Characterization of *katE* and its product, catalase HPII, from *Escherichia coli* by sequence analysis and site-directed mutagenesis

by

Ingemar Eric von Ossowski

A thesis
submitted to the Faculty of Graduate Studies
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Department of Microbiology
University of Manitoba
Winnipeg, Manitoba

ISBN 0-315-86063-4

Name Ingemar von Ossowski

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

Molecular Biology
SUBJECT TERM

0307
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.

SUJET

CODE DE SUJET

U·M·I

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-âtre 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
- PolymÇres 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 ther modynamique 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

CHARACTERIZATION OF katE AND ITS PRODUCT, CATALASE HPII, FROM *Escherichia coli* BY SEQUENCE ANALYSIS AND SITE-DIRECTED MUTAGENESIS

BY

INGEMAR ERIC VON OSSOWSKI

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

DOCTOR OF PHILOSOPHY

ABSTRACT

The complete nucleotide sequence of a 3,466 bp *PstI-ClaI* fragment containing the *katE* gene which encodes *E. coli* HPII catalase has been determined. An open reading frame of 2,259 bp encoding an amino acid sequence of 753 residues with a predicted N-terminal sequence in agreement with the results of direct amino acid sequencing was identified. A potential ribosome binding site 9 bp upstream and a transcriptional start site preceded by a putative promoter region similar to sequences preceding the KatF controlled *xthA* and *bolA* genes, 127 bp upstream of the initiation codon were identified. A preliminary comparison between the predicted sequence of HPII and the sequences of other catalases from various sources revealed significant sequence similarity as well as the conservation of several individual amino acids considered to play a role in the active site and in the binding of the heme group in bovine liver catalase.

Twenty catalase amino acid sequences from sources that include bacteria, fungi, animals, and plants, were compiled and used to reconstruct an evolutionary history. Phylogenetic reconstruction based on two different methods of tree building (neighbor-joining and parsimony) showed that animal and fungal catalases can be derived from a common ancestor. Bacterial catalases did not form a monophyletic group suggesting a polyphyletic origin. Plant

catalases apparently represent a distinct class of catalases that arose independently of other eukaryotes.

An investigation of the structure-function relationship of the catalase enzyme has been undertaken by testing the predicted roles for three active site residues and one proximal-side heme binding residue in HPII catalase. Site-directed mutagenesis was used to replace the active site residues His128 with Ala, Asn, Glu, and Gln, Ser167 with Ala, Thr, Cys, and Asn, and Asn201 with Ala, His, Asp, Gln, and Arg, and the proximal-side heme binding residue Tyr415 with Phe and His. Only the mutant enzymes H128A, H128N, N201A, N201H, N201D, and N201Q accumulated in amounts similar to the wild type enzyme. The His128 mutant enzymes were completely inactive and it was concluded that the His128 residue in HPII is essential to the catalytic mechanism. The Asn201 mutant enzymes were active, but at significantly lower levels than the wild type enzyme. They were used for activity based characterization and showed no change in maximal activity over a broad pH range indicative of the relatively pH independent nature of the catalytic mechanism. The sensitivity of the mutant enzymes to 2-mercaptoethanol and 3-amino-1,2,4-triazole varied with the residue substitution consistent with residue 201 sterically affecting the site of inhibitor interaction. It was concluded that the role of Asn201 in HPII is to enhance catalysis, but that the residue is not essential for

catalytic activity. No other mutant HPII protein including H128E, H128Q, S167A, S167T, S167C, S167N, N201R, Y415F, and Y415H accumulated in amounts sufficient for recovery indicating each replacement had in some way destabilized the protein structure.

to Debbie

you've been there with love and support
and made it worthwhile

ACKNOWLEDGEMENTS

The financial support of Dr. P.C. Loewen of the Department of Microbiology, University of Manitoba is gratefully acknowledged.

The postgraduate fellowship funding from NSERC and the University of Manitoba is also acknowledged.

Many thanks to my fellow graduate students for their friendship.

And a special thanks to my parents, Siegfried and Renate, for their never ending encouragement.

TABLE OF CONTENTS

Table of Contents (continued)

LIST OF FIGURES

List of Figures (continued)

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Absorbance
Amp^R	Ampicillin resistant
ATP	Adenosine 5′-triphosphate
AT	3-Amino-1,2,4-triazole
bp	Base pair(s)
BCIG	5-Bromo-4-chloro-3-indolyl-β-D-galactoside
Da	Dalton
dATP	2′-Deoxyadenosine 5′-triphosphate
DEAE	Diethylaminoethyl
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNaseI	DeoxyribonucleaseI
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
HPLC	High pressure liquid chromatography
IPTG	Isopropyl β-D-thiogalactoside
kb	Kilobase pair(s)
kDa	Kilodalton
Klenow	DNA polymerase I (Klenow fragment)
min	Minute
mRNA	Messenger ribonucleic acid
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
PEG	Polyethylene glycol
RNA	Ribonucleic acid

SDS	Sodium dodecyl sulfate
Tet^R	Tetracycline resistant
Tris	Tris(hydroxymethyl)aminomethane
U	Unit(s)
WT	Wild type
w/v	Weight per volume

INTRODUCTION

Reactive Oxygen Species

About two billion years ago the earth acquired an oxygenated atmosphere largely as a photosynthetic product resulting from the emergence of the first blue-green algae (Fridovich, 1978). Because of its paramagnetic nature which imposes a spin restriction with respect to molecular orbitals, molecular oxygen is a relatively unreactive molecule and does not display an overt toxicity (Demple, 1991; Farr and Kogoma, 1991). However, several active oxygen species of varying reactivity can be generated as by-products of normal aerobic respiration, and it is these reactive oxygen molecules that pose a threat to the viability of an organism. These active forms of oxygen include the superoxide anion radical, hydrogen peroxide, the hydroxyl radical, the hydroperoxyl radical, and singlet oxygen (Demple, 1991). As shown below, some of these oxygen molecules can be obtained through the sequential univalent reduction of molecular oxygen to water (Farr and Kogoma, 1991).

$$O_2 \xrightarrow{e^-} O_2^{\cdot} \xrightarrow{e^-,\ 2H^+} H_2O_2 \xrightarrow{e^-,\ H^+} OH\cdot \xrightarrow{e^-,\ H^+} H_2O$$

Oxygen	Superoxide Anion Radical	Hydrogen Peroxide	Hydroxyl Radical	Water

This four electron reduction of oxygen to water can also occur enzymatically, for example, when oxygen is employed as

the terminal electron acceptor in oxidative phosphorylation for the production of ATP through the electron transport chain (Farr and Kogoma, 1991). All of these active oxygen species, both individually and collectively, can cause specific damage to cellular macromolecules such as DNA, RNA, protein, and lipids, and have also been linked to several diseases such as rheumatoid arthritis, inflammatory bowel disorders, and atherosclerosis (Farr and Kogoma, 1991). Moreover, these oxygen molecules also appear to play a part in advancing mutagenesis, tumorigenesis, and aging (Farr and Kogoma, 1991). Consequently, survival in an oxygen rich environment meant that organisms had to rapidly evolve protective mechanisms to tolerate the level of oxidative stress brought on by the toxic oxygen species.

As described above, one of these toxic oxygen intermediates is hydrogen peroxide, and although it is one of the least potent of the reactive species of oxygen, hydrogen peroxide is still a strong oxidant and can be detrimental to the cell. For example, hydrogen peroxide can serve as a weak oxidizing agent attacking thiol groups of proteins and reduced glutathione. Hydrogen peroxide can also react with keto acids (Halliwell and Gutteridge, 1990) and can indirectly cause specific DNA strand breakage (Imlay and Linn, 1986). However, the exact nature of the effect of hydrogen peroxide on organic compounds remains poorly understood since hydrogen peroxide is able to rapidly react

with contaminating metals (iron and copper) to produce other more reactive forms of oxygen that then mask the actual role of hydrogen peroxide (Farr and Kogoma, 1991). Hydrogen peroxide can originate in the cell from numerous sources including as a product of oxidases such as the amino acid oxidases, as a photoproduct of near UV irradiation, and by the spontaneous and enzyme-catalyzed dismutation of the superoxide anion radical (Farr and Kogoma, 1991). Therefore, an example of one protective mechanism that must have developed early in evolution is a method to rid the cell of hydrogen peroxide. This is accomplished through the catalytic activity of the heme-enzyme, catalase. Catalase is ubiquitous in nature being found in virtually all aerobic organisms, from bacteria to man. However, it makes up only a small portion of the relatively complex and intricate defense system employed in the protection of organisms from toxic oxygen molecules.

Catalase Reactions

The catalase enzyme catalyzes a dismutation reaction in which hydrogen peroxide is decomposed and converted to the relatively innocuous products of molecular oxygen and water. The catalase reaction mechanism involves a two-electron oxidation-reduction of the heme prosthetic group by hydrogen peroxide and is carried out in two stages as shown below (Vainshtein et al., 1986).

$$H_2O_2 + Fe\text{-}E \text{ ----> } H_2O + O{=}Fe\text{-}E \text{ (Compound I)}$$

$$H_2O_2 + O{=}Fe\text{-}E \text{ ----> } H_2O + O_2 + Fe\text{-}E$$

In the first stage of the reaction, the first hydrogen peroxide reacts with the heme group forming a water molecule and the intermediate structure, compound I. Compound I, which contains an oxygen bound to the ferryl iron ($O{=}Fe^{IV}$) in the heme group, is an unstable structure and difficult to isolate. In the second stage of the reaction, compound I reacts with the second molecule of hydrogen peroxide to produce an oxygen molecule, a second water molecule, and the resting state of the enzyme. In the reaction, one molecule of hydrogen peroxide supplies hydrogen atoms while a second molecule accepts hydrogen atoms. Catalases efficiently carry out this reaction with a very high turnover number. For example, at a one molar concentration of hydrogen peroxide, one billion hydrogen peroxide molecules per active site per second can be decomposed by a typical catalase (Wang, 1955). As a result, catalases are very difficult to saturate with hydrogen peroxide (Beyer and Fridovich, 1988). Because hydrogen peroxide acts as the electron source, the reaction is independent of an external reducing source, and because the reaction is exothermic, ATP is not required (Farr and Kogoma, 1991). This suggests that catalases will protect the cell against hydrogen peroxide even during energy-depleted conditions.

In addition to catalyzing the catalatic reaction, some

catalases are capable of facilitating the peroxidatic reaction (i.e., $H_2O_2 + RH_2 \text{---}> 2H_2O + R$) with very low levels of hydrogen peroxide, but with a slower reaction rate. In this type of reaction, hydrogen peroxide is reduced to water by employing various short chain organic compounds (e.g., formate, methanol, ethanol, and phenol) as the electron donors (Beyer and Fridovich, 1988). This reaction is typically carried out by peroxidases which represent another enzyme employed in the detoxification of active oxygen species.

Eukaryotic and "Typical" Catalases

The catalase enzyme was first isolated with a high degree of purity from bovine liver extract and later crystallized (Sumner and Dounce, 1937). At the time, this also marked one of the first successful crystallizations of an intracellular enzyme. Catalase has subsequently been isolated and purified to homogeneity from several animal, plant, and microorganism sources. From physicochemical studies that followed, a general structure for catalase has emerged. Typically, catalases are structurally oligomeric consisting of four tetrahedrally arranged identical 60 kDa polypeptides such that the overall molecule has a molecular weight of 240 kDa (Deisseroth and Dounce, 1970). Each of the polypeptide subunits has bound a single protoheme IX group (ferric protoporphyrin IX) and they appear to function

independently of each other, giving catalase four separate active sites (Deisseroth and Dounce, 1970).

The heme prosthetic groups are not covalently bound to the subunits and are readily extracted from the protein with acidic acetone. In determining the three-dimensional structure of crystalline bovine liver catalase, Murthy et al. (1981) reported that the heme group is located 20 Å below the surface of the molecule in a hydrophobic pocket, only accessible through a hydrophobic channel. Two different subunits provide the residues that line this heme cavity. The significance of the association between the protein and the prosthetic group with respect to catalytic activity is demonstrated by a comparison of the relative rates for the enzymatic and non-enzymatic breakdown of hydrogen peroxide. For example, an intact catalase enzyme exhibits a turnover number (per sec) of 90,000 (in 10 mM H_2O_2), whereas turnover numbers for "prosthetic group analogs" such as hemin phosphate, ferric salts, and the iron chelate of triethylenetetramine are only 0.05, 1.0, and 22.7, respectively.

Some catalases, notably from bovine and human sources, also contain four molecules of tightly bound NADPH (Kirkman and Gaetani, 1984). The association of NADPH with catalase is not essential for enzymic activity, but appears to protect the enzyme from transformation into compound II by hydrogen peroxide (Kirkman et al., 1987). Recently, a

mechanism for NADPH inhibition of catalase compound II formation has been proposed which involves NADPH blocking the final formation of stable compound II by reacting as a 2-electron donor with a free radical generated during the initial formation of compound II from compound I (Hillar and Nicholls, 1992). Binding sites for NADPH on bovine liver catalase have been established (Fita and Rossmann, 1985b). Carbohydrate moieties have been found in certain mammalian catalases, suggesting a glycoprotein nature (Pegg et al., 1986).

The complex structural arrangement of catalase suggests that the biosynthesis of catalase is likely to occur in stages. For rat liver catalase, it has been proposed (de Duve, 1974) that the synthesis of the enzyme occurs in the following three steps: (i) the generation of the 60 kDa apoenzyme subunits, (ii) the intercalation of the prosthetic groups with the individual subunits, and (iii) the assembly of the subunits into the tetrameric structure. Murthy et al. (1981) had concluded that the biosynthesis of bovine liver catalase also involves a series of steps in which a holo-monomer is first generated, followed by the formation of a holo-dimer and a holo-tetramer in succeeding order.

The catalase enzyme is generally considered a stable enzyme. The assembly of the subunits into the quaternary structure is not dependent upon the presence of disulfide linkages (Mörikofer-Zwez et al., 1969) or covalent bonding

(Schonbaum and Chance, 1976), but rather on hydrophobic interactions (Furuta et al., 1974). Irreversible denaturation involving exposure to rather harsh conditions is normally necessary for tetramer dissociation. For example, treatment of catalase with an extremely alkaline pH (pH= >10) can cause the enzyme to dissociate into the individual subunits. Likewise, a highly acidic pH (pH= <3) can cause the prosthetic group to dissociate from the protein (Schonbaum and Chance, 1976). Other conditions that dissociate the oligomeric structure include detergent treatments, chemical modification of the apoenzyme, and storage, dilution, and freeze drying (Schonbaum and Chance, 1976; Beyer and Fridovich, 1988).

A number of compounds are known to affect catalase activity. For example, cyanide and azide have been reported to reversibly inhibit catalase activity by ligating at the heme iron (Beyer and Fridovich, 1988). Several thiol agents such as 2-mercaptoethanol have also been reported to inhibit the activity of catalase (Takeda et al., 1980). The compound 3-amino-1,2,4-triazole (AT) has been documented as a potent inhibitor of catalase activity (Margoliash and Novogrodsky, 1958) in catalases from numerous sources (Margoliash et al., 1960). The inhibition of human erythrocyte catalase activity by AT required low hydrogen peroxide concentrations (Margoliash and Novogrodsky, 1958). AT is believed to interact with the active site histidyl

residue located near the heme group thereby hindering further hydrogen peroxide binding and reaction of the compound I intermediate (Reid et al., 1981). Darr and Fridovich (1985) have reported that the compound 3,3′-diamino-benzidine is also a strong inhibitor and can cause reversible or irreversible inhibition depending upon the presence of hydrogen peroxide. The superoxide radical (Kono and Fridovich, 1982) and low concentrations of hydrogen peroxide (Kirkman and Gaetani, 1984) have also been demonstrated to inactivate catalase.

Structure of "Typical" Catalases

Since the first crystallization of bovine liver catalase by Sumner and Dounce (1937), other reports of crystalline catalases from a variety of sources have appeared. Several have been used for structural investigations by X-ray diffraction and electron microscopy. For example, the spatial organization of three tetrameric catalases from bovine liver, *Penicillium vitale*, and *Micrococcus lysodeikticus* have been determined at a high resolution (Murthy et al., 1981; Vainshtein et al., 1986; Murshudov et al., 1992). One of the best studied crystalline forms of catalase is the bovine liver enzyme (Eventoff et al., 1976), and therefore provides much of the detailed structural information on the catalases.

The three-dimensional structure of bovine liver

catalase has been established to within a 2.5 Å resolution (Murthy et al., 1981; Reid et al., 1981). Murthy et al. (1981) proposed that the tertiary structure of the 506 residue primary sequence (Schroeder et al., 1982a) can be divided into four domains. The first domain represents an extended non-globular amino terminal arm of about 70 residues that interacts with neighboring subunits to stabilize the quaternary structure. The second domain consists of an antiparallel, eight-stranded β-barrel (residues 76-320) containing the heme binding residues. The third domain (residues 321-436), known as the wrapping domain, is a structurally random region that wraps around the exterior of the subunit. The fourth and final domain encompasses the carboxy-terminal region (residues 437-506) and consists of four α-helices positioned on the outside of the subunit.

Efforts to elucidate the amino acid sequence of the catalase subunits were first carried out on bovine liver and erythrocyte catalases by rather laborious chemical means involving enzymatic hydrolysis methods (Schroeder et al., 1982a; 1982b). With the advances in molecular biology methodologies the determination of amino acid sequences from cloned DNA sequences became routine resulting in several catalase sequences becoming available for comparison. In general, the catalases represent a rather well-conserved group of proteins (Okada et al., 1987) exhibiting the

greatest degree of similarity in the region encompassing the β-barrel (domain two) and less similarity towards the carboxy-end of the protein.

Fita and Rossmann (1985a) examined the refined structure of bovine liver catalase for potential catalytic mechanisms and proposed a model describing functional roles for amino acids previously identified by Murthy et al. (1981) as catalytically significant. Three active site residues have been identified. These include His74 and Asn147, which are located on the distal side of the heme and are involved in substrate binding and catalysis, and Ser113, which forms hydrogen bonds with the His74 residue and a water molecule while the enzyme is in the resting state and unchallenged by hydrogen peroxide. Several other residues are involved in heme binding. Residues located on the distal side of heme include Val73, Thr114, Phe152, and Phe160, while residues found on the proximal side include Pro335 and Arg353. Tyr357 is also located on the proximal side of heme and serves as the fifth ligand for the heme iron with its phenolic oxygen in a deprotonated state. The predicted roles of the various amino acids involved in the active site and heme binding have yet to be tested by specific alteration of residues through site-directed mutagenesis. Probing structure-function relationships by site-directed mutagenesis has been conducted on various other heme proteins and has been thoroughly reviewed by

Dawson (1988) and Stayton et al. (1989).

From their study on the three-dimensional structure of bovine liver catalase, Fita and Rossmann (1985a) have proposed a pathway for the catalase mechanism. In their model, the first hydrogen peroxide molecule entering the heme pocket is sterically forced to make contact with the His74 and Asn147 residues while simultaneously interacting with the heme iron resulting in polarization of the peroxide O-O bond. Hydrogen bonding of the hydrogen peroxide to R-groups of His74 and Asn147 and electrostatic interactions of the oxygen(2) to the nearby heme iron (Fe^{3+}) reduce the pKa value of the peroxide OH group, thus weakening the O(2)-H bond. The stretching of the O(2)-H bond by the attraction of O(2) to the heme iron orientates O(1) in a position for hydrogen bonding with the histidyl-imidazole group. This suggests that proton transfer from O(2) to O(1) of the peroxide is facilitated by the imidazole side chain of His74. Once the O(2) is deprotonated, the interaction between O(2) and heme iron will be strengthened which causes the O(2)-O(1) bond to stretch and delocalize the negative charge on O(1). The continued interaction between O(2) and heme iron causes a free ligand orbital in O(1) to appear which may then be occupied by the transferred proton. As a positive charge develops on the heme group, the O-O bond will break resulting in O(2) transfer to the iron to generate compound I and one water molecule. Hydrogen

bonding between the Asn147 amide group and O(1), may also assist in the cleavage of the O-O bond. This completes stage one of the dismutation reaction. For the reduction of compound I in stage two, the second hydrogen peroxide molecule forms hydrogen bonds with the same two active site residues (His74 and Asn147) resulting in one hydrogen of the peroxide interacting with the oxo-oxygen of compound I. The transfer of this hydrogen molecule to the oxo-oxygen coupled with the transfer of a second hydrogen from the imidazole group of His74 completes the reaction so that an oxygen molecule and an additional water molecule are generated.

Other "Atypical" Catalases

Although the most common form of the typical catalase consists of the tetrameric structure containing protoheme as earlier described, several other variant forms of catalase have been reported. For example, catalases produced by *Escherichia coli* (HPII) (Loewen and Switala, 1986) and *Bacillus subtilis cat-1* (Loewen and Switala, 1988) are arranged in a hexameric structure composed of larger subunits. The fungus *Neurospora crassa* (Jacob and Orme-Johnson, 1979a) and *E. coli* (Loewen and Switala, 1986) produce catalases containing a heme *d*-isomer rather than the protoheme IX moiety found in the typical catalase. *E. coli* (Claiborne and Fridovich, 1979) and *Rhodopseudomonas capsulata* (Hochman and Shemesh, 1979) also produce a

tetrameric enzyme consisting of larger-sized subunits that is a bifunctional catalase-peroxidase. *Lactobacillus plantarum* (Kono and Fridovich, 1983) and *Thermoleophilum album* (Allgood and Perry, 1986) produce non-heme but manganese-containing catalases while *Klebsiella pneumoniae* (Goldberg and Hochman, 1989) produces a catalase with a dimeric structure.

The two catalases produced by *E. coli*, labeled hydroperoxidase I (HPI) and hydroperoxidase II (HPII), have been extensively studied revealing that each enzyme has unique structural properties, making them quite different from the typical catalases. The following sections review some of these properties.

Catalase HPI of *E. coli*

HPI catalase is a tetrameric enzyme composed of identical 80 kDa polypeptides (Claiborne and Fridovich, 1979) making it larger than the typical catalase. Only two protoheme IX moieties are bound to the HPI tetramer, half the number associated with typical catalases. As with a few of the typical catalases, HPI is a bifunctional catalase exhibiting an associated peroxidase activity (Claiborne, 1978). However, unlike most typical catalases, HPI displays a broad spectrum peroxidase activity where, in addition to oxidizing short chain organic alcohols and acids, the enzyme can oxidize compounds such as pyrogallol and dianisidine.

Optimum activities for the catalatic and peroxidatic reactions by HPI occur at pHs 7.5 and 6.5, respectively (Claiborne and Fridovich, 1979). Other properties of HPI catalase which incidentally are not shared by the typical catalase include heat lability at 70°C, a pH optimum of 6.8, and a resistance to inhibition by aminotriazole (AT) (Meir and Yagil, 1985). An interesting property of HPI is its ability to be separated electrophoretically into two isoforms identified as HPI-A and HPI-B (Loewen et al., 1985a; Meir and Yagil, 1985). Little information has been obtained on these two forms of HPI and consequently, their difference, if any, remains unknown.

The structural gene encoding the HPI subunit (*katG*), has been mapped at 89.2 minutes on the *E. coli* chromosome (Loewen et al., 1985b), cloned (Loewen et al., 1983), characterized (Triggs-Raine and Loewen, 1987), and sequenced (Triggs-Raine et al., 1988). An open reading frame of 2,181 bp encodes a 726 residue amino acid sequence that shares no similarity with any known catalase primary sequence except KatG (HPI) from *Salmonella typhimurium* (Loewen and Stauffer, 1990). However, *katG* is 48% similar to the *Bacillus stearothermophilus* peroxidase gene (*perA*) (Loprasert et al., 1989). This suggests that the bifunctional HPI is more closely related to the peroxidase family of enzymes than to the family of catalases, and perhaps, has evolved from a peroxidase enzyme (Loewen, 1992). This theory is further

supported by a recent sequence comparison (Welinder, 1992) which shows that HPI bears resemblance to the plant peroxidase family, and that the catalase activity associated with the bifunctional HPI may be merely an ancillary property of the active site.

This latter observation of Welinder (1992) will make it possible to carry out a more directed approach to studying the structure-function relationship of HPI. In the past, because HPI shared no similarity to the catalases for which the three-dimensional structure has been determined and because HPI itself had not been crystallized, the only available approach to investigate structure-function relationships was to carry out random mutagenesis followed by the identification of regions of the HPI protein affecting enzyme activity (Loewen et al., 1990). Similarity between HPI and plant peroxidases, for which three-dimensional information is available, will allow the alteration of specific amino acids using site-directed mutagenesis.

The synthesis of HPI catalase in *E. coli* increases during the exponential phase of cell growth and then declines during the stationary phase, but can also be induced by hydrogen peroxide and the hydrogen peroxide-generating compound, ascorbate (Richter and Loewen, 1982; Loewen et al., 1985a). The fact that hydrogen peroxide elicits production of HPI is not surprising considering the

protective role of catalase against hydrogen peroxide. HPI catalase is just one of at least 34 proteins whose synthesis is induced by the presence of hydrogen peroxide. The synthesis of 9 of these proteins, including HPI, alkyl hydroperoxidase, and glutathione reductase, are positively controlled by the *oxyR* gene product as part of the *oxyR* regulon (Christman et al., 1985; Morgan et al., 1986). A more detailed description of the *oxyR* regulon and its role in the protection of *E. coli* from oxidative stress can be found in recent reviews (Demple, 1991; Farr and Kogoma, 1991).

The *oxyR* gene encodes a 34.4 kDa protein whose 305 amino acid sequence displays significant sequence similarity to a group of bacterial regulatory proteins collectively known as the LysR family (Christman et al., 1989; Tao et al., 1989). The majority of these regulatory proteins function as transcriptional activators, but some also behave as repressors (Henikoff et al., 1988). In addition to activating transcription from the *katG* promoter, OxyR can also negatively control its own synthesis (Christman et al., 1989).

The link between hydrogen peroxide induction of HPI synthesis and OxyR regulation of *katG* expression has been studied by Ames and his colleagues. They have proposed that the oxidation state of the OxyR protein rather than just an increase in protein levels is the determining factor in *katG*

expression (Storz et al., 1990). In response to hydrogen peroxide, the OxyR protein is oxidized and changes conformation in such a way that it is able to activate transcription from the *katG* gene. In the reduced state the conformation of the OxyR protein is altered so that it can no longer activate transcription. These conformational changes likely affect the manner in which the OxyR protein can or cannot interact with the corresponding promoter regions of the gene but the nature of this interaction has not yet been elucidated. The OxyR protein is therefore a transcriptional activator following activation by an oxidant, hydrogen peroxide allowing it to act as both a sensor and transducer of an oxidative stress signal (hydrogen peroxide) within a global stress response (Farr and Kogoma, 1991; Loewen, 1992).

Catalase HPII of *E. coli*

HPII catalase is one of the larger-sized catalases owing to its unique quaternary structural arrangement and subunit size. HPII is a hexameric enzyme composed of identical polypeptide subunits with an estimated size of 93 kDa (Loewen and Switala, 1986). Each of the six subunits contains a *cis* heme d prosthetic group (Chiu et al., 1989). This heme d-isomer confers a characteristic green color on the enzyme and has been also identified in the green catalase of *Neurospora crassa* (Jacob and Orme-Johnson,

1979b). HPII functions only as a catalase, is unable to catalyze the peroxidatic reaction, and like typical catalases, is heat stable retaining activity at temperatures up to 70°C (Loewen and Switala, 1986). The stability of HPII is also demonstrated by its retention of activity in 0.1% sodium dodecyl sulfate or 7M urea, and over a broad pH range from pH 4 to 11 (Loewen and Switala, 1986). Also like typical catalases, but unlike HPI catalase, HPII is inhibited by AT.

The unique structural properties of HPII catalase, including its hexameric arrangement, large subunit size, and unusual prosthetic group, make it attractive for a more detailed structural investigation which is being pursued with the report of a preliminary X-ray diffraction analysis (Tormo et al., 1990). Further crystallographic investigation is underway and may eventually provide greater insight into the apparent structural differences between the hexameric HPII and the tetrameric typical catalases.

HPII catalase synthesis does not respond to hydrogen peroxide, unlike HPI catalase, but is induced as *E. coli* enters stationary phase. HPII synthesis can also be induced during exponential phase, by growth on tricarboxylic acid cycle intermediates (Loewen et al., 1985a), and by aromatic acids (Mulvey et al., 1990). The HPII subunit is encoded by *katE* which has been mapped at 37.8 minutes on the *E. coli* chromosome (Loewen, 1984). *katE* has been cloned (Mulvey et

al., 1988) and this thesis describes subsequent work in its characterization.

Unlike *katG* which is regulated by *oxyR*, *katE* expression is controlled by *katF* (also known as *rpoS*, *appR*, and *csi2*), mapping at 59 minutes on the chromosome, which is essential for *katE* expression (Loewen and Triggs, 1984). The *katF* gene has been cloned (Mulvey et al., 1988) and sequenced (Mulvey and Loewen, 1989) revealing a 1,086 bp open reading frame which encodes a 41.5 kDa protein. The amino acid sequence of the *katF* gene product (the KatF protein) closely resembles the predicted sequences of several known sigma transcription factors (Mulvey and Loewen, 1989), and the protein functions as a positive effector of *katE* expression by possibly assisting in RNA polymerase binding to the *katE* promoter.

In addition to regulating *katE* expression, KatF has been shown to control positively the genes for as many as 35 proteins including *xthA* encoding exonuclease III (Sak et al., 1989), *bolA* affecting cell morphology (Lange and Hengge-Aronis, 1991a), *appA* affecting an acid phosphatase and alternate cytochrome oxidase (Lange and Hengge-Aronis, 1991b), *glgS* affecting glycogen synthesis (Hengge-Aronis and Fischer, 1992), and the *pex* genes induced during starvation (McCann et al., 1991). However, the identity for most *katF*-regulated genes remain to be determined. Unlike the majority of bacterial genes which exhibit a decrease in

expression in response to nutrient limiting conditions (Siegele and Kolter, 1992), the various genes under KatF control show an increase in expression during starvation conditions and maximal expression in stationary phase (Mulvey et al., 1990; Lange and Hengge-Aronis, 1991b). Consistent with this and the roles of the numerous *katF*-regulated genes is the observation that *katF* mutants are more susceptible to death during starvation (Mulvey et al., 1990), presumably because they lack: resistance against stationary phase stresses including thermal, oxidative, and osmotic (Lange and Hengge-Aronis, 1991b; McCann et al., 1991); the ability to accumulate glycogen and trehalose (Lange and Hengge-Aronis, 1991b; Hengge-Aronis et al., 1991); and the ability to carry out the rod-to-sphere morphology conversion of cells during the transition to stationary phase (Lange and Hengge-Aronis, 1991a). Consequently, this has led to the suggestion that the *katF*-regulated genes represent a regulon responding to starvation conditions preparing *E. coli* for survival in nutrient-poor conditions (Loewen, 1992; Hengge-Aronis, 1993). The identity of the cellular signal responsible for inducing this regulon is still largely unknown. However, the results of several studies on the regulation of *katF* and *katE* expression (Schellhorn and Hassan, 1988; Mulvey et al., 1990; Schellhorn and Stones, 1992) suggest that certain weak acids, particularly acetate (Schellhorn and Stones, 1992),

are effective inducers and lead to the proposal that intracellular proton accumulation coupled to an internal pH reduction facilitates expression of *katF*-regulated genes. HPII catalase is just one of a large group of proteins controlled by KatF, and the likely primary role of HPII is to protect the cell from hydrogen peroxide during dormancy (Loewen, 1992).

The work described in this thesis focuses on the further characterization of *katE*. The sequence analysis is reported. The predicted amino acid sequence of HPII is described and compared to the sequences of other catalases from prokaryotic and eukaryotic origins. The predicted amino acid sequence has also allowed the initiation of a structure-function study of HPII, targeting specific residues by site-directed mutagenesis.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Plasmids, and Bacteriophage.

E. coli strain NM522 (Table 1) served as the host for all plasmids, routine cloning, and the generation of single-stranded DNA. *E. coli* strain UM255 (Table 1) was used for mutant *katE* expression and the subsequent production of mutant HPII proteins. *E. coli* strain CJ236 (Table 1) served as the host of phagemids for the generation of uracil-containing single-stranded DNA for use in site directed *in vitro* mutagenesis.

Plasmid pAMkatE72 (Table 1) was the source of the *katE* gene. The M13 Bluescript phagemids pKS+, pKS-, pSK+, and pSK- (Table 1) were used for subcloning, sequencing, and *in vitro* mutagenesis. The operon fusion vector pRS415 (Table 1) was used in constructing promoter fusions of the regulatory region of the *katE* gene.

The helper bacteriophage R408 (Table 1) was employed in the generation of single-stranded DNA template for use in DNA sequencing and *in vitro* mutagenesis.

