

Carbamoylphosphate synthetase (CPS) genes of *Neisseria gonorrhoeae* and other *Neisseria* species: novel gene organisation, variable intergenic sequences, characterisation of naturally occurring mutants, and evolution of CPS genes.

Les gènes codant pour la carbamoylphosphate synthetase (CPS) chez *Neisseria gonorrhoeae* et autres espèces de *Neisseria*: nouvelle organisation génétique, séquences intergéniques variables, caractérisation de mutants naturels, et évolution des gènes CPS.

FIONA S. LAWSON

Thesis submitted to the
School of Graduate Studies and Research
University of Ottawa
in partial fulfillment of the requirements for the
Ph.D. degree in the
Ottawa-Carleton Institute of Biology

Thèse soumise à
l'Ecole des études supérieures et de la recherche
Université d'Ottawa
en vue de l'obtention de la Ph.D. à
L'Institut de biologie d'Ottawa-Carleton

Copyright © 1996 by Fiona Lawson



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-16419-5

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

This work is dedicated to my parents. With unyielding support, love and friendship, (not to mention a hefty financial contribution!) they were my rock through these university years. I could not have done it without them and can't tell them enough how thankful I am for all they've done.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Dr. Jo-Anne R. Dillon. She has played the most influential role in my development as a researcher and I am forever grateful for her guidance.

I would like to thank my advisory committee members Dr. Linda Bonen, Dr. George Carmody and Dr. Doug Johnson, for valuable questions and comments, and also Dr. Lai-King Ng for help in figuring out those little problems. I would like to especially thank Finola Billowes, who with me performed many of the PCR reactions and Southern blots critical to studies of the sequence between *carA* and *carB*. I thank Dr. Craig Strathdee for supplying chromosomal DNA from *Neisseria meningitidis* and commensal *Neisseria spp.* isolates, and for insightful comments, and I also thank Dr. Ahmed T. Abdelal and Dr. Chung-dar Lu for supplying CPS synthetase sequences from *Pseudomonas stutzeri* and *Pseudomonas aeruginosa* before their submission to GenBank. I thank Jim Ng for supplying chromosomal DNA from a *Peptostreptococcus magus* strain, and also Jill Parsons for performing a PCR amplification to obtain a portion of *argJ*. I would especially like to thank Dr. Robert L. Charlebois for generously agreeing to give me access to the *S. solfataricus* CPS sequence as part of a collaboration.

I would also like to acknowledge my many fellow graduate students and co-workers, who always made me smile: amazing-amplicon-amplifying-Avni, “interesting”-stories-César, Finola-not-fiona-or-madonna, long-distance-relationships-forever-Franco, fellow-thesis-writing-maniac-Jim, Lang-Li (not evil), that’s-cute!-Tholib, Sir Stéphane (who ascertains and delineates a cornucopia of mayhem), and so many others... J’aimerais remercier tout particulièrement Stéphane Bernatchez pour l’aide appréciable qu’il m’a accordée lors de la traduction en français du résumé de ma thèse. A special

thanks goes out to Ray and Jeanne Burnham for putting me up during the final month of my degree. I would also like to acknowledge the many people at the National Laboratory for Sexually Transmitted Diseases for their unending help during those early days when our lab was with Health Canada.

Lastly, I wish to thank Ryan and my parents. Ryan, whose love, support and Thoughts kept me going throughout the writing of this thesis (ilum) and with whom I can now see the future... My parents, who have done so much for me and more.

This work was funded in part by the Canadian Bacterial Diseases Network, the Natural Sciences and Engineering Research Council, and the biotechnology initiative for federal laboratories.

ABSTRACT

Carbamoylphosphate synthetase (CPS) catalyses the formation of carbamoylphosphate from CO₂, ATP, and glutamine; the first committed step in the arginine and pyrimidine biosynthetic pathways. Initial interest in the CPS enzyme of *Neisseria gonorrhoeae* stemmed from the natural occurrence of CPS deficiencies in 10-20% of clinical isolates. This property, causing a concurrent citrulline and uracil auxotrophy, is employed in a typing scheme and there have been some indications that this auxotrophy may be beneficial to the gonococcus.

CPS has been studied in many prokaryotes as a model for investigations of gene and enzyme regulation. In all of the prokaryotes examined, CPS is a heterodimer encoded by two genes commonly called *carA* and *carB*. These genes are separated by up to 24 bp and are co-transcribed in most of the prokaryotes examined. However, recently it was reported that *Pseudomonas aeruginosa carA* and *carB* are separated by 682 bp, though they are still co-transcribed.

I have now determined the complete sequence of the *carA* and *carB* genes from *N. gonorrhoeae* CH811. *carA* (1125 bp) and *carB* (3237 bp) are similar in size to other prokaryotic CPS genes, and encode all the highly conserved regions present in other CPS's. However, these genes in strain CH811 are separated by a 3287 bp intergenic sequence which has no similarity to the corresponding sequence in *P. aeruginosa*, and putative transcription terminators are found downstream of both genes. Northern blot analysis has demonstrated that these gonococcal CPS genes are separately transcribed, a gene organisation unlike that observed in any other prokaryote. Several neisserial repetitive sequences were identified within the 9 kb sequenced, as well as novel 120 and 150 bp repeats (designated RS6 and RS7) which were found within the intervening sequence between *carA* and *carB*.

To determine whether the intervening sequence observed in *N. gonorrhoeae* CH811 was typical of gonococcal isolates, the sequence between *carA* and *carB* was PCR amplified from 30 isolates of *N. gonorrhoeae*. The intervening sequence was found to vary in size, from approximately 2.2 to 3.7 kb, although the *carA* and *carB* genes themselves did not vary in size in isolates with functional CPS. A similar large, variably sized, intervening sequence was also found between the *carA* and *carB* genes of 12 isolates of *Neisseria meningitidis* and 18 commensal *Neisseria* isolates comprising 9 species. This unexpected organization of the CPS genes in *N. gonorrhoeae* is therefore widespread throughout the genus *Neisseria*. This work suggests that CPS gene organization may be much more variable within the prokaryotes than was previously assumed.

The variably sized intergenic sequence was further studied in more gonococcal isolates using PCR and restriction enzyme analysis (REA), to evaluate the utility of this sequence for typing of gonococci. PCR-REA of the intergenic sequence was able to discriminate between many isolates of the same A/S/P class and had a level of discrimination analogous to ribotyping and chromosomal REA typing methods. Hybridization studies indicated that the RS6/RS7 repeats present in the intergenic sequence of *N. gonorrhoeae* CH811 may also be diagnostic for gonococci.

The presence of a variable sequence near *carA* and *carB*, and the numerous repetitive sequences found in this area in strain CH811, suggests that this region may be a hotspot for recombination and may in part explain the high frequency of occurrence of CPS mutants. I have begun a preliminary investigation of the naturally occurring CPS mutants in gonococcal isolates: PCR amplification of *carA* and *carB* from CPS deficient isolates indicates that more than one mutation is responsible for this phenotype. Transformation of a gonococcal CPS mutant with the cloned gonococcal CPS genes has demonstrated that other mutations are present in the arginine and

pyrimidine biosynthetic pathways of this isolate, which are hidden by the CPS deficiency. This study also suggests that isolates of auxotype CUH and OUH are closely related.

I also report an analysis of the first CPS gene sequence from an archaeon (*Sulfolobus solfataricus* P2; supplied by R.L. Charlebois). This analysis indicates that the archaeal CPS is similar in function, size and gene organization to other prokaryotic CPS's, contrary to a previous study of an archaeal CPS protein. This demonstrates that many features of CPS were present in the progenote. Internal similarity previously observed within other CPS enzymes was also observed in this archaeal CPS, confirming that this internal similarity was probably the result of an ancient gene duplication which preceded the divergence of Bacteria, Archaea and Eukarya.

A comprehensive phylogenetic analysis of all known CPS genes, including the gonococcal and archaeal sequences, was performed. Using the internal duplication as a root for phylogenetic tree construction, I was able to determine a branching order for the tree of life and provide support for the theory that this archaeon is more related to eukaryotes than bacteria. This is the first report of confirmation of this theory using a metabolic gene and is the first use of an internal duplication within a gene to root the tree of life. This analysis also showed that at least two, and possibly even three, separate gene duplication events led to the formation of the two CPS enzymes present in Gram-positive bacteria and most eukaryotes. Comparisons of the branching order with the organization of CPS genes in different organisms further indicated that the evolution of CPS has been complex, involving multiple deletions and insertions.

RESUME

La carbamoylphosphate synthétase (CPS) catalyse la formation du carbamoylphosphate à partir de CO₂, d'ATP et de glutamine, ce qui constitue la première étape engagée dans les sentiers biosynthétiques de l'arginine et des pyrimidines. L'intérêt initial pour la CPS de *Neisseria gonorrhoeae* a dérivé de l'incidence naturelle de déficiences en CPS chez 10-20% des isolats cliniques. Cette propriété, causant simultanément des auxotrophies à la citrulline et à l'uracile, est employée dans un système de typage et il y a eu des indications que cette auxotrophie pourrait être bénéfique au gonocoque.

La CPS a été étudiée chez de nombreux procaryotes en tant que modèle de recherche sur la régulation génétique et enzymatique. Chez tous les procaryotes examinés, la CPS est un hétérodimère encodé par deux gènes communément désignés *carA* et *carB*. Ces gènes sont séparés par jusqu'à 24 paires de bases (pb) et sont co-transcrits chez la plupart des procaryotes examinés. Cependant, il a été récemment rapporté que *carA* et *carB* étaient séparés par 682 pb chez *Pseudomonas aeruginosa* mais qu'ils étaient tout de même co-transcrits.

J'ai maintenant déterminé la séquence complète des gènes *carA* et *carB* de *N. gonorrhoeae* CH811. *carA* (1125 pb) et *carB* (3237 pb) ont une taille semblable à celle des autres CPS bactériennes et codent pour toutes les régions conservées présentes chez les autres CPS. Cependant, les gènes de la souche CH811 sont séparés par une séquence intergénique de 3287 pb n'ayant aucune similarité avec les séquences correspondantes de *P. aeruginosa*, et de présumés terminateurs transcriptionnels ont été identifiés en aval de chacun des deux gènes. L'analyse par hybridation Northern a montré que ces gènes CPS gonococaux sont transcrits séparément, ce qui constitue une organisation génétique différente de celle observée chez tout autre procaryote.

Plusieurs séquences répétitives neisseriales ont été identifiées à l'intérieur des 9 kilobases (kb) séquencés, avec en plus des nouvelles séquences répétées de 120 et 150 pb (désignées RS6 et RS7) qui étaient présentes dans la séquence intervenante entre *carA* et *carB*.

Afin de déterminer si la séquence intergénique observée chez *N. gonorrhoeae* CH811 était typique d'isolats gonococaux, la séquence entre *carA* et *carB* fut amplifiée par la PCR à partir de 30 isolats de *N. gonorrhoeae*. Il fut établi que la séquence intergénique était de taille variable, environ 2.2 à 3.7 kb malgré que les gènes *carA* et *carB* eux-mêmes ne changeaient pas de taille chez les isolats ayant une CPS fonctionnelle. Une large et similaire séquence intergénique de taille variable fut aussi trouvée entre les gènes *carA* et *carB* chez 12 isolats de *N. meningitidis* et 18 isolats commensaux comprenant 9 espèces de *Neisseria*. Cette organisation inattendue des gènes CPS chez *N. gonorrhoeae* est donc répandue à travers le genre *Neisseria*. Ce travail suggère que l'organisation des gènes CPS puisse être beaucoup plus variable que précédemment assumée chez les procaryotes.

La séquence intergénique de taille variable fut également examinée chez d'autres isolats gonococaux par la PCR et l'analyse de fragments de restriction (REA) afin d'évaluer l'utilité de cette séquence pour typer les gonocoques. Le PCR-REA de la séquence intergénique parvint à discriminer de nombreux isolats de même classe A/S/P et avait un niveau de discrimination analogue à celui obtenu avec d'autres méthodes de typage telles le ribotypage et l'analyse de restriction chromosomale. Des études par hybridation ont indiqué que les répétitions RS6/RS7 présentes dans la séquence intergénique de *N. gonorrhoeae* CH811 pourrait aussi être diagnostiques pour les gonocoques.

La présence d'une séquence variable près de *carA* et *carB* et des nombreuses séquences répétitives trouvées dans ce secteur dans la souche CH811 suggèrent que cette région pourrait être un point chaud pour la recombinaison et pourrait en partie expliquer la fréquence élevée de mutants CPS. J'ai débuté une étude préliminaire des mutants CPS rencontrés naturellement dans des isolats gonococaux: l'amplification par la PCR de *carA* et de *carB* à partir d'isolats à CPS déficiente indique que plus d'une mutation est responsable de ce phénotype. La transformation d'un mutant CPS gonococcal avec les gènes CPS gonococaux a montré que d'autres mutations sont présentes dans les sentiers biosynthétiques de l'arginine et des pyrimidines chez cet isolat, et ces mutations sont cachées par la déficience en CPS. Cette étude suggère également que les isolats d'auxotypes CUH et OUH sont étroitement liés.

Je rapporte également une analyse de la première séquence archéobactérienne d'un gène CPS (fournie par R.L. Charlebois). Cette analyse indique que la CPS archéobactérienne est semblable en fonction, en taille et en organisation génétique aux autres CPS procaryotiques, contrairement à une étude antérieure sur une protéine CPS archéobactérienne. Une similarité interne observée antérieurement chez d'autres enzymes CPS fut aussi observée dans cette CPS archéobactérienne, confirmant que cette similarité interne était probablement le résultat d'une ancienne duplication génique ayant précédé la divergence des eubactéries, des archéobactéries et des eucaryotes. Une analyse phylogénétique détaillée de tous les gènes CPS connus, incluant les séquences gonococcale et archéobactérienne, fut effectuée. En utilisant la duplication interne comme racine pour la construction d'un arbre phylogénétique, je suis arrivée à déterminer un ordre de ramification pour l'arbre de la vie et à appuyer la théorie voulant que les archéobactéries soient liées de plus près aux eucaryotes que le sont les eubactéries. C'est la première confirmation

de cette théorie en utilisant un gène métabolique et c'est la première fois qu'une duplication interne à un gène est utilisée pour enraciner l'arbre de la vie. Cette analyse a également démontré qu'au moins deux et peut-être même trois duplications génétiques séparées ont mené à la formation des deux enzymes CPS présentes chez les eubactéries Gram-positives et chez la plupart des eucaryotes. Des comparaisons de l'ordre de ramification avec l'organisation des gènes CPS dans différents organismes indiqua encore davantage que l'évolution de la CPS a été complexe, impliquant de multiples délétions et insertions.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	i
ABSTRACT.....	iii
RESUME.....	vi
TABLE OF CONTENTS.....	x
LIST OF TABLES.....	xvi
LIST OF FIGURES.....	xviii
LIST OF ABBREVIATIONS.....	xxi
1. INTRODUCTION.....	1
1.1 THE NEISSERIACEAE AND <i>NEISSERIA GONORRHOEAE</i> : OVERVIEW AND IMPORTANCE IN DISEASE.....	1
1.2 THE GONOCOCCUS: HETEROGENEITY AND GENE STRUCTURE.....	3
1.3 NATURALLY OCCURRING ARGININE-REQUIRING GONOCOCCI.....	7
1.4 GENETIC STUDIES OF ARGININE BIOSYNTHESIS AND CPS IN THE GONOCOCCUS.....	10
1.5 OVERVIEW OF ARGININE BIOSYNTHESIS AND THE UREA/ORNITHINE CYCLE: COMPARISONS BETWEEN SPECIES.....	12
1.6 OVERVIEW OF PYRIMIDINE BIOSYNTHESIS.....	17
1.7 CARBAMOYLPHOSPHATE SYNTHETASE: ENZYME FUNCTION AND STRUCTURE.....	21

1.8	CPS GENE ORGANIZATION AND REGULATION IN THE BACTERIA.....	26
1.9	CPS GENES IN THE EUKARYOTES: FEW STUDIES OF REGULATION.....	29
1.10	PHYLOGENETIC ANALYSIS OF CPS: ROOTING THE TREE OF LIFE AND THE NEED FOR AN ARCHAEOAL SEQUENCE.....	31
1.11	OBJECTIVES AND RATIONALE.....	35
1.11.1	Rationale for initial study.....	35
1.11.2	Objectives.....	36
2.	MATERIALS AND METHODS.....	38
2.1	BACTERIAL STRAINS, PLASMIDS AND PHAGES.....	38
2.2	MEDIA, GROWTH CONDITIONS AND STRAIN TYPING.....	43
2.3	DNA PURIFICATION, RESTRICTION ENDONUCLEASE DIGESTION, AND ELECTROPHORESIS.....	45
2.4	SUBCLONING OF DNA FRAGMENTS.....	46
2.5	SOUTHERN HYBRIDIZATION STUDIES.....	47
2.6	GENERATION OF PLASMIDS WITH NESTED DELETIONS.....	48
2.7	DNA SEQUENCING AND SEQUENCE ANALYSIS.....	49
2.8	POLYMERASE CHAIN REACTION (PCR).....	50
2.9	RNA EXTRACTION AND ANALYSIS BY NORTHERN BLOT.....	53

2.10	TRANSFORMATION OF MUTANT GONOCOCCAL ISOLATES WITH CPS GENES.....	55
2.11	PHYLOGENETIC ANALYSIS.....	56
2.11.1	DNA sequences.....	56
2.11.2	Alignment of sequences.....	57
2.11.3	Phylogenetic tree construction.....	58
3.	RESULTS.....	59
3.1	LOCATING AND SUBCLONING GONOCOCCAL DNA FRAGMENTS CONTAINING <i>CARA</i> AND <i>CARB</i>	59
3.2	ANALYSIS OF THE SEQUENCE OF PFL50 AND PFL35: <i>CARA</i> AND <i>CARB</i> ARE SIMILAR TO OTHER CPSII GENES, BUT ARE SEPARATED BY A UNIQUE INTERGENIC SEQUENCE AND SURROUNDED BY MANY REPETITIVE SEQUENCES.....	68
3.2.1	<i>carA</i> and <i>carB</i> in <i>N. gonorrhoeae</i> CH811 encode a protein very similar to previously characterized CPSII enzymes.....	68
3.2.2	<i>carA</i> and <i>carB</i> in <i>N. gonorrhoeae</i> CH811 are separated by a unique, large intervening sequence and, unlike other prokaryotes, putative transcription terminators are found downstream of both genes.....	72
3.2.3	Numerous previously characterised repetitive sequences are found close to <i>carA</i> and <i>carB</i> in <i>Neisseria gonorrhoeae</i> CH811, particularly upstream of <i>carA</i>	75
3.2.4	Novel repetitive sequences are found in the sequence between <i>carA</i> and <i>carB</i> in <i>Neisseria gonorrhoeae</i> CH811.....	76

3.3	NORTHERN ANALYSIS: <i>CAR</i> AND <i>CAR</i> ARE SEPARATELY TRANSCRIBED.....	79
3.4	A VARIABLY-SIZED INTERGENIC SEQUENCE IS PRESENT BETWEEN <i>CAR</i> AND <i>CAR</i> IN <i>N. GONORRHOEAE</i> AND OTHER <i>NEISSERIA</i> SPECIES.....	81
3.4.1	PCR amplification of the intergenic sequence from gonococcal isolates: it varies in size.....	81
3.4.2	PCR amplification of the intergenic sequence from other <i>Neisseria</i> species: a similar large, variable sequence is present between <i>car</i> and <i>car</i>	84
3.4.3	Comparison of the intergenic sequences between <i>car</i> and <i>car</i> from different <i>Neisseria</i> species by Southern Hybridization: similarities exist, but RS6 and RS7 may be gonococcal-specific.....	85
3.4.4	Comparison of the neisserial intergenic sequences by restriction enzyme analysis (REA) reveals more variability between the intergenic sequences, though a conserved region is present in the gonococcal intergenic sequences.....	91
3.4.5	Investigating the variable intergenic sequence from more closely related gonococcal isolates: possible utility for typing of <i>N. gonorrhoeae</i>	94
3.4.6	Comparison of the gonococcal intergenic sequences by PCR amplification of selected sub-regions: changes in the number of RS6/RS7 repeats explains some, but not all, changes in the size of the intergenic sequences.....	99
3.5	STUDIES OF NATURALLY OCCURRING GONOCOCCAL CPS MUTANTS: CAUSED BY MORE THAN ONE MUTATION AND HIDING OTHER MUTATIONS.....	101
3.5.1	<i>car</i> and <i>car</i> are conserved in size in wild type isolates but not in CPS deficient isolates: more than one mutation led to the formation of CPS deficiencies in gonococcal isolates.....	101

3.5.2	Transformation of gonococci deficient in CPS using cloned <i>carA</i> and <i>carB</i> : multiple mutations are present in the arginine and pyrimidine biosynthetic pathways of CUH auxotrophs.....	102
3.6	PHYLOGENETIC ANALYSIS OF CPS GENES: COMPLEX EVOLUTIONARY HISTORY INCLUDES AN INTERNAL DUPLICATION WITHIN A GENE WHICH CAN ROOT THE TREE OF LIFE.....	106
3.6.1	The phylogenetic analysis includes analysis of the first CPS genes from an archaeobacterium: <i>Sulfolobus solfataricus</i> CPS is found to be encoded by overlapping genes and contain protein motifs consistent with a CPSII enzyme.....	106
3.6.2	Phylogenetic analysis of complete CPS genes: at least two separate gene duplication events led to the formation of the two CPS enzymes found in Gram-positive bacteria and in eukaryotes.....	107
3.6.3	Phylogenetic trees constructed using the internal duplication within the synthetase domain of CPS: rooting the tree of life groups Archaea with eukaryotes.....	110
3.6.4	Evolution of the organization of CPS genes: the intervening sequence between the amidotransferase and synthetase domains of heterodimeric CPS involved multiple deletions and insertions.....	113
4.	DISCUSSION.....	115
4.1	NOVEL GENE ORGANIZATION FOR <i>CARA</i> AND <i>CARB</i> IN <i>N. GONORRHOEAE</i> : SEPARATE TRANSCRIPTION OF <i>CARA</i> AND <i>CARB</i> , AND THE PRESENCE OF A VARIABLE INTERGENIC SEQUENCE, LEADS TO MANY NEW HYPOTHESES REGARDING GONOCOCCAL CPS GENE REGULATION.....	115

4.2	<i>CARA</i> AND <i>CARB</i> MAY RESIDE WITHIN A HOTSPOT FOR RECOMBINATION: PARTIAL EXPLANATION FOR NATURALLY OCCURRING CPS DEFICIENCIES.....	120
4.3	COMPARISON OF THE SEQUENCE BETWEEN <i>CARA</i> AND <i>CARB</i> IN <i>NEISSERIA</i> SPECIES: REVEALING SEQUENCES WHICH MAY BE DIAGNOSTIC FOR <i>N. GONORRHOEAE</i>	123
4.4	UTILITY OF THE VARIABLE INTERGENIC SEQUENCE FOR TYPING OF GONOCOCCAL ISOLATES.....	125
4.5	STUDIES OF THE CPS DEFICIENT MUTANTS REVEAL NEW RELATIONSHIPS WITHIN AND BETWEEN AUXOTYPES.....	126
4.6	PHYLOGENETIC ANALYSIS OF CPS GENES: ROOTING THE TREE OF LIFE AND REVEALING A COMPLEX HISTORY FOR CPS GENE EVOLUTION.....	127
4.7	CONCLUSIONS AND IMPACTS.....	136
5.	REFERENCES.....	138
6.	APPENDIX.....	165

LIST OF TABLES

TABLE 1.	<i>N. gonorrhoeae</i> isolates selected for their varying A/S/P class, date of isolation and geographic origin.....	40
TABLE 2.	Isolates of <i>N. gonorrhoeae</i> selected for their similar A/S/P class (P/IA-6/2.6,4.5), but varying date of isolation and geographic origin within two Canadian provinces.....	41
TABLE 3.	Two clusters of isolates of <i>N. gonorrhoeae</i> with an A/S class similar to isolates associated with two separate outbreaks: 1) isolates of A/S class NR/IA-5, associated with an outbreak in Toronto; 2) isolates of the A/S/P class NR/IB-3/2.6/24.5 associated with an outbreak in Winnipeg which spread to Toronto....	42
TABLE 4.	<i>N. meningitidis</i> and commensal <i>Neisseria</i> species isolates.....	44
TABLE 5.	Accurate location of the priming sites for primers used in PCR amplifications, based on the total sequence obtained from pFL50 and pFL35.....	51
TABLE 6.	Amplicon size and REA pattern for the intergenic sequence between <i>carA</i> and <i>carB</i> from isolates of <i>N. gonorrhoeae</i> of varying A/S/P class, and location and date of isolation.....	83
TABLE 7.	Amplicon size and REA pattern for the intergenic sequence between <i>carA</i> and <i>carB</i> from isolates of <i>N. meningitidis</i> and commensal <i>Neisseria</i> species.....	86
TABLE 8.	Summary of hybridization studies comparing portions of the intergenic sequence from <i>N. gonorrhoeae</i> CH811 with the intergenic sequences from other <i>Neisseria</i> species and with chromosomal DNA from <i>E. coli</i> and <i>P. magnus</i>	90

TABLE 9.	Amplicon size and REA pattern for the intergenic sequence between <i>carA</i> and <i>carB</i> from <i>N. gonorrhoeae</i> isolates of similar A/S/P class (P/IA-6/2.6,4.5), but varying date of isolation and geographic origin within two Canadian provinces.....	95
TABLE 10.	Amplicon size and REA pattern for the intergenic sequence between <i>carA</i> and <i>carB</i> from <i>N. gonorrhoeae</i> isolates of similar A/S/P class, date of isolation, and geographic location as isolates associated with two outbreaks of gonorrhea.....	98

LIST OF FIGURES

FIGURE 1.	The arginine and pyrimidine biosynthetic pathways.....	9
FIGURE 2.	Physical map of the chromosome of <i>N. gonorrhoeae</i> FA1090.....	14
FIGURE 3.	Schematic diagram of the insert from pMC50 which contains <i>E. coli carAB</i>	22
FIGURE 4.	Schematic diagram of the organization of CPS genes from all organisms in which CPS gene sequence data has been obtained.....	27
FIGURE 5.	Location of <i>carA</i> and <i>carB</i> in the gonococcal inserts of phage and plasmid clones.....	39
FIGURE 6.	Schematic diagram of the sequence comprising <i>carA</i> and <i>carB</i> from <i>N. gonorrhoeae</i> CH811, showing the priming location for each primer used for PCR amplifications.....	52
FIGURE 7.	Hybridization of gonococcal clones with an <i>E. coli carAB</i> probe.....	60
FIGURE 8.	Confirmation that pFL35 contains the 3.5 kb <i>EcoRI/XhoI</i> fragment of pSTD18.....	62
FIGURE 9.	Hybridization of fragments from the gonococcal clone EFP10 with <i>E. coli carA</i>	63
FIGURE 10.	Confirmation of the origin of the insert in pFL50 including hybridisation of this insert to a specific fragment of gonococcal chromosomal DNA.....	65
FIGURE 11.	Hybridization of deletion derivatives of pFL50 with <i>E. coli carA</i>	66
FIGURE 12.	The complete sequence of pFL50 and pFL35, comprising the <i>carA</i> and <i>carB</i> genes of <i>N. gonorrhoeae</i> CH811.....	69

FIGURE 13.	Deduced amino acid sequence of the CPS subunit: encoded by <i>N. gonorrhoeae</i> CH811 <i>carA</i> and <i>carB</i>	71
FIGURE 14.	Proposed transcription terminators downstream of <i>carA</i> and <i>carB</i> in <i>N. gonorrhoeae</i> CH811.....	74
FIGURE 15.	Alignment of the four copies of RS6 and the three copies of RS7 found in the sequence between <i>carA</i> and <i>carB</i> of <i>N. gonorrhoeae</i> CH811.....	78
FIGURE 16.	Northern blot analysis of <i>N. gonorrhoeae</i> CH811 total RNA, using probes comprising cloned gonococcal <i>carA</i> and <i>carB</i> genes.....	80
FIGURE 17.	Electrophoresis and Southern blot analysis of PCR amplicons of the intergenic sequence from gonococcal isolates of varying A/S/P class.....	82
FIGURE 18.	Electrophoresis and Southern blot analysis of PCR amplicons of the sequence between <i>carA</i> and <i>carB</i> from 17 commensal <i>Neisseria</i> species isolates comprising 8 species.....	87
FIGURE 19.	Hybridization of the RS6/RS7 repeats with amplicons of the sequence between <i>carA</i> and <i>carB</i> from a selection of <i>Neisseria</i> species isolates.....	89
FIGURE 20.	Composite of all the different REA patterns of the intergenic sequence identified from a total of 60 <i>Neisseria</i> species isolates.....	92
FIGURE 21.	REA of amplicons of the intergenic sequence from two groups of isolates with similar A/S class as isolates associated with two outbreaks of gonorrhea.....	97
FIGURE 22.	PCR amplicons of subregions of the intergenic sequence between <i>carA</i> and <i>carB</i> from selected gonococcal isolates which have varying intergenic sequence size.....	100
FIGURE 23.	PCR amplification of <i>carA</i> and <i>carB</i> from 11 gonococcal isolates.....	103

FIGURE 24.	Phylogenetic tree constructed from alignments of deduced amino acid sequences of complete CPS genes, using the Fitch distance matrix method.....	109
FIGURE 25.	Phylogenetic tree constructed from alignments of the deduced protein sequence of the the first and second thirds of the synthetase domain of CPS's, using the Neighbor-joining distance matrix method.....	111
FIGURE 26.	A rooted phylogenetic tree, constucted using the method of maximum parsimony, with schematic representations of the organization of the CPS genes shown beside each organisms name.....	114
FIGURE A1.	Alignment of all currently known CPS sequences of the amidotransferase domain.....	166
FIGURE A2.	Alignment of all currently known CPS sequences of the synthetase domain.....	167
FIGURE A3.	Phylogenetic tree constructed from alignments of deduced amino acid sequences of amidotransferase domains of CPS, using the Neighbour-joining distance matrix method.....	168
FIGURE A4.	Phylogenetic tree constructed from alignments of deduced amino acid sequences of amidotransferase domains of CPS, using the method of maximum parsimony.....	169
FIGURE A5.	Phylogenetic tree constructed from alignments of deduced amino acid sequences of synthetase domains of CPS, using the Neighbour-joining distance matrix method.....	170
FIGURE A6.	Phylogenetic tree constructed from alignments of deduced amino acid sequences of synthetase genes/domains of CPS, using the method of maximum parsimony.....	171

LIST OF ABBREVIATIONS

A/S/P	auxotype/serotype/plasmid content
ATC	aspartate transcarbamylase
ATP	adenosine triphosphate
C	citrulline-requiring
cfu	colony forming units
CMP-NANA	CMP-N-acetylneuraminic acid
CPS	carbamoylphosphate synthetase
CPS (Arg)	carbamoylphosphate synthetase (arginine-specific)
CPS (Pyr)	carbamoylphosphate synthetase (pyrimidine-specific)
C-terminal	carboxyl-terminal
CTP	cytidine triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dNTPs	deoxynucleotides triphosphate
DEPC	diethylpyrocarbonate
DHO	dihydroorotase
DNA	deoxyribonucleic acid
ET	electrophoretic type
GCMB	GC medium base
GMP	guanosine monophosphate
GTP	guanosine triphosphate

H	hypoxanthine-requiring
IMP	inosine monophosphate
LB	Luteina Bertani
LOS	lipooligosaccharide
MOPS	200mM 4-morpholinopropanesulfonic acid, 50mM sodium acetate, 10mM EDTA
N-terminal	amino-terminal
NR	non-requiring
O	ornithine-requiring
OAT	ornithine acetyltransferase
ORF	open reading frame
OTC	ornithine transcarbamylase
P	proline-requiring
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
PID	pelvic inflammatory disease
PPRP	5'-phospho- α -D-ribosyl 1-pyrophosphate
REA	restriction enzyme analysis
RNA	ribonucleic acid
SDS	sodium dodecylsulfate
10xSSC	3 M sodium chloride, 0.3 M sodium citrate
STD	sexually transmitted disease
1xTAE	0.04 M Tris-acetate, 0.001 M EDTA

TSA	tryptic soy agar
U	uracil-requiring
UMP	uridine monophosphate
X-gal	5-bromo-4-chloro-3-indoyl- β -D-galactoside

1. INTRODUCTION

Neisseria gonorrhoeae is an important pathogen for which no vaccine has yet been developed. Research in gonococcal genetics has therefore been extensive, revealing many interesting mechanisms of genetic variation, however these studies have focused primarily on the study of the genes encoding pili and surface proteins. Few studies of gonococcal housekeeping genes have been reported. Carbamoylphosphate synthetase (CPS), which catalyses the first committed step in both the arginine and pyrimidine biosynthetic pathways, has been studied in other organisms as a model for gene and enzyme regulation, and so represents a good starting point for studies of gonococcal housekeeping genes. It is also often mutated in the gonococcus, a feature which some have suggested confers a benefit to this organism's survival. Just why or how this common CPS deficiency is occurring is as yet unclear.

1.1 THE NEISSERIACEAE AND *N. GONORRHOEAE*: OVERVIEW AND IMPORTANCE IN DISEASE

The Neisseriaceae comprise a group of Gram-negative bacteria (Proteobacteria) of the β -subdivision which are usually diplococcal in morphology (Vedros, 1984). The majority of *Neisseria* species are non-pathogenic and are frequently found in human nasal passages, however within this genus are also found the pathogens *Neisseria meningitidis* and *N. gonorrhoeae*. Both of these pathogens are noted for their more complex nutritional requirements relative to the commensal *Neisseria* species (Catlin, 1973; Vedros, 1984). They have also been found to be very closely related (over 98% similarity between housekeeping gene sequences from the two organisms; Zhou and Spratt,

1992) and are each others closest relative based on phylogenies determined from electrophoretic typing, DNA hybridisation, and rRNA sequence analysis (Chun *et al.*, 1985; Hoke and Vedros, 1982; Wolff *et al.*, 1993). They are both obligate human pathogens and share many structural features including pili (McGee and Stevens, 1984). The meningococcus is however notably different in having a capsule, which is not present on the gonococcus or any other commensal *Neisseria* species (Chen *et al.*, 1989).

Meningococci are considered endemic, with approximately 15% of the world population as carriers and 90% antibody positive. They are noted for causing asymptomatic infection in the nasopharyngeal environment, however, under certain unknown conditions they can rapidly penetrate the basal membrane and blood/brain barrier, causing life-threatening meningococcal meningitis or septicemia (DeVoe, 1982). On the other hand, the majority of gonococcal infections remain limited to the mucosa of the urogenital tract, and they rarely cross into the bloodstream (Hook and Holmes, 1985). *N. gonorrhoeae* much more frequently causes disease, though asymptomatic infections are common in women and are thought to represent an important reservoir of infection (Biro *et al.*, 1995; Hook and Holmes, 1985).

Though gonococcal infections are treatable, *N. gonorrhoeae* is still a significant health concern. It can cause a spectrum of disease of varying severity, ranging from asymptomatic infection to gonococcal arthritis to rare cases of bacteremia and meningitis (Kerle *et al.*, 1992). Complications of gonococcal disease are disproportionately borne by women; 15-30% of those with uncomplicated urogenital infections may develop pelvic inflammatory disease (PID) which can cause serious sequelae such as infertility and ectopic pregnancy (Padian and Washington, 1994). The total health care costs of all PID infections in Canada alone are estimated to be \$140 million

annually; gonococcal infections represent a substantial portion of these costs (World Health Organization, 1995). Gonococcal infections, primarily of urogenital origin, are the fourth most prevalent disease reported to Health Canada (Health Canada, 1993). The World Health Organization (WHO) estimates that the global burden of new cases of gonorrhea is 62 million for (World Health Organization, 1995). The WHO also estimates that treating 100 women with gonorrhea successfully will avert 25 cases of PID; similarly, the prevention or cure of 100 cases of gonorrhea is predicted to prevent 425 cases of HIV in core groups (World Health Organization, 1995). Gonorrhea can facilitate the transmission of HIV (Wasserheit, 1992). The growing resistance of *N. gonorrhoeae* to antibiotics is also a problem. The incidence of gonococci with high-level resistance to penicillin and/or tetracycline is so high that most countries, including Canada, no longer use these drugs for treatment (Dillon and Yeung, 1989; Easmon, 1990; Health Canada, 1991; Whittington and Knapp, 1988). Resistance is also now emerging against commonly used quinolone antibiotics (Centers for Disease Control and Prevention, 1994). As mentioned previously, a vaccine has yet to be successfully developed. All of these concerns have been the driving factor stimulating the study of the gonococcus and its genetics.

1.2 THE GONOCOCCUS: HETEROGENEITY AND GENE STRUCTURE

Studies of gonococcal genetics have revealed a complex array of mechanisms which this bacterium uses to manipulate its genetic information (Meyer, 1990 and Seifert and So, 1991). The gonococcus has been found to be highly heterogenous, such that it has been classified as a non-clonal bacterium (O'Rourke and Spratt, 1994; O'Rourke and Stevens, 1993).

The genetic variation characteristic of gonococci is thought in part due to its natural competence for DNA uptake and transformation, when pilated (Sparling, 1966). Transformation frequencies of as high as 10^{-3} are frequently observed (Sparling, 1966). Competent gonococci specifically take up linear DNA containing a 10 bp sequence 5'-GCCGTCTGAA-3', commonly called the gonococcal or neisserial uptake sequence (Elkins *et al.*, 1991; Goodman & Scocca, 1988). This sequence is found frequently within the gonococcal genome (approximately once every 1 kb; Goodman & Scocca, 1991), and is also present in the genomes of other *Neisseria* species (Dougherty *et al.*, 1979). This therefore results in a genus-specificity for the DNA which can transform the gonococcus (Dougherty *et al.*, 1979; Goodman & Scocca, 1991). Other bacteria, such as species of *Bacillus*, *Streptococcus*, and *Haemophilus*, also show a similar natural competence for transformation (Stewart, 1986), and some, such as *Haemophilus* species, have been shown to contain an analogous DNA uptake sequence, though the consensus sequence is different from the gonococcal/neisserial sequence (Kroll *et al.*, 1992; Smith *et al.*, 1995).

A reflection of this high propensity for transformation, gonococcal genes are often mosaic in structure, containing regions of DNA with higher similarity to certain *Neisseria* species (Halter *et al.*, 1989; Zhou and Spratt, 1992). Numerous cases of horizontal transfer between the *Neisseria* species have been documented (Kroll *et al.*, 1992; Smith *et al.*, 1995).

Gonococci are also noted for the high frequency of antigenic and phase variation which occurs in many of its surface antigens, most of which are required for host cell attachment and invasion (Seifert and So, 1991; Swanson *et al.*, 1992). The mechanisms for this heterogeneity have been determined in many cases. For example, antigenic variation in pili occurs when a variant pilin sequence from a silent locus recombines into an expression locus by a predominantly unidirectional, homologous

recombination which can be intrachromosomal or transformation mediated (Hagblom *et al.*, 1985; Segal *et al.*, 1985). Variation in pilin expression can also occur by the on/off switch in expression of PilC, a pilus biogenesis protein which has a stretch of G residues within the signal peptide-encoding region which can undergo slip-strand mispairing (Jonsson *et al.*, 1991). Interestingly, both PilC and pilin have been found to be essential for transformation competence in *N. gonorrhoeae* and are thought to possibly play a role in DNA recognition and/or outer membrane translocation (Rudel. *et al.*, 1995).

As studies of gonococcal genetics accumulate, trends in gonococcal gene organisation and structure have been revealed. For one, there are few operonic gene structures found in the gonococcus (Lawson *et al.*, 1995; West and Clark, 1989; this work). Genes involved in similar pathways which are co-ordinately regulated are often located in different regions of the chromosome (Dempsey and Cannon, 1994). Another observation has been that gonococcal uptake sequences are commonly found as part of inverted repeats which flank gonococcal genes (Elkins *et al.*, 1991; Goodman & Scoocca, 1988). These inverted repeats have been recently been shown to act as rho-independent transcription terminators in *Escherichia coli* (Barber *et al.*, 1994). Regulation of gonococcal genes is only now beginning to be adequately studied. No review of promoters involved in gonococcal gene expression has yet been reported, however, many of the consensus sequences for promoters commonly observed in other Proteobacteria are found upstream of transcription start sites for gonococcal genes. For example, promoters of the σ -70 type, with and without the -35 consensus sequence, have been identified upstream of *pilA*, *pilB*, and *pilE* genes (Meyer *et al.*, 1984; Taha *et al.*, 1988) and σ -54 type consensus sequences have been identified upstream of *pilC* loci (Jonsson *et al.*, 1995). Gearbox promoter sequences, thought to regulate genes active during the stationary phase of growth, have been identified upstream of surface protein-encoding genes expressed during anaerobic growth and also the

uvrB gene involved DNA repair (Black *et al.*, 1995; Hoehn and Clark, 1992). In some cases, for example in the expression of the *pilE* gene, promoter expression has been shown to be different in gonococcal backgrounds verses *E. coli* or other bacteria, a cause for concern for regulatory studies performed outside of the gonococcal environment (Fyfe *et al.*, 1995).

Of the few housekeeping genes which have been studied in the gonococcus, most have many of the features typical of gonococcal gene structure. Namely, the presence of gonococcal uptake sequences as part of inverted repeats flanking the genes, a lack of operonic gene organisation, and a mosaic gene composition. However, as previously mentioned, few studies of gonococcal housekeeping genes have been performed. The only amino acid biosynthesis genes which have been sequenced from the gonococcus to date are the *argF*, and *argJ* genes of the arginine biosynthetic pathway (Martin *et al.*, 1990; Martin and Mulks, 1992b). Preliminary studies of the proline and aromatic amino acid biosynthetic pathways in the gonococcus have been reported, however this work did not involve the use of any DNA sequence data Berry *et al.*, 1987; Stein *et al.*, 1984; Subramaniam *et al.*, 1994). Studies of the tricarboxylic acid cycle, glucose metabolism, and respiratory chain of *N. gonorrhoeae* have been reported, but again these studies did not include the study of any DNA sequences (Hebeler and Morse, 1976; Morse *et al.*, 1974; Winter and Morse, 1975). A glutamine synthetase gene has been reported in GenBank, however, the only report of its use has been in studies of gonococcal heterogeneity through restriction enzyme digestion of the gene's sequence (O'Rourke and Spratt, 1994). No significant regulatory studies have been performed on any gonococcal housekeeping genes.

1.3 NATURALLY OCCURRING ARGININE-REQUIRING GONOCOCCI

Initial interest in the metabolic genes of the gonococcus stemmed from the high natural occurrence of auxotrophic strains in clinical isolates of *N. gonorrhoeae* (Catlin, 1973). There have been some indications that this auxotrophy may be advantageous to the gonococcus: Kenyon (1978) found that a gonococcal proline-requiring strain passaged through mice gained genetic defects resulting in arginine and uracil auxotrophy. In human populations around the world, the prevalence of specific auxotypes changes, indicating selective pressures may play a role (Dillon *et al.*, 1990; Dillon and Carballo, 1990; Dillon and Pauzé, 1984). Analysis of *N. gonorrhoeae* from the pre-penicillin era has revealed that arginine- and uracil-requiring strains first appeared after penicillin was introduced (Catlin and Reyn, 1982). It has been proposed that the resulting slower growth of strains of certain auxotypes enables them to evade the penicillin (Mayer *et al.*, 1977). Another hypothesis has been that these slower-growing isolates could be causing lower grade infections and therefore causing patients to delay seeking treatment (Whittington *et al.*, 1993).

There have also been indications that auxotrophy is beneficial, or at least selected for, in other species of bacteria when pathogenic. In one study, 18% of *Pseudomonas aeruginosa* isolates from cystic fibrosis patients, or patients with bronchiectasis, had requirements for methionine, proline, arginine, thiamine or multiple requirements, while *P. aeruginosa* isolates from other sources showed no such auxotrophy (Taylor *et al.*, 1992). *Staphylococcus* species which are niche-specific and/or host-specific often require arginine for growth, while those showing a wide host range have no such auxotrophy (Emmett and Kloos, 1979). Many other species of bacteria, including *N. meningitidis*, *Lactobacillus* species, *Shigella* species, *Campylobacter* species, and *Streptococcus* species, also have

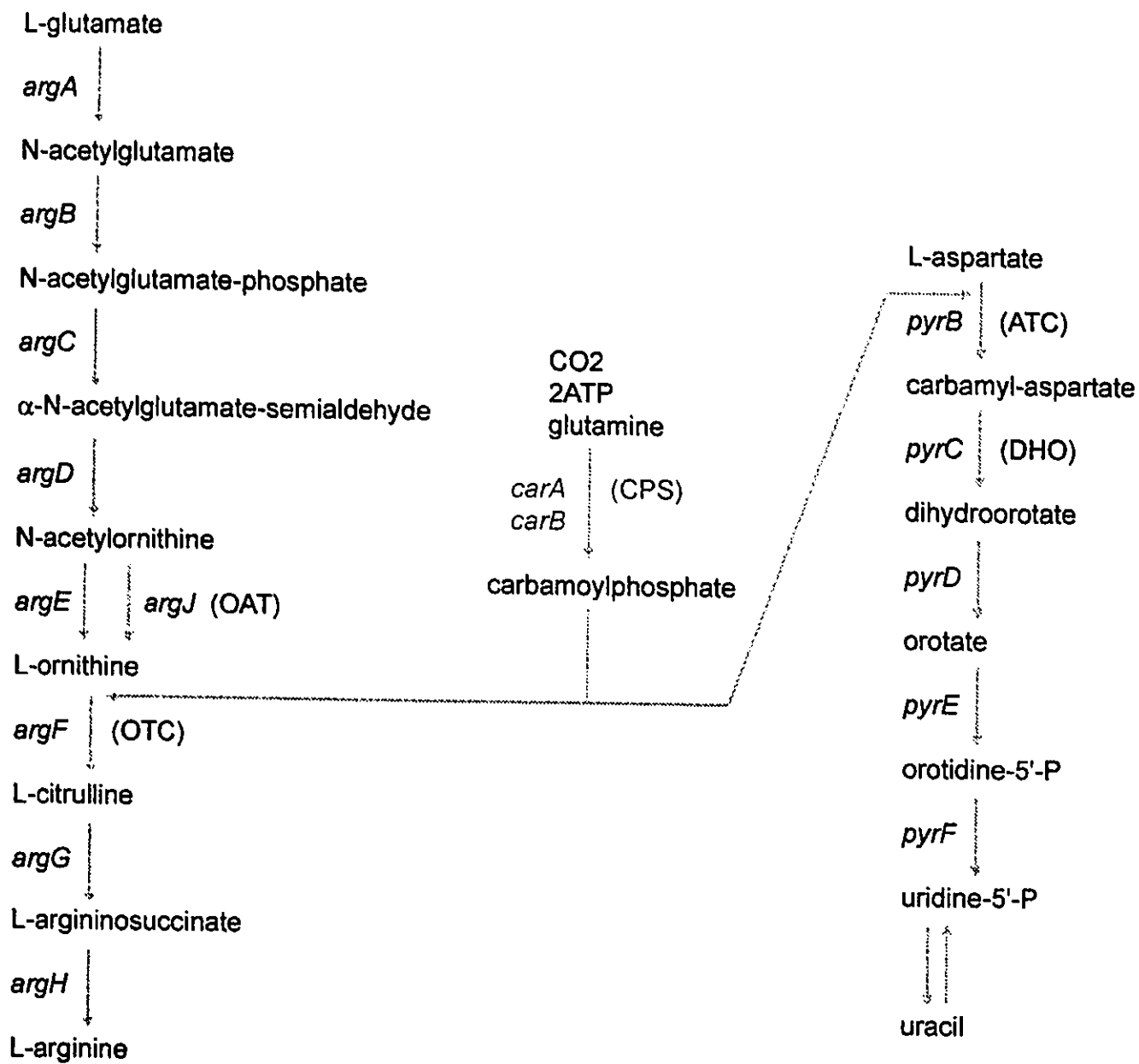
naturally occurring nutrient requirements (Ahmed *et al.*, 1988; Bringel, 1994, Tenover and Patton, 1987).

Gonococci are relatively unique, however, in that their varying nutrient requirements are monitored frequently through a very widely used typing scheme, auxotyping, which is used to type gonococcal isolates (Carifo and Catlin, 1973; Hendry and Stewart, 1979). Auxotypes most frequently observed with *N. gonorrhoeae* are non-requiring (NR), proline-requiring (P), proline, citrulline, and uracil-requiring (PCU), ornithine, uracil and hypoxanthine-requiring (OUH), and citrulline, uracil and hypoxanthine-requiring (CUH; Dillon *et al.*, 1990). PCU isolates are particularly prevalent in Canada and have been found to be relatively clonal in comparison to other gonococcal isolates (Dillon *et al.*, 1990; Ng and Dillon, 1993). CUH and OUH isolates have been of notable interest because of their association with cases of disseminated gonococcal infection in population studies (Brunham *et al.*, 1985). Since citrulline and ornithine are precursors of arginine (Figure 1), these isolates of auxotype PCU, CUH, and OUH all require arginine and are sometimes referred to as PAU and AUH isolates.

In total, approximately 40-60% of clinical isolates of *N. gonorrhoeae* are auxotrophic for arginine (Catlin and Nash, 1978; Dillon *et al.*, 1990; Picard and Dillon, 1989a). Further biochemical characterisation of these arginine-requiring strains has shown that the vast majority of them have gene defects involving ornithine acetyltransferase (OAT; ~55%), or carbamoylphosphate synthetase (CPS; ~45%) and that a deficiency in ornithine transcarbamylase (OTC) is extremely rare (Figure 1; Picard and Dillon, 1989a; Shinnars and Catlin, 1982; Li, 1993). The deficiency in OAT results in an ornithine-requirement, while the CPS deficiency results in a dual requirement for citrulline and uracil (Figure 1).

OAT deficient isolates have been further characterised at the molecular level by Martin and Mulks (1992a). In strains requiring ornithine, uracil, and hypoxanthine, a defect in the gene encoding

FIGURE 1. The arginine and pyrimidine biosynthetic pathways. Carbamoylphosphate synthetase (CPSase) and the genes which encode its subunits (*carA* and *carB*) are highlighted in red. Other enzymes frequently referred to, namely ornithine acetyltransferase (OAT), ornithine transcarbamylase (OTC), aspartate transcarbamylase (ATC), and dihydroorotase (DHO), are also labeled.



OAT (*argJ*) was identified, which is the result of one clonally derived mutation. This defect, a 3 bp deletion, results in the loss of a seemingly crucial alanine residue (Martin and Mulks, 1992a). Other, less common, *argJ* defective strains seem to have a different mutation which has not been fully characterised (Martin and Mulks, 1992a).

Preliminary studies using genetic complementation have revealed that there are likely between two and four different mutations causing CPS deficiencies (Copley, 1987; Mayer *et al.*, 1977). Unfortunately, the exact number of different mutations is hard to determine since these studies did not adequately differentiate between arginine-requirements satisfied by ornithine (possible OAT defect), and arginine-requirements not satisfied by ornithine (possible CPS defect).

1.4 GENETIC STUDIES OF ARGININE BIOSYNTHESIS AND CPS IN THE GONOCOCCUS

Though initial interest in the arginine biosynthetic pathway of the gonococcus was due to the appearance of naturally occurring mutants, studies of the organisation and regulation of the arginine biosynthesis genes would give the first detailed knowledge of a housekeeping gene family in the gonococcus. Currently, seven arginine biosynthesis genes have been cloned from the gonococcus (Picard, 1990; Picard and Dillon, 1989b). Genes encoding OTC (*argF*) and OAT (*argJ*) have been sequenced (Martin *et al.*, 1990; Martin and Mulks, 1992b) and found to have typical gonococcal gene structure: they are separately transcribed and unlinked and are flanked by inverted repeats containing the gonococcal uptake sequence. *argF* has been shown to have a mosaic gene structure (Zhou and Spratt, 1992). The regulation of these arginine pathway genes has not been studied, however “arg-box” sequences, indicative of regulation by a repressor similar to that of *E. coli* (see next section) are not present upstream of either gene (Martin *et al.*, 1990; Martin and Mulks, 1992b).

Since CPS catalyses the first committed step for both the arginine and pyrimidine pathways, and is highly regulated in many organisms, this enzyme forms a good starting point for the study of the arginine biosynthetic pathway in the gonococcus. A single glutamine-dependent CPS enzyme was previously demonstrated in *N. gonorrhoeae* (Shinners and Catlin, 1982). Two experiments regarding CPS regulation have been reported, in which CPS levels were determined for prototrophic gonococci grown on defined media with and without arginine and uracil (Catlin and Nash, 1978; Shinners and Catlin, 1982). In the first study, repression of CPS activity was detected, however in the second study they deduced that no repression was in fact occurring. However, the media used was one commonly used for auxotyping (Catlin, 1973) which contains other amino acids which may be contaminated with arginine. No further studies have been performed to confirm whether repression of gonococcal CPS does occur, however, a study of CPS repression has been reported for the closely related species *N. meningitidis*, reporting incomplete repression (20% reduction in activity) of CPS by arginine and uracil (Jyssum, 1983). There have been no reports examining the allosteric effectors of gonococcal CPS. There have been no reports examining the size and structure of CPS, which in other bacteria is a heterodimer.

A study of the genes encoding CPS has been initiated in our laboratory. A phage clone (EFP10), complementing *E. coli* CPS mutants, was isolated (Picard and Dillon, 1989b). A 5 kb *EcoRI* fragment from EFP10 was found to hybridise at high stringency to a probe consisting of the *E. coli* CPS genes *carA* and *carB* (Picard, 1990).

1.5 OVERVIEW OF ARGININE BIOSYNTHESIS AND THE UREA/ORNITHINE CYCLE: COMPARISONS BETWEEN SPECIES

The arginine biosynthetic pathway has been studied extensively in many organisms. This is in part due to its historical significance: Genetic control of arginine biosynthesis was the earliest example of the one gene, one enzyme theory in *Neurospora crassa* (Srb and Horowitz, 1944). The term “repression” was first coined soon after studies of the arginine pathway began in *E. coli* (Vogel, 1957). The term “regulon” was also developed through arginine pathway studies when the existence of a repressor was first discovered (Maas and Clark, 1964). Studies of this pathway have continued at a steady pace as the complexity of the regulatory mechanisms unfolded and today this pathway has become a paradigm for studies of basic enzyme function and gene regulation.

Arginine biosynthesis proceeds from glutamate in eight enzymatic steps (Figure 1; Cunin *et al.*, 1986). The first five steps involve N-acetylated intermediates leading to the production of ornithine. Ornithine is then converted to arginine in the final three steps, utilising carbamoylphosphate which is synthesised by CPS in a step also critical for pyrimidine production (Figure 1).

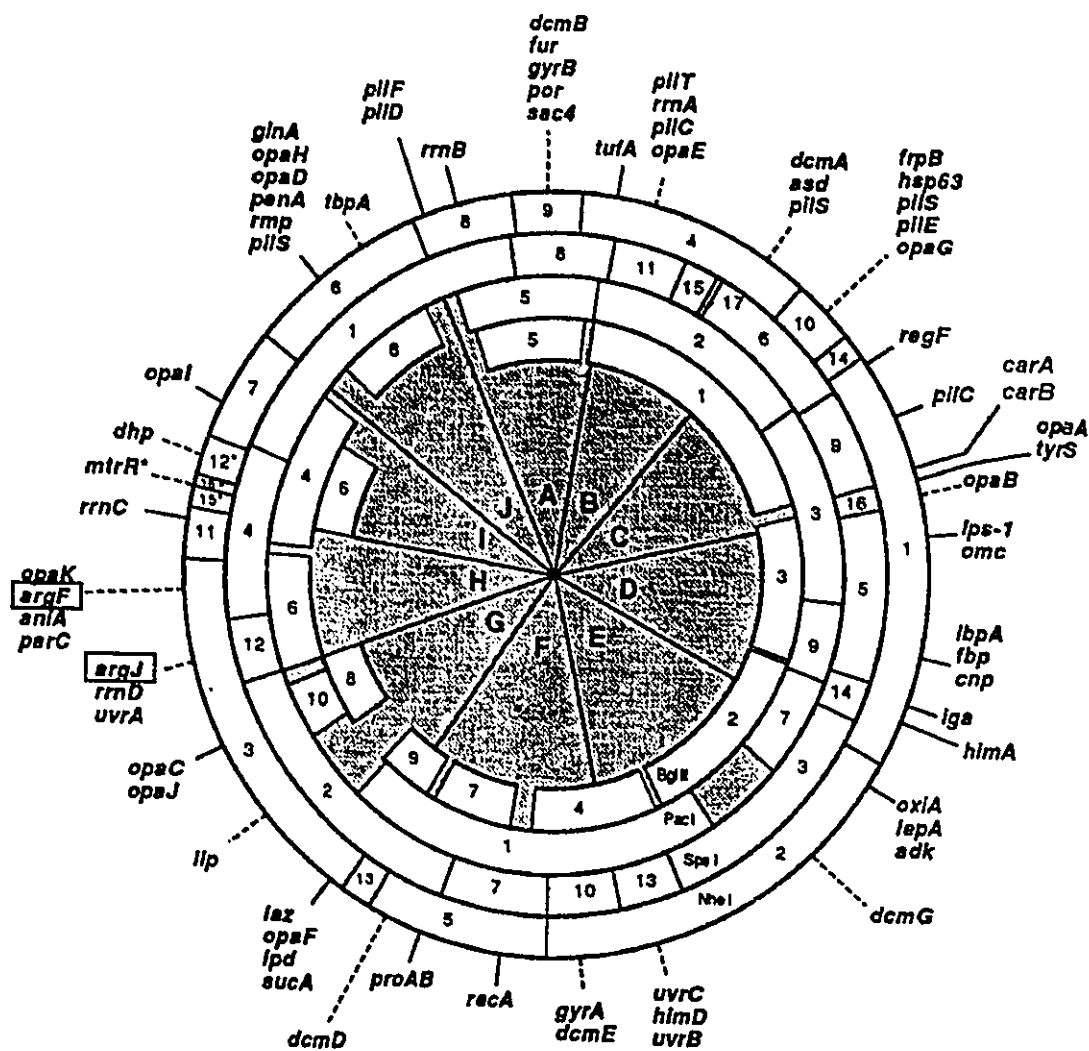
The enzymatic steps involved in the arginine biosynthetic pathway are similar in most bacteria, with the exception of the fifth step which deacetylates acetylornithine to form ornithine. In Enterobacteriaceae the pathway proceeds in a linear fashion, with the fifth step catalysed by the hydrolytic enzyme acetylornithinase (encoded by *argE*; Udaka, 1966). In these bacteria the enzyme catalysing the first step of the pathway, encoded by *argA*, is subject to feedback inhibition. In contrast, a variety of other bacteria including pseudomonads, *Bacillus spp.*, *Micrococcus spp.*, cyanobacteria, and *Thermus spp.*, recycle the acetyl group back to glutamate through a transacetylation reaction catalysed by OAT (encoded by *argJ*; Cunin *et al.*, 1986; Udaka and Kinoshita, 1958). In these species

the enzyme catalysing the second step of the pathway, encoded by *argB*, often undergoes feedback inhibition. *N. gonorrhoeae* has also been shown to recycle the acetyl group in this way (Picard and Dillon, 1989b; Shinnars and Catlin, 1978).

