T., Granger-Schnarr, M., and Schnarr, M., *J. Mol. Biol.*, **229**, 1-7 (1993)

Otwinowski, Z., Schevitz, R., Zhang, R.G., Lawson, C.L., Marmorstein, R.Q., Luisi, R.Q., and Sigler, P.B., *Nature,* **355**, 321-329 (1988)

Pabo, C.O., Krovatin, W., Jeffrey, A., and Sauer, R.T., *Nature,* **298**, 441-443 (1982)

Parekh, B.S., and Hatfield, G.W., *Proc. Natl. Acad. USA,* **93,** 1173-1177 (1996)

Pellicer, M-T, Badía, J., Aguilar, J., and Baldomà, L., *J. Bacteriol,* **178**, 2051-2059 (1996)

Peterson M.L., and Reznikoff, W.S., *J. Mol. Biol.*, **185**, 535-543, (1985)

Pfau, J., Arvidson, D.N., Youderian P., Pearson, L.L., and Sigman, D.S., *Biochemistry*, **33**, 11391-11403 (1994)

Phillips, S.E., Manfield, I., Parsons, I., Davidson, B.E., Rafferty, J.B., Somers, W.S., Margarita, D., Cohen, G.N., Saint-Girons, I., and Stockley, P.G., *Nature*, **341**, 711-715 (1989)

Pollock, R., and Treisman R., *Nucl. Acids Res.*, **18**, 6197-6204 (1990)

Ramseier, T.M., Negre, D., Cortay, J.C., Scarabel, M., Cozzone, A.J., and Saier, M.H., Jr., *J. Mol. Biol.* **234**, 28-44 (1993)

Raumann, B.E., Rould, M.A., Pabo, C.O., and Sauer, R.T., *Nature* **367**, 754-757 (1994)

Record, M.T., deHaseth, P.L., and Lohman, T.M., *Biochemistry* **16**, 4791-4796 (1977)

Resnik E., Pan, B., Ramani, N., Freundlich, M., and LaPorte, D.C. J. Bacteriol., 178, 2715-2717 (1996)

Reverchon, S., Nasser W., and Robert-Baudouy, J., *Mol. Microbiol.*, **5**, 2203-2216 (1991)

Riggs, A.D., Bourgeois, S., Newby, R.F., and Cohn, M., *J. Mol. Biol.* **34**, 365-368 (1968)

Robertson, D.L., and Joyce, G.F., *Nature*, **344**, 467-468 (1990)

Robson, B., and Suzuki, E., *J. Mol. Biol.*, **107**, 327-356, (1976)

Saier, M.H., Jr., and Chin, M., in *Bacterial Energetics* (Krulwich, T.A., ed) 273-299, Academic Press, NY (1990)

Saiki, R.K., in *PCR Technology*, H.A. Erlich Ed. (Stockton Press, New York) 7-16, (1989)

Sasmor, H.M., and Betz, J.L., *Gene*, **89**, 1-6, (1990)

Sasse-Dwight, S., and Gralla J.D. *J. Mol. Biol.* 202, **107**-119 (1988)

Schultz S.C., Shields G.C., and Steitz. T.A., *Science* **253**, 1001-1007 (1991)

Schumacher, M.A., Choi, K.Y., Zalkin, C.H., and Brennan, R.G., *Science* **266**, 763 (1994)

Smith, C.P., and Chater, K.F., *J. Mol. Biol.* **204**, 569-580, (1988)

Smith, T.L., and Sauer, R.T., *J.Mol. Biol.*, **249,** 729-742, (1995)

Somerville, R.L., In *Amino acid biosynthesis and regulation* (ed. R.M. Hermann and R.L. Somerville), pp351-378, Addison-Wesley, Reading, Massachusetts (1983)

Stanier, R.Y. *Ann. Rev. Microbiol* **5**, 53, (1951)

Strahs, D., and Brenowitz, M., *J. Mol. Biol.*, **244**, 494-510, (1994)

Sugimura, K., and Nishihara, T., *J. Bacteriol.*, **170**, 5625-5632, (1988)

Sunnaborg, A., Klumpp, D., Chung, T., and LaPorte, D.C. *J. Bacteriol,* **172**, 2642-2649 (1990)

Thiesen H-J., and Bach C., *Nucleic Acids Res.*, **18**, 3203-3209 (1990)

Thilveris, A.T., and Mount D.W., *Proc. Natl. Acad. USA,* **89**, 4500-4504 (1992)

Toledano, M.B., Kullik, I., Trinh, F., Baird, P.T., Schneider T.D., and Storz, G., *Cell,* **78**, 897-909, (1994)

Tuerk, C., and Gold, L., *Science,* **249**, 505-510 (1990)

Vershon, A.K., Liao, S-M., McClure, W.R., and Sauer, R., *J. Mol. Biol.,* **195**, 311-322 (1987)

Watson, J.D., and Crick F.H.C.; *Nature* **171**,737-738 (1953)

Whitson, P.A., Olson, J.S., and Matthews, K.S., *Biochemistry,* **25**, 3852-3858, (1986)

Wilson, D.S., Guenther, B., Desplan, C., and Kuriyan J., *Cell,* **82**, 709-719 (1995)

Winter, R.B., and vonHippel, P.H., *Biochemistry,* **20**, 6948-6960 (1981)

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., and Pabo, C.O.. *Cell,* **67**, 517-528 (1990)

Zhang, R.-G., Joachimiak, A., Lawson, C.L., Schevitz, R., Otiwinowski, Z., and Sigler, P.B. *Nature,* **327**, 591-597 (1987)

Zubay, G., Schwartz, D., Beckwith, *J. Proc. Natl. Sci. U.S.A.* **66**, 104 (1970)