Growth Media.

LB (Luria-Bertani) Medium (Miller, 1972)

Per litre:	Bacto-tryptone	10 g
	Bacto-yeast extract	5 g
	NaCl	5 g
For plates:	Bacto-agar	10 g

This type of growth medium in liquid and solid form was

Table 1. Bacterial strains, plasmids, and bacteriophage.

	Genotype or Characteristics	Source
Strains		
NM522	*supE thi* Δ(*lac-proAB*) *hsd-5* [F′ *proAB lacI*^{Q} *lacZ*Δ*15*]	Mead et al. (1985)
UM255	*pro leu rpsL hsdM hsdR endI lacY katG2 katE12*::Tn*10 recA*	Mulvey et al. (1988)
CJ236	*dut1 ung1 thi1 relA1*/ pCJ105 (*cam*^{R} F′)	Kunkel et al. (1987)
Plasmids		
pAMkatE72	Amp^{R}	Sorby (1989)
pRS415	Amp^{R} Tet^{R}	Simons et al. (1987)
pKS+, pKS-, pSK+, and pSK-	Amp^{R}	Stratagene Cloning Systems
Bacteriophage		
R408 (Helper Phage)		Stratagene Cloning Systems

routinely used for culturing all *E. coli* strains utilized in this study.

Ampicillin was supplemented at 250 μg/ml for the liquid medium and at 100 μg/ml for the solid medium to select for Amp^R plasmid-bearing cells.

Chloramphenicol was added at 40 μg/ml for the liquid medium and at 5 μg/ml for the solid medium to maintain the presence of the F′ episome in the CJ236 strain.

The color selection of subclones was accomplished by spreading 50 μl aliquots of 2% 5-bromo-4-chloro-3-indolyl-β-D-galactoside (BCIG) in NN dimethyl formamide and 200 mM isopropyl thio-β-D-galactoside (IPTG) in distilled water onto LB-ampicillin medium agar plates.

Glucose-minimal Medium (Miller, 1972)

Per litre:		
	$Na_2HPO_4 \cdot H_2O$	6 g
	KH_2PO_4	3 g
	NaCl	0.5 g
	NH_4Cl	1 g
	Bacto-agar	10 g

Autoclaved, and the following sterilized solutions added:

1 M $MgSO_4$	1 ml
Vitamin B_1 (1 μg/ml)	1 ml
Trace Elements	1 ml
30% Glucose	10 ml

Antibiotics were added as required.

The solution of trace elements is comprised of the following per litre: 2.5 g $FeSO_4 \cdot H_2O$, 2.9 g H_3BO_4, 1.2 g $CoSO_4 \cdot 7H_2O$, 0.1 g $CuSO_4 \cdot 5H_2O$, 0.09 g $MnCl_2 \cdot 4H_2O$, 2.5 g $Na_2MoO_4 \cdot 7H_2O$, 2.1 g $ZnSO_4 \cdot 7H_2O$, and 5 ml conc. H_2SO_4.

Glucose-minimal medium agar plates were used to grow the NM522 strain to ensure the presence of the F′ episome required for phagemid induced single-stranded DNA production.

* All media and solutions were made with distilled water unless otherwise stated.

Culturing and Storage Conditions.

All *E. coli* strains, in liquid and on solid media, were grown at 37°C unless stated otherwise. Liquid cultures were grown with vigorous aeration on reciprocating and gyrotory shakers. Estimates of the cell density were made with a Klett-Summerson colorimeter containing a blue filter.

Stock cultures for long term storage were kept in 50% glycerol at -20°C and 8% DMSO at -60°C.

Restriction Nuclease Digests.

All restriction nucleases and buffers used in this study were obtained from GIBCO-BRL, Pharmacia, or Boehringer-Mannheim. Restriction digestions were carried out at 37°C for 2-3 hours unless the enzyme required otherwise, e.g., *Sma*I at 30°C. Restriction reactions were carried out in total volumes that varied from 10-50 μl, and contained 5-10 μg DNA, ~10 units restriction nuclease, and the optimal buffer for the enzyme(s) in use. All restriction reactions were brought up to the desired volume

with sterilized distilled water.

Agarose Gel Electrophoresis.

Agarose gel electrophoresis was employed in the analysis of restriction nuclease digested DNA samples in the manner described by Sambrook et al. (1989). Generally, 0.7% agarose gels, prepared with TAE (40 mM Tris-acetate and 1 mM EDTA, pH 8) buffer and with dimensions of 15 cm x 15 cm (or 15 cm x 20 cm), were run submerged in a Bio-Rad Subcell Horizontal Slab Gel system containing TAE buffer. (1.2% agarose gels were used for the separation of smaller sized DNA fragments). The voltage was typically set at 1-10 volts/cm and the extent of the separation monitored by the migration of the bromophenol blue dye from the loading buffer. DNA was located by placing the agarose gel in TAE buffer containing 0.5 μg/ml ethidium bromide for 10-20 minutes and then exposing the gel to ultraviolet light illumination. A permanent record of the banding pattern was subsequently obtained by photographing the agarose gel using Polaroid Type 667 film. Sizing of DNA fragments was accomplished using 1 kb ladder DNA sizing markers (GIBCO-BRL).

DNA Ligations.

DNA ligations were carried out according to the protocol described by Sambrook et al. (1989). Insert and

vector DNAs were mixed at a ratio where the amount of insert DNA always exceeded that of the vector DNA (about 5:1). All restriction nucleases used to cut the vector DNA at the desired restriction site(s) were inactivated by a 10-15 minute incubation at 65°C prior to the ligation. DNA fragments to be used for subsequent insertion into a vector were extracted from agarose gels. These samples did not require a heat inactivating step prior to electrophoretic separation. Ligation reactions were generally carried out in a total volume of 15 μl with 1-3 units of T_4 DNA ligase (Pharmacia or GIBCO-BRL) and ligation buffer diluted five-fold from a 5x stock solution [250 mM Tris-HCl (pH 7.6), 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, and 25% (w/v) PEG 8000]. Ligations were typically carried out at room temperature for 2-4 hours and then placed on ice until further use.

Transformations.

Transformation of *E. coli* with cloned versions of the Bluescript phagemids and the plasmid pRS415 was carried out as described by Chung et al. (1989). In this method exponential phase cells were harvested and made competent for transformation by a calcium chloride treatment. DNA samples were combined with the competent cells and incubated at 0°C followed by a 90 sec heat shock at 42°C. Following recovery at 37°C in LB medium, the cells were plated onto solid medium containing a selective antibiotic and allowed

to grow overnight.

Recombinant Selection.

Two types of plasmids, the M13 Bluescript phagemid and the operon fusion vector pRS415, were used in this study and both shared similar modes of recombinant selection.

The Bluescript phagemid (2.95 kb), which was principally used for subcloning of the *katE* gene and the production of single-stranded DNA for sequencing and *in vitro* mutagenesis, contains a multiple cloning region adjacent to a *lacZ* gene fragment (which encodes part of the β-galactosidase enzyme), as well as the gene conferring ampicillin resistance. Cells harboring phagemids with inserts were selected based on the ability to grow on LB medium supplemented with ampicillin and the inability to produce a blue colour when grown on LB medium containing BCIG and IPTG.

The plasmid pRS415 (10.8 kb), employed in the analysis of 5′ end deletions of the *katE* gene, contained the ampicillin resistance gene and a promoterless *lacZ* gene preceded by a multiple cloning region. Cells with plasmids were grown on LB medium containing ampicillin and the plasmids containing the desired insert screened by restriction digest analysis. The fusion plasmids were then characterized using the β-galactosidase assay.

For both vector systems, confirmation of insert

containing plasmids was accomplished by plasmid isolation followed by restriction nuclease mapping.

Isolation and Purification of DNA.

Small and large quantities of plasmid DNA were obtained using two different methods. Isolation of smaller amounts of plasmid DNA as required in the screening of transformants was carried out according to the rapid plasmid DNA preparation protocol of Morelle (1989). Larger amounts of plasmid DNA were recovered using the method of Birnboim and Doly (1979). The isolation and purification of single-stranded M13 phagemid DNA required for DNA sequencing and *in vitro* mutagenesis was carried out as described by Vieira and Messing (1987).

DNA fragments produced from restriction digestions were purified from agarose electrophoretic gels using the GENECLEAN DNA extraction kit (Bio/Can Scientific Inc.). The region of agarose containing the DNA fragment for extraction was identified by ethidium bromide staining and excised from the gel. DNA was then extracted from the agarose using the kit according to the protocol provided by the manufacturer.

DNA Sequencing.

The dideoxy chain termination method of DNA sequencing described by Sanger et al. (1977) was used in the sequence determination of the *katE* gene, in identifying and

confirming base changes created by *in vitro* mutagenesis, and for a sequencing ladder in the determination of the transcriptional initiation site by primer extension analysis. All sequencing was carried out using single-stranded DNA template, [α^{32}-P]dATP (DuPont), and Klenow fragment (Pharmacia). The preparation of sequencing reagents and buffers, the taping and sealing of glass plates, pouring of gel mixtures, assembly and dismantling of sequencing apparatus, preparation and loading of samples, electrophoretic run, and exposure of gel to X-ray film were carried out as described by Sambrook et al. (1989).

Preparation and Isolation of Oligonucleotide Primers.

Oligonucleotides required for DNA sequencing, *in vitro* mutagenesis, and primer extension analysis were synthesized on a PCR-Mate Synthesizer (Applied Biosystems Inc.). Primers for *in vitro* mutagenesis were synthesized with a phosphate group attached at the 5′ end. The extraction of oligonucleotides was carried out according to the method suggested by Applied Biosystems Inc. Bound oligonucleotides were cleaved from the synthesis cartridge with concentrated ammonium hydroxide and incubated at 55°C overnight. The sample was then diluted with HPLC (high pressure liquid chromatography) distilled water, lyophilized, ethanol precipitated, washed with 70% ethanol, and finally resuspended with HPLC distilled water for storage at -40°C.

5′ end mRNA Mapping.

The identification of the transcriptional initiation site for the *katE* gene was accomplished with the primer extension method described by Ausubel et al. (1989). An oligonucleotide primer with a sequence complementary to a region of the *katE* gene encoding the amino terminus of the HPII primary sequence was annealed to the corresponding katE mRNA and elongated by avian reverse transcriptase (Pharmacia). 5′ end labelling of the oligonucleotide primer with [γ-P^{32}]ATP for the primer extension analysis was carried out as described by Maniatis et al. (1982). Total RNA was prepared from NM522 harboring the plasmid pAMkatE72 and isolated as described by Gilman and Chamberlin (1983), with the following procedural modifications. Cells were ruptured by extracting twice with an equal volume of redistilled phenol and the subsequent nucleic acid precipitate was redissolved in a 100 mM sodium acetate buffer (pH 5) containing 10 mM magnesium sulfate for the treatments with DNaseI and proteinase K.

***In vitro* Mutagenesis.**

The introduction of specific base changes for creating new restriction nuclease sites and amino acid substituted protein mutants was carried out according to the *in vitro* mutagenesis methodology described by Kunkel et al. (1987). Bluescript phagemids (rather than M13 bacteriophage)

containing the *katE* gene (or portions thereof) were propagated in the *dut⁻ ung⁻* CJ236 strain to generate uracil-containing single-stranded DNA template. Mutagenesis was carried out by annealing the phosphorylated mutagenic oligonucleotide primers to the template and a complementary DNA strand synthesized using unmodified T_7 DNA polymerase (New England Biolabs) and T_4 DNA ligase (Pharmacia). The double-stranded products were transformed into the NM522 strain so that the uracil-containing strand was degraded, and a new non-uracil-containing DNA strand generated. Transformants were then screened for anticipated base changes by single-stranded DNA sequencing.

Mutant HPII Protein Production and Purification.

To produce and isolate amino acid substituted HPII catalases devoid of the wild type form, plasmids carrying the mutant *katE* gene were transformed into the *katE⁻ katG⁻* strain UM255 and grown in LB medium supplemented with ampicillin for 16 hours at 37°C with shaking. Cells were recovered and mutant HPII protein isolated and purified according to Loewen and Switala (1986), but with DEAE cellulose DE52 (Whatman) rather than DEAE Sephadex A-25. All purification steps were carried out at 5°C. A crude extract was prepared using a French press at 20,000 psi and then precipitated using 2.5% streptomycin sulfate as described by Claiborne and Fridovich (1979), except with a

buffer that included 1 mM EDTA. Ammonium sulfate precipitation was carried out at 30, 40, 50, and 60% saturation with solid ammonium sulfate with gentle stirring for 30 minutes at each step to ensure complete mixing and precipitation. Centrifugation was used to collect the precipitate and additional ammonium sulfate added to the supernatants. HPII protein was isolated in the 50 and 60% pellets, which were then mixed together and fractionated once again with ammonium sulfate. The pooled ammonium sulfate fractions were dialyzed against 50 mM potassium phosphate buffer (pH 7) and then loaded onto a 2 x 15 cm column of DE52 DEAE cellulose equilibrated in the same buffer. The column was washed with the equilibrating buffer until the absorbance of the column fractions at 280 nm was below 0.05. The HPII protein was eluted by a 500 mM NaCl linear gradient in the same buffer. After this initial ion exchange chromatography step, the purified preparations of HPII protein were examined by SDS-polyacrylamide gel electrophoresis (Weber and Osborn, 1969) and usually found to be near homogeneity, *ca.* 85-90% pure. However, if further purification was warranted, either the DEAE-cellulose chromatography was repeated or a gel filtration through Bio-Gel A-1.5m was used. HPII protein was concentrated by ultrafiltration, dialyzed against 50 mM sodium phosphate (pH 7), and stored at -20°C. For mutant HPII protein with minimal or no catalase activity, the use

of the catalase assay to monitor the purification procedure was not feasible and SDS-polyacrylamide gel electrophoresis was used instead.

Following purification, the absorption spectrum of each of the mutant proteins was obtained using a Milton Roy MR3000 Spectrophotometer following dilution in 50 mM potassium phosphate buffer (pH 7).

β-Galactosidase and Catalase Assays.

Levels of β-galactosidase activity were assayed according to Miller (1972), but with the following two changes in the protocol. Cells were lysed by the addition of 20 μl of chloroform and a 4 minute incubation at 28°C. β-galactosidase activity was recorded in Miller Units which are proportional to the increase in *o*-nitrophenol per minute per bacterium.

HPII catalase activity was determined in a Gilson oxygraph equipped with a Clark electrode as described by Rørth and Jensen (1967). One unit of catalase activity is the amount of enzyme that decomposes one μmol of hydrogen peroxide per minute at 37°C.

Nucleic Acid and Protein Determinations.

Spectrophotometric measurements were used to determine nucleic acid and protein concentrations as described by Sambrook et al. (1989) and Layne (1957), respectively.

Phylogenetic Analysis.

An estimate of the evolutionary relationship of the 20 aligned catalase sequences was made using both parsimony and distance methods. Unrooted parsimony trees were constructed using programs within the PHYLIP package (Version 3.4, Felsenstein, 1989). The bootstrap analysis method (Felsenstein, 1985) was used to estimate the confidence levels of potential monophyletic clusterings of more than one species. The program SEQBOOT was used to obtain 100 bootstrap replicates, each of which was analyzed by the parsimony program PROTPARS. The program CONSENSE was used to construct a majority-rule consensus tree based on the 100 bootstrap replicates. Felsenstein (1985) reported that a consensus tree generated from a bootstrap analysis represents an overall estimate of a phylogeny and can be considered statistically relevant when tree branches are supported by greater than 95% of the bootstrap replicates. Programs within the CLUSTALV package (Higgins and Sharp, 1989) were used in estimating phylogenies for the catalase sequences based on distance criteria. The neighbor-joining method (Saitou and Nei, 1987) was employed in the construction of unrooted distance trees and estimating the level of confidence for the tree branches determined from 100 bootstrap replicates as derived from the "bootstrap tree" option.

Computer Programs.

The Microgenie (Beckman) and PC/gene (IntelliGenetics Inc., Mountain View, California) software packages were employed in the analysis of the *katE* sequence which included prediction of the amino acid sequence and secondary structures, determination of codon usage and amino acid content, as well as for primary sequence alignments.

Programs such as MASE, PHYLIP, and CLUSTALV employed in a phylogenetic survey of primary sequences were provided by BIRCH (Biological Research Computer Hierarchy; Fristensky, 1991), a collection of statistical analyses programs offered in the SUN UNIX system at the University of Manitoba.

RESULTS

Section I: Determination of the Nucleotide Sequence of *katE*

Sequencing Strategy

Mulvey et al. (1988) had previously cloned an *E. coli* chromosomal fragment encompassing the *katE* gene and constructed the plasmid pAMkatE6. For the construction of the plasmid pAMkatE72 to be used for sequence analysis, a 3.5 kb fragment containing *katE* was removed from pAMkatE6 with *Pst*I and *Cla*I, and inserted into the corresponding restriction sites in the multiple cloning region of pKS+ (Fig. 1). Numerous subclones of portions of the chromosomal insert in pAMkatE72 were prepared in pKS+, pKS-, pSK+, and pSK- using different combinations of restriction nucleases (Table 2). Sequencing was carried out using synthetic oligonucleotide primers corresponding to portions of the multiple cloning site and to portions of established *katE* sequence (see Fig. 2). A restriction map and the sequencing strategy for both strands of the *katE*-containing insert within pAMkatE72 are summarized in Figure 2.

DNA Sequence of *katE*

The complete DNA sequence of the *Pst*I-*Cla*I fragment, which includes the *katE* gene, was determined and found to be 3,466 bp in length. All restriction sites determined by digestion were also identified within this stretch of sequence. A single open reading frame of 2,259 bp was

Figure 1. Construction of pAMkatE72. (See text for details.) Restriction nucleases are abbreviated as follows: C, *Cla*I; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pst*I; and V, *Eco*RV.

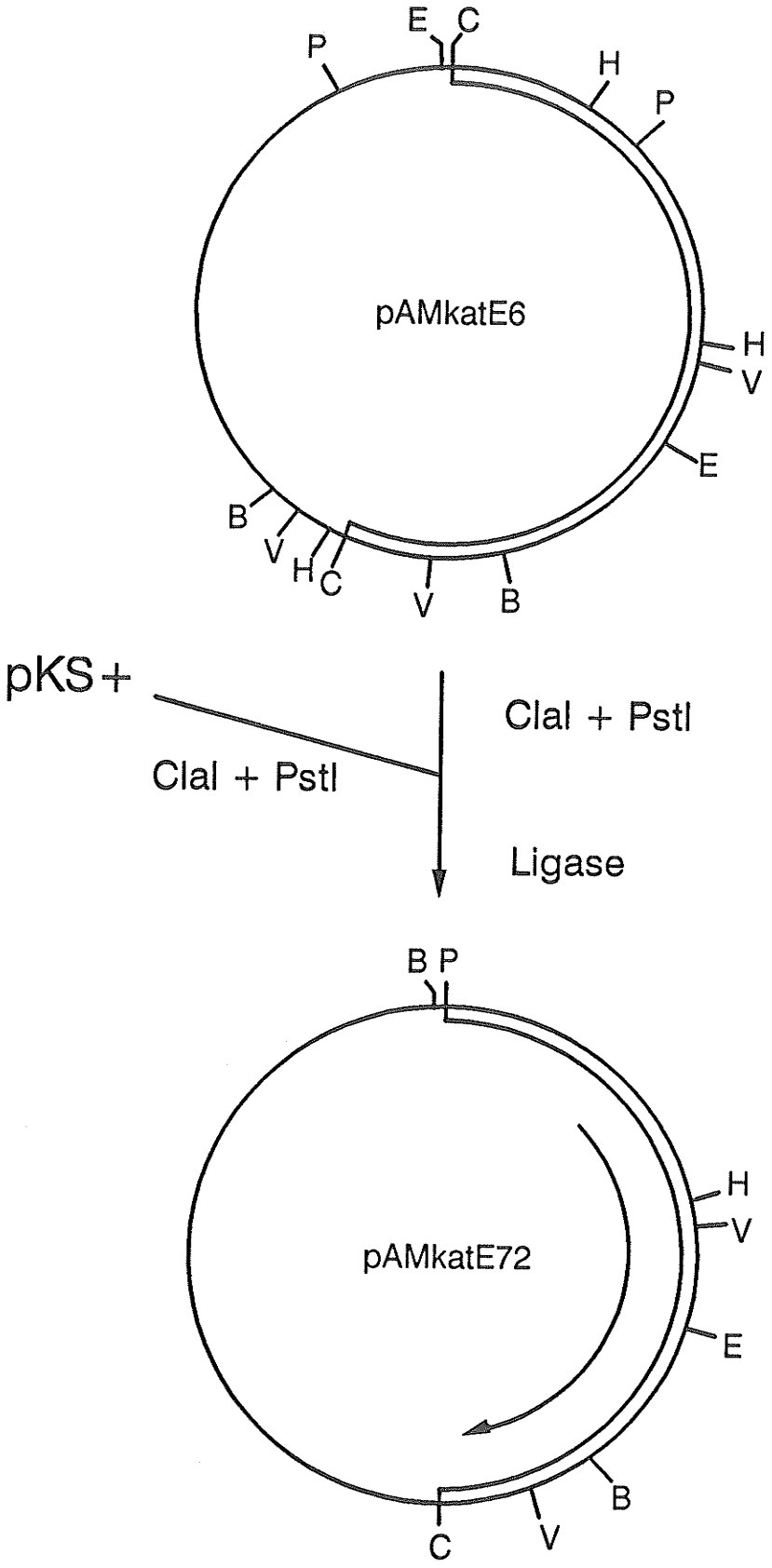
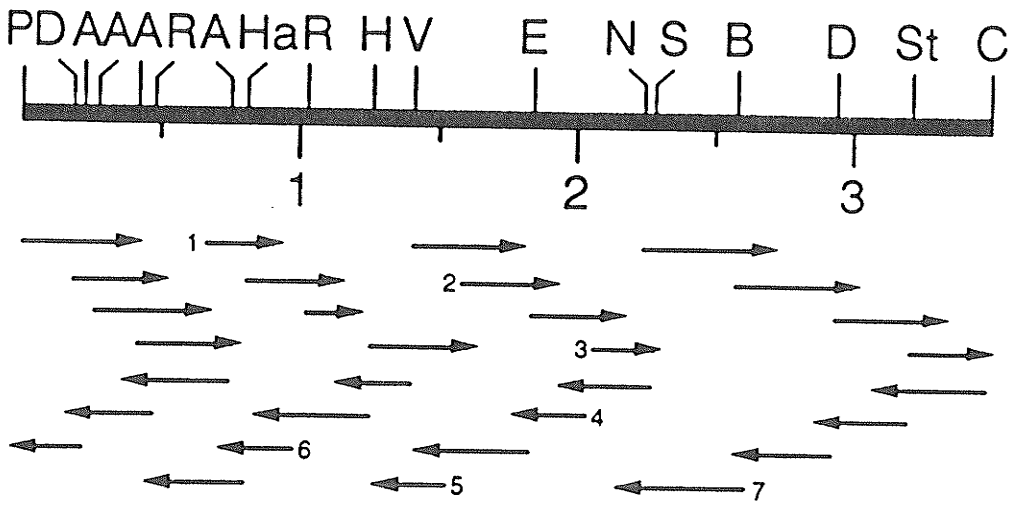

E
C
P
H
P
pAMkatE6
H
V
E
B
V
H
C
V
B
pKS+
ClaI + PstI
ClaI + PstI
Ligase
B
P
H
V
pAMkatE72
E
B
C
V

Table 2. Subclones used for nucleotide sequencing of *katE*.

Subclone	Restriction Fragment of *katE*	Fragment Size Limits[a] (bp)
pKS+6	*Pst*I-*Hind*III	1-1251
pKS+30	*Dra*I-*Hind*III	218-1251
pKS+29	*Sau*3A-*Hind*III	269-1251
pKS+28a	*Sau*3A-*Sau*3A	402-742
pKS+27	*Hae*III-*Hind*III	786-1251
pKS+31	*Rsa*I-*Hind*III	1031-1251
pKS-10	*Hind*III-*Ssp*I	1246-2333
pSK+4	*Eco*RV-*Bam*HI	1361-2577
pSK+5	*Eco*RI-*Bam*HI	1856-2577
pKS+24	*Nru*I-*Cla*I	2302-3466
pKS+11	*Bam*HI-*Cla*I	2572-3466
pKS+7	*Dra*I-*Cla*I	2930-3466
pKS+36	*Stu*I-*Cla*I	3206-3466
pSK+11	*Cla*I-*Bam*HI	3466-2572
pKS-38	*Stu*I-*Bam*HI	3211-2572
pKS-14	*Dra*I-*Bam*HI	2935-2572
pSK+1	*Ssp*I-*Eco*RV	2333-1361
pKS+32	*Eco*RI-*Hind*III	1861-1246
pSK+8	*Eco*RV-*Ssp*I	1366-40
pKS-9	*Hind*III-*Ssp*I	1251-40
pKS-26	*Hae*III-*Pst*I	789-1
pKS+28	*Sau*3A-*Pst*I	742-1
pKS-35	*Rsa*I-*Pst*I	479-1
pKS+28b	*Sau*3A-*Pst*I	251-1

[a] Numbers refer to the location of the fragments on the *katE* sequence presented in Figure 3.

Figure 2. Restriction map of the *katE*-containing *Pst*I-*Cla*I chromosomal insert (3.5 kb) and the DNA sequencing strategy for *katE*. The arrows indicate the direction and the extent of sequencing. All start sites of the fragments were obtained by digestion with various restriction nucleases (P, *Pst*I; D, *Dra*I; A, *Sau*3A; R, *Rsa*I; Ha, *Hae*III; H, *Hind*III; V, *Eco*RV; E, *Eco*RI; N, *Nru*I; S, *Ssp*I; B, *Bam*HI; St, *Stu*I; and C, *Cla*I). Synthetic oligonucleotide primers used in the chain elongation from the multiple cloning site into the subcloned *katE* regions included: the universal primer, 5′-GTAAAACGACGGCCAGT; the reverse primer, 5′-AACAGCTATGACCATG; the SK primer, 5′-TCTAGAACTAGTTGGATC; and the KS primer, 5′-CGAGGTGGCGACGGTATCG. Those arrows beginning with a number (1-7) were obtained using the following synthetic oligonucleotide primers of established sequence, 1, 5′-TTTGCGTGTATTTCATA; 2, 5′-TTCGGTATTCACACCTT; 3, 5′-CGCTGTTGCAGGGACGT; 4, 5′-ACGTCCCTGCAACAGCG; 5, 5′-CACGTTGTGCAGAGTTT; 6, 5′-GGTAGGTTGTGCACCTG; and 7, 5′-TACCACGCGACCTTTCA. (See text for further details.)


PDAAARAHaR HV E N S B D St C
1
2
3
1
2
3
4
6
5
7

found. This open reading frame beginning with an ATG codon at base 821 and ending with a TGA termination codon at base 3,080 predicted a protein sequence of 753 amino acids. The DNA sequence and the predicted amino acid sequence are presented in Figure 3.

Potential *katE* Control Sequences

Elements controlling both translational and transcriptional activities were identified within the 3,466 bp fragment containing the *katE* gene.

A putative ribosome binding sequence (Shine-Dalgarno) of AGGAG was identified 10 bp upstream of the initiation codon (Fig. 3). This site is very similar to the sequence of AGGAGG established as a consensus for ribosome binding in *E. coli* (Glover, 1984). As shown in Figure 4, the transcriptional start site for *katE* was identified as the G residue at position 695 in the sequence in Figure 3. (Weaker bands at about 696-705 were likely due to premature termination of the elongating DNA sequence.) Potential promoter sequences at -10 (ACGTCC) and -35 (GTTTAGC) were identified at 133 and 154 bp upstream from the initiation codon (Fig. 3) with a spacing of 15 bp.

Because *katE* expression is controlled and regulated by the *katF* gene product, which in all likelihood represents a novel sigma factor (Mulvey and Loewen, 1989), it is reasonable to anticipate that perhaps the promoter sequences

Figure 3. Nucleotide sequence of the *katE* gene region and the deduced amino acid sequence of HPII catalase. The open reading frame begins at base 821, and the HPII predicted amino acid sequence is given below the nucleotide sequence. The transcriptional start site is indicated by *. The putative promoter sequences (-10 and -35) are underlined and labeled upstream of the start site. The potential transcriptional terminator site is underlined and highlighted by boldfaced type. The potential Shine-Dalgarno ribosome binding site is underlined and labeled as SD.