Chromosomal organisation of the arginine biosynthesis genes varies significantly in different genera of Bacteria. In *Bacillus subtilis*, the first six genes involved in the formation of citrulline are clustered together in an operon, while the other two, involved in the formation of arginine from citrulline, form a separate operon (Mountain *et al.*, 1984). In *Escherichia coli*, these genes are organised into four operons, including an *argECBH* cluster which contains two divergent promoters for *argCBH* and *argE* (Bachmann, 1983). In *Pseudomonas aeruginosa*, the arginine biosynthesis genes are scattered around the chromosome (Haas *et al.*, 1977). *N. gonorrhoeae* seems most similar to the pseudomonads, since both co-complementation and chromosome mapping experiments indicate that most of the arginine biosynthesis genes in the gonococcus are unlinked (Figure 2; Bihlmaier *et al.*, 1991; Dempsey and Cannon, 1994; Picard and Dillon, 1989b).

The regulation of arginine biosynthesis genes has also been found to vary significantly between different bacterial genera. *E. coli*, *Salmonella typhimurium*, *B. subtilis*, and *P. aeruginosa* have been the most well studied. In *E. coli* and *S. typhimurium*, the genes involved in this pathway are subject to repression by arginine through the repressor ArgR (Lim *et al.*, 1987; Lu *et al.*, 1992; Maas, 1995). This repressor binds to two “arg boxes”, two 18 bp operator sequences found in the promoter regions upstream of *argECBH*, *argD*, *argF*, and *carAB*. A copy is also found upstream of the gene encoding the repressor itself, *argR*. The arginine repressor is unique in that it is a hexamer and not a dimer like other characterised repressors (Lim *et al.*, 1987; Lu *et al.*, 1992). This difference is thought to explain why this repressor usually binds two operator sequences rather than one (Maas, 1995).

FIGURE 2. Physical map of the chromosome of *N. gonorrhoeae* FA1090 (obtained from Dempsey and Cannon, 1994). The location of the *carA* and *carB* genes is highlighted in red, and other genes involved in arginine biosynthesis are boxed.



In *B. subtilis*, a similar repressor is found and sequences analogous to the “arg-boxes” are present upstream of the main arginine biosynthesis operon (North *et al.*, 1989; Smith *et al.*, 1989). This repressor can regulate the *E. coli* *arg* genes, however, curiously the *E. coli* repressor cannot repress the *B. subtilis* genes (Smith *et al.*, 1989). *Bacillus* species are relatively unique in that they have two CPS enzymes, one repressible by arginine and encoded within the cluster of genes involved in arginine biosynthesis, and the other repressible by uracil and encoded within an operon of pyrimidine biosynthetic genes (Ghim *et al.*, 1994; Mountain *et al.*, 1984; Paulus and Switzer, 1979; Quinn *et al.*, 1991). All other bacteria studied to date contain one CPS enzyme utilised in both pathways.

Neither the Enterobacteriaceae nor Bacillaceae show any evidence of attenuation for control of arginine pathway genes. However, in *P. aeruginosa* and *Pseudomonas stutzeri* attenuation seems to be an important factor in the control of *carAB* (Kwon *et al.*, 1994). The arginine pathways in these organisms show no indications of a repressor analogous to the enterics and no arg-box sequences are found upstream of any genes (Wong and Abdelal, 1990). The focus of studies of the arginine biosynthesis pathway in *P. aeruginosa* has been primarily on the CPS enzyme. This is due not only to CPS’s entry level position in the pathway, but because *P. aeruginosa* also contains a carbamate kinase enzyme which can degrade carbamoylphosphate, essentially performing the reverse reaction of CPS (Abdelal *et al.*, 1982). Questions regarding the organism’s ability to avoid a futile cycle have been the primary reason for the study of CPS enzyme in this organism (Abdelal *et al.*, 1982; Kwon *et al.*, 1994).

De novo arginine biosynthesis can also occur in some eukaryotes, however the number of different organisms with this ability, and the distribution of such organisms throughout the eukaryotes, is notably lacking in the literature. The arginine biosynthesis pathway has, however, been extensively studied in the two fungi *Saccharomyces cerevisiae* and *Neurospora crassa* (Davis, 1986). In these

organisms, two CPS enzymes are present as in the Bacillaceae, one for the arginine biosynthetic pathway and the other for pyrimidine biosynthesis. Compartmentalisation of the pathway plays a role and actually differs between the two eukaryotes: In *S. cerevisiae* the steps leading to the production of ornithine occur in the mitochondrion, while biosynthesis of arginine from ornithine proceeds in the cytosol. In *N. crassa* the steps leading to the production of citrulline are mitochondrial and the last two steps are cytosolic. This is important in the context of the CPS enzyme because in both cases the pyrimidine-specific enzyme is cytosolic. This difference in compartmentalisation of the arginine pathway enzymes therefore results in *S. cerevisiae* having both its arginine- and pyrimidine-specific CPS enzymes in the cytosol, while in *N. crassa* the two CPS enzymes are separately compartmentalized. Stark differences in the regulation of this pathway are seen between the two organisms, in part a reflection of these compartmentalisation differences; In *S. cerevisiae*, a general amino acid control system is involved, along with an arginine-specific repression system mediated by at least three ARGR proteins (Béchet *et al.*, 1970; De Rijcke *et al.*, 1992; Qiu *et al.*, 1990). In *N. crassa* compartmentalisation and enzyme inhibition play more prominent roles, and no ARGR repression system is present (Davis, 1986).

In the vertebrates and some other eukaryotes, the first 5 enzymes of the arginine biosynthetic pathway leading to the synthesis of ornithine have been lost and these organisms cannot synthesise arginine *de novo*. However, these organisms all produce large amounts of the enzyme arginase which cleaves urea from arginine and regenerates ornithine. This forms the basis of the urea/ornithine cycle which enables the organism to convert ammonia to urea. Unlike studies of arginine biosynthesis in eukaryotes, the distribution of the urea cycle in organisms is relatively well studied (Mommensen and Walsh, 1989; Morris, 1992; Paulus, 1983). This cycle seems to have developed early in evolution since

a functional ornithine cycle has been identified in the primitive flatworm *Bipalium kewense* (Campbell, 1965). The function of this cycle varies significantly in different organisms: In marine elasmobranchs (sharks, skates and rays) and the coelacanth, urea is used as an osmolyte, whereas in lungfishes and amphibians urea is synthesised to detoxify ammonia during periods of water stress (Mommsen and Walsh, 1989). Most mammals synthesise urea to detoxify ammonia, though some hibernating mammals and ruminants use urea to recycle nitrogen between the liver and the gut (Mommsen and Walsh, 1989). Like the fungi described earlier, the vertebrates studied all have two CPS enzymes, one specific for pyrimidine biosynthesis and the other specific for the urea cycle. The urea cycle-specific CPS is mitochondrial in all cases and the pyrimidine-specific CPS is cytosolic (Anderson, 1980; Lusty, 1978). Arginase, however, changes in location, being mitochondrial in the elasmobranchs and being cytosolic in the amphibians and mammals (Anderson, 1980; Lusty, 1978). This change in location of arginase correlates with a change in the properties of CPS (from being glutamine dependent to ammonia dependent - discussed further below) and with a change in function of the urea cycle from that of osmolyte production to ammonia detoxification.

1.6 OVERVIEW OF PYRIMIDINE BIOSYNTHESIS

The CPS enzyme is not only critical for arginine biosynthesis, but also forms an entry level step in the biosynthesis of pyrimidines (Figure 1). Though gonococcal isolates with uracil auxotrophy are frequently observed (an attribute tested for in auxotyping), no studies of pyrimidine biosynthesis have yet been performed in *N. gonorrhoeae*. Studies performed in other organisms, however, have yielded some interesting results regarding gene regulation and enzyme evolution.

De novo synthesis of UMP, the precursor of all pyrimidines, is accomplished in six enzymatic

steps (Figure 1). Though these same six reactions occur in all organisms studied, the number of genes encoding these functions decreases from prokaryotes to eukaryotes, as some genes became fused over the course of evolution (Davidson *et al.*, 1993). These fusions occurred within the first three enzymes of the pathway, and so these three enzymes have been the most studied: carbamoylphosphate synthetase (CPS), aspartate transcarbamylase (ATC), and dihydroorotase (DHO). The regulation and gene organisation within the bacteria also varies as discussed below.

The genetics of *E. coli* pyrimidine biosynthesis has been the most extensively studied, and its six enzymes are encoded by six unlinked loci which are non-coordinately regulated by the levels of various nucleotides pools (Neuhard and Nygaard, 1987). In this organism ATC is a dodecamer composed of two different subunits, a catalytic subunit encoded by *pyrB*, and regulatory subunit encoded by *pyrI* (Pauza *et al.*, 1982). These genes comprise an operon, *pyrBI*, which undergoes regulation through a well studied UTP-sensitive attenuation mechanism and also through an as yet unidentified attenuator-independent mechanism (Liu and Turnbough, 1989). Dihydroorotase is encoded by *pyrC*, a gene regulated by the ratio of intracellular concentrations of CTP and GTP through a novel mechanism involving 2 bp changes in the location of initiation of transcription resulting in varying hairpin formation and translational initiation (Wilson *et al.*, 1992). *pyrD* and *pyrE* are regulated by other attenuation mechanisms, and the mechanisms controlling *pyrF* have yet to be elucidated (Bonekamp *et al.*, 1984). The regulation and organisation of pyrimidine biosynthetic genes in *S. typhimurium* is similar (Neuhard and Nygaard, 1987).

Pseudomonas species also have unlinked loci for pyrimidine biosynthesis, however the ATC and CPS which have been studied show significant differences from their enteric counterparts. ATC has been shown to be a dodecamer composed of two different subunits which differ significantly from

those of enterics (Bergh and Evans, 1993). The *pyrB*-encoded subunit is bigger than its counterpart in *E. coli*, and it contains both the catalytic and regulatory activities of the enzyme (Schurr *et al.*, 1995). The second subunit is an inactive dihydroorotase which is transcribed from a gene which overlaps *pyrB* called *pyrC'* (Schurr *et al.*, 1995). This second subunit seems to have no enzymatic activity, yet seems to be required for the dodecameric assembly of the ATC enzyme (Schurr *et al.*, 1995). Another unlinked *pyrC* gene, encoding a functional dihydroorotase, is located elsewhere on the chromosome (Schurr *et al.*, 1995). The regulation of these genes has not yet been studied and little is known about the other pyrimidine biosynthetic genes. The CPS genes of *P. aeruginosa* and *P. stutzeri* undergo a different mode of regulation from the enterics, as described below in the section (Section 1.8) discussing CPS gene organisation and regulation.

In *B. subtilis*, these genes are organised in a gene cluster of the order *pyrR*, *pyrP*, *pyrB*, *pyrC*, *pyrAA*, *pyrAB*, *pyrD*, *pyrF*, and *pyrE*, where *pyrAA* and *pyrAB* encode the pyrimidine-regulated CPS (Quinn *et al.*, 1991) and *pyrR* and *pyrP* encode a regulatory protein and uracil permease, respectively (Turner *et al.*, 1994). These genes are co-ordinately regulated by uracil through a complex attenuation/repression mechanism involving the protein expressed from *pyrR*, three terminator/antiterminator structures, and the availability of pyrimidines (Turner *et al.*, 1994). In this mechanism, the PyrR protein binds to a particular sequences in the *pyr* mRNA and disrupts the antiterminator hair-pin formation, allowing transcription termination (Turner *et al.*, 1994). *B. caldolyticus* has the same gene organisation for its pyrimidine biosynthetic genes (Ghim *et al.*, 1994).

Most of the eukaryotes examined possess multifunctional enzymes for catalysis of the early steps of pyrimidine biosynthesis. These multifunctional complexes presumably arose as a result of ancient gene fusions (Jones, 1980). There are differences between fungi and other higher eukaryotes,

however some interesting similarities are observed: In the slime mold *Dictyostelium discoideum*, as well as *Drosophila melanogaster* and Syrian hamster, the first three steps of the pyrimidine pathway are encoded together in a trifunctional protein containing the enzymes in domains of the order NH₂-CPS-DHO-ATC-COOH (Davidson *et al.*, 1993; Jones, 1980). In *S. cerevisiae* and *N. crassa* the first two steps of the pathway (CPS and ATC) are encoded by one gene, *ura2*, which contains an inactive DHO-like domain between the CPS and ATC domains of the protein (Souciet *et al.*, 1989). A functional DHO enzyme is encoded by another gene, *ura4*, which is present on a separate chromosome (Gyonvarch *et al.*, 1988).

While most eukaryotes studied contain multifunctional proteins for catalysing early pyrimidine biosynthesis, exceptions are found within the apicomplexan protozoan's *Plasmodium falciparum*, *Plasmodium berghei*, and *Babesia bovis*. In these organisms the enzymatic activities for each step of the pathway can be separated and a monofunctional CPS gene has been characterised for both *P. falciparum* and *B. bovis* (Chansiri and Bagnara, 1995; Flores *et al.*, 1994). These protozoan's are, however, also different from the other eukaryotes studied in that they have only one CPS enzyme instead of two. The enzymes of the pyrimidine biosynthetic pathway, including CPS, are of particular interest to researchers studying these protozoan parasites, most notably the malarial parasite *P. falciparum*, because these protozoans can only synthesise pyrimidines *de novo* and do not have any salvage pathways (Sherman, 1979; Gero *et al.*, 1983). The enzymes involved in this pathway are therefore attractive targets for antiparasitic/antimalarial drugs (Krungkrai *et al.*, 1992; Seymour *et al.*, 1994).

While the organisation of the pyrimidine biosynthetic genes is well known in eukaryotes, transcriptional regulation of these genes has not been well studied. Genes encoding the multifunctional

proteins undergo regulation involving low level repression/derepression by pyrimidines by an as yet unknown mechanism (Brustilow *et al.*, 1989). The transcriptional regulation of pyrimidine genes in the protozoans has not yet been studied.

1.7 CARBAMOYLPHOSPHATE SYNTHETASE: ENZYME FUNCTION AND STRUCTURE

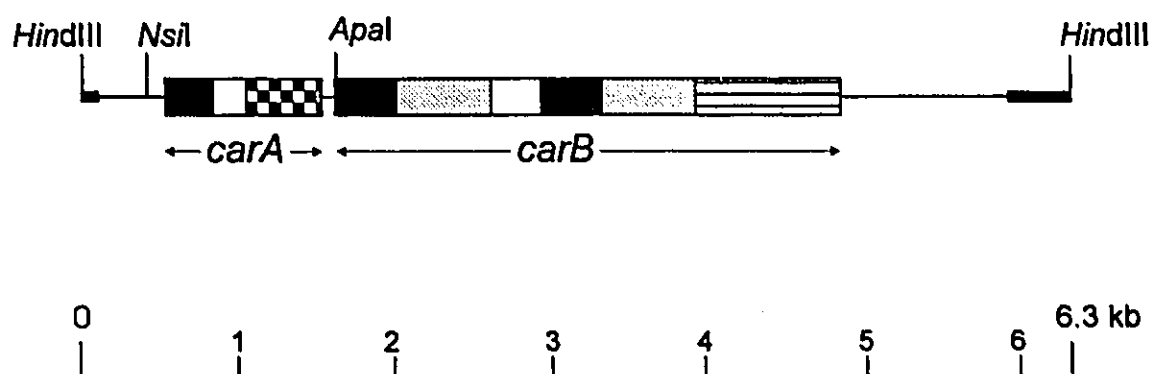
The majority of studies of arginine and pyrimidine biosynthesis have focused on the CPS enzyme. This enzyme is the first committed step for both pathways and so is often subject to multiple levels of regulation.

In bacteria, CPS catalyses the formation of carbamoylphosphate from CO₂, ATP, and glutamine. The enzyme is a heterodimer, consisting of a small subunit of approximately 40 kDa and a large subunit of approximately 120 kDa (Trotta *et al.*, 1974). The small, or amidotransferase subunit, binds glutamine and releases an amide group which is subsequently shuttled to the large subunit. The large, or synthetase subunit, then uses this amide group and bicarbonate in a complex reaction, involving two phosphorylations, to form a molecule of carbamoylphosphate (Anderson and Meister, 1966; Trotta *et al.*, 1974).

The structural organisation and functional domains of CPS have been most extensively studied in *E. coli*. In *E. coli* the genes encoding the small and large subunits are named *carA* and *carB*, respectively (Nyunoya and Lusty, 1983; Piette *et al.*, 1984). These genes are depicted in Figure 3 with the location of functional domains for the encoded protein, described below, mapped on these genes.

The small subunit, encoded by *carA*, contains what is commonly called the amidotransferase domain of CPS and is thought to have evolved from an ancestral amidotransferase (Werner *et al.*, 1985). The C-terminal end of this subunit contains features in common with other glutamine aminotransferases (*trpG*-type), including residues known to bind glutamine and/or be reactive

FIGURE 3. Schematic diagram of the 6.3 kb insert from the clone pMC50, which contains the *carAB* operon from *E. coli*, showing the location of functional domains in the resulting CPS protein. Note that these genes were originally isolated in the phage clone λ *dcarAB*37-9 and so some λ sequences (indicated by thick black lines) are present in this clone (*Hind*III restriction enzyme sites within the phage sequence were used to subclone the 6.3 kb fragment into pBR322 to produce pMC50; Piette *et al.*, 1984). Boxes denote the location of the *carA* and *carB* genes which are co-transcribed and separated by a small (17 bp) intervening sequence. Functional domains encoded by these genes are indicated as follows: glutaminase domain (checkered), domains of subunit interaction (solid), ATP-CO₂ domains (lightly shaded), and allosteric binding domain (striped). Restriction enzyme sites relevant to this work are noted.



in glutamine/amide transfer (Mullins *et al.*, 1991; Miran *et al.*, 1991). The most studied residue of this protein has been cysteine-269 (of the *E. coli* protein) which when substituted with a serine or glycine results in a CPS enzyme unable to utilise glutamine but able to utilise ammonia as a substrate (Rubino *et al.*, 1986). The N-terminal region of this subunit is thought to be involved in small subunit/large subunit interactions, since deletion of this region abolishes any interactions between subunits (Guillou *et al.*, 1989).

The large subunit (encoded by *carB*) contains what is commonly called the synthetase domain of CPS. This subunit/domain contains internal similarity between the first and second thirds of the protein (Nyunoya and Lusty, 1983). Since these similar regions both contain an ATP binding site (Kim *et al.*, 1991; Post *et al.*, 1990), they are thought to have evolved from an ancient duplication of an ancestral kinase gene (Nyunoya and Lusty, 1983). ATP and CO₂ binding regions have not yet been accurately mapped, and there has been some controversy regarding their exact location, but the general location is known (Figure 3; Guillou *et al.*, 1989) and glycine rich regions consistent with ATP binding sites have been noted in these two homologous regions of the protein (Post *et al.*, 1990). Interactions with the small subunit are thought to occur at the N-terminal of the protein and within the central region of this synthetase subunit (Guillou *et al.*, 1989; Figure 3). The terminal third of the protein has been defined as the regulatory region, where all allosteric effectors have been found to bind (Rubio *et al.*, 1991). Interestingly, while the activators and inhibitors of CPS vary significantly between organisms, they all bind in the same regulatory region, even in the eukaryotic enzymes (Rodriguez-Aparicio *et al.*, 1989).

All other studied bacteria have a similar CPS, based on sequence analysis and enzymatic activity studies. Even the two CPS enzymes identified in *Bacillus* species are similar in structure and

activity (Paulus and Switzer, 1979). The only major differences noted are the allosteric effectors which control each enzyme. The CPS of *E. coli* is activated by IMP (inosine monophosphate) and ornithine and inhibited by UMP (Meister and Powers, 1978), while the CPS of *P. aeruginosa* is activated by N-acetylmethionine and ornithine and feedback inhibited by UMP (Abdelal *et al.*, 1983). This difference is likely due to *P. aeruginosa*'s capacity for both catabolic and anabolic arginine pathways. In *B. subtilis*, the CPS encoded by genes found in the arginine biosynthetic operon is not subject to allosteric effectors, however the CPS encoded within the pyrimidine biosynthetic operon is severely inhibited by uridine nucleotides and activated by PRPP (5'-phospho- α -D-ribose 1-pyrophosphate) and GMP (Paulus and Switzer, 1979).

In most eukaryotes the arginine-specific CPS enzyme is a monomer, the result of a proposed ancient gene fusion (Nyunoya *et al.*, 1985). *S. cerevisiae* and *N. crassa* are the only currently known exceptions to this rule, they both have an arginine-specific CPS which is heterodimeric (Davis *et al.*, 1980; Lacroute *et al.*, 1965). The pyrimidine-specific CPS's of eukaryotes are usually part of multifunctional proteins as mentioned previously (Davidson *et al.*, 1993; Jones, 1980). The protozoan's, which do not have a multifunctional protein, have a monomeric CPS, however, this CPS is quite unusual in that it contains large translated insertions within functional domains of the protein (Chansiri and Bagnara, 1995; Flores *et al.*, 1994).

Variation in enzyme substrate specificity is observed within the arginine-specific CPS enzymes of eukaryotes, as this pathway evolved from that of arginine biosynthesis to ammonia detoxification. In eukaryotes such as *S. cerevisiae*, which can synthesise arginine *de novo*, the arginine-specific CPS is similar to bacterial CPS, however in the ureotelic organisms studied such as human and rat, the corresponding CPS has lost the ability to bind glutamine and instead uses ammonia as a primary

substrate (Marshall, 1976). This forms one of the critical pieces of the urea cycle since this CPS enzyme is now used to harvest ammonia for conversion into urea. In this CPS, commonly called CPSI, the reactive cysteine residue known to play a role in the glutamine binding/amide transfer is replaced with a serine (Nyumoya *et al.*, 1985). This was originally proposed to be the sole factor governing this enzymes use of ammonia over glutamine as a substrate. However, recently the CPSI gene from the bullfrog *Rana catesbeiana* was sequenced and the deduced protein sequence found to contain this reactive cysteine residue (Helbing and Atkinson, 1994). This would suggest that other residues have also played a role in the loss of glutamine utilisation. In ureosmotic organisms such as the spiny dogfish shark, *Squalus acanthias*, another CPS is used in the urea cycle called CPSIII. This enzyme uses glutamine as a substrate and contains the reactive cysteine residue, however it still seems closely related to CPSI in that its role is similar (*i.e.* it is part of the urea cycle) and that it is subject to the same activator (Hong *et al.*, 1994). This activator, N-acetylglutamate, is an absolute requirement for CPSI or CPSIII activity (Anderson, 1980). All the other CPS enzymes found in eukaryotes (and bacteria) have up until now been given the name CPSII. CPSII enzymes do not require N-acetylglutamate for activation.

Within the CPSII group, however, there are some differences in the presence of other activators and inhibitors, suggesting that this group is not so homogeneous, even within the eukaryotes. For example, in *S. cerevisiae*, the arginine-specific CPS has no allosteric effectors and is a heterodimeric protein, yet it is grouped with the same name (CPSII) as the pyrimidine specific CPS's of eukaryotes which are inhibited by UTP and often part of multifunctional proteins.

Clearly, there is a need to organise these different CPS enzymes to clarify relationships between the many different CPS's and to unravel the evolution of this enzyme.

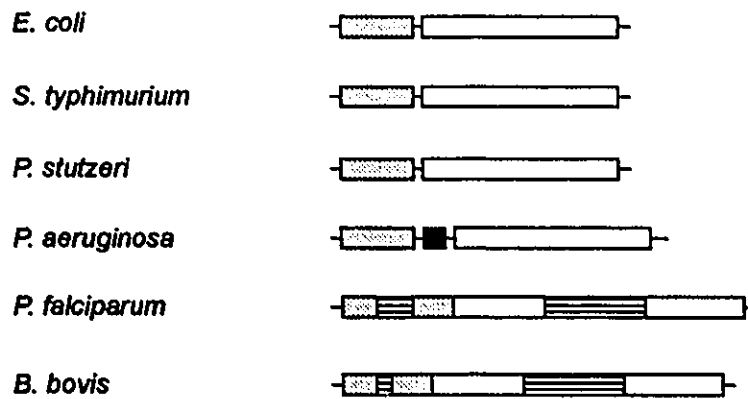
1.8 CPS GENE ORGANISATION AND REGULATION IN THE BACTERIA

For reference, a schematic diagram is shown in Figure 4 summarising the organisation of CPS genes for all organisms in which CPS gene sequences have been obtained. CPS genes have been fully or partially sequenced in 6 bacteria, and in all instances are found to be co-transcribed in the order *carA-carB*. In most prokaryotes examined these genes overlap or are separated by a small intervening sequence. In *E. coli*, *S. typhimurium*, and *P. stutzeri* the genes are separated by 17, 18, and 24 bp, respectively (Piette *et al.*, 1984; Kilstrup *et al.*, 1988, Kwon *et al.*, 1994). The corresponding pyrimidine-specific CPS genes in *Bacillus caldolyticus*, and pyrimidine and arginine-specific CPS genes of *B. subtilis*, overlap by between 8 and 11 bp (Ghim *et al.*, 1994; Mountain *et al.*, 1984; Quinn *et al.*, 1991). *CarA* and *carB* in *P. aeruginosa*, however, are separated by an intervening sequence of 682 bp which contains an unidentified open reading frame (Kwon *et al.*, 1994).

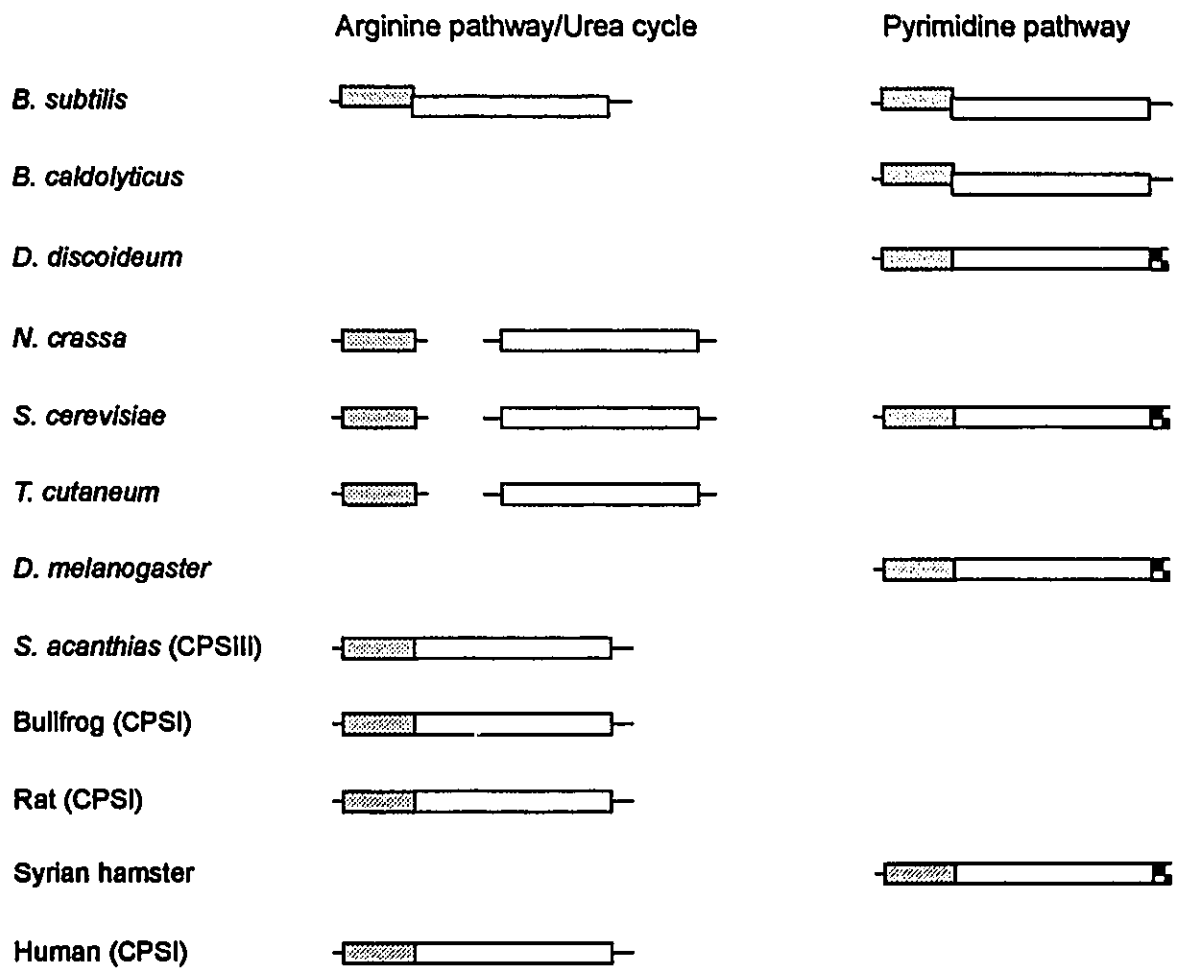
Regulation of CPS differs significantly in the bacteria studied. In *E. coli* and *S. typhimurium*, for which most information is available regarding the regulation of these genes (Cunin *et al.*, 1986), *carAB* is cumulatively repressed by arginine and uracil (Piette *et al.*, 1984; Kilstrup *et al.*, 1988) from two promoters named P1 and P2. P1, which is upstream of P2, is repressed by pyrimidine nucleotides by an as yet incompletely defined complex which involves use of the integration host factor and a regulatory protein CarP (Charlier *et al.*, 1995). The *carP* gene has recently been cloned and sequenced and shown to be identical to the *xerB/pepA* gene encoding the protein aminopeptidase A which is involved in Xer site-specific recombination in ColEI-type plasmids. CarP was found to undergo negative self-regulation and had no effect on any of the other genes involved in pyrimidine biosynthesis (Charlier *et al.*, 1995). The second promoter, P2 is repressed by arginine through the

FIGURE 4. Schematic diagram of the organization of CPS genes from all organisms in which CPS gene sequence data has been obtained. Regions encoding the common amidotransferase and synthetase domains of CPS shown as shaded and white boxes, respectively. Introns are not included. Note that the *N. crassa*, *S. cerevisiae* and *T. cutaneum* arginine-specific CPS genes are encoded by unlinked genes, and that the CPS genes in *B. subtilis* and *B. caldolyticus* overlap. These amidotransferase and synthetase domains have fused in the urea cycle-specific CPS enzymes (CPSI/CPSIII) of *S. acanthias*, bullfrog, rat and human. Checkered boxes denote a portion of the dihydroorotase domain which is fused to the pyrimidine-specific CPS genes of eukaryotes. Striped boxes denote large translated sequences which are present between functional domains of the CPS of *P. falciparum* and *B. bovis*. The black box represents an unidentified open reading frame found between the *P. aeruginosa* CPS genes.

Organisms with one CPS enzyme



Organisms with two CPS enzymes



arginine repressor, which binds to two arg-boxes which overlap P2 (Piette *et al.*, 1984). Interestingly, the arginine repressor has been shown to play an obligate role in the same site-specific recombination for ColEI plasmids as XerB/PepA and is therefore necessary for stable plasmid inheritance (Stirling *et al.*, 1988). This is apparently a completely unrelated function of this protein from its arginine pathway repressing abilities. The arginine repressor's effect on P2 is also affected by the levels of pyrimidines, such that the presence of both pyrimidines and arginine represses P2 more efficiently than arginine alone (Charlier *et al.*, 1988; Lu *et al.*, 1989). Pyrimidines have been shown not to directly interact with the arginine repressor, however it is possible that another factor mediates this contact (Lu *et al.*, 1992).

The CPS of *P. aeruginosa* on the other hand, undergoes only partial cumulative repression by arginine and uracil (Abdelal *et al.*, 1983; Wong and Abdelal, 1990). Studies of the regulation of *carAB* in this organism have demonstrated that these genes are transcribed from just one promoter which is controlled by both arginine and pyrimidines (Wong and Abdelal, 1990; Kwon *et al.*, 1994). Arginine seems to control expression at the level of transcription initiation., while there is strong evidence to suggest that pyrimidines exert their control through an attenuation mechanism (Kwon *et al.*, 1994). This putative attenuation mechanism shows possibly significant differences from the attenuation mechanism known to control the well characterised *pyrBI* operon (Kwon *et al.*, 1994). No repressor similar to the arginine repressor of the Enteric bacteria has been found in *P. aeruginosa* and no arg-boxes are present in the sequence upstream of these CPS genes (Wong and Abdelal, 1990). Between the *carA* and *carB* genes of *P. aeruginosa* lies an ORF of unknown function which encodes a protein of 216 amino acids (Kwon *et al.*, 1994). In *P. stutzeri* the CPS genes are very similar but are not separated by an ORF (Kwon *et al.*, 1994). Similar structural features found upstream of *carA* in *P.*

stutzeri suggest that its CPS genes are also regulated by a similar attenuation mechanism (Kwon *et al.*, 1994).

As mentioned earlier, *B. subtilis* contains two CPS's, one repressible by arginine and encoded within a cluster of genes involved in arginine biosynthesis, and the other regulated by uracil and encoded within an operon of pyrimidine biosynthetic genes (Paulus and Switzer, 1979; Mountain *et al.*, 1984; Quinn *et al.*, 1991). An arginine repressor, analogous to the repressor found in Enteric bacteria, is found in *B. subtilis* (North *et al.*, 1989; Smith *et al.*, 1989). This repressor binds to a promoter region upstream of *argC*, the first gene of the arginine biosynthetic gene operon which contains *carAB*, and so does not specifically act on the CPS genes. The same applies for the pyrimidine-pathway CPS genes, which are also contained within the pyrimidine biosynthetic operon and so are not subject to any sort of unique regulation. Because these two CPS enzymes are present in the same environment (*i.e.* are not separately compartmentalised as in most eukaryotes) this system is somewhat unique in that each CPS can act in both pathways and the total pool of CPS is actually a mixture of the two (Paulus and Switzer, 1979).

1.9 CPS GENES IN THE EUKARYOTES: FEW STUDIES OF REGULATION

CPS genes have currently been sequenced from 12 eukaryotes. As mentioned previously, all eukaryotes examined contain two CPS enzymes, with the exception of the apicomplexan protozoans studied (Chansiri and Bagnara, 1995; Flores *et al.*, 1994). For these organisms studied which have two CPS enzymes, most are compartmentalised so that pyrimidine-specific CPS is located in the nucleus, and arginine-specific CPS is located in the mitochondrion. The only confirmed exception found so far is in *S. cerevisiae* where both CPS's are cytoplasmic (Urrestarazu *et al.*, 1977; Lusty *et al.*, 1983). In all

cases the CPS genes are chromosomally encoded - no CPS genes have been found on mitochondrial genomes.

The pyrimidine-specific CPS of eukaryotes seems to be fairly conserved in structure in the eukaryotes examined. It is usually part of a multifunctional protein catalysing the first two or three steps of the pyrimidine biosynthetic pathway (Shoaf and Jones, 1973; Mori *et al.*, 1975; see section 1.6). Few regulation studies have been performed on these pyrimidine-specific multifunctional complexes, and no studies of regulation have been reported for the monofunctional, monomeric CPS's of the protozoans. A transcription start site has been mapped for the hamster multifunctional protein containing CPS, called CAD. Regulation involves low level repression/derepression by pyrimidines (Kollmar *et al.*, 1994). The focus of studies of CAD, however, were directed at elucidating the boundaries of functional domains and determining the ability of these domains to act as discrete enzymes (Rubio, 1993).

The arginine-specific CPS's vary quite significantly in structure, and even function, within the eukaryotes. In *S. cerevisiae* and *N. crassa*, the arginine-specific CPS is a CPSII-type enzyme, with a functionality similar to the pyrimidine-specific enzymes of eukaryotes and the bacterial CPS's utilised for both pathways. This fungal CPS is a heterodimer encoded by two unlinked genes similar in size to bacterial CPS genes (Werner *et al.*, 1985; Orbach *et al.*, 1990). In *S. cerevisiae*, the gene encoding the synthetase subunit seems to be constitutively expressed, while the gene encoding the amidotransferase subunit is regulated by a complex arginine repression system comprising at least three arginine repressor proteins (Béchet *et al.*, 1970; De Rijcke *et al.*, 1992; Qiu *et al.*, 1990). This essentially results in the small subunit behaving like a rate limiting co-factor for the large subunit. There have been indications in yeast that the large subunit becomes active when arginine is not present, but only when it

aggregates with the small subunit (Piérard *et al.*, 1979). In all other eukaryotes studied, the arginine-specific CPS (CPSI or CPSIII) is monomeric (Nyunoya *et al.*, 1985) but no significant studies of regulation have been performed. Significant functional differences have been noted for these enzymes: *i.e.*, the arginine-specific CPS of ureotelic vertebrates, CPSI, has lost its ability to bind glutamine and instead utilises ammonia directly (Marshall, 1976; Nyunoya *et al.*, 1985). Studies on CPS enzymes involved in the urea cycle has primarily involved gene sequencing to examine structural similarities between CPSI, CPSII and CPSIII enzymes (Hong *et al.*, 1994; Nyunoya *et al.*, 1985), and characterisation of naturally occurring CPSI mutants in humans, which cause a genetic disease leading to hyperammonemia (Haraguchi *et al.*, 1991).

1.10 PHYLOGENETIC ANALYSIS OF CPS: ROOTING THE TREE OF LIFE AND THE NEED FOR AN ARCHAEL CPS SEQUENCE

Clearly the organization and regulation of CPS in different organisms varies significantly. Organisms may contain one or two CPS enzymes, the enzymes may be heterodimeric or monomeric, and the heterodimeric enzymes can be encoded by genes which are co-transcribed or on separate chromosomes. Phylogenetic investigations of CPS sequences could help order these complex findings and with respect to the evolution of CPS.

Because of a duplication within the synthetase domain of CPS genes, phylogenetic trees constructed from CPS sequences could also be rooted. Within the synthetase domain of CPS an internal homology has been postulated to have evolved from an ancient duplication of a kinase gene (Nyunoya and Lusty, 1983). This internal homology has been found in the CPS genes of all the bacteria and eukaryotes examined so far, and has led to suggestions that the original gene duplication occurred

before the divergence of the main domains of Bacteria, Archaea and Eukarya (Nyunoya *et al.*, 1985). Genes which duplicated in the progenote before the divergence of the three domains, have previously been used to root phylogenetic trees (Iwabe *et al.*, 1989; Gogarten *et al.*, 1989; Brown and Doolittle, 1995) and so this duplication within the synthetase domain could be used in a similar manner.

Rooting a tree constructed from CPS genes has obvious benefits including the ability to determine branching order within the low branching organisms. However, if an archaeal sequence were also included in this analysis, the initial branching order for Bacteria, Archaea and Eukarya could also be determined, essentially finding a root for the tree of life. The tree of life was first rooted using elongation factors Tu/1 α and G/2 in addition to ATPase α and β subunits (Iwabe *et al.*, 1989; Gogarten *et al.*, 1989). Both analyses suggested that the Archaea are more related to the Eukarya than Bacteria. However, the statistical reliability of the trees produced using the elongation factors has been criticized (Creti *et al.*, 1994). Also there has been concern that the full family of genes has not yet been clearly determined for the ATPases, and that horizontal gene transfer cannot be discounted (Forterre *et al.*, 1993; Hilario and Gogarten, 1993). Subsequently, aminoacyl-tRNA synthetase genes were used to root the tree of life and also showed the “sistership” of the eukaryotes and Archaea (Brown and Doolittle, 1995). However, there is a need to further confirm this rooting using other genes. In particular there is an interest in rooting the tree of life using genes involved in functions other than transcription/translation since the similarity of the archaeal genes with their eukaryotic counterparts may be in part a reflection of the similarity in the transcription/translation machinery between these two organisms (J.R. Brown, personal communication). Archaea are noted for having a transcription/translation system which

shows many similarities to eukaryotes (Keeling *et al.*, 1995; Keeling and Doolittle, 1995; Langer *et al.*, 1995).

No complete archaeal CPS gene sequence has yet been obtained. Obtaining an archaeal CPS sequence is also of interest to determine if these organisms have a CPS structure similar to those of Bacteria and Eukarya. Recently a CPS protein was studied from the archaeon *Pyrococcus furiosus* and found to be much smaller in size than any other CPS previously characterized (Mr 70,000; Legrain *et al.*, 1995). This CPS was also found to utilize ammonia and not glutamine as a substrate (Legrain *et al.*, 1995). A partial DNA sequence of the synthetase domain of CPS was also obtained from the archaeon *Methanosarcina barkeri* during the course of sequencing an *argG* gene (Morris and Reeve, 1988). Schofield (1993) was then able to extend this sequence by PCR. The results show that the internal duplication present in other CPS genes is also present in the *M. barkeri* sequence. This sequence is unavailable, so determination of a complete archaeal CPS gene sequence is still a hot issue.

Simmer *et al.* (1990) performed a limited evolutionary analysis of CPS genes by constructing dendrograms using a selection of CPS gene sequences. This paper, which focused on the characterization of a mammalian CPS sequence, was unable to come to any strong conclusions from this analysis because of the limited number of CPS sequences available at the time. The authors did suggest that the arginine-specific or CPSI enzymes of eukaryotes could have evolved by diffusion from the pyrimidine-specific CPS which is part of a multifunctional protein.

A phylogenetic analysis of CPS genes was also recently reported by van den Hoff *et al.* at the end of 1995. They used a selection of previously reported CPS sequences to primarily perform an analysis of introns in eukaryotic CPS genes, suggesting supporting evidence for the “introns early”

theory. Their analysis, which included some bacterial sequences, demonstrated that two separate gene duplications led to the formation of the two CPS enzymes in Gram-positive bacteria and eukaryotes.

Recently released sequences (*i.e.*, the apicomplexan protozoan sequences) now enable more conclusions to be made on the evolution of CPS genes through a phylogenetic analysis. There is also a need to include an archaeal sequence in such an analysis, to not only root the tree of life, but to aid in demonstrating which aspects of CPS are conserved in all living organisms.

Currently a project is underway to sequence the entire genome of the Crenarchaeote *Sulfolobus solfataricus*, strain P2 (Sensen *et al.*, in press). During the course of their investigations, they happened to come across a CPS gene sequence and agreed to give me prior access to it as part of a collaboration with Dr. Robert L. Charlebois. This sequence, combined with the other recently obtained protozoan and bacterial sequences, including the gonococcal sequence reported in this work, forms the basis for a comprehensive phylogenetic analysis of all currently known complete CPS genes.

1.11 OBJECTIVES AND RATIONALE

1.11.1 Rationale for initial study

Initial interest in the CPS genes of *N. gonorrhoeae* stemmed from the high prevalence of naturally occurring mutants, and a possible pathogenic advantage that these mutations may confer. However, I also realised that these genes are an excellent candidate for primary studies of gonococcal biosynthetic gene structure and regulation. Few housekeeping genes have been studied in the gonococcus and no significant studies of the organisation or regulation of gonococcal housekeeping genes have been reported. CPS has been studied as a model for gene and enzyme regulation in many other bacteria such as *E. coli*, *S. typhimurium*, *P. aeruginosa*, *B. subtilis*, *S. cerevisiae*, spiny dogfish shark, hamster, rat and human (Cunin *et al.*, 1986; Maas, 1991; Hong *et al.*, 1994). This is in part because regulation of the CPS genes and the CPS enzyme is complex in most organisms, due to the enzyme's effect on both arginine and pyrimidine biosynthesis and its entry level position in each pathway. The regulation and organisation of CPS genes varies significantly between different genera, and Neisseriaceae are on a different evolutionary branch than other well studied Gram-negative bacteria (Woese, 1987), therefore the regulation of CPS genes in *N. gonorrhoeae* may be different. Studies of the gonococcal CPS genes may also give insight into why these genes are so often mutated in the gonococcus.

As suggested in reports by Copley (1987) and Mayer *et al.* (1977), more than one mutation may be causing the CPS deficiencies seen in gonococcal isolates. Evaluation of the gonococcal CPS mutants at a genetic level could resolve this question. If different mutations could be distinguished, subclassification of the auxotrophs would be possible. In the long term, these naturally occurring CPS mutants could also be useful for learning more about the function and regulation of CPS.

No archaeal CPS genes have previously been isolated. Analysis of an archaeal CPS sequence would complete a characterisation of CPS from all three domains of life and provide insights into what features of CPS were present in the progenote before the divergence of Bacteria, Archaea, and Eukarya. Phylogenetic analysis of all CPS sequences, including an archaeal sequence, could provide new insights into the evolution of this enzyme. The internal duplication within CPS could be used for rooting phylogenetic trees and verifying previously postulated theories on early evolutionary relationships. The tree of life has not been previously rooted using a duplication internal to a gene or by using a 'metabolic' gene.

1.11.2 Objectives

The first two objectives were initially proposed, with an aim to begin studies of CPS gene regulation in the gonococcus. However, as unexpected results were obtained, most notably the discovery of a unique CPS gene organisation in the gonococcus and a variable sequence with possible diagnostic/typing applications, the initial objectives were modified and new additional objectives were proposed, as listed below.

1. To confirm the existence of, and determine the primary sequence for, the *carA* and *carB* genes of *Neisseria gonorrhoeae* CH811. To analyse the sequence for known regulatory elements, gonococcal uptake sequences, significant repeats and any other notable motifs, and to perform comparisons with the CPS genes found in other organisms. *Hypotheses:* Gonococcal CPS genes are present and have significant sequence similarity to those of *E. coli*. They have similar gene organisation to other Gram-

negative bacteria, but have different regulatory sequences and have the uptake sequences characteristic of gonococcal genes.

2. To characterise, at the molecular level, the CPS defects found naturally in the gonococcal population. *Hypothesis*: There are at least two different defects causing CPS deficiencies. The CPS defective gonococci can be transformed with cloned, functional CPS genes to produce isolates with functional CPS and novel auxotypes.

3. To characterise the unique intervening sequence between *carA* and *carB* in *N. gonorrhoeae* CH811, investigating its prevalence in other isolates of *N. gonorrhoeae*, and other *Neisseria* species. *Hypothesis*: A similar unique intergenic sequence is present in other isolates of *N. gonorrhoeae*.

4. To perform a comprehensive phylogenetic analysis of all CPS genes, including the use of the internal similarity within CPS to root trees. To perform an analysis of an archaeal CPS sequence, looking for features in common with other CPS genes. *Hypotheses*: The evolution of the CPS genes is complex. Two separate gene duplications led to the formation of the two CPS enzymes in *Bacillus spp.* and eukaryotes. The internal duplication can be used to root trees. Organisms with a similar gene organisation and regulation will branch in an order consistent with other previously determined phylogenetic relationships. The archaeal CPS sequence will have some features common to other CPS enzymes, such as an internal duplication within the synthetase domain.

2. MATERIALS AND METHODS

2.1 BACTERIAL STRAINS, PLASMIDS AND PHAGES

The *carA* and *carB* genes were previously cloned from *N. gonorrhoeae* strain CH811 [auxotype/serotype/plasmid content class (A/S/P class): non-requiring/IB-2/plasmid free; Moreno *et al.*, 1987] into EMBL4 to produce clone EFP10 (Figure 5; Picard and Dillon, 1989b). A 5 kb *EcoRI* fragment from this clone had been previously subcloned into pBR329 to produce pSTD18 (Dillon and Bigelow, unpublished data; Figure 5). Both EFP10 and pSTD18 were used for subcloning DNA fragments containing *carA* and *carB*. pMC50, a plasmid containing the *E. coli carAB* operon on a 6.3 kb *HindIII* fragment, was used for comparative studies and was provided by R. Cunin (Vrije Universiteit Brussel, Brussels, Belgium; Figure 3). The vector used for all cloning experiments was the plasmid pBluescriptII KS(+) (Stratagene, LaJolla, California). *E. coli* NM539 (F⁻, *supF*, *hsdR*, *lacY*, (P2)) was used for phage propagation. *E. coli* JM83 (*ara*, Δ *lac*, *pro*, *thi*, *strA*, f80 Δ *lacZ* Δ M15) was used in all transformation experiments to select recombinants. Chromosomal DNA from *E. coli* JM240, *E. coli* HB101 and *Peptostreptococcus magnus* ATCC 53516 (the latter supplied by J. Ng of our laboratory) were used as controls in hybridisation experiments. *N. gonorrhoeae* strain FS62 (Str^R) was used as a control in the gonococcal transformation experiments (Sparling, 1966; West and Clark, 1989; Dillon, unpublished data).

Other strains of *N. gonorrhoeae*, which were used for PCR-based studies of the intergenic sequence between *carA* and *carB*, are listed in Tables 1, 2 and 3 (Reasons for their use are discussed below). All isolates were chosen from the culture collection of the National Laboratory for Sexually Transmitted Diseases (NLSTD), Laboratory Centre for Disease Control (LCDC), Ottawa, Canada,

FIGURE 5. Location of *carA* and *carB* in the gonococcal inserts of phage and plasmid clones. Arrows indicate the progression in the isolation of each clone, which all originate from EFP10. Relevant restriction enzyme sites are indicated above each clonal insert. The relative position of *carA* and *carB*, determined by complete DNA sequencing of pFL50 and pFL35, is shown below the clones in blue.

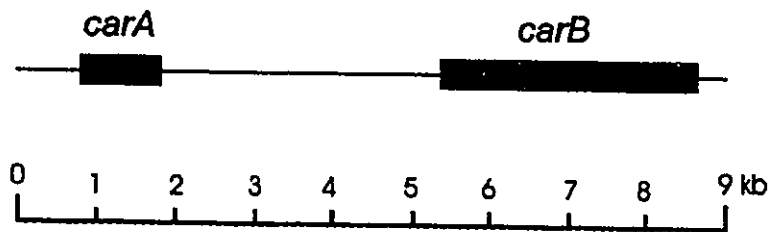
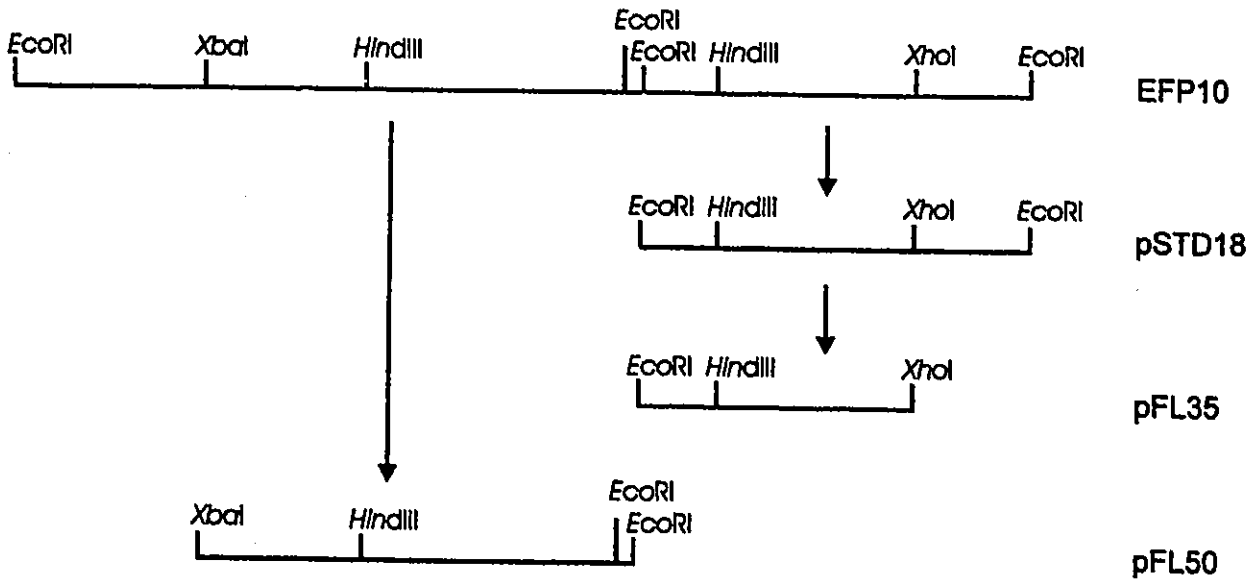


TABLE 1. *N. gonorrhoeae* isolates selected for their varying auxotype^a, serotype, plasmid content^b (A/S/P class), date of isolation and geographic origin.

Isolate	A/S/P class	Date of isolation	Geographic Source
CH811	NR/IB-2/ plasmid free	1982	Chile
NS466	NR/IB-5/2.6	88/03	Manitoba, Canada
NS2117	NR/IB-5/2.6	88/08	Manitoba, Canada
NS568	NR/IB-7/2.6	88/04	British Columbia, Canada
NS1095	NR/IB-7/2.6,24.5	88/06	New Brunswick, Canada
MS11-ms	NR/IB-9/2.6	1976	North Carolina, USA
11110	NR/IA-1/2.6,25.2	92/07	Nova Scotia, Canada
9228	NR/IA-2/2.6,3.2,25.2	92/03	Ontario, Canada
11844	NR/IA-2/2.6,3.2,25.2	92/11	Ontario, Canada
9345	NR/IA-5/2.6,3.2,24.5	92/03	Ontario, Canada
9881	NR/IA-5/2.6,3.2,24.5	92/04	Ontario, Canada
9257	P/IA-6/2.6,4.5	92/03	British Columbia, Canada
10830	P/IA-6/2.6,4.5,25.2	92/08	British Columbia, Canada
11860	P/IA-6/2.6,3.2,25.2	92/11	Ontario, Canada
FA1090	P/IB-3/2.6	1970	New York, USA
NS4702	P/IB-5/2.6,24.5	89/01	Ontario, Canada
NS5846	P/IB-5/2.6,24.5	90/06	Ontario, Canada
NS667	P/IB-7/2.6	88/05	Ontario, Canada
NS3019	P/IB-7/2.6	89/01	Ontario, Canada
NS8318	P/IB-7/2.6	91/11	Québec, Canada
9073	OUH/IA-2/2.6	92/02	Alberta, Canada
9315	OUH/IA-2/2.6	92/03	Ontario, Canada

continued...

TABLE 1. *continued...*

Isolate	A/S/P class	Date of isolation	Geographic Source
NS303 ^c	CUH/IA-2/2.6	88/04	Ontario, Canada
NS686 ^c	CUH/IA-2/2.6	88/05	Nova Scotia, Canada
NS716 ^c	CUH/IA-2/2.6	88/05	Québec, Canada
NS1061 ^c	CUH/IA-2/2.6	88/06	Yukon, Canada
NS791 ^c	PCU/IB-3/2.6	88/05	Saskatchewan, Canada
NS2152 ^c	PCU/IB-2/plasmid free	88/10	Alberta, Canada
NS384 ^c	PCU/IB-2/plasmid free	88/05	Nova Scotia, Canada
NS348 ^c	PCU/IB-7/plasmid free	88/04	Alberta, Canada

^a The following abbreviations have been used for auxotype: NR, non-requiring; P: proline-requiring; OUH: ornithine-, uracil-, and hypoxanthine-requiring; CUH: citrulline-, uracil-, and hypoxanthine-requiring; PCU: proline-, citrulline-, and uracil-requiring.

^b Plasmid sizes are given in MegaDaltons.

^c Isolates previously shown by enzyme assays to be CPS deficient (Li, 1993).

TABLE 2. Isolates of *N. gonorrhoeae* selected for their similar A/S/P class (P/IA-6/2.6,4.5) but varying in date of isolation and geographic origin within two Canadian provinces.

Isolate	A/S/P class	Date of isolation	Province of Isolation
709	P/IA-6/2.6,4.5	84/11	Ontario
767	P/IA-6/2.6,4.5	85/01	Ontario
1114	P/IA-6/2.6,4.5	86/02	Ontario
2082	P/IA-6/2.6,4.5	87/10	Ontario
2318	P/IA-6/2.6,4.5	88/04	Ontario
4165	P/IA-6/2.6,4.5	89/12	Ontario
4555	P/IA-6/2.6,4.5	90/02	Ontario
5988	P/IA-6/2.6,4.5	91/01	Ontario
6384	P/IA-6/2.6,4.5	91/04	Ontario
6769	P/IA-6/2.6,4.5	91/06	Ontario
7121	P/IA-6/2.6,4.5	91/08	Ontario
8266	P/IA-6/2.6,4.5	91/12	Ontario
1537	P/IA-6/2.6,4.5	86/10	British Columbia
2525	P/IA-6/2.6,4.5	88/07	British Columbia
5774	P/IA-6/2.6,4.5	90/12	British Columbia
5788	P/IA-6/2.6,4.5	91/01	British Columbia
6339	P/IA-6/2.6,4.5	91/05	British Columbia
7921	P/IA-6/2.6,4.5	91/09	British Columbia
7944	P/IA-6/2.6,4.5	91/10	British Columbia

TABLE 3. Two clusters of isolates of *N. gonorrhoeae* with an A/S class similar to isolates associated with two separate outbreaks: 1) isolates of A/S class NR/IA-5, associated with an outbreak in Toronto; 2) isolates of the A/S/P class NR/IB-3/2.6/24.5 associated with an outbreak in Winnipeg which spread to Toronto.

Isolate	A/S/P Class	Year Isolated	Geographic Source
3523	NR/IA-5/2.6,24.5	1989	Toronto, Ontario
2852	NR/IA-5/2.6,3.2,24.5	1988	Toronto, Ontario
3735	NR/IA-5/2.6,3.2,24.5	1989	Toronto, Ontario
3928	NR/IA-5/2.6,3.2,24.5	1989	Toronto, Ontario
4020	NR/IA-5/2.6,3.2,24.5	1989	Toronto, Ontario
5265	NR/IA-5/2.6,3.2,24.5	1990	Toronto, Ontario
5707	NR/IA-5/2.6,3.2,24.5	1990	Toronto, Ontario
6525	NR/IA-5/2.6,3.2,24.5	1991	Toronto, Ontario
6595	NR/IA-5/2.6,3.2,24.5	1991	Toronto, Ontario
5283	NR/IA-5/2.6,3.2,25.2	1990	Toronto, Ontario
88-1286	NR/IB-3/2.6,24.5	1988	Winnipeg, Manitoba
88-1919	NR/IB-3/2.6,24.5	1988	Winnipeg, Manitoba
88-2351	NR/IB-3/2.6,24.5	1988	Winnipeg, Manitoba
88-2802	NR/IB-3/2.6,24.5	1988	Winnipeg, Manitoba
88-3482	NR/IB-3/2.6,24.5	1989	Winnipeg, Manitoba
88-3511	NR/IB-3/2.6,24.5	1989	Winnipeg, Manitoba
3644	NR/IB-3/2.6,24.5	1989	Toronto, Ontario
3648	NR/IB-3/2.6,24.5	1989	Toronto, Ontario
4743	NR/IB-3/2.6,24.5	1990	Toronto, Ontario
5837	NR/IB-3/2.6,24.5	1990	Toronto, Ontario
5840	NR/IB-3/2.6,24.5	1990	Toronto, Ontario
5855	NR/IB-3/2.6,24.5	1990	Toronto, Ontario

with the exception of isolates MS11-ms and FA1090 which were obtained from J. Cannon (University of North Carolina). Isolates listed in Table 1 were chosen to vary in their geographic source in Canada, isolation site, A/S/P class, and date of isolation from 1988 to 1992. Exceptions include isolate CH811 which was first isolated in Chile in 1982 (Moreno *et al.*, 1987), and isolates FA1090 and MS11-ms which are described by West and Clark (1989). Isolates listed in Table 2 were selected to have the same A/S/P class, but still vary in location within two provinces and also vary in date of isolation from 1984 to 1991. Isolates listed in Table 3 comprise clusters of more closely related isolates which have a similar A/S/P class, location of isolation, and date of isolation as isolates associated with two gonorrhea outbreaks (Li, 1993). One outbreak comprised isolates of A/S class NR/IA-5 and occurred in Toronto, Ontario. The other unrelated outbreak comprised isolates of A/S/P class NR/IB-3/2.6/24.5 which were first isolated in Winnipeg, Manitoba, and later noted in Toronto, Ontario.

The sequence between *carA* and *carB* was also investigated from 12 *N. meningitidis* isolates and 18 commensal *Neisseria* isolates (comprising 9 species) listed in Table 4. Chromosomal DNA from these species was kindly provided by Dr. Craig Strathdee (National Laboratory for Special Pathogens, LCDC). These isolates had been speciated using standard identification criteria (Vedros, 1984) at the National Laboratory for Bacteriology (LCDC, Canada).

2.2 MEDIA, GROWTH CONDITIONS AND STRAIN TYPING

Confirmation of speciation of *N. gonorrhoeae* isolates was performed by visual identification of the pure culture combined with the oxidase test, and were previously speciated at the NLSTD (LCDC, Ottawa, Canada). *N. gonorrhoeae* isolates were stored frozen at -70°C in Brain-Heart Infusion

TABLE 4. *N. meningitidis* and commensal *Neisseria* species isolates.

Isolate	Species	Serovar	Isolate	Species	Serovar
1503	<i>N. meningitidis</i>	A 4:P1.9	1685	<i>N. mucosa</i>	N/A
1504	<i>N. meningitidis</i>	A 4:P1.9	1686	<i>N. mucosa</i>	N/A
1500	<i>N. meningitidis</i>	B 4:P1.15	1687	<i>N. lactamica</i>	N/A
1501	<i>N. meningitidis</i>	B 2b:P1.2	1688	<i>N. lactamica</i>	N/A
1514	<i>N. meningitidis</i>	C 2a:P1.2 ^a	1689	<i>N. lactamica</i>	N/A
1518	<i>N. meningitidis</i>	C 2a:P1.2 ^a	1690	<i>N. perflava/sicca</i>	N/A
1525	<i>N. meningitidis</i>	C 2a:P1.1 ^a	1691	<i>N. perflava/sicca</i>	N/A
1526	<i>N. meningitidis</i>	C 2a:P1.1 ^a	1692	<i>N. perflava/sicca</i>	N/A
1527	<i>N. meningitidis</i>	C 2a:P1.1 ^b	1693	<i>N. flavescens</i>	N/A
1339	<i>N. meningitidis</i>	W135	1694	<i>N. flavescens</i>	N/A
1497	<i>N. meningitidis</i>	Y	1695	<i>N. polysacchareae</i>	N/A
1498	<i>N. meningitidis</i>	Y	1696	<i>N. polysacchareae</i>	N/A
1682	<i>N. flava</i>	N/A	1697	<i>N. polysacchareae</i>	N/A
1683	<i>N. subflava</i>	N/A	1698	<i>N. cinerea</i>	N/A
1684	<i>N. elongata</i>	N/A	1699	<i>N. cinerea</i>	N/A

^aET-15 (electrophoretic type 15) isolates.^bnon-ET-15 isolate.

containing 20% glycerol. Except where indicated they were subcultured for 18-24 hours on GC medium base (Difco) supplemented with Kellogg's defined supplement, as modified by Dillon (1983), in the presence of 5% CO₂, in a humid environment, at 35°C. For some transformation experiments, gonococcal isolates were grown on GC medium base (GCMB) supplemented with 500 mg/l streptomycin under the same conditions. For auxotyping and some other transformation experiments, the isolates were grown on chemically defined media as described by Hendry and Stewart (1979). For RNA analysis, gonococci were grown on a modified form of this chemically defined media (as further described in Section 3.3). *E. coli* strains were grown on tryptic soy agar (TSA) at 37°C, for 18-24 hours, with supplements added at the following final concentrations: ampicillin, 100 mg/L; tetracycline, 10 mg/L; X-gal 80 mg/L. LB broth was used for all liquid cultures with final concentrations of 10mM MgSO₄ and 0.2% maltose used in phage propagations.

The auxotype of each gonococcal isolate listed in Tables 1, 2, and 3 was determined at the NLSTD using a modified method (Dillon, 1983) of the method developed by Hendry and Stewart (1979). The serotype and plasmid content of each gonococcal isolate was also determined at the NLSTD using the methods of Knapp *et al.* (1984), and Dillon *et al.*, (1985), respectively. *N. meningitidis* isolates were typed by the National Laboratory for Bacteriology (LCDC) using methods reviewed by Frasch *et al.* (1985).

2.3 DNA PURIFICATION, RESTRICTION ENDONUCLEASE DIGESTION, AND ELECTROPHORESIS

DNA was extracted from *N. gonorrhoeae* and *E. coli* as described by Dillon *et al.* (1985) and Birnboim and Doly (1979), respectively. DNA was purified by caesium chloride density gradient

ultracentrifugation (Sambrook *et al.*, 1989), except for small scale multiple plasmid isolations, where DNA was purified using the Magic Minipreps™ DNA purification system of Promega (Fisher Scientific, Ottawa, Ontario), or for some PCR reactions where purification entailed a phenol:chloroform extraction followed by ethanol precipitation (Sambrook *et al.*, 1989). However, for most PCR reactions involving *N. gonorrhoeae* isolates, the template DNA was not purified and instead overnight cultures of *N. gonorrhoeae* were simply suspended in water to a MacFarlands 0.5 standard (supplied by Unipath, Nepean, Ontario) and samples for PCR were directly obtained from this crude lysate. Phage DNA was isolated from polyethylene glycol-precipitated phage stocks using the method described by Sambrook *et al.* (1989). Restriction endonucleases were obtained from Boehringer Mannheim (Laval, Québec) or GIBCO BRL (Gaithersburg, Maryland) and used according to the manufacturer's instructions. DNA was electrophoresed in 1% agarose gels in TAE buffer (40mM Tris-acetate/1mM EDTA) using a constant voltage of 50 Volts, with the exception of the restriction enzyme digestions of PCR amplicons which were electrophoresed in a 1.5% agarose gels under the same conditions (Sambrook *et al.*, 1989). DNA was visualised in the agarose gel by ethidium bromide staining (1mg/l), and photographed using the GelPrint 2000i digital imaging system (BIO/CAN Scientific, Mississauga, Ontario). Fragments of DNA from restriction endonuclease digestions were purified from agarose gels using the QIAEX gel extraction kit (Qiagen, Chatsworth, CA).

2.4 SUBCLONING OF DNA FRAGMENTS

A portion of the *carB* gene was initially subcloned from pSTD18 into pBluescriptII KS(+) producing the plasmid pFL35 as follows (Figure 5). pSTD18 (Amp^R, Tet^R) and pBluescriptII KS(+) (Amp^R) were digested with *Eco*RI and *Xho*I, co-precipitated with ethanol, and ligated with T4 DNA

ligase (Boehringer Mannheim) at 16°C overnight. *E. coli* JM83 was transformed using the calcium chloride method described by Sambrook *et al.* (1989), and clones which were ampicillin resistant, tetracycline sensitive, and unable to utilise X-gal, were selected. Plasmid DNA from the selected clones was then screened by gel electrophoresis for plasmid size, and putative positive clones were confirmed by high stringency Southern hybridisation with the fragment of interest (see section 2.5 for high stringency conditions).

The *carA* gene, and the rest of *carB*, were subcloned from EFP10 into pBluescriptII KS(+) producing the plasmid pFL50 (Figure 5): EFP10 and pBluescriptII KS(+) were digested with *EcoRI* and *XbaI*, and treated as described above for the subcloning from pSTD18, with the exception that screening for tetracycline resistance was not performed.

2.5 SOUTHERN HYBRIDIZATION STUDIES

Probes were prepared by either PCR amplification, or isolation of DNA fragments from an agarose gel using the QIAEX gel extraction kit (Qiagen, Chatsworth, CA). Probes were labelled with digoxigenin by the random priming method using the GeniusTM kit (Boehringer Mannheim, Laval, Québec). DNA electrophoresed in an agarose gel as described above was transferred to positively charged nylon membrane (Boehringer Mannheim) by the method of Southern (1975) or by using the BioRad Model 785 Vacuum Blotter (BioRad Laboratories, Hercules, California) as per the manufacturer's instructions. Hybridisation with labelled probe at high stringency was completed at 68°C in 5x SSC overnight, followed by two 15 minute washes at 68°C in 0.1x SSC, 0.1% SDS. Hybridisations at lower stringencies were done using the same solutions and incubation times but with

lower hybridisation and washing temperatures ranging from 55 to 66°C as indicated. Blots were visualised using the colourimetric method of the Genius™ kit protocol (Boehringer Mannheim).

2.6 GENERATION OF PLASMIDS WITH NESTED DELETIONS

Deletion derivatives of pFL50 and pFL35 were prepared, in both directions, using the Erase-a-base™ kit of Promega (Fisher Scientific, Ottawa, Ontario). First the plasmid DNA was digested with either *SacI* or *KpnI* (depending on the direction of the deletions) to generate one DNA end protected from Exonuclease III digestion. Restriction enzymes with sites corresponding to ends of the insert of pFL50 (*XbaI* or *EcoRI*) or pFL35 (*EcoRI* or *XhoI*) were used to generate an end susceptible to Exonuclease III. Exonuclease III digestion was performed at 30°C and aliquots were taken from the reaction mixture at 30 second intervals for a total of 20 time points. The remainder of the protocol was followed as per the manufacturers instructions. Clones containing deletions of pFL35 from the *EcoRI* site were named EB1-1 to EB1-10 for those obtained from the first time point, EB2-1 to EB2-10 for those obtained from the second time point, and so on until the last time point in which colonies were obtained, which was time point 17. Clones containing deletions of pFL35 from the *XhoI* site were named EBB1-1 etc.. in a similar fashion and colonies were obtained up until time point 16. A strong correlation was not observed between the size of the deletion and the time point from which the clone was obtained, and so subsequent deletion clones obtained were just numbered consecutively. Clones with deletions of pFL50 from the *EcoRI* site were named EBC1 to EBC228. Clones with deletions of pFL50 from the *XbaI* site were named EBD1 to EBD41. All clones were used for sequencing of pFL50 and pFL35 and so the exact nature of the deletion in each case has been determined.

2.7 DNA SEQUENCING AND SEQUENCE ANALYSIS

DNA sequences were determined by the dideoxynucleotide-chain termination method of Sanger *et al.* (1977) using a combination of the nested deletion clones described above and synthetic oligonucleotide primers prepared by General Synthesis and Diagnostics (Toronto, Ontario). Primers were selected using the PrimerDesigner computer program (Scientific and Educational Software). Both manual and automated DNA sequencing methods were used. For manual sequencing, SequenaseTM (United States Biochemical, Cleveland, Ohio) and TaqTrackTM (Promega, Fisher Scientific, Ottawa, Ontario) DNA sequencing kits were used with ³⁵S-dATP purchased from Dupont (Mississauga, Ontario). Electrophoresis was performed using the Model S2 sequencing gel apparatus of GIBCO BRL with 6% polyacrylamide denaturing gels. For automated sequencing, the PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing kit (Applied Biosystems Canada Inc., Mississauga, Ontario) and Centri-Sep Spin Columns (Princeton Separations, Adelphia, New Jersey) were used to prepare reactions for the Applied Biosystems Model 373A DNA Sequencing System. Enough sequence was obtained for a coverage redundancy of 2 in all regions (both strands of the DNA were completely sequenced), with some regions having a coverage redundancy of three. DNA sequences were compiled and analysed using the MicrogenieTM (Beckman Instruments, Palo Alto, California) and PC Gene (IntelliGenetics Inc., Geneva, Switzerland) computer programs. Alignments of the deduced protein sequences from the CPS gene sequences were obtained using the ClustalV multiple sequence alignment method of the GDE (genetic data environment) version 2.2 computer program. Default settings were used with the exception that a fixed gap penalty of 40 was used.

2.8 POLYMERASE CHAIN REACTION (PCR)

Primers were selected and synthesised as described for DNA sequencing. A list of primers used to amplify portions of the sequence comprising *carA* and *carB* is given in Table 5, and the priming location for each is shown schematically in Figure 6. The sequence between *carA* and *carB* was amplified (using primers 61 and 62) from all *Neisseria* species isolates listed in Tables 1, 2, 3 and 4. Amplification of *carA* (primers 38 and 71), and *carB* (primers 78 and 22) was performed with the following 11 *N. gonorrhoeae* isolates: CH811, MS11-ms, FA1090, 11860, 9345, NS791, NS384, NS303, NS716, NS1061, NS686. A portion of the intergenic sequence containing RS6/RS7 sequences (primers 66 and 72) was amplified from the following five *N. gonorrhoeae* isolates: CH811, MS11-ms, FA1090, 11860, and 9345. Five gonococcal isolates used for amplification of other portions of the intergenic sequence (primers 53 and 74, and primers 75 and 63; see Table 5 for details) were: CH811, FA1090, 11110, NS5846, and 9315. PCR amplification of a 299 bp portion of *argJ* was performed on isolates CH811, 9315, NS686 and NS686C⁺, using the primers 5'-ACCGTCGGTGCAGTCTTCA-3' and 5'-TCGTTCCAGAAGGCAGGCT-3' (not listed in Table 5) which are homologous to positions 135-154 and 414-433 of the *argJ* gene (co-ordinates start at the beginning of the gene sequence reported by Martin and Mulks [1992b]).

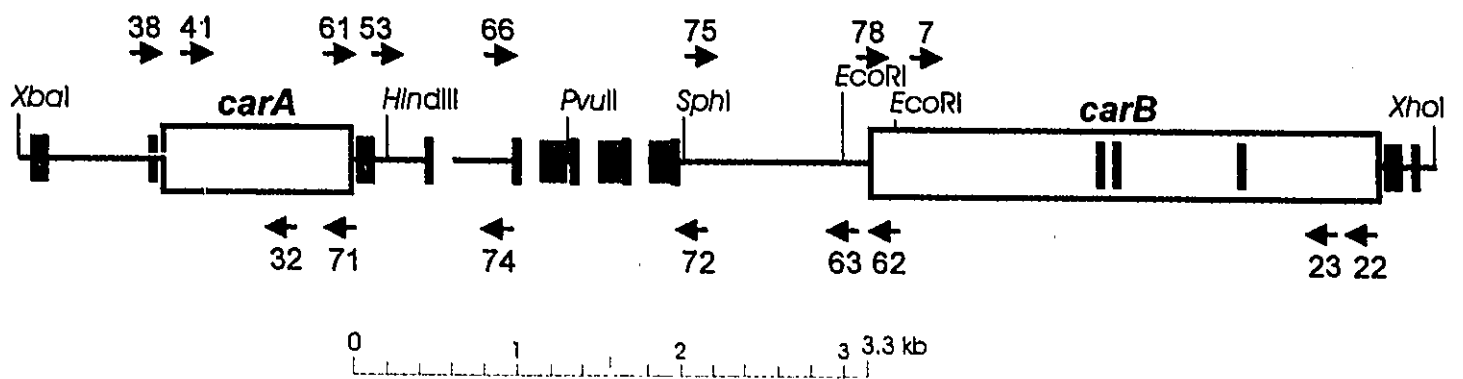
The GeneAmpTM PCR kit (Perkin Elmer Cetus, Norwalk, Connecticut) was used for all PCR experiments with the following reaction conditions: 1x PCR BufferII, 200 mM dNTPs, 1.5 mM MgCl₂, 2.5 U/100ml *Taq* polymerase, 0.2 mM primer, and 0.1 mg template DNA in a 100 ml total volume. Note that for PCR amplifications from most gonococcal isolates, the DNA was not purified and instead 10 µl of a crude lysate (described above in Section 2.3) was used as a template DNA source. For all other amplifications, the template DNA had been purified by either CsCl density

TABLE 5. Accurate location of the priming sites for primers used in PCR amplifications, based on the total sequence obtained from pFL50 and pFL35.

Primer	Priming location coordinates ^a	Complementary strand
7	5965-5982	no
22	8575-8592	yes
23	8467-8484	yes
32	1530-1546	yes
38	651-666	no
41	923-940	no
53	2104-2120	no
61	1908-1927	no
62	5342-5365	yes
63	4965-4984	yes
66	2981-2999	no
71	2063-2081	yes
72	3921-3940	yes
74	2879-2897	yes
75	3966-3984	no
78	5290-5309	no

^a coordinates were determined based on the top strand of the total sequence of pFL50 and pFL35 (shown in Figure 12).

FIGURE 6. Schematic diagram of the sequence comprising *carA* and *carB* from *N. gonorrhoeae* CH811, showing the priming location for each primer used for PCR amplifications (small arrows). Additional restriction enzyme sites relevant for the production of some probes are also indicated. Yellow and purple boxes denote copies of the newly identified repetitive sequences, RS6 and RS7, respectively. Thick vertical bars mark the location of gonococcal uptake sequences.



gradient centrifugation or phenol:chloroform/ethanol precipitation (Section 2.3). For the PCR reactions, all components were kept at 4°C (*i.e.* "cold start" method) until placed in the thermocycler that had been preheated to 94°C. The following thermal profile was performed in a Perkin-Elmer 9600 Thermocycler unless otherwise described: An initial hold at 94°C for 3 min.; 35 cycles of 94°C for 15 sec., 60°C for 15 sec., and 72°C for 3 min. 30 sec.; a final hold at 72°C for 5 min.. All amplifications were done in duplicate. Isolates not producing amplicons of the intergenic sequence using the thermal profile described above were further investigated by using the same protocol with a lower primer annealing temperature (45°C). A negative control comprising all components except the template DNA (which was replaced by water) was included with all PCR experiments.

2.9 RNA EXTRACTION AND ANALYSIS BY NORTHERN BLOT

N. gonorrhoeae cultures were grown on specially formulated chemically defined media (see results Section 3.3) under the conditions described in Section 2.2, with the exception that the cultures were only incubated for 9 to 12 hours. Total RNA was extracted from this culture using RNeasy Total RNA Extraction Kit (Qiagen, Chatsworth, California) as per the manufacturers instructions. This kit uses guanidinium isocyanate in the initial cell lysis buffer, followed by a patented column purification method. The RNA was stored at -70°C in RNase-free water until use. RNase-free water was prepared by adding 1 ml of diethylpyrocarbonate (DEPC) to 1 L of sterile distilled water, incubating at 37°C overnight, and then autoclaving for 30 minutes to remove all traces of the DEPC.

Immediately before electrophoresis of the RNA, 4 volumes (12 µl) of sample preparation buffer was added to the RNA sample (3 µl), and the RNA was denatured at 65°C for 10 minutes. This sample preparation buffer comprised a freshly prepared stock of 250 µl of formamide, 83 µl of formaldehyde,

50 µl 10xMOPS (200mM 4-morpholinopropanesulfonic acid, 50 mM sodium acetate, 10 mM EDTA, pH 7), 8 µl of 0.08g/ml bromophenol blue, and 9 µl of 10 mg/ml ethidium bromide. Electrophoresis of the RNA was performed in a 1% agarose gel prepared with 1xMOPS and 6% formaldehyde as described by Sambrook *et al.* (1989) and the RNA visualized directly under UV light (due to the presence of ethidium bromide in the sample preparation buffer). The gel was then rinsed with RNase-free water for at least one hour to remove the formaldehyde prior to Northern transfer. The RNA was then transferred to positively charged nylon membrane (Boehringer Mannheim) using the BioRad Model 785 Vacuum Blotter (BioRad Laboratories) as per the manufacturer's instructions but using a longer transfer time of 2.5 hours.

Probes comprising portions of the *carA* and *carB* genes of *N. gonorrhoeae* CH811 were prepared by PCR amplification from the plasmid clones pFL50 and pFL35 using primers 41 and 32 (*carA*) and primers 7 and 23 (*carB*; Figure 6). For further purification, the PCR amplicons were electrophoresed on a 1% agarose gel, and the bands excised. These amplicons were then labelled by the random prime method using the rediprime DNA labelling system of Amersham International (Buckinghamshire, England). Their protocol was followed for the labelling of DNA fragments in low melting point agarose, and Redivue [$\alpha^{32}\text{P}$]dCTP, 250 µCi (Amersham) was used in the labelling reaction.

The blots were prehybridised for 15 minutes, and then hybridized for 2 hours with freshly denatured labeled probe, at 65°C, using the Rapid-Hyb buffer of Amersham. The blots were then washed for 20 minutes in 5xSSC, 0.1%SDS at room temperature, followed by two 15 minutes washes in 0.1xSSC, 0.1% SDS at 65°C.

2.10 TRANSFORMATION OF MUTANT GONOCOCCAL ISOLATES WITH CPS GENES

The following method, modified from Janik *et al.* (1976), was developed. First pure cultures of gonococci with the T2 colony phenotype were obtained by repeated selective subculturing (T2 colonies contain piliated gonococci which are naturally competent for DNA uptake): After primary growth of the culture, colonies of phenotype T2 were visually identified according to descriptions by Kellogg *et al.* (1963). These colonies were selectively subcultured onto GCMB every 18 to 20 hours until a pure culture with this phenotype was obtained.

Then approximately 10^9 cfu of each strain (the cfu was approximated by determining the OD of a sample) was cultured onto a final plate of non-selective media (GCMB) and the plate divided into 6 regions. To each region was added 5 μ g of the different DNA treatments such as pFL35 and pFL50 DNA, and positive and negative controls (see below). After 6 hours, the DNA-treated cultures were transferred to selective media comprising the chemically defined media used for auxotyping (lacking citrulline or uracil or other nutrients as controls). After 24 and 48 hours the plates were scored for growth, and any growth was subcultured onto fresh selective media, to confirm the positive result. Controls included checking: 1. DNA sterility by plating the DNA solutions alone; 2. culture purity by plating each culture on non-selective media; 3. auxotyping by the method of Hendry and Stewart (1979); 4. ability to be transformed to streptomycin resistance using chromosomal DNA from the streptomycin resistant gonococcal strain FS62 (to investigate this transformation to streptomycin resistance, both treated and untreated cultures were plated onto GCMB plus streptomycin as the selective media); 5. ability to be transformed by DNA from the prototrophic gonococci strain CH811; 6. lack of spontaneous mutation by plating the cultures alone without DNA treatment; and finally 7. accuracy of the selective media by plating control strains of known auxotype.

Test strains initially examined included the CPS deficient isolates NS303, NS686, NS716, NS348, and NS2152.

2.11 PHYLOGENETIC ANALYSIS

2.11.1 DNA sequences

21 completely sequenced CPS genes were used for the phylogenetic analysis, including the sequence I had obtained from *N. gonorrhoeae* CH811. The *P. stutzeri* and *P. aeruginosa* synthetase protein sequences were kindly supplied by Dr. A.T. Abdelal (Georgia University, Atlanta, Georgia) before publication or submission to GenBank. The *S. solfataricus* P2 CPS gene sequences were included as part of a collaboration with Dr. Robert L. Charlebois of the Sulfolobus Genome Project. Under this collaboration, Dr. Charlebois supplied the sequence which I analyzed for notable CPS sequence motifs and also included in my phylogenetic analysis. All other sequences were obtained by screening the GenBank and EMBL databases (as of August 1995) for all sequences reported to contain complete CPS genes.

In total, 20 sequences from the following organisms were used (accession numbers in brackets). Note that for organisms in which there are two CPS enzymes, the particular CPS gene sequence obtained is designated as either CPSI, CPSIII, Arg (CPSII-arginine-specific), or Pyr (CPSII-pyrimidine-specific). For organisms in which there is only one CPS enzyme, no such designation is given: *Escherichia coli* (J01597), *Salmonella typhimurium* (X13200), *Pseudomonas stutzeri* (U04993), *Pseudomonas aeruginosa* (U04992), *Neisseria gonorrhoeae* (U11295), *Bacillus caldolyticus* (Pyr, X73308), *Bacillus subtilis* (Arg and Pyr, Z26919, M59757), *Sulfolobus solfataricus* (U33768), *Plasmodium falciparum* (L32150), *Babesia bovis* (U18792), *Saccharomyces cerevisiae*

(Arg and Pyr; K02132, K01178, M27174), *Neurospora crassa* (Arg; J05512), *Trichosporin cutaneum* (Arg; L08965), *Dictyostelium discoideum* (Pyr; X14533, X55433), *Squalus acanthias* or spiny dogfish shark (CPSIII; L31362), *Rana catesbeiana* or bullfrog (CPSI; U05193), Syrian hamster (Pyr; J05503), rat (CPSI; M11710, M12318-28), and human (CPSI; D90282).

Note that only the amidotransferase gene has been sequenced from *N. crassa* and *S. typhimurium* (Davis *et al.*, 1980; Kilstrup *et al.*, 1988) and only the synthetase gene has been sequenced from *T. cutaneum* (Reiser, 1994). For organisms with two CPS enzymes, *S. cerevisiae* and *B. subtilis* are the only organisms with two CPS enzymes from which both CPS genes sequences have been obtained. The sequence from *Drosophila melanogaster* was not included due to possible mistakes noted by Simmer *et al.* (1990). No partial sequences were included in this study.

2.11.2 Alignment of sequences

Deduced protein sequences from the DNA sequences obtained were aligned using the ClustalV multiple sequence alignment method of GDE version 2.2, using the identity matrix or PAM 250 matrix with a fixed gap penalty of 40, and a floating gap penalty of 10. After initial alignment of the sequences, the central amidotransferase and synthetase domains were determined and used for subsequent phylogenetic analysis. These domains correspond to residues 38 to 367 of the deduced protein from the *E. coli carA* sequence (amidotransferase domain) and residues 12 to 916 of the deduced protein from *E. coli carB* (synthetase domain). The synthetase domain has internal similarity (Nyunoya and Lusty, 1983), and so additional alignments were also performed using the two similar regions of this synthetase domain. These regions correspond to residues 12 to 333 in the N-terminus, and residues

563 to 875 in the C-terminus, of the deduced protein from *E. coli carB*. All protein alignments generated were used as a template to produce a corresponding alignment of the DNA sequences.

2.11.3 Phylogenetic tree construction

Phylogenetic trees were produced using Phylip (Phylogeny Inference Package) version 3.5 (Felsenstein, 1993). Parsimonious trees were compiled using the dnaps and protpars methods of Phylip. Distance matrices were developed using dnadist with the Kimura 2-parameter model, and protdist using the categories model of Hall. Trees were constructed from distance matrices using Fitch, Kitsch and Neighbor-joining methods. All bootstrap analysis was done using 100 multiple data sets with a jumbling factor of 1 ($J=1$). A jumbling factor of 50 was used when no bootstrapping was performed. The maximum likelihood method was also employed, using Molphy (maximum likelihood inference of protein phylogeny) version 2.1.2 (protml version 1) with automatic default settings (Adachi and Hasegawa, 1993).

3. RESULTS

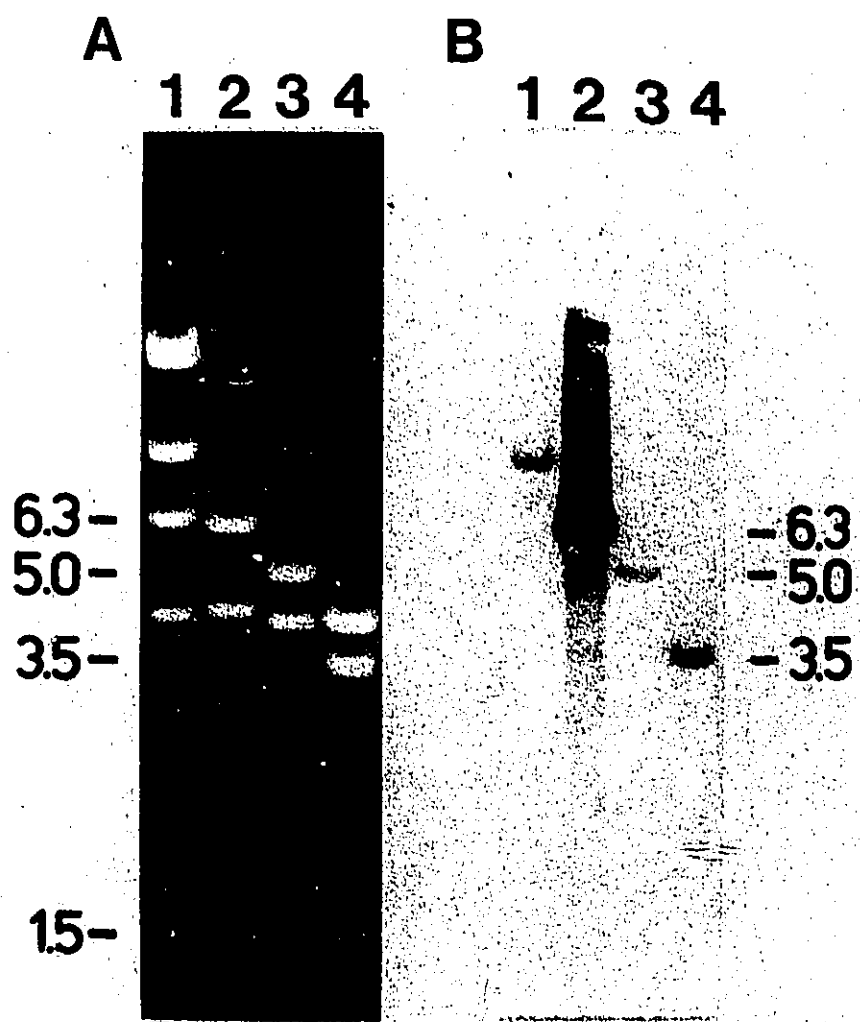
3.1 LOCATING AND SUBCLONING GONOCOCCAL DNA FRAGMENTS CONTAINING *CARA* AND *CARB*

EFP10, a lambda clone containing DNA from *N. gonorrhoeae* strain CH811, was previously shown to complement *E. coli carA* and *carB* mutants (Picard & Dillon, 1989b). A 5 kb *EcoRI* fragment from EFP10 was also shown to hybridise at high stringency to a clone (pMC50) containing the *E. coli carAB* genes (Picard and Dillon, 1989b). This 5 kb fragment was cloned into pBR329 to produce pSTD18 (Dillon and Bigelow, unpublished data; Figure 5).

In order to locate the *carA* and *carB* genes on these clones, direct sequencing was the preferred method, however EFP10 and pSTD18 were not suitable templates for sequencing and so further subcloning of DNA fragments was necessary. pBluescriptII KS(+) was chosen as a suitable vector and initially it was thought that the 5 kb *EcoRI* fragment from pSTD18 would be the best candidate for subcloning. However, there was concern that cloning of this 5 kb fragment would be difficult, both because of its size and the fact that its ends were generated using one restriction enzyme. So restriction enzyme digests of pSTD18 were probed with *E. coli carAB* (a 6.3 kb *HindIII* fragment from pMC50 which contains the cloned *E. coli* genes; Figure 3) in the hopes of finding a smaller fragment which may still contain *carA* or *carB* gene sequences.

One such fragment was found. A 3.5 kb *EcoRI XhoI* fragment hybridised to the probe comprising *E. coli carAB* at 62°C (Figure 7). However, *carA* and *carB* in other prokaryotes are approximately 1.1 kb and 3.2 kb in size, respectively, and the remaining 1.5 kb fragment from the insert of pSTD18 did not hybridise (Figure 7), therefore this was the first indication that maybe not all of

FIGURE 7. Hybridization of restriction enzyme digestions of gonococcal clones with an *E. coli carAB* probe. Figure 7A: Electrophoresis of the following DNA: Lane 1, λ -*Hind*III; Lane 2, pMC50-*Hind*III; Lane 3, pSTD18-*Eco*RI; Lane 4, pSTD18-*Eco*RI/*Xho*I. Figure 7B: Southern blot of 7A probed with the 6.3 kb *Hind*III fragment from pMC50 which contains the *E. coli carAB* operon (hybridization temperature: 62°C). Note that some hybridization with lambda sequences was observed. This is because the insert in pMC50 actually originated from a lambda clone and, as a result of the cloning procedure used, small segments of lambda DNA are still flanking the cloned *E. coli* DNA in the pMC50 clone.



carAB was present in the 5 kb insert of pSTD18. Nevertheless, obtaining some DNA sequence would indicate exactly where the genes were located on this fragment and within EFP10, and so the 3.5 kb fragment was subcloned into pBluescriptII KS (+) for sequencing. The resulting clone was named pFL35 (Figure 5). Concurrent attempts were also made to subclone the remaining 1.5 kb *EcoRI XhoI* fragment of pSTD18 into pBluescriptII KS(+) as well, however this was not pursued once the location of *carB* was confirmed, because this 1.5 kb fragment was then deduced to have no sequences of interest. The pFL35 clone was verified by checking that its insert hybridised at high stringency to the fragment in pSTD18 which it came from (Figure 8). The 3.5 kb insert of pFL35 hybridised to the probe containing the *E. coli carAB* genes at 62°C but not at 68°C, and hybridised at 68°C to gonococcal chromosomal DNA, but not *E. coli* chromosomal DNA (data not shown).

After initial sequence from pFL35 was obtained at the *EcoRI* end of the insert, alignment of this sequence with *E. coli carB* revealed that pFL35 (and pSTD18) was missing approximately the first 146 bp from the 5' end of *carB*, and all of *carA*. pFL35 was completely sequenced in both directions, and shown to contain 3091 bp of the gonococcal *carB* gene. In order to obtain the remaining ~146 bp of *carB*, and the *carA* gene, hybridisations were performed at lower stringency between the original clone EFP10 and a probe containing the *E. coli carA* gene (a 1.4 kb *ApaI NsiI* fragment from pMC50; see Figure 3 for location of restriction enzyme sites in the *E. coli* clone). It was found that at a lower stringency (55°C instead of 62°C) the *E. coli carA* gene did hybridise to another fragment within EFP10 (a 7.5 kb *EcoRI* fragment; Figure 9) which according to restriction enzyme mapping was adjacent to the 5 kb *EcoRI* fragment which was known to contain *carB* sequences. Again there was concern that cloning of such a large fragment (7.5 kb) would prove difficult and so further restriction enzyme digests of EFP10 were probed with *E. coli carAB*. An approximately 5 kb *XbaI EcoRI*

FIGURE 8. Confirmation that pFL35 contains the 3.5 kb *EcoRI/XhoI* fragment of pSTD18. Figure 8A. Electrophoresis of the following: Lane 1, λ -*HindIII*; Lane 2, pSTD18-*EcoRI*; Lane 3, pSTD18-*EcoRI/XhoI*, Lane 4, pBluescriptII KS(+)-*EcoRI/XhoI*; Lane 5, pFL35-*EcoRI/XhoI*. Figure 8B. Southern blot of 8A probed with the 3.5 kb *EcoRI/XhoI* fragment of pSTD18 (hybridization performed at high stringency: 68°C).

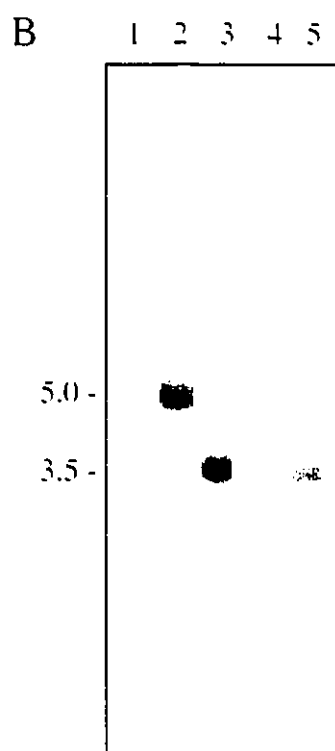
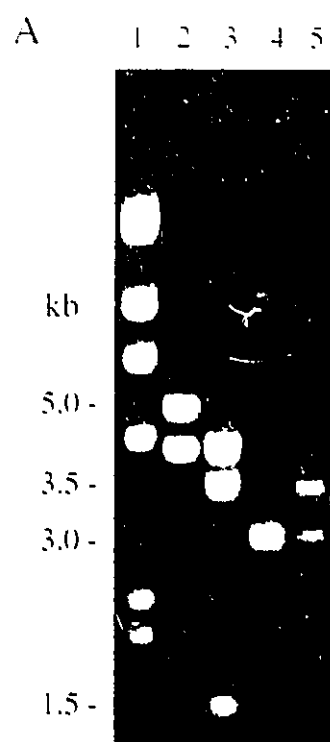
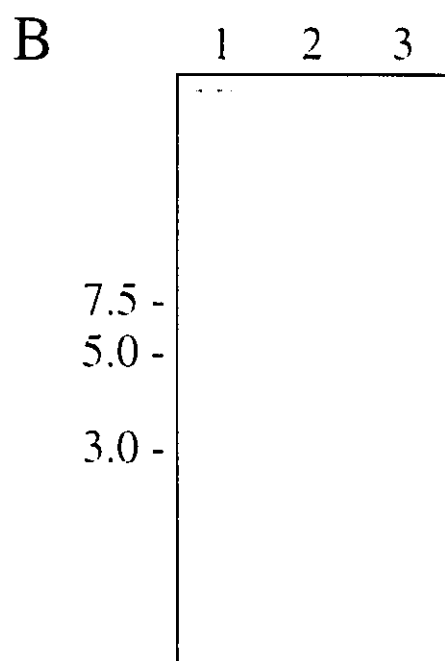
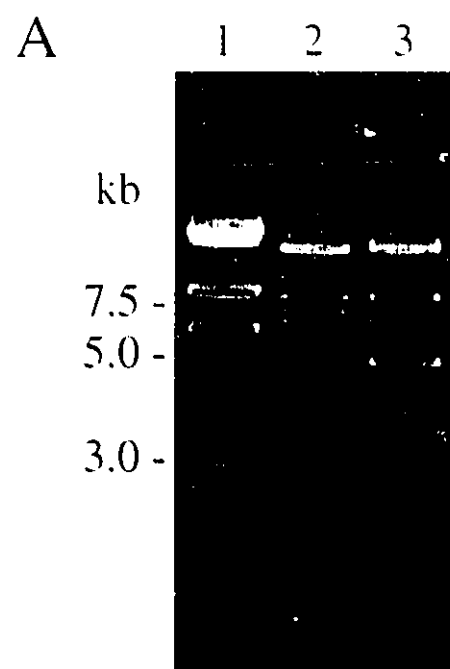


FIGURE 9. Hybridization of restriction enzyme fragments from the gonococcal clone EFP10 with *E. coli carA*. Figure 9A Electrophoresis of the following: Lane 1, λ -*Hind*III; Lane 2, EFP10-*Eco*RI; Lane 3, EFP10-*Eco*RI/*Xba*I. Figure 9B. Southern blot of 9A probed with the 1.4 kb *Apa*I/*Nsi*I fragment from pMC50 which contains *E. coli carA* (hybridisation performed at 55°C).



fragment, which hybridised to the *E. coli* genes, was found to be the smallest suitable fragment for cloning (Figure 9). This fragment was successfully cloned into pBluescriptII KS(+) to produce clone pFL50 (Figure 5). This clone was verified by checking that its insert hybridised at high stringency to both the fragment it came from and to *N. gonorrhoeae* CH811 chromosomal DNA (Figure 10). Notably, the 5 kb *Xba*I *Eco*RI insert from pFL50 hybridised to a fragment of the same size from gonococcal chromosomal DNA digested with *Xba*I *Eco*RI (Figure 10). This 5 kb fragment hybridised to the probe containing the *E. coli carA* genes at 55°C but not at 60°C (data not shown).

Initial DNA sequencing of pFL50 revealed that the first 146 bp of *carB* were found on this clone. However, sequences directly upstream of this region had no homology with *E. coli carA*. This was surprising because in all prokaryotes examined at the time (*E. coli*, *S. typhimurium*, *P. stutzeri*, and *B. subtilis*), *carA* and *carB* overlapped or were separated by up to 24 bp. In order to facilitate locating the *carA* gene on pFL50, deletion derivatives of pFL50 were constructed in both directions and probed with the DNA fragment containing *E. coli carA*. Hybridisation of clones containing deletions from the *Eco*RI end is shown in Figure 11. In this case, clones with inserts of 1.1 kb or larger did hybridise with the *E. coli carA* probe (Figure 11, lanes 3 to 13), while clones with inserts 0.5 kb or smaller did not hybridise (Figure 11, lanes 14 to 19). This would suggest that one end of *carA* is between 0.5 and 1.1 kb from the *Xba*I end of the insert of pFL50. All clones containing deletions from the *Xba*I end did not hybridise (data not shown). The smallest deletion obtained in this direction (from the *Xba*I end) was almost 2 kb, suggesting that *carA* was located within 2 kb of the *Xba*I end of the insert. Initial sequencing in that region did confirm the presence of *carA* sequences. The precise location of *carA* relative to *carB* was determined when the complete DNA sequence of both strands of pFL50 were completed (Figure 5; see below). Restriction enzyme digests of pFL50 and pFL35 were done using

FIGURE 10. Confirmation of the origin of the insert in pFL50 including hybridisation of this insert to a specific restriction enzyme fragment of gonococcal chromosomal DNA. Figure 10A. Electrophoresis of the following: Lane 1, λ -HindIII; Lane 2, EFP10-*Eco*RI; Lane 3, EFP10-*Eco*RI/*Xba*I; Lane 4, pFL50-*Eco*RI/*Xba*I; Lane5, Chromosomal DNA from *N. gonorrhoeae* CH811-*Eco*RI/*Xba*I; Lane 6, pFL35-*Eco*RI/*Xho*I. Figure 10B. Southern blot of 10A probed with the 5.0 kb *Eco*RI/*Xba*I fragment from pFL50 (high stringency: 68°C).

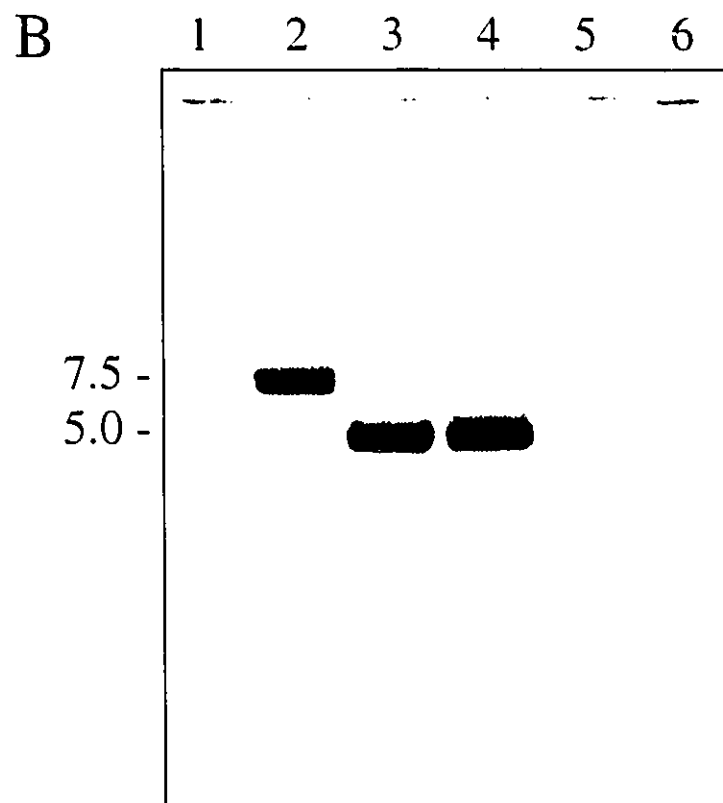
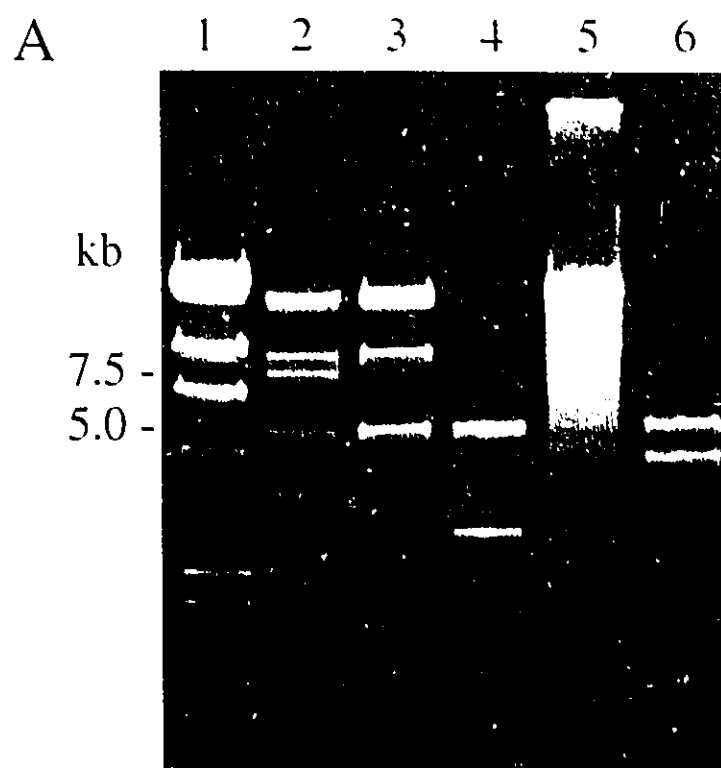
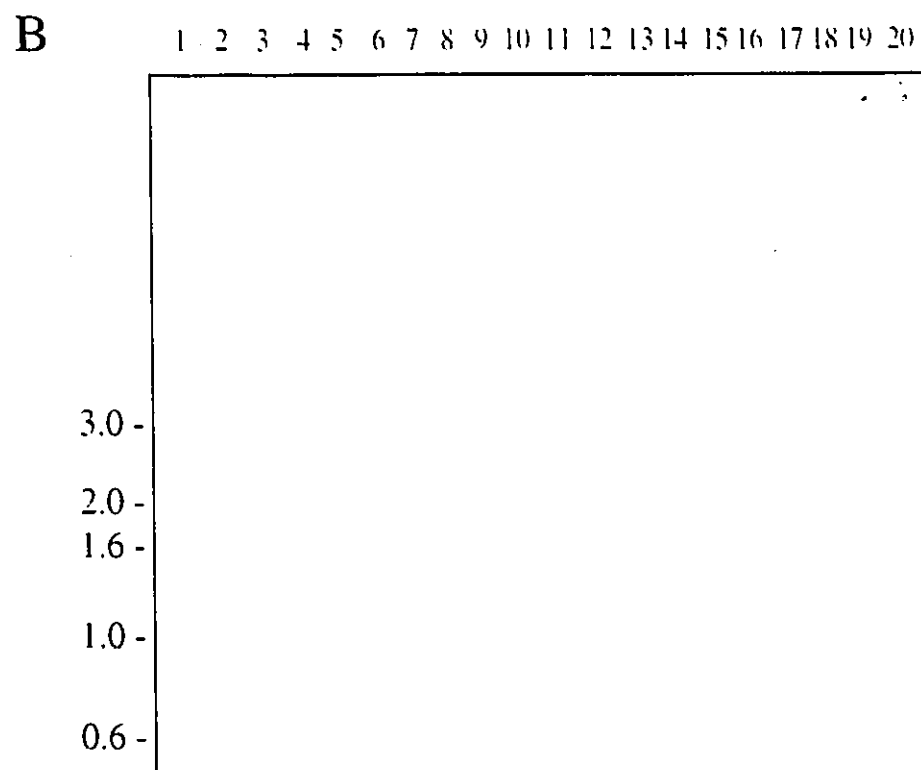
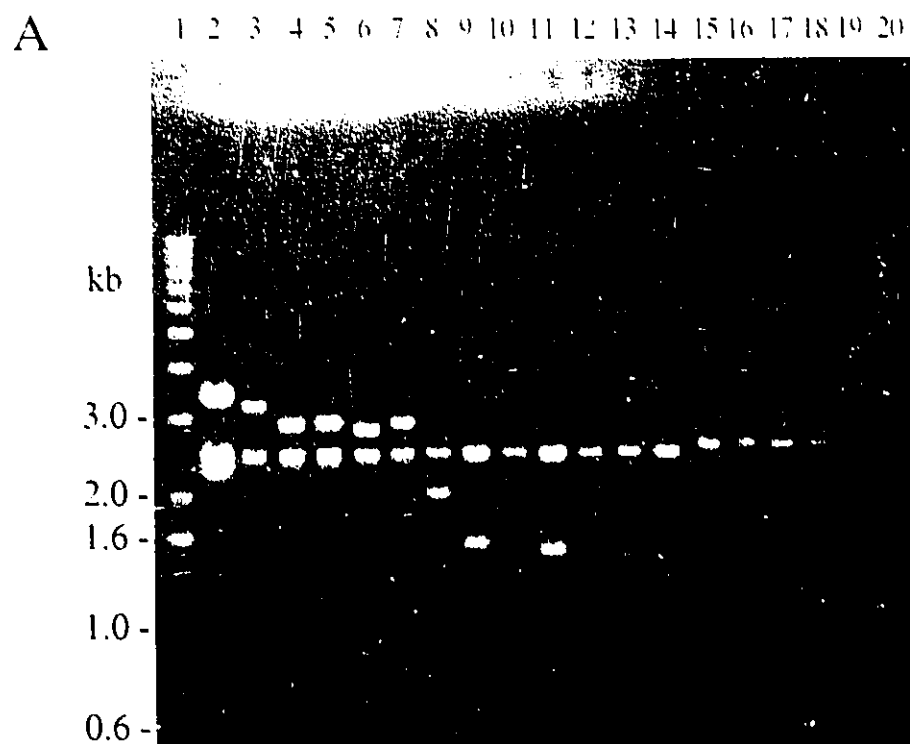


FIGURE 11. Hybridization of restriction enzyme digestions of deletion derivatives of pFL50 with *E. coli carA*. Clones shown are deletions from the *EcoRI* end of the insert of pFL50. Note that the clones were digested with *PvuII* which releases the insert, plus an additional 0.4 kb of vector, therefore the sizes of the inserts are actually 0.4 bp smaller than indicated by this digestion. Figure 11A: . Electrophoresis of the following: Lane 1, 1 kb ladder; Lane 2, pFL50-*PvuII*; Lane 3, EBC206-*PvuII*; Lane 4, EBC214-*PvuII*; Lane 5, EBC217-*PvuII*; Lane 6, EBC211-*PvuII*; Lane 7, EBC203-*PvuII*; Lane 8, EBC234-*PvuII*; Lane 9, EBC215-*PvuII*; Lane 10, EBC216-*PvuII*; Lane 11, EBC189-*PvuII*; Lane 12, EBC205-*PvuII*; Lane 13, EBC228-*PvuII*; Lane 14, EBC227-*PvuII*; Lane 15, EBC219-*PvuII*; Lane 16, EBC232-*PvuII*; Lane 17, EBC231-*PvuII*; Lane 18, EBC204-*PvuII*; Lane 19, EBC209-*PvuII*; Lane 20, 1 kb ladder. Figure 11B. Southern blot of 11A probed with the 1.4 kb *ApaI/NsiI* fragment from pMC50 which contains *E. coli carA* (hybridisation performed at 55°C).



various enzymes to help confirm the obtained DNA sequence for these clones. All restriction enzyme digests produced fragment sizes as predicted by the DNA sequence obtained (data not shown).

I supplied DNA fragments containing these cloned *carA* and *carB* genes to two different research groups who mapped these genes to *Neisseria spp.* chromosomes. *carB* was mapped to the chromosome of *N. gonorrhoeae* strain MS11-N198 by Dr. Carol Gibbs (Bihlmaier *et al.*, 1991; Gibbs, personal communication), and both *carA* and *carB* were mapped to the chromosome of *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strain Z2491 (hybridisation stringency: 60°C) by Ms. Jo-Anne Dempsey and Dr. Janne Cannon (University of North Carolina; Dempsey and Cannon, 1994; Dempsey *et al.* , 1995; Dempsey, personal communication; Figure 2). In all cases the genes mapped to the same single area, and they did not map close to any previously mapped biosynthetic genes, notably any other arginine biosynthesis genes (Figure 2).

3.2 ANALYSIS OF THE SEQUENCE OF pFL50 AND pFL35: *CAR*A AND *CAR*B ARE SIMILAR TO OTHER CPSII GENES, BUT ARE SEPARATED BY A UNIQUE INTERGENIC SEQUENCE AND SURROUNDED BY MANY REPETITIVE SEQUENCES

The complete DNA sequence of pFL50 and pFL35, containing the complete sequence of the *carA* and *carB* genes of *N. gonorrhoeae* CH811, was submitted to GenBank (accession number U11295) and is shown in Figure 12. In this sequence, *carA* (1125 bp) and *carB* (3237 bp) are found in the same orientation and are separated by 3290 bp. An additional 876 bp upstream of *carA* and an additional 456 bp downstream of *carB* has been sequenced from these clones. A schematic diagram of the complete sequence has also been illustrated in Figure 6.

3.2.1 *carA* and *carB* in *N. gonorrhoeae* CH811 encode a protein very similar to previously characterized CPSII enzymes.

DNA sequence analysis of the *carA* and *carB* genes of *N. gonorrhoeae* CH811 shows they have a GC content and codon usage similar to the other gonococcal genes (West & Clark, 1989) notably *argJ* (Martin & Mulks, 1992) and *argF* (Martin *et al.*, 1990; data not shown). Also primary sequence analysis indicates that these gonococcal CPS genes are similar in size and sequence to the corresponding genes of other prokaryotes. *CarA*, with an ATG start codon and TAA stop codon, encodes a protein of 375 amino acids (predicted molecular weight: 40551). *CarB*, with an ATG start codon and TGA stop codon, encodes a protein of 1079 amino acids (predicted molecular weight: 119614). Alignment of the deduced protein sequence from *carA* with deduced CPS sequences from the corresponding genes from other prokaryotes shows that the gonococcal sequence is equidistantly related to the pseudomonad and enteric homologues, having approximately 65% similarity with *P.*

FIGURE 12. The complete sequence of pFL50 and pFL35, comprising the *carA* and *carB* genes of *N. gonorrhoeae* CH811. The sequence of the *carA* and *carB* genes is highlighted in blue text and putative ribosome binding sites upstream of these genes are circled. Relevant restriction enzyme sites are in bold black text, with the exception of the *EcoRI* site at the border between pFL50 and pFL35 which is specially highlighted in dark blue. Gonococcal uptake sequences with no more than one mismatch with the consensus sequence are surrounded by yellow boxes, with mismatches indicated in dark/bold text. Arrows denote inverted repeats. One inverted repeat which contains an internal inverted repeat within it (at 448-547 bp), is marked with different arrowheads for clarity. Copies of the RS3 repeat sequence are highlighted in green text and copies of RS5 and RS2 are shown in black italics and yellow (non-italics), respectively. A polyT region upstream of *carA* is underscored with a dotted line. Copies of the novel repeats RS6 and RS7, found in the intergenic sequence between *carA* and *carB*, are marked in pink and purple text, respectively.