PstI

60

CTGCAGCCTTTCTTTAAAAGAGTCGAAAGCCAGGCTTTTAATATTTAAATCACCATAATT

120

ACTCTGTATTAAGTTTGTAGAAAACATCTCCCGCCTCATATTGTTAACAAAATTATTATC

180

TCATTTAAATCTAAGTCATTTACAATATAAGTTTAAGAGCGACGCCACAGGATGAACTAT

240

CAAAAATAGCTCATCATGATTAGCAAAACTTAACCATTTTAAAATAAATAAACAATTAAA

300

GAAAAAAGATCACTTATTTATAGCAATAGATCGTCAAAGGCAGCTTTTTGTTACAGGTGG

360

TTTGAATGAATGTAGCAACGAAATACAGAATTTCAGGTCATGTAACTCCCGGCAAACCGG

420

GAGGTATGTAATCCTTACTCAGTCACTTCCCCTTCCTGGCGGATCTGATTTGCCCAACGT

480

TGGGCAGATTCAGGCACAGTAAACGCCGGTGAGCGCAGAAATGACTCTCCCATCAGTACA

540

AACGCAACATATTTGCCACGCAGCATCCAGACATCACGAAACGAATCCATCTTTATCGCA

600

TGTTCTGGCGGCGCGGGTTCCGTGCGTGGGACATAGCTAATAATCTGGCGGTTTTGCTGG

660

CGGAGCGGTTTCTTCATTACTGGCTTCACTAAACGCATATTAAAAATCAGAAAAACTGTA

720

GTTTAGCCGATTTAGCCCCTGTACGTCCCGCTTTGCGTGTATTTCATAACACCGTTTCCA

-35 -10 *

780

GAATAGTCTCCGAAGCGGGATCTGGCTGGTGGTCTATAGTTAGAGAGTTTTTTGACCAAA

835

ACAGCGGCCCTTTCAGTAATAAATTAAGGAGACGAGTTCA ATG TCG CAA CAT AAC

SD *met ser gln his asn*

(1)

880

GAA AAG AAC CCA CAT CAG CAC CAG TCA CCA CTA CAC GAT TCC AGC
glu lys asn pro his gln his gln ser pro leu his asp ser ser

925

GAA GCG AAA CCG GGG ATG GAC TCA CTG GCA CCT GAG GAC GGC TCT
glu ala lys pro gly met asp ser leu ala pro glu asp gly ser

970

CAT CGT CCA GCG GCT GAA CCA ACA CCG CCA GGT GCA CAA CCT ACC
his arg pro ala ala glu pro thr pro pro gly ala gln pro thr

1015

GCC CCA GGG AGC CTG AAA GCC CCT GAT ACG CGT AAC GAA AAA CTT
ala pro gly ser leu lys ala pro asp thr arg asn glu lys leu

1060

AAT TCT CTG GAA GAC GTA CGC AAA GGC AGT GAA AAT TAT GCG CTG
asn ser leu glu asp val arg lys gly ser glu asn tyr ala leu

1105

ACC ACT AAT CAG GGC GTG CGC ATC GCC GAC GAT CAA AAC TCA CTG
thr thr asn gln gly val arg ile ala asp asp gln asn ser leu

1150

CGT GCC GGT AGC CGT GGT CCA ACG CTG CTG GAA GAT TTT ATT CTG
arg ala gly ser arg gly pro thr leu leu glu asp phe ile leu

1195

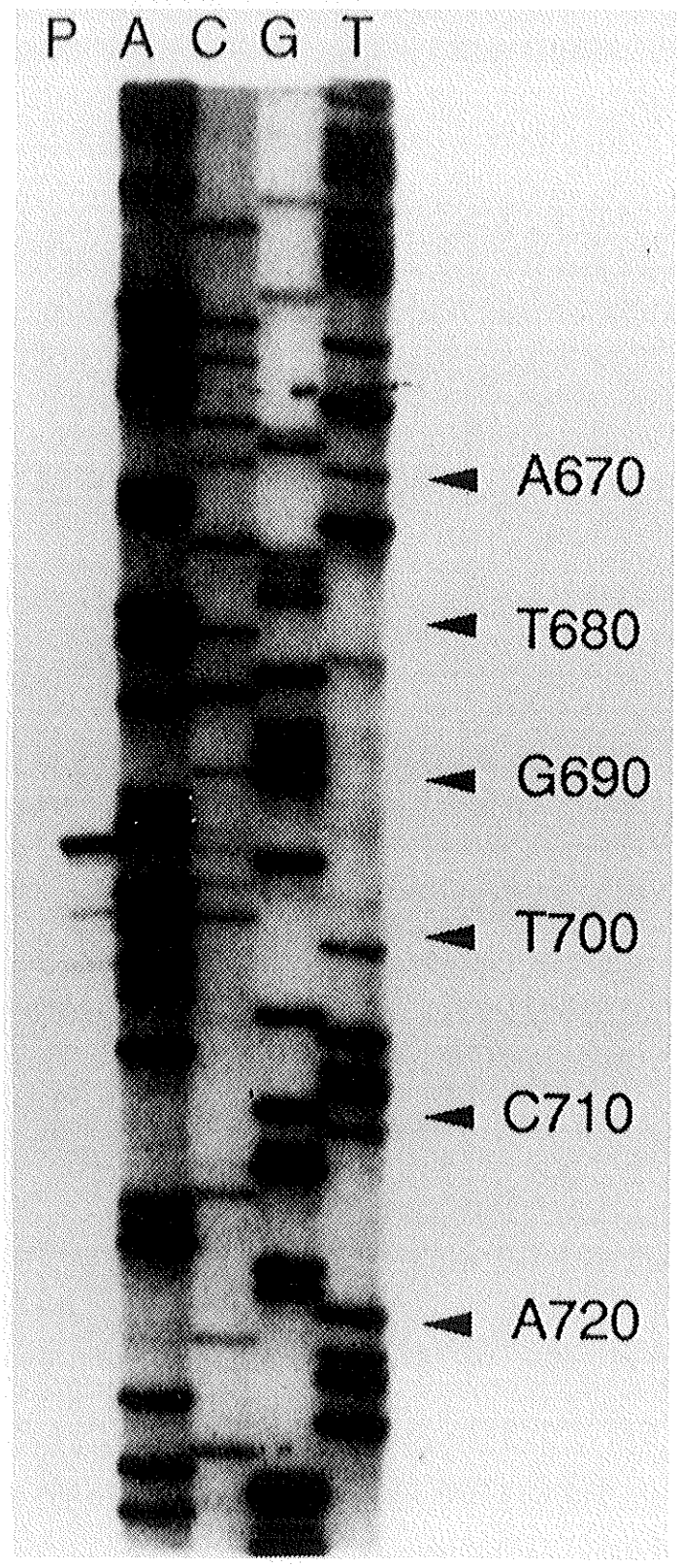
CGC GAG AAA ATC ACC CAC TTT GAC CAT GAG CGC ATT CCG GAA CGT
arg glu lys ile thr his phe asp his glu arg ile pro glu arg

1240
ATT GTT CAT GCA CGC GGA TCA GCC GCT CAC GGT TAT TTC CAG CCA
ile val his ala arg gly ser ala ala his gly tyr phe gln pro
1285
TAT AAA AGC TTA AGC GAT ATT ACC AAA GCG GAT TTC CTC TCA GAT
tyr lys ser leu ser asp ile thr lys ala asp phe leu ser asp
1330
CCG AAC AAA ATC ACC CCA GTA TTT GTA CGT TTC TCT ACC GTT CAG
pro asn lys ile thr pro val phe val arg phe ser thr val gln
(156)
1375
GGT GGT GCT GGC TCT GCT GAT ACC GTG CGT GAT ATC CGT GGC TTT
gly gly ala gly ser ala asp thr val arg asp ile arg gly phe
1420
GCC ACC AAG TTC TAT ACC GAA GAG GGT ATT TTT GAC CTC GTT GGC
ala thr lys phe tyr thr glu glu gly ile phe asp leu val gly
1465
AAT AAC ACG CCA ATC TTC TTT ATC CAG GAT GCG CAT AAA TTC CCC
asn asn thr pro ile phe phe ile gln asp ala his lys phe pro
1510
GAT TTT GTT CAT GCG GTA AAA CCA GAA CCG CAC TGG GCA ATT CCA
asp phe val his ala val lys pro glu pro his trp ala ile pro
1555
CAA GGG CAA AGT GCC CAC GAT ACT TTC TGG GAT TAT GTT TCT CTG
gln gly gln ser ala his asp thr phe trp asp tyr val ser leu
1600
CAA CCT GAA ACT CTG CAC AAC GTG ATG TGG GCG ATG TCG GAT CGC
gln pro glu thr leu his asn val met trp ala met ser asp arg
1645
GGC ATC CCC CGC AGT TAC CGC ACC ATG GAA GGC TTC GGT ATT CAC
gly ile pro arg ser tyr arg thr met glu gly phe gly ile his
1690
ACC TTC CGC CTG ATT AAT GCC GAA GGG AAG GCA ACG TTT GTA CGT
thr phe arg leu ile asn ala glu gly lys ala thr phe val arg
1735
TTC CAC TGG AAA CCA CTG GCA GGT AAA GCC TCA CTC GTT TGG GAT
phe his trp lys pro leu ala gly lys ala ser leu val trp asp
1780
GAA GCA CAA AAA CTC ACC GGA CGT GAC CCG GAC TTC CAC CGC CGC
glu ala gln lys leu thr gly arg asp pro asp phe his arg arg
(306)
1825
GAG TTG TGG GAA GCC ATT GAA GCA GGC GAT TTT CCG GAA TAC GAA
glu leu trp glu ala ile glu ala gly asp phe pro glu tyr glu
1870
CTG GGC TTC CAG TTG ATT CCT GAA GAA GAT GAA TTC AAG TTC GAC
leu gly phe gln leu ile pro glu glu asp glu phe lys phe asp
1915
TTC GAT CTT CTC GAT CCA ACC AAA CTT ATC CCG GAA GAA CTG GTG
phe asp leu leu asp pro thr lys leu ile pro glu glu leu val
1960
CCC GTT CAG CGT GTC GGC AAA ATG GTG CTC AAT CGC AAC CCG GAT
pro val gln arg val gly lys met val leu asn arg asn pro asp
2005
AAC TTC TTT GCT GAA AAC GAA CAG GCG GCT TTC CAT CCT GGG CAT
asn phe phe ala glu asn glu gln ala ala phe his pro gly his

2050
ATC GTG CCG GGA CTG GAC TTC ACC AAC GAT CCG CTG TTG CAG GGA
ile val pro gly leu asp phe thr asn asp pro leu leu gln gly
2095
CGT TTG TTC TCC TAT ACC GAT ACA CAA ATC AGT CGT CTT GGT GGG
arg leu phe ser tyr thr asp thr gln ile ser arg leu gly gly
2140
CCG AAT TTC CAT GAG ATT CCG ATT AAC CGT CCG ACC TGC CCT TAC
pro asn phe his glu ile pro ile asn arg pro thr cys pro tyr
2185
CAT AAT TTC CAG CGT GAC GGC ATG CAT CGC ATG GGG ATC GAC ACT
his asn phe gln arg asp gly met his arg met gly ile asp thr
2230
AAC CCG GCG AAT TAC GAA CCG AAC TCG ATT AAC GAT AAC TGG CCG
asn pro ala asn tyr glu pro asn ser ile asn asp asn trp pro
(456)
2275
CGC GAA ACA CCG CCG GGG CCG AAA CGC GGC GGT TTT GAA TCA TAC
arg glu thr pro pro gly pro lys arg gly gly phe glu ser tyr
2320
CAG GAG CGC GTG GAA GGC AAT AAA GTT CGC GAG CGC AGC CCA TCG
gln glu arg val glu gly asn lys val arg glu arg ser pro ser
2365
TTT GGC GAA TAT TAT TCC CAT CCG CGT CTG TTC TGG CTA AGT CAG
phe gly glu tyr tyr ser his pro arg leu phe trp leu ser gln
2410
ACG CCA TTT GAG CAG CGC CAT ATT GTC GAT GGT TTC AGT TTT GAG
thr pro phe glu gln arg his ile val asp gly phe ser phe glu
2455
TTA AGC AAA GTC GTT CGT CCG TAT ATT CGT GAG CGC GTT GTT GAC
leu ser lys val val arg pro tyr ile arg glu arg val val asp
2500
CAG CTG GCG CAT ATT GAT CTC ACT CTG GCC CAG GCG GTG GCG AAA
gln leu ala his ile asp leu thr leu ala gln ala val ala lys
2545
AAT CTC GGT ATC GAA CTG ACT GAC GAC CAG CTG AAT ATC ACC CCA
asn leu gly ile glu leu thr asp asp gln leu asn ile thr pro
2590
CCT CCG GAC GTC AAC GGT CTG AAA AAG GAT CCA TCC TTA AGT TTG
pro pro asp val asn gly leu lys lys asp pro ser leu ser leu
2635
TAC GCC ATT CCT GAC GGT GAT GTG AAA GGT CGC GTG GTA GCG ATT
tyr ala ile pro asp gly asp val lys gly arg val val ala ile
2680
TTA CTT AAT GAT GAA GTG AGA TCG GCA GAC CTT CTG GCC ATT CTC
leu leu asn asp glu val arg ser ala asp leu leu ala ile leu
(606)
2725
AAG GCG CTG AAG GCC AAA GGC GTT CAT GCC AAA CTG CTC TAC TCC
lys ala leu lys ala lys gly val his ala lys leu leu tyr ser
2770
CGA ATG GGT GAA GTG ACT GCG GAT GAC GGT ACG GTG TTG CCT ATA
arg met gly glu val thr ala asp asp gly thr val leu pro ile
2815
GCC GCT ACC TTT GCC GGT GCA CCT TCG CTG ACG GTC GAT GCG GTC
ala ala thr phe ala gly ala pro ser leu thr val asp ala val

2860
ATT GTC CCT TGC GGC AAT ATC GCG GAT ATC GCT GAC AAC GGC GAT
ile val pro cys gly asn ile ala asp ile ala asp asn gly asp
2905
GCC AAC TAC TAC CTG ATG GAA GCC TAC AAA CAC CTT AAA CCG ATT
ala asn tyr tyr leu met glu ala tyr lys his leu lys pro ile
2950
GCG CTG GCG GGT GAC GCG CGC AAG TTT AAA GCA ACA ATC AAG ATC
ala leu ala gly asp ala arg lys phe lys ala thr ile lys ile
2995
GCT GAC CAG GGT GAA GAA GGG ATT GTG GAA GCT GAC AGC GCT GAC
ala asp gln gly glu glu gly ile val glu ala asp ser ala asp
3040
GGT AGT TTT ATG GAT GAA CTG CTA ACG CTG ATG GCA GCA CAC CGC
gly ser phe met asp glu leu leu thr leu met ala ala his arg
3086
GTG TGG TCA CGC ATT CCT AAG ATT GAC AAA ATT CCT GCC TGATGGG
val trp ser arg ile pro lys ile asp lys ile pro ala
(753) 3146
AGCGCGCAATTGCGCCGCCTCAATGATTTACATAGTGCGCTTTGTTTATGCCGGATGCGC
3206
GTGAACGCCTTATCCGGCCTACAAAACTGTGCAAATTCAATATATTGCAGGAAACACGTA
3266
G**GCCTGAT**AAGCGAAGCC**ATCAGGC**AGTTTTGCGTTTGTCAGCAGTCTCAAGCGGCGGCA
3326
GTTACGCCGCCTTTGTAGGAATTAATCGCCGGATGCAAGGTTCACGCCGATCTGGCAAAC
3386
ATCCTCACTTACACATCCCGATAACTCCCCAACCGATAACCACGCTGAGCGATAGCACCT
3446
TTCAACGACGCTGATGTCAACACATCCAGCTCCGTTAAGCGTGGGAAACAGTAAGCACTC
*Cla*I 3466
TGACGGATAGTATTATCGAT

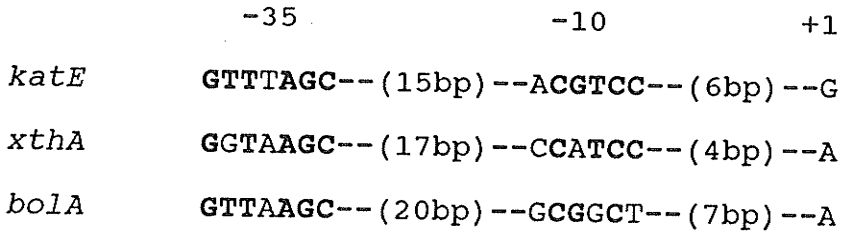
Figure 4. Determination of the *katE* transcriptional start site by primer extension mapping. A ^{32}P-labeled oligonucleotide primer (5'-CTGATGTGGGTTCTTTTCGTTATGT TGCGA) corresponding to a complementary stretch of sequence (bases 824 to 853) within the *katE* region was annealed to 25 μg of total RNA from NM522 carrying pAMkatE72, elongated by reverse transcriptase, and electrophoresed on a 6% polyacrylamide sequencing gel in lane P. Lanes A, C, G, and T represent a sequencing ladder generated by the same primer described above, and the ladder is the complementary sequence of that given in Fig. 3. The numbered bases indicated on the right correspond to the sequence in Fig. 3. The primer extension product (lane P) migrating adjacent to the C is complementary to G-695 of the sequence in Fig. 3.


P A C G T
A670
T680
G690
T700
C710
A720

governing *katE* expression are different from the -10 (TATAAT) and -35 (TTGACA) consensus regions recognized by σ^{70} (Harley and Reynolds, 1987). Because *xthA* and *bolA* are also genes under KatF regulation (Sak et al., 1989; Lange and Hengge-Aronis, 1991a), the promoter sequences of *katE*, *xthA*, and *bolA* taken together may possibly form a consensus region specific for KatF recognition. A comparison of the *katE* -10 and -35 sequences with the *xthA* and *bolA* genes revealed the presence of similar sequences upstream of the start sites established in *xthA* (Saporito et al., 1988) and *bolA* (Aldea et al., 1989). As shown by an alignment of the putative promoter sequences for *katE*, *xthA*, and *bolA* in Figure 5, several bases are identical within the -10 and -35 regions. However, as expected, the *katE*, *xthA*, and *bolA* putative promoter sequences are different from the consensus -10 and -35 sequences for σ^{70}. Despite the clear differences from the σ^{70} sequences, these three potential promoter regions are too small a sampling to produce a consensus sequence of any confidence, and an examination of other genes controlled by KatF is necessary to establish a "true" consensus sequence for KatF recognition.

At the opposite end of the *katE* gene, an element of two-fold symmetry representing a potential transcriptional termination sequence was identified about 125 bp downstream of the termination codon. The 7 bp inverted repeat displays a secondary structure with a predicted stability of -9.1

Figure 5. Comparison of putative promoter sequences for *katE*, *xthA*, and *bolA*. The start sites for each of the genes are indicated by +1, and the -10 and -35 regions labeled accordingly. The spacing between the start site and the -10 region, as well as between the -10 and -35 regions is also given. Nucleotides in the -10 and -35 regions for *katE* also found in *xthA* and *bolA* are indicated by boldfaced type.


-35
-10
+1
katE
GTTTAGC--(15bp)--ACGTCC--(6bp)--G
xthA
GGTAAGC--(17bp)--CCATCC--(4bp)--A
bolA
GTTAAGC--(20bp)--GCGGCT--(7bp)--A

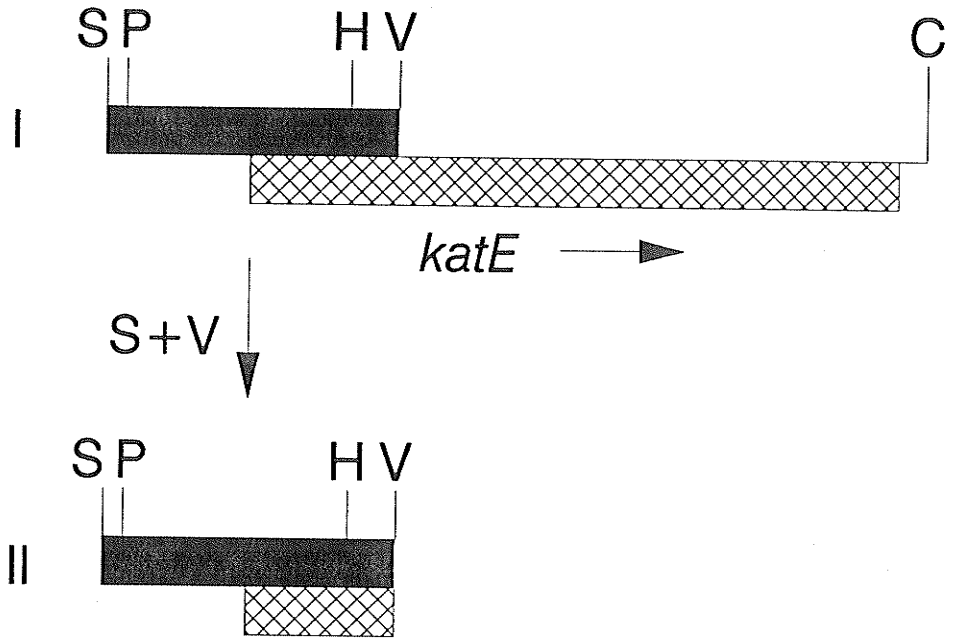
kcal/mol and is followed by four T residues, thereby exhibiting the typical structure of a rho-independent terminator of moderate strength.

5′ end Deletions of *katE*

As part of a preliminary characterization of the *katE* regulatory region, a series of deletions was created upstream of the *katE* gene. A 660 bp segment of DNA precedes the putative *katE* promoter (Fig. 3) and the object was to use deletions to establish the minimum length of DNA necessary for regulating *katE* expression and to determine if KatF is the sole regulatory factor for *katE* expression. The truncated *katE* promoter segments were fused to *lacZ* in pRS415 (Simons et al., 1987) and β-galactosidase levels were used as a measure of expression. Mulvey et al. (1990) reported the construction of a fusion plasmid (pRSkatE16) containing the complete sequence upstream of *katE* by inserting a 1.4 kb *Sma*I-*Eco*RV fragment from pAMkatE72 into pRS415 (Fig. 6). The *Sma*I site used for cloning is from the pKS+ vector and in pAMkatE72 and the *Eco*RV site is at base 1,361 within *katE*. Expression studies using pRSkatE16 confirmed that KatF was required for *katE* expression (Mulvey et al., 1990).

Because the 5′ noncoding region of *katE* lacked unique restriction sites for the excision of various sized fragments, new *Pst*I sites were inserted into pKS+6 by *in*

Figure 6. A general outline for the construction of *lacZ*-fusion plasmids containing the *katE* 5′ noncoding region. The *Pst*I to *Cla*I *katE* containing DNA fragment, including the *Sma*I site from the KS+ vector, is presented in I. The *katE* coding region is indicated by the cross-hatched region. A *Sma*I-*Eco*RV fragment is obtained by restriction nuclease digestion as indicated by II and then inserted into the *Sma*I site of pRS415 to produce fusion plasmids containing the *katE* 5′noncoding region. The direction of transcription is shown by the horizontal arrow. Abbreviations for restriction nucleases are as given in Figures 1 and 2. (Not drawn to scale.)


S P
H V
C
I
katE
S+V
S P
H V
II

vitro mutagenesis using the primers in Table 3. Cleavage at the two *Pst*I sites followed by religation permitted the removal of the intervening sequences as shown in Figure 7. The larger bands in each lane were recovered by the GENECLEAN protocol and treated with ligase to generate the deletion mutants. Next, the *Pst*I-*Hind*III fragments were transferred from the deletion plasmids into pAMkatE72 cut with *Pst*I and *Hind*III to produce *katE* clones with different sized deletions in the 5′ noncoding region. Subsequently, the *Sma*I-*Eco*RV fragment (as outlined in Figure 6) from each mutant was transferred to pRS415 to generate the *katE*::*lacZ* fusions for which the end points are shown in Figure 8.

Five fusion plasmids with deletions extending to 593, 623, 655, 753, and 790 bp in the *katE* sequence in Figure 3 (designated pRSkatEΔP1, pRSkatEΔP2, pRSkatEΔP4, pRSkatEΔP3, and pRSkatEΔP5, respectively) were isolated and β-galactosidase activity assayed during growth from exponential phase (3 hours in Table 4) into stationary phase (12 hours in Table 4). The three fusion plasmids, pRSkatEΔP1, pRSkatEΔP2, and pRSkatEΔP4, express β-galactosidase in exponential and stationary phase similar to pRSkatE16 both in pattern and extent (Table 4). None of the three deletions had removed the putative -10 and -35 regions. However, the two fusion plasmids, pRSkatEΔP3 and pRSkatEΔP5, deleted to positions 753 and 790, respectively, exhibited greatly reduced expression levels. Approximately

Table 3. Synthetic oligonucleotide primers employed in the construction of new *Pst*I sites within the 5′ noncoding region upstream of *katE*.

Primer	Oligonucleotide Sequence[a]	Location[b] (bp)
P1	5′-GGCGGTT*C*T*GC*A*G*GCGGAGC	587 to 606
P2	5′-ACTGG*CTGCAG*TAAACGCAT	619 to 638
P3	5′-TGGTGGT*CTGCAG*TTAGAGA	747 to 766
P4	5′-AAAAA*CTGCAG*TTTAGCCGA	651 to 670
P5	5′-GGCCC*CTGCAG*TAATAAATT	786 to 805

[a] Sequences for *Pst*I sites are italicized and bases changed to produce the new *Pst*I sequences are underlined.

[b] Numbers refer to the location of the primers on the *katE* sequence presented in Figure 3.

Figure 7. *Pst*I restriction nuclease digestion of subclone pKS+6 and derivatives (pP1, pP2, pP3, pP4, and pP5) containing one additional *Pst*I site within the *katE* 5′ noncoding region. Restriction nuclease digestion and electrophoresis were carried out as described in Experimental Procedures.

Lane 1: pKS+6
Lane 2: pP1
Lane 3: pP2
Lane 4: pP3
Lane 5: pP4
Lane 6: pP5

DNA bands representing the various *Pst*I fragments (III) and the KS+ vector with the remaining portion of *katE* (II) are indicated. Linearized pKS+6 is indicated by I.

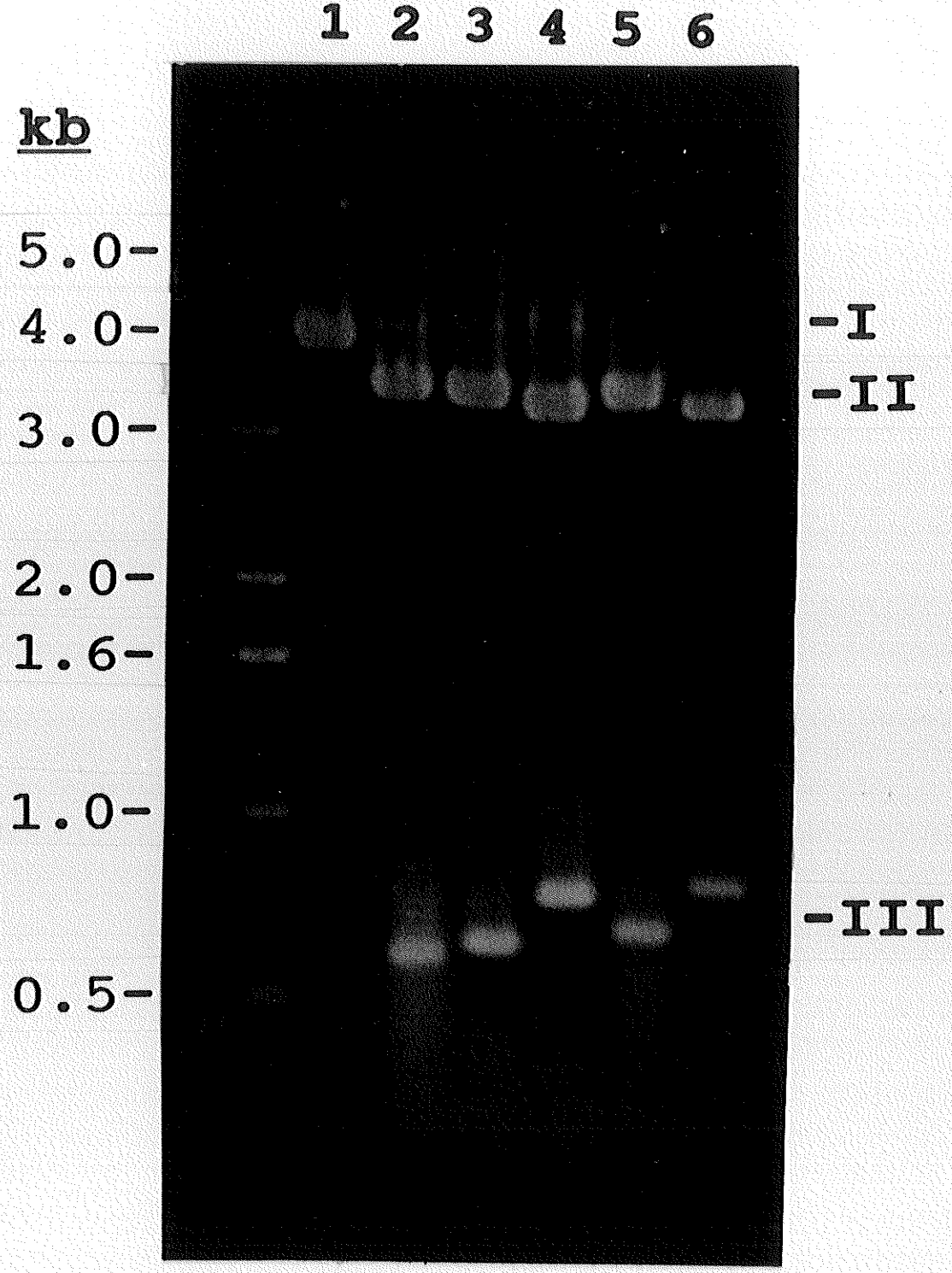

1 2 3 4 5 6
kb
5.0-
4.0-
3.0-
2.0-
1.6-
1.0-
0.5-
-I
-II
-III

Figure 8. 5′ end deletions of the *katE* noncoding region. The 5′ noncoding region is indicated by a solid line, with the location of the putative -10 and -35 sequences indicated by the hatched region. The *katE* coding region is indicated by the solid broad line. Both the undeleted form (wild type) of the *katE* region and the various deletions of the 5′ noncoding region brought on by the primers P1, P2, P3, P4, and P5 listed in Table 3 are indicated. The location (in bp) along the sequence (see Fig. 3) at which the truncation begins is indicated by the values within the parentheses. (Not drawn to scale.)

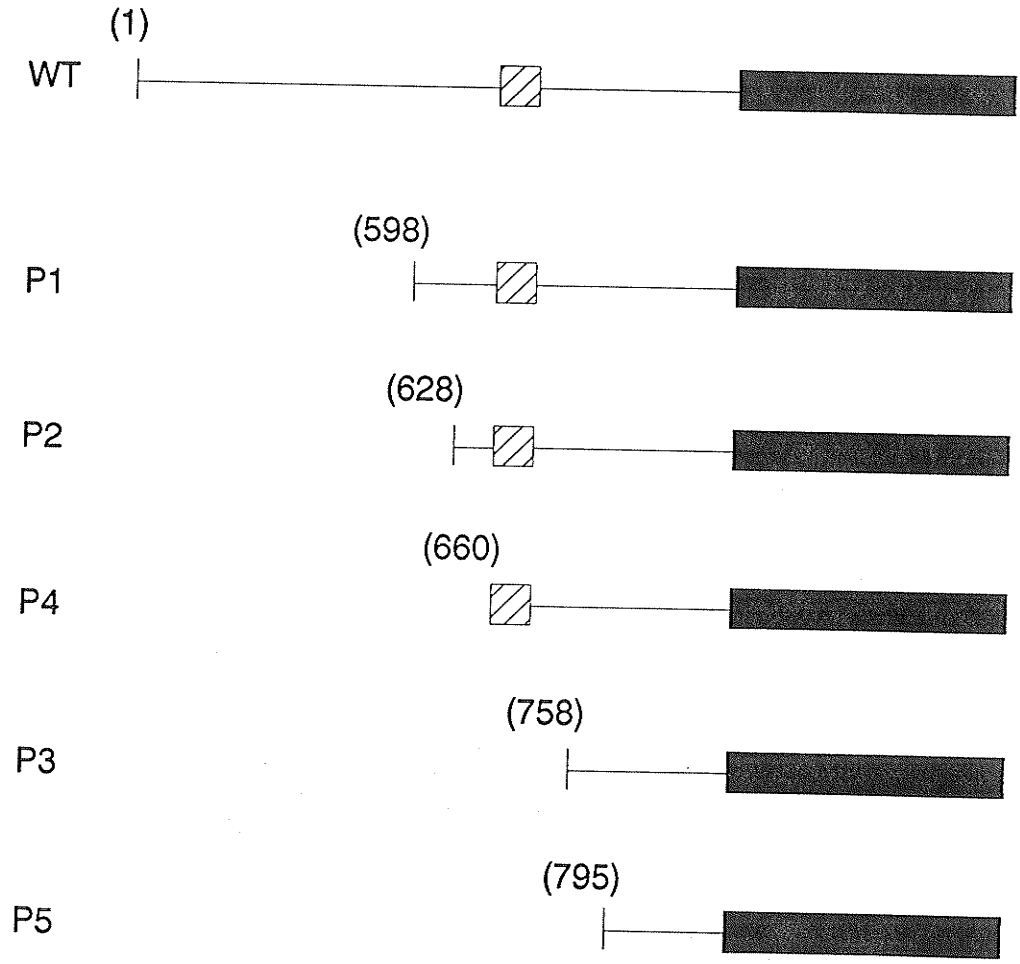

(1)
WT
(598)
P1
(628)
P2
(660)
P4
(758)
P3
(795)
P5

Table 4. β-Galactosidase levels from fusion plasmids containing deletions of various lengths upstream of *katE*.

Fusion Plasmid	Time of sampling (hrs) 3	12	27
pRSkatE16	88.0[a] (0.70)[b]	10,900 (4.78)	10,100 (4.57)
pRSkatEΔP1	111 (0.76)	9,100 (3.85)	11,900 (4.33)
pRSkatEΔP2	52.0 (0.57)	12,200 (3.48)	12,300 (4.57)
pRSkatEΔP4	54.7 (0.74)	11,800 (4.68)	10,300 (4.39)
pRSkatEΔP3	14.8 (0.80)	29.5 (4.07)	36.7 (3.90)
pRSkatEΔP5	10.7 (0.64)	38.1 (3.83)	42.5 (3.91)

[a] β-Galactosidase activity in Miller Units.
[b] A_{600} readings of cell density.

equal amounts of plasmid DNA were present in transformants with the different plasmids confirming that the variations in expression were not the result of variations in plasmid copy number. These results suggest that the *katE* promoter region is located between positions 655 and 753 in Figure 3. Furthermore, there cannot be a secondary regulatory factor acting upstream of position 655 to control *katE* expression. Deletions downstream of the putative promoter region will have to be constructed to determine if there are regulatory sequences between the promoter and the open reading frame.

The Predicted HPII Amino Acid Sequence

N-terminal sequence analysis of purified HPII protein confirmed the accuracy of the first 11 amino acids in the predicted HPII sequence. HPII catalase was purified (Loewen and Switala, 1986) from the *E. coli* strain UM255[pAMkatE72] (Mulvey et al., 1988), and the N-terminal sequence determined by the Tripartite Microanalytical Centre (University of Victoria, British Columbia) to be xxx-Ser-Gln-His-Asn-Glu-Lys-Asn-Pro-His-Gln. This is identical to residues 2 to 11 of the N-terminus deduced from the nucleotide sequence. For technical reasons, the identity of the first amino acid was not determined.

The molecular mass for the predicted protein, as calculated from the deduced amino acid sequence, is 84,173 Da and is considerably smaller than the 93,000 Da size

estimated from SDS-polyacrylamide electrophoresis (Loewen and Switala, 1986). The size discrepancy may be due to the moderately acidic nature of the HPII protein (pI=6.08), which may affect its migration on SDS-polyacrylamide gels.

Comparison of the amino acid composition of the predicted HPII sequence with the experimentally determined composition (Table 5) revealed close similarity between the two, a further substantiation that the correct reading frame had been obtained. The distribution of acidic, basic, aromatic, and hydrophobic residues was calculated from the predicted sequence and is given in Table 6. The HPII protein consists of 25% charged amino acids, with a slightly greater proportion of acidic (14.2%) over basic (10.8%) residues.

Comparison of the Predicted HPII Sequence with other Catalase Sequences

A preliminary comparison of the deduced amino acid sequence for the HPII subunit with the amino acid sequences of eukaryotic catalases from mammalian, fungal, and plant sources is presented in Figure 9, and reveals significant similarity between HPII and other catalases. The high degree of similarity is evident only with the core of HPII because of its greater length (753 amino acids) compared to the average length of about 500 residues found for most eukaryotic catalases. Consequently, segments of HPII at the

Table 5. Comparison of predicted and actual amino acid compositions of HPII catalase.

Amino Acid	Predicted No. (%)[a]	Actual No. (%)[b]
Alanine	64 (8.5)	61.2 (8.1)
Arginine	44 (5.8)	41.4 (5.5)
Asparagine	33 (4.4)	----------
Aspartate	58 (7.7)	95.1 (12.6)[c]
Cysteine[d]	2 (0.3)	2.2 (0.3)
Glutamine	26 (3.5)	----------
Glutamate	49 (6.5)	86.3 (11.5)[c]
Glycine	53 (7.0)	57.0 (7.6)
Histidine	28 (3.7)	22.7 (3.0)
Isoleucine	43 (5.7)	40.7 (5.4)
Leucine	60 (8.0)	57.7 (7.7)
Lysine	37 (4.9)	36.6 (4.9)
Methionine[e]	12 (1.6)	11.5 (1.5)
Phenylalanine	40 (5.3)	31.9 (4.2)
Proline	59 (7.8)	60.0 (8.0)
Serine[f]	40 (5.3)	31.8 (4.2)
Threonine	37 (4.9)	39.0 (5.2)
Tryptophan[g]	9 (1.2)	12.5 (1.7)
Tyrosine	19 (2.5)	13.8 (1.8)
Valine	40 (5.3)	43.7 (5.8)

[a] Predicted values were calculated by the Microgenie software program.

[b] Actual values were determined according to the following protocol: One nmol HPII was mixed with constant boiling hydrochloric acid, sealed within a glass tube, and incubated at 110°C for 24, 48, and 72 hours. A LKB 4151 Alpha Plus amino acid analyzer (Department of Animal Science, University of Manitoba) was used to carry out the amino acid analyses.

[c] Because of amide hydrolysis, asparagine and glutamine values are combined with the aspartate and glutamate values, respectively.

[d] Determined as cysteic acid by oxidizing one nmol HPII for 20 hours at 5°C in performic acid (a mixture of 0.2 ml of 30% H_2O_2 and 1.8 ml of 90% formic acid), followed by lyophilization and hydrolysis.

[e] Determined as methionine sulfoxide and methionine sulfone following performic acid oxidation.

[f] Hydrolysis time extrapolated to zero to compensate for for hydrolytic destruction.

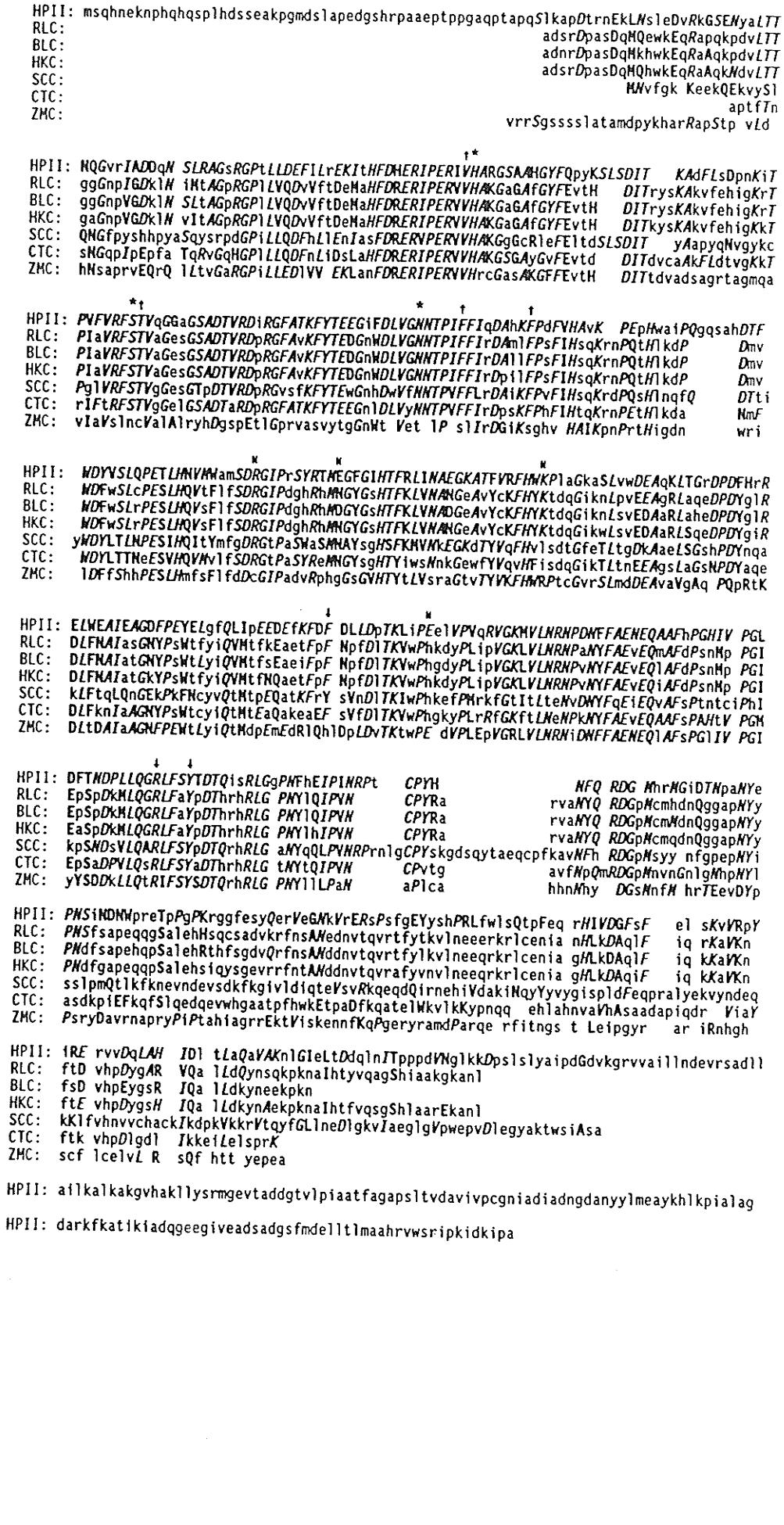
[g] As determined by the method of Edelhoch (1967).

Table 6. Amino acid distribution of the predicted HPII sequence[a].

Amino Acid	No.	(%)
Acidic[b]	117	14.2
Basic[c]	81	10.8
Aromatic[d]	68	9.0
Hydrophobic[e]	223	29.6

[a] As determined by the Microgenie software package.
[b] Aspartic and glutamic acids.
[c] Arginine and lysine.
[d] Phenylalanine, tryptophan, and tyrosine.
[e] Isoleucine, leucine, methionine, valine, and aromatic amino acids.

Figure 9. Comparison of the predicted amino acid sequence of the HPII subunit with the catalase sequences from rat liver (RLC), bovine liver (BLC), human kidney (HKC), *Saccharomyces cerevisiae* Type T (SCC), *Candida tropicalis* (CTC), and maize (ZMC). Residues identical in HPII and any of the other catalases are indicated by upper case italicized boldfaced type. Conservative replaced residues between HPII and any of the other catalases are indicated by upper case boldfaced type. Conservative replacements are defined according to the following groups: (S and T), (A and G), (F, Y, and W), (H, K, and R), (D, E, N, and Q) and (I, L, M, and V). Residues involved in the active site of bovine liver catalase are indicated by *. Residues involved in heme binding on the distal (↑) and proximal (↓) sides are also indicated. Residues involved in NADPH binding sites are indicated by **N**. Sequence homology was enhanced by the insertion of gaps.



N-terminus (57 residues) and the C-terminus (168 residues) could not be aligned having no similarity to any part of the shorter catalases or to any protein sequence contained in the GENBANK and PIR databases. With both identical residues and conservative replacements considered, much of the similarity between HPII and the other catalases is confined to the N-terminal half of the core sequence, lessening towards the C-terminal. Amino acids identified in bovine liver catalase (Fita and Rossmann, 1985a) as participating directly in catalysis (His74, Ser113, and Asn147) are conserved in HPII as His128, Ser167, and Asn201. Various residues interacting with the heme group in bovine liver catalase on the distal (Val73, Thr114, Phe152, and Phe160) and proximal (Pro335, Arg353, and Tyr357) sides of the heme are also conserved in HPII as Val127, Thr168, Phe206, Phe214, Pro393, Arg411, and Tyr415, respectively. Finally, NADPH binding-site residues reported in bovine liver catalase by Fita and Rossmann (1985b) include Arg202, Asp212, Lys236, and His304, but only three of these are conserved in HPII catalase (Arg260, Glu270, and Lys294), while the fourth (the His304 residue) was replaced by Glu362 in HPII. The conservation of these residues within HPII and other catalases emphasizes their functional importance in the catalase structure and catalysis. The testing of the predicted roles for some of these residues in HPII is described in Section III. A more complete and detailed

comparison of the catalase sequences for evolutionary purposes is described in Section II.

Section II: A Phylogenetic Survey of HPII and Related Catalases

The ancient and widespread nature of the catalase enzyme, as well as its conservation of amino acid sequence, makes it a potential candidate for use in a phylogenetic reconstruction. Nothing in the literature has been reported on the evolution of the catalase enzyme, although a few phylogenetic analyses have already been carried out on the oxygen radical scavenging superoxide dismutase enzyme (Lee et al., 1985; Smith and Doolittle, 1992). With the determination of the amino acid sequence for HPII catalase, twenty catalase primary sequences from sources that include the four kingdoms of bacteria, fungi, animals, and plants, are now available for analysis. (Because HPI catalase bears no sequence similarity with HPII or the other catalases, it was not included in the phylogenetic study.) In this preliminary investigation, these catalase sequences are subjected to a phylogenetic reconstruction by two different methods of tree building including an assessment of the branching for statistical reliability.

The amino acid sequences encoded by the different catalase genes, in addition to *katE*, were obtained from the literature, as well as from GENBANK and PIR databases (Table 7). An initial alignment of the sequences was made using the program CLUSTAL (PC/gene; IntelliGenetics Inc., Mountain View, California), followed by further adjustments based on

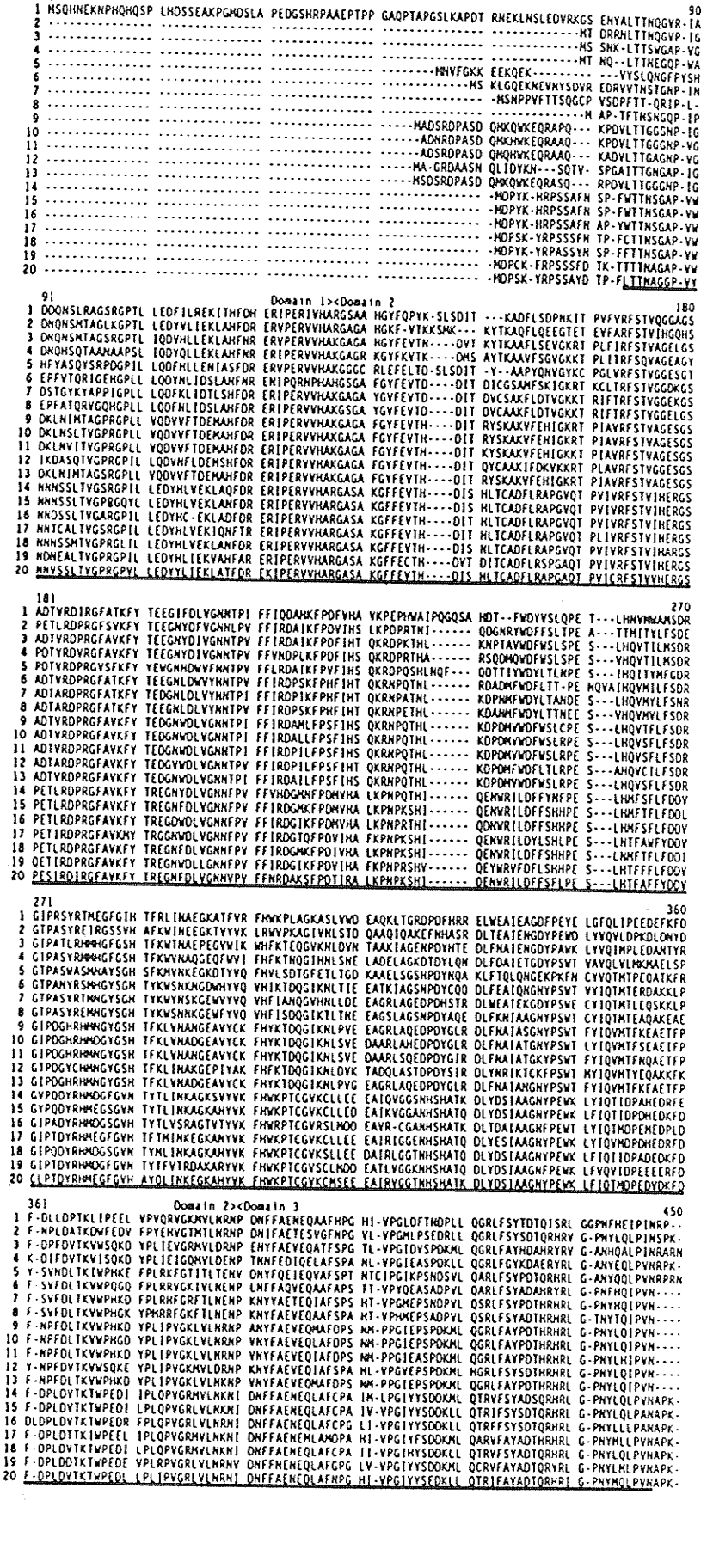
Table 7. Abbreviations, organisms, and sources of catalase amino acid sequences compiled for phylogenetic reconstruction.

	Abbreviation	Organism	Source
1.	eco	*Escherichia coli* K12	This study
2.	lse	*Listeria seeligeri*	Haas et al. (1991)
3.	bsu	*Bacillus subtilis*	Bol & Yasbin (1991)
4.	lsa	*Lactobacillus sake*	Knauf et al. (1992)
5.	sct	*Saccharomyces cerevisiae* (Type T)	Hartig & Ruis (1986)
6.	sca	*Saccharomyces cerevisiae* (Type A)	Cohen et al. (1988)
7.	hpo	*Hansenula polymorpha*	Didion & Rogenkamp (1992)
8.	ctr	*Candida tropicalis*	Okada et al. (1987)
9.	rno	*Rattus norvegicus* (Rat)	Furuta et al. (1986)
10.	bpr	*Bos primigenius* (Bovine)	Schroeder et al. (1982a)
11.	hsa	*Homo sapiens* (Human)	Bell et al. (1986)
12.	dme	*Drosophila melanogaster*	Orr et al. (1990)
13.	mmu	*Mus musculus* (Mouse)	Shaffer et al. (1990)
14.	psa	*Pisum sativum* (Pea)	Isin & Allen (1991)
15.	ghi	*Gossypium hirsutum* (Cottonseed)	Ni et al. (1990)
16.	zma	*Zea mays* (Maize)	Bethards et al. (1991)
17.	iba	*Ipomoea batatas* (Sweet potato)	Sakajo et al. (1987)
18.	ath	*Arabidopsis thaliana*	PIR Bank S18972
19.	osa	*Oryza sativa* (Rice)	Mori et al. (1992)
20.	les	*Lycopersicon esculentum* (Tomato)	GENBANK M93719

already published alignments and by eye using the program MASE (Multiple Aligned Sequence Editor; Faulkner and Jurka, 1988).

The multiple alignment for the amino acid sequences of the twenty catalases from various sources is given in Figure 10. All sequences were readily aligned with a remarkable degree of similarity considering their evolutionarily distant sources. Gaps were inserted to enhance overall sequence homology, and were largely due to the greater sequence lengths from the *E. coli* HPII and *S. cerevisiae* Type T catalases. Table 8 shows the percentage of amino acid similarity exhibited between the primary sequences for all twenty catalases. As indicated by the similarity values and the complete alignment, a high degree of sequence similarity is maintained within each of the groupings of animal, plant, and fungal catalases. A greater degree of heterogeneity exists amongst the bacterial catalases, with the catalases from *E. coli* and *Listeria seeligeri* demonstrating the least similarity to all the other catalases. As demonstrated previously (section I), the alignment of catalase sequences (Fig. 10) shows that the similarity among the various catalases is greatest near the amino-terminal region and begins to lessen at the carboxyl end of the protein. This produces a well conserved core sequence consisting of identical residues and conservative replacements. Because portions of sequence at the amino and

Figure 10. The aligned amino acid sequences of twenty catalases. Positions representing the core segment of conserved sequence used in the analysis are underlined.



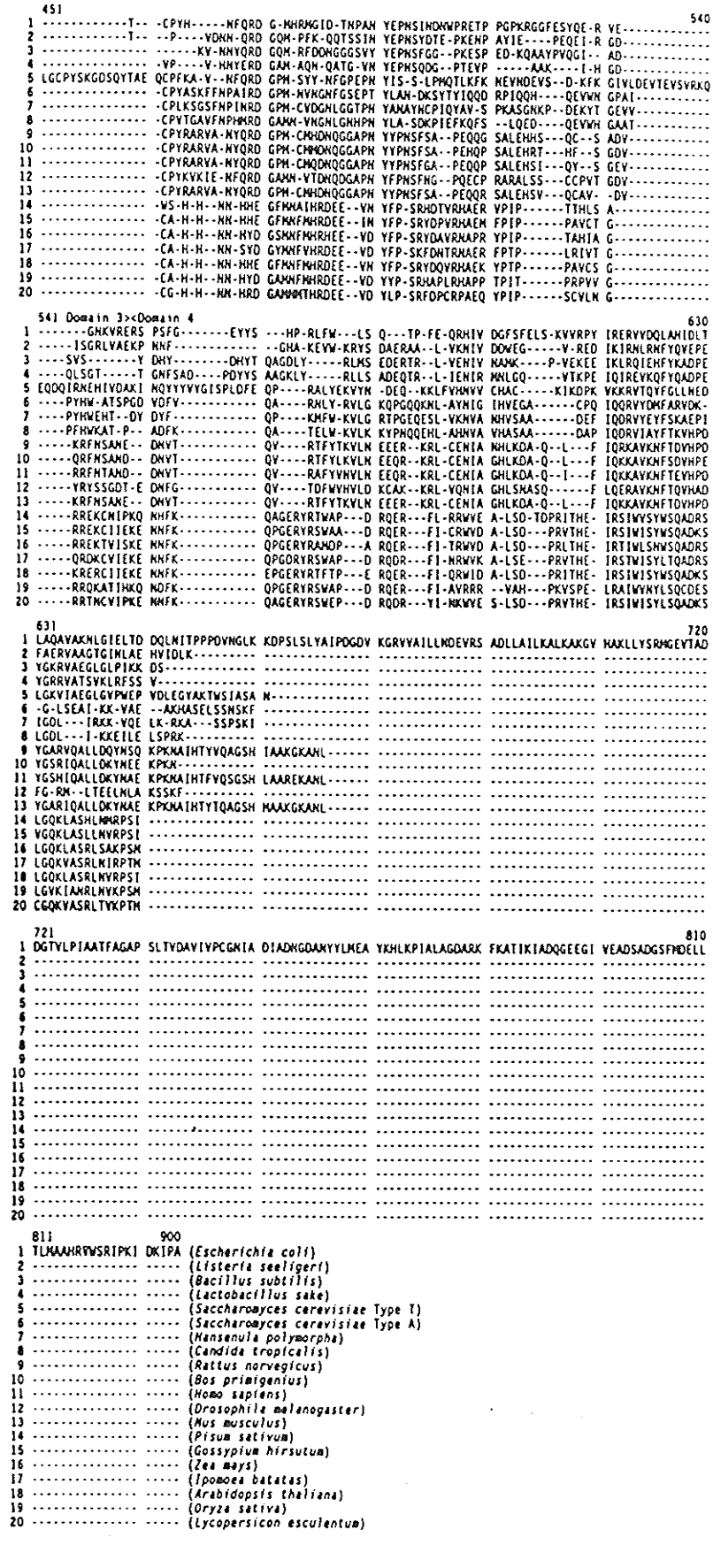


Table 8. Percent amino acid sequence similarity for the core segment of catalase sequence.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. *E. coli*	100																			
2. *L. seeligeri*	51	100																		
3. *B. subtilis*	52	57	100																	
4. *L. sake*	46	50	68	100																
5. *S. cerevisiae* T	42	45	49	47	100															
6. *S. cerevisiae* A	46	45	57	54	50	100														
7. *H. polymorpha*	49	51	57	54	54	66	100													
8. *C. tropicalis*	47	48	57	55	56	74	78	100												
9. *R. norvegicus*	50	51	68	61	50	61	64	63	100											
10. *B. primigenius*	50	51	67	61	50	60	63	62	95	100										
11. *H. sapiens*	49	50	67	61	51	60	64	62	93	94	100									
12. *D. melanogaster*	47	49	65	59	51	60	64	66	76	77	77	100								
13. *M. musculus*	50	51	68	60	50	60	64	63	98	95	93	76	100							
14. *P. sativum*	49	55	53	50	42	47	49	51	54	56	54	53	55	100						
15. *G. hirsutum*	52	57	53	49	43	47	49	50	54	55	53	52	54	90	100					
16. *Z. mays*	51	57	53	49	43	47	50	50	54	56	55	53	55	83	85	100				
17. *I. batatas*	51	53	54	47	43	46	49	49	54	56	54	52	55	81	81	76	100			
18. *A. thaliana*	52	57	54	50	43	47	49	50	55	56	54	53	54	89	93	82	81	100		
19. *O. sativa*	50	53	54	48	43	46	48	48	54	56	54	51	54	77	79	78	77	78	100	
20. *L. esculentum*	51	55	53	48	41	46	48	49	53	55	53	50	53	80	83	76	79	81	75	100

carboxyl termini lack significant similarity and are of varying lengths, the alignment of the core segment of conserved sequence, underlined in Figure 10, was used in the phylogenetic estimates. Murthy et al. (1981) previously described bovine liver catalase as being composed of four structural domains (see Introduction). The truncated segment of sequence used in the analysis contains portions of the first and third domains, and all of the second domain. Furthermore, residues implicated in the active site and heme binding (Fita and Rossmann, 1985a) conserved in all twenty sequences can be found within this core segment including His128, Ser171, and Asn206 (from the active site), and Pro404, Arg423, and Tyr427 (proximal side heme binding). On the distal side, the Phe residues at 211 and 219 are conserved in all twenty sequences but Val at 127 is not conserved in *S. cerevisiae* Type A and Thr at 172 is not conserved in *L. sake*.

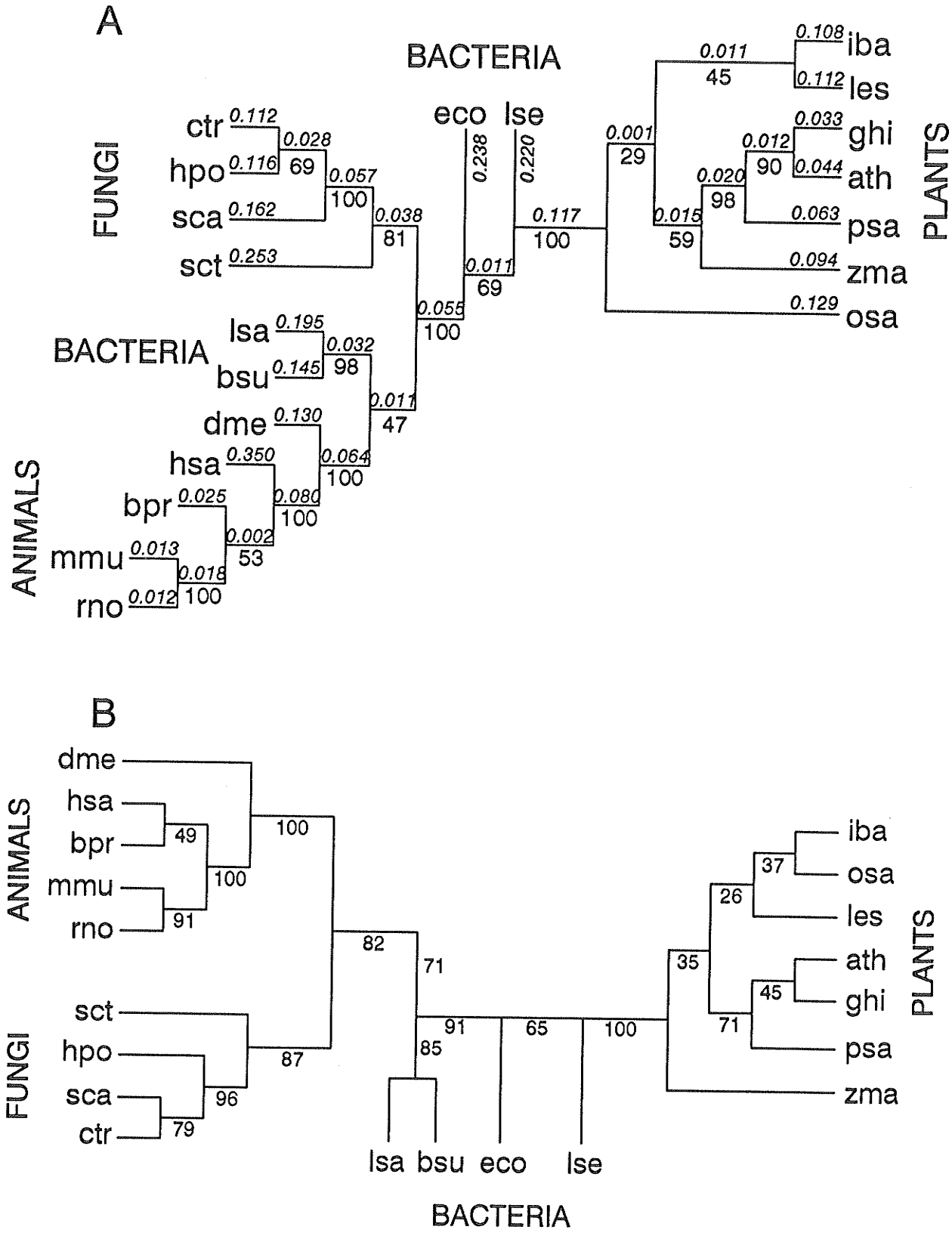
An estimate of the phylogenetic relationship of the catalases from the various sources was carried out using the truncated segment of the sequence described above. To maintain a level of confidence in the phylogenetic reconstruction, two different methods of tree building, the distance neighbor-joining method and protein parsimony, were used, both of which were subjected to a bootstrapping analysis involving 100 replicates to estimate the level of confidence in the branching. Because of the small number of

catalase sequences available for the study, the phylogenetic trees were not rooted, although the *E. coli* sequence would probably have been selected for the rooting option on the basis of branch lengths.

Generally, both methods of tree construction, neighbor-joining (Fig. 11A) and parsimony (Fig. 11B), give similar topologies with respect to the monophyletic clustering of groups of organisms, although some minor internal anomalies are evident. Furthermore, bootstrap values greater than 95% suggest that the branching order for some of these major clusterings are reliable. For example, the plant catalases can be considered as a monophyletic group owing to the bootstrap probability of 100%. This level of confidence is obtained from both the distance and parsimony methods, although the two methods give different internal branching orders with less reliable bootstrap support, and thus lower internal resolution. With protein parsimony, catalases from sweet potato and rice form part of the same internal branch despite their respective dicot and monocot nature. With the distance method, this grouping does not occur, but rather a more likely pairing of catalases from the two dicots, sweet potato and tomato, is found. Surprisingly, the internal branching of the catalases from the two monocots, rice and maize, does not occur in either of the two methods of tree building.

For the mammalian catalases, both methods also give

Figure 11. Unrooted phylogenetic trees based on the amino acid sequences of twenty catalases generated by the neighbor-joining (**A**) and parsimony (**B**) methods. The numbers at the nodes below the line indicate the level of confidence for the major branches as derived by bootstrap analysis. The numbers on top of the branches of the distance tree (**A**) represent the branch lengths as determined by the neighbor-joining program.


A
BACTERIA
eco
lse
0.238
0.220
0.117
100
0.055
100
0.011
69
0.011
45
0.108
iba
0.112
les
0.001
29
0.033
ghi
0.012
90
0.044
ath
0.020
98
0.063
psa
0.015
59
0.094
zma
0.129
osa
PLANTS
FUNGI
ctr
0.112
0.028
hpo
0.116
69
0.057
100
sca
0.162
0.038
81
sct
0.253
BACTERIA
lsa
0.195
0.032
bsu
0.145
98
0.011
47
dme
0.130
hsa
0.350
0.064
100
bpr
0.025
0.080
100
0.002
53
mmu
0.013
0.018
rno
0.012
100
ANIMALS
B
dme
hsa
bpr
49
100
mmu
91
rno
100
82
71
sct
hpo
87
sca
96
ctr
79
91
85
65
100
35
iba
37
osa
26
les
ath
45
ghi
71
psa
zma
PLANTS
FUNGI
ANIMALS
lsa
bsu
eco
lse
BACTERIA

bootstrap support of 100%, indicating this monophyletic grouping is reliable. Internally, the two tree building methods display the expected pairing of catalases from rat and mouse, although parsimonic analysis exhibits a lower confidence level of 91%, in contrast to 100% in the distance tree. Both methods also strongly support with greater than 95% confidence the grouping of the catalase from *Drosophila melanogaster* with the mammalian catalases to produce a monophyletic grouping based on animal sequences.

In the case of the fungal catalases, both of the tree building methods clearly demonstrate that catalases from the three yeasts, *Hansenula polymorpha*, *Candida tropicalis*, and *Saccharomyces cerevisiae* Type A, form a monophyletic grouping with a high level of confidence (>95%), but with a few variations in the internal branching order. The fourth catalase used in the analysis, catalase Type T from *S. cerevisiae*, also clusters with the yeasts, but the reliability for this branching is not quite as strong as for the other three catalase sequences. This weakened linkage is expected because the primary sequence from catalase Type T is longer than the other three fungal catalase sequences and the subcellular locations also differ. *Hansenula*, *Candida*, and *S. cerevisiae* Type A catalases are all considered to be peroxisomal and would naturally be expected to exhibit a greater similarity with one another than with *S. cerevisiae* Type T catalase which is a cytosolic enzyme.

The lower confidence for this branching of the fungal monophyletic group and the branch lengths from the distance tree support these conclusions. It is apparent that the phylogenetic reconstruction of the fungal catalase sequences confirms differences based on cellular localization. The concept that peroxisomal proteins form an evolutionarily homogeneous group is also supported by a recent phylogeny reconstruction based on the thiolase enzyme (Igual et al., 1992).

Unlike the plant, animal, and fungal catalases, the four bacterial catalases fail to cluster together to form a single monophyletic group, thereby suggesting a polyphyletic origin. Both types of trees display a grouping of the catalases from the gram-positive organisms, *Bacillus subtilis* (catalase-1) and *Lactobacillus sake*, although only the distance method has a confidence level exceeding 95% for this grouping. It is likely that insufficient information is available to reliably resolve these bacterial sequences. The two bacterial catalases are found between the animal and fungal catalases in the tree constructed with the distance method, in contrast to the parsimony method which places them outside a node linking the animal and fungal catalases. The grouping of these catalases is expected considering that both *the Bacillus* and *Lactobacillus* catalases are reported to have hexameric structures (Loewen and Switala, 1987; Knauf et al., 1992).

The two remaining bacterial catalases from *E. coli* and *Listeria seeligeri* do not group with any other catalases and are joined by a branch in the distance and parsimony trees with confidence levels of 69% and 65%, respectively. This is understandable for the *E. coli* HPII catalase because it has several anomalous physical properties when compared to most other typical catalases including its quaternary structure, subunit size, and type of heme. Therefore, *E. coli* HPII catalase is expected to remain outside of any monophyletic grouping. On the other hand, *Listeria seeligeri* catalase, which is enzymatically and structurally very similar to the typical catalases (Haas et al., 1991), cannot have its lack of clustering convincingly accounted for. Possibly these results suggest that the bacterial catalases are evolving at a more rapid rate than catalases from other sources.

Section III: Site-Directed Mutagenesis of HPII Catalase

The discovery that the predicted amino acid sequence of HPII catalase shows striking similarity to the sequence of bovine liver catalase including the conservation of residues implicated in the active site and heme binding will facilitate an investigation of the structure-function relationship of the catalase enzyme. Based on the extensive sequence similarity between HPII and the bovine liver enzyme, the 3-dimensional structural arrangement of the active center proposed for bovine liver catalase (Fita and Rossmann, 1985a) may also exist in HPII catalase. In this study, the predicted roles for three active site residues (His128, Ser167, and Asn201) and one proximal-side heme binding residue (Tyr415) in HPII catalase were tested by substituting the various residues using site-directed mutagenesis. Table 9 gives a listing of the various amino acid substitutions chosen for the His128, Ser167, Asn201, and Tyr415 residues, and summarizes the rationale for their selection.