XbaI
1 TCTAGACATT CAATGCTAAG GCAATTTATC GGAATGACT GAAACTCAAA AAACCGGATT **RS3**
RS3
61 CCCACTTTCG TGGGAATGAC GAATTTTAGG TTTCTGATT TGGTTTTCTG TTTTCGTGGG
121 AATGACGGTT CACGCGAAAA TCCCGTCCGA AAGGTTTTCG GACGGCATT TTTGCGTTTT
RS5
181 TCGGGAGAGG GCGGGTTCGT ATAGCGGATT AACAAAAATC GGGACAAGGC GGCGGGCCGC
RS3
241 AGACAGTACG AATGGTACGG AACCGGTTCTG CCCGGTGCTT GGGCGCCTTA GGAATCGTT
301 CCCTTTGAGC CGGGGCGGGG CAACCCGTAC CGGTTTTTGT TAATCCGCTA TAAAGGCGG
361 GCTACAGGGT AGGCTTCATC CCGCCAATCT CAACTGAATC CGTCAATTC CGCAATTCAA
RS3 **RS3**
421 TTAAATACCG TCAAACCGAT GCCGTCAATT CCCGCGCAGG CGGGAAT
481 CCGTGCCGAC AGGTCCGATT CCCGCTGCG **RS3**
RS3
541 CGGGAATGAC GGTTCGGGTA TTCCCATAGT TTCGCCGGC GGACGTGGG AAATGCGTAA
601 CGGGCATAGG GGGCGCGAG CGGGCGGTTT TATGCCCCG ATTTCGTTT TCGCGCGAAC
661 ATATCAGCCC GCCTGCCGCG TTTGCGCTTG AAATCGGGTA TGTTTCGTCT TAAATATGC
721 TGCTTTCAGG GTATAGGCAC TTGCCGAAA AGCACGTTAC GCGTCTATCT TGCGCGCGT
781 GTTTTTTTTT GACCGGATTT TTCCGACCGG ATGCCCCCTG CCGAAGTCCC TTCAGACGGC
RS3 **carA**
841 ATTGTCAAGA ATTTTATTAA AAACAGGATT CCCATCATGA ACACCCCGC CCTCCTCGTC
901 CTCGCTGACG GCAGCGTATT CCACCGGACA TCAATCGGTT ACGAAGGTT GGCTTCCGGC
961 GAAGTCGTGT TCAATACTTC GATGACCGG TATCAGGAAA TCCTGACCGA CCCGTCCTAC
1021 TGCAAACAAA TCGTTACCTT CACCTACCCC CACATCGGCA ATGCCCGCAC CAACGCCGAA
1081 GATGAAGAAA GCCGCACGGT TTTACCGGCA GGCCTGATTA TCCGCGACCT GCCGCTCTTG
1141 CATAGCAACT TCCGCGCCTC CGAAAGCCTG CACGACTATT TGGTACGCAA CGAAACCGTC
1201 GCCATCGCCG ACATCGACAC CCGCCGCTG ACCATGCTGT TGCGCGAAAA AGGCCAAGGC
1261 GGCGCGATTG TGACAGGTGC GGATGCCACC GTTGAAAAAG CGCAAGAACT CATCGCCGCA
1321 TTCGGCAGCA TGGTCGGCAA AGATTTGGCA AAAGAAGTTT CCTGCACGGA AACTTACGAA
1381 TGGACGGAAG GCGAATGGGA ATTGGGCAA GGTTCGTTA CCCCTGACAA ACAGCCTTAC
1441 CACGTCGTCG CCTACGATTT CGGCGTGAAA ACCAACATCC TGCGTATCTC GCCTCGCGGC
1501 TGCCGCTGA CCGTCGTCCC CGCACAACG AGCGCGGAAG ACGTGTGGC ACTCAATCCC
1561 GACGGCGTGT TCCTGTCCCA ACGGCCCGG GACCCCGAGC CTTGCACCTA CGCCATCGAA
1621 GCCGTGCAAA AACTGATGGA AAGCGGCAA CCGATTTTCG GCATCTGCTT GGGACACCAG
1681 CTCATCAGCC TCGCCATCGG CGCGAAAACC CTGAAAATGC GCTTCAGCCA CCACGGTGCG
1741 AACCATCCTG TGCAAGATTT GGACAGCGG AAAGTCGTCA TTACCAGCCA AAACCACGGT
1801 TTCGCCGTCG ATGCCGACAC CCTGCCCGCC AACGCACGCA TTACCACAAA ATCCTTGTTT
1861 GACAATACCT TGCAAGGCAT CGAGCTGACC GACAAACCCG TGTTCTGTTT CCAAGGACAC
1921 CCCGAAGCCA GCGCGGTCC GCAAGATGTC GGCTATTTGT TCGACAAAT CATTTGCAAT
1981 ATGAAGGCTG CAAAACAAGC ATAATGATTT TCAGACGGCA ACAGTATGCT GCTCCCGTCT
2041 GAAAGAGAAA GCTGGAAATG AAAATTAGCG CACTCGACCA TCTAGTACTG ACTGTTGCCG
2101 ACATTGACCG AACCATCGCG TTTTACACAC AAGTTTGGG CATGGAAGAA GTTTCATTTG
HindIII
2161 GCAACAACCG TAAAGCTTTG TTGTTTGGCA GTCAGAAAT CAACCTACAC GGGCGCGGTG
2221 CGGAAATTCA GCCTAACGCG CAACACGCCG CCTGCGGCAC AGCGGATTTA TGCTGTCTGA
2281 CCGATACGCC CCACTGGAAA CGGTTTTACA GGAATTATCC GCACACGGCA TCAAACCTTT

2341 AAGCGGCATC GTAGCGCGCA CAGGCGCAAT GGGCAAATC CAATCGGTTT ACCTGCGCGA
 2401 TCCCGATGGC AACCTGCTGG AAATCAGCAG TTATTAATTT TCAGACGGAT TATGCAAAAT
 2461 AAAAAACAGC CTGCACAAGC TGTTTTCTTT GCAGCCCCTT TAACCCCAAC AGCCGACCCG
 2521 TCCCCTCTCC CTGTGGGAAA GAGTTAGAGA GAGGGCAACA AGCTGTAAGG CTTACACAAA
 2581 CAGCAACCCG ATAACAAAAA TGAACCCGCC CGAGAACTA TTGACCGCCG AAAACCCCGC
 2641 CCTGCGCCAA CGCGCCAAAG CCATGCGTCA AGAAATGAGC GAGGCGGAAG CAAAATTGTG
 2701 GCAACACCTG CGGGCAGGCC GTCTGAAAGG CTATAAATTC CGCCGACAGC AGCCGATGGG
 2761 GAATTATATT GTTGATTTTA TGTGCGTAAC GCCCAAGCTG ATTGTGGAAG CAGACGGCGG
 2821 GCAGCACGCG GAACAAGCCG TATACGACCA CGCGCGGACG GTATATCTCA ACAGCTTGGG
 2881 CTTTACCGTG CTGCGTTTTT GGAATCACGA AATTTCAGC AGACAAACGA TGTACTGGCG
 2941 GAAATCCTGC GCGTATTGCA GGAATTGGAA AAGCAGCCAG CACGGTAGCA AACGGTTGAT
 3001 TTTGATTAGA TTTTGAAAAA AATAAGGTAA AATTAAGTGA ATTAAGCGTT TAAAAAGTTC
 3061 GTCTGAAAG CAAAAAGCA ACCTGCACAA CCTGTTTTCC TTGCAGCAA CCTCAACCCC
 3121 AACAGCCGAC CCGTCCTCTC TCCCTGTGGG AGAGAGCTAG AGAGAGGGCA ACAAGCCGCA
 3181 AGGCTTGAT TTGGGCGGCG AAGGTATAG GAAAGATTGC CACCATTGGG AGAATGCCCT
 3241 CTCCCCAGCC CTCCCCACA GGGGAGGGAG TGGGTTTCGG CAGATTTTTC GGTTGCAGGC
 3301 GATTTGAAAG GCAACTTAGA TTTGCAGCTG TTGTTTCAAG TCGTCTGAAA AATAAAAAGC
 3361 AACCTGCACA ACCTGTTTTT CTGTCAGCAA ACCTCAACCC CAACAGCCGA CCCGTCCTCT
 3421 CTCCCGTGG GAGAGAGCTA GAGAGAGGGC AACAAGCCGC AAGGCTTGTA TTTGGGCGGC
 3481 GAAGGTATTA GGAAAGATTG CCACCATTGC GAGAATGCCC TCTCCCCAGC CCTCCCCAC
 3541 GGGGAGGGA GTGGGTTTCG GCAGATTTT CGGTTGCCGG TGGTTTGAAA AATAACTTAA
 3601 ATTTGCAACC GTTGATTCA GTCGCTGA AAAATAAAAA GCAACCTGCA CAACCTGTTT
 3661 TCCTTGCAGC AAACCTCAAC CCCAACAGCC GACCCGTCCT CTCTCCCTGT GGGAGAGAGC
 3721 TGGAGAGAGG GCAACAAGCC GCAAGGCTTG TATTGGGCG GCGAAGGTAT TAGGAAAGAT
 3781 TGCCACCATT CGGAGAATGC CCTCTCCCCA GCCCTCCCCC ACGGGGAGG GAGTGGGTTT
 3841 CGGCAGATTT TTCGGTTGCC GGTGGTTTGA AAAATAACTT AAATTTGCAA CCGTTGATTT
 3901 CAGGTCGTCT GAAAGAAT GCACGAAATA TCAACAGCGG GAATTTTCA GGTAGCCTTT
 3961 ATTGTCAAGC AGATGGAACA AACGCCGCGA GCGTTTTTTC AGACAATCTT TGAATCATC
 4021 GGCAGAGAGT GTGCCGAAG GCGCGCATGC GGTGGGTTGG GGTACAGGG AAAATGAAGA

4081 ACGCGTGCCT GCGTACCGCA CATACCCTAC ATGCGGGCTA CGGCTTGCTA GGATTGCTCG
 4141 TTATACTTAA TTTTGACAAG GAGTCAATGA AATGGCGATA GAACTAAATG GTAACCTGGAC
 4201 AAAAGGATTT GCTTTAGACC ATCATATGCA AAAAAGTATT TTGTTAGGTT ATGATGAATT
 4261 AGGGCATCCT CACTTTGATA GCAAGAGAAG TGTGATTGGA GAATGGGTTT ATCAATTAAA
 4321 ATATCGTAAT CAAACCCAAA ATGTTTCTTT GCTGGTAGAC TATATTTTAC AAAAATTTAG
 4381 TGGTTTAGAA TCTATAAATT TAATTGTTCC CGCCCCTTTT ACAACAGAGC GTATTAATCA
 4441 ACCAGTTCAA TTAATTGCCA AAGAACTAAG TGATCGCTTG AATATTCCAT ATTCTCCTAT
 4501 TTTAAGAAAA TCTAGCTCAC ACACACCACT CAAAAATATT GAAGAAAAAT CAGAAAAACT
 4561 AGAAATACTG AATAATTCCA TCACAATTGA CAATATAGAT CTATCCAATA AGAATATTTT
 4621 AGTAATCGAT GATTTATTTG ATAGCGGAGC AACATTGGAA ATATCAACTG AAAAATTGTT
 4681 TTCAAAAAAT GCTAGATCTG TAATTGTTTT AGCAATGACA AAGACTAGAG GATAAAAAATG
 4741 CGCACTATTT TCTTTTCTGG TTCACGCAGC ATTCACGAC TAAATCCACA AATACGAGAA
 4801 AGAATTAACA ATATTCTTAG CAATAATTTT GATATTGTTA TTGGAGATGC TAATGGGGCT
 4861 GATAAAGCTA TCCAAGAATT TCTACAAGAA CAAGATTATG CTAATGTGCA TATTTACTTT
 4921 TCTGGCAAAA TTTACCGAAA TAATGTGGGA AATTGGCAGT TTGTCCAAGT AGACAGCAAA
 4981 GGAAGTGAAG GGTATTTTAT ACAGCAAAAG ATAAGAAAAT GGCAGAAATT GCTGATTATG
 5041 GCTTTATTTT ATGGGATGGT AAGAGTATTG GTTCATTAAA TAATATAGCA GAACTTTTAC
 5101 AGCTAAATAA ACCTTCTCTT GTATACCATT CTCAAACAAA **EcoRI** AGAATTCTTT AAAATTAAT
 5161 CAAGCGCAGA TTTAGAGAAT ATTTTATCTA ACATAGAAGA TGATGTTTTA GCTTCAATTT
 5221 TAGAAAAGGG AAATACTTTT CTTAAAAGCT ATGTAACAAA ACAACCTTCT TTAATTCAGG **RBS**
 5281 ~~AGTA~~ACCACC CATGCCCAA CGTACCGACC TAAATCCAT CCTTATCATC GGCGCCGGCC
 5341 ~~CTATCGTTAT~~ CGGTCAGGCC TGCGAATTTG ACTATTCGGG CGCACAGGCC TGCAAAGCCT
EcoRI site at border between pFL50 and pFL35
 5401 TGCGTGAAGA AGGCTATAAA GTCATTTTGG TGAATTCCAA CCCC GCCACG ATTATGACCG
 5461 ACCCCGAAAT GGCGGATGTT ACCTACATCG AGCCGATTAT GTGGCAGACG GTGGAGAAGA
 5521 TTATCGCCAA GGAGCGGCCC GATGCGATTG TGCCACGAT GGGCGGTACA GACCGGCTGA
 5581 ACTGTGCGCT GGATTGGGCC GGTAACGGCG TGCTGGCGAA ATACGATGTC GAGTTAATCG
 5641 GCGCAACGGA AGACGCGATC GACAAGGCGG AAGACCGGG CCGCTTTAAA GAAGCGATGG
 5701 AAAAAATCGG CCTCTCTTGC CCGAAATCTT TGTTCTGCCA CACCATGAAC GAAGCCTTGG
 5761 CGGCGCAAGA ACAGGTGCGC TTTCCGACGC TGATTGCTCC GTCTTTCACG ATGGGCGGTT
 5821 CGGGCGGCGG CATTGCCTAC AATAAGGATG AGTTTTTGGC GATTTGCGAA CGCGGTTTCG
 5881 ATGCGTCGCC TACGCATGAG CTGCTGATTG AGCAGTCTGT GCTCGGCTGG AAAGAGTACG
 5941 AGATGGAAGT GGTGCGCGAT AAGGCGGACA ACTGCATCAT CATCTGTTTC ATTGAAACT
 6001 TCGACCGGAT GGGCGTTCAT ACGGGCGACT CGATTACGGT TGCGCCGGCC GAAACGCTGA
 6061 CGGACAAGGA ATACCAAATC ATGCGCAACG CTTTCGTTGGC GGTATTGCGC GAAATCGGCG
 6121 TGGACACGGG CGGCTCGAAC GTGCAGTTTG CGGTGAACCC TGAAAAGCGG CGAGATGATT
 6181 GTGATCGAGA TGAACCCGCC GCCGTGAGCC GTTCGTCCGC GCTGGCTTCC AAAGCAACGG
 6241 GCTTCCCGAT TGCGAAGGTG GCGGCGAACT GGGCGGTCGG CTTTACGCTG GACGAGTTGC
 6301 GCAACGACAT CACCGGCCGC CGCACGCCCG CGTCGTTTCA GCCTTCCATC GACTATGTGG
 6361 TAACCAAAAT GGCCCGTTTC GCGTTTGAAA AATTCCCCGC CGCAGACGAC CGCCTGACCA

6421 CGCAGATGAA ATCTGTGGGC GAAGTAATGG CGATGGGCCG TACCATTGAG GAAAGCTTCC
6481 AAAAAGCCTT GCGCGGTTTG GAAACCGGCT TTGTGCTGGT TTCAATGCCG GCCAGGCGAA
6541 GACAAACGGA AATCCGCCGC GAACTGGCCA ACCCGGCCCG CGAACGTATG CTGTTTGTGG
6601 GCAAAGCCTT CCGCGCGGCC TTCACGCCCG AAGAAATCCA CGAGATCTCG GCCATCGACC
6661 CTTGGTTCTT GGCGCAAATC GAAGACTTGA TGAAGGAAGA AAAATCGGTI TCAGACGGCC
6721 AGTTGCAAGA TTTGGATTAC GCCGCCCTAC ACCGTCTGAA ACGCAAAGGC TTCTCCGACA
6781 AACGTTTGGC ACAATTGTTG AACGTAAGCG AAAAAGAAGT TCGCGAAACC CGTACACCGG
6841 TGAAGCTCGA TCCGGTTTAC AAACCGGTCG ATACCTCGGC CGCCGAGTTC GCCACCGAAA
6901 CCGCCTATCT TTAATCCACT TACGAAGAAG AATGCCGAAT CCGTCCCTT CCCGACCGTA
6961 AAAAAGTGAT GATTCTCGGC GCGGGCCCGA ACCCGATCGG TCAGGGCATC GAGTTTGAAT
7021 ACTGCTCGGT TCACGCCGCA CTGCCCCCTGC GCGAATCGGG CTTTGAAACG ATTATGGTCA
7081 ACTGCAACCC CGAAACCGTG TCCACCGACT TCGACACCAG CGACCGCCTC TATTTGAGC
7141 CGCTGACATT GGAAGACGTG TTGGAAATCG TCCGCACCGA AAACCGGTGG GCGGTGATTG
7201 TTCATTACGG CGGTCAAACC CCGCTGAAAC TCGCCAACGC ATTGGTTGAA AACGGCGTGA
7261 ACATCATCGG CACATCCGCC GACAGCATCG ATGCCGCCGA AGACCGCGAA CGCTTCCAAA
7321 AAGTGTGAA CGACTTAGGC CTGCGCCAAC CGCCCAACCG CATCGCCAC AACGAAGAAG
7381 AAGCACTCGT CAAAGCCGAA GAAATCGGTT ATCCGCTGGT CGTGCGCCCG TCTTATGTTT
7441 TGGGCGGACC GCGGATGCAG ATTGTCCACT CCGCCGAAGC CTTGCAAAAA TACATGCGCG
7501 AACCGGTGCA GGTTCGCGAA GACAGCCCCG TGCTGCTCGA TTTCTTCCTC AACACGCGA
7561 TTGAAGTCGA TGTGGACTGC GTTTCAGACG GCAAGATGT GGTATCGGC GGCATCATGC
7621 AGCACGTCGA ACAGGCGGGC ATCCACTCCG GCGACTCCGG CTGCTCGCTG CCGCCATACT
7681 CCTTAAGCGA AGAAATCCAA GACGAAATCC GCGGCCAAAC CAAAGCGATG GCGTACGCGC
7741 TGGGCTGGT CGGACTGATG AACGTGCAGT TTGCCGTGCA GGACGCGGTG GTGTTGCTGT
7801 TGGAAGTGAA CCGCGCGCGC ACGCGTACCG TCCCCTTTGT CTCCAAAGCC ACCGGGCGAGC
7861 GCCTCGCCAA AGTCGGCGCG CGCTGCATGG CAGGCATTTC CCTGAAAGAA CAAGGCGTGG
7921 AAAAAGAAGT CGTCCCCGAT TTCTATGCCG TTAAAGAAGC CGTGTTCCTC TTCATCAAAT
7981 TCCCGGGCGT GGATACGATT TTGAATCCGG AAATGCGCTC CACCGGCGAA GTGATGGGCG
8041 TGGGCCGAAG TTTCCGCGAA GCGTACTACA AAGCCCAACT CGGCGCGGGC GAACGCCTCA
8101 ACCCGACCGG CAAAATCTTC CTCTCCGTGC GCGAAGAAGA CAAAGAACGC GTCATTAAAA
8161 CCGCCAAAAA CTTCCAAGCC TTAGGCTACG GGCATCTGCC CACCCGGCAC GCGCAATAC
8221 CTGACCGAAC ACGGGCTGAT TGTGCAGGCC ATCAACAAAG TCCCGAAGG CCGCCCGCAC
8281 ATCGGCGACG CGCTGAAAAA CGGCGAAATC GCACTGGTCG TAAATACCGT TTCCAGCGAT
8341 CCGCAATCCG TGTCCGACAG CCACATCATC CGCCAAAGCG CATTCGAGCA ACGCGTGCCG
8401 CAATACACCA CCACCGCCGG CGGCGAAGCG ATGAGCGAAG GCGCGAAAAG CCGGGATTAT
8461 CTGGGTGTGT ACAGCGTTCA AGAGCTGCAC GGTGTTTTGA AAAACCGCAA CTGATGCCTG
8521 AATCAGGTTG AACAATCCCG TCTGAAGCCG TTTTGCGGTT TCAGACGGCA TTTTGTGCA
8581 CATAGCCTTC ACTTTCTTTT GACCTTTAAG CACCCTAATT TGGGGCGGTG GATTATCCAA
8641 CTGCCGGACA TCCATAAAAA TACCGCCCGA ACCACGGTTC AGACGGCATC TGCCTTGCAA
8701 CGGACAGCAG CAGTTTAAAG CCCGCCCCCT TCGTTCCGCC CCAAAACCAT CGCATCTCCG
8761 TAGCTGAAGA AACGGTATTC ACGTTCAATC GCATGACGGT ACACGGCGCG GATATGACCC
8821 ATACCCGAAA ACGCGCCGAC CAACATCAGC AGCGTCGATT TCGGCAGATG AAAATTGGTA
8881 ACCAGCCTGT CGACAACATT AAAACGGTAG CCCGGCGTGA TGAAAATATC GGTGTGCCCC
8941 TGTCCGTCTT TCAAATATCC CGTTGCGCGC GCGGCAGACT CGAG

aeruginosa, *P. stutzeri*, *E. coli* and *S. typhimurium* CPS. The pseudomonad and enteric sequences also show approximately 65% similarity with one another. The gonococcal sequence shows 44% similarity with *B. subtilis* CPS (pyrimidine-regulated enzyme).

Alignment of deduced proteins from *carB* genes reveals that the gonococcal sequence has approximately 66% similarity with *E. coli*, and 46% similarity with *B. subtilis* (pyrimidine-regulated). The last 342 bp of *carB*, however, is notable for its lack of conservation between species and the gonococcal sequence is no exception: the deduced protein sequence of this region of the gene shares only 49% similarity with the corresponding *E. coli* sequence, while the remaining sequence from this gene actually shares 70% similarity with the *E. coli* sequence. Internal similarity is found between the first third and second third of the gonococcal *carB* gene (34% similarity at the protein level), similar to the initial report for *E. coli carB* (Nyunoya & Lusty, 1983).

Figure 13 shows the deduced sequence of gonococcal CPS with conserved domains I have identified by alignment with 6 prokaryotic and 7 eukaryotic CPS sequences. Simmer *et al.* (1990) also identified conserved domains, however they did not look at conserved domains using complete CPS gene sequences, since they were missing the 5' end of the hamster CPS gene sequence they were primarily investigating. In my analysis, highly conserved domains present in other CPS's, including the conserved domains reported by Simmer *et al.* (1990), are conserved in gonococcal CPS. The conserved residues include those previously found reactive in the catalysis of glutamine amide transfer which are found in all CPSII enzymes (Amuro *et al.*, 1985; Rubino *et al.*, 1986; Miran *et al.*, 1991; Figure 13). Regions of sequence thought to be involved in ATP binding (Post *et al.*, 1990), and present in all CPS enzymes, are also present in the deduced gonococcal sequence (Figure 13). Cysteine

FIGURE 13. Deduced amino acid sequence of the CPS subunits encoded by *N. gonorrhoeae* CH811 *carA* (Figure 13A) and *carB* (Figure 13B). Residues of groups of three or more which are conserved or are conservative substitutions in all known CPS sequences are boxed. Residues implicated in glutamine binding and found in all glutamine-dependent CPS enzymes (CPSII) are marked in red text. Red italics highlight the cysteine residue thought critical for formation of the γ -glutamyl thioester intermediate. Regions proposed to contain the two ATP binding sites are shown in blue text, and regions with similarity to glycine rich loops found important for catalysis in other ATP binding proteins are shown in green text. The two cysteine residues conserved in acetylglutamate-dependent CPS's (CPSI, CPSIII) are not present in the gonococcal CPS: residues in their corresponding location are marked in yellow.

A

MNTPALLVLADGSVFHGTSGIGYEGSASGEVVENTSMGTGYQEI
 LTDPSYCKDIVTTTYPHIGNAGTNAEDEESRTVLPAGLIIRD
 LPLLHSEFRASESLHDYLVARNETVAIADIDTRRLTMLLREKG
 QGGAILTGADATVEKAQELIAAFGSMVGKDLAKEVSCETETYE
 WTEGEWELGKGFVTPDKQPYHVVAYDFGVKTNILRISPRGCR
 LTVVPAQTS AEDVLALNPDGVFLSQRPGDPEPCTYAIEAVQK
 LMESGKPIFGICLGHQLISLAIGAKTLKMRSSHGANHPVQD
 LD SGKVVITSONHGEFAVDADTLPANARITHKSLFDNTLQIE
 LTDKPVFCFQGHFEASPGPQDVGYLFDKFIGNMKAAKQA

B

MPKRTDLKSILIIAGPIVIGQAEFDYSGAQACKALREEGY
 KVLVNSNPATIMTDPEMADVITYIEPIMWQVVEKIIAKERP
 AILPTMGGTDRLNCALDLAGNGVLAKYDVELIGATEDAIDKA
 EDPGRFKEAMEKIGLSCP KSLFCHTMNEALAAQEQVGFP
 TLI RPSFTMGSGGGIAYNKDEFLAICERGF DASP
 THELLIEQSV IGWKEYEMEVVRDKADNCIIICSIENF
 DMGVHTGDSITVAP AETLTDKEYQIMRNASLAVLRE
 IGVD TGGSNVQFAVNPEKRR DDCDRDEPA
 VSRSSALASKATGFPIAKVAANWAVGFTLDEL
 RNDITGR RTPASFEP SIDYVVT KMARFAFEK
 FPAADDRLTQ MKSVGEVMAMGRITIOESFO
 KALRGLETGFVLVSM PARRROTE IRRELANP
 GPERMLFVGKAFRAAFTPEEIH EISAIDWFLAQ
 IEDLMKEEKSVSDGQLQDLDYAALHRLKRKGFSD
 KRLAQLLN VSEKEVRETRTPVKLDPVYKEVDT
 SAAEFATETAYLYSTYEE ECRIPSLPDRKKVM
 ILGAGNP IGGIEFDYCSVHAALPLRE SGFET
 IMVNCNPETVSTD FDTSDRLYFEPLTLEDVLEI
 VRTE NPWGVIVHYGGCTPIKLANALVENG VNI
 IGT SADSIDAEDR ERFQKVLNDLGLRQPPN
 RIAHNEEEALVKAEETGYPLVVRPS YVLGG
 PAMQIVHSAEALOKYMRP VQVSEDSPVLLD
 FFLNNA IEVDVDCVSDGKDVVIGGIMQHVEQ
 AGIHSGDSGCSLPPYSL SEEIQDEIRROTK
 AMAYALGLVGLMNVQFAVQDAVVFVLEVN
 PRATRTVFEVSKATGORLAKVGARCMAGISLKE
 QGVEKEVVP D FYAVKEAVFPFIKFP
 GVD TILNEMRSTGEVMVGRSFG
 EAYYKAQLGAGERLNPTGKIFLSVREEDK
 ERVIKTAKNFQALGY GHLPTRHGAI
 PDRTRADCAGHQQSRRPPAHRRAEKRRNRT
 GRKYRFQRSAIRVRQPHHPKRIAATRAAII
 HHHRRRRSDERR REKPGLSGCVQRSRAARS
 FEKPQLMPESG

residues implicated in acetylglutamate binding, which are found in CPSI and CPSIII enzymes (Geschwill and Lumper, 1989), are not present (Figure 13).

3.2.2 *carA* and *carB* in *N. gonorrhoeae* CH811 are separated by a unique, large intervening sequence and, unlike other prokaryotes, putative transcription terminators are found downstream of both genes.

Although the CPS genes of *N. gonorrhoeae* CH811 are similar in size and sequence to those of other prokaryotes, their organization is quite different. Firstly, these genes are separated by an unusually large intervening sequence of 3290 bp. Most other prokaryotic CPS genes either overlap, or are separated by, up to 24 bp. Only *P. aeruginosa* PAO1 is an exception, having *carA* and *carB* genes which are separated by a 682 bp sequence which contains an unidentified open reading frame (Kwon *et al.*, 1994). However, there is no significant similarity between this 682 bp intervening sequence and the corresponding intervening sequence of *N. gonorrhoeae* CH811, according to a similarity search using Microgenie's "databank search" algorithm (set to recognize 7 or more similar nucleotides). In fact, the gonococcal intervening sequence has no significant similarity to any sequences found in GenBank and EMBL (as of December, 1995), according to a search using the program blastn.

One unusual and surprising exception was however identified. A 211 bp region of the gonococcal sequence near the end of *carA* was found to have 99% similarity to a region within a GenBank sequence (accession no. NGOFUSAA). This was further investigated and it was found that the sequence reported in NGOFUSAA actually comprised a composite of three gonococcal sequences previously shown to be in separate regions of the chromosome. This GenBank sequence (NGOFUSAA) is apparently the result of an error by Genbank staff who accidentally fused two

separate sequence submissions together in a mosaic fashion (Robert Belland, personal communication). This will not be mentioned further since the identity of the similar sequence region has not yet been confirmed. Recently another sequence of *carA* from *N. gonorrhoeae* has been submitted to GenBank by an independent group which confirms the sequence that I obtained (Rudel, 1995).

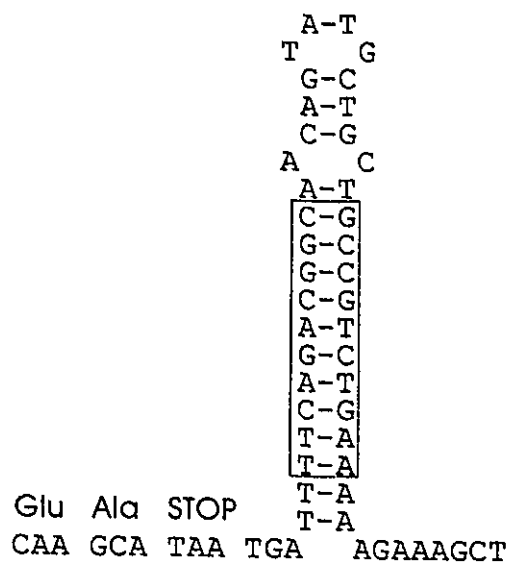
Ten ORFs, ranging in size from 75 to 560 bp, were identified in this gonococcal intervening sequence (only ORFs of 75 bp or more were considered for analysis), however, based on primary sequence analysis, no obvious function could be determined for these ORFs and none had a codon usage consistent with a gonococcal gene. The codon usage inconsistency is likely a reflection of the significantly lower GC content of this region; *carA* and *carB* both have a GC content of 57% while the sequence between *carA* and *carB* is 43% GC.

Not only is the intervening sequence between *carA* and *carB* in *N. gonorrhoeae* CH811 quite unique, but in all other previously characterized prokaryotes no transcription terminators are found between *carA* and *carB*. I have identified putative transcription terminators downstream of both the *carA* and *carB* genes of *N. gonorrhoeae* CH811. These terminators consist of inverted repeats containing the gonococcal uptake sequence with additional A or T residues (Figure 14). The terminators have an estimated ΔG of -27 (downstream of *carA*) and -24 (downstream of *carB*). Similar inverted repeats are often found downstream of other gonococcal genes, and they resemble the Rho-independent transcription terminators identified in *E. coli* (Elkins *et al.*, 1991).

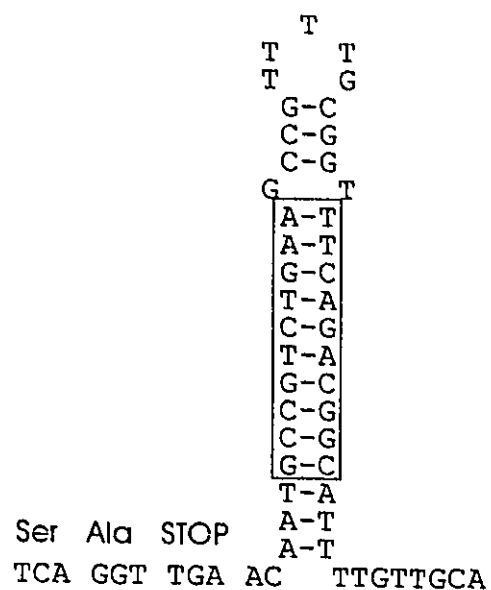
No obvious similarities are seen between the sequence upstream of *carB* and the sequence upstream of *carA* where promoter sequences may reside. No obvious similarities are also found between these sequences upstream of *carA* and *carB*, and sequences obtained upstream of the gonococcal *argF* and *argJ* genes which are involved in the same biosynthetic pathway. However, one interesting feature noted 88 bp upstream of *carA* was a string of 9 thiamine bases which would become

FIGURE 14. Proposed transcription terminators downstream of *carA* (Figure 14A) and *carB* (Figure 14B) in *N. gonorrhoeae* CH811. Gonococcal uptake sequences are boxed.

A



B



uracil's if transcribed (Figure 12). It is enticing to hypothesise that they could be involved in regulation by pyrimidines. No inverted repeats were found in this region, however, so no indications of an attenuation mechanism could be identified. A few possible σ -70 consensus promoter sequences were identified upstream of both *carA* and *carB*, however mapping of the transcription start site(s) is needed before any functional promoter sequences are identified.

3.2.3 Numerous previously characterised repetitive sequences are found close to *carA* and *carB* in *N. gonorrhoeae* CH811, particularly upstream of *carA*.

Three different previously characterised repeats (RS3, RS2 and RS5), commonly found near genes encoding gonococcal surface proteins and pilin, were identified in the 876 bp of sequence obtained upstream of *carA*. Some of these repeats contained significant secondary structure. The sequence 5'-GGGAAT-3', part of the RS3 class of repeats proposed by Haas and Meyer (1986), is found in no less than 10 copies upstream of *carA*, and often as part of inverted repeats (Figure 12). RS2, also initially identified by Haas and Meyer (1986), is found at position 462 to 510 of the sequence shown in Figure 12, surrounded by 21 bp inverted repeats. These 21 bp repeats each contain 8 bp inverted repeats containing copies of the 6 bp sequence of RS3. At position 200 to 352 of the sequence shown in Figure 12 is found a copy of the 152 bp repeat first reported by Correia *et al.* (1988). This sequence has been subsequently given a number of names including NR and Ng-rep. For more consistency with the naming of repeats found in *N. gonorrhoeae*, the name RS5 is proposed for this sequence (there are already currently repeats designated RS1 to RS4). This RS5 sequence contains 19 bp inverted repeats and one copy of the RS3 sequence.

Nine copies of the 10 bp gonococcal uptake sequence were identified within the complete 9 kb of sequence comprising *carA* and *carB*. Eight more copies containing one mismatch (these mismatches are not necessarily in the same location) were also found. Upstream of *carA*, at position 137 to 171 of the sequence shown in Figure 12, two copies of the gonococcal uptake sequence were found as part of 16 bp inverted repeats, a possible transcription terminator for an upstream gene. These gonococcal uptake sequences each contain one mismatch which are complementary so that the inverted repeat is conserved. Copies of the gonococcal uptake sequence were also associated with the putative transcription terminators found downstream of *carA* and *carB* as mentioned above. Most of the other copies of the gonococcal uptake sequence surround the novel repeats found in the intergenic sequence, as described below.

3.2.4 Novel repetitive sequences are found in the sequence between *carA* and *carB* in *N. gonorrhoeae* CH811.

Complex, novel, direct repeats, designated RS6 and RS7, were found in this sequence between *carA* and *carB* (Figure 6; Figure 12). RS6 is approximately 130 bp and RS7 is approximately 150 bp. Both RS6 and RS7 showed no significant homologies with sequences from GenBank or EMBL. A single copy of RS6 (copy 1 in Figure 12) was found 450 bp downstream of *carA*, associated with a gonococcal uptake sequence that is on the complementary strand and which has one mismatch. Three more copies of RS6 (copies 2, 3 and 4 in Figure 12) are found further downstream, each associated with a copy of RS7 (RS7 copies are designated A, B, and C in Figure 12). These three units of RS6/RS7 are each flanked by gonococcal uptake sequences and are grouped together. Note that these flanking gonococcal uptake sequences are not found on the complementary strand, as is seen with the

single copy of RS6 (copy 1). These gonococcal uptake sequences also have one mismatch, however it is not the same mismatch seen in the uptake sequence associated with copy 1 of RS6.

Complex secondary structure may form within these RS6/RS7 repeats. Within copies 2, 3 and 4 of RS6 are found 9 bp inverted repeats ($\Delta G = -15$). Within the copies of RS7 are seen inverted repeats of 14 bp (copy A) or 15 bp (copies B and C). These repeats are GC rich, with estimated ΔG 's of -24. These 9 bp and 14/15bp inverted repeats found within RS6/RS7 actually form part of a larger imperfect 40 bp inverted repeat (74% similarity, $\Delta G = -69$). In fact, these copies of RS6/RS7 are even roughly inverted repeats with themselves (59% similarity): the RS6 (copy 2), RS7 (copy A), and RS6 (copy 3) form one half of the inverted repeat, while RS7 (copy B), RS6 (copy 4) and RS7 (copy C) form the other half.

Alignment of the four copies of RS6 (Figure 15) shows that the isolated copy of RS6 (copy 1), which is closest to *carA*, shows the least similarity with the determined RS6 consensus sequence (88.9%). The other three RS6 sequences further downstream from *carA* (copies 2, 3 and 4) show increasing levels of homology with the consensus (98.3%, 99.2%, 99.2%, respectively). The last two copies of RS6 are identical except for a single change of an adenine to a guanine. The three copies of RS7 show a similar trend when aligned. Copy A, which is closest to *carA* (and associated with RS6 copy 2), shows 92.9% similarity with the RS7 consensus sequence, while copies B and C are identical.

These RS6/RS7 repeats were provided to Ms. Jo-Anne Dempsey (University of North Carolina) for mapping to the chromosome of *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strain Z2491 (hybridisation stringency: 60°C). In both cases they mapped only in one location and to the same area as *carA* and *carB* (Dempsey, personal communication).

FIGURE 15. Alignment of the four copies of RS6 (Figure 15A) and the three copies of RS7 (Figure 15B) with the derived consensus sequence. Mismatches are highlighted in red text and arrows mark the inverted repeats found within these sequences. Note that the copies were numbered according to the order in which they appeared when proceeding from *carA* to *carB*.

A

consensus 1 AA ATAAAAAGCAACCTGCACAACCTGTTTTCTTGCAGCAAACCTCAACCC 51
 copy1 1 AA ATAAAAAACAGCCTGCACAAGCTGTTTTCTTGCAGC CCCTTTAACCC 50
 copy2 1 AAGCAAAAAAGCAACCTGCACAACCTGTTTTCTTGCAGCAAACCTCAACCC 52
 copy3 1 AA ATAAAAAGCAACCTGCACAACCTGTTTTCTTGCAGCAAACCTCAACCC 51
 copy4 1 AA ATAAAAAGCAACCTGCACAACCTGTTTTCTTGCAGCAAACCTCAACCC 51

consensus 52 CAACAGCCGACCCGTCCTCTCTCCCTGTGGGAGAGAGCTAGAGAGAGGGCAA 103
 copy1 51 CAACAGCCGACCCGTCCTCTCTCCCTGTGGGAAAGAGTTAGAGAGAGGGCAA 102
 copy2 53 CAACAGCCGACCCGTCCTCTCTCCCTGTGGGAGAGAGCTAGAGAGAGGGCAA 104
 copy3 52 CAACAGCCGACCCGTCCTCTCTCCCTGTGGGAGAGAGCTAGAGAGAGGGCAA 103
 copy4 52 CAACAGCCGACCCGTCCTCTCTCCCTGTGGGAGAGAGCTGGAGAGAGGGCAA 103

consensus 104 CAAGCCGCAAGGCTT 118
 copy1 103 CAAGCTGTAAGGCTT 117
 copy2 105 CAAGCCGCAAGGCTT 119
 copy3 104 CAAGCCGCAAGGCTT 118
 copy4 104 CAAGCCGCAAGGCTT 118

B

consensus 1 GTATTTGGGCGGCGAAGGTATTAGGAAAGATTGCCACCATTCCGAGAATGCC 52
 copyA 1 GTATTTGGGCGGCGAAGGTATTAGGAAAGATTGCCACCATTCCGAGAATGCC 52
 copyB 1 GTATTTGGGCGGCGAAGGTATTAGGAAAGATTGCCACCATTCCGAGAATGCC 52
 copyC 1 GTATTTGGGCGGCGAAGGTATTAGGAAAGATTGCCACCATTCCGAGAATGCC 52

consensus 53 CTCTCCCCAGCCCTCCCCACGGGGGAGGGAGTGGGTTTCGGCAGATTTTTC 104
 copyA 53 CTCTCCCCAGCCCTCCCCACGGGGGAGGGAGTGGGTTTCGGCAGATTTTTC 104
 copyB 53 CTCTCCCCAGCCCTCCCCACGGGGGAGGGAGTGGGTTTCGGCAGATTTTTC 104
 copyC 53 CTCTCCCCAGCCCTCCCCACGGGGGAGGGAGTGGGTTTCGGCAGATTTTTC 104

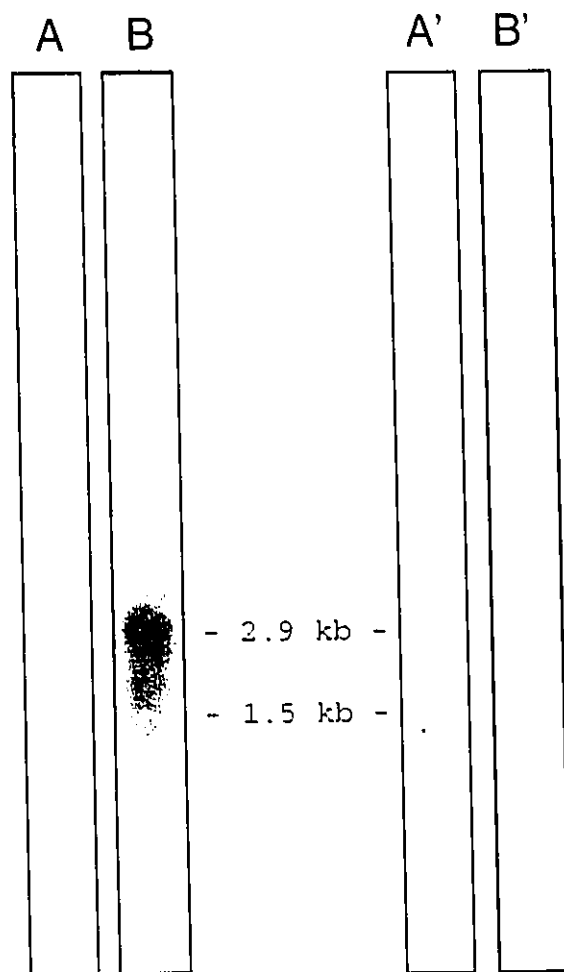
consensus 105 GGTGCGCGGTGGTTTGAAAAATAACTTAAATTTGCAACCGTTGATTTTCAG 154
 copyA 105 GGTGCGAGCGGATTTGAAAGGCAACTTAGATTTGCAGCTGTTG TTTTCAG 153
 copyB 105 GGTGCGCGGTGGTTTGAAAAATAACTTAAATTTGCAACCGTTGATTTTCAG 154
 copyC 105 GGTGCGCGGTGGTTTGAAAAATAACTTAAATTTGCAACCGTTGATTTTCAG 154

3.3 NORTHERN ANALYSIS: *CAR*A AND *CAR*B ARE SEPARATELY TRANSCRIBED

In order to maximize the amount of RNA produced containing *carA* and *carB* transcripts, it was first necessary to develop a minimal media which lacked the end products and intermediates of the arginine and pyrimidine biosynthetic pathway which may repress transcription of these CPS genes. Also, since many amino acids are contaminated with other by-products (*i.e.* ornithine is usually contaminated with citrulline and arginine), it was decided that a very minimal media, which was devoid of all amino acids except those needed for growth, would be most suitable. Various formulations adapted from the chemically defined media used for auxotyping were evaluated and it was determined that auxotyping media lacking all amino acids except for cysteine would effectively support growth. *N. gonorrhoeae* CH811 was grown on this media for total RNA extraction.

Northern blots of the RNA were separately probed with DNA fragments containing *carA* (fragment produced using primers 41 and 32; see Figure 6) and *carB* (primers 7 and 23; Figure 6). As seen in Figure 16, the *carA* and *carB* probes hybridized to differently sized RNA transcripts. The migration of these bands was compared with the migration of the ribosomal RNA bands which were visualized on the corresponding ethidium bromide stained gel. Based on the known sizes of these gonococcal ribosomal RNA bands (2.9 kb and 1.5 kb), the size of *carA* transcript was estimated to be about 1.3 kb, and size of the *carB* transcript was found to be approximately 3 kb. These sizes correspond approximately to the size of each gene, suggesting that promoters are located immediately upstream of both genes. This is particularly notable for *carB* because it indicates that there are no other products within the intergenic sequence which are co-transcribed with this gene.

FIGURE 16. Northern blots of *N. gonorrhoeae* CH811 total RNA, probed with fragments of the cloned gonococcal *carA* (Figure 16A) and *carB* (Figure 16B) genes (see text for details of probe production and bacterial growth conditions). Computer enhanced contrasting of the same figure, shown to help discern the bands, is displayed in Figures 16A' and 16B'. Size markers indicate the location of the 2.9 kb and 1.5 kb ribosomal RNA bands which were visualized in the corresponding ethidium bromide stained gel (not shown).



3.4 A VARIABLY-SIZED INTERGENIC SEQUENCE IS PRESENT BETWEEN *CAR*A AND *CAR*B IN *N. GONORRHOEAE* AND OTHER *NEISSERIA* SPECIES

3.4.1 PCR amplification of the intergenic sequence from gonococcal isolates: it varies in size.

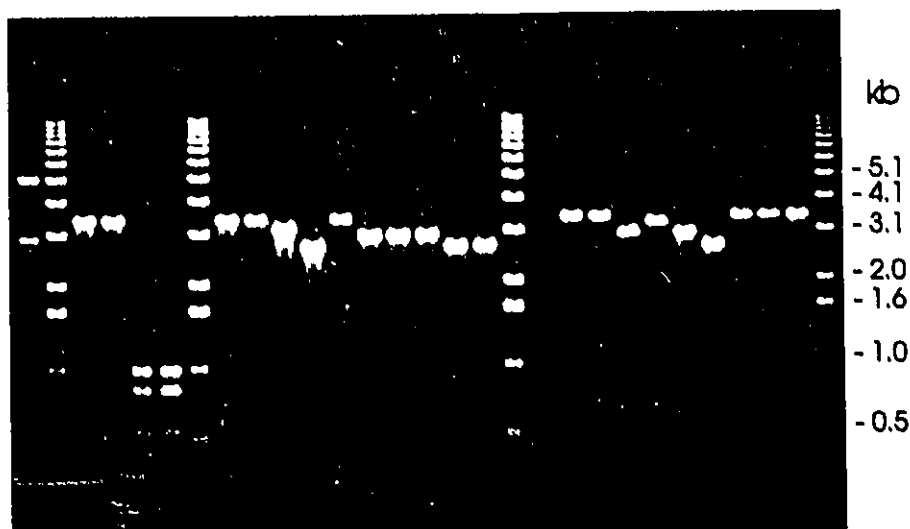
In order to determine the universality of the intervening sequence between *carA* and *carB* in other gonococcal isolates, a PCR protocol was developed to amplify this sequence using primers homologous to the end of *carA* and the beginning of *carB* (Primers 61 and 62). The amplicons produced from a template of pFL50 DNA and *N. gonorrhoeae* CH811 chromosomal DNA were found to be the same size, approximately 3450 bp, corresponding with the size (3473 bp) predicted from the DNA sequence of pFL50 (Figure 17A, lanes 3 and 4). Restriction enzyme analysis of these amplicons with *DdeI/HinfI* produced 7 similar fragments of predicted size (Figure 17A, lanes 5 and 6). Both amplicons hybridized at high stringency with a 3 kb *HindIII/EcoRI* probe from pFL50 (restriction enzyme sites shown in Figure 6) containing sequences from between *carA* and *carB* from *N. gonorrhoeae* CH811 (Figure 17B, lanes 3 and 4).

When 29 other gonococcal isolates of varying A/S/P class (Table 1) were examined, the resulting amplicons were found to vary in size from approximately 3.8 to 2.4 kb (Table 6). These amplicons contain an estimated 74 bp of the 3' end of *carA*, and 111 bp of the 5' end of *carB*, therefore the size of the intervening sequences for the isolates is estimated to vary from between 3.7 and 2.2 kb. All of the amplicons produced hybridized at high stringency to the 3 kb *HindIII/EcoRI* probe containing sequences from between *carA* and *carB* from pFL50. As an example, Figure 17 shows amplicons produced from isolates of auxotype NR and P.

FIGURE 17. Electrophoresis and Southern blot analysis of PCR amplicons of the intergenic sequence from gonococcal isolates of varying A/S/P class. Figure 17A. Electrophoresis of the following: Lane 1: pFL50-*HindIII/EcoRI*; Lanes 2, 7, 18 and 29: 1 kb DNA ladder; Lane 3: amplicon from pFL50; Lane 4: amplicon from isolate CH811; Lane 5: amplicon from pFL50-*DdeI/HinfI*; Lane 6: amplicon from isolate CH811; Lanes 8 to 17: amplicons from isolates NS466 (lane 8), NS2117 (lane 9), NS568 (lane 10), NS1095 (lane 11), MS11-MS (lane 12), 11110 (lane 13), 9228 (lane 14), 11844 (lane 15), 9345 (lane 16), 9881 (lane 17); Lane 19: Negative control (no DNA template); Lanes 20 to 28: amplicons from isolates 9257 (lane 20), 10830 (lane 21), 11860 (lane 22), FA1090 (lane 23), NS4702 (lane 24), NS5846 (lane 25), NS667 (lane 26), NS3019 (lane 27), and NS8318 (lane 28). Figure 17B. Southern blot of 17A probed with a 3 kb *HindIII/EcoRI* from pFL50 which contains sequences from between *carA* and *carB* from *N. gonorrhoeae* CH811. Note that only the results obtained from isolates of auxotype NR and P are shown here.

A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29



B

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

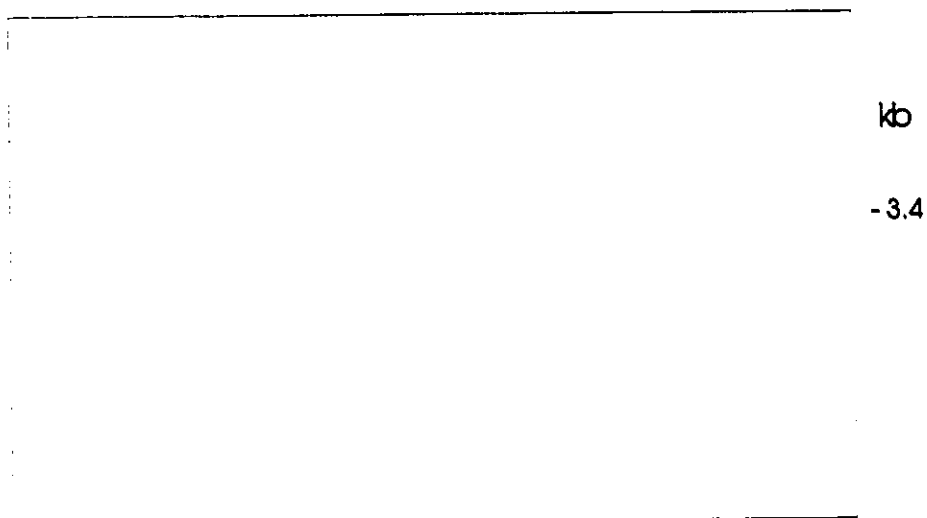


TABLE 6. Amplicon size and REA pattern for the intergenic sequence between *carA* and *carB* from isolates of *N. gonorrhoeae* of varying A/S/P class, and location and date of isolation.

Isolate type ^a	A/S/P class	Estimated amplicon size (kb)	REA pattern
CH811	NR/IB-2/ plasmid free	3.4	G2
NS466	NR/IB-5/2.6	3.4	G4
NS2117	NR/IB-5/2.6	3.4	G4
NS568	NR/IB-7/2.6	3.2	G5
NS1095	NR/IB-7/2.6,24.5	2.7	G9
MS11-ms	NR/IB-9/2.6	3.4	G3
11110	NR/IA-1/2.6,25.2	2.9	G6
9228	NR/IA-2/2.6,3.2,25.2	2.9	G6
11844	NR/IA-2/2.6,3.2,25.2	2.9	G6
9345	NR/IA-5/2.6,3.2,24.5	2.7	G9
9881	NR/IA-5/2.6,3.2,24.5	2.7	G9
9257	P/IA-6/2.6,4.5	3.4	G4
10830	P/IA-6/2.6,4.5,25.2	3.4	G4
11860	P/IA-6/2.6,3.2,25.2	2.9	G7
FA1090	P/IB-3/2.6	3.2	G6 ^d
NS4702	P/IB-5/2.6,24.5	2.9	G6
NS5846	P/IB-5/2.6,24.5	2.7	G8
NS667	P/IB-7/2.6	3.4	G2
NS3019	P/IB-7/2.6	3.4	G3
NS8318	P/IB-7/2.6	3.4	G3
9073	OUH/IA-2/2.6	3.5	G1
9315	OUH/IA-2/2.6	3.8	G1 ^d

continued....

TABLE 6. *continued*....

Isolate type ^a	A/S/P class	Estimated amplicon size (kb)	REA pattern
NS303 ^b	CUH/IA-2/2.6	3.5	G1
NS686 ^b	CUH/IA-2/2.6	- ^c	n/a
NS716 ^b	CUH/IA-2/2.6	- ^c	n/a
NS1061 ^b	CUH/IA-2/2.6	- ^c	n/a
NS791 ^b	PCU/IB-3/2.6	2.4	G10
NS2152 ^b	PCU/IB-2/plasmid free	- ^c	n/a
NS384 ^b	PCU/IB-2/plasmid free	- ^c	n/a
NS348 ^b	PCU/IB-7/plasmid free	- ^c	n/a

^a REA patterns were developed by *DdeI/HinfI* digestion of the PCR amplicons. Each pattern type is illustrated in Figure 20.

^b Isolates known to be CPS deficient (Li, 1993).

^c PCR amplicons not produced, even when a lower stringency PCR was performed (see Methods). Amplicons were produced of a portion of the *carB* gene, indicating the template DNA was of suitable quality.

^d Isolates which according to REA with *DdeI/HinfI* produce patterns similar to those of other isolates with a different amplicon size. These isolates' intergenic sequences do produce different, unique patterns when digested with other restriction enzymes.

While amplicons were obtained from all isolates which are known to have a functional CPS enzyme (auxotype NR, P, and OUH), only 2 out of 8 isolates with CPS deficiencies (auxotype CUH and PCU) produced amplicons (Table 6). These isolates which did not produce amplicons of the intergenic sequence, did produce an amplicon when a portion of the *carB* gene was amplified, indicating the template DNA was of acceptable quality. A lower primer annealing temp. of 45°C was also used and still no amplicon of the intergenic sequence was obtained. Of the two CPS deficient isolates which did produce amplicons, one isolate, of auxotype PCU, had an atypical serotype (IB-3) and plasmid content (2.6) for a PCU isolate. However the other isolate, auxotype CUH, was of a common A/S/P class and the same A/S/P class as the another 3 isolates. This group is therefore being further discriminated by this analysis.

No obvious correlations were seen between the size of the amplicons produced and the date of isolation, isolation site, or geographic origin of the isolate. There was a possible correlation between amplicon size and A/S/P class, however the sample size within each A/S/P class was not large enough to make any definitive conclusion. Some isolates of the same A/S/P class could be further discriminated by amplicon size, for example the two isolates of A/SP class P/IB-5/2.6,24.5 had different amplicon sizes.

3.4.2 PCR amplification of the intergenic sequence from other *Neisseria* species: a similar large, variable sequence is present between *carA* and *carB*.

In order to determine whether the large intervening sequence observed in gonococcal isolates was also present in other *Neisseria* species, the protocol used to amplify the intervening sequence from gonococcal isolates was also used to investigate 12 isolates of *Neisseria meningitidis* and 18 isolates of

commensal *Neisseria* which comprise 9 species. Amplicons were found to vary in size, from approximately 1.4 to 4.1 kb (Table 7).

All twelve isolates of *N. meningitidis*, comprising 5 different serovars, produced amplicons, ranging in size from 2.8 to 3.3 kb (Table 7). Though the sample size was small, amplicon size was associated with serogroup, except for 2 isolates from serogroup B, which produced differently sized amplicons of 2.8 and 3.3 kb. Seventeen of the eighteen commensal *Neisseria* isolates examined produced amplicons of twelve different sizes, ranging from 1.4 to 4.1 kb (Figure 18A; Table 7). The one isolate not producing an amplicon was the only isolate of *N. elongata* examined, a species that is significantly different from the others in that it is a rod and not a diplococcus (Vedros, 1984). Intraspecies variation in amplicon size was noted in all species where more than one isolate had been examined. A single amplicon was produced from each isolate with the exception of *N. perflava/sicca* 1690 which produced an amplicon of 3.0 kb and an amplicon of lesser intensity of 2.8 kb in size (Figure 18A, Lane 8). Amplicons produced from commensal *Neisseria* species isolates were in general more variable in size than those from the pathogenic *Neisseria* isolates, however, more species were examined in the commensal group.

3.4.3 Comparison of the intergenic sequences between *carA* and *carB* from different *Neisseria* species by Southern hybridization: similarities exist, but RS6 and RS7 may be gonococcal-specific sequences.

The 3 kb *HindIII* *EcoRI* probe comprising most of the intervening sequence from *N. gonorrhoeae* CH811 was also used to investigate the degree of similarity of this sequence with the intergenic sequences of the various *Neisseria* species. This probe hybridized at high stringency to all

TABLE 7. Amplicon size and REA pattern for the intergenic sequence between *carA* and *carB* from isolates of *N. meningitidis* and commensal *Neisseria* species.

Isolate	Species	Serovar	Estimated amplicon size (kb)	REA pattern type ^a
1503	<i>N. meningitidis</i>	A 4:P1.9	3.3	M1
1504	<i>N. meningitidis</i>	A 4:P1.9	3.3	M1
1500	<i>N. meningitidis</i>	B 4:P1.15	2.8	M5
1501	<i>N. meningitidis</i>	B 2b:P1.2	3.3	M2
1514	<i>N. meningitidis</i>	C 2a:P1.2 ^a	3.0	M3
1518	<i>N. meningitidis</i>	C 2a:P1.2 ^a	3.0	M3
1525	<i>N. meningitidis</i>	C 2a:P1.1 ^a	3.0	M3
1526	<i>N. meningitidis</i>	C 2a:P1.1 ^a	3.0	M3
1527	<i>N. meningitidis</i>	C 2a:P1.1 ^b	3.0	M4
1339	<i>N. meningitidis</i>	W135	2.8	M5
1497	<i>N. meningitidis</i>	Y	2.3	M5
1498	<i>N. meningitidis</i>	Y	2.8	M5
1682	<i>N. flava</i>	N/A	2.8	C4
1683	<i>N. subflava</i>	N/A	2.7	C5
1684	<i>N. elongata</i>	N/A	- ^b	n/a
1685	<i>N. mucosa</i>	N/A	2.8	C4
1686	<i>N. mucosa</i>	N/A	2.4	C7
1687	<i>N. lactamica</i>	N/A	3.0	C3
1688	<i>N. lactamica</i>	N/A	2.4	C8
1689	<i>N. lactamica</i>	N/A	2.4	C8
1690	<i>N. perflava/sicca</i>	N/A	3.0 (2.8) ^c	C3
1691	<i>N. perflava/sicca</i>	N/A	4.1	C1
1692	<i>N. perflava/sicca</i>	N/A	1.6	C10

continued...

TABLE 7. *continued...*

Isolate	Species	Serovar	Estimated amplicon size (kb)	REA pattern type ^a
1693	<i>N. flavescens</i>	N/A	3.0	C2
1694	<i>N. flavescens</i>	N/A	2.5	C6
1695	<i>N. polysacchareae</i>	N/A	2.1	C9
1696	<i>N. polysacchareae</i>	N/A	1.7	C10
1697	<i>N. polysacchareae</i>	N/A	1.8	nd.
1698	<i>N. cinerea</i>	N/A	1.5	C11
1699	<i>N. cinerea</i>	N/A	1.4	C12

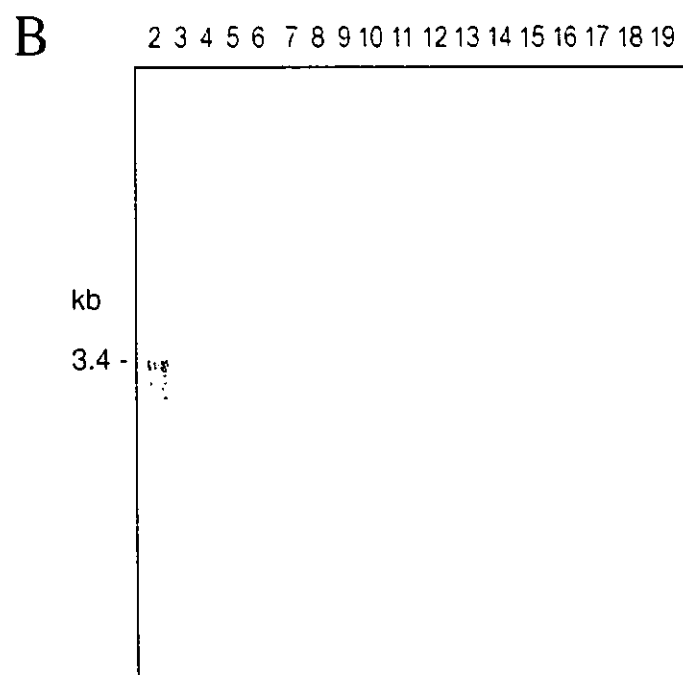
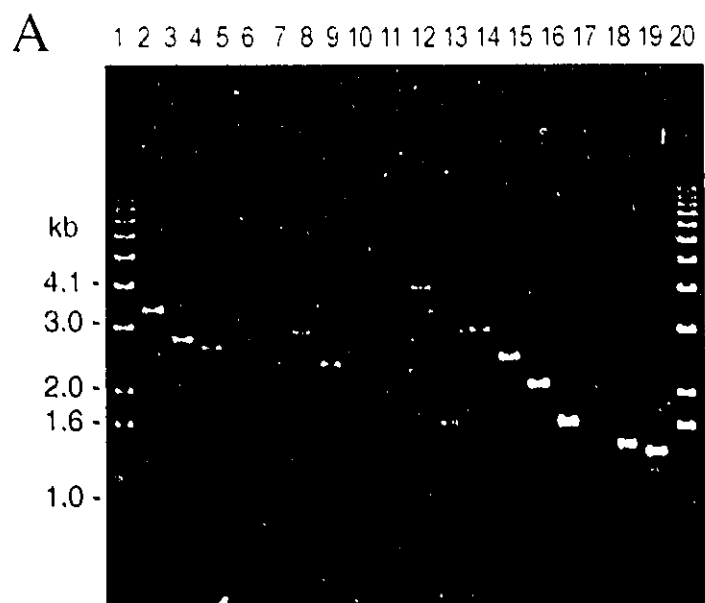
^a REA patterns were developed by *DdeI/HinfI* digestion of the PCR amplicons. Each pattern type is illustrated in Figure 20.

^b PCR amplicon not produced, even when a lower stringency PCR was performed (see Methods).

^c A second less intense amplicon was produced by this isolate of approximately 2.8 kb in size (see Figure 18).

nd. not done.

FIGURE 18. Electrophoresis and Southern blot analysis of PCR amplicons of the sequence between *carA* and *carB* from 17 commensal *Neisseria* species isolates comprising 8 species. Figure 18A. Electrophoresis of the following: Lane 1 and 20, 1 kb ladder; Lane 2, amplicon from pFL50 template DNA; Lanes 3 to 19 contain amplicons from a template of chromosomal DNA from the following different isolates: Lane 3, *N. flava* 1682; Lane 4, *N. subflava* 1683; Lane 5, *N. mucosa* 1685; Lane 6, *N. mucosa* 1686; Lane 7, *N. lactamica* 1687; Lane 8, *N. lactamica* 1688; Lane 9, *N. lactamica* 1689; Lane 10, *N. perflava/sicca* 1690; Lane 11, *N. perflava/sicca* 1691; Lane 12, *N. perflava/sicca* 1692; Lane 13, *N. flavescens* 1693; Lane 14, *N. flavescens* 1694; Lane 15, *N. polysacchareae* 1695; Lane 16, *N. polysacchareae* 1696; Lane 17, *N. polysacchareae* 1697; Lane 18, *N. cinerea* 1698; Lane 19, *N. cinerea* 1699. Figure 18B. Southern blot of 18A probed with the 3.0 kb *Hind*III *Eco*RI fragment from pFL50 which contains most of the intervening sequence from *N. gonorrhoeae* CH811. The hybridization was performed under conditions of low stringency (60°C).



the amplicons of the intergenic sequence produced from the *N. meningitidis* isolates, however a lower stringency (60°C) was required for hybridization with any amplicons produced from commensal *Neisseria* species. At this lower stringency, 13 out of the 17 amplicons produced from the commensal *Neisseria* isolates strongly hybridized to the probe, while the remaining 4 out of 17, all with the smallest amplicon sizes, weakly hybridized (Figure 18B).

A 0.7 kb *Pvu*II *Sph*I fragment from pFL50 (see restriction enzyme sites in Figure 6) was also used to probe the amplicons of the intergenic sequence in a similar fashion. This probe only contains copies of the RS6/RS7 repeats and was used to look at the species specificity of these sequences. Since none of the intergenic sequences from the commensal *Neisseria* species isolates hybridized to the whole intergenic sequence at high stringency, there was no need to attempt to hybridize the RS6/RS7 repeats at high stringency to these isolates, however all of the *N. gonorrhoeae* and *N. meningitidis* isolates were investigated. The RS6/RS7 probe hybridized to the intergenic sequences from all the gonococcal isolates at this high stringency, but did not hybridize to any of the corresponding sequences in the meningococcal isolates (Figure 19). This indicates that these repeats may be species-specific for *N. gonorrhoeae* at this stringency. At a lower stringency (60°C) these RS6/RS7 sequences did hybridize to the amplicons of the meningococcal intergenic sequences indicating there were some analogous sequences present (data not shown). The commensal *Neisseria* species isolates were not investigated at this lower stringency. These RS6/RS7 repeats did not hybridize at high stringency to chromosomal DNA from two *E. coli* strains (JM240 and HB101) and one *Peptostreptococcus magnus* strain (ATCC 53516), though they hybridized with the expected fragment from *Xho*I/*Xba*I digested gonococcal chromosomal DNA (data not shown). These hybridization studies are summarized in Table 8.

FIGURE 19. Hybridization of the RS6/RS7 repeats with amplicons of the sequence between *carA* and *carB* from a selection of *N. gonorrhoeae* and *N. meningitidis* isolates. Note that some lanes have been omitted from the gel for brevity. Figure 19A. Electrophoresis of the following: Lanes 1, 8 and 14, 1 kb ladder; Lane 2, pFL50 template DNA; Lane 3, negative control; Lanes 4 to 7, *N. gonorrhoeae* chromosomal DNA template from the following isolates: Lane 4, NS5846; Lane 5, NS4702; Lane 6, FA1090; Lane 7, MS11-ms; Lanes 9 to 13, *N. meningitidis* chromosomal DNA template from the following isolates: Lane 9, 1500; Lane 10, 1501; Lane 11, 1504; Lane 12, 1497; Lane 13, 1525. Figure 19B. Southern blot of 19A probed at high stringency with the 0.7 kb *Pvu*II *Sph*I fragment from pFL50 which predominantly contains copies of the RS6/RS7 repeats found between *carA* and *carB* in *N. gonorrhoeae* CH811.

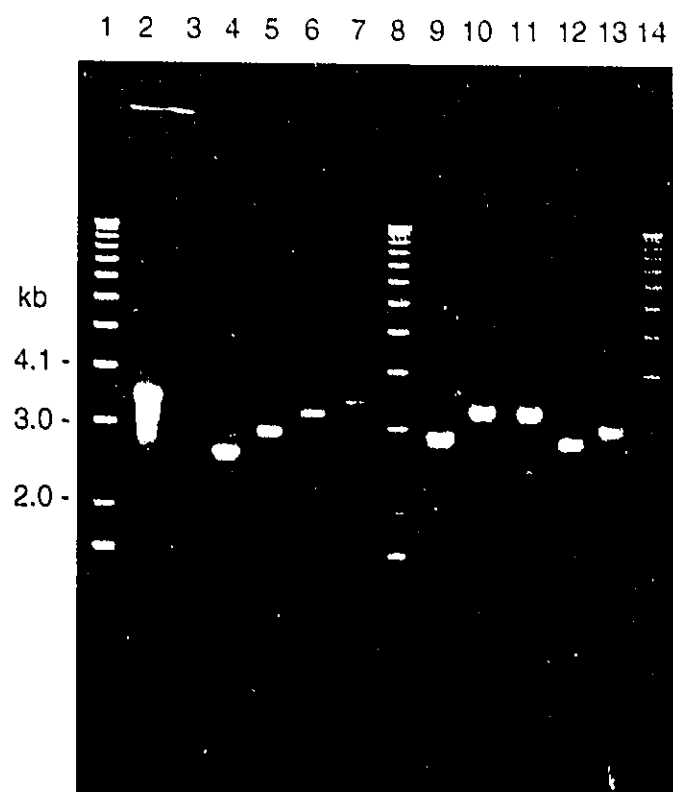
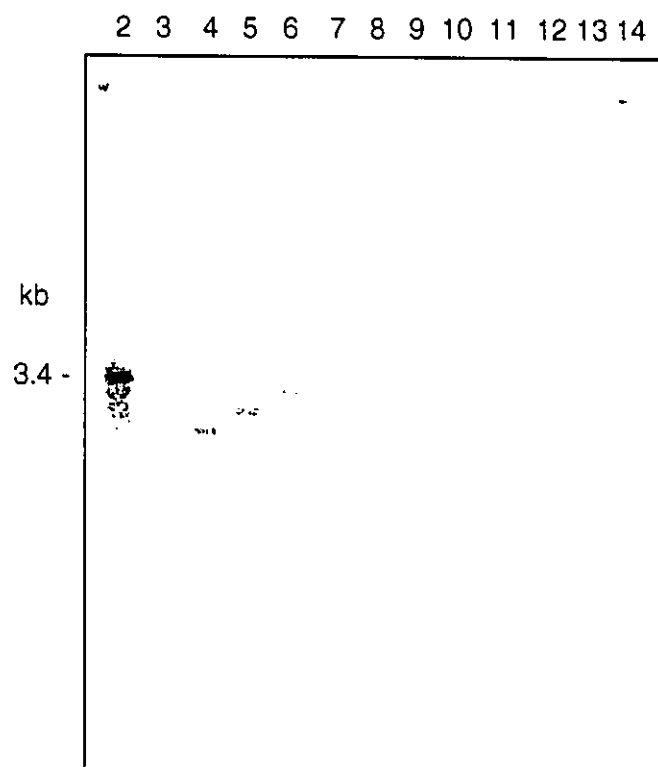
A**B**

TABLE 8. Summary of hybridization studies comparing portions of the intergenic sequence from *N. gonorrhoeae* CH811 with the intergenic sequences from other *Neisseria* species and with chromosomal DNA from *E. coli* (strain HB101 and JM240) and *P. magnus* (ATCC 53516).

Species	Proportion of isolates with sequences hybridizing to the following probes under conditions of high (68°C) or low (60°C) stringency.			
	<i>N. gonorrhoeae</i> CH811 intergenic sequence probe ^a		RS6/RS7 repeats probe ^b	
	High Str.	Low Str.	High Str.	Low Str.
<i>N. gonorrhoeae</i>	24/24	nd.	24/24	nd.
<i>N. meningitidis</i>	12/12	nd.	0/12	12/12
Commensal <i>Neisseria</i> spp.	0/17	13/17	0/17	nd.
<i>E. coli</i>	nd.	nd.	0/2	nd.
<i>P. magnus</i>	nd.	nd.	0/1	nd.

^a Probe consists of a 3.0 kb *Hind*III *Eco*RI fragment from pFL50 which contains most of the intervening sequence from *N. gonorrhoeae* CH811.

^b Probe comprises a 0.7 kb *Pvu*II *Sph*I fragment from pFL50 which predominantly contains copies of the RS6/RS7 repeats found in the intergenic sequence of *N. gonorrhoeae* CH811.

nd. not done.

3.4.4 Comparison of the neisserial intergenic sequences by restriction enzyme analysis (REA) reveals more variability between the intergenic sequences, though a conserved region is present in the gonococcal intergenic sequences

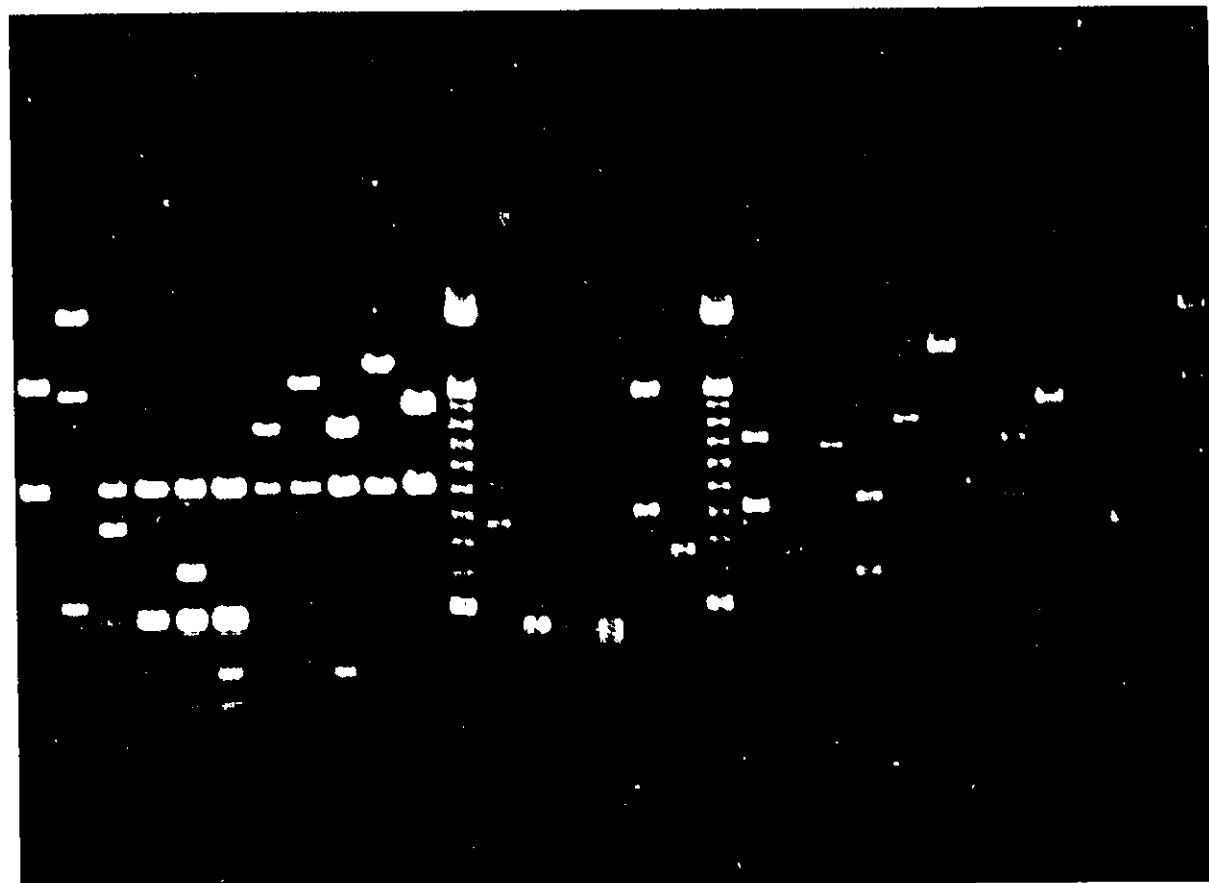
Restriction enzyme analysis of the amplicons containing the intergenic sequences was performed using the enzymes *DdeI* and *HinfI*. These enzymes were initially chosen based on their ability to produce a pattern of suitable complexity for gonococcal isolates.

In the 24 gonococcal isolates (of varying A/S/P) examined, 10 different REA patterns (Figure 20, lanes G1 to G10; Table 6) were observed. Many PCR amplicons of similar size produced different REA patterns (*i.e.* either patterns G2, G3, or G4 are produced from isolates with amplicons of 3.4 kb in size), indicating that these amplicons are more variable in sequence than can be observed just by amplicon size. No less than 4 out of the 7 different amplicon sizes observed could be further discriminated by the REA analysis. Isolates even of the same A/S/P class and same amplicon size could be further discriminated by this REA analysis, for example the three isolates of A/S/P class P/IB-7/2.6 could be further subgrouped into patterns G2 or G3.

However, conversely, isolates with patterns G1 were found to have either a 3.5 or 3.8 kb amplicon size. A similar observation was made within the group which produced pattern G6 (comprising isolates with an amplicon size of either 2.9 or 3.2 kb). Analysis using other restriction enzymes (data not shown) showed that this was due to the presence of co-migrating bands in the isolates with the larger amplicon size (*i.e.* the isolates with amplicon sizes of 3.8 and 3.2 kb). Interestingly, the extra co-migrating bands found in the 3.8 kb/pattern G1 isolate and the 3.2 kb/pattern G6 isolate were of the same size (just under 300 bp) which corresponded to the size of the RS6/RS7 repeat unit.

FIGURE 20. Composit of all the different REA patterns of the intergenic sequence identified from a total of 60 *Neisseria* species isolates (using the restriction enzymes *Dde*I and *Hinf*I). REA was performed on PCR amplicons of the sequence between *carA* and *carB*, using the restriction enzymes *Dde*I and *Hinf*I. A total of 10 patterns were identified from the 30 *N. gonorrhoeae* isolates examined (patterns G1 to G10), a total of 6 patterns were identified within the 12 *N. meningitidis* isolates examined (patterns M1 to M6), and a total of 12 patterns were identified in a group of 17 commensal *Neisseria* isolates which comprise 8 species (patterns C1 to C12). Lanes containing 100 bp DNA ladders are marked with L.

G1 G2 G4 G6 G8 G10 M1 M3 M5 L C2 C4 C6 C8 C10 C12
 L G3 G5 G7 G9 L M2 M4 M6 C1 C3 C5 C7 C9 C11 L



kb
 - 1.0
 - 0.8
 - 0.6
 - 0.4
 - 0.2

When looking at all of the patterns in general, the patterns did have some bands in common between them, in particular one band corresponding to a fragment of approximately 1 kb in size, which was present in all patterns. Mapping of the *Dde*I and *Hinf*I restriction enzyme sites showed that this 1 kb band is a conserved fragment which is located immediately downstream of *carA*. According to this REA study, the variability in the intergenic sequence in gonococcal isolates seems to reside within the middle of the intergenic sequence and within the end closest to *carB*.

In the 12 meningococcal isolates examined, 6 different REA patterns were observed (Figure 20, lanes M1 to M6; Table 7). These patterns were clearly different from those of the gonococcal isolates. Single restriction enzyme digests also confirmed that the patterns were quite different between the two species. Several bands were common to most meningococcal isolates, however there was no band conserved in all isolates. Identical patterns were found within, but not between, different serogroups. Some serogroups could be further discriminated. Notably, within the serovar C group, the non-ET-15 type isolate (Figure 20 pattern M4) was differentiated from the ET-15 type isolates (Figure 20, pattern M3), the latter of which have been associated with some significant outbreaks of meningitis in Canada (Ashton *et al.*, 1991; Strathdee *et al.*, 1993).

In 16 commensal *Neisseria* species isolates examined, 12 patterns were observed (Figure 20, lanes C1 to C12; Table 7). These patterns were also very different from those of the gonococci and meningococci. Virtually all isolates produced completely different REA patterns, indicating that this group is quite heterogenous. Identical patterns were only seen between 2 isolates of *N. lactamica*, an isolate of *N. lactamica* and *N. perflava/sicca*, and an isolate of *N. mucosa* and *N. flava*.

Overall, interspecies variability was greater than intraspecies variability for this REA of the sequence between *carA* and *carB* in *Neisseria* species.

3.4.5 Investigating the variable intergenic sequence from more closely related gonococcal isolates: possible utility for typing of *N. gonorrhoeae*.

REA patterns of the sequence between *carA* and *carB* were also obtained from sets of gonococcal isolates of similar A/S/P class (Tables 2 and 3), to examine the degree of variation in this sequence and to evaluate whether this sequence may be useful for epidemiological typing of *N. gonorrhoeae*.

Isolates of similar A/S/P class (P/IA-6/2.6,4.5), but varying in date and location of isolation were found to have an intergenic sequence REA pattern of predominantly type G2 (Table 9). Exceptions included four isolates obtained from Ontario, of which three clustered together when the isolates were ordered by date of isolation (Table 9; isolates 2082, 2318, and 4165; patterns G6, G3, and G6, respectively). The last isolate within this Ontario group (when ordered by date of isolation) also had a different pattern (isolate 8266; pattern G8). Of the isolates obtained from British Columbia, only one isolate had a pattern different from G2 and this isolate appeared second when ordered by date of isolation (Table 9, isolate 2525; pattern G3). In general the variability of the intergenic sequence within this group was less than that of the group of isolates of varying A/S/P class discussed above, though some variability was observed, indicating these isolates were not homogeneous.

Two other sets of isolates were also examined which had a similar A/S class, date of isolation and location of isolation as isolates associated with two outbreaks of gonorrhea. Note

TABLE 9. Amplicon size and REA pattern for the intergenic sequence between *carA* and *carB* from *N. gonorrhoeae* isolates of similar A/S/P class (P/IA-6/2.6,4.5), but varying date of isolation and geographic origin within two Canadian provinces.

Isolate	Date of isolation	Estimated amplicon size (kb)	REA pattern type ^a
709	84/11 ^b	3.4	G2
767	85/01 ^b	3.4	G2
1114	86/02 ^b	3.4	G2
2082	87/10 ^b	2.9	G6
2318	88/04 ^b	3.4	G3
4165	89/12 ^b	2.9	G6
4555	90/02 ^b	3.4	G2
5988	91/01 ^b	3.4	G2
6384	91/04 ^b	3.4	G2
6769	91/06 ^b	3.4	G2
7121	91/08 ^b	3.4	G2
8266	91/12 ^b	2.7	G8
1537	86/10 ^c	3.4	G2
2525	88/07 ^c	3.4	G3
5774	90/12 ^c	3.4	G2
5788	91/01 ^c	3.4	G2
6339	91/05 ^c	3.4	G2
7921	91/09 ^c	3.4	G2
7944	91/10 ^c	3.4	G2

^a REA patterns were developed by *DdeI/HinfI* digestion of the PCR amplicons. Each pattern type is illustrated in Figure 20.

^b Isolated in Ontario.

^c Isolated in British Columbia.

that these isolates were obtained from population studies and not patient-pair epidemiological studies. One cluster was of A/S class NR/IA-5 and the other unrelated cluster was of A/S class NR/IB-3 (Table 3). These isolates had been previously typed by ribotyping, chromosomal REA using the enzyme *Sma*I, and pulsed field gel electrophoresis (PFGE) using the enzyme *Spe*I (Li, 1993; Li, personal communication).

The isolates of A/S class NR/IA-5 produced amplicons of the same size with the same REA pattern for 9/10 isolates (Table 10; Figure 21, lanes 2 to 10; pattern G9). One isolate produced a different amplicon size and REA pattern (Figure 21, lane 11; pattern G6). This isolate is notably different from the others in that it contains a 25.2 MDa tetM-containing plasmid (verses a 24.5 MDa plasmid found in the other isolates). When these isolates were characterised using ribotyping and chromosomal REA (using *Sma*I), all isolates were found to be of the same type (Li, 1993). Therefore REA of the sequence between *carA* and *carB* from these isolates was more discriminatory than these previously published molecular methods. PFGE, however, separated these isolates into 5 groups and so was the most discriminatory (Li, personal communication).

The isolates of A/S class NR/IB-3 were actually all of the same A/S/P class (plasmid content: 2.6,24.5). When arranged in chronological order by date of isolation, the first 8 isolates produced PCR amplicons of one size, while the following 4 isolates produced amplicons of another size (Table 10). The REA patterns from these amplicons were correspondingly different (patterns G2 and G8), as shown in Figure 21 (lanes 13 to 24). This grouping of isolates correlated exactly with what was observed previously by the chromosomal REA method (Li, 1993). Ribotyping classified the first 6 isolates as one type and the remaining 6 isolates as another (Li, 1993). PFGE differentiated these isolates into 5 groups (Li, personal communication).

FIGURE 21. REA (using restriction enzymes *DdeI* and *HinfI*) of amplicons of the intergenic sequence from two groups of isolates with similar A/S class as isolates associated with two outbreaks of gonorrhea. Lane 1, 1 kb ladder; Lanes 2 to 11, chromosomal DNA template from the following isolates which are all of A/S class NR/IA-5: Lane 2, 3523; Lane 3, 2852; Lane 4, 3785; Lane 5, 3928; Lane 6, 4020; Lane 7, 5265; Lane 8, 5707; Lane 9, 6525; Lane 10, 6595; Lane 11, 5283; Lane 12, 100 bp ladder; Lanes 13 to 24, chromosomal DNA template from the following isolates which are all of A/S class NR/IB-3: Lane 13, 88-1286; Lane 14, 88-1919; Lane 15, 88-2351; Lane 16, 88-2802; Lane 17, 88-3482; Lane 18, 88-3511; Lane 19, 3644; Lane 20, 3648; Lane 21, 4743; Lane 22, 5837; Lane 23, 5840; Lane 24, 5855.

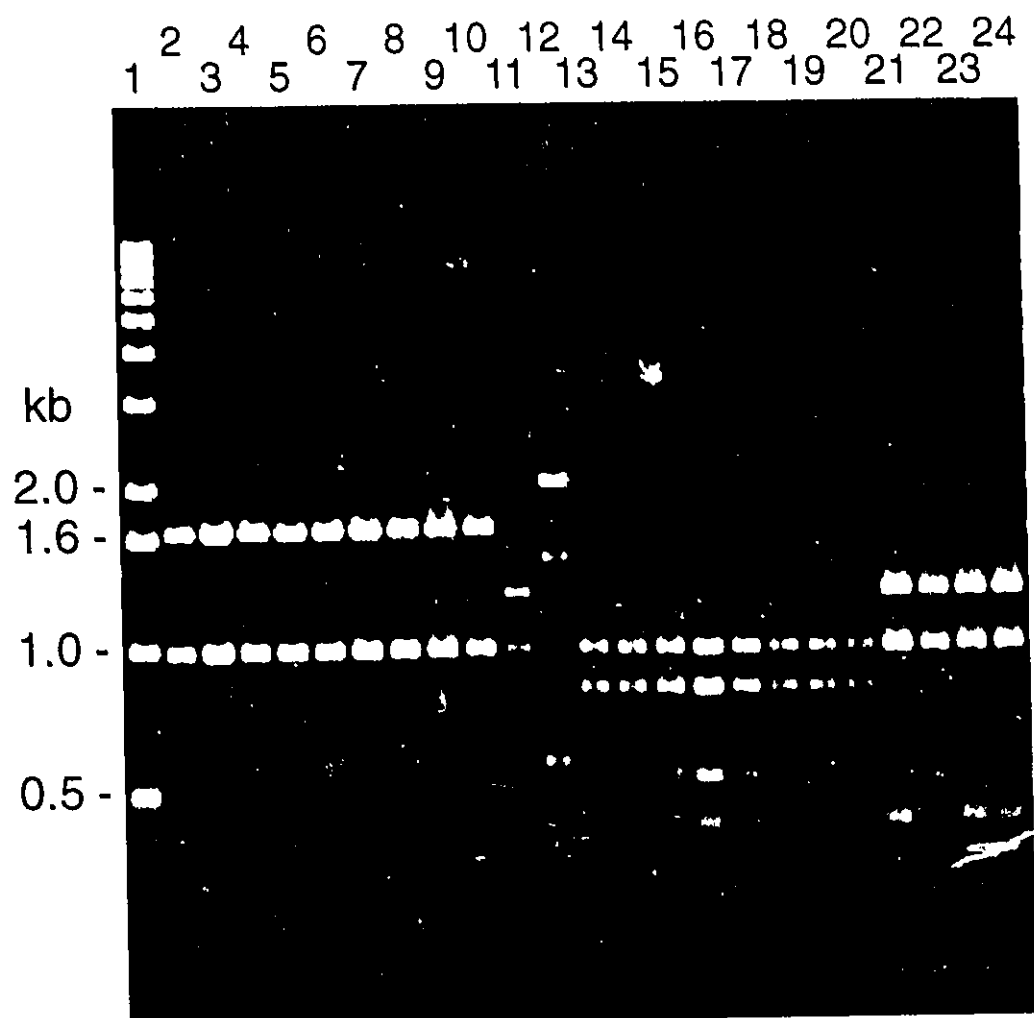


TABLE 10. Amplicon size and REA pattern for the intergenic sequence between *carA* and *carB* from *N. gonorrhoeae* isolates with a similar A/S/P class, date of isolation, and geographic location as isolates associated with two outbreaks of gonorrhea (one outbreak comprised isolates with A/S class NR/IA-5 and the other outbreak involved isolates of A/S class NR/IB-3).

Isolate	A/S/P Class	Year Isolated	Estimated amplicon size (kb)	REA pattern type ^a
3523	NR/IA-5/2.6,24.5	1989	2.7	G9
2852	NR/IA-5/2.6,3.2,24.5	1988	2.7	G9
3735	NR/IA-5/2.6,3.2,24.5	1989	2.7	G9
3928	NR/IA-5/2.6,3.2,24.5	1989	2.7	G9
4020	NR/IA-5/2.6,3.2,24.5	1989	2.7	G9
5265	NR/IA-5/2.6,3.2,24.5	1990	2.7	G9
5707	NR/IA-5/2.6,3.2,24.5	1990	2.7	G9
6525	NR/IA-5/2.6,3.2,24.5	1991	2.7	G9
6595	NR/IA-5/2.6,3.2,24.5	1991	2.7	G9
5283	NR/IA-5/2.6,3.2,25.2	1990	3.4	G6
88-1286	NR/IB-3/2.6,24.5	1988	3.4	G2
88-1919	NR/IB-3/2.6,24.5	1988	3.4	G2
88-2351	NR/IB-3/2.6,24.5	1988	3.4	G2
88-2802	NR/IB-3/2.6,24.5	1988	3.4	G2
88-3482	NR/IB-3/2.6,24.5	1989	3.4	G2
88-3511	NR/IB-3/2.6,24.5	1989	3.4	G2
3644	NR/IB-3/2.6,24.5	1989	3.4	G2
3648	NR/IB-3/2.6,24.5	1989	3.4	G2
4743	NR/IB-3/2.6,24.5	1990	2.7	G8
5837	NR/IB-3/2.6,24.5	1990	2.7	G8
5840	NR/IB-3/2.6,24.5	1990	2.7	G8
5855	NR/IB-3/2.6,24.5	1990	2.7	G8

^a REA patterns created by *DdeI/HinfI* digestion of the PCR amplicons. Pattern types are shown in Figure 20.

In general, REA of the sequence between *carA* and *carB* showed analogous levels of discrimination as compared to the chromosomal REA and ribotyping methods, and was not as discriminatory as PFGE. Amplicon size correlated with REA pattern for all of these closely related isolates. The degree of variation was less within these sets of isolates in comparison to the other sets of isolates examined which had been obtained over a wider period of time and are thought to be less related.

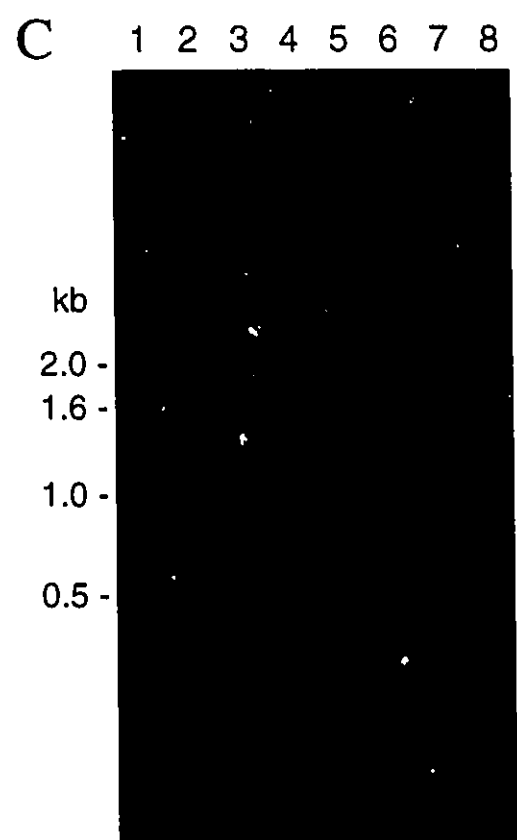
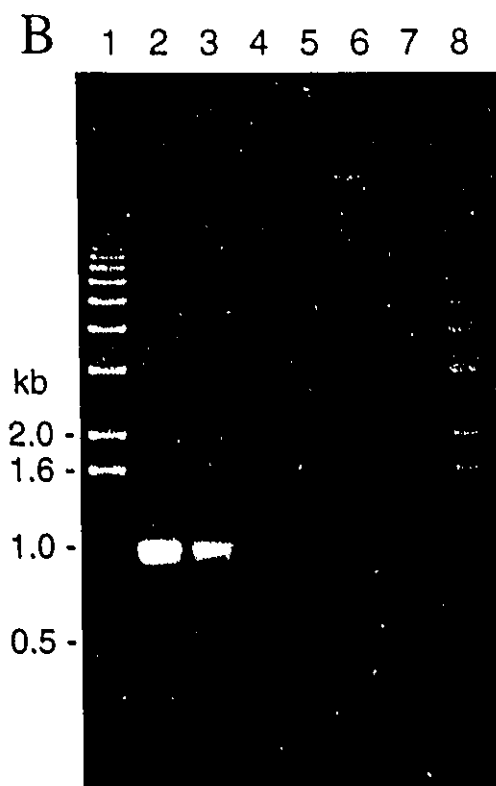
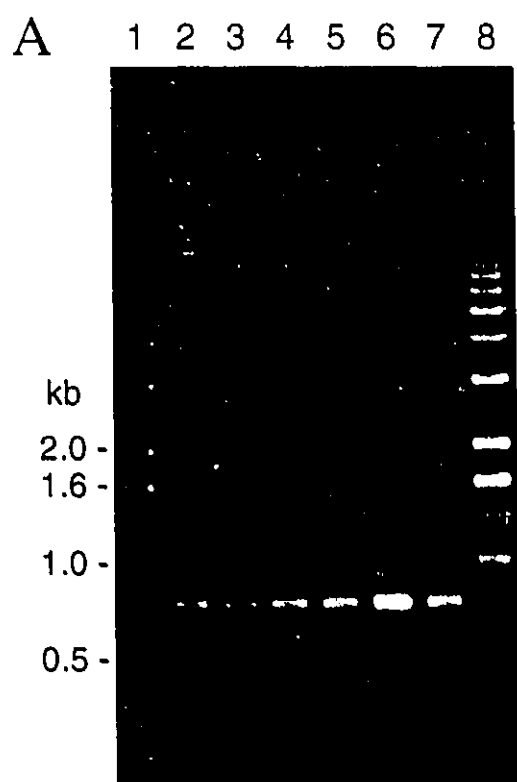
3.4.6 Comparison of the gonococcal intergenic sequences by PCR amplification of selected sub-regions: changes in the number of RS6/RS7 repeats explain some, but not all, changes in the size of the intergenic sequences.

In order to preliminarily investigate the cause of the variation in the intergenic sequence between the *carA* and *carB* genes, and attempt to narrow down or confirm the location of the variable region, selected portions of the intergenic sequence were amplified by PCR. A selection of 5 gonococcal isolates (see Methods Section 2.8 or Figure 22 legend for details) whose intergenic sequences represented a range of sizes, were used for these studies.

When a portion of the intergenic sequence immediately downstream of *carA* was amplified using primers 53 and 74 (see Figure 6 for primer binding sites) all amplicons were of the same size, 0.8 kb, and this size correlated exactly with the size predicted from the sequence of pFL50 (Figure 22A). This provided further indication that this region of the intergenic sequence immediately downstream of *carA* is conserved in gonococcal isolates.

When a portion of the intergenic sequence immediately upstream of *carB* was amplified in a similar manner using primers 75 and 63, amplicons were only obtained from pFL50 and the isolate

FIGURE 22. PCR amplicons of subregions of the intergenic sequence between *carA* and *carB* from selected gonococcal isolates which have a varying intergenic sequence size. Regions were chosen based on the sequence of pFL50 (see Figure 6 for primer-binding locations). Figure 22A. Electrophoresis of PCR amplifications using primers 53 and 74 which amplifies a portion of the intergenic sequence immediately downstream of *carA* pFL50: Lanes 1 and 8, 1 kb ladder; Lanes 2 to 7, PCR amplicons obtained from a template of the following plasmid DNA or organism's chromosomal DNA: Lane 2, pFL50; Lane 3, CH811; Lane 4, FA1090; Lane 5, 11110; Lane 6, NS5846; Lane 7, 9315. Figure 22B: Electrophoresis of PCR amplifications using primers 75 and 63 which amplifies a portion of the intergenic sequence immediately upstream of *carB* in pFL50. Lanes are the same as described for Figure 22A. Figure 22C: Electrophoresis of PCR amplifications using primers 66 and 72 which amplifies a portion of the intergenic sequence containing the RS6/RS7 sequences in pFL50. Lanes are the same as described for Figure 22A with the exception of a few changes in the isolates examined in Lanes 4 to 7: Lane 4, MS11-ms; Lane 5, FA1090; Lane 6, 11860; Lane 7, 9345.



CH811 (Figure 22B, lanes 2 and 3). Some faint bands were detected for isolate FA1090 and NS5846 (Figure 22B, lanes 4 and 6). This lack of amplicon production suggested that the sequence where these primers were binding in pFL50 was either somewhat different or not present at all in the other isolates.

When a central portion of the intergenic sequence was amplified from isolates with a varying intergenic sequence size (using primers 66 and 72; amplifies the region containing the RS6/RS7 repeats in pFL50), amplicons were produced from all isolates, most of which varied in size (Figure 22C). The variation observed correlated closely with changes in size predicted from changes in the number of RS6/RS7 sequences. However, there was also some indication that changes were occurring elsewhere in the intergenic sequence. For example, two isolates which had different amplicon sizes for the whole intergenic sequence (CH811 and FA1090), produced similar amplicon sizes when primers 66 and 72 were used (Figure 22C, lanes 4 and 6). Since the size of the sequence immediately upstream of *carA* seemed to be conserved in these isolates (Figure 22A, Lanes 3 and 4), this would therefore suggest that the remaining changes would be occurring in the region close to *carB*.

3.5 STUDIES OF NATURALLY OCCURRING GONOCOCCAL CPS MUTANTS: CAUSED BY MORE THAN ONE MUTATION AND HIDING OTHER MUTATIONS

3.5.1 *carA* and *carB* are conserved in size in wild type isolates but not in CPS deficient isolates: more than one mutation led to the formation of CPS deficiencies in gonococcal isolates.

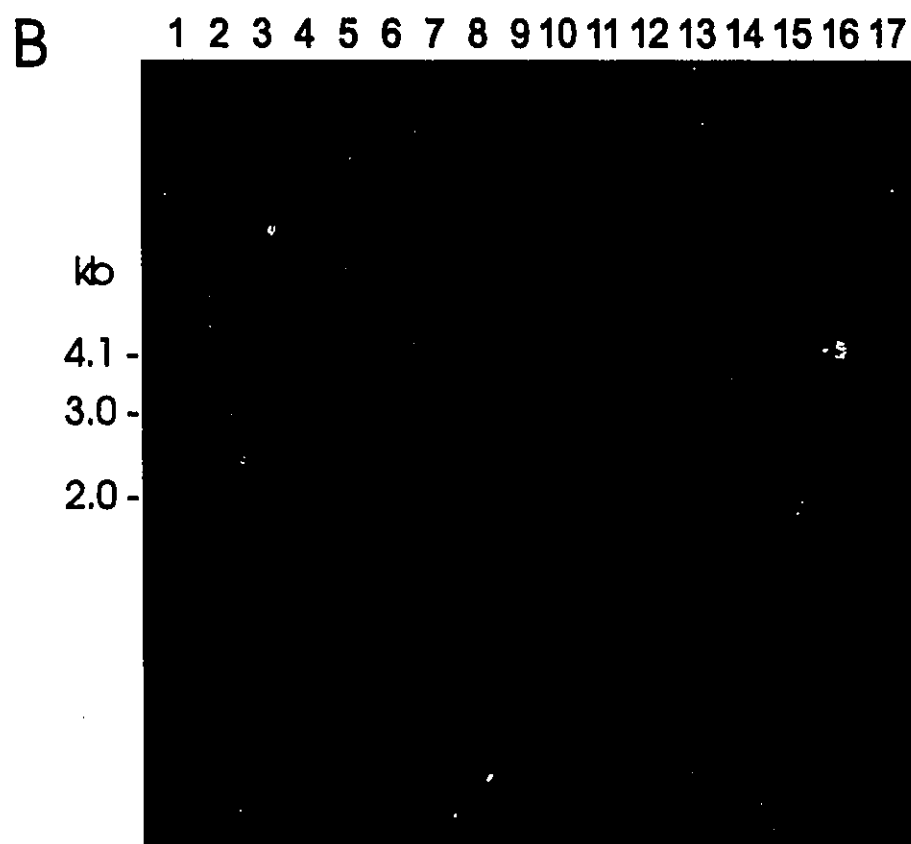
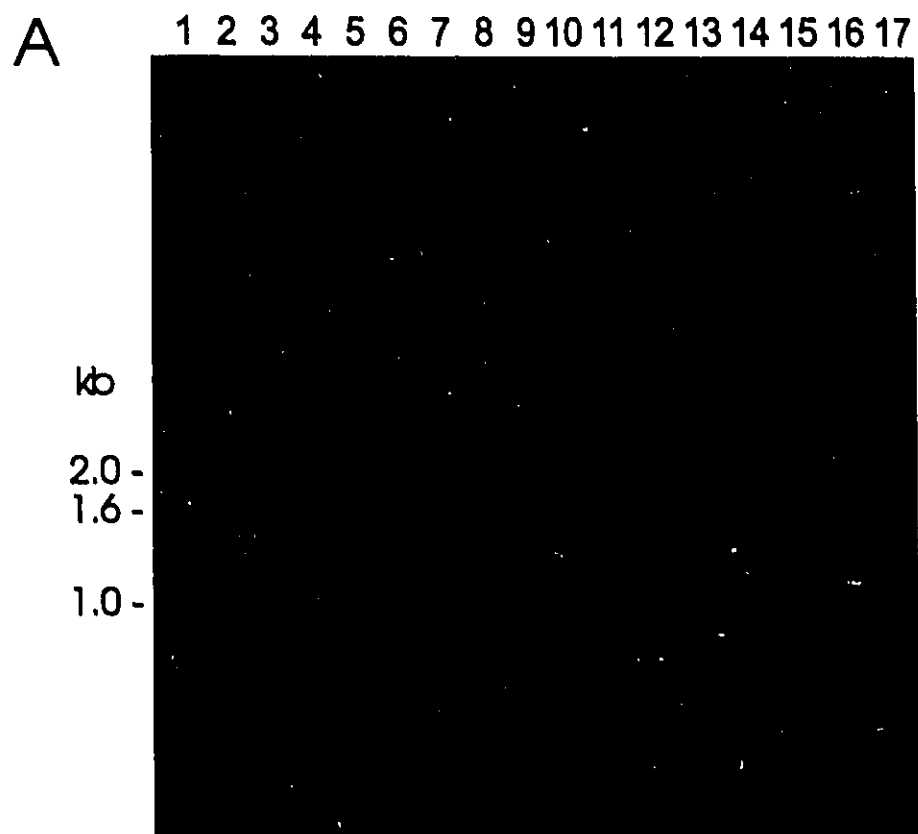
In order to determine whether the *carA* and *carB* genes varied in size like the intergenic sequence did, and to examine these genes in the naturally occurring gonococcal CPS mutants, primers

were selected to amplify these genes from 11 gonococcal isolates. Five were selected that vary in A/S/P class and intergenic sequence size, and 6 were selected because of their CPS deficiencies (see Methods Section 2.8 or Figure 23 legend for details). Of the 5 isolates known to have functional CPS, all produced amplicons containing *carA* of similar and predicted size (1431 bp; Figure 23A, lanes 3 to 7), and all amplicons containing *carB* were also of similar, predicted size (3303 bp; Figure 23B, lanes 3 to 7). Therefore, the size of the *carA* and *carB* genes seems to be conserved, though the intervening sequence between these genes varies. Of the 6 isolates known to be CPS deficient, all produced an amplicon of predicted size containing *carB* (Figure 23B, lanes 9 to 14). However, when amplifying *carA* (Figure 23A, lanes 9 to 14), amplicons from the CPS deficient isolates varied from either no or faint amplicons of predicted size (isolates NS716, NS1061, and NS686; Figure 23A, lanes 12 to 14), to a smaller amplicon (approximately 850 bp; isolate NS303; Figure 23A, lane 11), to an amplicon of expected size and intensity (NS791 and NS384; Figure 23A, lanes 9 and 10). This variability indicates that more than one mutation may be responsible for causing the CPS deficiency in these isolates.

3.5.2 Transformation of gonococci deficient in CPS using cloned *carA* and *carB*: multiple mutations are present in the arginine and pyrimidine biosynthetic pathways of CUH auxotrophs.

A modified version of the method reported by Janik *et al.* (1976) was developed to transform competent gonococci with ligated pFL50/pFL35 DNA fragments containing the cloned *carA* and *carB* genes. The gonococcal isolates were first selectively subcultured to enrich for the T2 colony type (which comprises piliated gonococci which are naturally competent for linear DNA uptake) and then the DNA was mixed with the bacteria on a plate of non-selective media. The bacteria were then transferred to selective media and putative transformants were isolated. Note that allowing the

FIGURE 23. PCR amplification of *carA* (Figure 23A; primers 38 and 71) and *carB* (Figure 23B; primers 78 and 22) from 11 gonococcal isolates. 5 gonococcal isolates were selected for their varying intergenic sequence size (Lanes 3 to 7) and another 6 isolates were selected because they have known CPS deficiencies (Lanes 9 to 14). PCR amplicons were obtained from the same isolates in both cases: Lane 3, CH811; Lane 4, MS11-ms; Lane 5, FA1090; Lane 6, 11860; Lane 7, 9345; Lane 9, NS791; Lane 10, NS384; Lane 11, NS303; Lane 12, NS716; Lane 13, NS1061; Lane 14, NS686; Lane 15, NS686C⁺. Lanes 1, 8 and 17, 1 kb ladder; Lane 2, pFL50; Lane 16, no DNA template control. Note that for the amplification of *carB*, one of the primers lies outside the sequence contained in pFL50 and so no amplicon was produced in this control as expected.



transformants to grow on non-selective media for 6 hours before selection is applied, instead of the 3-4 hours previously reported, greatly increased the number of transformants. Initial experiments were performed with 6 gonococcal isolates known to be CPS deficient (and from which T2 colonies could be obtained), as well as the control strains (see Section 2.10). All control strains produced the expected results, and preliminary results showed that isolates NS686 and NS716 could be transformed to streptomycin resistance (a control) and transformed with the pFL50/pFL35 DNA. The other 4 CPS deficient isolates were not transformed by any DNA, and were not further pursued since this lack of transformation was most likely due to problems with obtaining the T2 phenotype and therefore competent gonococci (gonococcal isolates when grown *in vitro* rapidly lose this phenotype if it has not been selected for; Kellogg *et al.*, 1963). Isolates NS686 and NS716 (which are known by auxotyping to be citrulline, uracil and hypoxanthine requiring) were therefore used as the main test strains. In the controls, these isolates were successfully transformed to prototrophy using chromosomal DNA from strain CH811 and also could be transformed to streptomycin resistance using chromosomal DNA from strain FS62 (Str^R). These isolates did not change in auxotype in a negative control where they were exposed to pBluescriptII KS(+) DNA which had been purified from *E. coli* cells in the same manner as the pFL35/pFL50 DNA.

When the ligated mixture of pFL50/pFL35 DNA was mixed with the competent cells of each isolate (NS686 and NS716), some cells successfully grew on media not containing citrulline indicating successful transformation. No transformants were obtained using uracil media for the selection for both isolates. Selected colonies of transformants which did grow on the citrulline media were auxotyped and surprisingly all were of auxotype OUH (ornithine, uracil, and hypoxanthine requiring). That is, the uracil requirement was still present and transformation to citrulline⁺ had revealed a hidden

ornithine requirement in both isolates (see Figure 1 for the relationship between ornithine and citrulline biosynthesis). PCR amplification of *carA* and *carB*, and the intergenic sequence between these genes, was performed on isolate NS686 and the transformant which was named NS686C⁺. Isolate NS686 produced a faint amplicon of expected size for *carA* (Figure 23A, lane 15), no amplicon for the intergenic sequence (data not shown), and a strong amplicon of expected size for *carB* (Figure 23A, lane 15). The transformant, NS686C⁺, produced amplicons of expected size (as would be predicted from the sequence of pFL50 and pFL35) for *carA* (Figure 23A, lane 16), *carB* (Figure 23B, lane 16), and the intergenic sequence (data not shown).

This transformation of isolates of auxotype CUH, to an auxotype OUH, suggests a possible close relationship between these auxotypes. Previously a 3 bp deletion in the *argJ* gene had been found to cause the ornithine requirement in most OUH isolates (Martin and Mulks, 1992a), and so the isolates NS686 and NS686C⁺ were further investigated to see if they contained this same mutation in *argJ*. PCR amplification of a 299 bp portion of *argJ*, using primers homologous to positions 135-154 and 414-433 in the gene, was performed and the resulting amplicon sequenced in both directions using the PCR primers. The resulting sequence clearly showed the presence of the 3 bp deletion within *argJ* at the expected location (307 bp downstream of the start of the gene) in both isolates NS686 and NS686C⁺, as well as a control strain, 9315, which is of OUH auxotype (data not shown). Another control, the prototrophic strain CH811, did not contain this mutation when analyzed in a similar manner. This further indicates a close relationship between CUH and OUH isolates.

3.6 PHYLOGENETIC ANALYSIS OF CPS GENES: COMPLEX EVOLUTIONARY HISTORY INCLUDES AN INTERNAL DUPLICATION WITHIN A GENE WHICH CAN ROOT THE TREE OF LIFE

3.6.1 The phylogenetic analysis includes analysis of the first CPS genes from an archaeon: *Sulfolobus solfataricus* CPS is found to be encoded by overlapping genes and contain protein motifs consistent with a CPSII enzyme.

Initially, phylogenetic trees were constructed using the gonococcal sequence and all other known CPS genes which at the time did not include an archaeal sequence. However, as part of a subsequent collaboration, Dr. Robert L. Charlebois (Dept. of Biology, University of Ottawa) supplied an archaeal CPS gene sequence, which he had obtained from *S. solfataricus* as part of the *Sulfolobus* genome project, for the phylogenetic analysis. This DNA sequence was analyzed by me and a brief description of my findings is presented below.

The amidotransferase and synthetase domains of the *S. solfataricus* CPS were found to be encoded by two genes, *carA* and *carB*, of sizes 1101 and 3153 bp, respectively. These genes are similar in size to those encoding heterodimeric CPS in other organisms, with the deduced protein sequences showing approximately 37% and 43% identity with *E. coli* and bullfrog (CPSI) sequences, respectively (as reference, *E. coli* and bullfrog CPS share 37% identity, and *E. coli* and *P. aeruginosa* CPS share 68% identity). The *S. solfataricus* CPS genes were found to be transcribed in the same direction in the order *carA-carB* and overlap by 4 bp.

Analysis of the deduced protein sequences from this heterodimeric CPS of *S. solfataricus* showed that it contained conserved domains found in other CPS enzymes. Alignment of the amidotransferase and synthetase domains with all other known CPS sequences is shown in Figures

A1 and A2, respectively. Within the synthetase domain of the CPS, two cysteine residues implicated in acetylglutamate binding, which are found in the synthetase domain of CPSI and CPSIII enzymes (Geschwill and Lumper, 1989), are not present. The location where these residues would be located is highlighted in Figure A2. Within the amidotransferase domain, residues which have been implicated in glutamine binding for CPSII, in particular the reactive cysteine residue (Rubino *et al.*, 1986; Miran *et al.*, 1991), are present, suggesting that this enzyme is most similar to CPSII (see Figure A1 for the location of the reactive cysteine residue). The residues thought to be involved in ATP binding, and found in the synthetase domain of all CPS's (Post *et al.*, 1990), are present.

Notably, this archaeal enzyme contained internal similarity between the first and second thirds of the synthetase domain, as has been observed in all bacteria and eukaryotes. The presence of this internal similarity in the archaeal sequence confirms that the synthetase domain likely evolved from an ancient gene duplication and shows that this duplication event occurred before the divergence of Bacteria, Eukarya and Archaea.

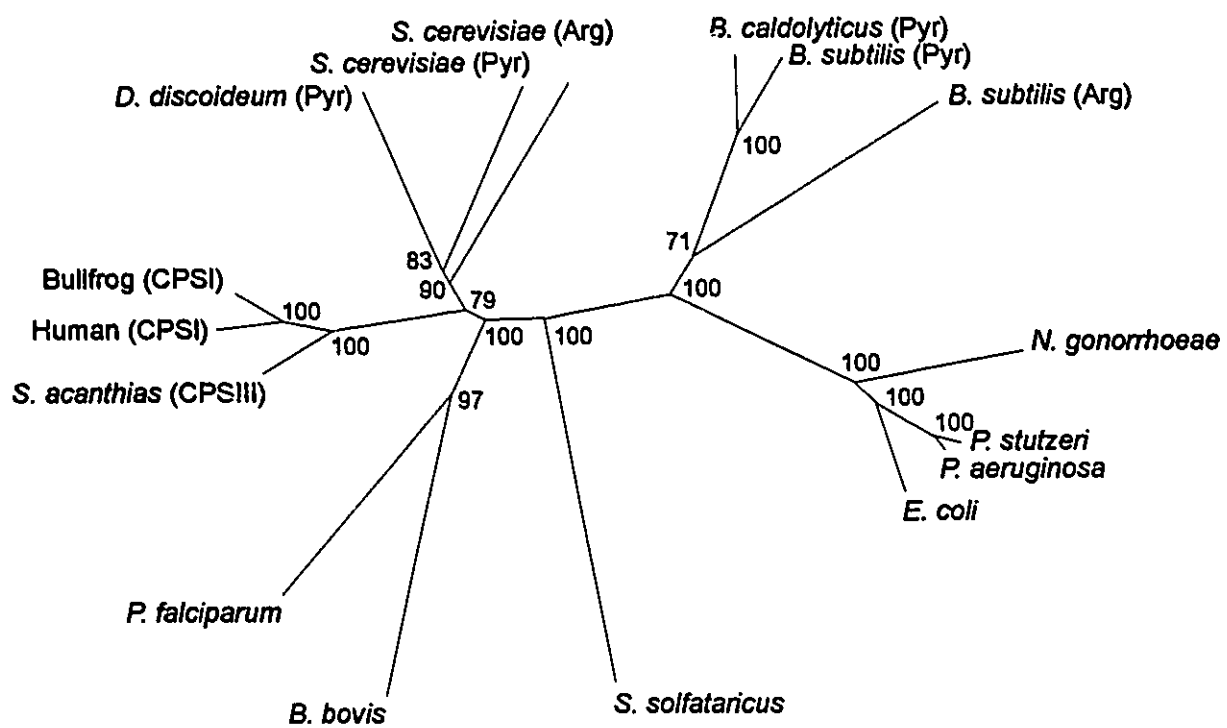
3.6.2 Phylogenetic analysis of complete CPS genes: at least two separate gene duplication events led to the formation of the two CPS enzymes found in Gram-positive bacteria and in eukaryotes.

A total of 40 phylogenetic trees were constructed using five different methods (maximum parsimony, Kitsch, Fitch and neighbour-joining distance matrix methods, and maximum likelihood) to analyze both protein and DNA sequences of four different alignments (the amidotransferase domain of CPS, the synthetase domain of CPS, complete CPS sequences, and the duplication within the synthetase domain). Alignments of the deduced protein sequences of all

the amidotransferase and synthetase domain sequences are shown in Figures A1 and A2, respectively.