Construction and Production of Mutant HPII Catalases

The plasmid pAMkatE72, which contains a 3,466 bp *Pst*I-*Cla*I DNA fragment encompassing the *katE* gene, was used to construct the subclones necessary for the generation of the mutant HPII protein. A 0.6 kb DNA fragment containing the

Table 9. Amino acid substitutions of HPII catalase and rationale for selection.

Amino Acid Substitution	Rationale for Replacement
His128-Ala	o No role in the reaction o No steric hindrance o Functions as a null mutant
His128-Glu	o Possible H-bonding with peroxide or Ser167, but not both
His128-Gln	o Same as Glu replacement
His128-Asn	o Same as Glu and Gln replacements, but with less steric hindrance
Ser167-Ala	o Lacks the hydroxyl group for H-bonding o Functions as a null mutant
Ser167-Thr	o Maintains the hydroxyl group, but is larger sized and creates greater steric hindrance
Ser167-Cys	o Possible mimicking of H-bonding, but with less efficiency
Ser167-Asn	o Maintains an interaction with His128, but will not react with the heme group
Asn201-Ala	o Functions as a null mutant
Asn201-His	o Possible mimicking of Asn, but larger sized causing greater steric hindrance
Asn201-Asp	o Lacks the amide group, but may still function normally
Asn201-Gln	o Extends the side chain by one methylene group and may function normally, but less efficiently
Asn201-Arg	o Found in place of Asn in peroxidases
Tyr415-Phe	o Lacks the hydroxyl group, but maintains the normal sizing
Tyr415-His	o Found in place of Tyr in peroxidases

Ser167 and Asn201 codons was removed from pAMkatE72 with *Hind*III and *Eco*RI and then ligated into the corresponding restriction sites of pKS- producing the subclone pKS-HE. Similarily for the substitution of Tyr415, a 1.6 kb *Eco*RI-*Cla*I DNA fragment was removed from pAMkatE72 and cloned into pKS+ creating the subclone pKS+EC. The subclone, pKS+6 containing a 1.25 kb *Pst*I-*Hind*III fragment from pAMkatE72 in pKS+, was used for the replacement of His128 and had been previously constructed for *katE* sequencing (Table 2). The locations of the restriction sites for the DNA fragments used to construct the three subclones are indicated in Figure 2. The advantage of using smaller segments of *katE* for mutagenesis is that only these smaller DNA segments will have to be sequenced following mutagenesis to confirm that only the desired base changes were obtained.

Each of the three subclones (pKS+6, pKS-HE, and pKS+EC) were transformed into the *dut*$^-$ *ung*$^-$ strain CJ236 to produce the uracil-containing single-stranded DNA template required for mutagenesis. Site-directed mutagenesis was carried out according to the Kunkel method using the phosphorylated mutagenic primers listed in Table 10. Screening for the anticipated base changes was carried out by single-stranded DNA sequencing. The autoradiograms showing the actual base changes obtained for the respective amino acid substitutions are presented in Figure 12. The various DNA fragments identified as having the proper base changes were recovered

Table 10. Synthetic oligonucleotide primers used to create base changes for the production of amino acid substituted mutants of HPII catalase.

Primer	Oligonucleotide Sequence[a]	Location[b] (bp)
H128A	5′-GCTATTGTT*GCT*GCACGCGG	1193 to 1212
H128N	5′-GCTATTGTT*AAT*GCACGCGG	1193 to 1212
H128E	5′-GCTATTGTT*GAA*GCACGCGG	1193 to 1212
H128Q	5′-GCTATTGTT*CAA*GCACGCGG	1193 to 1212
S167A	5′-ACGTTTC*GCT*ACCGTTCAGGGT	1312 to 1333
S167C	5′-ACGTTTC*TGT*ACCGTTCAGGGT	1312 to 1333
S167N	5′-ACGTTTC*AAT*ACCGTTCAGGGT	1312 to 1333
S167T	5′-ACGTTTC*ACT*ACCGTTCAGGGT	1312 to 1333
N201A	5′-CTCGTTGGC*GCT*AACACGCCA	1412 to 1432
N201H	5′-CTCGTTGGC*CAT*AACACGCCA	1412 to 1432
N201D	5′-CTCGTTGGC*GAT*AACACGCCA	1412 to 1432
N201Q	5′-CTCGTTGGC*CAA*AACACGCCA	1412 to 1432
N201R	5′-CTCGTTGGC*CGT*AACACGCCA	1412 to 1432
Y415F	5′-GTTCTCC*TTT*ACCGATACA	2056 to 2074
Y415H	5′-GTTCTCC*CAT*ACCGATACA	2056 to 2074

[a] Target codon sites are italicized and bases changed to produce the new amino acids are underlined.

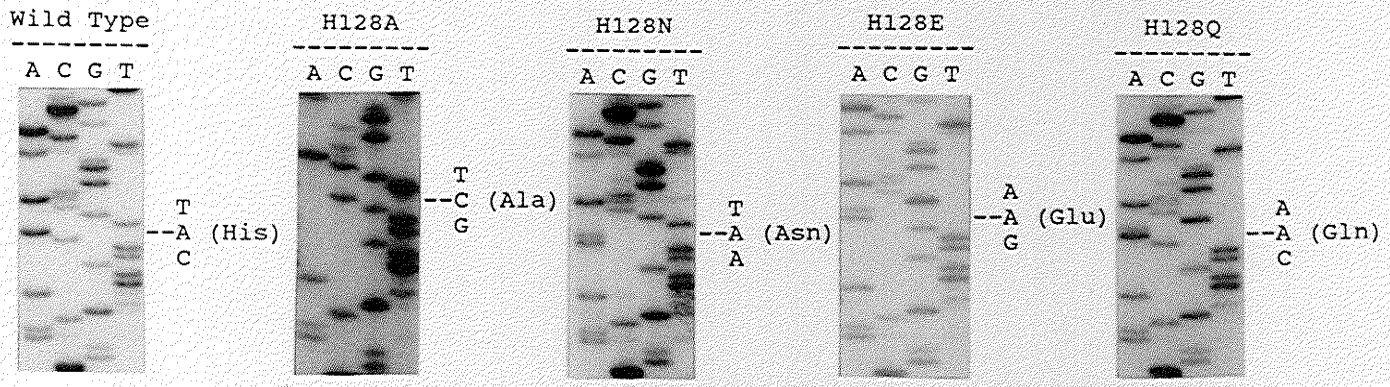
[b] Numbers refer to the location of the primers on the *katE* sequence presented in Figure 3.

Figure 12. Autoradiograms of sequencing gels revealing base changes in the codon sites for His128, Ser167, Asn201, and Tyr415 in HPII catalase.

(A) Base changes required for the replacement of His128 (CAT) in the wild type enzyme with the amino acids Ala (GCT), Asn (AAT), Glu (GAA), and Gln (CAA).

(B) Base changes required for the replacement of Ser167 (TCT) in the wild type enzyme with the amino acids Ala (GCT), Cys (TGT), Asn (AAT), and Thr (ACT).

(A) His128



(B) Ser167

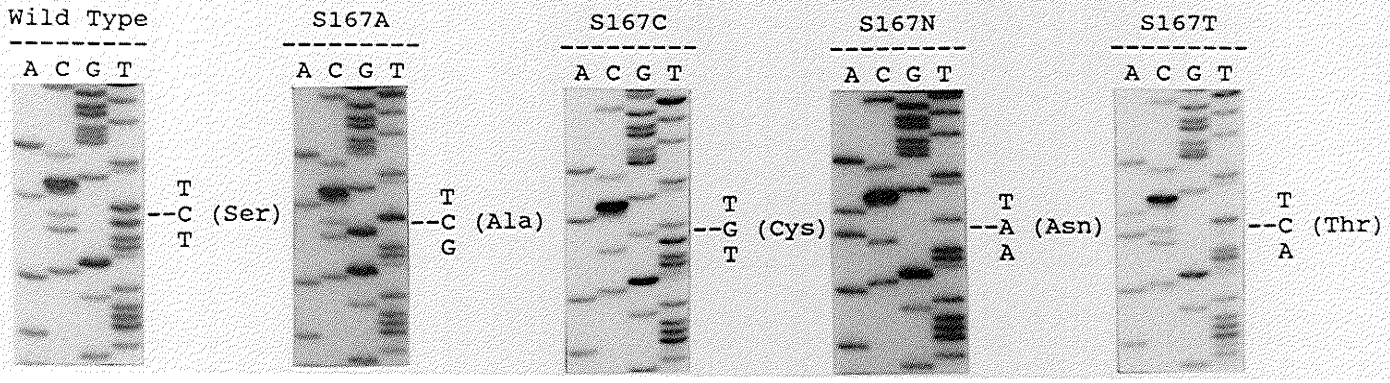
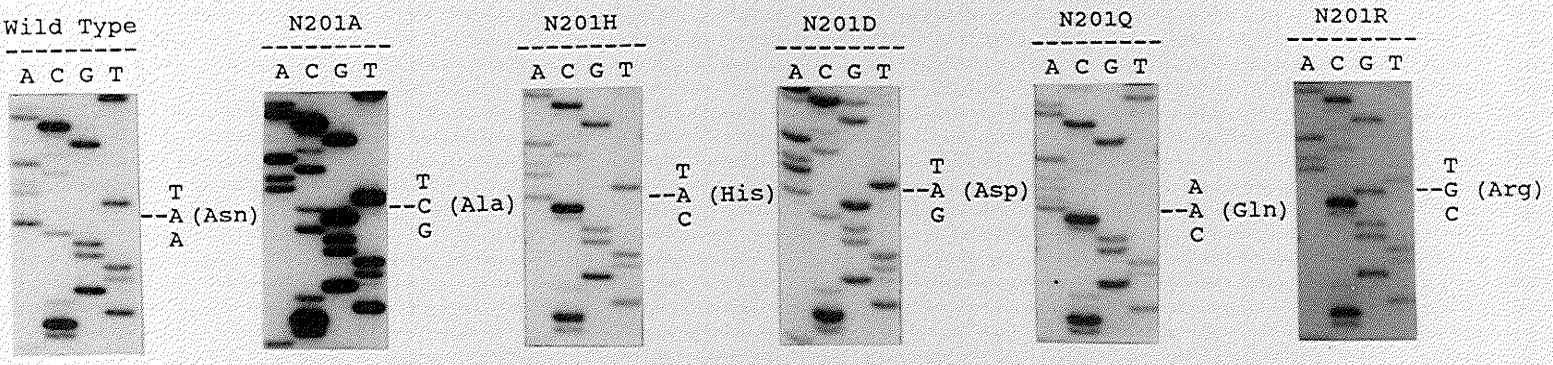


Figure 12. (continued)

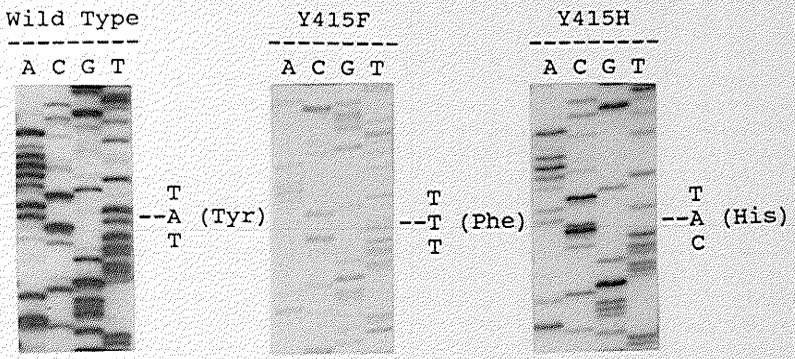
(C) Base changes required for the replacement of Asn201 (AAT) in the wild type enzyme with the amino acids Ala (GCT), His (CAT), Asp (GAT), Gln (CAA), and Arg (CGT).

(D) Base changes required for the replacement of Tyr415 (TAT) in the wild type enzyme with the amino acids Phe (TTT) and His (CAT).

(C) Asn201



(D) Tyr415



from the subclones and inserted back into pAMkatE72 cut with appropriate enzymes to generate the mutant *katE* gene. In this way plasmids containing mutated *katE* genes were generated with base changes only at the codon sites for His128 (pEkatH128A, pEkatH128N, pEkatH128E, and pEkatH128Q), Ser167 (pEkatS167A, pEkatS167C, pEkatS167T, and pEkatS167N), Asn201 (pEkatN201H, pEkatN201D, pEkatN201A, pEkatN201Q, and pEkatN201R), and Tyr415 (pEkatY415F and pEkatY415H).

To produce the mutant HPII catalases, the fifteen mutant plasmids were transformed into the *katE*⁻ and *katG*⁻ strain UM255, which does not produce either HPII or HPI. In this strain, the only catalase (mutant or wild type) produced will be that expressed from the resident plasmid. Confirmation of HPII protein production in each of the mutant clones was obtained by screening crude extracts of overnight cultures for catalase activity and HPII-like protein on SDS-polyacrylamide gels. For comparison, crude extracts were prepared from untransformed UM255 and from UM255 containing pAMkatE72, which produces wild type HPII.

As indicated in Table 11, four Asn201 substituted mutants (N201A, N201H, N201D, and N201Q) and two His128 substituted mutants (H128A and H128N) accumulated HPII-like protein in amounts similar to the wild type enzyme, although only the crude extracts of the four Asn201 mutants exhibited catalase activity. The remaining nine mutants did not accumulate detectable HPII-like protein (Table 11),

Table 11. Summary of catalase activity and HPII-like protein content in crude extracts from strains transformed with plasmids containing mutant *katE* genes.

Mutant	Protein	Activity
H128A	yes	no
H128N	yes	no
H128E	no	no
H128Q	no	no
S167A	no	no
S167C	no	no
S167N	no	no
S167T	no	no
N201A	yes	yes
N201H	yes	yes
N201D	yes	yes
N201Q	yes	yes
N201R	no	no
Y415F	no	no
Y415H	no	no

suggesting that the amino acid replacements had destabilized the protein structure resulting in increased sensitivity to proteolysis.

Characterization of Mutant HPII Catalases

The six mutant HPII catalases and the wild type enzyme were purified as described in Experimental Procedures. The purified preparations of the wild type and mutant HPII catalases were examined on SDS-polyacrylamide gels and found to be 85-90% pure. The specific activities of wild type HPII catalase as reported in the literature (14,800 U/mg) (Loewen and Switala, 1986), and as reported in this study, are in good agreement. All four Asn201 mutant enzymes exhibited greatly reduced specific activities compared to the wild type enzyme (Table 12). Of the mutants, the N201D and N201A mutant enzymes displayed the highest activity (about 10% of wild type), the N201H and N201Q mutants displayed significantly lower activities (less than 1% of wild type), and the H128A and H128N mutants displayed no catalase activity. Consequently, characterization of these latter two mutant enzymes on the basis of activity changes was not possible. Nonetheless, the above results indicate that each of the six amino acid substitutions affected the normal catalytic activity of HPII.

For three of the mutant enzymes, a noteworthy difference was observed in the coloration of the enzyme.

Table 12. Kinetic parameters of the wild type and the Asn201 and His128 mutant HPII catalases.

Enzyme	Activity (U/mg)	Km[a] (mM)	Vmax[a] (mol/min/g)	k_{cat} (min^{-1})	$\Delta\Delta G_S$ [b] (kcal/mol)	$\Delta\Delta G_T'$ [c] (kcal/mol)
WT	15,200	214 ± 41	68.7 ± 8.1	(3.47 ± 0.42) x 10^7	-	-
N201H	100	309 ± 60	0.8 ± 0.1	(4.04 ± 0.53) x 10^5	-0.23	-3.0
N201D	1,700	166 ± 35	6.7 ± 0.8	(3.38 ± 0.41) x 10^6	+0.16	-1.3
N201A	1,300	143 ± 30	5.8 ± 0.7	(2.93 ± 0.35) x 10^6	+0.25	-1.3
N201Q	50	91 ± 20	0.2 ± 0.02	(1.01 ± 0.10) x 10^5	+0.53	-3.1
H128A	<0.1	-	-	-	-	-
H128N	<0.1	-	-	-	-	-

[a] Apparent Km and Vmax.

[b] Calculated from $\Delta\Delta G_S = RT\ln\{Km_{(wt)}/Km_{(mut)}\}$ (Moody and Wilkinson, 1990).

[c] Calculated from $\Delta\Delta G_T' = RT\ln\{(k_{cat}/Km)_{mut}/(k_{cat}/Km)_{wt}\}$ (Moody and Wilkinson, 1990).

The color of purified wild type HPII catalase is normally dark green, a result of the *cis*-heme d-isomer bound to the enzyme (Chiu et al., 1989). Three of the mutant enzymes (N201Q, N201A, and N201D) maintained this coloration, but N201H, H128A, and H128N exhibited a dark reddish-brown color, characteristic of protoheme-containing enzymes such as HPI and bovine catalase.

Steady-State Kinetic Properties of Asn201 Mutant Catalases

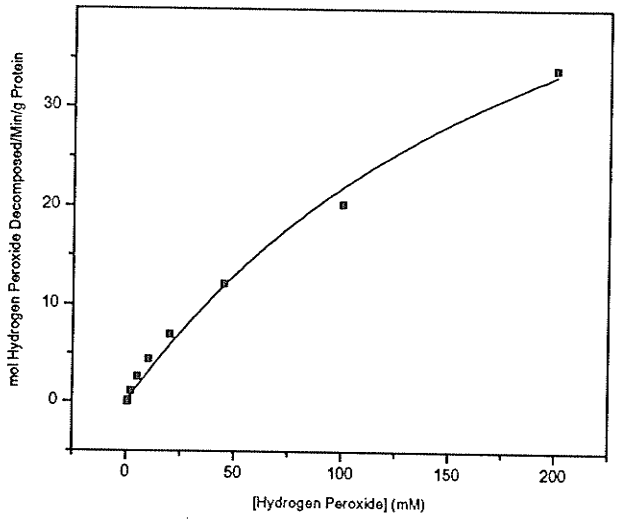
As described in the preceding section, all four Asn201 mutant enzymes displayed reduced specific activities relative to the wild type form, suggesting that differences in their other kinetic properties might also be evident. Aside from the preliminary determination of an apparent Km of 30.9 mM for HPII catalase in a crude extract (Loewen and Triggs, 1984), no further kinetic analysis of HPII catalase has ever been carried out. Catalases generally do not lend themselves to easy kinetic analysis since the normal pattern of Michaelis-Menten saturation kinetics is not followed (Abei, 1974). For example, enzyme saturation with the hydrogen peroxide substrate cannot be obtained within a workable concentration range with 5 M hydrogen peroxide being required to saturate some catalases. Moreover, hydrogen peroxide concentrations above 100 mM cause autoinactivation of the enzyme making determination of enzyme activity under saturating conditions very difficult.

A rapid-flow, two mixer assay method has been used to bring about a quenching effect that alleviates the inhibition caused by higher substrate concentrations (Ogura, 1955), but the present study did not utilize such an assay method.

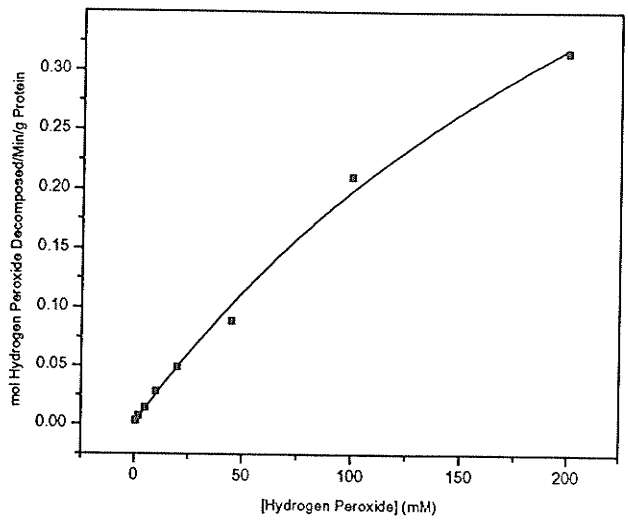
As a means of comparing the mutant and wild type enzymes, the dependence of the enzymic reaction rates on substrate concentration was examined and used to obtain steady-state kinetic and binding energy parameters. Figures 13A-E show that the velocity versus substrate concentration curves for the wild type and four Asn201 mutant enzymes resemble the hyperbolic pattern typical of Michaelis-Menten kinetic theory. Saturation curves were generated by fitting the data using the non-linear least squares method offered within the Origin (version 2) software package (MicroCal Software Inc., North Hampton, MA.) The apparent Vmax (the maximum initial reaction velocity) and Km (the substrate concentration at which the initial reaction velocity is equivalent to half the maximum initial reaction velocity) values for each of the enzymes (Table 12) were calculated directly from the saturation curves using the Origin computer program. Similar Vmax and Km values were obtained using the Sigma Plot software package (Jandel, San Rafael, CA.). Since the rates on the oxygraph at high hydrogen peroxide concentrations exhibited no falling off, enzyme inactivation does not appear to be present and, consequently, the kinetic parameters obtained from the curve

Figure 13 (A-E). The effect of hydrogen peroxide concentrations on the initial velocity of wild type and Asn201 mutant HPII catalases. (A, Wild Type; B, N201H; C, N201D; D, N201A; E, N201Q)

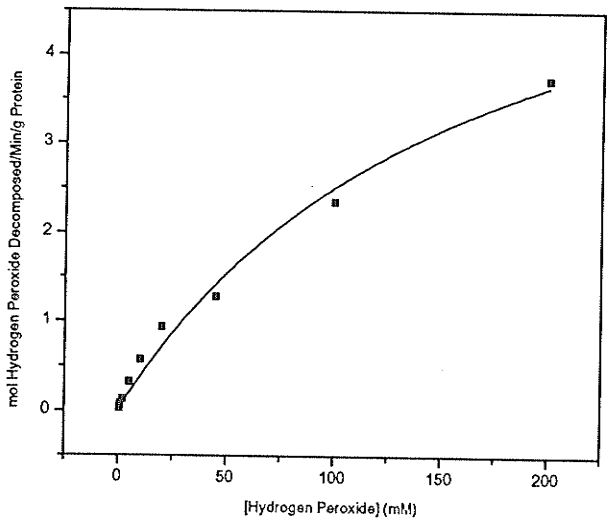
A


mol Hydrogen Peroxide Decomposed/Min/g Protein
0
10
20
30
0
50
100
150
200
[Hydrogen Peroxide] (mM)

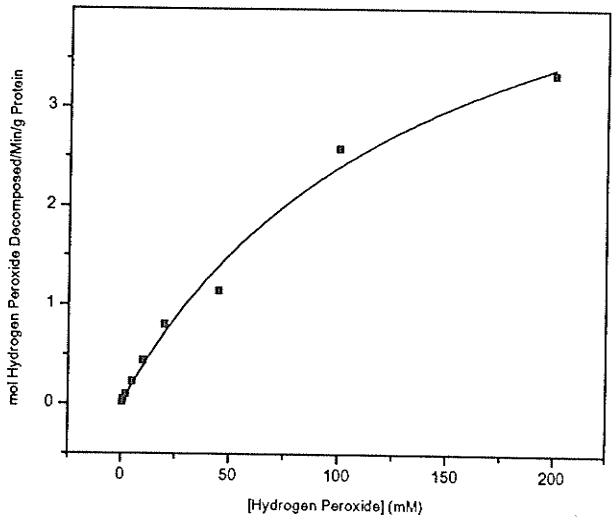
B


mol Hydrogen Peroxide Decomposed/Min/g Protein
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
50
100
150
200
[Hydrogen Peroxide] (mM)

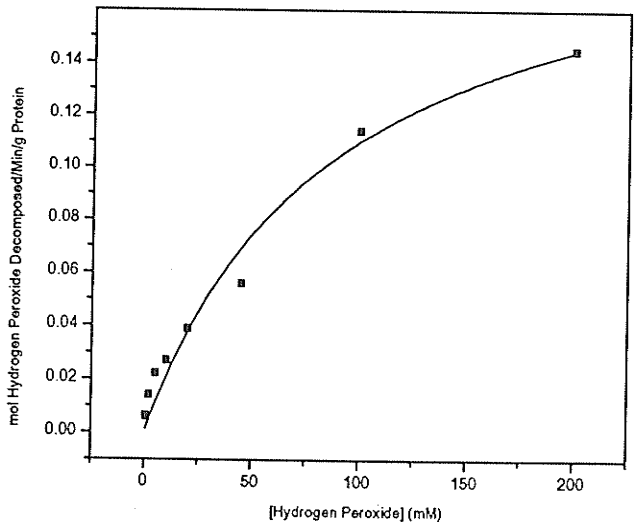
C


mol Hydrogen Peroxide Decomposed/Min/g Protein
0
1
2
3
4
0
50
100
150
200
[Hydrogen Peroxide] (mM)

D


mol Hydrogen Peroxide Decomposed/Min/g Protein
0
1
2
3
0
50
100
150
200
[Hydrogen Peroxide] (mM)

E


mol Hydrogen Peroxide Decomposed/Min/g Protein
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
50
100
150
200
[Hydrogen Peroxide] (mM)

are likely valid. The Vmax and Km values were subsequently used to calculate the k_{cat}, $\Delta\Delta G_B$, and $\Delta\Delta G_T^{\ddagger}$ values for each of the enzymes. k_{cat} represents the number of substrate molecules converted to product per unit time by each active site of the enzyme at saturating substrate concentrations (the turnover number) and when compared between mutant and wild type enzymes, can be used for interpreting the structural effects of the mutation (Moody and Wilkinson, 1990). $\Delta\Delta G_B$ is the difference in free energy of binding between mutant and wild type enzymes and represents the change in enzyme-substrate binding energy caused by the mutation. Therefore, if the mutant enzyme binds the substrate less efficiently than the wild type enzyme, the value for $\Delta\Delta G_B$ is negative (Moody and Wilkinson, 1990). $\Delta\Delta G_T^{\ddagger}$ is the difference in enzyme-transition state binding energies between the mutant and wild type enzymes, and reflects the change in the enzyme-transition state binding energy caused by the mutation. Consequently, $\Delta\Delta G_T^{\ddagger}$ is negative when the mutation has a destabilizing effect on transition state binding (Moody and Wilkinson, 1990).

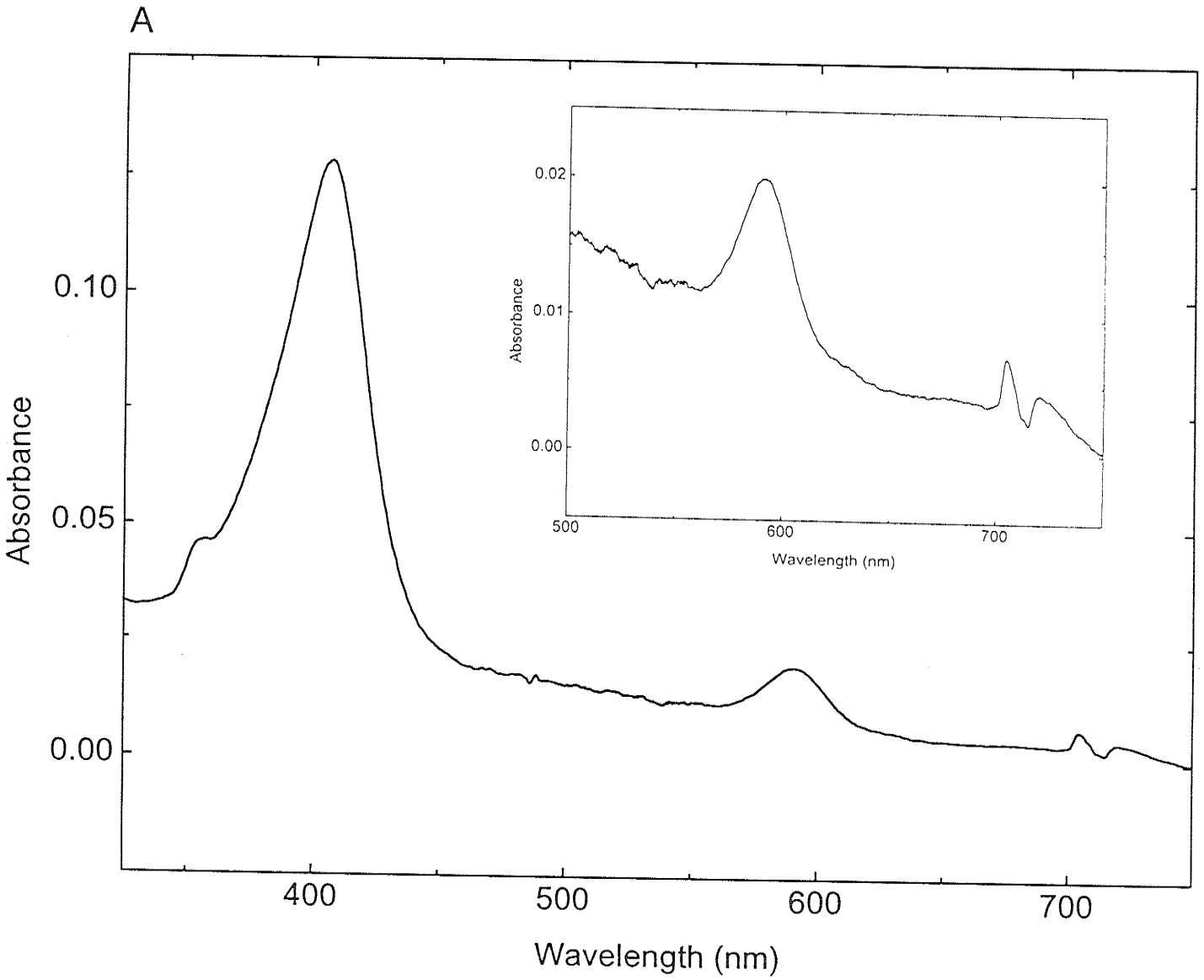
As shown in Table 12, the apparent Km for wild type HPII catalase is inconsistent with the literature value (see above) and may reflect differences in the way the kinetic data was generated. With the level of error taken into account, the apparent Km values for three of the mutant enzymes (N201H, N201D, and N201A) are within the same range

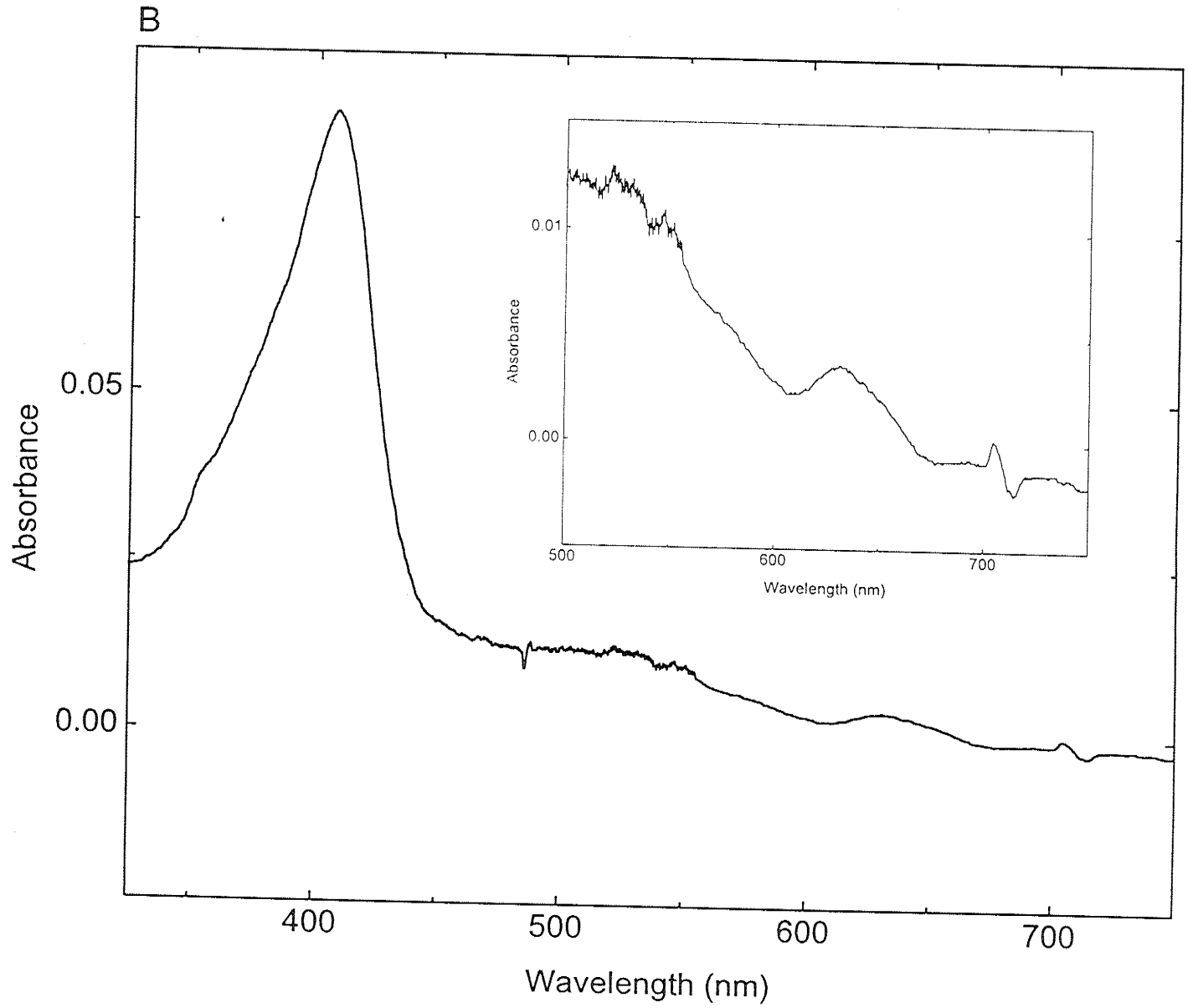
as the value for the wild type enzyme. Only the Km for N201Q was smaller. All of the mutant enzymes displayed smaller Vmax values ranging from 9.8% and 8.4% of the wild type value for N201D and N201A to 1.2% and 0.3% of the wild type value for N201H and N201Q. A similar pattern in reduced turnover numbers was also evident. Based on the free energy values, it appears that each of the mutations had caused a change in substrate and transition state binding. However, because of the high level of error in the variables (Vmax and Km) used in the free energy calculations ($\Delta\Delta G_B$ and $\Delta\Delta G_T^{\ddagger}$), the accuracy of these binding energies is doubtful.