Trees constructed using complete CPS sequences (*i.e.* combined amidotransferase and synthetase domain) showed high confidence in branching order for almost all organisms with 9 out of 14 nodes consistently showing bootstrap values of 100. An example of one tree, constructed using the Fitch distance-matrix method with protein sequences, is shown in Figure 24. In this tree the organisms grouped into six clusters comprising the Proteobacteria (*P. stutzeri*, *P. aeruginosa*, *E. coli*, *N. gonorrhoeae*), Gram-positive bacteria (*B. caldolyticus*, *B. subtilis*), Archaea (*S. solfataricus*), Apicomplexa (*P. falciparum*, *B. bovis*), eukaryotic CPSII arginine and pyrimidine-specific enzymes (*S. cerevisiae* Arg, *S. cerevisiae* Pyr, *D. discoideum* Pyr), and eukaryotic CPSI and CPSIII enzymes (*S. acanthias* CPSIII, bullfrog CPSI, human CPSI). Similar branching order was noted in all trees constructed, with some exceptions which are illustrated in some examples of trees constructed using just the amidotransferase or just the synthetase domain of CPS (Appendix, Figures A3 through A6). These exceptions included the *B. subtilis* arginine-specific (Arg) CPS sometimes forming its own monophyletic group (Figure A3 and A4), the arginine-specific (Arg), pyrimidine-specific (Pyr) and urea cycle-specific (CPSI, CPSIII) CPS's of eukaryotes forming three separate clusters which varied in branching order relative to one another (Figure A3 and A6; discussed further in Section 3.6.3), and a variable branching order for the pyrimidine-specific CPS enzymes of eukaryotes (Figure A5 and A6). These figures shown in the Appendix were specifically chosen to illustrate some of these differences in branching order and do not reflect the most common branching orders observed. Other variations were specifically observed in trees constructed from DNA sequences rather than protein, which were directly

FIGURE 24. Phylogenetic tree constructed from alignments of deduced amino acid sequences of complete CPS genes, using the Fitch distance matrix method. Bootstrap values (out of 100 replicates) are shown at each node with the scale for branch lengths shown below the figure. For organisms which contain two CPS enzymes, the sequence is identified by the letters "Arg" (for arginine-specific CPSII), "Pyr" (for pyrimidine-specific CPSII), "CPSI", and "CPSIII".



— 0.1 substitutions

attributable to differences in GC content among organisms.

When trees were constructed using just the amidotransferase or synthetase domains of CPS, other sequences could be included such as the *N. crassa* (Arg) sequence and the *T. cutaneum* (Arg) sequence, which both clustered with *S. cerevisiae* (Arg) as a monophyletic group. Similarly, *S. typhimurium* clustered with *E. coli*, and rat (CPSI) grouped with human (CPSI), as expected.

Of particular interest in all phylogenetic trees constructed is that the two CPS's present in the Gram-positive bacteria and eukaryotes did not cluster together, confirming that separate gene duplication events led to the formation of two CPS enzymes in these organisms. Furthermore, the eukaryotic CPS enzymes used in arginine-biosynthesis (Arg) and the urea cycle (CPSI, CPSIII) never clustered together, suggesting a possible separate evolutionary origin for the two.

3.6.3 Phylogenetic trees constructed using the internal duplication within the synthetase domain of CPS: rooting the tree of life groups an Archaea with eukaryotes.

Internal similarity is observed within the first and second thirds of the synthetase domain of all CPS sequences studied to date, including within the newly reported archaeal sequence. Phylogenetic trees were constructed using alignments of these two similar regions of the synthetase domain and an example is shown in Figure 25. Sequences corresponding to the first half of the duplicated region clustered together as one monophyletic group and the sequences corresponding to the second half of this region formed a similar cluster. This demonstrates that the duplication which formed the ancestral synthetase domain preceded divergence of the Bacteria, Eukarya, and Archaea, and shows that this duplication could be used to root the tree of life. Within all trees constructed using this duplication, the archaeal sequence grouped with high confidence with the eukaryotes, and always

FIGURE 25. Phylogenetic tree constructed from alignments of the deduced protein sequence of the the first and second thirds of the synthetase domain of CPS's, using the Neighbor-joining distance matrix method. Bootstrap values (out of 100 replicates) are shown at each node with the scale for branch lengths shown below the figure. The number preceding the name of each organism indicates whether that sequence was derived from 1: the first third of the synthetase domain or 2: the second third of the synthetase domain. For organisms which contain two CPS enzymes, the sequence is further identified by the letters "Arg" (for arginine-specific CPSII), "Pyr" (for pyrimidine-specific CPSII), "CPSI", and "CPSIII". Note that the branching order determined from both thirds of the synthetase domain are similar, with the exception that the eukaryotic "Pyr" sequences cluster together in one case (second third of the synthetase domain) and in the other case (first third) they branch separately from one another. This branching order determined from the first third of the synthetase domain is not observed using any other method of tree construction, although the branching order for these eukaryotic "Pyr" sequences still varies.

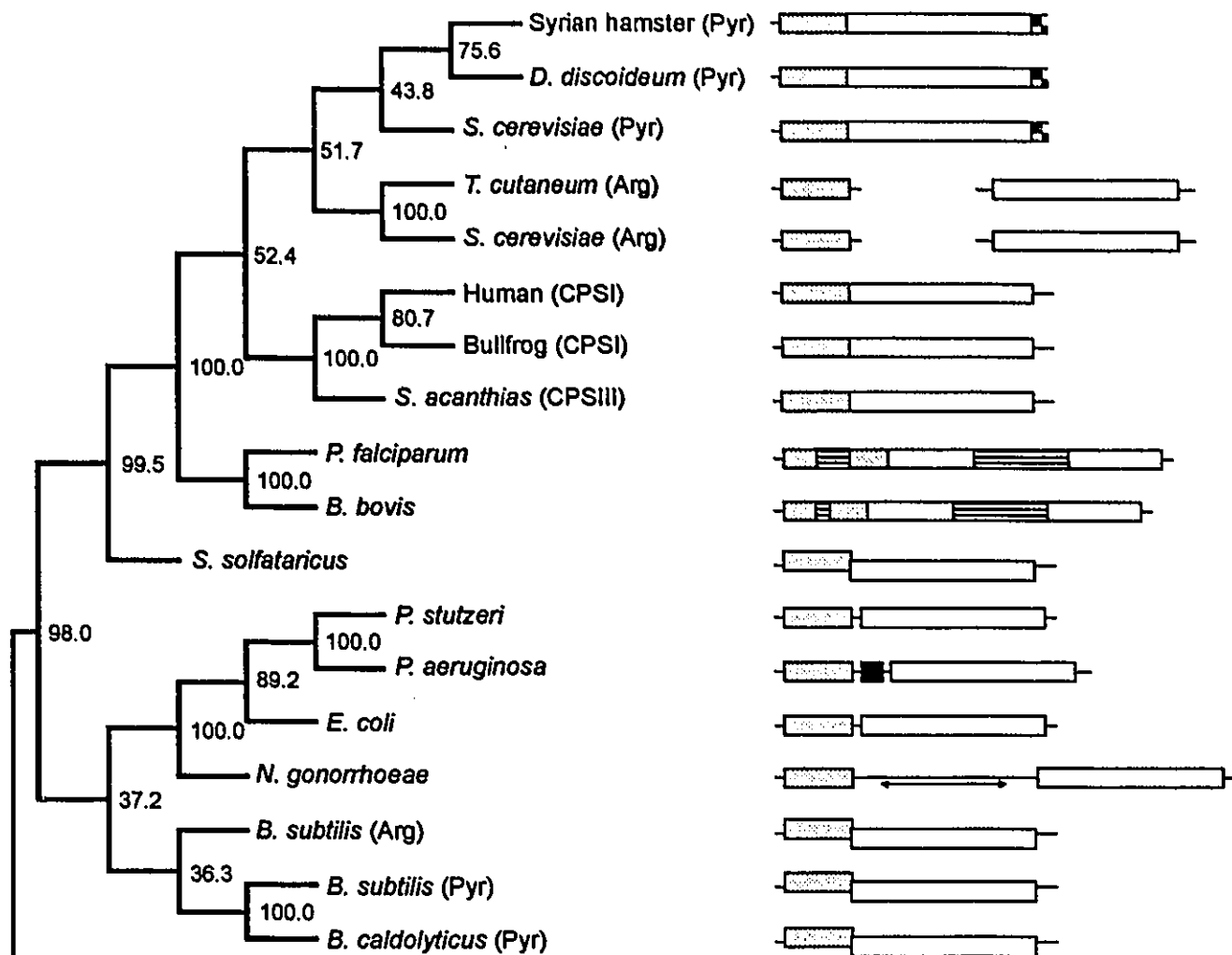
as the deepest branch in the clade. The eukaryotes clustered together, with the apicomplexan protozoans branching at the deepest point in most trees. This is notable since these protozoans have only one CPS enzyme, while all other eukaryotes have two. For the eukaryotes which contain two CPS enzymes, three clusters were commonly found within this group as previously described (the arginine-specific CPS's, the urea cycle-specific CPS's (CPSI, CPSIII), and the pyrimidine-specific enzymes). The branching order within these groups varied such that sometimes the urea cycle-specific CPS's would form the deepest branch within this group and sometimes the pyrimidine-specific or arginine-specific CPS's would form the deepest branch. A more variable branching order was sometimes observed for the eukaryotic pyrimidine-specific sequences of the first third of the synthetase domain, as shown in Figure 25, however, this was not observed in the trees constructed from the second third of the synthetase domain. Also, even when the branching order for these pyrimidine-specific sequences was variable, the eukaryotic arginine-specific CPS's and CPSI/CPSIII enzymes never clustered together.

This duplication within the synthetase domain of CPS was also used as a root for determining phylogenetic relationships among bacteria. Gram-positive bacteria and Proteobacteria formed two distinct groups and branching order for the Proteobacteria reflected previously determined phylogenetic relationships between organisms (Woese, 1987) suggesting that the proteobacterial CPS genes evolved from a common ancestor. The branching order for the two Gram-positive bacteria suggests that an isolated gene duplication occurred after the divergence of the Gram-positives from the proteobacteria to form the two CPS enzymes found in these organisms.

3.6.4 Evolution of the organization of CPS genes: the intervening sequence between the amidotransferase and synthetase domains of heterodimeric CPS involved multiple deletions and insertions.

The general organization of CPS genes, not including introns, is shown schematically in Figure 26 alongside one of the phylogenetic trees constructed. Comparison of the intervening sequences between the amidotransferase and synthetase domains of CPS in different organisms shows that this sequence has varied significantly over time, particularly within the Proteobacteria examined so far. Now confirming the branching order for these lineages, one can see that multiple deletions and insertions must have occurred in this intervening region. For example an insertion event likely led to the formation of an open reading frame between the amidotransferase and synthetase genes of *P. aeruginosa*, since the corresponding *E. coli* and *P. stutzeri* genes have a similar structure comprising a very small intervening sequence. The unlinked amidotransferase and synthetase genes of the *S. cerevisiae*, and *T. cutaneum* arginine-specific CPS's were likely due to a separate gene duplication or translocation event since all the other CPS genes in eukaryotes, including the single CPS enzymes in the apicomplexan protozoans which branch lower on the tree, are fused. A few other different scenarios for the evolution of the organization of CPS genes and the intervening sequence between them is possible, but all involve multiple deletions and insertions.

FIGURE 26. Phylogenetic tree constructed from alignments of the deduced protein sequence of the first and second thirds of the synthetase domain of CPS's, using the method of maximum parsimony. Only the branching order determined from the second third of the protein is shown, with the first third used as a root. Bootstrap values (out of 100 replicates) are shown at each node. For organisms which contain two CPS enzymes, sequences are further identified by the letters "Arg" (arginine-specific CPSII), "Pyr" (pyrimidine-specific CPSII), "CPSI" and "CPSIII". The organization of the CPS genes in each organism is also shown schematically, with regions encoding the common amidotransferase and synthetase domains of CPS shown as shaded and white boxes, respectively. Introns are not included. Note that the *S. cerevisiae* (Arg) and *T. cutaneum* (Arg) amidotransferase and synthetase domains are encoded by unlinked genes, and that the corresponding genes in *S. solfataricus* and *Bacillus* species overlap. Checkered boxes denote a portion of the aspartate transcarbamylase domain which is fused to the pyrimidine-specific CPS genes of eukaryotes. Striped boxes denote large translated sequences which are present between functional domains of the CPS of *P. falciparum* and *B. bovis*. The black box represents an unidentified open reading frame found between the *P. aeruginosa* CPS genes. The arrow denotes a variable intergenic sequence present in isolates of *N. gonorrhoeae*. Note that the *N. gonorrhoeae* CPS genes are separately transcribed, while all other bacterial CPS genes are part of operons.



Tree rooted by N-terminal region
of synthetase domain of CPS

4. DISCUSSION

4.1 NOVEL GENE ORGANIZATION FOR *CAR*A AND *CAR*B IN *N. GONORRHOEAE*: SEPARATE TRANSCRIPTION OF *CAR*A AND *CAR*B, AND THE PRESENCE OF A VARIABLE INTERGENIC SEQUENCE, LEADS TO MANY NEW HYPOTHESES REGARDING GONOCOCCAL CPS GENE REGULATION

Initially I hypothesised that the CPS genes of *N. gonorrhoeae* would be similar in gene organisation and structure to that of the other bacteria which had been examined at the time (namely *E. coli*, *S. typhimurium* and *P. stutzeri*), though the regulation of these genes would be different. However, some surprising findings were made over the course of this project with respect to gonococcal CPS gene organisation, somewhat changing the focus of this study.

While the size and sequence of the *carA* and *carB* genes in *N. gonorrhoeae* are similar to those of other prokaryotes and encode protein motifs consistent with a prokaryotic CPSII enzyme, the organisation of the *carA* and *carB* genes in *N. gonorrhoeae* is significantly different from that of other previously characterised prokaryotes. The presence of the 3.7 to 2.2 kb intervening sequence between *carA* and *carB* in the gonococcal isolates examined, contrasts with the 17, 18, and 24 bp of sequence seen between these genes in *E. coli*, *S. typhimurium*, and *P. stutzeri*, respectively (Piette *et al.*, 1984; Kilstrup *et al.*, 1988, Kwon *et al.*, 1994). The 3290 bp intervening region sequenced from *N. gonorrhoeae* CH811, with its novel complex repeats, differs noticeably from the 682 bp intervening sequence recently found in *P. aeruginosa* (Kwon *et al.*, 1994), which has no comparable secondary structure and no significant sequence similarity to the gonococcal sequence. In *P. aeruginosa*, and all other prokaryotes, *carA* and *carB* are co-transcribed (Piette *et al.*, 1984; Kilstrup *et al.*, 1988, Kwon *et*

al., 1994; Quinn *et al.*, 1991; Ghim *et al.*, 1994). Based on DNA sequence and Northern blot analysis, I have shown that *carA* and *carB* are separately transcribed in *N. gonorrhoeae* CH811.

This separate transcription of *carA* and *carB* leads to many interesting hypotheses. The size of the RNA transcripts encoding *carA* and *carB* in *N. gonorrhoeae* CH811 do not seem to be much bigger than the size of the genes they encode. More accurate mapping of the transcription start sites is needed, and a clearer Northern blot is desirable, however the size of the transcripts estimated from the Northern blot analysis performed does suggest that the main promoter and terminator for each gene is located immediately upstream and downstream of each gene, respectively. Indeed, according to the sequence obtained from pFL50 and pFL35, strong putative transcription terminators are found immediately downstream of both genes. These putative transcription terminators contain gonococcal uptake sequences with additional A or T residues. Similar inverted repeats are found downstream of other gonococcal genes, and they resemble the Rho-independent transcription terminators identified in *E. coli* (Elkins *et al.*, 1991). Recently, it has been demonstrated that these inverted repeats containing the gonococcal uptake sequence can terminate transcription in *E. coli* in an orientation independent manner (Barber *et al.*, 1994).

The location of the *carB* promoter(s) is of particular interest, since promoters could have been located further upstream within the intergenic sequence between the two genes. If the main promoter is not located significantly further upstream of *carB*, no other products may be co-transcribed with this gene. It is possible, in fact, that no products are produced at all from this sequence between *carA* and *carB* since sequence analysis of this region does not indicate the presence of any open reading frames which have codon usage's consistent with a gonococcal gene, and there is no significant sequence similarity between this region and any other currently known genes (with the exception of the unusual

region of sequence found in GenBank sequence NGOFUSAA; see Section 3.2.2). Furthermore, preliminary *in vitro* transcription-translation studies of the plasmid pFL50 performed in our laboratory (Bernatchez and Dillon, unpublished data) indicate that no products are produced from the gonococcal sequence between *carA* and *carB* in an *E. coli* background (though a product of approximately 40 kDa, corresponding to the size of *carA*, was produced). However, these studies may not be able to detect very small products or the products may only be expressed in a gonococcal background. Northern blots of gonococcal RNA probed with the intergenic sequence may help resolve this issue. This intergenic region varies significantly in sequence between the different gonococcal isolates, and so it is possible that products are produced from this region in other gonococcal isolates. My REA and PCR amplification studies indicate that the region directly upstream of *carB* is significantly different in some gonococcal isolates from the sequence obtained from pFL50. The promoter for *carB* may therefore be correspondingly varying in location in these other isolates, or there may be selection for different promoters in different isolates.

Another point worth investigating would be whether there are one or two (or more) promoters upstream of *carA* and/or *carB*. In *E. coli* and *S. typhimurium* there are two promoters upstream of *carA*, one regulated by pyrimidines and the other regulated mostly by arginine but also by pyrimidines to a lesser degree (Kilstrup *et al.*, 1988; Lu *et al.*, 1989; Piette *et al.*, 1984). In *P. aeruginosa* there is only one promoter, with a proposed attenuation mechanism which regulates expression by pyrimidines (Kwon *et al.*, 1994). Preliminary primer extension experiments which I recently performed suggested the presence of a transcription start site immediately upstream of *carA*, however it was not clear whether there were one or two promoters present, due to unoptimal reaction conditions. Without conclusive primer extension experiments to map transcription start sites, no promoters could be

conclusively identified in the gonococcal sequence, however the lack of similarity between the sequence upstream of *carA* in the gonococcus and these other organisms suggests that the regulation of the gonococcal CPS genes is different from other bacteria. Notably, no arg-boxes are present upstream of the gonococcal *carA* gene, or *carB*, indicating that an arginine repressor analogous to that found in the enterics and *B. subtilis* is not present in the gonococcus. One interesting feature noted upstream of *carA*, however, was a string of 9 thymidine residues which would be uracil nucleotides if transcribed. No significant inverted repeats could be found in this region and so no possible attenuation mechanism has been identified, however, this poly-T region may still be involved in the regulation of *carA* by some other mechanism. For example, if these thymidines varied in number close to a promoter, they could affect the promoter's expression. Recently, the Opc outer membrane protein of *N. meningitidis* has been shown to undergo phase variation in this way, through size-variation of a polycytidine region associated with a promoter (Sarkari *et al.*, 1994). If this polyT region is transcribed and translated, a slip-strand mispairing mechanism could be involved. Many gonococcal genes encoding virulence factors undergo phase variation via slip-strand mispairing of a repetitive sequence at the 5' end of the gene; most notably this occurs in the opacity surface-protein genes (variation in the number of CTCTT repeats), in the pilus biogenesis factor gene *pilC* (variation in a polyG region), and in lipooligosaccharide genes (variation in a polyG region; Gotschlich, 1994; Murphy *et al.*, 1989; Belland *et al.*, 1989). Interestingly, my preliminary primer extension experiments did suggest that a transcription start site was located in the vicinity of this polyT region.

The separate transcription of gonococcal CPS genes shows similarities with the yeast and *Neurospora crassa* arginine-specific CPS genes, which are the only other documented cases of separate transcription of CPS (Lacroute *et al.*, 1965; Davis, 1986). In these fungal CPS's, the small

subunit is repressible by arginine while the large subunit is not, and the small subunit essentially behaves as a rate-limiting co-factor for the large subunit (Cybis and Davis, 1975; Piérard *et al.*, 1979). It would be interesting to see if, in the gonococcus, different promoter sequences are present upstream of *carA* and *carB*, and possibly a similar asymmetrical regulation of the subunits is occurring. Asymmetric regulation of subunits is seen with other enzymes which, like CPS, contain glutamine amidotransferase subunits. The glutamine aminotransferase encoded by *trpG* is amphibolic, being utilised as part of enzymes involved in both tryptophan and folate biosynthesis (Kane *et al.*, 1972; Slock *et al.*, 1990).

Clearly, there is mounting evidence that studies of the expression and regulation of the CPS genes in *N. gonorrhoeae* could prove interesting. This first documented case of separate transcription of CPS genes in a prokaryote suggests that CPS gene organisation may be even more variable than previously thought. The separate transcription of gonococcal CPS genes could, however, also be a reflection of a general lack of operonic structures seen in the gonococcal genome. The gonococcus is known for its heterogeneity (O'Rourke and Spratt, 1994; O'Rourke and Stevens, 1993), and it could be advantageous to the gonococcus to have compact genetic units so that recombination in the genome would be less likely to have a deleterious effect.

Recently in our laboratory, Billowes *et al.* (unpublished) sequenced the intergenic sequences between *carA* and *carB* from gonococcal isolates MS11 and FA1090 and also sequenced this region from a meningococcal and a commensal *Neisseria* species isolate. Putative transcription terminators were found downstream of *carA* in all of these organisms. The sequence immediately upstream of *carB* was found to be significantly different, while the sequence downstream of *carA* was found to be almost identical in the gonococcal isolates, as has been indicated in this work.

4.2 *CAR*A AND *CAR*B MAY RESIDE WITHIN A HOTSPOT FOR RECOMBINATION: PARTIAL EXPLANATION FOR NATURALLY OCCURRING CPS DEFICIENCIES

Numerous novel and previously characterised repetitive sequences are found around the CPS genes of *N. gonorrhoeae* CH811, and the size of the intervening sequence between these genes varies in gonococcal isolates, suggesting that this region of the gonococcal genome may be a hotspot for recombination. This could explain in part why CPS deficiencies are so often found in the gonococcus. The repeats found around these genes include copies of RS2 and RS3, sequences previously found around silent and expressed *pil* loci. They are thought to confer more homology between these sites, possibly aiding in their recombination (Haas and Meyer, 1986). My cloned *carA* and *carB* genes map near *pil* loci on both *N. gonorrhoeae* MS11 and *N. gonorrhoeae* FA1090 physical maps (Gibbs, 1991; Dempsey and Cannon, 1994), and according to recent sequence submissions to GenBank there is a *pilC* locus located directly upstream of *carA* in *N. gonorrhoeae* MS11 (Rudel *et al.*, 1995).

The repetitive sequences found around *carA* and *carB* also show significant secondary structure. A copy of RS2 found upstream of the *carA* in *N. gonorrhoeae* CH811 is flanked by complex inverted repeats, reminiscent of a mobile element. A copy of RS5, also found upstream of *carA*, contains 19 bp inverted repeats at either end. Gonococcal uptake sequences are found throughout the sequence containing *carA* and *carB*, most prominently as inverted repeats upstream and downstream of *carA* and downstream of *carB*. Notably, gonococcal uptake sequences flank the novel direct repeats, RS6 and RS7, found in the sequence between *carA* and *carB*. RS6 and RS7 contain multiple inverted repeats, adding to their complexity. Mapping of these repeats suggests that they are not located elsewhere on the gonococcal genome and so these repeats may be an isolated string of duplications. Alignment of the repeats does show a possible succession of origin for these repeats, as if

the repeats were formed due to successive recombination events which resulted in sequence duplications. These repeats are likely candidates for explaining most of the variability seen in the intervening sequence between *carA* and *carB* in different gonococcal isolates. The size of the variation in the intergenic sequence in most cases corresponds to the size of these repeats. The mechanism for this variation in the number of repeats can not be conclusively determined at this time, however large-scale slip-strand mispairing, unequal cross-over between newly replicated DNA strands, or transformation and recombination with DNA from another strain could have led to these changes. Slip-strand mispairing, though most commonly used to explain deletions of very short repeats, has been postulated to occur between repeats which are within as much as 400 bp of each other (Chédin *et al.*, 1994). Palindromic sequences immediately flanked by direct repeats strongly stimulate deletion formation (Glickman and Ripley, 1984) and RS6 and RS7 both contain inverted repeats and are flanked by direct repeats of the gonococcal uptake sequence. The inverted repeats are thought to aid in bringing the ends of the repeats together. Transformation and recombination with DNA from another strain is also a likely mechanism, since gonococci are naturally competent for DNA containing the gonococcal uptake sequence, and so frequently undergo recombination by this mechanism (Halter *et al.*, 1989; Zhou and Spratt, 1992).

Interestingly, some variation also seems to occur in the region directly upstream of *carB* which, in gonococcal isolate CH811, does not contain these repeats. The degree of this variation is unknown and could be a fairly infrequent occurrence relative to the frequency of changes in the number of RS6/RS7 repeats. It could be that RS6/RS7 repeats are present in this region in some isolates or it could be that other repeats are responsible for this variation.

Nevertheless, if this sequence region is a hotspot for recombination, one could see how variations in the intergenic sequence could have occurred and the RS6/RS7 repeats could have evolved. One could also see how CPS deficient mutants could have naturally occurred. PCR amplifications have provided preliminary evidence that more than one mutation occurred to form the naturally occurring CPS deficiencies.

Though this region of the gonococcal genome seems to be a hotspot for recombination, it should be noted that many studies have indicated that there is more involved in the occurrence of CPS deficiencies in the gonococcus. That is, that there is a selective advantage for the productions of these CPS deficiencies in gonococcal isolates. Gonococcal isolates passaged through mice develop nutrient requirements consistent with a CPS deficiency (Kenyon, 1978). Clinical isolates of *N. gonorrhoeae* with auxotypes consistent with a CPS deficiency did not start to appear until after penicillin was first used for treatment of gonorrhea (Catlin and Reyn, 1982). These isolates with CPS deficiencies grow slower and it is thought that this slower growth may help them evade the effects of penicillin (Mayer *et al.*, 1977). Isolates with auxotypes consistent with a CPS deficiency have also been found in one study to cause a lower grade infection (Whittington *et al.*, 1993). Whittington *et al.* propose that there may be selection of these gonococcal auxotrophs because patients with a lower grade infection will delay seeking treatment and therefore the chances of its spread to others are higher. In 1995, Rudel submitted a *carA* sequence to GenBank (accession no. Z54242) with an attached unpublished publication reference entitled "The *Neisseria gonorrhoeae carA* gene: Generation of mutants and implication of pyrimidine biosynthesis for the infection of epithelial cells." Unfortunately, the results of this study are not yet known.

4.3 COMPARISON OF THE SEQUENCE BETWEEN *CARA* AND *CARB* IN *NEISSERIA* SPECIES: REVEALING SEQUENCES WHICH MAY BE DIAGNOSTIC FOR *N. GONORRHOEAE*

This is the first investigation of CPS gene organization from multiple isolates of the same species, and it has produced the first report of intraspecies variation in a sequence proximate to CPS genes. This variable sequence is not limited to gonococcal isolates, but is also present in other *Neisseria* species.

Hybridization studies have demonstrated that this variable intergenic sequence is somewhat conserved between the *Neisseria* species. Using portions of the intergenic sequence of *N. gonorrhoeae* CH811 as a basis for comparison, gonococcal intergenic sequences were found to be the most similar, followed by meningococcal sequences, followed by the commensal *Neisseria* species intergenic sequences. The meningococcal sequences all seemed to have a region of sequence which had significant similarity with CH811's sequence, since hybridization between the two did occur at high stringency. This similar region did not comprise the RS6/RS7 repeats, however, since the repeats only hybridized with the meningococcal intergenic sequences at a lower stringency. The commensal *Neisseria* species had no highly similar region in their intergenic sequence when compared against CH811, however at lower stringency some hybridization was observed with most isolates. Commensal *Neisseria* isolates with a small intergenic sequence size showed little to no similarity between their sequence and that of CH811, however that does not discount them having similarity with the intergenic sequences of other gonococcal isolates or other *Neisseria* species.

The hybridization studies also indicated that the RS6/RS7 repeats found in *N. gonorrhoeae* CH811 may be species-specific for *N. gonorrhoeae* under conditions of high stringency. This combined with the fact that RS6 and RS7 have no significant similarity to any

other sequences in GenBank, and did not hybridize at high stringency to chromosomal DNA from *E. coli* or *P. magnus*, provides preliminary evidence that these sequences may be diagnostic for *N. gonorrhoeae*. Other genera need to be screened, though, to confirm this species specificity. The hybridization studies do suggest that the meningococcal isolates have analogous sequences, however, these sequences are likely an isolated series of repeats, since RS6/RS7 only map to one location (with *carA* and *carB*) on both the gonococcal and meningococcal chromosome maps (at a low stringency). I would predict that other *Neisseria* species isolates could have analogous repeats, however, it is still possible that these repeats (or analogous ones) are limited to the Neisseriaceae.

While the hybridization studies showed that some similarities existed between the intergenic sequences of the different neisserial species, REA of these amplicons (using *DdeI* and *HinfI*) revealed noticeable differences between the species. Similar banding patterns were seen within species, however no similar patterns were seen between species. There were however a couple of notable exceptions within the commensal *Neisseria* group: Identical patterns were seen between an isolate of *N. lactamica* and *N. perflava/sicca*, and an isolate of *N. mucosa* and *N. flava*. The production of identical patterns from these different species, when in all other cases the patterns showed no interspecies similarity, suggests that these isolates may be more related than is indicated by classical speciating criteria and in fact could be being misclassified. As mentioned, similarities in REA patterns were observed within species. In particular a common band was found in all gonococcal isolates examined, which mapping and PCR experiments revealed corresponds to a 1 kb fragment which is located immediately downstream of *carA*. Just why this sequence is conserved in the gonococcal isolates is unknown.

4.4 UTILITY OF THE VARIABLE INTERGENIC SEQUENCE FOR TYPING OF GONOCOCCAL ISOLATES

The size of the amplicons of the intergenic sequence, and the REA patterns produced from these amplicons, has further elucidated relationships between gonococcal isolates. For example, according to amplicon size, isolates of A/S/P class CUH/1A-2/2.6 and P/1B-5/2.6/24.5 are not homogeneous. Both amplicon size and REA were in fact able to significantly further discriminate many isolates of the same A/S/P class, even isolates associated with outbreaks of gonorrhea which may be closely related. This method of PCR-REA of the intergenic sequence would not likely be useful for typing of very unrelated isolates, since pattern reversion is probably common, however it could be useful for typing of more closely related isolates, such as isolates associated with an outbreak or within a small geographic location. It could also be useful as an additional marker when confirmation of similarity between isolates is needed. Sets of patient-pair, epidemiologically linked, isolates would be worth investigating. Levels of discrimination using this PCR-REA method seem to be analogous to that of the chromosomal REA and ribotyping methods performed by Li (1993), since the two sets of isolates examined by all three methods were often discriminated into a similar number of groups. This PCR-REA of the intergenic sequence is, however, more easy to interpret than the chromosomal REA method and ribotyping is a longer procedure to perform. PFGE was more discriminatory than the PCR-REA method, separating the isolates into more groups in all cases, however there have been reports that the level of discrimination with PFGE is so high that isolates which only have one to two band differences in their PFGE patterns should still be considered to be very closely related and from the same outbreak (Li and Dillon, 1995; Tenover *et al.*, 1995).

REA of this intergenic region using other restriction enzymes may also be worth investigating. I have also found preliminary evidence that the variable portion of the intergenic sequence may be amplified from all gonococcal isolates, including isolates with CPS deficiencies, by using other primer pairs.

Interestingly, REA of the intergenic sequence of meningococcal isolates discriminated between ET-15 isolates and non-ET-15 isolates of *N. meningitidis* serovar C. Though the sample size is small, this may be worth further study, since electrophoretic typing is cumbersome and the serovar C/ET-15 isolates are noted for causing significant outbreaks of meningitis in Canada (Ashton *et al.*, 1991; Strathdee *et al.*, 1993). The REA also suggested there were significant similarities between some commensal *Neisseria* species, which are often hard to speciate. This method could also be useful for investigating relationships within and between these other pathogenic and commensal *Neisseria* species.

4.5 STUDIES OF THE CPS DEFICIENT MUTANTS REVEAL NEW RELATIONSHIPS WITHIN AND BETWEEN AUXOTYPES

PCR amplification of the *carA* and *carB* genes themselves revealed heterogeneity within the CPS deficient mutants. More than one mutation seems to have caused gonococcal CPS deficiencies. This contrasts with studies of gonococcal OAT (*argJ*) mutants, which seem to have evolved clonally from one mutation (Martin and Mulks, 1992a). This could be a consequence of selection for CPS mutations but, regardless, is likely a reflection of the proposed hot spot for recombination present in this region. This occurrence of multiple mutations causing CPS deficiencies could enable certain auxotypes of gonococci to be further differentiated.

Transformation of a CPS mutant with the cloned CPS genes also revealed relationships between auxotypes. The addition of functional CPS genes to CPS deficient isolates of auxotype CUH (isolates NS686 and NS716) converted their auxotype to OUH. Previously the C and U components of the CUH auxotrophs were thought to be caused solely by the CPS deficiency. With this scenario the auxotype would have become just H after transformation. However, these studies have revealed that other mutations are hidden by this CPS deficiency and that in fact the CUH isolates examined have another mutation in the uracil biosynthetic pathway and also in the arginine biosynthetic pathway (before the biosynthesis of ornithine). This leads to the question of just how many other mutations may be hidden in these pathways. This also suggests a possible close relationship between isolates of auxotype OUH and CUH (both common auxotypes as opposed to auxotype H which has never been identified). The mutation in *argJ* thought responsible for the ornithine-requirement in most OUH isolates (Martin and Mulks, 1992a), has also been found to be present in the *argJ* gene of both isolate NS686 (CUH) and the same isolate transformed with CPS genes, NS686C⁺(OUH). This further indicates a close relationship between the CUH and OUH auxotrophs.

4.6 PHYLOGENETIC ANALYSIS OF CPS GENES: ROOTING THE TREE OF LIFE AND REVEALING A COMPLEX HISTORY FOR CPS GENE EVOLUTION

After initial studies of gonococcal CPS gene structure were completed, I realized that no comprehensive phylogenetic analysis of all known CPS genes had yet been performed, even though the synthetase domain of CPS had an internal duplication which could be used to root the trees constructed. A phylogenetic analysis was performed using all available complete CPS gene sequences, including my gonococcal sequence and the *S. solfataricus* sequence, described below.

As part of this study I performed an analysis of the sequence containing the *S. solfataricus* CPS genes. This represents the first analysis of a complete CPS gene sequence from an archaeon. The analysis shows that this archaeal CPS is heterodimeric and encoded by overlapping genes (*carA-carB*), which are likely co-transcribed. Charlebois (personal communication) has also observed that immediately upstream of *carA* are genes encoding the final two steps of arginine biosynthesis (*argG* and *argH*) in an arrangement suggesting co-transcription with *carAB*.

These *S. solfataricus* CPS genes are similar in size to those of other organisms with heterodimeric CPS, encoding an amidotransferase subunit of predicted Mr 41,480 and a synthetase subunit of Mr 118,204. This is notable since Legrain *et al.* (1995) recently characterized the CPS enzyme from the archaeon *Pyrococcus furiosus*, and found it to be atypical in size (Mr 70,000). They also found that this *P. furiosus* CPS uses ammonia and not glutamine as its nitrogen donor. My analysis of the deduced protein sequence from the *S. solfataricus* genes shows the existence of the reactive cysteine residue known to be involved in glutamine binding for glutamine-dependent CPS enzymes, along with other nearby histidine residues also known to play a part (Rubino *et al.*, 1986; Miran *et al.*, 1991). However, without further study of the *S. solfataricus* CPS protein, few conclusions can be made at this point regarding its ability to bind glutamine, especially since recent findings (the CPSI of bullfrog contains the cysteine residue known to bind glutamine and yet is ammonia-dependent) have shown that other as yet unidentified residues must also play a part in glutamine-binding (Helbing and Atkinson, 1994). However, all CPSI and CPSIII enzymes have an acetylglutamate binding site and an absolute requirement for acetylglutamate in order to function, and the *S. solfataricus* sequence does not contain this acetylglutamate binding site. Therefore, even though there is controversy regarding all

the residues needed for glutamine-utilization by CPS, the *S. solfataricus* CPS seems to be most similar to CPSII enzymes, since it does contain residues found in CPSII which are known to be involved in glutamine binding and does not contain residues involved in acetylglutamate binding which are not found in CPSII enzymes.

One of the most notable features of this archaeal CPS sequence is the presence of internal similarity between the first and second thirds of the synthetase domain, encoded by *carB*. This confirms a previous report by Schofield (1993), who obtained enough partial sequence of a CPS gene from the archaeon *Methanosarcina barkeri* to determine that this internal similarity was present. This, combined with my phylogenetic trees constructed with this duplication, demonstrate that the synthetase domain of CPS evolved from an ancient gene duplication which occurred before the divergence of Bacteria, Eukarya and Archaea. My phylogenetic analysis further shows, with high confidence, that the archaeal *S. solfataricus* sequence is more related to those of the eukaryotes than the bacteria.

This is the first use of a gene involved in a metabolic pathway to root the tree of life, and so represents an important confirmation of these ancient phylogenetic relationships. Previously, the tree of life has been rooted using elongation factors (Iwabe *et al.*, 1989), ATPases (Gogarten *et al.*, 1989) and aminoacyl-tRNA synthetase genes (Brown and Doolittle, 1995), however there has been concern that the full family of genes has not yet been clearly determined for the ATPases, and that horizontal gene transfer and comparisons between paralogs cannot be discounted (Forterre *et al.*, 1993; Hilario and Gogarten, 1993). Also the statistical reliability of the trees produced using the elongation factors has been criticized (Creti *et al.*, 1994). My phylogenetic trees, however, are rooted by a duplication which occurred not to form separate

genes, but to form one gene. Therefore, since horizontal transfer of part of a gene (*i.e.* encoding the first third of the synthetase domain) to fuse with another part of a gene (*i.e.* encoding the second third of the synthetase domain) is far less likely, this means that I know I am at least comparing orthologous sequences. Also, this use of an internal duplication may mean mutation rates are very similar between the two sets of sequences. This is reflected in the level of congruence between the two halves of the trees constructed from the synthetase domain of CFS.

This phylogenetic analysis also confirms the recent report by van den Hoff *et al.* (1995) that separate gene duplications led to the formation of the two CPS enzymes found in Gram-positive bacteria and in eukaryotes. The branching order shows that the duplication event which formed two CPS enzymes in the Gram-positives occurred after their divergence from the Proteobacteria. Only two Gram-positive bacteria have been examined so far, both of the genus *Bacillus*, and so it will be interesting to see if all Gram-positive bacteria have two enzymes, or if this feature is limited to certain genera. There have been preliminary reports that *Lactobacillus* species, which are Gram-positive, have two CPS enzymes (Bringel, 1994), however CPS genes from these organisms have not yet been cloned. Within the eukaryotes themselves, separate gene duplications may also have formed the two CPS enzymes found in this group, since the arginine-specific CPS's of *S. cerevisiae*, *N. crassa*, and *T. cutaneum* do not cluster together with the urea-cycle specific CPS's of human, rat, bullfrog, and *S. acanthias*. However, these enzymes do perform fairly different functions, one involved in biosynthesis of an amino acid and the other involved in ammonia detoxification, and so it cannot be discounted that the separate clusters for these sequences are a consequence of functional divergence. In general, more CPS gene sequences need to be acquired before a clear picture can be obtained regarding the number of

duplication events which formed two CPS enzymes in eukaryotes. However, with an enzyme like CPS being utilized for two metabolic pathways, one can see how advantageous it would be for an organism to have two CPS enzymes, and therefore for a gene duplication event such as this to occur a few times in evolution.

van den Hoff *et al.* (1995) have proposed that the duplication of CPS genes in the eukaryotes occurred between the branching off of plants and fungi, since yeasts contain two CPS enzymes and there has been a report suggesting the pea plant *Pisum sativum* contains only one CPS (Doremus, 1986). However, to date no conclusive genetic studies have yet been performed on any plant to confirm the number of CPS enzymes in these organisms. Further study of plant CPS is needed. Further study of the CPS enzymes of other eukaryotes is in fact also needed, since the organisms studied do not currently reflect the true diversity present within the eukaryotes. Of the "microbial" eukaryotes, only *S. cerevisiae* and *N. crassa* have been confirmed to have two CPS enzymes, and only *P. falciparum* and *B. bovis* have been confirmed to have one CPS (Chansiri and Bagnara, 1995; Davis, 1986; Flores *et al.*, 1994). Based on the present evidence it is concluded that a gene duplication forming the two CPS's seen in eukaryotes probably occurred after the divergence of the apicomplexan protozoans and that possibly more than one duplication event was involved.

Other theories regarding CPS evolution are supported by my phylogenetic analysis. For example, recently it was determined through comparisons of percent sequence similarity for CPS genes that the glutamine-dependent CPSIII of the spiny dogfish shark, *S. acanthias*, is more related to the ammonia-dependent CPSI of ureotelic vertebrates than to the glutamine-dependent CPSII used in pyrimidine biosynthesis in eukaryotes (Hong *et al.*, 1994). This relationship is also

observed in all phylogenetic trees I constructed, where the CPSIII sequence clusters closely with the CPSI sequences. The position of the CPSIII sequence also supports the theory that CPSI evolved from CPSIII (Hong *et al.*, 1994).

Comparison of the branching order determined from CPS sequences with the organisation of the CPS genes in the different organisms reveals that multiple deletions and insertions were involved in the evolution of the CPS genes and in particular the sequence between the regions encoding the amidotransferase and synthetase domains. Based on the branching order, the sequence between the *P. aeruginosa* CPS genes likely evolved as an insertion event. It still cannot be determined, however, whether the original CPS genes of bacteria were in an operonic structure or were separately transcribed genes. Most bacterial CPS genes are co-transcribed, and the archaeal genes are co-transcribed; however, the *N. gonorrhoeae* sequence is the first branch off the tree for the Proteobacteria examined and I have found that these gonococcal genes are separately transcribed. CPS gene sequences need to be obtained from more organisms before a clear pattern can emerge. However, all organisms branching lowest on the phylogenetic trees have a heterodimeric CPS rather than monomeric CPS, and CPS genes clustered together are always found in the order *carA-carB*, and so it is likely that the ancestral CPS was a heterodimer which was encoded by genes closely linked in this order. Eukaryotes are the only organisms found to have monomeric CPS and that is more consistent with the general observation that the genes became fused in eukaryotes during the transition from prokaryotic to eukaryotic translational machinery as the old prokaryotic signals were no longer recognized. The branching order and gene organisation within the eukaryotes also indicates that the arginine-specific CPS genes of yeast were probably initially fused and then re-divided and translocated to form the unlinked amidotransferase and synthetase genes now observed. A similar unlinked gene organization for the

arginine-specific CPS genes of *N. crassa* and *T. cutaneum* has also been observed, though only the amidotransferase gene has been sequenced from *N. crassa* (Orbach *et al.*, 1990) and the synthetase gene sequenced from *T. cutaneum* (Reiser *et al.*, 1994). Previously, van den Hoff *et al.* (1995) concluded that separate gene fusion events led to the monomeric eukaryotic enzymes CPSI and CPSII. However, they did not have access to the data showing that apicomplexan protozoan CPS is monomeric. In light of these recent findings, the more likely hypothesis is that a single gene fusion event occurred in eukaryotes before the divergence of the apicomplexan protozoans, forming a monomeric CPS, and then a subsequent re-division of these genes occurred in the fungal lineage. The apicomplexan protozoan CPS's do have the unusual feature of containing large translated insertions between functional domains of the CPS (Flores *et al.*, 1994; Chansiri and Bagnara, 1995). However this is likely just a reflection of these particular organisms, which are noted for having large polypeptide insertions between the functional domains of other enzymes (Flores *et al.*, 1994). The pyrimidine specific CPS enzymes of eukaryotes have fused with other enzymes of the pyrimidine biosynthetic pathway, and based on the branching order observed in most trees constructed, this is likely due to a gene fusion event which occurred after the duplication of the CPS genes in eukaryotes.

I therefore propose the following summary for the description of the evolution of the CPS genes, based on this phylogenetic analysis and the studies of others (such as Nyunoya and Lusty, 1983; Nyunoya *et al.*, 1985; Hong *et al.*, 1994). First an amidotransferase gene became associated with a kinase gene which duplicated to form the synthetase domain, or the kinase gene duplicated and then associated with the amidotransferase gene, in either case resulting in genes encoding the first primeval heterodimeric CPS. Then, within the Bacteria, the Proteobacteria and Gram-positive

bacteria diverged and a gene duplication event led to the formation of two CPS enzymes in the Gram-positives. Significant insertions and deletions later occurred in the sequence between the genes encoding the heterodimeric CPS of the Proteobacteria. Meanwhile, significant evolution of CPS was occurring in the eukaryotes after the Archaea diverged: First there was a fusion of the amidotransferase and synthetase genes to form a monomeric CPS in the eukaryotes. Then, after the divergence of the apicomplexan protozoans, the first gene duplication event occurred within the eukaryotes to form the two CPS enzymes present in these organisms. Whether one or two gene duplications led to the formation of the arginine-specific and urea-cycle-specific CPS's is as yet unclear, however, each subsequently underwent significant change: The arginine-specific CPS genes of *S. cerevisiae*, *N. crassa*, and *T. cutaneum* re-divided to encode a heterodimeric CPS enzyme once again, and the arginine-specific CPS of human, rat, bullfrog, and *S. acanthias* obtained acetylglutamate binding ability to become similar to CPSIII. Then in human, rat and bullfrog, glutamine-binding ability was lost and this CPS became the CPSI used to harvest ammonia for the urea cycle. During this evolution of the eukaryotes, the pyrimidine specific CPS was relatively unchanged, though it did undergo a fusion with other enzymes involved in the pyrimidine biosynthetic pathway to become part of a multifunctional protein. This fusion occurred early in the evolution of the eukaryotes, since the multifunctional enzyme in *D. discoideum* is the same as that found in *D. melanogaster* and hamster. The same structure is also found in *S. cerevisiae*, although the central DHO enzymatic domain has become inactive.

In conclusion, the evolution of CPS has involved gene duplications, gene fusions, re-division of previously fused genes, gene translocations, deletions and insertions in sequence surrounding the genes, and mutations within the genes resulting in changes in function. Further

investigations of this enzyme and its gene sequence(s) in other organisms should continue to prove interesting. In particular, it would be beneficial to examine the placement of other Archaea in trees rooted by the internal gene duplication, to help resolve the debate over the monophyletic or polyphyletic origin of the Archaea. Also CPS needs to be examined in more eukaryotes to help pinpoint when the gene duplication(s) occurred leading to the formation of arginine- and pyrimidine-specific CPS's in eukaryotes, and to help determine when fusion of the amidotransferase and synthetase domains of CPS occurred.

In general, CPS genes seem amenable to phylogenetic analysis: they show a branching order consistent with other phylogenetic analyses for genes which are related in function, they reveal gene duplications which have occurred through clustering of these genes within the trees, and they have formed from an initial ancient gene duplication which can be used to root the tree of life.

4.7 CONCLUSIONS AND IMPACTS

This work shows that the organization of *carA* and *carB* in *Neisseria gonorrhoeae* is significantly different from that of other previously characterized prokaryotes. A unique, large, variably sized intervening sequence separates the two genes, and the genes are separately transcribed. A similar gene organization seems to be present in other *Neisseria* species. This provides notable evidence that the CPS gene organization in prokaryotes is more variable than previously assumed. This identification of a variable sequence further demonstrates the genetic heterogeneity of the gonococcus. There are many indications that studies of the regulation of gonococcal CPS genes could prove interesting, for example because there is a variable sequence in the region where a promoter for *carB* may reside.

The variable sequence between the CPS genes has been found to be useful for the typing of gonococcal isolates, and possibly other *Neisseria* species. Relationships between isolates have been confirmed, or new relationships have been suggested. This variable sequence could have particular use in the typing of fairly closely related isolates. Novel repetitive sequences have also been identified in this region (RS6 and RS7) which seem to be species-specific for *N. gonorrhoeae* and so could possibly be used for diagnostic purposes.

The gonococcal CPS genes seem to reside within a hotspot for recombination which may explain the high rate of naturally occurring CPS deficiencies which are observed in the gonococcus. Initial characterization of natural CPS mutants indicates that more than one mutation may have caused this deficiency. This property could enable certain auxotypes to be further differentiated. Transformation of one CPS mutant with functional CPS genes has also demonstrated a link between gonococcal isolates of auxotype CUH and isolates of auxotype OUH.

This work also reports the first characterisation of a complete CPS gene sequence from an archaeon, showing that the size and sequence of these genes is quite conserved with those of the bacteria and eukaryotes, in particular the heterodimeric CPSII genes. This provides significant evidence that CPS structure and function in the progenote was most like heterodimeric CPSII.

Phylogenetic analysis of all known CPS genes was performed and revealed further complexity in the series of duplications, deletions and insertions which have played a role in the evolution of these genes. I was also able to root the tree of life and support the theory that the archaeon *S. solfataricus* is more related to Eukarya than Bacteria. This was the first support of this theory using a “metabolic” gene and also the first use of an internal duplication within a gene to root the tree of life. It will be of interest to determine the CPS gene sequences from more Archaea to further investigate hypotheses regarding the tree of life, and also from other organisms to further unravel the complex story behind the evolution of CPS genes.

5. REFERENCES

- Abdelal, A.T., W.F. Bibb, and O. Nainan.** 1982. Carbamate kinase from *Pseudomonas aeruginosa*: purification, characterization, physiological role, and regulation. *J. Bacteriol.* **151**:1411-1419.
- Abdelal, A.T., L. Bussey, and L. Vickers.** 1983. Carbamoylphosphate synthetase from *Pseudomonas aeruginosa*: subunit composition, kinetic analysis, and regulation. *Eur. J. Biochem.* **129**:697-702.
- Adachi, J., and M. Hasegawa.** 1993. Molphy (maximum likelihood inference of protein phylogeny) version 2.1.2. Graduate University for Advanced Study, and Institute for Statistical Mathematics, Tokyo, Japan.
- Ahmed, Z.U., M.R. Sarker, and D.A. Sack.** 1988. Nutritional requirements of *Shigellae* for growth in a minimal medium. *Infect. Immun.* **56**:1007-1009.
- Amuro, N., J.L. Paluh, and H. Zalkin.** 1985. Replacement by site-directed mutagenesis indicates a role for histidine 170 in the glutamine amide transfer function of anthranilate synthase. *J. Biol. Chem.* **260**:14844-14849.
- Anderson, P.M.** 1980. Glutamine- and N-acetylglutamate-dependent carbamoyl phosphate synthetase in elasmobranchs. *Science* **208**:291-293.
- Anderson, P.M. and A. Meister.** 1966. Bi-carbonate dependent cleavage of adenosine triphosphate and other reactions catalyzed by *Escherichia coli* carbamoylphosphate synthetase. *Biochemistry* **5**:3157-3163.

- Ashton, F.E., J.A. Ryan, A. Borczyk, D.A. Caugant, L. Mancino, and D. Huang. 1991. Emergence of a virulent clone of *Neisseria meningitidis* serotype 2a that is associated with meningococcal group C disease in Canada. *J. Clin. Microbiol.* **29**:2489-93.
- Bachmann, B. 1983. Linkage map of *Escherichia coli* K-12, edition 7. *Microbiol. Rev.* **47**:180-230.
- Barber, M., S. Cheng-Chang, D. Friedman, and M. Koomey. 1994. Gonococcal uptake sequences act as transcription terminators in *E. coli*, abstr. D-219, p. 135. Abstr. 94th Annu. Meet. Amer. Soc. Microbiol. 1994.
- Béchet, J., M. Greenson, and J.M. Wiame. 1970. Mutations affecting the repressibility of arginine biosynthetic enzymes in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **12**:31-39.
- Belland, R.J, S.G. Morrison, P. van der Ley, and J. Swanson. 1989. Expression and phase variation of gonococcal P.II genes in *Escherichia coli* involves ribosomal frameshifting and slipped-strand mispairing. *Mol. Microbiol.* **3**:777-786.
- Bergh, S.T. and D.R. Evans. 1993. Subunit structure of class A aspartate transcarbamylase from *Pseudomonas fluorescens*. *Proc. Natl. Acad. Sci. USA* **90**:9819-9822.
- Biro, F.M., S.L. Rosenthal, and M. Kiniyalocts. 1995. Gonococcal and chlamydial genitourinary infections in symptomatic and asymptomatic adolescent women. *Clin. Pediatrics* **34**:419-423.
- Berry, A., R.A. Jensen, and A.T. Hendry. 1987. Enzymatic arrangement and allosteric regulation of the aromatic amino acid pathway in *Neisseria gonorrhoeae*. *Arch. Microbiol.* **149**:87-94.
- Bihlmaier, A., U. Römling, T.F. Meyer, B. Tümmier, and C.P. Gibbs. 1991. Physical and genetic map of the *Neisseria gonorrhoeae* strain MS11-N198 chromosome. *Mol. Microbiol.* **5**:2529-39.

- Birnboim, H.C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7:1513-1523.
- Black, C.G., J.A. Fyfe and J.K. Davies.** 1995. A promoter associated with the neisserial repeat can be used to transcribe the *uvrB* gene from *Neisseria gonorrhoeae*. *J. Bacteriol.* 177:1952-1958.
- Bonekamp, F.K., O. Clemmesen, O. Karlstrom, and K.F. Jensen.** 1984. Mechanism of UTP-modulated attenuation at the *pyrE* gene of *Escherichia coli*: an example of operon polarity control through the coupling of translation to transcription. *EMBO J.* 3:2857-2861.
- Bringel, F.** 1994. Isolation of Lactobacilli naturally deficient in carbamyl phosphate biosynthesis. XIVth International Arginine Symposium. Sept 12-14, 1994. Paris, France.
- Brown, J.R., and W.F. Doolittle.** 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc. Natl. Acad. Sci. USA* 92:2441-2445.
- Brunham, R.C., F. Plummer, L. Stanley, F. Rand, and W. DeWitt.** 1985. Correlation of auxotype and protein type 1 with expression of disease due to *Neisseria gonorrhoeae*. *J. Infect. Dis.* 152:339-343.
- Brustilow, S.W., and A.L. Horwich.** 1989. Urea cycle enzymes. p. 629-663. *In* C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle (ed.), *The metabolic basis of inherited disease*, 6th ed.. McGraw-Hill, New York.
- Campbell, J.W.** 1965. Arginine and urea biosynthesis in the land planarium: its significance in biochemical evolution. *Nature* 208:1299-1301.
- Carifo, K., and B.W. Catlin.** 1973. *Neisseria gonorrhoeae* auxotyping: differentiation of clinical isolates based on growth responses on chemically defined media. *Appl. Microbiol.* 26:223-30.

- Catlin, B.W. 1973. Nutritional profiles of *Neisseria gonorrhoeae*, *Neisseria meningitidis* and *Neisseria lactamica* in chemically defined media and the use of growth requirements for gonococcal typing. J. Infect. Dis. 128:178-194.
- Catlin, B.W., and E.H. Nash. 1978. Arginine biosynthesis in gonococci isolated from patients, p.1-7. In G.F. Brooks, E.C. Gotschlich, K.K. Holmes, W.D. Sawyer, and F.E. Young (ed.), Immunobiology of *Neisseria gonorrhoeae*. American Society for Microbiology, Washington, D.C.
- Catlin, B.W., and A. Reyn. 1982. *Neisseria gonorrhoeae* isolated from disseminated and localised infections in pre-penicillin era. Brit. J. Vener. Dis. 58:158-165.
- Centers for Disease Control and Prevention. 1994. Decreased susceptibility of *Neisseria gonorrhoeae* to fluoroquinolones - Ohio and Hawaii, 1992-1994. Morb. Mort. Week. Rep. 43:325-327.
- Chansiri, K., and A.S. Bagnara. 1995. The structural gene for carbamoyl phosphate synthetase from the protozoan parasite *Babesia bovis*. Mol. Biochem. Parasitol. 74:239-243.
- Charlier, D., G. Hassanzadeh, A. Khoti, D. Gigot, A. Pierard, and N. Glansdorff. 1995. carP, involved in the pyrimidine regulation of the *Escherichia coli* carbamoylphosphate synthetase operon encodes a sequence-specific DNA-binding protein identical to XerB and PepA, also required for resolution of ColEI multimers. J. Mol. Biol. 250:392-406.
- Charlier, D., G. Weyens, M. Roovers, J. Piette, A. Bocquet, A. Piérard, and N. Glansdorff. 1988. Molecular interactions in the control region of the *carAB* operon encoding *Escherichia coli* carbamoyl-phosphate synthetase. J. Mol. Biol. 204:876-77.

- Chédin, F., E. Dervyn, R. Dervyn, S.D. Ehrlich, and P. Noirot.** 1994. Frequency of deletion formation decreases exponentially with distance between short direct repeats. *Mol. Microbiol.* **12**:561-569.
- Chen, C.Y., C.A. Genco, J.P. Rock, and S.A. Morse.** 1989. Physiology and metabolism of *Neisseria gonorrhoeae* and *Neisseria meningitidis*: implications for pathogenesis. *Clin. Microbiol. Rev.* **2**:S35-40.
- Chun, P.K., G.F. Sensabaugh, and N.A. Vedros.** 1985. Genetic relationships among *Neisseria* species assessed by comparative enzyme electrophoresis. *J. Gen. Microbiol.* **131**:3105-3115.
- Copley, C.G.** 1987. *Neisseria gonorrhoeae*: subdivision of auxogroups by genetic transformation. *Genitourin. Med.* **63**:153-156.
- Correia, F.F., S. Inouye, and M. Inouye.** 1988. A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae*. *J. Bio. Chem.* **263**:12194-12198.
- Cunin, R., N. Glansdorff, A. Piérard, and V. Stalon.** 1986. Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* **50**:314-352.
- Creti, R., E. Ceccaralli, M. Bocchetta, A.M. Sanangelantoni, O. Tiboni, P. Palm, and P. Cammarano.** 1994. Evolution of translational elongation factor (EF) sequences: reliability of global phylogenies inferred from EF-1 alpha (Tu) and EF-2(G) proteins. *Proc. Natl. Acad. Sci. USA* **91**:3255-3259.
- Cybis, J., and R.H. Davis.** 1975. Organisation and control in the arginine biosynthetic pathway of *Neurospora*. *J. Bacteriol.* **123**:196-202.