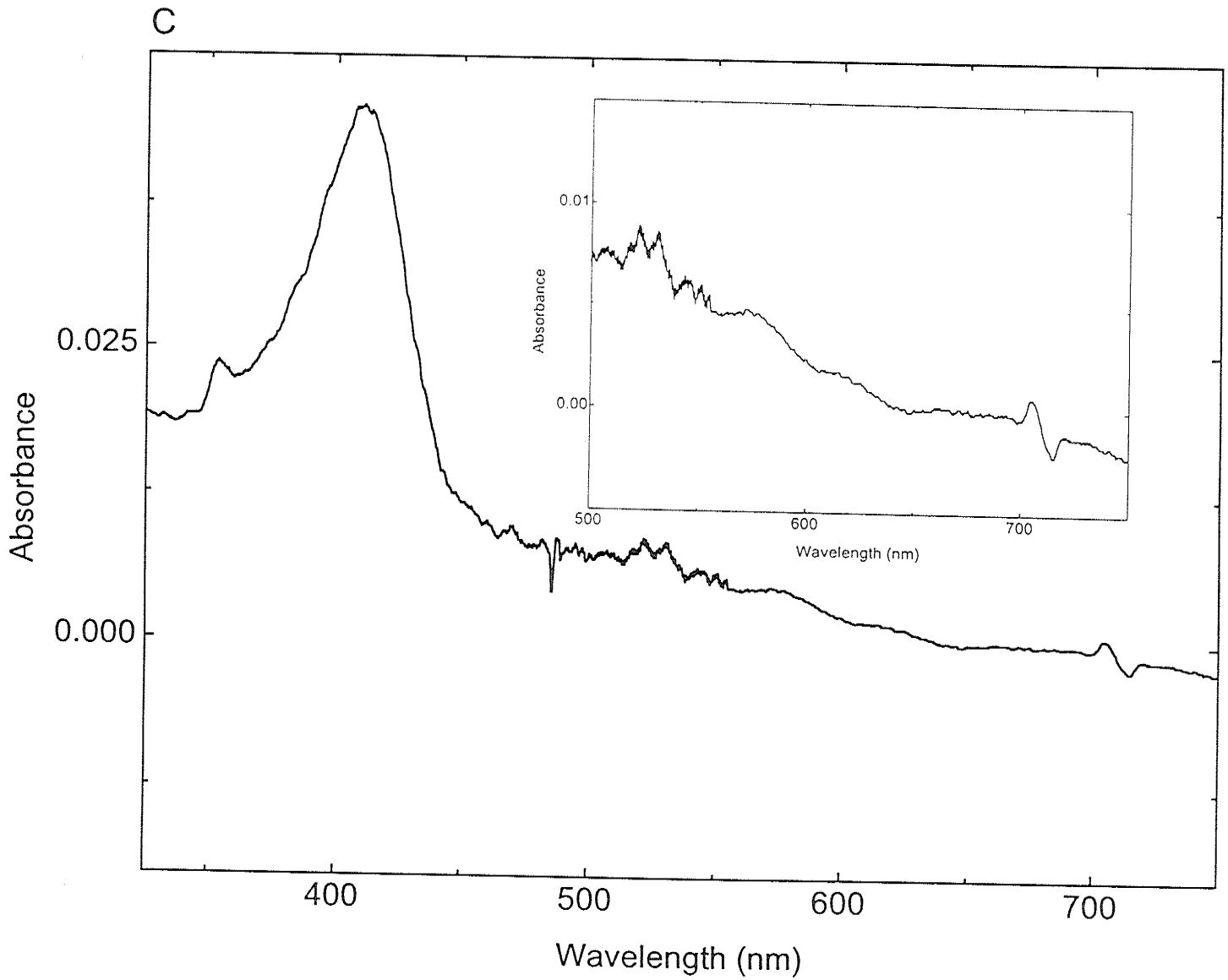
Spectral Properties of the Mutant Catalases

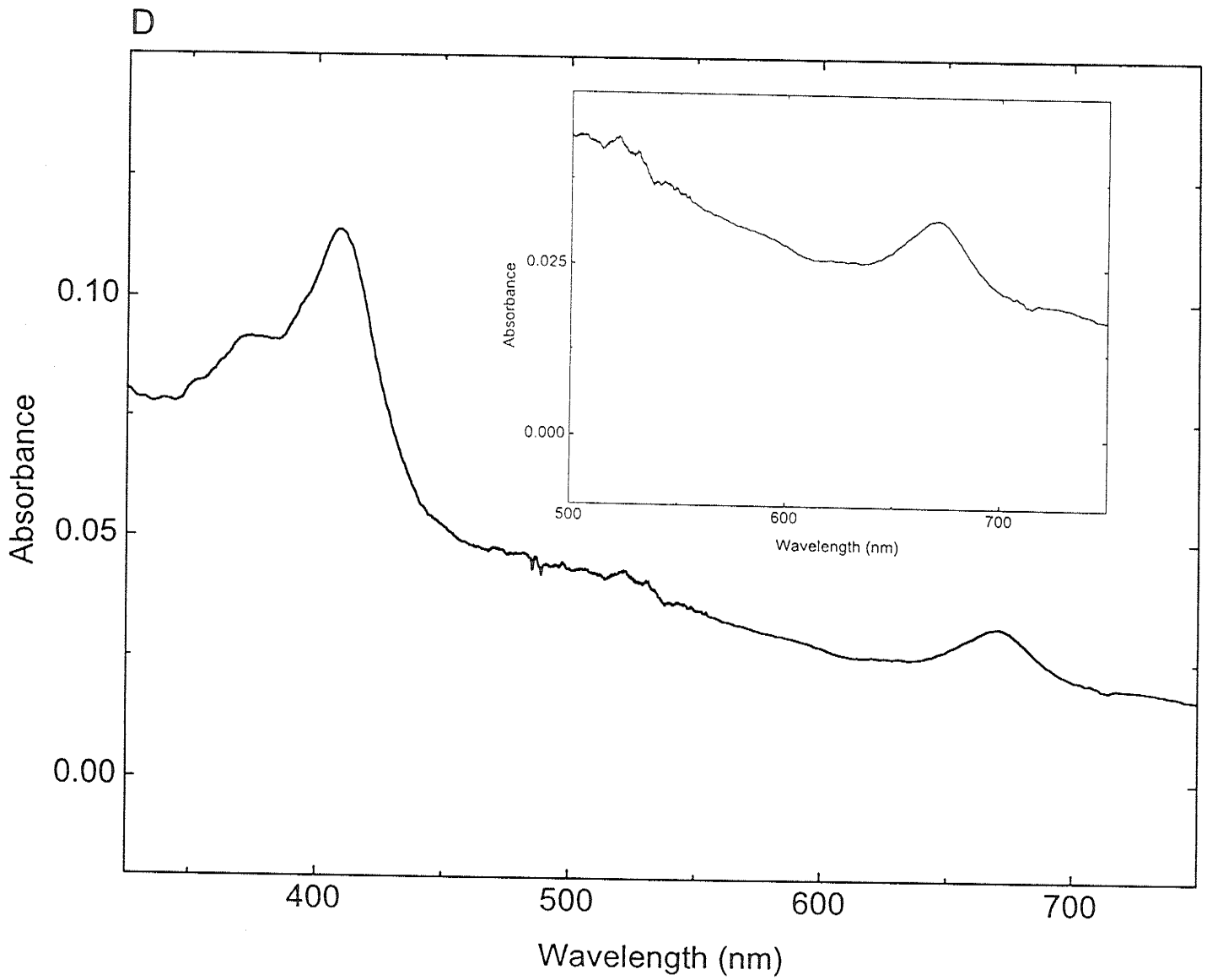
Loewen and Switala (1986) reported that the absorption spectrum for HPII catalase consists of a prominent Soret peak at 406 nm, as well as smaller peaks at 590, 630, and 715 nm. A similar spectrum was obtained in this work (Fig. 14A), although the 630 nm band was missing. The UV and visible spectra for each of the Asn 201 and His128 mutant enzymes were also determined (Figs. 14B-G) and surprisingly each was found to be different. When compared to the spectrum for wild type HPII which contains the *cis* heme d-isomer (Chiu et al., 1989) and HPI which contains protoheme IX (Claiborne and Fridovich, 1979), the spectra of the mutant enzymes appear to represent mixtures of the two

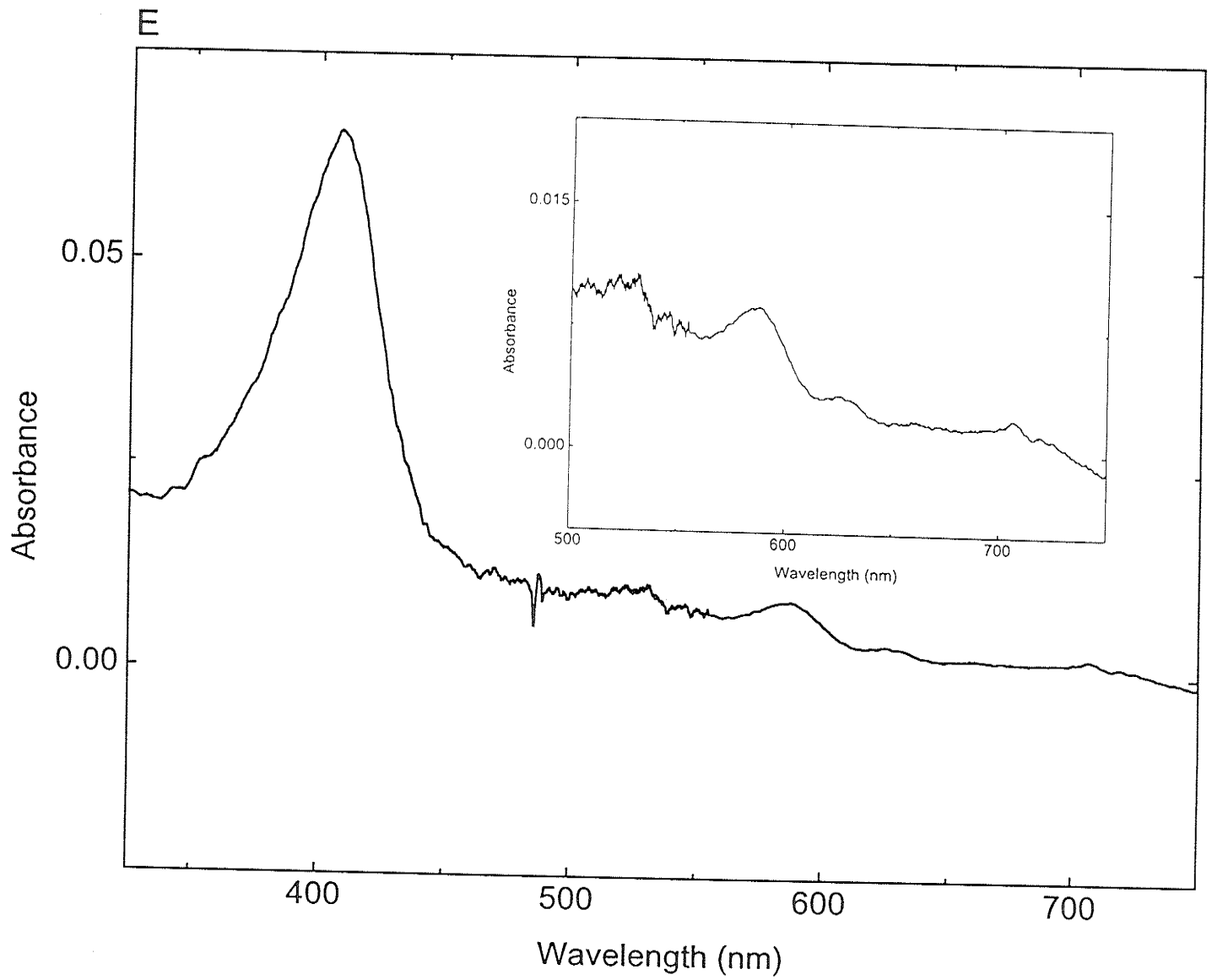
Figure 14 (A-G). Absorption spectra of wild type, His128, and Asn201 mutant HPII catalases. (A, Wild Type; B, N201H; C, N201D; D, N201A; E, N201Q; F, H128A; G, H128N)

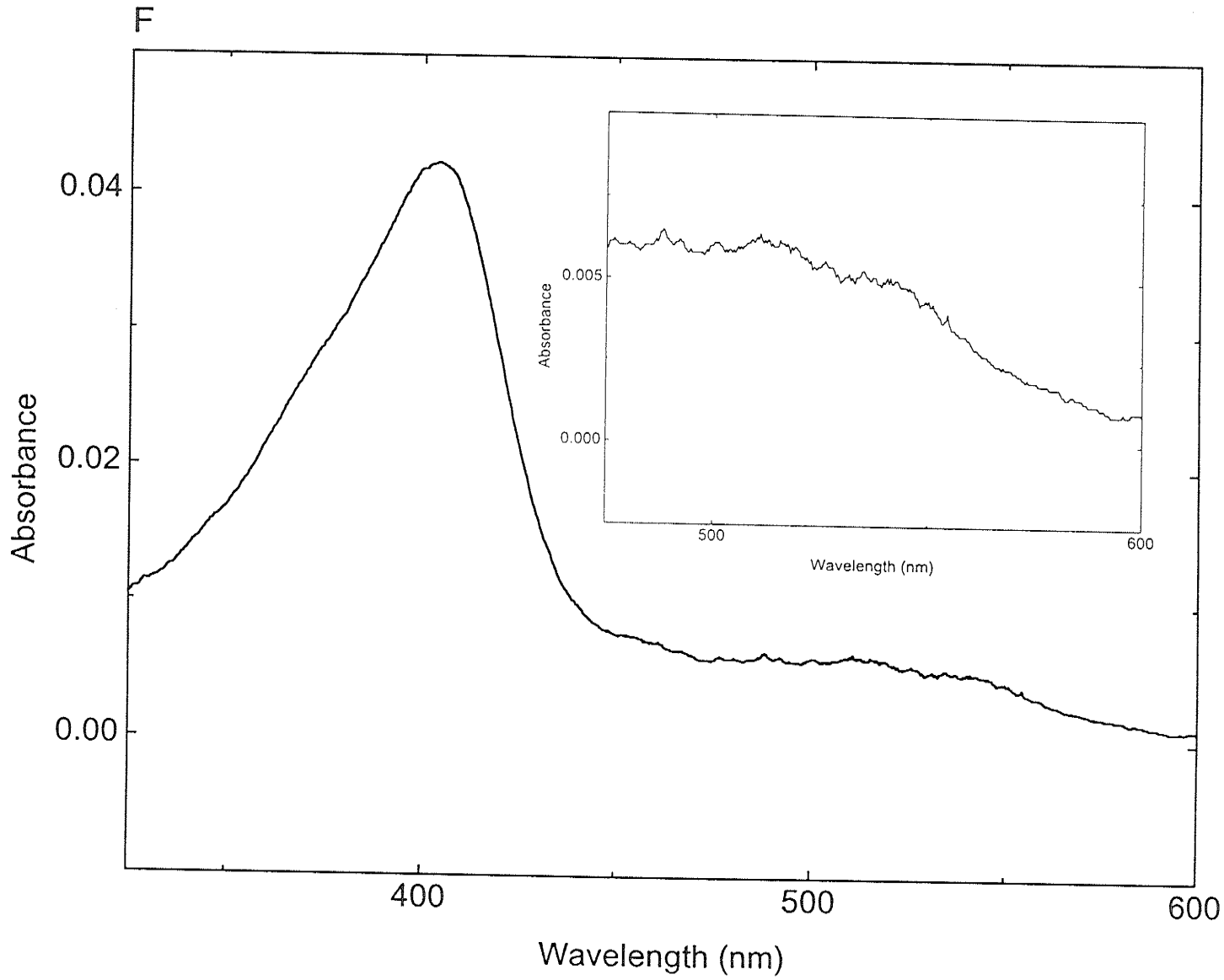

A
Absorbance
0.10
0.05
0.00
400
500
600
700
Wavelength (nm)
0.02
0.01
0.00
Absorbance
500
600
700
Wavelength (nm)

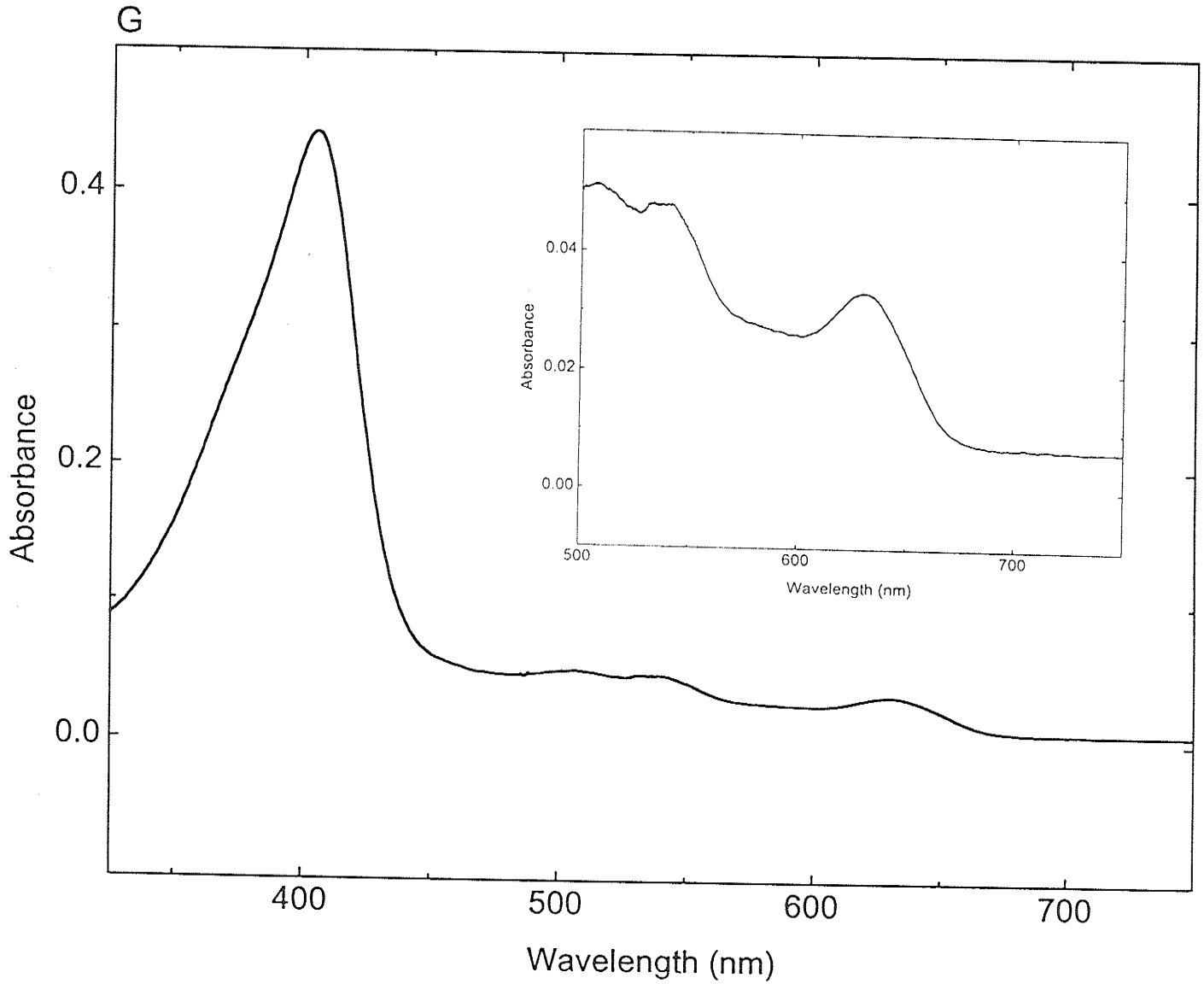

B
Absorbance
0.05
0.00
400
500
600
700
Wavelength (nm)
0.01
0.00
Absorbance
500
600
700
Wavelength (nm)


C
Absorbance
0.025
0.000
400
500
600
700
Wavelength (nm)
0.01
0.00
Absorbance
500
600
700
Wavelength (nm)


D
Absorbance
0.10
0.05
0.00
400
500
600
700
Wavelength (nm)
Absorbance
0.025
0.000
500
600
700
Wavelength (nm)


E
0.05
0.00
Absorbance
400
500
600
700
Wavelength (nm)
0.015
0.000
Absorbance
500
600
700
Wavelength (nm)


F
0.04
0.02
0.00
Absorbance
400
500
600
Wavelength (nm)
0.005
0.000
Absorbance
500
600
Wavelength (nm)


G
0.4
0.2
0.0
Absorbance
400
500
600
700
Wavelength (nm)
0.04
0.02
0.00
Absorbance
500
600
700
Wavelength (nm)

hemes. Based on the characteristic peaks for heme d at 590 and 715 nm and for protoheme IX at 540 and 630 nm, these mixtures range from predominantly heme d as found in N201Q to predominantly protoheme IX as found in N201H. These spectra were also consistent with the characteristic color differences described previously. The existence of protoheme IX-containing mutant HPII catalases suggests that the normal protoheme-to-heme d conversion has been affected indicating that the certain substitutions of the His128 and Asn201 residues prevent the *cis*-hydroxylation of protoheme on ring III to produce the heme d-isomer.

Effect of pH and Temperature on Asn201 Mutant Catalases

HPII catalase is a very stable enzyme exhibiting near maximal activity over a broad pH range and during incubation at 70°C. Activity of the wild type enzyme was maintained from pH 4 to pH 11 with maximal activity at pH 11. A comparison of the pH profile for the wild type enzyme with those determined for the Asn201 mutant enzymes (Fig. 15) revealed that all mutant enzymes retained activity over a similar broad pH range, but with less activity at the high pH extreme.

Figure 16 shows the effect of incubation at 65°C on the activity of wild type and Asn201 mutant catalases. The wild type enzyme retained full activity after one hour whereas the four mutant enzymes exhibited some thermal instability.

Figure 15. Activity of wild type and Asn201 mutant HPII catalases as a function of pH. Each of the enzymes was incubated in 50 mM potassium phosphate buffer at the various pH values and 37°C for one minute before starting the assay by adding hydrogen peroxide. Data are expressed as a percentage of the maximum activity (± 10%).

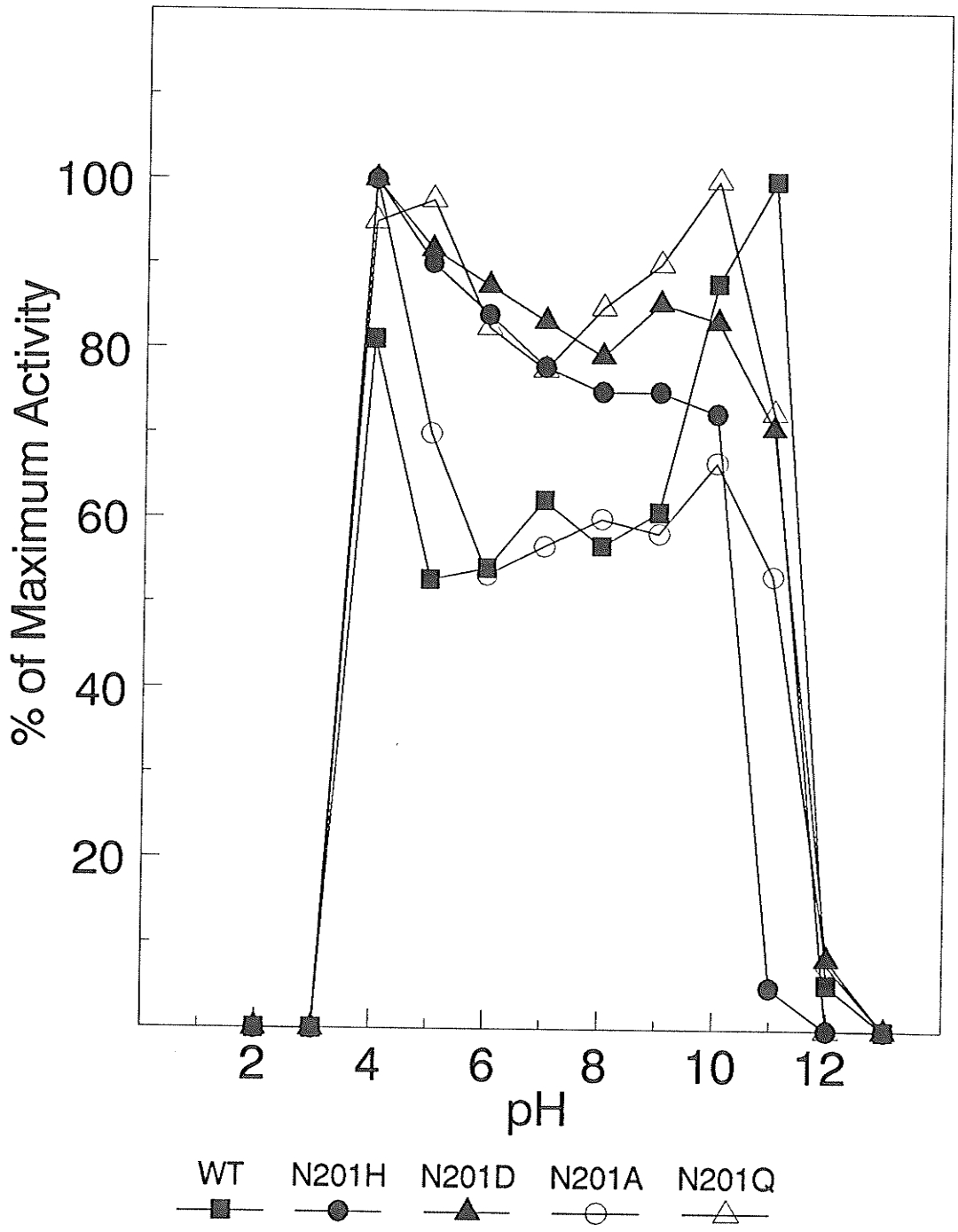
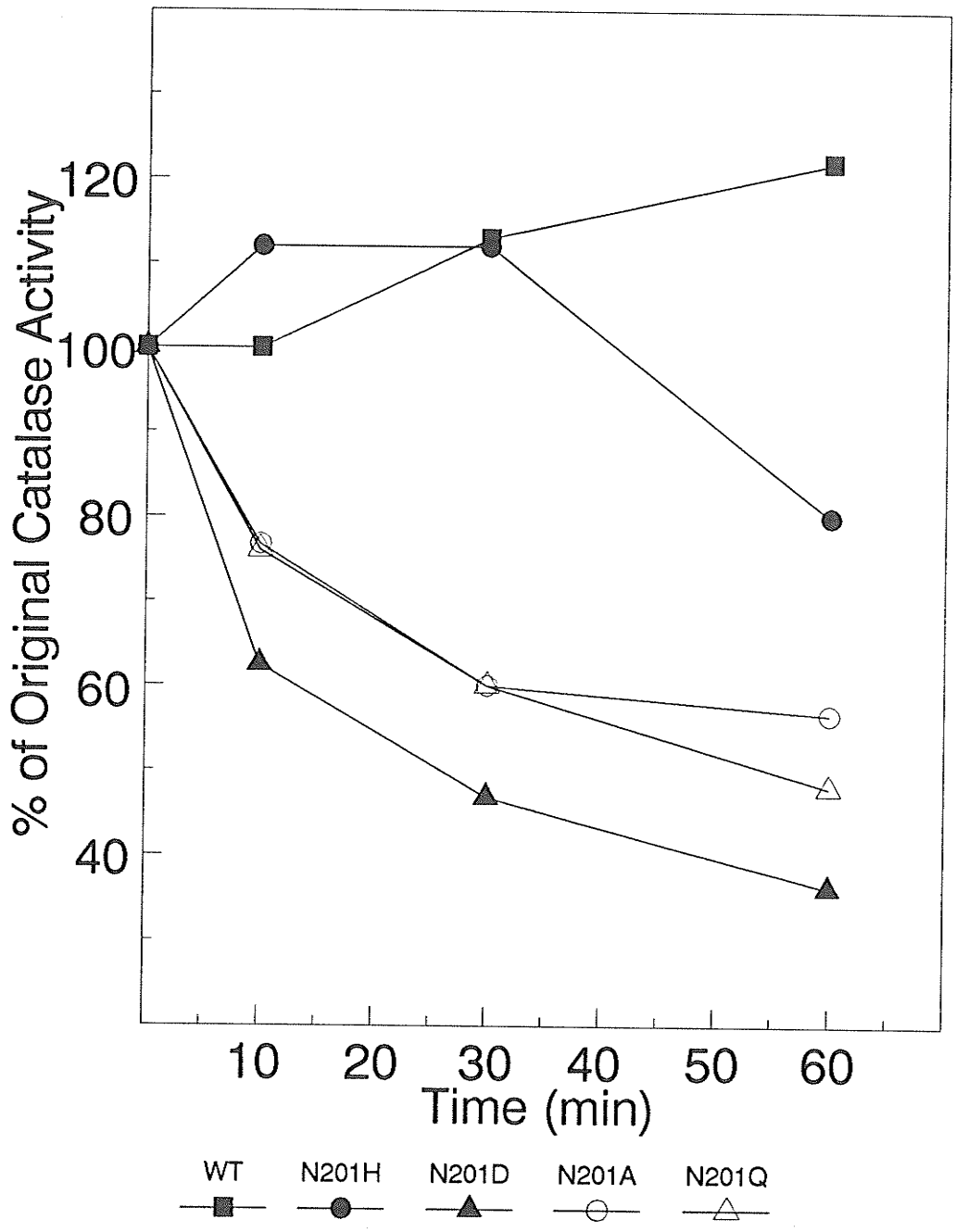

100
80
60
40
20
% of Maximum Activity
2
4
6
8
10
12
pH
WT
N201H
N201D
N201A
N201Q

Figure 16. Determination of the activity of wild type and Asn201 mutant HPII catalases during incubation at 65°C. Each of the enzymes was incubated at 65°C in 50 mM sodium phosphate buffer (pH 7), and aliquots were removed at various times and assayed. Data are expressed as a percentage of original activity (± 10%).

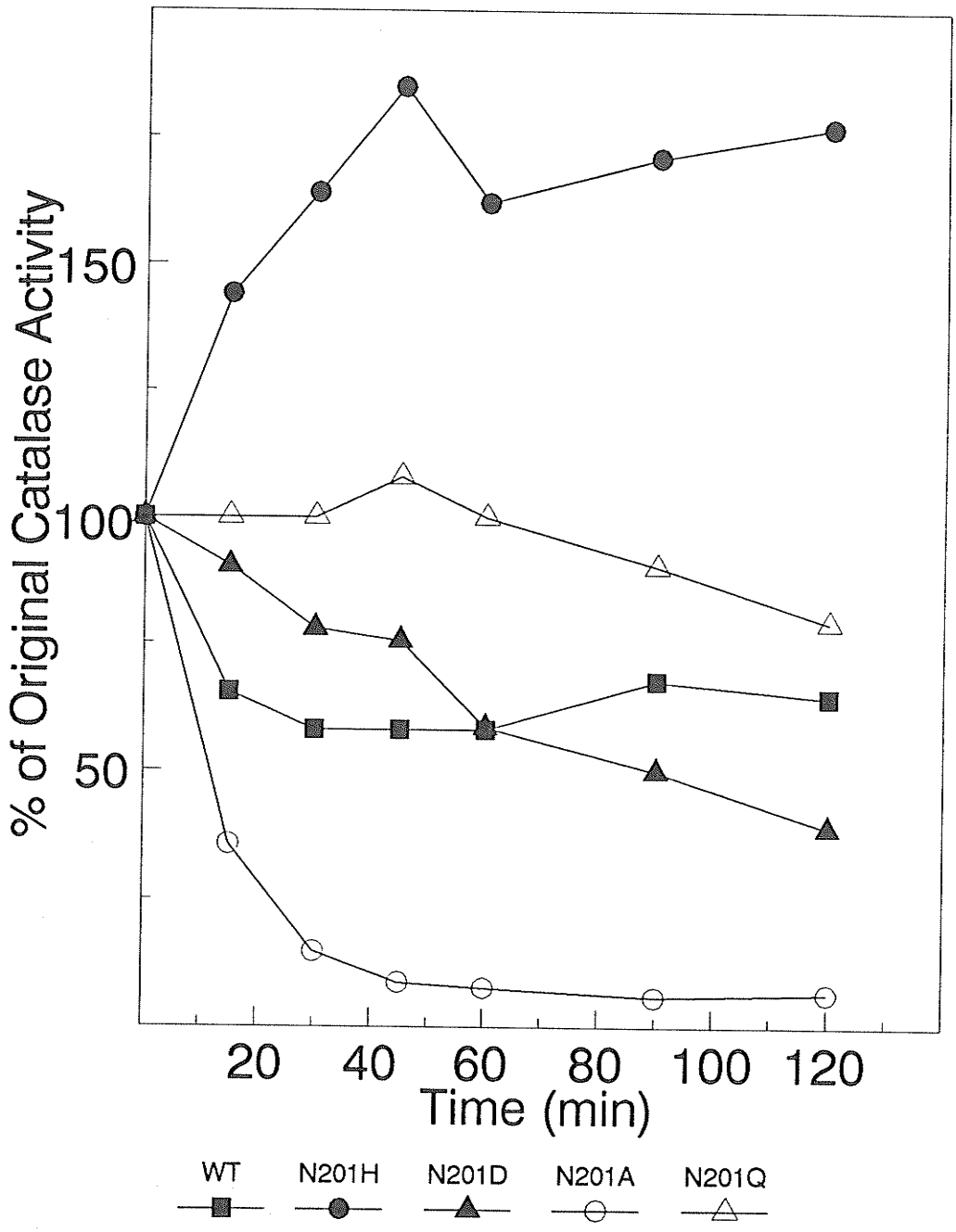

% of Original Catalase Activity
120
100
80
60
40
10
20
30
40
50
60
Time (min)
WT
N201H
N201D
N201A
N201Q

N201D, N201A, and N201Q retained portions of original activity whereas N201H retained almost full activity under the same conditions and can be considered the most heat stable of the mutant enzymes.

Effect of 2-Mercaptoethanol on Asn201 Mutant Catalases

Several thiol compounds have been reported to inhibit catalase activity (Takeda et al., 1980) possibly through modification by the superoxide anion radical. To further characterize the mutant HPII catalases, their activities were determined during incubation with 5mM 2-mercaptoethanol, as shown in Figure 17. The wild type enzyme exhibited a reduction in activity after 15 minutes of incubation but was not inhibited further, suggesting that 2-mercaptoethanol has only a mild inhibitory effect on activity. The mutant enzymes displayed a myriad of responses from little effect on the N201Q enzyme to a slow but steady reduction in activity of the N201D enzyme, and an immediate and extensive inhibition of the N201A enzyme within 30 minutes reaching near complete inhibition after two hours incubation. Surprisingly, the N201H enzyme exhibited an increase in activity of nearly two-folds of the original activity. This enhancement of activity may be a result of partial protoheme-to-heme d conversion utilizing hydrogen peroxide generated from 2-mercaptoethanol oxidation. The range of effects displayed by the mutant

Figure 17. Effect of 2-mercaptoethanol on the activity of wild type and Asn201 mutant HPII catalases. Each of the enzymes was incubated at 37°C in 50 mM sodium phosphate buffer (pH 7) in the presence of 5 mM 2-mercaptoethanol, and the aliquots were removed at various times and assayed. Data are expressed as a percentage of original activity (± 10%).


% of Original Catalase Activity
150
100
50
20
40
60
80
100
120
Time (min)
WT
N201H
N201D
N201A
N201Q

enzymes may reflect the residue at position 201 sterically controlling access of the thiol agent to the active site.

Effect of Cyanide and Azide on Asn201 Mutant Catalases

Both cyanide and azide have been reported to reversibly inhibit catalase activity by binding to the heme iron (Beyer and Fridovich, 1988) but their effect on HPII catalase has not been determined. Both reagents were examined for their inhibitory effect on the mutant and wild type HPII catalases. As shown in Figure 18, the wild type enzyme is very sensitive to sodium cyanide, with close to complete inhibition of catalase activity at 0.1 mM and no activity being evident at 2.5 mM. All four mutant enzymes are less sensitive to cyanide with N201D being the most sensitive and the remaining three mutant enzymes (N201H, N201A, and N201Q) being mildly inhibited. The inhibition pattern for sodium azide differs only marginally from that obtained with cyanide. Figure 19 shows that the wild type and N201D enzymes are strongly inhibited by azide whereas the N201Q enzyme is inhibited similarly at low azide but exhibits less sensitivity at high azide. The N201H and N201A enzymes are less sensitive to 0.5 mM azide but similarily sensitive to high azide. Each of the amino acid substitutions appears to prevent cyanide and azide from exerting their full inhibitory effect relative to the wild type enzyme.

Figure 18. Activity of wild type and Asn201 mutant HPII catalases in the presence of sodium cyanide. Each of the enzymes was incubated with the inhibitor in 50 mM sodium phosphate buffer (pH 7) at 37°C for one minute before starting the assay with the addition of hydrogen peroxide. Data are expressed as a percentage of original activity (± 10%).

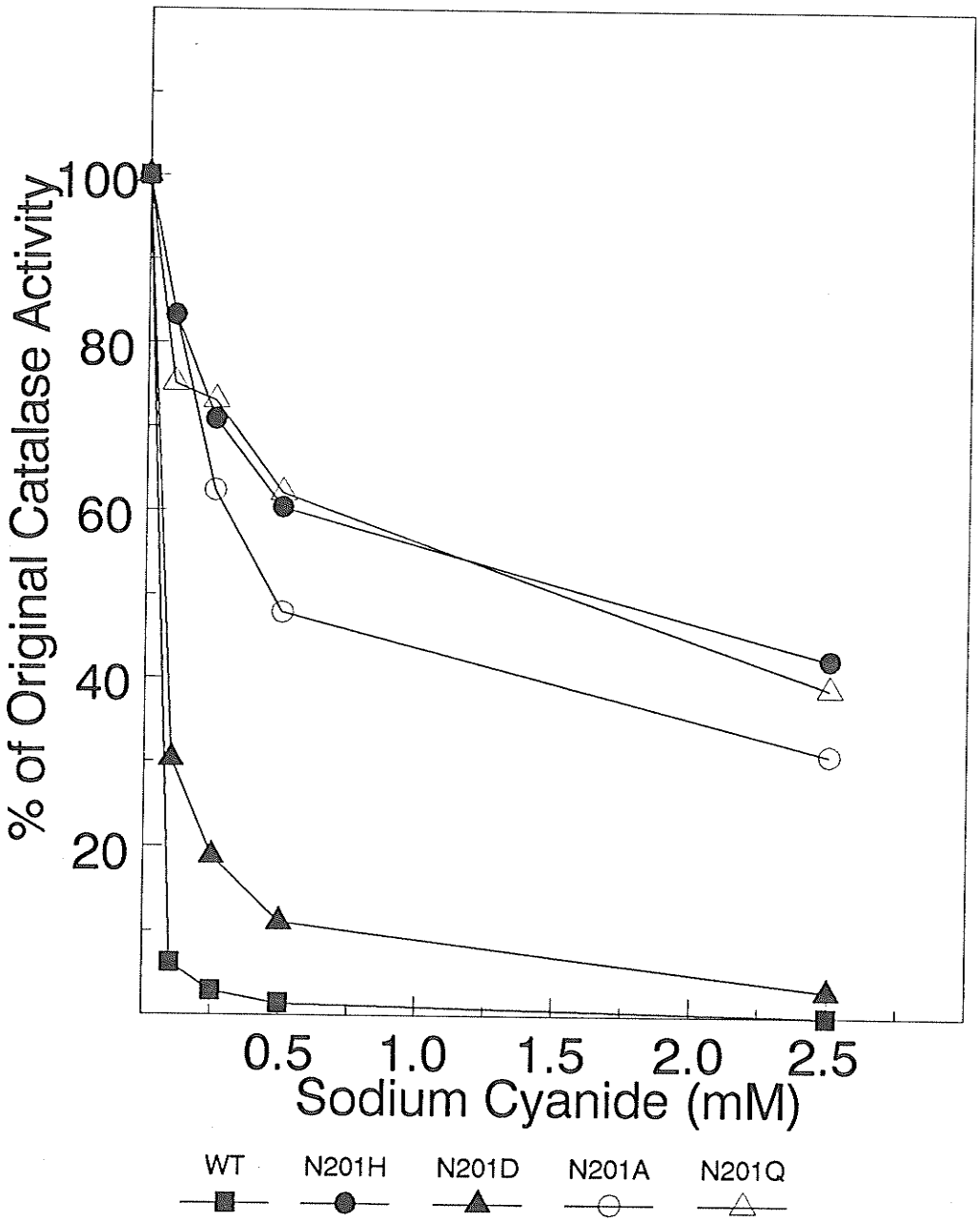
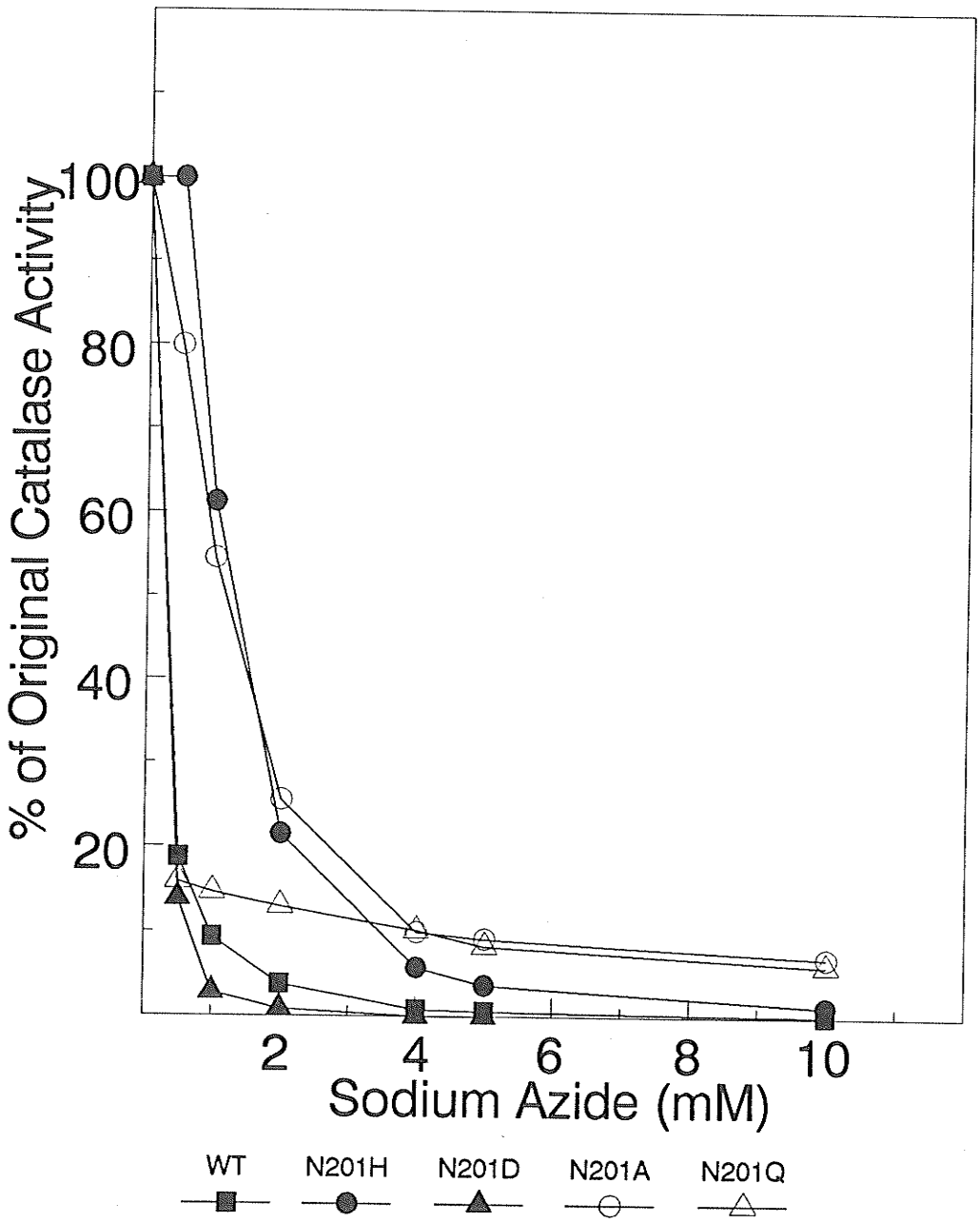

% of Original Catalase Activity
100
80
60
40
20
0.5
1.0
1.5
2.0
2.5
Sodium Cyanide (mM)
WT
N201H
N201D
N201A
N201Q

Figure 19. Activity of wild type and Asn201 mutant HPII catalases in the presence of sodium azide. Each of the enzymes was incubated with the inhibitor in 50 mM sodium phosphate buffer (pH 7) at 37°C for one minute before starting the assay with the addition of hydrogen peroxide. Data are expressed as a percentage of original activity (± 10%).


% of Original Catalase Activity
100
80
60
40
20
2
4
6
8
10
Sodium Azide (mM)
WT
N201H
N201D
N201A
N201Q

Effect of 3-Amino-1,2,4-Triazole on Asn201 Mutant Catalases

The compound 3-amino-1,2,4-triazole (AT) has been reported to be a specific inhibitor of catalase activity (see Introduction), although its effect on HPII has not been reported. In 8 mM AT, the N210A, N201Q, and N201D mutant enzymes are all inhibited to a greater extent than the wild type enzyme, whereas the N201H mutant enzyme is relatively unaffected (Fig. 20). The N201H substitution may sterically hinder access to the active site histidyl group (His128). Margoliash and Novogrodsky (1958) found that inhibition of human erythrocyte catalase activity by AT was enhanced by low concentrations of hydrogen peroxide. When ascorbate (a source of hydrogen peroxide) was included with AT, the only significant change was observed in the N201H mutant enzyme for which activity was enhanced. Therefore, unlike eukaryotic catalases low concentrations of hydrogen peroxide produced *in situ* from ascorbate did not significantly affect the inhibition of HPII catalases by AT (Fig. 21).

The apparent increase in activity for the "protoheme IX-containing" N201H mutant enzyme from the low concentrations of hydrogen peroxide generated by ascorbate is a rather interesting observation and may possibly result from a change in the heme species. For example, if the *cis* heme d-isomer is the preferred heme in HPII, the conversion of the protoheme IX in N201H to heme d may cause the rise in enzyme activity. This implies that heme d is generated

Figure 20. Effect of 3-amino-1,2,4-triazole on the activities of wild type and Asn201 mutant HPII catalases. Each of the enzymes was incubated with 8 mM 3-amino-1,2,4-triazole in 50 mM sodium phosphate buffer (pH 7) at 37°C, and aliquots were removed at various times and assayed. Data are expressed as a percentage of original activity (± 10%).

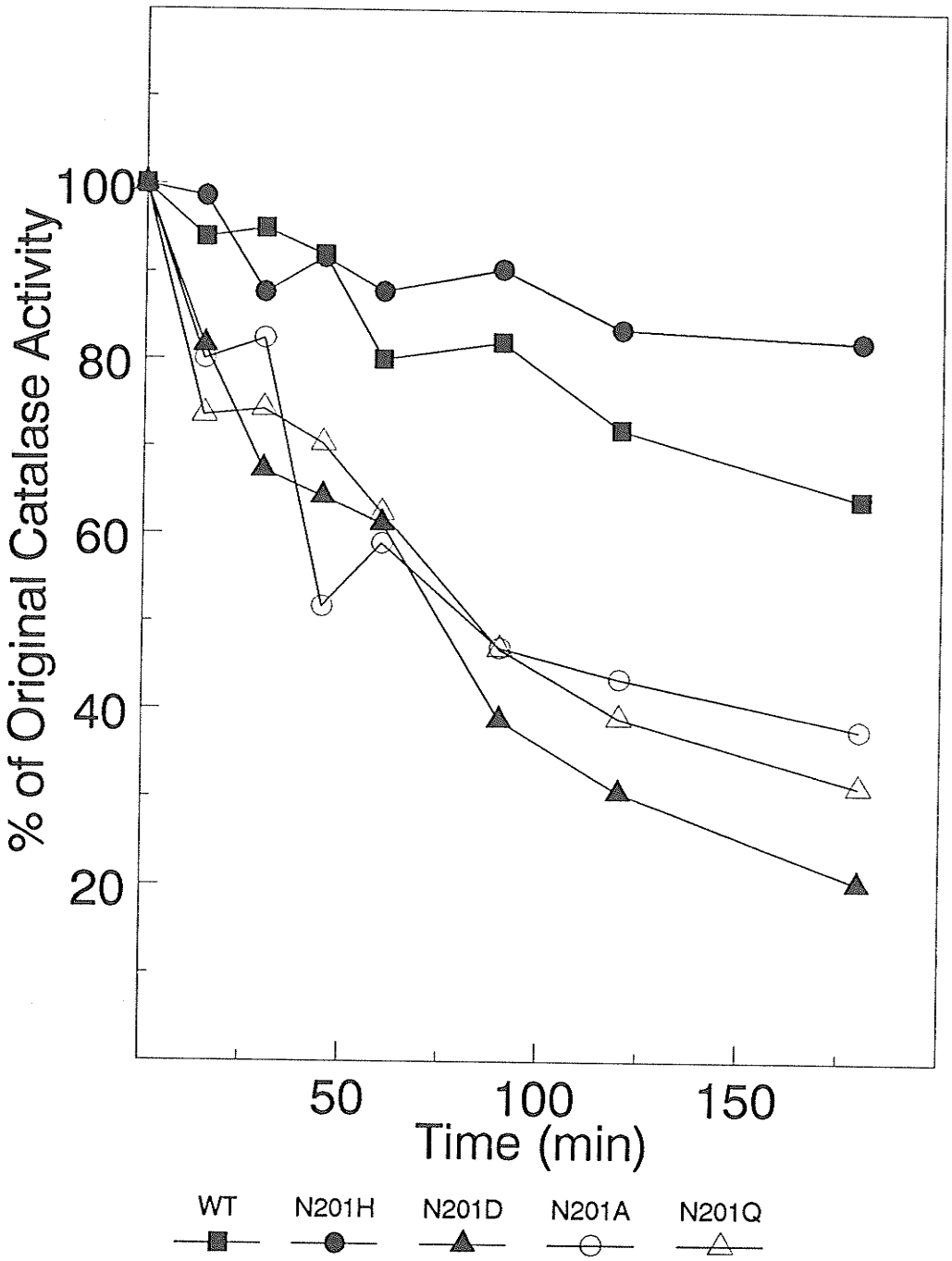
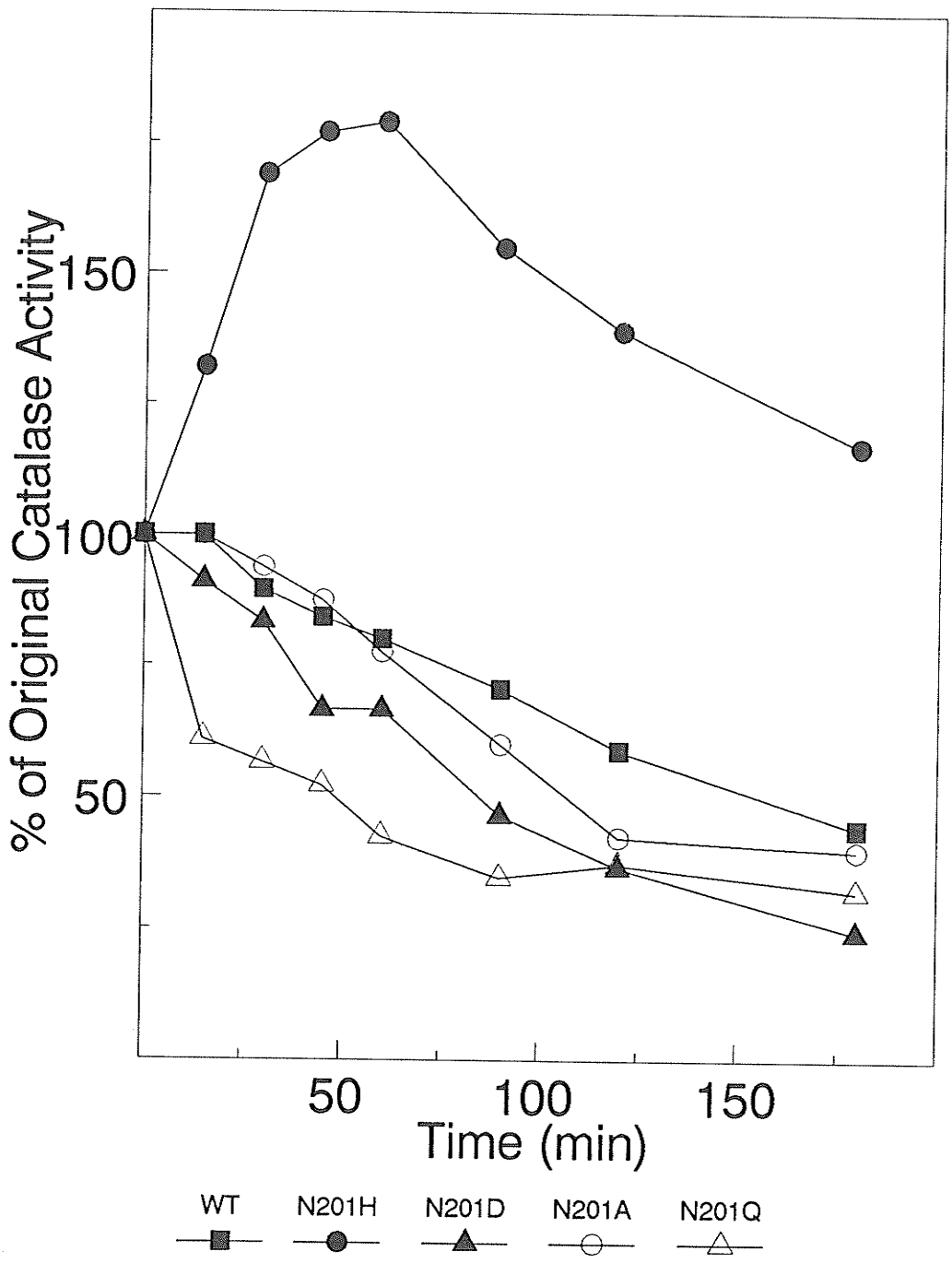

% of Original Catalase Activity
100
80
60
40
20
50
100
150
Time (min)
WT
N201H
N201D
N201A
N201Q

Figure 21. Effect of 3-amino-1,2,4-triazole and ascorbate on the activities of wild type and Asn201 mutant HPII catalases. Each of the enzymes was incubated with 8 mM 3-amino-1,2,4-triazole and 1 mM ascorbate in 50 mM sodium phosphate buffer (pH 7) at 37°C, and aliquots were removed at various times and assayed. Data are expressed as a percentage of original activity (± 10%).


% of Original Catalase Activity
150
100
50
50
100
150
Time (min)
WT
N201H
N201D
N201A
N201Q

directly from protoheme IX in a reaction catalyzed by HPII. Consequently, the N201H mutant may prove to be a valuable tool in elucidating the mechanism for heme d biosynthesis.

DISCUSSION

The nucleotide sequence for the *katE* gene, which encodes HPII catalase in *E. coli*, has been determined and found to display features characteristic of a typical prokaryotic gene including transcriptional start and termination sites, a ribosome binding site, a putative -10 and -35 promoter region, and an open reading frame encoding the protein. A preliminary sequence comparison revealed significant similarity between the sequence of HPII and the sequences of other catalases from various sources, which spawned a more thorough and comprehensive comparison for use in an evolutionary study. In addition, several individual amino acids considered to play a role in the active site of bovine liver catalase and the binding of the heme group were conserved in the HPII sequence, facilitating a site-directed mutagenesis study using HPII as a model to test the predicted roles for some of these specific residues.

The discovery of sequence similarity between HPII and other catalases was a surprising and unexpected result in view of the many apparent differences in physical properties (see Introduction). For example, the active quaternary structure of HPII is hexameric, whereas most other catalases have been found to be tetrameric; the HPII subunits with a predicted size of 84,173 Da are substantially larger than the average subunit size of 60,000 Da for the "typical" catalases; and HPII catalase contains a heme d-isomer rather

than protoheme IX associated with the "typical" catalases. This latter difference in heme content results in the color difference, HPII being green in color while other catalases are reddish-brown. It was the core of HPII that was similar to other catalases and the extra sequence in the larger HPII subunit was located at the N- and C- termini, possibly allowing extra subunit-subunit interactions necessary for formation of the hexameric structure. However, *B. subtilis* and *L. sake* catalases have also been characterized as hexameric, but with smaller subunits (483 residues for *B. subtilis* and 478 residues for *L. sake*) which would therefore lack the extra N- and C- terminal sequence of HPII. This inconsistency may be explained once the crystal structure of HPII has been completely determined.

The binding of NADPH by HPII catalase has yet to be demonstrated. However an examination of HPII for the individual amino acids implicated in NADPH binding by bovine liver catalase suggests that NADPH is not bound by HPII. For example, Fita and Rossmann (1985b) proposed that the binding of NADPH by the bovine catalase involves a water molecule in association with the Lys236, His234 and Tyr214 residues, and the interaction of the pyrophosphate group of NADPH with the His304 residue. Although Lys236 and His234 are conserved as Lys294 and His292 in HPII, the replacement of Tyr214 with Phe272 in HPII eliminates the phenolic hydroxyl group necessary for an interaction with the water

molecule, and therefore would undoubtly affect the predicted binding of water. Moreover, the counterpart for the His304 residue in HPII is a negatively charged Glu362 residue which would not interact favorably with the negatively charged pyrophosphate group of NADPH.

As part of the characterization of the *katE* gene, a putative promoter region has been proposed. Based on sequence similarity in the regions directly upstream of the transcriptional start sites in *katE*, *xthA*, and *bolA*, potential -10 and -35 sequences upstream of *katE* were identified, consistent with all three genes being controlled by KatF. However, a recent review by Hengge-Aronis (1993) noted that other genes (*cyxA*, *osmB*, and *glgS*) controlled by KatF protein do not share this similar sequence, implicating the involvement of a secondary regulatory factor for transcriptional initiation. An alternative consensus sequence called the "gearbox" promoter proposed to be involved in stationary-phase induced gene regulation (Aldea et al., 1990) has also been shown to not be a factor in KatF regulation of *bolA* (Lange and Hengge-Aronis, 1991a), further supporting the concept of an additional regulatory factor. In my study, deletions of the *katE* sequence made upstream of the putative -10 and -35 sequences did not effect *katE* expression, ruling out the possibility that this upstream region was the site of action for an unidentified regulatory factor. However, a regulatory factor could still be active

at a site downstream of the putative promoter sequence.

The gathering and analysis of common and unique characteristics has been traditionally used in the determination of an evolutionary history for a group of organisms. Life forms demonstrating greater similarity are generally considered to have descended more recently from a common ancestor than those with less similarity. Generally, phenotypical data based on morphological, chemical, metabolic, or behavioral traits, have been used for phylogenetic reconstruction. With the relatively recent developments in protein and DNA sequencing, genotypical data can now be used for further phylogenetic analysis of organisms. Establishing an evolutionary history for a set of organisms based on such sequences is attractive because a heightened resolution of the inferred evolutionary relationships usually results. However, drawbacks such as assuming that a sequence is unique in a population, and that the sequence reliably represents the complete genome must be considered (Cedergren et al., 1988).

For this study, a phylogenetic reconstruction (or evolutionary history) was carried out based on catalase amino acid sequences. The phylogenetic analysis generated unrooted networks that support with confidence levels greater than 95% the monophyly of the following groups of organisms: animals (mammals and insects), plants (dicots and monocots), and fungi (ascomycetous yeasts). Although

this preliminary phylogenetic survey supports the monophyletic groups previously identified from the analysis of well established "molecular chronometers" such as ribosomal genes (Cedergren et al., 1988; Sogin et al., 1986), several anomalies were observed within the groupings. Plant or animal catalases were grouped together with 100% confidence by the distance and parsimony methods, but with different internal topologies consisting of different branches without bootstrap support (see Results, Section II).

Within the grouping for plants, the catalase sequences could not distinguish between the monocot and dicot plants with the rice and maize (monocots) sequences branching within dicot sequences. The divergence of monocot and dicot plants is considered to have occurred about 200 million years ago (Wolfe et al., 1989), and it is therefore surprising that catalase sequences have not diverged enough to be recognized as two routes of evolutionary progression. This peculiarity may be explained in two possible ways. Plant catalases may have evolved at such a slow rate that insufficient "informative changes" have arisen to produce statistically acceptable phyletic estimates. Alternatively, the peculiar branching order of the monocot and dicot sequences may result from homoplasy, wherein rather than the sequence similarity stemming from a common ancestor, the apparent resemblance of the catalase sequences of the

monocot and dicot plants results from other events such as convergence.

Although the branching order for the plant catalases lacks consistency, it is noteworthy that the *Arabidopsis thaliana* catalase was grouped with the plant catalases, unlike cytochrome c and histone H3 sequences which had placed *A. thaliana* within a fungal grouping (Kemmerer et al., 1991). Therefore, it appears that catalase sequences provide a more conservative phyletic estimate for the taxonomic position of *A. thaliana* and also support the findings of Pasternak and Glick (1992) where *A. thaliana* was grouped within the mustard plant family, Brassicaceae, based on an analysis of ribulose-1,5-bisphosphate carboxylase/oxygenase sequences.

For the four fungal catalases (three peroxisomal and one cytosolic), the phylogenetic reconstruction substantiated their differences based on subcellular location. Peroxisomal proteins are believed to constitute a homogeneous group in terms of evolution (Igual et al., 1992), although whether this is a result of homoplasy or an endosymbiotic origin of peroxisomes remains unclear. The three peroxisomal catalases from *Hansenula polymorpha*, *Candida tropicalis*, and *Saccharomyces cerevisiae* Type A formed a monophyly of high confidence which is consistent with the concept that peroxisomal proteins form an evolutionarily homogeneous group. However, the observation

that the cytosolic catalase (*S. cerevisiae* Type T) consistently groups with the other fungal catalases (although with lower confidence) suggests that duplication of a host gene is more likely responsible for the formation of the Type T catalase gene than a separate origin such as homoplasy or endosymbiosis. The cytosolic catalase may have been exposed to evolutionary pressures different from the catalases found within the peroxisomal environment such that with the force of natural selection operating, the cytosolic enzyme has diverged at a more rapid rate and consequently undergone its own history of independent mutation. Such a recombination event with the peroxisomal form would result in a greater degree of similarity between the cytosolic and peroxisomal enzymes than would be expected had the peroxisomal form originated through endosymbiosis.

Based on the phylogenetic reconstruction, catalase sequences from animal, fungal, and the two bacterial (*B. subtilis* and *L. sake*) sources can be derived from one internal node (or common ancestor) to form a common lineage being supported by both methods of tree construction. This potential class of catalases is located beside the *E. coli* and *L. seeligeri* sequences which do not group with other catalases and are connected by a branch with low bootstrap support. Moreover, a branch uniting all plant catalases for both methods is found next to the *L. seeligeri* sequences. Such a branching pattern suggests that at least two distinct

classes of catalases, the animal/fungal and the plant catalases, arose from different prokaryotic (or endosymbiont) ancestors. As catalase sequences from prokaryotes, as well as cyanobacteria and green algae become available, a better understanding of the evolution and potential origin of plant catalases can be made.

To conclude, catalase amino acid sequences may be helpful in further reconstructing the evolutionary history within the animals and fungi. Contrastingly, the bacterial catalases are of diverse origin and consequently were not used for estimating phylogenies. Plant catalases appear to have evolved slowly and formed a distinct class that arose independently of other eukaryotes from a possible prokaryotic ancestor.