- Davidson, J.N., K.C. Chen, R.S. Jamison, L.A. Musmanno, and C.B. Kern.** 1993. The evolutionary history of the first three enzymes in pyrimidine biosynthesis. *Bioessays* 15:157-164.
- Davis, R.H.** 1986. Compartmental and regulatory mechanisms in the arginine pathways of *Neurospora crassa* and *Saccharomyces cerevisiae*. *Microbio. Rev.* 50:280-313.
- Davis, R.H., J.L. Ristow, and B.A. Hanson.** 1980. Carbamoylphosphate synthetase A of *Neurospora crassa*. *J. Bacteriol.* 141:144-155.
- Dempsey, J.F., and J.G. Cannon.** 1994. Locations of genetic markers on the physical map of the chromosome of *Neisseria gonorrhoeae* FA1090. *J. Bacteriol.* 176:2055-2060.
- Dempsey, J.F., A.B. Wallace, and J.G. Cannon.** 1995. The physical map of the chromosome of a serogroup A strain of *Neisseria meningitidis* shows complex rearrangements relative to the chromosomes of the two mapped strains of closely related species of *N. gonorrhoeae*. *J. Bacteriol.* 177:6390-6400.
- De Rijcke, M., S. Seneca, B. Punyammalee, N. Glansdorff, and M. Crabeel.** 1992. Characterisation of the DNA target site for the yeast ARGR regulatory complex, a sequence able to mediate repression or induction by arginine. *Mol. Cell. Biol.* 12:68-81.
- DeVoe, I.W.** 1982. The meningococcus and mechanisms of pathogenicity. *Microbiol. Rev.* 46:162-90.
- Dillon, J.R.** 1983. Laboratory methods for *Neisseria gonorrhoeae*. Publication H47-58/1983E, Health and Welfare Canada, Ottawa, Ontario, Canada.
- Dillon, J.R., G.S. Bezanson, and K.-H. Yeung.** 1985. Basic Techniques. pp. 1-26. *In* J.R. Dillon, A. Nasim, and E.R. Nestmann (ed.), *Recombinant DNA methodology*. John Wiley and Sons, New York.

- Dillon, J.R., and M. Carballo.** 1990. Molecular epidemiology and novel auxotype/serovar/plasmid content combinations in tetracycline-resistant *Neisseria gonorrhoeae* (TRNG) isolated in Canada. *Can. J. Microbiol.* **36**:64-67.
- Dillon, J.R., S. Hickey, K.-H. Yeung, C. Drixler, M. Pauzé, M. Carballo, N. Bigelow, and the National Co-operative Study Group.** 1990. Final Report, National Study to determine clinical and biological factors influencing *N. gonorrhoeae* infections and treatment in children and adults, 1 March, 1988 - 31 March, 1989. Health Canada.
- Dillon, J.R., and M. Pauzé.** 1984. Resistance to antimicrobial agents: What next for *N. gonorrhoeae*. *Sexual. Trans. Dis.* **11**:353-359.
- Dillon, J.R., and K.-H. Yeung.** 1989. Evolution of antibiotic resistance in pathogenic *Neisseria*. *Clin. Microbiol. Rev.* **2**:S125-S133.
- Doremus, H. D.** 1986. Organization of the pathway of *de novo* pyrimidine nucleotide biosynthesis in Pea (*Pisum sativum* L. cv progress no. 9) leaves. *Arch. Biochem. Biophys.* **250**:112-119.
- Dougherty, T.J., A. Asmus, and A. Tomasz.** 1979. Specificity of DNA uptake in genetic transformation of gonococci. *Biochem. Biophys. Res. Commun.* **86**:97-104.
- Easmon, C.S.F.** 1990. The changing pattern of antibiotic resistance of *Neisseria gonorrhoeae*. *Genitourin. Med.* **66**:55-56.
- Elkins, C., C.E. Thomas, H.S. Seifert, P.F. Sparling.** 1991. Species specific uptake of DNA by gonococci is mediated by a 10 bp sequence. *J. Bacteriol.* **173**:3911-3913.
- Emmett, W., and E. Kloos.** 1979. The nature of arginine auxotrophy in cutaneous populations of staphylococci. *Gen. Microbiol.* **110**:305-314

- Felsenstein, J.** 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Dept. of Genetics, University of Washington, Seattle, USA.
- Flores, M.V.C., W.J. O'Sullivan, and T.S. Stewart.** 1994. Characterization of the carbamoyl phosphate synthetase gene from *Plasmodium falciparum*. Mol. Biochem. Parasitol. **68**:315-318.
- Forterre, P., N. Benachenhou-Lahfa, F. Confalonieri, M. Duguet, C. Elie, and B. Labedan.** 1993. The nature of the last universal ancestor and the root of the tree of life, still open questions. Biosystems **28**:15-32.
- Frasch, C.E., W.D. Zollinger, and J.T. Poolman.** 1985. Serotype antigens of *Neisseria meningitidis* and a proposed scheme for designation of serotypes. Rev. Infect. Dis. **7**:504-510.
- Fyfe, J.A., C.S. Carrick, and J.K. Davies.** 1995. The *pile* gene of *Neisseria gonorrhoeae* MS11 is transcribed from a sigma-70 promoter during growth In Vitro. J. Bacteriol. **177**:3781-3787.
- Gero, A.M., W.J. O'Sullivan, I.G. Wright, and D.F. Mahoney.** 1983. The enzymes of pyrimidine biosynthesis in *Babesia bovis* and *Babesia bigemina*. Austral. J. Exper. Biol. Med. Sci. **61**:239-243.
- Geschwill, K., and L. Lumper.** 1989. Identification of cysteine residues in carbamoyl-phosphate synthase I with reactivity enhanced by N-acetyl-L-glutamate. Biochem. J. **260**:573-576.
- Ghim, S.-Y., P. Nielsen, and J. Neuhard.** 1994. Molecular characterization of pyrimidine biosynthesis genes from the thermophile *Bacillus caldolyticus*. Microbiology **140**:479-491.

- Glickman, B.W., and L.S. Ripley. 1984. Structural intermediates of deletion mutagenesis: a role for palindromic DNA. *Proc. Natl. Acad. Sci. USA* **81**:512-516.
- Gogarten, J.P., H. Kilbak, P. Dittrich, L. Taiz, E.J. Bowman, B.J. Bowman, M.F. Manolson, R.J. Poole, T. Date, and T. Oshima. 1989. Evolution of vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA* **86**:6661-6665.
- Goodman, S.D., and J.J. Scocca. 1988. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc. Natl. Acad. Sci. USA* **85**:6982-6986.
- Goodman, S.D., and J.J. Scocca. 1991. Factors influencing the specific interaction of *Neisseria gonorrhoeae* with transforming DNA. *J. Bacteriol.* **173**:5921-5923.
- Gotschlich, E.C. 1994. Genetic locus for the biosynthesis of the variable portion of *Neisseria gonorrhoeae* lipooligosaccharide. *J. Exper. Med.* **180**:2181-2190.
- Guillou, F., S.D. Rubino, R.S. Markovitz, D.M. Kinney, and C.J. Lusty. 1989. *Escherichia coli* carbamoyl-phosphate synthetase: domains of glutaminase and synthetase subunit interaction. *Proc. Natl. Acad. Sci. USA* **86**:8304-8309.
- Gyonvarch, A., M. Nguyen-Juilleret, J.C. Hubert, and F. Lacroute. 1988. Structure of the *Saccharomyces cerevisiae* URA4 gene encoding dihydroorotase. *Mol. Gen. Genet.* **212**:134-141.
- Haas, D., B.W. Holloway, A. Schambock, and T. Leisinger. 1977. The genetic organisation of arginine biosynthesis in *Pseudomonas aeruginosa*. *Mol. Gen. Genet.* **154**:7-22.

- Haas, R., and T.F. Meyer.** 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. *Cell* **44**:107-115.
- Hagblom, P., E. Segal, E. Billyard, and M. So.** 1985. Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. *Nature* **315**:156.
- Halter, R., J. Pohler and T.F. Meyer.** 1989. Mosaic-like organisation of IgA protease genes in *Neisseria gonorrhoeae* generated by horizontal genetic exchange *in vivo*. *EMBO J.* **8**:2737-2744.
- Haraguchi, Y., T. Uchino, M. Takiguchi, F. Endo, M. Mori, and I. Matsuda.** 1991. Cloning and sequence of a cDNA encoding human carbamyl phosphate synthetase I: molecular analysis of hyperammonemia. *Gene* **107**:335-340.
- Health Canada.** 1991. Interim guidelines for the treatment of uncomplicated gonococcal infection. *Can. Dis. Week. Rep.* **17**:281-282.
- Health Canada.** 1991. Notifiable disease annual summary 1990. Government of Canada.
- Health Canada.** 1993. Notifiable disease annual summary 1992. Government of Canada.
- Hebeler, B.H., and S.A. Morse.** 1976. Physiology and metabolism of pathogenic *Neisseria*: tricarboxylic acid cycle activity in *Neisseria gonorrhoeae*. *J. Bacteriol.* **128**:192-201.
- Helbing, C., and B.G. Atkinson.** 1994. Heat shock stabilizes T3-induced urea cycle enzyme gene expression in the liver of *Rana catesbeiana* tadpoles. *J. Biol. Chem.* **269**:11743-11750.
- Hendry, A.T., and L.O. Stewart.** 1979. Auxanographic grouping and typing of *Neisseria gonorrhoeae*. *Can. J. of Microbiol.* **25**:512-552.

- Hilario, E., and J.P. Gogarten.** 1993. Horizontal transfer of ATPase genes - the tree of life becomes a net of life. *Biosystems* **31**:111-119.
- Hoehn, G.T., and V.L. Clark.** 1992. Isolation and nucleotide sequence of the gene *aniA* encoding the major anaerobically induced outer membrane protein of *Neisseria gonorrhoeae*. *Infect. Immun.* **60**:4695-4703.
- van den Hoff, M.J.B., A. Jonker, J.J. Beintema, and W.H. Lamers.** 1995. Evolutionary aspects of the carbamoylphosphate synthetase genes. *J. Mol. Evol.* **41**:813-832.
- Hoke, C., and N.A. Vedros.** 1982. Taxonomy of the Neisseriae: Deoxyribonucleic acid base composition, interspecific transformation, and deoxyribonucleic acid hybridization. *Int. J. Sys. Bacteriol.* **32**:57-66.
- Hong, J., W.L. Salo, C.J. Lusty, and P.M. Anderson.** 1994. Carbamyl phosphate synthetase III, an evolutionary intermediate in the transition between glutamine-dependent and ammonia-dependent carbamyl phosphate synthetases. *J. Mol. Biol.* **243**:131-140.
- Hook, E.W., and K.K. Holmes.** 1985. Gonococcal infections. *Ann. Intern. Med.* **102**:229-243.
- Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata.** 1989. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* **86**:9355-9359.
- Janik, A., E. Juni, and G.A. Heym.** 1976. Genetic transformation as a tool for detection of *Neisseria gonorrhoeae*. *J. Clin. Microbiol.* **4**:71-81.
- Jones, M.E.** 1980. Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. *Annu. Rev. Biochem.* **49**:253-279.

- Jonsson, A.-B., G. Nyberg, and S. Normark.** 1991. Phase variation of gonococcal pili by frame shift mutation in *pilC*, a novel gene for pilus assembly. *EMBO J.* **10**:35-43.
- Jonsson, A.-B., M. Rahman, and S. Normark.** 1995. Pilus biogenesis gene, *pilC*, of *Neisseria gonorrhoeae*: *pilC1* and *pilC2* are each part of a larger duplication of the gonococcal genome and share upstream and downstream sequences with *opa* and *pil* loci. *Microbiology.* **141**:2367-2377.
- Jyssum, S.** 1983. Pyrimidine biosynthesis in *Neisseria meningitidis*. *Acta. Path. Microbiol. Immunol. Scand. Sect. B.* **91**:257-260.
- Kane, J.F., W.M. Holmes, and R.A. Jensen.** 1972. Metabolic interlock - the dual function of a folate pathway gene as an extra-operonic gene of tryptophan biosynthesis. *J. Biol. Chem.* **247**:1587-1596.
- Keeling, P.J., R.L. Charlebois and W.F. Doolittle.** 1995. Archaeobacterial genomes: eubacterial form and eukaryotic content. *Curr. Opinion Gen. Dev.* **4**:816-822.
- Keeling, P.J., and W.F. Doolittle.** 1995. Archaea: Narrowing the gap between prokaryotes and eukaryotes. *Proc. Natl. Acad. Sci.* **92**:5761-5764.
- Kellogg, D.S., Jr., W.L. Peacock, Jr., W.E. Deacon, L. Brown, and C.L. Pirkle.** 1963. *Neisseria gonorrhoeae*. I. Virulence genetically linked to clonal variation. *J. Bacteriol.* **85**:1274-1279.
- Kenyon, J.E.** 1978. *Neisseria gonorrhoeae*: Acquisition of auxotrophy in the mouse. *Curr. Microbiol.* **1**:253-256.
- Kerle, K.K., J.R. Mascola, and T.A. Miller.** 1992. Disseminated gonococcal infection. *Amer. Fam. Phys.* **45**:209-214.

- Kilstrup, M., C-D. Lu, A.T. Abdelal, and J. Neuhard.** 1988. Nucleotide sequence of the *carA* gene and regulation of the *carAB* operon in *Salmonella typhimurium*. *Eur. J. Biochem.* **176**:421-429.
- Kim, H., L. Lee, and D.R. Evans.** 1991. Identification of the ATP binding sites of the carbamoylphosphate synthetase domain of the syrian hamster multifunctional protein CAD by affinity labelling with 5'-[p-(fluorosulfonyl)benzoyl]adenosine. *Biochemistry* **30**:10322-10329.
- Kollmar, R., K.A. Sukow, S.K. Sponagle, and P.J. Farnham.** 1994. Start site selection at the TATA-less carbamoyl-phosphate synthase (glutamine-hydrolysing)/aspartate transcarbamylase /dihydroorotase promoter. *J. Biol. Chem.* **269**:2252-2257.
- Kroll, J.S., B.M. Loynds, and P.R. Langford.** 1992. Palendromic *Haemophilus* species DNA uptake sequences in presumed transcription terminators from *H. influenzae* and *H. parainfluenzae*. *Gene* **114**:151-152.
- Krungskrai, J., S.R. Krungskrai, K. Phakanont.** 1992. Antimalarial activity of orotate analogs that inhibit dihydroorotase and dihydroorotate dehydrogenase. *Biochem. Pharm.* **43**:1295-1301.
- Knapp, J.S, M.R. Tam, R.C. Nowinski, K.K. Holmes, and E.G. Sandstrom.** 1984. Serological classification of *Neisseria gonorrhoeae* with the use of monoclonal antibodies to gonococcal outer membrane protein I. *J. Infect. Dis.* **150**: 44-48.
- Kwon, D.-H., C.-D. Lu, D.A. Walthall, T.M. Brown, J.E. Houghton, and A.T. Abdelal.** 1994. Structure and regulation of the *carAB* operon in *Pseudomonas aeruginosa* and *Pseudomonas stutzeri*: no untranslated region exists. *J. Bacteriol.* **176**:2532-2542.
- Lacroute, F., A. Pierard, M. Grenson, and J.M. Wiame.** 1965. The biosynthesis of carbamoylphosphate in *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **40**:127-142.

- Langer, D., J. Hain, P. Thuriaux, and W. Zillig.** 1995. Transcription in Archaea: similarity to that in Eucarya. *Proc. Nat. Acad. Sci. USA* 92:5768-5772.
- Lawson, F.S., F.M. Billowes, and J.R. Dillon.** 1995. Organization of carbamoyl-phosphate synthase genes in *Neisseria gonorrhoeae* includes a large, variable intergenic sequence which is also present in other *Neisseria* species. *Microbiology* 141:1183-1191.
- Legrain, C., M. Demarez, N. Glansdorff, and A. Pierard.** 1995. Ammonia-dependent synthesis and metabolic channeling of carbamoyl phosphate in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Microbiology*. 141:1093-1099.
- Li, H.** 1993. M.Sc. thesis. University of Ottawa, Ottawa, Canada.
- Li, H., and J.R. Dillon.** 1995. Utility of ribotyping, restriction endonuclease analysis and pulsed-field gel electrophoresis to discriminate between isolates of *Neisseria gonorrhoeae* of serovar 1A-2 which require arginine, hypoxanthine or uracil for growth. *J. Med. Microbiol.* 43:208-215.
- Lim, D., J.D. Oppenheim, T. Eckhardt, and W.K. Maas.** 1987. Nucleotide sequence of the *argR* gene of *Escherichia coli* K12 and isolation of its product, the arginine repressor. *Proc. Natl. Acad. Sci. USA*. 84:6697-701.
- Liu, C., and C.L. Turnbough, Jr.** 1989. Multiple control mechanisms for pyrimidine-mediated regulation of *pyrBI* operon expression in *Escherichia coli* K-12. *J. Bacteriol.* 171:3337-3342.
- Lu, C., J.E. Houghton, and A.T. Abdelal.** 1992. Characterisation of the arginine repressor from *Salmonella typhimurium* and its interactions with the *carAB* operator. *J Mol. Biol.* 225:11-24.
- Lu, C., M. Kilstrup, J. Neuhaard, and A. Abdelal.** 1989. Pyrimidine regulation of tandem promoters for *carAB* in *Salmonella typhimurium*. *J. Bacteriol.* 171:5436-5442.

- Lusty, C.J.** 1978. Carbamoylphosphate synthetase I of rat-liver mitochondria. Purification, properties, and polypeptide molecular weight. *Eur. J. Biochem.* **85**:373-383.
- Lusty, C.J., E.E. Widgren, K.E. Broglie, and H. Nyunoya.** 1983. Yeast carbamyl phosphate synthetase. Structure of the yeast gene and homology to *Escherichia coli* carbamyl phosphate synthetase. *J. Biol. Chem.* **258**:14466-14472.
- Maas, W.K.** 1991. The regulation of arginine biosynthesis: Its contribution to understanding the control of gene expression. *Genetica* **128**:489-494.
- Maas, W.K.** 1995. The arginine repressor of *Escherichia coli*. *Microbiol. Rev.* **58**:631-640.
- Maas, W.K., and A.J. Clark.** 1964. Studies on the mechanism of repression of arginine biosynthesis in *Escherichia coli*. II. Dominance of repressibility in diploids. *J. Mol. Biol.* **8**:365-370.
- Marshall, M.** 1976. Carbamoyl-phosphate synthetase I from frog liver. p. 133-142. *In* S. Grisolia, R. Baguena, and F. Mayor (ed.), *The Urea Cycle*. John Wiley and Sons, New York.
- Martin, P.R., and M.H. Mulks.** 1992a. Molecular characterisation of the *argJ* mutation in *Neisseria gonorrhoeae* strains with requirements for arginine, hypoxanthine, and uracil. *Infect. Immun.* **60**:970-975.
- Martin, P.R. and M.H. Mulks.** 1992b. Sequence analysis and complementation studies of the *argJ* gene encoding ornithine acetyltransferase from *Neisseria gonorrhoeae*. *J. Bacteriol.* **174**: 2694-2701.
- Martin, P.R., J.W. Cooperider, and M.H. Mulks.** 1990. Sequence of the *argF* gene encoding ornithine transcarbamylase from *Neisseria gonorrhoeae*. *Gene* **94**:139-140.
- Mayer, L.W., G.K. Schoolnik, and S. Falkow.** 1977. Genetic studies on *Neisseria gonorrhoeae* from disseminated gonococcal infections. *Infect. Immun.* **18**:165-172.

- McGee, Z.A., and D.S. Stevens.** 1984. Common pathways of invasion of mucosal barriers by *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Surv. Synth. Path. Res.* **3**:1-10.
- Meister, A., and S.G. Powers.** 1978. Glutamine-dependent carbamoylphosphate synthetase: catalysis and regulation. *Adv. Enzyme Regul.* **16**:289-315.
- Meyer, T.F., E. Billyyard, R. Haas, S. Storzbach, and M. So.** 1984. Pilus genes of *Neisseria gonorrhoeae*: chromosomal organization and DNA sequence. *Proc. Natl. Acad. Sci. USA* **81**:6110-6114.
- Meyer, T.F.** 1990. Pathogenic *Neisseriae* - a model of bacterial virulence and genetic flexibility. *Zbl. Bakt.* **274**:135-154.
- Miran, S.G., S.H. Chang, and F.M. Raushel.** 1991. Role of four conserved histidine residues in the amidotransferase domain of carbamoyl phosphate synthetase. *Biochemistry* **30**:7901-7907.
- Mommsen, T.P., and P.J. Walsh.** 1989. Evolution of urea synthesis in vertebrates: the piscine connection. *Science* **243**:72-75.
- Moreno, J.G., J.R. Dillon, R. Arroyave, A. Maldonado, F. Fitch, A. Salvo, D. Villalobos, and M. Pauzé.** 1987. Identification of penicillinase-producing *N. gonorrhoeae* (PPNG) in Chile during clinical and microbiological investigations of gonococcal susceptibility to antimicrobial agents. *Genitourin. Med.* **63**:6-12.
- Mori, M., H. Ishida, and M. Tatibana.** 1975. Aggregation states and catalytic properties of the multienzyme complex catalysing the initial steps of pyrimidine biosynthesis in rat liver. *Biochemistry* **14**:2622-2630.

- Morris, C.J., and J.N. Reeve.** 1988. Conservation of structure in the human gene encoding argininosuccinate synthetase and the *argG* genes of the archaeobacteria *Methanosarcina barkeri* MS and *Methanococcus vannielli*. *J. Bacteriol.* **170**:3125-3130.
- Morris, S.M.** 1992. Regulation of the enzymes of urea and arginine biosynthesis. *Annu. Rev. Nutr.* **12**:81-101.
- Morse, S.A., S. Stein, and J. Hines.** 1974. Glucose metabolism in *Neisseria gonorrhoeae*. *J. Bacteriol.* **120**:702-714.
- Mountain, A., N.H. Mann, R.N. Munton, and S. Baumberg.** 1984. Cloning of a *Bacillus subtilis* restriction fragment complementing auxotrophic mutants of eight *Escherichia coli* genes of arginine biosynthesis. *Mol. Gen. Genet.* **197**:82-89.
- Mullins, L.S., C.J. Lusty, and F.M. Raushel.** 1991. Alterations in the energetics of the carbamoylphosphate synthetase reaction by site-directed modification of the essential sulfhydryl group. *J. Biol. Chem.* **266**:8236-8240.
- Murphy, G.L., T.D. Connell, D.S. Barritt, M. Koomey, and J.G. Cannon.** 1989. Phase variation of gonococcal protein PII: Regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. *Cell* **56**:539-547.
- Neuhard, J., and P. Nygaard.** 1987. Purines and pyrimidines. p. 445-473. *In* F.C. Neidhardt, J.L. Ingraham, K.B. Law, S. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.

- Ng, L.-K., and J.R. Dillon. 1993. Typing by serovar, antibiogram, plasmid content, riboprobing, and isoenzyme typing to determine whether *Neisseria gonorrhoeae* isolates requiring proline, citrulline, and uracil for growth are clonal. J. Clin. Microbiol. **31**:1555-1561.
- North, A.K., M.C.M. Smith, and S. Baumberg. 1989. Nucleotide sequence of a *Bacillus subtilis* regulatory gene and homology of its product to the *Escherichia coli* arginine repressor. Gene **80**:29-38.
- Nyunoya, H., and C. J. Lusty. 1983. The *carB* gene of *Escherichia coli*: a duplicated gene coding for the large subunit of carbamoyl-phosphate synthetase. Proc. Natl. Acad. Sci. USA. **80**:4629-4633.
- Nyunoya, H., K.E. Broglie, and C.J. Lusty. 1985. The gene coding for carbamoylphosphate synthetase I was formed by fusion of an ancestral glutaminase gene and a synthetase gene. Proc. Natl. Acad. Sci. USA. **82**:2244-2246.
- O'Neal, T.D., and A.W. Naylor. 1976. Partial purification and properties of carbamoyl phosphate synthetase of Alaska pea (*Pisum sativum* L. cultivar Alaska). Biochem. J. **113**:271-279.
- Orbach, M.J., M.S. Sachs, and C. Yanofsky. 1990. The *Neurospora crassa arg-2* locus. Structure and expression of the gene encoding the small subunit of arginine-specific carbamoyl phosphate synthetase. J. Biol. Chem. **265**:10981-10987.
- O'Rourke, M., and B.G. Spratt. 1994. Further evidence for the non-clonal population structure of *Neisseria gonorrhoeae*: extensive genetic diversity within isolates of the same electrophoretic type. Microbiology **140**:1285-1290.

- O'Rourke, M., and E. Stevens.** 1993. Genetic structure of *Neisseria gonorrhoeae* populations: a non-clonal pathogen. *J. Gen. Microbiol.* **139**: 2603-2611.
- Padian, N.S., and A.E. Washington.** 1994. Pelvic inflammatory disease - a brief overview. *AEP* **4**:129-132.
- Paulus, H.** 1983. The evolutionary history of the ornithine cycle as a determinant of its structure and regulation. *Curr. Topics Cell. Reg.* **22**:177-200.
- Paulus, T.J., and R.L. Switzer.** 1979. Characterisation of pyrimidine-repressible and arginine-repressible carbamyl phosphate synthetases from *Bacillus subtilis*. *J. Bacteriol.* **137**:82-91.
- Pauza, C.D., M.J. Karels, M. Navre, and H.K. Schachman.** 1982. Genes encoding *Escherichia coli* aspartate transcarbamoylase: the *pyrB-pyrI* operon. *Proc. Natl. Acad. Sci. USA* **79**:4020-4024.
- Picard, F.J.** 1990. Ph.D. thesis. University of Ottawa, Ottawa, Canada.
- Picard, F.J., and J.R. Dillon.** 1989a. Biochemical and genetic studies with arginine and proline auxotrophs of *Neisseria gonorrhoeae*. *Can. J. Microbiol.* **35**:1069-1075.
- Picard, F.J., and J.R. Dillon.** 1989b. Cloning and organisation of seven arginine biosynthesis genes from *Neisseria gonorrhoeae*. *J. Bacteriol.* **171**:1644-1651.
- Pierard, A., F. Messenguy, A. Feller, and F. Hilger.** 1979. Dual regulation of the synthesis of the arginine pathway carbamoylphosphate synthetase of *Saccharomyces cerevisiae* by specific and general controls of amino acid biosynthesis. *Mol. Gen. Genet.* **174**:163-171.
- Piette, J., H. Nyunoya, C.J. Lusty, R. Cunin, G. Weyens, M. Crabeel, D. Charlier, N. Glansdorff and A. Pierard.** 1984. DNA sequence of the *carA* gene and the control region of *carAB*: tandem promoters, respectively controlled by arginine and the pyrimidines, regulate the

- synthesis of carbamoyl-phosphate synthetase in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA. **81**:4134-4138.
- Post, L.E., D.J. Post, and F.M. Raushel. 1990. Dissection of the functional domains of *Escherichia coli* carbamoyl phosphate synthetase by site-directed mutagenesis. J. Biol. Chem. **265**:7742-7747.
- Qiu, H., E. Dubois, P. Broën, and F. Messenguy. 1990. Functional analysis of ARGRI and ARGRIII regulatory proteins involved in the regulation of arginine metabolism in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **222**:192-200.
- Quinn, C.L., B.T. Stephenson, and R.L. Switzer. 1991. Functional organisation and nucleotide sequence of the *Bacillus subtilis* pyrimidine biosynthetic operon. J. Biol. Chem. **266**:9113-9127.
- Reiser, J., V. Glumoff, U. Ochsner, and A. Fiechter. 1994. Molecular analysis of the *Trichosporin cutaneum* DSM 70698 *argA* gene and its use for DNA-mediated transformations. J. Bacteriol. **176**:3021-3032.
- Rodriguez-Aparicio, L.B., A.M. Guadalajara, and V. Rubio. 1989. Physical location of the site for N-acetyl-L-glutamate, the allosteric activator of carbamoyl phosphate synthetase, in the 20kDa COOH-terminal domain. Biochemistry **28**:3070-3074.
- Rubino, S.D., H. Nyunoya, and C.J. Lusty. 1986. Catalytic domains of carbamyl phosphate synthetase. Glutamine-hydrolyzing site of *Escherichia coli* carbamyl phosphate. J. Biol. Chem. **261**:11320-11327.
- Rubio, V. 1993. Structure-function studies in carbamoyl phosphate synthetases. Biochem. Soc. Trans. **21**:198-202.

- Rubio, V., J. Cervera, C.J. Lusty, E. Bendala, and H.G. Britton. 1991. Domain structure of the large subunit of *Escherichia coli* carbamoyl phosphate synthetase. Location of the binding site for the allosteric inhibitor UMP in the COOH-terminal domain. *Biochemistry* 30:1068-1075.
- Rudel, T. 1995. *N. gonorrhoeae carA* gene, complete cds. Sequence submission to GenBank, accession number Z54242.
- Rudel, T., D. Farcus, R. Barten, I. Scheuerpflug, E. Nonnenmacher, and T.F. Meyer. 1995. Role of pili and the phase-variable PilC protein in natural competence for transformation of *Neisseria gonorrhoeae*. *Proc. Natl. Acad. Sci. USA* 92:7986-7990.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, Cold Spring Harbor, N.Y.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Sarkari, J., N. Pandit, E.R. Moxon, and M. Achtman. 1994. Variable expression of the Opc outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine. *Mol. Microbiol.* 13:207-217.
- Schofield, J.P. 1993. Molecular studies on an ancient gene encoding for carbamoyl-phosphate synthetase. *Clin. Sci.* 84:119-128.
- Schurr, M.J., J.F. Vickrey, A.P. Kumar, A.L. Campbell, R. Cunin, R.C. Benjamin, M.S. Shanley, and G.A. O'Donovan. 1995. Aspartate transcarbamylase genes of *P. putida*: Requirement for an inactive dihydroorotase for assembly into the dodecameric holoenzyme. *J. Bacteriol.* 177:1751-1759.

- Segal, E., E. Billyard, M. So, S. Storzbach, and T.F. Meyer. 1985. Role of chromosomal arrangement in *Neisseria gonorrhoeae* pilus phase variation. *Cell* 40:293.
- Seifert, H.S., and M. So. 1991. Genetic systems in pathogenic Neisseriae, p.342-57. *In* Methods in Enzymology, Vol.204. Academic Press Inc., New York.
- Sensen, C.W., R.L. Charlebois, R.K. Singh, H.-P. Klenk, M.A. Ragan, and W.F. Doolittle. Sequencing the genome of *Sulfolobus solfataricus* P2. In press. *In* F.J. de Bruijn, J.R. Lupski, and G. Weinstock (ed.), Bacterial Genomes: Physical Structure and Analysis. Chapman & Hall, New York.
- Seymour, K.K., S.D. Lyons, L. Phillips, K.H. Rieckmann, and R.I. Christopherson. 1994. Cytotoxic effects of inhibitors of *de novo* pyrimidine biosynthesis upon *Plasmodium falciparum*. *Biochemistry* 33:5268-5274.
- Sherman, L.W. 1979. Biochemistry of *Plasmodium* (Malarial Parasites). *Microbiol. Rev.* 43:453-495.
- Shinners, E.N., and B.W. Catlin. 1978. Arginine biosynthesis in *Neisseria gonorrhoeae*: enzymes catalyzing the formation of ornithine to citrulline. *J. Bacteriol.* 136:131-135.
- Shinners, E.N., and B.W. Catlin. 1982. Arginine and pyrimidine biosynthetic defects in *Neisseria gonorrhoeae* strains isolated from patients. *J. Bacteriol.* 151:295-302.
- Shoaf, W.T., and M.E. Jones. 1973. Uridylic acid synthesis in ehrlich ascites carcinoma. Properties, subcellular distribution, and nature of enzyme complexes of the six biosynthetic enzymes. *Biochemistry* 12:4039-51.
- Simmer, J.P., R.E. Kelly, A.G. Rinker Jr., J.L. Scully, and D.R. Evans. 1990. Mammalian carbamyl phosphate synthetase (CPS). cDNA sequence and evolution of the CPS domain of the syrian hamster multifunctional protein CAD. *J. Biol. Chem.* 265:10395-10402.

- Slock, J., D.P. Stahly, C.Y. Han, E.W. Six, and L.P. Crawford.** 1990. An apparent *Bacillus subtilis* folic acid biosynthetic operon containing *pab*, an amphibolic *trpG* gene, a third gene required for synthesis of para-aminobenzoic acid, and the dihydropteroate synthase gene. *J. Bacteriol.* 172:7211-7226.
- Smith, H.O., J.F. Tomb, B.A. Dougherty, R.D. Fleischmann, and J.C. Venter.** 1995. Frequency and distribution of the DNA uptake signal sequence in the *H. influenzae* Rd genome. *Science* 269:538-540.
- Smith, M.C.M., L. Czaplewski, A.K. North, S. Baumberg, and P.G. Stockley.** 1989. Sequences required for regulation of arginine biosynthesis promoters are conserved between *Bacillus subtilis* and *Escherichia coli*. *Mol. Microbiol.* 3:23-28.
- Souciet, J.L., M. Nagy, M. Le Gouar, F. Lacroute, and S. Potier.** 1989. Organization of the yeast *URA2* gene: identification of a defective dihydroorotase-life domain in the multifunctional carbamoylphosphate synthetase-aspartate transcarbamoylase complex. *Gene* 79:59-70.
- Southern, E.M.** 1975. Selection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
- Sparling, P.F.** 1966. Genetic transformation of *Neisseria gonorrhoeae* to streptomycin resistance. *J. Bacteriol.* 92:1364-1371.
- Srb, A.M., and N.H. Horowitz.** 1944. The ornithine cycle in *Neurospora* and its genetic control. *J. Biol. Chem.* 154:129-139.
- Strathdee, C.A., S.D. Tyler, J.A. Ryan, W.M. Johnson, and F.E. Ashton.** 1993. Genomic fingerprinting of *Neisseria meningitidis* associated with group C meningococcal disease in Canada. *J. Clin. Microbiol.* 31:2506-2508.

- Stein, D.C., L.E. Silver, V.L. Clark, and F.E. Young. 1984. Cloning genes for proline biosynthesis from *Neisseria gonorrhoeae*: identification by interspecific complementation of *Escherichia coli* mutants. *J. Bacteriol.* **158**:696-700.
- Stewart, G.J. 1986. The biology of natural transformation. *Annu. Rev. Microbiol.* **40**:211-235.
- Subramaniam, P., R. Bhatnagar, A. Hooper, and R.A. Jensen. 1994. The dynamic progression of evolved character states for aromatic amino acid biosynthesis in Gram-negative bacteria. *Microbiology* **140**:3431-3440.
- Stirling, C.J., G. Szatmari, G. Stewart, M.C.M. Smith, and D.J. Sherratt. 1988. The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the *ColEI cer* locus. *EMBO J.* **7**:4389-4395.
- Swanson, J., R.J. Belland, and S.A. Hill. 1992. Neisserial surface variation: how and why? *Curr. Opinion Genet. Dev.* **2**:805-811.
- Taha, M.K., M. So, H.F. Seifert, E. Billyard, and C. Marchal. 1988. Pilin expression in *Neisseria gonorrhoeae* is under both positive and negative transcriptional control. *EMBO J.* **7**:4367-4378.
- Taylor, R.F.H., M.E. Hodson, and T.L. Pitt. 1992. Auxotrophy of *Pseudomonas aeruginosa* in cystic fibrosis. *FEMS Microbiol. Lett.* **92**:243-246.
- Tenover, F.C., R.D. Arbeit, R.V. Goering, P.A. Mickelsen, B.E. Murray, D.H. Persing, and B. Swaminathan. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J. Clin. Microbiol.* **33**:2233-2239.
- Tenover, F.C., and C.M. Patton. 1987. Naturally occurring auxotrophs of *Campylobacter jejuni* and *Campylobacter coli*. *J. Clin. Microbiol.* **25**:1659-1661.

- Trotta, P.P., L.M. Pinkus, R.H. Hashemeyer, and A. Meister.** 1974. Reversible dissociation of the monomer of glutamine-dependent carbamoylphosphate synthetase into catalytically active heavy and light subunits. *J. Biol. Chem.* 249:492-499.
- Turner, R.J., Y. Lu, and R.L. Switzer.** 1994. Regulation of the *Bacillus subtilis* pyrimidine biosynthetic (*pyr*) gene cluster by an autogenous transcriptional attenuation mechanism. *J. Bacteriol.* 176:3708-3722.
- Udaka, S.** 1966. Pathway-specific pattern of control of arginine biosynthesis in bacteria. *J. Bacteriol.* 91:617-621.
- Udaka, S., and S. Kinoshita.** 1958. Studies on L-ornithine fermentation. II. The change of fermentation product by a feedback type mechanism. *J. Gen. Appl. Microbiol.* 4:283-288.
- Urrestarazu, L.A., S. Vissers, and J.M. Wiame.** 1977. Change in location of ornithine carbamoyltransferase and carbamoylphosphate synthetase among yeasts in relation to the arginase/ornithine carbamoyl-transferase regulatory complex and the energy status of cells. *Eur. J. Biochem.* 79:473-481.
- Vedros, N.A.** 1984. *Neisseria*, p. 290-296. In N.R. Krieg (ed.), *Bergey's manual of systematic bacteriology*, vol. I. Williams & Wilkins, Baltimore.
- Vogel, H.J.** 1957. Repression and induction as control mechanisms of enzyme biogenesis: the "adaptive" formation of acetylornithinase, p. 276-296. In W.D. McElroy and B. Glass (ed.), *The chemical basis of heredity*. The John Hopkins Press, Baltimore.
- Wasserheit, J.N.** 1992. Epidemiological synergy. Relationships between human immunodeficiency virus infection and other sexually transmitted diseases. *Sexual. Trans. Dis.* 19:61-77.

- Werner, M., A. Feller, and A. Piérard.** 1985. Nucleotide sequence of the yeast gene *CPA1* encoding the small subunit of arginine-pathway carbamoylphosphate synthetase. Homology of deduced amino acid sequence to other glutamine amidotransferases. *Eur. J. Biochem.* **146**:371-381.
- Werner, M., A. Feller, F. Messenguy, and A. Piérard.** 1987. The leader peptide of yeast gene CPA1 is essential for the translational repression of its expression. *Cell* **49**:805-813.
- West, S.E.H., and V.L. Clark.** 1989. Genetic loci and linkage associations in *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Clin. Microbiol. Rev.* **2**:S92-S103.
- Whittington, W.L., and J.S. Knapp.** 1988. Trends in Resistance of *Neisseria gonorrhoeae* to antimicrobial agents in the United States. *Sex. Trans. Dis.* **15**:202-210.
- Whittington, W.L., R. Rice, J. Hale, and K.K. Holmes.** 1993. Relationship of AHU and PAU gonococcal auxotrophs to cervicitis, pelvic inflammatory disease (PID), and potential for transmission. abstr. 55., Tenth Inter. Meet. Inter. Soc. STD Res.
- Wilson, H.R., C.D. Archer, J. Liu, and C.L. Turnbough, Jr.** 1992. Translational control of *pyrC* expression mediated by nucleotide-sensitive selection of transcriptional start sites in *Escherichia coli*. *J. Bacteriol.* **174**:514-524.
- Winter, D.B., and S.A. Morse.** 1975. Physiology and metabolism of pathogenic *Neisseria*: partial characterization of the respiratory chain of *Neisseria gonorrhoeae*. *J. Bacteriol.* **123**:631-636.
- Woese, C.R.** 1987. Bacterial evolution. *Microbiol. Rev.* **51**:221-271.
- Wolff, K., S. Sperka, and A. Stern.** 1993. Phylogeny and nucleotide sequence of a 23S rRNA gene from *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Nuc. Acids Res.* **20**:4657.

- Wong, S.C., and A.T. Abdelal.** 1990. Unorthodox expression of an enzyme: evidence for an untranslated region within *carA* from *Pseudomonas aeruginosa*. J. Bacteriol. 172:630-642.
- Zhou, J., and B.G. Spratt.** 1992. Sequence diversity within the *argF*, *fbp* and *recA* genes of natural isolates of *Neisseria meningitidis*: interspecies recombination within the *argF* gene. Mol. Microbiol. 6:2135-2146.

6. APPENDIX

FIGURE A1. Alignment of all currently known CPS sequences of the amidotransferase domain, showing the region used for phylogenetic tree construction (corresponding to residues 36 to 367 of the deduced protein from *E. coli carA*). For organisms which contain two CPS enzymes, the sequence is identified by the letters “Arg” (for arginine-specific CPSII), “Pyr” (for pyrimidine-specific CPSII), “CPSI”, and “CPSIII”. A row of stars denotes where there is a large insertion of approximately 200 residues within the *P. falciparum* and *B. bovis* sequences. The location where a reactive cysteine residue has been found necessary for glutamine-binding is highlighted in pink.

<i>S. solfataricus</i>	GYVESLTDPS	YKGQILIITH	PLVGNYGVPE	KKYEQGILTN	FESE-----
<i>N. gonorrhoeae</i>	GYQEILTDPS	YCKQIVTLTY	PHIGNAGTNA	EDEESRT---	-----
<i>E. coli</i>	GYQEILTDPS	YSRQIVTLTY	PHIGNVGTND	ADEESSQ---	-----
<i>S. typhimurium</i>	GYQEIVTDPS	YSRQIVTLTY	PHIGNVGTNK	ADEESSQ---	-----
<i>P. aeruginosa</i>	GYQEILTDPS	YAQQIVTLTY	PHIGNTGTP	EDAEANR---	-----
<i>P. stutzeri</i>	GYQEILTDPS	YAKQIVTLTY	PHVNTGTTP	QDAESWK---	-----
<i>B. subtilis</i> (Pyr)	GYQEILSDPS	YCGQIVTLTY	PLIGNYGINR	DDFESIT---	-----
<i>B. subtilis</i> (Arg)	GYQEVLTDP	YKGQIIVFTY	PLIGNYGINE	KDFESKK---	-----
<i>B. caldolyticus</i> (Pyr)	GYQEILSDPS	YCGQIVTMTY	PLIGNYGINR	DDFEAIR---	-----
<i>N. crassa</i> (Arg)	GYPESMTDPS	YRGQILVFTQ	PLIGNYGVPS	NE-RDEFN--	-----
<i>S. cerevisiae</i> (Arg)	GYPESMTDPS	YRGQILVFTQ	PLIGNYGVPS	GEARDEYN--	-----
<i>S. cerevisiae</i> (Pyr)	GYPESVTDPS	YEGQILVITY	PLVGNYGVPD	MHLRDELVEE	-----
<i>D. discoideum</i> (Pyr)	GYNESISDPS	YTGQILVFSF	PLIGNYGVPS	FRERDPESG-	-----
<i>P. falciparum</i>	GYPEALTDPS	YFGQILVLT	PSIGNYGIEK	VKHDETFG--	-----
<i>B. bovis</i>	GYAEALTDPS	FLGQILVLT	PSVGNTGVPP	SEQVSVIELD	ITLVQDELRC
<i>S. acanthias</i> (CPSIII)	GYPETLTDPS	YRGQILTLTS	PMAGNYGVPD	TKKLDELG--	-----
<i>R. catesbeiana</i> (CPSI)	GYVEAVTDPS	YHGQILTLTN	PIIGNGGAPD	TKARDAYG--	-----
Rat (CPSI)	GYSEALTDPA	YKGQILTMAN	PIIGNGGAPD	TTARDELG--	-----
Human (CPSI)	GYPEAITDPA	YKGQILTMAN	PIIGNGGAPD	TTSLDELG--	-----
<i>S. solfataricus</i>	-----R	IQVEGLIVAE	HTYPSKWN	LTLDEWLKSE	NVPGVFDVDT
<i>N. gonorrhoeae</i>	-----V	LPAGLIIRD	PLLHSNFRAS	ESLHDYLVN	ETVAIADIDT
<i>E. coli</i>	-----V	HAQGLVIRD	PLIASNFRNT	EDLSSYLKRH	NIVAIADIDT
<i>S. typhimurium</i>	-----V	HAQGLVIRD	PLIASNFRNT	EDLSSYLKRH	NIVAIADIDT
<i>P. aeruginosa</i>	-----V	WAAGLIIRD	PLIASNWRSK	QSLPDYLRN	GTVAIAGIDT
<i>P. stutzeri</i>	-----V	WAAGLVIRD	PLIASNWRNK	QSLPDYLRN	DTVAIAGIDT
<i>B. subtilis</i> (Pyr)	-----P	FVKGLIIEK	CELPSNWRSA	YTLDEYLKMK	NIPGLQGIDT
<i>B. subtilis</i> (Arg)	-----P	QVKAADVVEA	CDHFSHYEAV	YSLKEYLQKW	NIPLLTHVDT
<i>B. caldolyticus</i> (Pyr)	-----P	HVHGFIKVEA	CVKPSNWRGE	LTLDEYLKEK	GIPGLSGIDT
<i>N. crassa</i> (Arg)	LLSTSSLPTS	SAPVSLSPMS	LPSTATGLLF	RAWASGAPAR	AFPPLSGVDT
<i>S. cerevisiae</i> (Arg)	LLKYFESPHI	HVVGIVVAEY	AYQYSHWTAV	ESLAQWCQRE	GVAAITGVDT
<i>S. cerevisiae</i> (Pyr)	LPRYFESNRI	HIAGLVISHY	TDEYSHYLAK	SSLGKWLQNE	GIPAVYGVDT
<i>D. discoideum</i> (Pyr)	LAVNFESDKA	HVQAIICSEY	CDEYSHWAAE	KSLSEWLKES	NIPGLYGIDT
<i>P. falciparum</i>	LVQNFESNKI	QVQGLVICEY	SKQSYHYNSY	ITLSEWLKIY	KIPCIGGIDT
<i>B. bovis</i>	LRSGFESSRI	HVNGFVCCDY	SIYESHWSSC	KSLSSWLREE	RIPAIAGIDT
<i>S. acanthias</i> (CPSIII)	LMKYVESEFI	QVSGLLVQDY	SYEYSHWNSV	KSLGEWLHEE	KIPALYGIDT
<i>R. catesbeiana</i> (CPSI)	LMKYIESENI	QASGLLVLDY	SHEYSHWGAV	KSLSEWLHEE	KVPALCGIDT
Rat (CPSI)	LNKYMESDGI	KVAGLLVLNY	SHDYNHWLAT	KSLGQWLQEE	KVPAIYGVDT
Human (CPSI)	LSKYLESNGI	KVSGLLVLDY	SKDYNHWLAT	KSLGQWLQEE	KVPAIYGVDT

<i>S. solfataricus</i>	RMIVKKIRTY	GTMMGIIASE	LEIDDPKRYL	EKK-----	-----
<i>N. gonorrhoeae</i>	RRLTMLLREK	G-QGGAILTG	ADATVEKAQE	LIAAFGSMVG	KDLAKEVSCT
<i>E. coli</i>	RKLTRLLEK	GAQNGCIIAG	DNPDAALALE	KARAFPGLNG	MDLAKEVTTA
<i>S. typhimurium</i>	RKLTRLLEK	GAQNGCIIAG	DSPDAKLALALE	KAKAFPGLNG	MDLAKEVTTA
<i>P. aeruginosa</i>	RRLTRILREK	GSQNGCILAG	ADATEERALE	LARAFPGLKG	MDLAKEVTTA
<i>P. stutzeri</i>	RRLTRILREK	GAQNGCILAG	ADATEEKALE	LARSFPGLKG	MDLAKEVSVK
<i>B. subtilis (Pyr)</i>	RKLTRMIRTA	GALKGTFASS	DEDIEAVLKR	LNETELPRNQ	VS-----
<i>B. subtilis (Arg)</i>	RAVVVKIRAN	GTMGATVTAS	KEGAEIALQP	ENVAE-----	-----
<i>B. caldolyticus (Pyr)</i>	RKLTRLIRQY	GTLKGMICGL	DVDPVEAAAK	LRAMEWPRDQ	VR-----
<i>N. crassa (Arg)</i>	RAIVTYLREQ	GSSLARISIG	DEYDAD----	-----ED	EGFIDPGQIN
<i>S. cerevisiae (Arg)</i>	RELVOYLREQ	GSSLGRITLA	DH-----	-----DP	VPYVNPMTN
<i>S. cerevisiae (Pyr)</i>	RSLTKHLRDA	GSMLGRSLSE	KSGSDRTISR	SSSWRSADFV	PEWVDPNVQN
<i>D. discoideum (Pyr)</i>	RALITKIREK	GSLKGKVIIG	DFDESKL----	-----	-EFEDINLRN
<i>P. falciparum</i>	RALTLLREK	GSMLGKIVII	KNRQHI--NK	LYKEINLFDP	GNIDTLKYVC
<i>B. bovis</i>	RALTKHLRNC	GSTLARIIG	PKSRGLVSPR	LLESSSFYDT	NNPDLMRSLP
<i>S. acanthias (CPSIII)</i>	RMLTKIVRD-	KGTLGKIEF	EGDPV-----	-----	-EFVDPNLRN
<i>R. catesbeiana (CPSI)</i>	RMLAKKIRDN	KGAVLGKIEF	EGQPV-----	-----	-EFIDPNKRN
<i>Rat (CPSI)</i>	RMLTKIIRD-	KGTMKGKIEF	EGQSV-----	-----	-DFVDPNKQN
<i>Human (CPSI)</i>	RMLTKIIRD-	KGTMKGKIEF	EGQPV-----	-----	-DFVDPNKQN
<i>S. solfataricus</i>	-----	-----	-----	-----*	-----Y DEIDFTQFTS
<i>N. gonorrhoeae</i>	-----	-----	-----	-----*	ETYEW TEGEWELGKG
<i>E. coli</i>	-----	-----	-----	-----*	EAYSW TQGSWTLTGG
<i>S. typhimurium</i>	-----	-----	-----	-----*	ETYRW TQGSWTLKDG
<i>P. aeruginosa</i>	-----	-----	-----	-----*	ERYEW RSSVWNLESD
<i>P. stutzeri</i>	-----	-----	-----	-----*	ERYEW RYSLWNLETD
<i>B. subtilis (Pyr)</i>	-----	-----	-----	-----*	-----QVS
<i>B. subtilis (Arg)</i>	-----	-----	-----	-----*	-----QAS
<i>B. caldolyticus (Pyr)</i>	-----	-----	-----	-----*	-----RVS
<i>N. crassa (Arg)</i>	LVKRVSTKAP	-----	-----	-----*	-----
<i>S. cerevisiae (Arg)</i>	LVAQVTTKKP	-----	-----	-----*	-----
<i>S. cerevisiae (Pyr)</i>	LVSKVSINEP	KLYVPPADNK	HIELQTGPDG	KVLR*-----	-----
<i>D. discoideum (Pyr)</i>	LVAEVSTKEI	KEYKAAENNK	KTGEKRKNKK	-----*	-----
<i>P. falciparum</i>	NHFIRVIKLN	NITYNYKNKE	EFNYTNEMIT	NDSS*STYIR	KMKKNEEFLN
<i>B. bovis</i>	DPHPVLYTMS	EID-GERYVT	SYEFTVAELD	DILS*-----	-----S
<i>S. acanthias (CPSIII)</i>	LMAEISTKEI	KVYGKGNNAIK	-----	-----*	-----
<i>R. catesbeiana (CPSI)</i>	LIAEVSTKET	KVFGKGNPVR	-----	-----*	-----
<i>Rat (CPSI)</i>	LIAEVSTKDV	KVFGKGNPTK	-----	-----*	-----
<i>Human (CPSI)</i>	LIAEVSTKDV	KVYGKGNPTK	-----	-----*	-----

<i>S. solfataricus</i>	PKSPIFHPNT	GDMIVVVDCG	IKHGILYGLY	KRGFSIVRVP	CSFSASKIIE
<i>N. gonorrhoeae</i>	F---VTPDKQ	PYHVVAIDFG	VKTNILRI-S	PRGCRLTVVP	AQTSADVLVA
<i>E. coli</i>	LPEAKKEDEL	PFHVVAIDFG	AKRNILRMLV	DRGCRLTIVP	AQTSADVLVK
<i>S. typhimurium</i>	LPEAKSEDDL	PFHVVAIDFG	AKRNILRMLV	DRGCRLTVVP	AQTSAAEVLK
<i>P. aeruginosa</i>	SHPEIPAGEL	PYHVVAIDYG	VKLNILRMLV	ARGCRLTVVP	AQTPASEVLA
<i>P. stutzeri</i>	SHPEIPASEL	PYHVVAIDYG	VKYNILRMLV	ARGCRVTVLP	AQTPASEALA
<i>B. subtilis (Pyr)</i>	AKTAYPSPGR	GKRIVLVDFG	MKHGILRELN	KRKCDVIVVP	YNITAAEVLQ
<i>B. subtilis (Arg)</i>	AQEISTFGDG	NKHIALIDFG	YKKSIASSLV	KRGCKVTVVP	YQQM-EAVYN
<i>B. caldolyticus (Pyr)</i>	TKSAYPSPGR	GERIVLIDFG	MKHGILRELN	KRNCDVIVLP	YNATAAEVLG
<i>N. crassa (Arg)</i>	FVVTN--PNA	KFHVALIDCG	VKENILRSLV	SRGASVTVFP	--YNYPIHKV
<i>S. cerevisiae (Arg)</i>	FHVSALPGKA	KANVALIDCG	VKENIIRCLV	KRGANVTVFP	--YDYRIQDV
<i>S. cerevisiae (Pyr)</i>	-----	---ILAIIDVG	MKYNQIRCFI	KRGVELKVVP	WNYDFTKE--
<i>D. discoideum (Pyr)</i>	-----	---VIVLDCG	IKNNQIRCLL	NRGVDLKVVP	WDYDVVANES
<i>P. falciparum</i>	LVNKRKVDHK	EKIIVIVDCG	IKNSIIKNLI	RHGMDLPLTY	IIVPYYYNFN
<i>B. bovis</i>	LYSSSLSLKG	VTLVVIVDCG	IKSNIIRLFL	RMS-FVQVRA	LVPHPNFDNF
<i>S. acanthias (CPSIII)</i>	-----	---IVAVDCG	IKHNIIRLLV	KRGVELHLVP	WDYDFTNME-
<i>R. catesbeiana (CPSI)</i>	-----	---IVAVDCG	VKHNIIRQLV	KRGAEVHLVP	WNHDFSQME-
Rat (CPSI)	-----	---VAVDCG	IKNNVIRLLV	KRGAEVHLVP	WNHDFQMD-
Human (CPSI)	-----	---VAVDCG	IKNNVIRLLV	KRGAEVHLVP	WNHDFTKME-
<i>S. solfataricus</i>	YNP-KGIVFS	NGPGNPNLLE	NQIKTFSELV	EYK----IPI	LGICLGHQIA
<i>N. gonorrhoeae</i>	LNP-DGVFLS	QRPGDPEPCT	YAIEAVQKLM	ESG----KPI	FGICLGHQLI
<i>E. coli</i>	MNP-DGIFLS	NGPGDPAPCD	YAITAIQKFL	ETD----IPV	FGICLGHQLL
<i>S. typhimurium</i>	MNP-DGIFLS	NGPGDPAPCD	YAITAIQKFL	ETD----IPL	FGICLGHQLL
<i>P. aeruginosa</i>	LNP-DGIFLS	NGPGDPEPCD	YAIQAIREFL	DTE----IPV	FGICLGHQLL
<i>P. stutzeri</i>	LNP-DGIFLA	NGPGDPEPCD	YAIRAIQEV	ETD----IPV	FGICLGHQLL
<i>B. subtilis (Pyr)</i>	LKP-DGIMLS	NGPGDPKDVP	EAIEMIKGVL	GK-----VPL	FGICLGHQLF
<i>B. subtilis (Arg)</i>	IKP-DGIVLS	NGPGDPKAIQ	PYLGKIKSII	SR-----FPT	LGICLGHQLI
<i>B. caldolyticus (Pyr)</i>	WHP-DGVMLS	NGPGDPKDVP	EAIEMIRGIL	GK-----VPL	FGICLGHQLF
<i>N. crassa (Arg)</i>	AENFDGVFIS	NGPGDPTHQ	ETVYNLAKLM	ETSP---IPI	MGICLGHQLL
<i>S. cerevisiae (Arg)</i>	ASEFDGIFLS	NGPGNPELCQ	ATISNVRELL	NNPVYDCIPI	FGICLGHQLL
<i>S. cerevisiae (Pyr)</i>	--DYDGLFIS	NGPGDPSVLD	DLSQRLSNVL	EAKKT---PV	FGICLGHQLI
<i>D. discoideum (Pyr)</i>	INDYDGVFIS	NGPGDPSLCG	KAIVENIRKVL	ALPVAK--AV	FGVCMGNQLL
<i>P. falciparum</i>	HIDYDAVLLS	NGPGDPKKCD	FLIKNLKDSL	TKNK----II	FGICLGNQLL
<i>B. bovis</i>	RIPYDGLIIS	NGPGDPSDAT	VTIANLRRAM	ERTT----PI	FGICLGHQLM
<i>S. acanthias (CPSIII)</i>	---YDGLFIS	NGPGDPTLAG	QLVANVHKVI	GSNRPE--PV	FGICMGNQLT
<i>R. catesbeiana (CPSI)</i>	---YDGLLIT	SGPGNPELAK	PLIQNLKKVF	QSDRPE--PI	FGICKGNEIA
Rat (CPSI)	---YDGLLIA	GGPGNPALAQ	PLIQNVKKIL	ESDRKE--PL	FGISTGNIIT
Human (CPSI)	---YDGILIA	GGPGNPALAE	PLIQNVQKIL	ESDRKE--PL	FGISTGNLIT

<i>S. solfataricus</i>	TLALGGKIKK	MKFGHRAIN	PVIESNSNKC	YISTHNGY	IISKNDIPP-
<i>N. gonorrhoeae</i>	SLAIGAKTLK	MRFSHHGANH	PVQDLDSGKV	VITSQNHGFA	VDADTLPAN-
<i>E. coli</i>	ALASGAKTVK	MKFGHHGGNH	PVKDVEKNV	MITAQNHGFA	VDEATLPAN-
<i>S. typhimurium</i>	ALASGAKTVK	MKFGHHGGNH	PVKDMDRNV	MITAQNHGFA	VDEDSL PAN-
<i>P. aeruginosa</i>	ALASGAKTLK	MHGHHGANH	PVQDLDSGV	MITSQNHGFA	VDESTLPD-
<i>P. stutzeri</i>	ALASGAKTVK	MPSGHHGANH	PVQDLETGV	MITSQNHGFC	ADEASLPAN-
<i>B. subtilis (Pyr)</i>	ALACGANTEK	MKFGHRGSNH	PVKELATGKV	ALTSQNHGYT	VS--SISKT-
<i>B. subtilis (Arg)</i>	ALAFGGNTFK	LPFGHRGANH	PVIDRKT KRV	FMTSQNHSYV	VDEQSINEE-
<i>B. caldolyticus (Pyr)</i>	ALACGANTEK	MKFGHRGSNH	PVKDLRTGKV	AITSQNHGYT	VTHESLSGT-
<i>N. crassa (Arg)</i>	ALAVGAKTIK	LSMVIRAHNI	PALDLTTGQC	HITSQNHGYA	VDISTLPD-
<i>S. cerevisiae (Arg)</i>	ALASGASTHK	LKYGNRAHNI	PAMDLTGQC	HITSQNHGYA	VDPETLPKD-
<i>S. cerevisiae (Pyr)</i>	ARAAGASTLK	LKFGNRGHNI	PCTSTISGR	YITSQNHGFA	VDVDTLT--S
<i>D. discoideum (Pyr)</i>	GLAAGAQTHK	MAFGNRGLNQ	PCVDQISGR	HITSQNHGFV	IDSNSLPAGS
<i>P. falciparum</i>	GISLGCDTYK	MKYGNRGVNQ	PVIQLVDNIC	YITSQNHGYC	LKKKSILKRR
<i>B. bovis</i>	GLAAGAKTYK	MRYGHRGFNQ	PCVDLRTSK	YMTSQNHGYA	IDEETLPSEW
<i>S. acanthias (CPSIII)</i>	ALAVGANCYK	LPMGNRGQNO	PVVNMNGQA	FITAQNHGYA	IDSSTLP EG-
<i>R. catesbeiana (CPSI)</i>	ALAAGGKTYR	LPMANRGQNO	PVMITLNGQA	FITAQNHAYA	VDNNSLPAG-
<i>Rat (CPSI)</i>	GLAAGAKSYK	MSMANRGQNO	PVLNITNRQA	FITAQNHGYA	LDN-TLPAG-
<i>Human (CPSI)</i>	GLAAGAKTYK	MSMANRGQNO	PVLNITNKQA	FITAQNH CYA	LDN-TLPAG-
<i>S. solfataricus</i>	NTKIWFYNPD	DYTI EGLIHE	KLPIITTOFH	PEARPGP WDT	TWVF
<i>N. gonorrhoeae</i>	-ARITHKSLF	DNTLQGI ELT	DKPVFCFQGH	PEASPGPQDV	GYLF
<i>E. coli</i>	-LRVTHKSLF	DGTLQGIHRT	DKPAFSFQGH	PEASPGPHDA	APLF
<i>S. typhimurium</i>	-LRVTHKSLF	DGTLQGIHRT	DKPAFSFQGH	PEASPGPHDA	APLF
<i>P. aeruginosa</i>	-LRATHKSLF	DGTLQGIERT	DKVAFSFQGH	PEASPGPHDV	APLF
<i>P. stutzeri</i>	-LRATHKSLF	DGTLQGIERT	DKVAFGFQGH	PEASPGPCDV	APLF
<i>B. subtilis (Pyr)</i>	ELEVTHIAIN	DDTIEGLKHK	TLPAFTVQYH	PEASPGPEDA	NHLF
<i>B. subtilis (Arg)</i>	ELTIRFHHVN	DTSVEGLAHK	KLPVMSVQFH	PEAHPGA AES	EWIF
<i>B. caldolyticus (Pyr)</i>	RLEVTHIALN	DGTVEGLRHL	DVPAFTVQYH	PEASPGPEDA	NPLF
<i>N. crassa (Arg)</i>	-FKEYFVN LN	DGSNEGMMHK	TRPIFSTQFH	PEAKGGPMD S	SYLF
<i>S. cerevisiae (Arg)</i>	QWKPYFVN LN	DKSNEGMIHL	QRPIFSTQFH	PEAKGGPLD T	AILF
<i>S. cerevisiae (Pyr)</i>	GWKPLFVNAN	DDSNEGIYHS	ELPYFSVQFH	PESTPGPRD T	EFLF
<i>D. discoideum (Pyr)</i>	GWKTYFINAN	DASNEGIYHE	SKPWFSVQFH	PEAMAGPTD T	EYLF
<i>P. falciparum</i>	ELAISYINAN	DKSIEGISHK	NGRFYSVQFH	PEGNNGPED T	SFLF
<i>B. bovis</i>	---LRYCDAN	DGCVEGIIHM	TYPWFSLQFH	PEASGGPTD T	LFLM
<i>S. acanthias (CPSIII)</i>	-WKPMFVNAN	DGTNEGIVHD	SKPIFTAQFH	PEAKGGPTD T	EFLF
<i>R. catesbeiana (CPSI)</i>	-WKPLFVNIN	DQSN EGIMHE	TKPIFTSQFH	PEANPGPVD T	EFLF
<i>Rat (CPSI)</i>	-WKPLFVN VN	DQTNEGIMHE	SKPFFAVQFH	PEVSPGPTD T	EYLF
<i>Human (CPSI)</i>	-WKPLFVN VN	DQTNEGIMHE	SKPFFAVQFH	PEVTPGPID T	EYLF

FIGURE A2. Alignment of all currently known CPS sequences of the synthetase domain, showing the region used for phylogenetic tree construction (corresponding to residues 12 to 916 of the deduced protein from *E. coli carB*). For organisms which contain two CPS enzymes, the sequence is identified by the letters “Arg” (for arginine-specific CPSII), “Pyr” (for pyrimidine-specific CPSII), “CPSI”, and “CPSIII”. A row of stars denotes where there is a large insertion of approximately 600 residues within the *P. falciparum* and *B. bovis* sequences. The location where two cysteine residues are present in CPS enzymes which bind acetylglutamate (CPSI and CPSIII-type enzymes) is highlighted in blue.