Site-directed mutagenesis was used to replace the active site residues His128 with Ala, Asn, Glu, and Gln, Ser167 with Ala, Thr, Cys, and Asn, and Asn201 with Ala, His, Asp, Gln, and Arg, and the proximal side heme binding residue Tyr415 with Phe and His. Of these fifteen amino acid substitutions in HPII, only the mutant enzymes H128A, H128N, N201A, N201H, N201D, and N201Q had accumulated in amounts similar to the wild type enzyme. As well, only the Asn201 mutant proteins exhibited catalase activity, although at significantly reduced levels relative to the wild type enzyme. As a result, the N201A, N201H, N201D, and N201Q mutant enzymes were characterized based on activity changes.

Fita and Rossmann (1985a) have proposed that the catalytic role for the Asn147 residue in bovine liver catalase involves the binding of hydrogen peroxide and the weakening of the O-H bonds through hydrogen bonding to promote proton transfer between oxygen atoms. Based on the results obtained in this study, this role is probably the same for the counterpart residue (Asn201) in HPII. For example, the Km values for the mutant enzymes were largely the same as the wild type enzyme, except for N201Q which is lower than the wild type and suggests the affinity for the substrate has increased. The N201Q mutant enzyme contains the same functional amide group as Asn, but with one extra methylene group in the side chain. However, the turnover numbers (k_{cat}) for all mutant enzymes are lower than the wild type value which suggests that hydrogen peroxide was positioned unfavorably for the formation of other interactions necessary for normal catalysis. Consequently, the role of Asn201 in HPII may include the binding of hydrogen peroxide to form a more stable enzyme-substrate complex and participation in the reaction perhaps by distorting the hydrogen bond of hydrogen peroxide. However, it is not absolutely required for catalytic activity.

The broad pH range of maximal activity characteristic of the wild type enzyme was unchanged in the mutant enzymes and is indicative of the relatively pH independent nature of

the catalytic mechanism. The sensitivity of the mutant enzymes to 2-mercaptoethanol and 3-amino-1,2,4-triazole varied with the amino acid substitution consistent with the residue at position 201 sterically affecting the site of inhibitor interaction. For example, the extent of inhibition by 2-mercaptoethanol decreased as amino acids at position 201 increased in size (Ala > Asp > Asn > Gln > His). Likewise, the changes in aminotriazole binding by the N201A and N201H mutant enzymes can also be explained in terms of steric interference.

Of the His128 mutants only the H128A and H128N mutant enzymes were sufficiently stable to accumulate in the cell allowing subsequent purification. Both mutant proteins were catalytically inactive consistent with the essential role for His74 (equivalent to His128) in catalysis proposed by Fita and Rossmann (1985a). His74 is also considered to be involved in the binding of hydrogen peroxide, but, more importantly, provides the imidazole group necessary to facilitate the transfer of a proton between oxygens of the peroxide molecule (see Introduction). Based on the absence of catalase activity in both mutants, it is concluded that the His128 residue in HPII is essential to the catalytic mechanism.

None of the remaining HPII mutant proteins, including H128E, H128Q, S167A, S167T, S167C, S167N, N201R, Y415F, and Y415H, accumulated in amounts sufficient for recovery

indicating that each replacement had in some way destabilized the protein structure. In bovine liver catalase, Ser113 (the counterpart to Ser167) forms hydrogen bonds with the His74 through an interaction with a water molecule, which itself also interacts with the propionate of pyrrole III of the heme group. Moreover, Fita and Rossmann (1985a) suggested that this pattern of hydrogen bonding forms a chain of interactions that alters the alkalinity of the histidyl imidazole group to assist with proton transfer to and from this group. Consequently, the replacement of Ser167 in HPII with residues unable to form the normal hydrogen bonds without steric hindrance will likely affect this chain of interactions and perhaps disrupt the normal stability and folding of the protein. Likewise, Tyr357 in the bovine enzyme interacts with the heme iron on the proximal side and forms hydrogen bonds with Arg353 (Arg411 in HPII) through its deprotonated phenolic oxygen (Fita and Rossmann, 1985a). The replacement of the equivalent Tyr415 in HPII with Phe and His, both lacking this phenolic hydroxyl group, will no longer permit the normal interactions, probably affecting heme binding, and the stability and folding of the protein.

In conclusion, the predicted roles of the His128 and Asn201 have been tested and indirectly substantiated by physical evidence providing a base from which further testing of other functional residues may be undertaken.

FUTURE RESEARCH PERSPECTIVES

With the determination of the *katE* nucleotide sequence, it was possible to propose promoter sequences that may be recognized by the KatF protein and that are quite unlike those recognized by the σ^{70} factor. The nature of the regulatory mechanism governing *katE* expression is poorly understood. For example, whether *katE* expression is controlled only by KatF or by KatF in conjunction with other factors has yet to be determined. In my study, deletions of the regulatory region have localized the *katE* promoter within a specific segment of sequence with upstream deletions suggesting that a secondary factor cannot be acting upstream of the promoter sequences. Consequently, the next logical step in characterizing the *katE* regulatory region is to determine whether an additional factor recognizes regulatory sequences between the promoter and the open reading frame. This may be carried out by deleting segments of sequence downstream of the promoter in a manner similar to what I have presented in this thesis. Such an experiment would help establish whether KatF is the sole regulatory factor in *katE* expression. Experiments may also be performed to confirm the validity of the putative promoter sequences. This may be determined by substituting the various bases encompassing the -10 and -35 sequences and observing the effect on gene expression. Moreover, as nucleotide sequences of other KatF-controlled genes become

available, it will be possible to determine whether similar -10 and -35 sequences proposed for *katE* also exist in other genes controlled by KatF and, therefore, form a "true" consensus sequence for KatF recognition.

The phylogenetic study using the catalase enzyme as presented in this thesis has helped determine the evolutionary position of *E. coli* HPII catalase with respect to other catalases and supported previously identified monophyletic groups. This has suggested the usefulness of catalase for inferring evolutionary relationships and provided a base from which further phylogenetic reconstruction may be undertaken once catalase sequences from other sources become available. Phyletic estimates based on a larger compilation of sequences will test the strength of the inferred relationships generated from this study. Since the phylogenetic reconstruction presented in this study is completely based on sequences from Eubacteria and Eukaryotes, it would be rather interesting to include catalase sequences from the Archaebacteria in future reconstructions. The evolutionary position of catalases from such sources may establish an ancestral origin for *E. coli* HPII catalase and clarify its unique physical properties.

Site-directed mutagenesis has proven to be a useful tool in confirming the functional roles for two active site residues (His128 and Asn201) in HPII catalase. (This also

suggests that HPII is a good model for testing the predicted roles for residues proposed in bovine liver catalase.) The testing of two additional residues (Ser167 and Tyr415) could not be undertaken since it was not possible to produce mutant protein in amounts sufficient for recovery. An ongoing investigation has revealed that cell growth at lower incubation temperatures increases the level of recoverable mutant protein, thereby making it possible to obtain and characterize substitutions of the Ser167 and Tyr415 residues. Furthermore, the production of other mutant proteins by site-directed mutagenesis will be more attainable. For example, several other residues have been implicated in heme binding and the testing of their roles would be useful. Although this study has suggested that NADPH is not likely associated with HPII, it would still be of interest to confirm this conclusion by directly assaying the HPII enzyme for the presence of NADPH, as well as use site-directed mutagenesis to substitute the residues implicated in NADPH binding.

The work presented in this thesis has provided several avenues of additional research that may be undertaken at both the genetic and biochemical level.

REFERENCES

Aebi, H. (1974) Catalase. In *Methods of Enzymatic Analysis* (Bergmeyer, H.U., ed) vol. 2, pp. 673-684, Academic Press, New York

Aldea, M., Garrido, T., Hernández-Chico, C., Vicente, M., and Kushner, S.R. (1989) Induction of a growth-phase-dependent promoter triggers transcription of *bolA*, an *Escherichia coli* morphogene. *EMBO J.* **8,** 3923-3931

Aldea, M., Garrido, T., Pla, J., and Vicente, M. (1990) Division genes in *Escherichia coli* are expressed coordinately to cell septum requirements by gearbox promoter. *EMBO J.* **9,** 3787-3794

Allgood, G.S., and Perry, J.J. (1986) Characterization of a manganese-containing catalase from the obligate thermophile *Thermoleophilium album*. *J. Bacteriol.* **168,** 563-567

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (eds) (1989) *Current Protocols in Molecular Biology,* vol. 1, Wiley Interscience, New York

Bell, G.I., Najarian, R.C., Mullenbach, G.T., and Hallewell, R.A. (1986) cDNA sequence coding for human kidney catalase. *Nucleic Acids Res.* **14,** 5561-5562

Bethards, L.A., Skadsen, R.W., and Scandalios, J.G. (1987) Isolation and characterization of a cDNA clone for the *Cat2* gene in maize and its homology with other catalases. *Proc. Natl. Acad. Sci. USA* **84,** 6830-6834

Beyer, W.F., and Fridovich, I. (1988) Catalases: with and without heme. In *Oxygen Radicals in Biology and Medicine* (Simic, M.G., Taylor, K.A., Ward, J.F., and von Sonntag, C., eds) pp. 651-661, Plenum Press, New York

Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screeening recombinant plasmid DNA. *Nucleic Acids Res.* **7,** 1513-1523

Bol, D.K., and Yasbin, R.E. (1991) The isolation, cloning and identification of a vegetative catalase gene from *Bacillus subtilis*. *Gene* **109,** 31-37

Cedergren, R., Gray, M.W., Abel, Y., and Sankoff, D. (1988) The evolutionary relationships among known life forms. *J. Mol. Evol.* **28,** 98-112

Chiu, J.T., Loewen, P.C., Switala, J., Gennis, R.B., and Timkovich, R. (1989) Proposed structure for the prosthetic group of the catalase HPII from *Escherichia coli*. *J. Am. Chem. Soc.* **111,** 7046-7050

Christman, M.F., Morgan, R.W., Jacobson, F.S., and Ames, B.N. (1985) Positive control of a regulon for defenses against oxidative stress and some heat shock proteins in *Salmonella typhimurium*. *Cell* **41,** 753-762

Christman, M.F., Storz, G., and Ames, B.N. (1989) OxyR, a positive regulator of hydrogen peroxide-inducible genes in *Escherichia coli* and *Salmonella typhimurium*, is homologous to a family of bacterial regulatory protein. *Proc. Natl. Acad. Sci. USA* **86,** 3484-3488

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989) One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* **86,** 2172-2175

Claiborne, A. (1978) Relationship between catalase and dianisidine peroxidase in *Escherichia coli* B. *Fed. Proc.* **37,** 1513

Claiborne, A., and Fridovich, I. (1979) Purification of the *o*-dianisidine peroxidase from *Escherichia coli* B. *J. Biol. Chem.* **254,** 4245-4252

Cohen, G., Rapatz, W., and Ruis, H. (1988) Sequence of the *Saccharomyces cerevisiae* CTA1 gene and amino acid sequence of catalase A derived from it. *Eur. J. Biochem.* **176,** 159-163

Darr, D., and Fridovich, I. (1985) Inhibition of catalase by 3,3'-diamino-benzidine. *Biochem. J.* **226,** 781-787

Dawson, J.H. (1988) Probing structure-function relations in heme-containing oxygenases and peroxidases. *Science* **240,** 433-439

de Duve, C. (1974) in *Alcohol and Aldehyde Metabolizing Systems* (Thurman, R.G., et al., eds) pp. 161, Academic Press, New York

Deisseroth, A., and Dounce, A.L. (1970) Catalase: physical and chemical properties, mechanism of catalysis, and physiological role. *Physiol. Rev.* **50,** 319-375

Demple, B. (1991) Regulation of bacterial oxidative stress genes. *Annu. Rev. Genet.* **25,** 315-337

Didion, T., and Roggenkamp, R. (1992) Targeting signal of the peroxisomal catalase in the methylotrophic yeast *Hansenula polymorpha*. *FEBS Lett.* **303,** 113-116

Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6,** 1948-1954

Eventoff, W., Tanaka, N., and Rossmann, M.G. (1976) Crystalline bovine liver catalase. *J. Mol. Biol.* **103,** 799-801

Farr, S.B., and Kogoma, T. (1991) Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* **55,** 561-585

Faulkner, D.V., and Jurka, J. (1988) Multiple aligned sequence editor (MASE). *Trends Biochem. Sci.* **13,** 321-322

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39,** 783-791

Felsenstein, J. (1989) *PHYLIP 3.4 User manual*. University of Washington, Seattle

Fita, I., and Rossmann, M.G. (1985a) The active center of catalase. *J. Mol. Biol.* **185,** 21-37

Fita, I., and Rossmann, M.G. (1985b) The NADPH binding site on beef liver catalase. *Proc. Natl. Acad. Sci. USA* **82,** 1604-1608

Fridovich, I. (1978) The biology of oxygen radicals. *Science* **201,** 875-880

Fristensky, B. (1991) Biological research computer hierarchy (BIRCH). User manual. University of Manitoba, Winnipeg

Furuta, H., Hachimori, A., Ohta, Y., and Samejima, T. (1974) Dissociation of bovine liver catalase into subunits on acetylation. *J. Biochem.* **76,** 481-491

Furuta, S., Hayashi, H., Hijikata, M., Miyazawa, S., Osumi, T., and Hashimoto, T. (1986) Complete nucleotide sequence of cDNA and deduced amino acid sequence of rat liver catalase. *Proc. Natl. Acad. Sci. USA* **83,** 313-317

Gilman, M.Z., and Chamberlin, M. (1983) Development and genetic regulation of *Bacillus subtilis* genes translated by σ^{28}-RNA polymerase. *Cell* **35,** 285-293

Glover, D.M. (1984) *Gene Cloning*. Chapman and Hall, New York

Goldberg, I., and Hochman, A. (1989) Purification and characterization of a novel type of catalase from the bacterium *Klebsiella pneumoniae*. *Biochim. Biophys. Acta* **991,** 330-336

Haas, A., Brehm, K., Kreft, J., and Goebel, W. (1991) Cloning, characterization, and expression in *Escherichia coli* of a gene encoding *Listeria seeligeri* catalase, a bacterial enzyme highly homologous to mammalian catalases. *J. Bacteriol.* **173,** 5159-5167

Halliwell, B., and Gutteridge, J.M.C. (1990) Role of free radicals and catalytic metal ions in human disease: an overview. *Methods Enzymol.* **186,** 1-85

Harley, C.B., and Reynolds, R.P. (1987) Analysis of the *E. coli* promoter sequences. *Nucleic Acids Res.* **15,** 2343-2361

Hartig, A., and Ruis, H. (1986) Nucleotide sequence of the *Saccharomyces cerevisiae CTT1* gene and deduced amino acid sequence of yeast catalase T. *Eur. J. Biochem.* **160,** 487-490

Hengge-Aronis, R. (1993) Survival of hunger and stress: the role of *rpoS* in early stationary phase gene regulation in *E. coli*. *Cell* **72,** 165-168

Hengge-Aronis, R., and Fischer, D. (1992) Identification and molecular analysis of *glgS*, a novel growth phase-regulated and *rpoS*-dependent gene involved in glycogen synthesis in *Escherichia coli*. *Mol. Microbiol.* **6,** 1877-1886

Hengge-Aronis, R., Klein, W., Lange, R., Rimmele, M., and Boos, W. (1991) Trehalose synthesis genes are controlled by the putative sigma factor encoded by *rpoS* and are involved in stationary-phase thermotolerance in *Escherichia coli*. *J. Bacteriol.* **173,** 7918-7924

Henikoff, S., Haughn, G.W., Calvo, J.M., and Wallace, J.C. (1988) A large family of bacterial activator proteins. *Proc. Natl. Acad. Sci. USA* **85,** 6602-6606

Higgins, D.G., and Sharp, P.M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *CABIOS* **5,** 151-153

Hillar, A., and Nicholls, P. (1992) A mechanism for NADPH inhibition of catalase compound II formation. *FEBS Lett.* **314,** 179-182

Hochman, A., and Shemesh, A. (1987) Purification and characterization of a catalase-peroxidase from the photosynthetic bacterium *Rhodopseudomonas capsulata*. *J. Biol. Chem.* **262,** 6871-6876

Igual, J.C., González-Bosch, C., Dopazo, J., and Pérez-Ortin, J.E. (1992) Phylogenetic analysis of the thiolase family. Implications for the evolutionary origin of peroxisomes. *J. Mol. Evol.* **35,** 147-155

Imlay, J.A., and Linn, S. (1986) Bimodal pattern of killing of DNA-repair-defective or anoxically grown *Escherichia coli* by hydrogen peroxide. *J. Bacteriol.* **166,** 797-799

Isin, S.H., and Allen, R.D. (1991) Isolation and characterization of a pea catalase cDNA. *Plant Mol. Biol.* **17,** 1263-1265

Jacob, G.S., and Orme-Johnson, W.H. (1979a) Catalase of *Neurospora crassa*. 1. Induction, purification, and physical properties. *Biochemistry* **18,** 2967-2975

Jacob, G.S., and Orme-Johnson, W.H. (1979b) Catalase of *Neurospora crassa*. 2. Electron paramagnetic resonance and chemical properties of the prosthetic group. *Biochemistry* **18,** 2975-2980

Kemmerer, E.C., Lei, M., and Wu, R. (1991) Structure and molecular evolutionary analysis of a plant cytochrome *c* gene: surprising implications for *Arabidopsis thaliana*. *J. Mol. Evol.* **32,** 227-237

Kirkman, H.N., and Gaetani, G.F. (1984) Catalase: a tetrameric enzyme with four tightly bound molecules of NADPH. *Proc. Natl. Acad. Sci. USA* **81,** 4343-4348

Kirkman, H.N., Galiano, S., and Gaetani, G.F. (1987) The function of catalase-bound NADPH. *J. Biol. Chem.* **262,** 660-666

Knauf, H.J., Vogel, R.F., and Hammes, W.P. (1992) Cloning, sequence, and phenotypic expression of *katA*, which encodes catalase of *Lactobacillus sake* LTH677. *Appl. Environ. Microbiol.* **46,** 549-552

Kono, Y., and Fridovich, I. (1982) Superoxide radical inhibits catalase. *J. Biol. Chem.* **257,** 5751-5754

Kono, Y., and Fridovich, I. (1983) Isolation and characterization of the pseudocatalase of *Lactobacillus plantarum*. *J. Biol. Chem.* **258,** 6015-6019

Kunkel, T.A., Roberts, J.D., and Zakour, R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154,** 367-382

Lange, R., and Hengge-Aronis, R. (1991a) Growth phase-regulated expression of *bolA* and morphology of stationary-phase *Escherichia coli* cells are conmtrolled by the novel sigma factor σ^s (*rpoS*). *J. Bacteriol.* **173,** 4474-4481

Lange, R., and Hengge-Aronis, R. (1991b) Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. *Mol. Microbiol.* **5,** 49-59

Layne, E. (1957) Spectrophotometric and turbidimetric methods for measuring proteins. *Methods Enzymol.* **3,** 447-454

Lee, Y.M., Friedman, D.J., and Ayala, F.J. (1985) Superoxide dismutase: an evolutionary puzzle. *Proc. Natl. Acad. Sci. USA* **82,** 824-828

Loewen, P.C. (1984) Isolation of catalase-deficient *Escherichia coli* mutants and genetic mapping of *katE*, a locus that affects catalase activity. *J. Bacteriol.* **157,** 622-626

Loewen, P.C. (1992) Regulation of bacterial catalase synthesis. In *Molecular Biology of Free Radical Scavenging Systems* (Scandalios, J.G., ed) pp. 97-115, Cold Spring Harbor Laboratory Press, New York

Loewen, P.C., and Stauffer, G.V. (1990) Nucleotide sequence of *katG* of *Salmonella typhimurium* LT2 and characterization of its product, hydroperoxidase I. *Mol. Gen. Genet.* **224,** 147-151

Loewen, P.C., and Switala, J. (1986) Purification and characterization of catalase HPII from *Escherichia coli* K12. *Biochem. Cell Biol.* **64,** 638-646

Loewen, P.C., and Switala, J. (1987) Purification and characterization of catalase-1 from *Bacillus subtilis*. *Biochem. Cell Biol.* **65,** 939-947

Loewen, P.C., and Switala, J. (1988) Purification and characterization of spore-specific catalase-2 from *Bacillus subtilis*. *Biochem. Cell Biol.* **66,** 707-714

Loewen, P.C., Switala, J., Smolenski, M., and Triggs-Raine, B.L. (1990) Molecular characterization of three mutations in *katG* affecting the activity of hydroperoxidase I of *Escherichia coli*. *Biochem. Cell Biol.* **68,** 1037-1044

Loewen, P.C., Switala, J., and Triggs-Raine, B.L. (1985a) Catalases HPI and HPII in *Escherichia coli* are induced independently. *Arch. Biochem. Biophys.* **243,** 144-149

Loewen, P.C., and Triggs, B.L. (1984) Genetic mapping of *katF*, a locus that with *katE* affects the synthesis of a second catalase species in *Escherichia coli*. *J. Bacteriol.* **160,** 668-675

Loewen, P.C., Triggs, B.L., George, C.S., and Hrabarchuk, B.E. (1985b) Genetic mapping of katG, a locus that affects synthesis of the bifunctional catalase-peroxidase I in *Escherichia coli*. *J. Bacteriol.* **162,** 661-667

Loewen, P.C., Triggs, B.L., Klassen, G.R., and Weiner, J.H. (1983) Identification and physical characterization of a Col E1 hybrid plasmid containing a catalase gene in *Escherichia coli*. *Can. J. Biochem. Cell. Biol.* **61,** 1315-1321

Loprasert, S., Negro, S., and Okada, H. (1989) Cloning, nucleotide sequence, and expression in *Escherichia coli* of the *Bacillus stearothermophilus* peroxidase gene (*perA*). *J. Bacteriol.* **171,** 4871-4875

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

Margoliash, E., and Novogrodsky, A. (1958) A study of the inhibition of catalase by 3-amino-1:2:4-triazole. *Biochem. J.* **68,** 468-475

Margoliash, E., Novogrodsky, A., and Schejter, A. (1960) Irreversible reaction of 3-amino-1:2:4-triazole and related inhibitors with the protein of catalase. *Biochem. J.* **74,** 339-350

McCann, M.P., Kidwell, J.P., and Matin, A. (1991) The putative σ factor KatF has a central role in development of starvation-mediated general resistance in *Escherichia coli*. *J. Bacteriol.* **173,** 4188-4194

Mead, D.A., Skorupa, E.S., and Kemper, B. (1985) Single-stranded DNA SP6 promoter plasmids for engineering mutant RNAs and proteins: synthesis of a "stretched" preparathyroid hormone. *Nucleic Acids Res.* **13,** 1103-1118

Meir, E., and Yagil, E. (1985) Further characterization of the two catalases in *Escherichia coli*. *Curr. Microbiol.* **12,** 315-320

Miller, J.H. (1972) *Experiments in Molecular Genetics,* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Moody, P.C.E., and Wilkinson, A.J. (1990) in *Protein Engineering,* pp. 28-33, IRL Press, Oxford

Morelle, G. (1989) A plasmid extraction procedure on a miniprep scale. *Focus* **11,** 7-8

Morgan, R.W., Christman, M.F., Jacobson, F.S., Storz, G., and Ames, B.N. (1986) Hydrogen peroxide-inducible proteins in *Salmonella typhimurium* overlap with heat shock and other stress proteins. *Proc. Natl. Acad. Sci. USA* **83,** 8059-8063

Mori, H., Higo, K., Higo, H., Minobe, Y., Matsui, H., and Chiba, S. (1992) Nucleotide and derived amino acid sequence of a catalase cDNA isolated from rice immature seeds. *Plant Mol. Biol.* **18,** 973-976

Mörikofer-Zwez, S., Cantz, M., Kaufman, H., von Wartburg, J.P., and Aebi, H. (1969) Heterogeneity of erythrocyte catalase. Correlations between sulfhydryl group content, chromatographic and electrophoretic properties. *Eur. J. Biochem.* **11,** 49-57

Mulvey, M.R., and Loewen, P.C. (1989) Nucleotide sequence of *katF* of *Escherichia coli* suggests KatF protein is a novel σ transcription factor. *Nucleic Acids Res.* **17,** 9979-9991

Mulvey, M.R., Sorby, P.A., Triggs-Raine, B.L., and Loewen, P.C. (1988) Cloning and physical characterization of *katE* and *katF* required for catalase HPII expression in *Escherichia coli*. *Gene* **73,** 337-345

Mulvey, M.R., Switala, J., Borys, A., and Loewen, P.C. (1990) Regulation of transcription of *katE* and *katF* in *Escherichia coli*. *J. Bacteriol.* **172,** 6713-6720

Murshudov, G.N., Melik-Adamyan, W.R., Grebenko, A.I., Barynin, V.V., Vagin, A.A., Vainshtein, B.K., Dauter, Z., and Wilson, K.S. (1992) Three-dimensional structure of catalase from *Micrococcus lysodeikticus* at 1.5 Å resolution. *FEBS Lett.* **312,** 127-131

Murthy, M.R.N., Reid, T.J., Sicignano, A., Tanaka, N., and Rossmann, M.G. (1981) Structure of beef liver catalase. *J. Mol. Biol.* **152,** 465-499

Ni, W., Turley, R.B., and Trelease, R.N. (1990) Characterization of a cDNA encoding cottonseed catalase. *Biochim. Biophys. Acta* **1049,** 219-222

Ogura, Y. (1955) Catalase activity at high concentration of hydrogen peroxide. *Arch. Biochem. Biophys.* **57,** 288-300

Okada, H., Veda, M., Sugaya, T., Atomi, H., Mozaffer, S., Hishida, T., Teranishi, Y., Okazaki, K., Takechi, T., Kamiryo, T., and Tanaka, A. (1987) Catalase gene of the yeast *Candida tropicalis*. Sequence analysis and comparison with peroxisomal and cytosolic catalases from other sources. *Eur. J. Biochem.* **170,** 105-110

Orr, E.C., Bewley, G.C., and Orr, W.C. (1990) cDNA and deduced amino acid sequence of *Drosophila* catalase. *Nucleic Acids Res.* **18,** 3663

Pasternak, J.J., and Glick, B.R. (1992) Molecular evolutionary analyses of the small and large subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Can. J. Bot.* **70,** 715-723

Pegg, M., Crane, D., and Masters, C. (1986) Confirmation that catalase is a glycoprotein. *Biochem. Int.* **12,** 831-838

Reid, T.J. III, Murthy, M.R.N., Sicignano, A., Tanaka, N., Musick, W.D.L., and Rossmann, M.G. (1981) Structure and heme environment of beef liver catalase at 2.5 Å resolution. *Proc. Natl. Acad. Sci. USA* **78,** 4767-4771

Richter, H.E., and Loewen, P.C. (1982) Catalase synthesis in *Escherichia coli* is not controlled by catabolite repression. *Arch. Biochem. Biophys.* **215,** 72-77

Rørth, M., and Jensen, P.K. (1967) Determination of catalase activity by means of the Clark oxygen electrode. *Biochim. Biophys. Acta* **139,** 171-173

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406-425

Sak, B.D., Eisenstark, A., and Touati, D. (1989) Exonuclease III and the catalase hydroperoxidase II in *Escherichia coli* are both regulated by the *katF* product. *Proc. Natl. Acad. Sci. USA* **86,** 3271-3275

Sakajo, S., Nakamura, K., and Asahi, T. (1987) Molecular cloning and nucleotide sequence of full-length cDNA for sweet potato catalase mRNA. *Eur. J. Biochem.* **165,** 437-442

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sanger, F.S., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463-5467

Saporito, S.M., Smith-White, B.J., and Cunningham, R.P. (1988) Nucleotide sequence of the xthA gene of *Escherichia coli* K-12. *J. Bacteriol.* **170,** 4542-4547

Schellhorn, H.E., and Hassan, H.M. (1988) Transcriptional regulation of *katE* in *Escherichia coli* K-12. *J. Bacteriol.* **170,** 4286-4292

Schellhorn, H.E., and Stones, V.L. (1992) Regulation of *katF* and *katE* in *Escherichia coli* K-12 by weak acids. *J. Bacteriol.* **174,** 4769-4776

Schonbaum, G.R., and Chance, B. (1976) Catalase. In *The Enzymes* (Boyer, P.D., ed) 3rd ed., vol. 13, pp. 363-408, Academic Press, New York

Schroeder, W.A., Shelton, J.R., Shelton, J.B., Apell, G., Evans, L., Bonaventura, J., and Fang, R.S. (1982b) The partial amino acid sequence of human erythrocyte catalase. *Arch. Biochem. Biophys.* **214,** 422-424

Schroeder, W.A., Shelton, J.R., Shelton, J.B., Robberson, B., Apell, G., Fang, R.S., and Bonaventura, J. (1982a) The complete amino acid sequence of bovine liver catalase and the partial sequence of bovine erythrocyte catalase. *Arch. Biochem. Biophys.* **214,** 397-421

Shaffer, J.B., Preston, K.E., and Shepard, B.A. (1990) Nucleotide and deduced amino acid sequences of mouse catalase: molecular analysis of a low affinity mutant. *Nucleic Acids Res.* **18,** 4941

Siegele, D.A., and Kolter, R. (1992) Life after log. *J. Bacteriol.* **174,** 345-348

Simons, R.W., Houman, F., and Kleckner, N. (1987) Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene* **53,** 85-96

Smith, M.W., and Doolittle, R.F. (1992) A comparison of evolutionary rates of the two major kinds of superoxide dismutase. *J. Mol. Evol.* **34,** 175-184

Sogin, M.L., Elwood, H.J., and Gunderson, J.H. (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc. Natl. Acad. Sci. USA* **83,** 1383-1387

Sorby, P.A. (1989) Cloning and partial characterization of *katE* encoding HPII catalase in *Escherichia coli*. M.Sc Thesis. University of Manitoba, Winnipeg

Stayton, P.S., Atkins, W.M., Springer, B.A., and Sligar, S.G. (1989) Site-directed mutagenesis of heme proteins. In *Metal Ions in Biological Systems: Interrelations Among Metal Ions, Enzymes, and Gene Expression* (Sigel, H., and Sigel, A., eds) vol. 25, pp. 417-475, Marcel Dekker, Inc., New York

Storz, G., Tartaglia, L.A., and Ames, B.N. (1990) Transcriptional regulator of oxidative stress-inducible genes: direct activation by oxidation. *Science* **248,** 189-194

Sumner, J.B., and Dounce, A.L. (1937) Crystalline catalase. *J. Biol. Chem.* **121,** 417-424

Takeda, A., Miyahara, T., Hachimori, A., and Samejima, T. (1980) The interactions of thiol compounds with porcine erythrocyte catalase. *J. Biochem.* **87,** 429-439

Tao, K., Makino, K., Yonei, S., Nakata, A., and Shinagawa, H. (1989) Molecular cloning and nucleotide sequencing of *oxyR*, the positive regulator gene of a regulon for an adaptive response to oxidative stress in *Escherichia coli*: homologies between OxyR protein and a family of bacterial activator proteins. *Mol. Gen. Genet.* **218,** 371-376

Tormo, J., Fita, I., Switala, J., and Loewen, P.C. (1990) Crystallization and preliminary X-ray diffraction analysis of catalase HPII from *Escherichia coli*. *J. Mol. Biol.* **213,** 219-220

Triggs-Raine, B.L., Doble, B.W., Mulvey, M.R., Sorby, P.A., and Loewen, P.C. (1988) Nucleotide sequence of *katG* encoding catalase HPI of *Escherichia coli*. *J. Bacteriol.* **170,** 4415-4419

Triggs-Raine, B.L., and Loewen, P.C. (1987) Physical characterization of *katG* encoding catalase HPI of *Escherichia coli*. *Gene* **52,** 121-128

Vainshtein, B.K., Melik-Adamyan, W.R., Barynin, V.V., Vagin, A.A., Grebenko, A.I., Borisov, V.V., Bartels, K.S., Fita, I., and Rossmann, M.G. (1986) Three-dimensional structure of catalase from *Penicillium vitale* at 2.0 Å resolution. *J. Mol. Biol.* **188,** 49-61

Vieira, J., and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods Enzymol.* **153,** 3-11

Wang, J.H. (1955) On the detailed mechanism of a new type of catalase-like action. *J. Am. Chem. Soc.* **77,** 4715-4719

Weber, K., and Osborn, M. (1969) The reliability of molecular weight determinations by dodecyl sulfate polyacrylamide gel electrophoresis. *J. Biol. Chem.* **244,** 4406-4412

Welinder, K.G. (1992) Superfamily of plant, fungal and bacterial peroxidases. *Curr. Opin. Struc. Biol.* **2,** 388-393

Wolfe, K.H., Gouy, M., Yang, Y.-W., Sharp, P.M., and Li, W.-H. (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. USA* **86,** 6201-6205