<i>S. solfataricus</i>	VIGSGPIKIA	EAAEFDYSGS	QALKALKEEG	IETVLVNSNV	ATVQTSKK--
<i>N. gonorrhoeae</i>	IIGAGPIVIG	QACEFDYSGA	QACKALREEG	YKVILVNSNP	ATIMTDPE--
<i>E. coli</i>	ILGAGPIVIG	QACEFDYSGA	QACKALREEG	YRVILVNSNP	ATIMTDPE--
<i>P. aeruginosa</i>	ILGAGPIVIG	QACEFDYSGA	QACKALREEG	YRVILVNSNP	ATIMTDPA--
<i>P. stutzeri</i>	ILGAGPIVIG	QACEFDYSGA	QACNGLREEG	FRVILVNSNP	ATIMTDPA--
<i>B. subtilis</i> (Arg)	VIGSGPIIIG	QAAEFDYSGT	QGCIALKEEG	YRVILVNSNP	ATIMTDEA--
<i>B. subtilis</i> (Pyr)	VIGSGPIIIG	QAAEFDYAGT	QACLALKEEG	YEVILVNSNP	ATIMTDTE--
<i>B. caldolyticus</i> (Pyr)	VIGSGPIVIG	QAAEFDYAGT	QACLALKEEG	YKVILVNSNP	ATIMTDTE--
<i>S. cerevisiae</i> (Arg)	VIGSGGLSIG	QAGEFDYSGS	QAIKALKEDN	KFTILVNPNI	ATNQTSHS--
<i>S. cerevisiae</i> (Pyr)	VLGSGGLSIG	QAGEFDYSGS	QAIKALKEEG	IYTLINPNI	ATIQTSGK--
<i>D. discoideum</i> (Pyr)	ILGSGGLSIG	QAGEFDYSGS	QAIKALKEEG	IKTILINPNI	ATVQTSFG--
<i>P. falciparum</i>	LLGSGGLCIG	QAGEFDYSGT	QAIKSLKECG	IYVILVNPNI	ATVQTSKG--
<i>B. bovis</i>	LLSSGGISIG	QAGEFDYSGS	QAILALKESG	AEVILVNPNI	ATVQTNHG--
<i>T. cutaneum</i> (Arg)	VVGSGGLSIG	QAGEFDYSGS	QAIKALRESN	IETILINPNI	ATIQTSHH--
<i>R. catesbieana</i> (CPSI)	ILGSGGLSIG	QAGEFDYSGS	QAVKAMKEEN	VKTVLMNPNI	ASVQTNEVGL
<i>S. acanthias</i> (CPSIII)	VLGSGGLSIG	QAGEFDYSGS	QAIKALKEEN	LKTVLMNPNI	ASVQTNEVGT
Hamster (Pyr)	ILGSGGLSIG	QAGEFDYSGS	QAIKALKEEN	IQTLINPNI	ATVQTSQG--
Human (CPSI)	ILGSGGLSIG	QAGEFDYSGS	QAVKAMKEEN	VKTVLMNPNI	ASVQTNEVGL

<i>S. solfataricus</i>	-FADKLYMLP	VVWVAWEKVI	EKERPDGIMI	GFGGQTALNV	GVDLHKKGV
<i>N. gonorrhoeae</i>	-MADVTYIEP	IMWQTVEKII	AKERPDAILP	TMGGTDRLNC	ALDLAGNGVL
<i>E. coli</i>	-MADATYIEP	IHWVVRKII	EKERPDVLP	TMGGQTALNC	ALEJERQGV
<i>P. aeruginosa</i>	-MADATYIEP	IKWATVAKII	EKERPDALLP	TMGGQTALNC	ALDLERHGV
<i>P. stutzeri</i>	-MADATYIEP	IKWQTVAKII	EKERPDALLP	TMGGQTALNC	ALDLERHGV
<i>B. subtilis</i> (Arg)	-FADEIYFEP	LTAESLTAII	KKERPDGLLA	NLGGQTALNL	AVELEETGV
<i>B. subtilis</i> (Pyr)	-MADRVYIEP	LTPEFLTRII	RKERPDAILP	TLGGQTGLNL	AVELSERGV
<i>B. caldolyticus</i> (Pyr)	-IADKVMEP	LTLDFVARI	RKERPDAILP	TLGGQTGLNL	AVELAKAGV
<i>S. cerevisiae</i> (Arg)	-LADKIYYLP	VTPEYITYII	ELERPDAILL	TFGGQTGLNC	GVALDESGV
<i>S. cerevisiae</i> (Pyr)	-LADKVYFVP	VTAEFVRKVI	LHERPDAIYV	TFGGQTALSV	GIAM--KDEF
<i>D. discoideum</i> (Pyr)	-LADKVYFLP	VNASSVQKVI	ENENPDGILV	TFGGQTALNC	GIELYKSGIL
<i>P. falciparum</i>	-LADKVYFLP	VNCEFVEKII	KKEKPDFILC	TFGGQTALNC	ALMLDQKKV
<i>B. bovis</i>	--LADVYFE	LLLLIVSNII	EKERPDGIMC	SFGGQTALNC	GIELYKSGIL
<i>T. cutaneum</i> (Arg)	-LASEIYFLP	VTADYVAYVL	EKERPDGILL	TFGGQSALNV	GIQLEKMGV
<i>R. catesbieana</i> (CPSI)	KQADTVYFLP	ITPQFVTEVI	KAETDGIIL	GMGGQTALNC	GVELFKRGV
<i>S. acanthias</i> (CPSIII)	KQADTVYFLP	VTPEFVTDII	KTEKPDGILL	SMGGQTALNC	GVALYKQGV
Hamster (Pyr)	-LADKVYFLP	ITPHYVTQVI	RNERPDGVLL	TFGGQTALNC	GVELTKAGV
Human (CPSI)	KQADTVYFLP	ITPQFVTEVI	KAEQPDGLIL	GMGGQTALNC	GVELFKRGV

<i>S. solfataricus</i>	QKYNVKVLGT	QIDGIEKALS	REKFRETMI	NNLPVPPSL	ARSEEEAIKN
<i>N. gonorrhoeae</i>	AKYDVELIGA	TEDAIDKAED	PGRFKEAMEK	IGLSCPksLF	CHTMNEALAA
<i>E. coli</i>	EEFGVTMIGA	TADAIDKAED	RRRFDVAMKK	IGLETARSGI	AHTMEEALAV
<i>P. aeruginosa</i>	EKFGVEMIGA	NADTIDKAED	RSRFDKAMKD	IGLACPRSGI	AHSMEAYGV
<i>P. stutzeri</i>	EKFGVEMIGA	NADTIDKAED	RSRFDKAMRD	IGLACPVSGI	AHNMDAYGV
<i>B. subtilis</i> (Arg)	KEHGVKLLGT	SVETIQKGED	REKFRSLMNE	LKQVPPESEI	VDNEADALHF
<i>B. subtilis</i> (Pyr)	AECGVEVLGT	KLSAIQQAED	RDLFRTLME	LNPEVPPESEI	IHSLEEAEKF
<i>B. caldolyticus</i> (Pyr)	EECGVEILGT	KLEAIEKAED	REQFRALMNE	LGEPVPESAI	IHSLEEAYAF
<i>S. cerevisiae</i> (Arg)	AKYNVKVLGT	PIKTLITSED	RDLFASALKD	INIPIAESFA	CETVDEALEA
<i>S. cerevisiae</i> (Pyr)	EALGVKVLGT	PIDTIITTED	RELFSNAIDE	INEKCAKSQA	ANSVDEALAA
<i>D. discoideum</i> (Pyr)	EKYNCKVLGT	PIETIATED	RGIFAEKLSE	INERAPIPSMA	CNSLEESLIE
<i>P. falciparum</i>	KKNNCQCLGT	SLESIRITEN	RTLFAEKLKE	INERAPIYGS	AKNVNQAI
<i>B. bovis</i>	SKYNCEVLGT	PIETIINTED	RALFNRLKAE	IGERCAPSKV	GTDVGCISIA
<i>T. cutaneum</i> (Arg)	ERLGVQVLGT	PIRTLEISED	RDLFVQALNE	IDIPAAQSTA	VSTIQDALDA
<i>R. catesbieana</i> (CPSI)	KEYGVRVLGT	SVESIMPTED	RQLFSDKLNE	IKEPIAPSFA	VESVKDALEA
<i>S. acanthias</i> (CPSIII)	EKYGVQVLGT	PITSVMATED	RQLFADKLLE	INEKIAPSFA	VETLEDAFQA
Hamster (Pyr)	ARYGVRVLGT	PVETIELTED	RRFAARMAE	IGEHVAPSEA	ANSLEQAQAA
Human (CPSI)	KEYGVKVLGT	SVESIMATED	RQLFSDKLNE	INEKIAPSFA	VESIEDALKA

<i>S. solfataricus</i>	AKIVGYPVMV	RVSFNLGGRG	SMVAWTEEDL	KKNIRRALSQ	SYIGEVILLEK
<i>N. gonorrhoeae</i>	QEQVGFPPTLI	RPSFTMGGS	GGIAYNKDEF	LAICERGFDA	SPTHELLIEQ
<i>E. coli</i>	AADVGFPCII	RPSFTMGGS	GGIAYNREEF	EEICARGLDL	SPTKELLIDE
<i>P. aeruginosa</i>	LEQVGFPFCII	RPSFTMGGTG	GGIAYNREEF	EEICARGLDL	SPTNELLIDE
<i>P. stutzeri</i>	LDKVGFPFCII	RPSFTMGGTG	GGIAYNREEF	EEICARGLDL	SPTSSELLIDE
<i>B. subtilis</i> (Arg)	AESIGFPVII	RPAYTLGGKG	GGIAPDKEAF	TAMIKQALLA	SPINQCLVEK
<i>B. subtilis</i> (Pyr)	VSQIGFPVIV	RPAYTLGGTG	GGICSNETEL	KEIVENGLKL	SPVHQCLLEK
<i>B. caldolyticus</i> (Pyr)	VEQIGYPVIV	RPAFTLGGTG	GGICTNEEEL	VEIVSTGLKL	SPVHQCLLER
<i>S. cerevisiae</i> (Arg)	AERVKYPVIV	RSAYALGGLG	SGFANNASEM	KELAAQSLSL	AP--QILVEK
<i>S. cerevisiae</i> (Pyr)	VKEIGFPVIV	RAAYALGGLG	SGFANNEKEL	VDLCNVAFFS	SP--QVILVEK
<i>D. discoideum</i> (Pyr)	AEKIGYPVIV	RAAYCLGGLG	SGFADNKEQL	TALVTEAMAT	SS--QVILVEK
<i>P. falciparum</i>	ANKIGYPILV	RTTFSLGGLN	SSFINNEEEL	IEKCNKIFLQ	TD-NEIFIDK
<i>B. bovis</i>	AQELGYPVLV	RTNYALGGFG	SSGLASDESEL	RSILSNIFST	SSCRKVVYIDK
<i>T. cutaneum</i> (Arg)	AKTIGYPIIL	RSASFSLGGLG	S-FPHDEEEL	RNLAAKSLSL	SP--QVILIEK
<i>R. catesbieana</i> (CPSI)	ADKIGYPVMI	RSAYALGGLG	SGLCPDKETL	TDLATKALAM	TN--QILVER
<i>S. acanthias</i> (CPSIII)	AEKIGYPVMV	RAAYALGGLG	SGLCNNKEKL	NEIAGNALAM	TT--QILVEQ
Hamster (Pyr)	AERLGPVLV	RAAFALGGLG	SGFASTKEEL	SALVAPAFAH	TS--QVLIDK
Human (CPSI)	ADTIGYPVMI	RSAYALGGLG	SGICPNRETL	MDLSTKAFAM	TN--QILVEK

<i>S. solfataricus</i>	YLYHWIELEY	EVMRDKKGN	AVIACIENLD	PMGVHTGEST	VVAPCQTLDN
<i>N. gonorrhoeae</i>	SVLGWKEYEM	EVVRDKADNC	IIICSIENFD	PMGVHTGDSI	TVAPAETLTD
<i>E. coli</i>	SLIGWKEYEM	EVVRDKNDNC	IIVCSIENFD	AMGIHTGDSI	TVAPAQTLTD
<i>P. aeruginosa</i>	SLIGWKEYEM	EVVRDKKDC	IIVCSIENFD	PMGVHTGDSI	TVAPAQTLTD
<i>P. stutzeri</i>	SLIGWKEYEM	EVVRDKKDC	IIVCSIENFD	PMGVHTGDSI	TVAPAQTLTD
<i>B. subtilis</i> (Arg)	SIAGFKEIEY	EVMRDSNNTC	ITVCNMENID	PVGVTGDSI	VVAPSQTLTD
<i>B. subtilis</i> (Pyr)	SIAGYKEIEY	EVMRDSQDHA	IVVCNMENID	PVGHTGDSI	VVAPSQTLSD
<i>B. caldolyticus</i> (Pyr)	SIAGYKEIEY	EVMRDANDNA	IVVCNMENID	PVGHTGDSI	VVAPSQTLSD
<i>S. cerevisiae</i> (Arg)	SLKGWKEVEY	EVVRDRVGN	ITVCNMENFD	PLGVHTGDSM	VVAPSQTLSD
<i>S. cerevisiae</i> (Pyr)	SMKGWKEVEY	EVVRDAFDC	ITVCNMENFD	PLGIHTGDSI	VVAPSQTLSD
<i>D. discoideum</i> (Pyr)	SLKGWKEIEY	EVLRDSKDC	ITVCNMENFD	PLGIHTGDSI	VVAPSQTLSD
<i>P. falciparum</i>	SLQGWKEIEY	ELLRDKNKNC	IAICNMENID	PLGIHTGDSI	VVAPSQTLSD
<i>B. bovis</i>	ALKGWKEIEF	EIIRDNDNC	ISPASMENFD	PLGIHTGDSI	VVAPAQTLTN
<i>T. cutaneum</i> (Arg)	SLKGWKEVEY	EVVRDAADNT	IICCNMENFD	PLGHTGDSI	VVAPSQTLTD
<i>R. catesbieana</i> (CPSI)	SVVGWKEIEY	EVVRDAADNC	VTVCNMENVD	AMGVHTGDSI	VVAPCQTLSD
<i>S. acanthias</i> (CPSIII)	SLLGWKEVEY	EVVRDAADNC	VTVCNMENFD	PLGIHTGDSV	VVAPSQTLSD
Hamster (Pyr)	SLKGWKEIEY	EVVRDAYGNC	VTVCNMENLD	PLGIHTGESI	VVAPSQTLND
Human (CPSI)	SVTGWKEIEY	EVVRDADDNC	VTVCNMENVD	AMGVHTGDSV	VVAPAQTLSD

<i>S. solfataricus</i>	LEYQNMRTYT	IEVARSINL-	IGECNVQFAL	NPRGYEYII	ETNPR-MSRS
<i>N. gonorrhoeae</i>	KEYQIMRNAS	LAVLREIGVD	TGGSNVQFAV	NPEKRRDDCD	RDEPAAVSRS
<i>E. coli</i>	KEYQIMRNAS	MAVLREIGVE	TGGSNVQFAV	NPKNGRLIVI	EMNPR-VSRS
<i>P. aeruginosa</i>	KEYQIMRNAS	LAVLREIGVE	TGGSNVQFGI	CPNTGRMVVI	EMNPR-VSRS
<i>P. stutzeri</i>	KEYQIMRNAS	LAVLREIGVE	TGGSNVQFGI	HPDTGRMVVI	EMNPR-VSRS
<i>B. subtilis</i> (Arg)	EDYQMLRTAS	LTIISALDV-	VGGCNIQFAL	DPFSKYQYVI	EVNPR-VSRS
<i>B. subtilis</i> (Pyr)	REYQLLRNVS	LKLIRALGI-	EGGCNVQLAL	DPDSFQYII	EVNPR-VSRS
<i>B. caldolyticus</i> (Pyr)	REYQLLRNAS	LKIIRALGI-	EGGCNVQLAL	DPDSFRYYVI	EVNPR-VSRS
<i>S. cerevisiae</i> (Arg)	EEFHMLRSAA	IKIIRHLGV-	IGECNVQYAL	QPDGLDYRVI	EVNAR-LSRS
<i>S. cerevisiae</i> (Pyr)	EDYNMLRTTA	VNVIRHLGV-	VGECNIQYAL	NPVSKDYCII	EVNAR-LSRS
<i>D. discoideum</i> (Pyr)	REYQMLRETA	IKTVRHLGV-	IGECNIQYSL	NPYSEEYCII	EVNAR-LSRS
<i>P. falciparum</i>	YEYYKFREIA	LKVITHLNI-	IGECNIQFGI	NPQTGEYCII	EVNAR-LSRS
<i>B. bovis</i>	GELYKYREIA	FKIVRYLGS-	VGECNVQFAV	NPDTDDYFIV	ELNAR-LSRS
<i>T. cutaneum</i> (Arg)	E-YHMLRSAA	IKIVRHVGV-	VGECNVQYAL	DPNSRDYRVI	EMNAR-LSRS
<i>R. catesbieana</i> (CPSI)	EECQMLRAVS	IKVVRHLGI-	VGECNIQFAL	HPTSLEYVII	EVNAR-LSRS
<i>S. acanthias</i> (CPSIII)	EEYHMLRETA	IKVVRHLEI-	VGECNIQYAL	HPLSLEYCII	EVNAR-LSRS
Hamster (Pyr)	REYQLLRRTA	IKVTQHLGI-	VGECNVQYAL	NPESEQYII	EVNAR-LSRS
Human (CPSI)	AEFQMLRRTS	INVVRHLGI-	VGECNIQFAL	HPTSMEYCII	EVNAK-MSPN

<i>S. solfataricus</i>	SALASKATGY	PLAYVSAKLA	LGYELHEVIN	KVSG-RTCAC	FEPSLDYIVT
<i>N. gonorrhoeae</i>	SALASKATGF	PIAKVAANWA	VGFTLDELRLN	DITGRRTPAS	FEPSIDYVVT
<i>E. coli</i>	SALASKATGF	PIAKVAAKLA	VGYTLDLMN	DITGGRTPAS	FEPSIDYVVT
<i>P. aeruginosa</i>	SALASKATGF	PIAKIAAKLA	VGYTLDLQN	DITGGRTPAS	FEPIDYVVT
<i>P. stutzeri</i>	SALASKATGF	PIAKIAAKLA	VGYTLDLQN	DITGGRTPAS	FEPIDYVVT
<i>B. subtilis</i> (Arg)	SALASKATGY	PIAKMAAKLA	VGYTLDLKN	PLTG-STYAS	FEPALDYVIV
<i>B. subtilis</i> (Pyr)	SALASKATGY	PIAKLAAKIA	VGLSLDEMNM	PVTG-KTYAA	FEPALDYVVS
<i>B. caldolyticus</i> (Pyr)	SALASKATGY	PIAKLAAKIA	VGLTLDEMNM	PVTG-KTYAC	FEPALDYVVT
<i>S. cerevisiae</i> (Arg)	SALASKATGY	PLAYTAAKIG	LGYTLPELNP	PITK-TTVAN	FEPSLDYIVA
<i>S. cerevisiae</i> (Pyr)	SALASKATGY	PLAYTAAKLG	LNIPLNEVKN	SVTK-STCAC	FEPSLDYCVV
<i>D. discoideum</i> (Pyr)	SALASKATGY	PLAFISAKVA	LGYDLAALRN	TITK-KTTAC	FEPSLDYLVV
<i>P. falciparum</i>	SALASKATGY	PLAYISAKIA	LGYDLISLKN	SITK-KTTAC	FEPSLDYITT
<i>B. bovis</i>	SALASKATGY	PLAYFAARIA	LGFDLVQMRN	AITL-VTTAC	FEPSLDYIVV
<i>T. cutaneum</i> (Arg)	SALASKATGY	PLAYTAAKIA	LGHTLRELNP	AVTK-STTAC	FEPSLDYIVT
<i>R. catesbieana</i> (CPSI)	SALASKATGY	PLAFIAAKIA	LGIPLPEIKN	VVSG-KTTAC	FEPSLDYMT
<i>S. acanthias</i> (CPSIII)	SALASKATGY	PLAFVAAKLA	LGIPLPDIKN	AVSG-KTTAC	FEPSLDYIVT
Hamster (Pyr)	SALASKATGY	PLAYVAAKLA	LGIPLPELRN	SVTG-GTAA-	FEPSLDYCVV
Human (CPSI)	SALASKTTGY	PLAFIAAKIA	LGIPLPGIKN	VVSG-KTSAC	FEPSLDYMT

<i>S. solfataricus</i>	KIPRWDL SKF	ENVDQSLATE	MMSVGEVMSI	GRSFEESLQK	AIRMLDIGEP
<i>N. gonorrhoeae</i>	KMARFAFEKF	PAADDRLTQ	MKSVGEVMAM	GRTIQESFQK	ALRGLETGFV
<i>E. coli</i>	KIPRFNFEKF	AGANDRLTQ	MKSVGEVMAI	GRTQQESLQK	ALRGLEV GAT
<i>P. aeruginosa</i>	KIPRF AFEKF	PKADARLTQ	MKSVGEVMAI	GRTFQESVQK	ALRGLEV GAT
<i>P. stutzeri</i>	KIPRF AFEKF	PKADARLTQ	MKSVGEVMAI	GGPSQESVQK	ALRGLEV GAT
<i>B. subtilis</i> (Arg)	KFPRWPFDFK	KNADRKLGTK	MKATGEVMAI	ERNLEAAIQK	AGASLELKNI
<i>B. subtilis</i> (Pyr)	KIPRWPFDFK	ESANRKLGTQ	MKATGEVMAI	GRTLEESLLK	AVRSLEADVY
<i>B. caldolyticus</i> (Pyr)	KIPRFPFDFK	ESANRRLLGTQ	MKATGEVMAI	GRTFEESLLK	AVRSLEIGVH
<i>S. cerevisiae</i> (Arg)	KIPKWDL SKF	QYVDRSIGSS	MKSVGEVMAI	GRNYEEAFQK	ALRQVDPSLL
<i>S. cerevisiae</i> (Pyr)	KMPRWDLK KF	TRVSTELSSS	MKSVGEVMSI	GRTFEEAIQK	AIRSTEYANL
<i>D. discoideum</i> (Pyr)	KMPRWDLK KF	TRVSNKISSS	MKSVGEVMSI	GRKFEEAIQK	AIRMVMDGAV
<i>P. falciparum</i>	KIPRWDLNKF	EFASNTMNSS	MKSVGEVMSI	GRTFEESI QK	SLRCIDDNYL
<i>B. bovis</i>	KIPKWDLRKF	EYADNLLGSS	MKSVGEVMSI	GRTFEEAMQK	ALRM----QV
<i>T. cutaneum</i> (Arg)	KIPKWDLAKF	QHVERN VGSA	MKSVGEVMAI	GRTFEESLQK	AIRQVDPNFA
<i>R. catesbieana</i> (CPSI)	KIPRWDLDRF	HGASGLIGSS	MKSVGEVMAI	GRTFEESFQK	ALRMCHPSVD
<i>S. acanthias</i> (CPSIII)	KIPRWDLDRF	HGASREIGSS	MKSVGEVMAI	GRTFEESFQK	ALRMCHPSVE
Hamster (Pyr)	KIPRWDL SKF	LRVSTKIGSC	MKSVGEVMGI	GRSFEESFQK	ALRMVDENCV
Human (CPSI)	KIPRWDLDRF	HGTSSRIGSS	MKSVGEVMAI	GRTFEESFQK	ALRMCHPSIE

<i>S. solfataricus</i>	GVVGGKVYES	NMSKEE-ALK	YLKERRPYWF	LYAAKAFKEG	ATINEVVEVT
<i>N. gonorrhoeae</i>	----LVSMFA	RRRQTE-IRR	ELANPGPERM	LFVGKAFRAA	FTPEEIH EIS
<i>E. coli</i>	GFDPKVSLDD	PEALTK-IRR	ELKDAGADRI	WYIADAFRAG	LSVDGVFNLT
<i>P. aeruginosa</i>	GFDPKLDLND	PEADSI-LKR	ELTVPSAERV	WYVADAFRAG	KSV EEFELT
<i>P. stutzeri</i>	GFDPKLNLGD	AEAEST-LTR	ELTVPGADRI	WYVADAFRAG	KSIDDV FALT
<i>B. subtilis</i> (Arg)	GTH--LPELS	GLSIDT-LWD	LAITPDDRRF	FVVMELLSRS	VSIDDIHEKT
<i>B. subtilis</i> (Pyr)	HLE--LKDAA	DISDEL-LEK	RIKKAGDERL	FYLAEAYRRG	YTVEDLHEFS
<i>B. caldolyticus</i> (Pyr)	HLE--LNEAK	TAADDV-MEK	RIRKAGDERL	FYIAEALRRG	VTVELTHEWS
<i>S. cerevisiae</i> (Arg)	GFQGSTEFGD	QLDEALRPT	DRVLAIQQA	L-----IHEN	YTVERVNELS
<i>S. cerevisiae</i> (Pyr)	GFNET-DLDI	DIDYELNNPT	DMRVFAIANA	F-----AKKG	YSVDKVMEMT
<i>D. discoideum</i> (Pyr)	EGFQA-GVFP	TSDEELEHPT	NNRILVLASA	F-----KDG	YSIDRVHQLT
<i>P. falciparum</i>	GFSNTYCIDW	DEKKIIE---	ELKNPSPKRI	DAIHQA FHLN	MPMDKIH ELT
<i>B. bovis</i>	RVLGFNSGVM	SGADSEAITE	ALRHPTPDHV	AAIARAFELG	MTVSDIHGLT
<i>T. cutaneum</i> (Arg)	GFEAYWK-PE	DMITALTNNN	DRRLFAIAHA	M-----LNLD	YSVDYLHDLT
<i>R. catesbieana</i> (CPSI)	GFTSNLPMNK	AWSSDVNLRK	EMAETSTRM	YSMAKAIQSG	ISLDEINKLT
<i>S. acanthias</i> (CPSIII)	GFVARLPMKK	SWSDDFDLQK	DLSVPSIHRI	YSLAKALHSG	ISVDEIYDLT
Hamster (Pyr)	-GFDH-TVKP	SVSVELETPT	DKRIFVAAA	L-----WAG	YSVERLYELT
Human (CPSI)	GFTPRLPMNK	EWPSNLDLRK	ELSEPSSTRI	YATAKAIDDN	MSLDEIEKLT

<i>S. solfataricus</i>	GINEFFLNKI	KGLVDFYETL	RKLK---EID	KETLKLAKKL	GFSDEQISKA
<i>N. gonorrhoeae</i>	AIDPWFLAQI	EDLMKEEKS	SDGQ-LQDLD	YAALHRLKRR	GFSDKRLAQL
<i>E. coli</i>	NIDRWFLVQI	EELVRLEEKV	AEVG-ITGLN	ADFLRQLKRR	GFADARLAKL
<i>P. aeruginosa</i>	RIDEWFLVQI	EDLVKDEEKV	KTLG-LSSID	RELMYKLRK	GFSDARLAKL
<i>P. stutzeri</i>	KIDRWFLVQI	EDLVKEEEQV	KTLG-LSSID	RDRMYKLRK	GFSDARLAKL
<i>B. subtilis</i> (Arg)	KIDPFFLHTF	DNI IKLENRL	MEAG--SDLS	FELLKKAKEK	GFSDATIASL
<i>B. subtilis</i> (Pyr)	AIDVFFLHKL	FGIVQFEKEL	KANA--GDT-	-DVLRRAKEL	GFSDQYISRE
<i>B. caldolyticus</i> (Pyr)	QIDRFFLHKI	QNI IEMETVL	KNHP--GDL-	-DVLKKAKGL	GFSDAAIAAL
<i>S. cerevisiae</i> (Arg)	KIDKWFLYKC	MNIVNIYKEL	ESVKSLSDSL	KDLLQRAKKL	GFSDKQIAVT
<i>S. cerevisiae</i> (Pyr)	RIDKWFLNKL	HDLVQFAEKI	SSFSGTKEELP	SLVLRQAKQL	GFDDRQIARF
<i>D. discoideum</i> (Pyr)	KIDKWFLTKL	KAIIDLNLHL	STYKEPSQIP	SEILKFSKQQ	GFSDKQIARA
<i>P. falciparum</i>	HIDYWFLHKF	YNIYNLQNKL	KTLK-LEQLS	FNDLK YFKKH	GFSDKQIAHY
<i>B. bovis</i>	KIDPWFLHRL	HHLHLNLAHL	SILPSLSSFT	PAMMRYKVVY	GFSDRQISRE
<i>T. cutaneum</i> (Arg)	KIDKWFLYKL	ENIVAVHKQL	RSTP-FEQLD	KELVMTAKKT	GFSDLQIAQL
<i>R. catesbieana</i> (CPSI)	AIDKWFLYKM	QGILNMEKTL	KGSR-SESV	EETLRRAKQI	GFSDRYIGKC
<i>S. acanthias</i> (CPSIII)	AIDKWFLYRL	KQIVNLEKEL	TKQN-SETIT	DELLKAKQD	GFSDRQIGDC
Hamster (Pyr)	RIDCWFLHRM	KRIVTHAQLL	EQHR-GQPLS	QDLLHQAKCL	GFSDKQIALA
Human (CPSI)	YIDKWFLYKM	RDILNMEKTL	KGLN-SESMT	EETLKRAKEI	GFSDKQISKC

<i>S. solfataricus</i>	LN-----KS	TEYVRKIRYE	TNTIPVVKLI	DTLAGEWPAV	TNYMYLTYNG
<i>N. gonorrhoeae</i>	LN-----VS	EKEVRETRTP	VKLDPVYKPV	DTSAAEFATE	TAYLYSTY-E
<i>E. coli</i>	AG-----VR	EAEIRKL RDQ	YDLHPVYKRV	DTCAAEFATD	TAYMYSTY-E
<i>P. aeruginosa</i>	LG-----VT	EKNLRSHRHK	LKVLPVYKRV	DTCAAEFATD	TAYMYSTY-E
<i>P. stutzeri</i>	LG-----VT	EKNLRSHRHK	LKVLPVYKRV	DTCAAEFATD	TAYMYSTY-E
<i>B. subtilis</i> (Arg)	IS-----KT	EEEVRR LRKE	MGITPSFKIV	DTCAAEFDAK	TNYFYSTYFG
<i>B. subtilis</i> (Pyr)	WK-----MK	ESELYSLRQ	AGIAPVFKMV	DTCAAEFESE	TPYFYSTY--
<i>B. caldolyticus</i> (Pyr)	WN-----KT	ERDIYAVRRQ	RGIMPVYKMV	DTCAAEFTSE	TPYYYSTY--
<i>S. cerevisiae</i> (Arg)	INKHASTNIN	ELEIRSLRKT	LGIIPFVKRI	DTLAAEFPAQ	TNYLYTTTNA
<i>S. cerevisiae</i> (Pyr)	LDS-----N	EVAIRRLRKE	YGITPFVKQI	DTVAAEFPAQ	TNYLYMTYNA
<i>D. discoideum</i> (Pyr)	VGT-----T	ELNVRDYRKK	MGII PCTKHI	DTVAAEFPAQ	NNYLYMTYNG
<i>P. falciparum</i>	LSFNTSDNVT	ENDVMKYREK	LGLFPHIKVI	DTLSAEFPAL	TNYLYLTYQG
<i>B. bovis</i>	IVKST---VS	EDDVRELRSK	WGIVPFVKVI	DTMAAEYPAK	TNYCYLTYNG
<i>T. cutaneum</i> (Arg)	TGK-----T	EAEVRTL RKQ	FGVTPFVKRI	DTLAAEFPAQ	TNYLYTSYNA
<i>R. catesbieana</i> (CPSI)	LG-----LS	ETQTRELRLN	KNVKPWVKQI	DTLAAEYPAI	TNYLYLTYNG
<i>S. acanthias</i> (CPSIII)	IG-----LT	ELDARKLRIS	RNIKPQIKQI	DTLAAEYPAI	TNYLYSTYHG
Hamster (Pyr)	VLS-----T	ELAVRKL RQE	LGICPAVKQI	DTVAAEWPAQ	TNYLYLTYWG
Human (CPSI)	LG-----LT	EAQTRELRLK	KNHPWVKQI	DTLAAEYPSV	TNYLYVTYNG

<i>S. solfataricus</i>	TEDDIE-*SQ	GNKLLIIGAG	GFRIGVSVEF	DWSVXSLMEA	GSKYXDEVAV
<i>N. gonorrhoeae</i>	EECRIPS*PD	RKKVMILGAG	PNPIGQGIEF	DYCSVHAALP	LRESGFETIM
<i>E. coli</i>	EECEANP*TD	REKIMVLGGG	PNRIGQGIEF	DYCCVHASLA	LREDGYETIM
<i>P. aeruginosa</i>	EECEANP*-S	REKIMILGGG	PNRIGQGIEF	DYCCVHAALA	MREDGYETIM
<i>P. stutzeri</i>	EECEANP*-G	RDKIMILGGG	PNRIGQGIEF	DYCCVHAALA	MREDGYETIM
<i>B. subtilis</i> (Arg)	ETDGDIS*KE	KKRALIIGSG	PIRIGQGV EF	DYSAVHGVLT	LQELGFETIM
<i>B. subtilis</i> (Pyr)	EEENESV*TD	KKSVMVLGSG	PIRIGQGV EF	DYATVHSVWA	IKQAGYEAI I
<i>B. caldolyticus</i> (Pyr)	EEENESI*TE	KPSVIVLGSG	PIRIGQGV EF	DYATVHCVWA	IKQAGYEAI I
<i>S. cerevisiae</i> (Arg)	TKNDV---*FN	ENGMLVLGSG	VYRIGSSVEF	DWCAVNTAKT	LRDQGGKKTIM
<i>S. cerevisiae</i> (Pyr)	DSHDL---*FD	DHGVMVLGSG	VYRIGSSVEF	DWCAVTAVRT	LRANNIKTIM
<i>D. discoideum</i> (Pyr)	ETNDV---*IN	EKS YITLGSG	SYRIGSSVEF	DWCAVSCIRT	LRLGLKKSIM
<i>P. falciparum</i>	QEHDVLP*EK	KKCFIVLGCG	CYRIGSSVEF	DWSAIHCVKT	IRKLNHKAIL
<i>B. bovis</i>	IESDVLP*VS	ATSIVVLGCG	PYRIGSSIEF	DWVCCSCVKA	LRLSLGHA VI
<i>T. cutaneum</i> (Arg)	TTHDV---*F-	DNGTMVLGSG	VYRIGSSVEF	DWCAVTC SRA	IRDLGKKTIM
<i>R. catesbieana</i> (CPSI)	QEHD I---*FD	DHGMMVLGCG	PYHIGSSVEF	DWCAVSSIRT	LRHVCKKTTV
<i>S. acanthias</i> (CPSIII)	LEHDL---*FN	DHGIMVLGCG	PYHIGSSVEF	DWCAVSSIRT	LRQLGKKTIV
Hamster (Pyr)	NTHDL---*FR	TPHVLVLGSG	VYRIGSSVEF	DWCAVGCIQQ	LRKMGYKTIM
Human (CPSI)	QEHDV---*FD	DHGMMVLGCG	PYHIGSSVEF	DWCAVSSIRT	LRQLGKKTIV

<i>S. solfataricus</i>	LNYNPETVST	DWDIARKLYL	DEISVERVLD	LIKKEKFRY-	--VATFSGGQ
<i>N. gonorrhoeae</i>	VNCNPETVST	DFDTSRDLFY	EPLTLEDVLE	IVRTENPW--	-GVIVHYGGQ
<i>E. coli</i>	VNCNPETVST	DYDTSRDLFY	EPVTLEDVLE	IVRIEKP--	-GVIVQYGGQ
<i>P. aeruginosa</i>	VNCNPETVST	DYDTSRDLFY	EPVTLEDVLE	IVRVEQPK--	-GVIVQYGGQ
<i>P. stutzeri</i>	VNCNPETVST	DYDTSRDLFY	EPVTLEDVLE	IVRVEQPK--	-GVIVQYGGQ
<i>B. subtilis</i> (Arg)	INNNPETVST	DYEIADRIFY	EPMTTEHIVN	VREQENID--	-FAIVQFGGQ
<i>B. subtilis</i> (Pyr)	VNNNPETVST	DFSISDKLYF	EPLTIEDVMH	IIDLEQPM--	-GVVVQFGGQ
<i>B. caldolyticus</i> (Pyr)	INNNPETVST	DFSTSDKLYF	EPLTAEDVMH	VIDLEQPI--	-GVIVQFGGQ
<i>S. cerevisiae</i> (Arg)	INYNPETVST	DFDEVDRIFY	EELSYERVMD	IYELEQSE--	-GCIISVGGQ
<i>S. cerevisiae</i> (Pyr)	VNYPETVST	DYDEADRIFY	ETINLERVLD	IYEIENSS--	-GVVVSMMGGQ
<i>D. discoideum</i> (Pyr)	INFPETVST	DYDECPLYF	EELSLERVLD	IYERGGPNSN	HGVILSVGGQ
<i>P. falciparum</i>	INCNPETVST	DYDESRLIFY	DEITTEVIKF	IYNFENS--	-GVIIAFGGQ
<i>B. bovis</i>	VNCNPETVST	DYDVRDRIFY	DELTVEIVDA	IYHFENPK--	-GIVISVGGQ
<i>T. cutaneum</i> (Arg)	INYNPETVST	DFDEADRIFY	EELGFERVMD	IYDLEGAS--	-GVVVSVGGQ
<i>R. catesbieana</i> (CPSI)	VNCNPETVST	DFDECCKLYF	EELSQERIMD	VFQLEQCD--	-GCIISVGGQ
<i>S. acanthias</i> (CPSIII)	VNHNPETVST	DFDECRLIFY	EELSRLERLD	IYEQEGCK--	-GSIISVGGQ
Hamster (Pyr)	VNYPETVST	DYDMCDRLIFY	DEISFEVMD	IYELENPD--	-GVILSMGGQ
Human (CPSI)	VNCNPETVST	DFDECCKLYF	EELSRLERLD	IYHQEACG--	-GCIISVGGQ

<i>S. solfataricus</i>	IGNSIAKELE	ENGVRLLGTS	GSSVDIAENR	EKFSKLLDKL	GISQPDWISA
<i>N. gonorrhoeae</i>	TPLKLANALV	ENGVNIIGTS	ADSIDAAEDR	ERFQKVLNDL	GLRQPPNRIA
<i>E. coli</i>	TPLKLARALE	AAGVPVIGTS	PDAIDRAEDR	ERFQHAVERL	KLKQAPANATV
<i>P. aeruginosa</i>	TPLKLCRALE	EAGVPIIGTS	PDAIDRAEDR	ERFQQMVQRL	NLRQAPANATA
<i>P. stutzeri</i>	TPLKICRALE	EAGVPIIGTS	PDAIDRAEDR	ERFQQMVQRL	NLRQPQNATA
<i>B. subtilis</i> (Arg)	TAINAAEAL	KAGITLLGTS	FQTLDVLEDK	DQFYQLLDEL	GLKHAKGEIA
<i>B. subtilis</i> (Pyr)	TAINLADELS	ARGVKILGTS	LEDLDRAEDR	DKFEQALGEL	GVPQPLGKTA
<i>B. caldolyticus</i> (Pyr)	TAINLAAELE	ARGVRLGTT	LEDLDRAEDR	DKFEQALSEL	GIPKPAGKTA
<i>S. cerevisiae</i> (Arg)	LPQNIALKLY	DNGCNIMGTN	PNDIDRAENR	HKFSSILDSI	DVDQPEWSEL
<i>S. cerevisiae</i> (Pyr)	TSNNIAMTLH	RENVKILGTS	PDIDSANR	YKFSRMLDQI	GVDQPAWKE
<i>D. discoideum</i> (Pyr)	IPNNLAIPLS	RCNVKVLGTH	PDIDSANR	YKFSRLLDTI	GIDQPLWKEL
<i>P. falciparum</i>	TSNNLVFSLY	KNNVNILGSV	HKVLIVVKIG	INFRTYV-IL	KIDQPKWNKF
<i>B. bovis</i>	TANNLALQFH	SLGLPILGTS	VESIDSCEDR	YSFSEVVCSF	GHDQTCMEEF
<i>T. cutaneum</i> (Arg)	LPQNIALRLK	KAGVQVLGTD	PEMIDSAEDR	HKFSSILDSI	GVDQPAWTEA
<i>R. catesbieana</i> (CPSI)	IPNNLAVPLY	KNGVKIMGTS	PMQIDRAEDR	SIFSALVDEL	QIAQAPWKAV
<i>S. acanthias</i> (CPSIII)	IPNNLAVPLY	KYGVNILGTN	PMQIDRAEDR	AVFSALVDEL	QVSQAPWTAV
Hamster (Pyr)	LPNNMAMALH	RQCCRVLGTS	PEAIDSAENR	EKFSRLLDTI	GISQPQWREL
Human (CPSI)	IPNNLAVPLY	KNGVKIMGTS	PLQIDRAEDR	SIFSALVDEL	KVAQAPWKAV

<i>S. solfataricus</i>	TSLGEIKKFA	NEVGFPVLVR	PSYVLGSSM	KIAYSEELY	EYVRRATEIS
<i>N. gonorrhoeae</i>	HNEEEALVKA	EEIGYPLVVR	PSYVLGGPAM	QIVHSAEALQ	KYMREPVQVS
<i>E. coli</i>	TAIEMAVEKA	KEIGYPLVVR	PSYVLGGRAM	EIVYDEADLR	RYFQTAVSVS
<i>P. aeruginosa</i>	RSEDEALAAS	KAIGYPLVVR	PSYVLGGRAM	EIVYQEEELK	RYMREAVQVS
<i>P. stutzeri</i>	RSEEEAIAAS	KAIGYPLVVR	PSYVLGGRAM	EIVYDEEELK	RYMREAVQVS
<i>B. subtilis</i> (Arg)	YTKEEAASKA	SEIGYPVLIR	PSYVIGGMGM	IIVDSQAQLS	QLLNDEDSM-
<i>B. subtilis</i> (Pyr)	TSVNQAVSIA	SDIGYPVLVR	PSYVLGGRAM	EIVYHEEELL	HYMKNAVKIN
<i>B. caldolyticus</i> (Pyr)	VSVEEAVAIA	EEIGYPVLVR	PSYVLGGRAM	EIVYNRGELL	HYMEHAVRVN
<i>S. cerevisiae</i> (Arg)	TSVEEAKLFA	SKVNYPVLIR	PSYVLGAAM	SVVNNEELK	AKLTASDVS
<i>S. cerevisiae</i> (Pyr)	TSMDEAESFA	EKVGYPVLVR	PSYVLGAAM	NTVYSKNDLE	SYLNQAVEVS
<i>D. discoideum</i> (Pyr)	TSVSDTKDFC	ESVGFPCLVR	PSYVLGAAM	NVVHSSQDLE	TFLTEAAVS
<i>P. falciparum</i>	TKLSKAIQFA	NEVKFPVLVR	PSYVLGAAM	RVVNCFEELK	NFLMKAIVS
<i>B. bovis</i>	TSFEGAKQFC	TKVSFPVLVR	PSYVLGASM	RVIVSFEELE	KYLQTSAVVN
<i>T. cutaneum</i> (Arg)	SSLASAKEFA	NRVGYPVLIR	PSYVLGAAM	NVVWDEAQL	HNLTLATNVS
<i>R. catesbieana</i> (CPSI)	NSLDDALQFT	KTVGYPCLLR	PSYVLGSAM	NVVGEEELK	TFLAEATRVS
<i>S. acanthias</i> (CPSIII)	NTLDDALSFA	ETVGYPCLLR	PSYVLGSAM	NVAHDAVEMK	KFLAEAAARVS
Hamster (Pyr)	SDLESARQFC	QTVGYPCVVR	PSYVLGAAM	NVAYTDGDLE	RFLSSAAAVS
Human (CPSI)	NTLNEALEFA	KVDYPCLLR	PSYVLGSAM	NVVFSEDEM	KFLAEATRVS

<i>S. solfataricus</i>	PKYPVVISKY	TENAIEAIED	GVSDGNKVLG	IT--LEHIEE	AGVHSGDATM
<i>N. gonorrhoeae</i>	EDSPVLLDFF	LNNAAIEVDVD	CVS-DGKDVV	IGGIMQHVEQ	AGIHSGDSGC
<i>E. coli</i>	NDAPVLLDHF	LDDAVEVDVD	AIC-DGEMVL	IGGIMEHIEQ	AGVHSGDSAC
<i>P. aeruginosa</i>	NDSPVLLDHF	LNCAIEVDID	AVC-DGEIVV	IGAIMQHIEQ	AGVHSGDSAC
<i>P. stutzeri</i>	NDSPVLLDHF	LNCAIEVDVD	AVC-DGEQVV	IGGIMQHIEQ	AGVHSGDSAC
<i>B. subtilis</i> (Arg)	-PYPILIDQY	VS-GKEVEID	LIS-DGEEVF	IPTYTEHIER	AGVHSGDSFA
<i>B. subtilis</i> (Pyr)	PQHPVLIDRY	LT-GKEIEVD	AVS-DGETVV	IPGIMEHIER	AGVHSGDSIA
<i>B. caldolyticus</i> (Pyr)	PQHPVLVDRY	IT-GKEVEVD	AIA-DGETVV	IPGIMEHIER	AGVHSGDSIA
<i>S. cerevisiae</i> (Arg)	PDHPVVM SKF	IEGAQEIDVD	AVAYNGNVLV	HA-ISEHVEN	AGVHSGDASL
<i>S. cerevisiae</i> (Pyr)	RDYPVVITKY	IENAKEIEMD	AVARNGELVM	HV-VSEHVEN	AGVHSGDATL
<i>D. discoideum</i> (Pyr)	RDHPVVISKF	IQEAKEIEID	AVADNGRIVL	FA-ISEHVEN	AGVHSGDATI
<i>P. falciparum</i>	KDNPVVISKF	IENAKEIEID	CVSKNGKIIN	YA-ISEHVEN	AGVHSGDATL
<i>B. bovis</i>	REHPVVISKF	IEKANEVEVE	TVWHLGYYTE	TTPLVEHVEH	AGTHSGDATL
<i>T. cutaneum</i> (Arg)	PDHPVVISQF	IDNAQEIDVD	AVAHKGKLLV	HA-VSEHVEN	AGVHSGDATL
<i>R. catesbieana</i> (CPSI)	QEHPVVITKF	IEGAREVEMD	AVGKEGRVIS	HA-ISEHVED	AGVHSGDATL
<i>S. acanthias</i> (CPSIII)	QEHPVVITKF	IEGAREVEVD	AVSKDGKVL A	HA-ITEHVED	AGVHSGDATL
Hamster (Pyr)	KEHPVVISKF	IQEAKEIDVD	AVACDGVVSA	IA-ISEHVEN	AGVHSGDATL
Human (CPSI)	QATPVVLTKE	VEGAREVEMD	AVGKDGRVIS	HA-ISEHVED	AGVHSENATL
<i>S. solfataricus</i>	SIPFRKLS EN	NVNRMRENV L	NIARXLNIK G	PXNVQXVVK-	-----ENTPY
<i>N. gonorrhoeae</i>	SLPPYSLSEE	IQDEIRRTQK	AMAYALGLVG	LMNVQFAVQ-	-----DAVVF
<i>E. coli</i>	SLPAYTLSQE	IQDVMRQQVQ	KLAFELQVRG	LMNVQFAVK-	-----NNEVY
<i>P. aeruginosa</i>	SLPPYSLPQH	IQDEIREQVK	KMALELGVVG	LMNVQMAVQ-	-----GEDIF
<i>P. stutzeri</i>	SLPAYSV PQH	IQDEIREQVR	KMALELDVIG	LMNVQMAVQ-	-----GEDIY
<i>B. subtilis</i> (Arg)	ILGPSITSG	LQQGMKDAAQ	KIARKLSFKG	IMNIQFVID-	-----NGNIL
<i>B. subtilis</i> (Pyr)	VYPPQSLTED	IKKKIEQYTI	ALAKGLNIVG	LLNIQFVLS-	-----QGEVY
<i>B. caldolyticus</i> (Pyr)	VYPPQTL SAE	VIDKIADYTI	RLARGLHIVG	LLNIQFVVS-	-----GSDVY
<i>S. cerevisiae</i> (Arg)	VLPPQHLSD D	VKIALKD IAD	KVAKAWKITG	PFNMQI IKD-	----GEHTLK
<i>S. cerevisiae</i> (Pyr)	IVPPQDLAPE	TVDRIVVATA	KIGKALKITG	PYNIQFI AK-	-----DNEJK
<i>D. discoideum</i> (Pyr)	VCPAQDLDLA	TILKVEETAR	KIAEALNVSG	PFNIQFI AK-	-----NNEIK
<i>P. falciparum</i>	ILPAQNIYVE	THRKKIKISE	KISKSLNISG	PFNIQFICH-	-----QNEIK
<i>B. bovis</i>	ILPAQNI FVG	THRAVKKITR	EFSRYLNYDG	PFNVQYLCK-	-----NNE--
<i>T. cutaneum</i> (Arg)	VLPPFSVNQH	DLGRLKTIAE	KVAQAFQISG	PFNMQIIRKP	PTGEEDAELK
<i>R. catesbieana</i> (CPSI)	MIPTQSI SQG	AI EKVKIATK	KIATAFAISG	PFNVQFLVR-	-----GNDVL
<i>S. acanthias</i> (CPSIII)	ILPTQTISQ G	ALEKVKIATR	KISEAFEISG	PFNVQFLVK-	-----GNDVM
Hamster (Pyr)	VTPPQDITPK	TLERIKAIVH	AVGQELQVTG	PFNLQLIAK-	-----DDQLK
Human (CPSI)	MLPTQTISQ G	AIEKVKDATR	KIAKAFAISG	PFNVQFLVK-	-----GNDVL
<i>S. solfataricus</i>	IIELNLRASR	SMPFSSKAKG	INLINESMKA	IFDGLDFSED	YYEPP-----
<i>N. gonorrhoeae</i>	VLEVNPRATR	TVPFVSKATG	QRLAKVGARC	MAGISLKEQG	VEKEVVVPD--
<i>E. coli</i>	LIEVNPR AAR	TVPFVSKATG	VPLAKVAARV	MAGKSLAEQG	VTKEVI PP--
<i>P. aeruginosa</i>	VIEVNPRASR	TVPFVSKCVG	ESLAKVAARV	MAGKTLAEVG	FTQEII PP--
<i>P. stutzeri</i>	VLEVNPRASR	TVPFVSKCIG	ESLAKVAARV	MAGKTLAEVG	FTKEII PP--
<i>B. subtilis</i> (Arg)	VLEVNPRASR	TVPVVS KVMG	VPRATRLLAG	ASLKDLNPAV	QNHH-----
<i>B. subtilis</i> (Pyr)	VLEVNPRSSR	TVPFLSKITG	IPMANLATKI	ILGQKLAAFG	YTEGLQPEQQ
<i>B. caldolyticus</i> (Pyr)	VLEVNPRSSR	TVPFLSKITG	VPMANLATKA	ILGTKLAEMG	YETGVCVPRP
<i>S. cerevisiae</i> (Arg)	VIECNIRASR	SFPFVSKVLG	VNFIEIAVKA	FLGGDIVPKP	VDLMLNKK--
<i>S. cerevisiae</i> (Pyr)	VIECNVRASR	SFPFISKVVG	VNLIELATKA	IMGLPLTPYP	VEKLP-----
<i>D. discoideum</i> (Pyr)	VIECNLRCSR	SFPFVSKTLN	INFIELATKI	IIKHQYD-LP	VVNP-----
<i>P. falciparum</i>	IECNLRASR	TFPFISKALN	LNFIDLATRI	LMGYDV KPIN	ISLID-----
<i>B. bovis</i>	--ICNLRASR	TLPF-----LN	VNFIDQATRV	MVGSPARVHN	IQLMD-----
<i>T. cutaneum</i> (Arg)	VIECNLRASR	SFPFVSFVLG	HN FIDTASAA	IMDTNVPAPI	D--LMAQK--
<i>R. catesbieana</i> (CPSI)	VIECNLRASR	SFPFVSKTLG	VDFIDVATKV	MIGE KIDESS	LPTLERPVIP
<i>S. acanthias</i> (CPSIII)	VIECNLRASR	SCFPVSKTIG	VDLINVATRV	MTGESIDESS	LPSLENPIIP
Hamster (Pyr)	VIECNVRVSR	SFPFVSKTLG	VDLVALATRI	IMGEKVEPIG	---LMTG---
Human (CPSI)	VNECNLRASR	SFPVS KTLG	VDFIDVATKV	LIGENVDEKH	LPTLDHPIIP

<i>S. solfataricus</i>	SKYWAVKSAQ	FSWSQLRGAY	PFLGPEMKST	GEAASFGVTF	YDALLK
<i>N. gonorrhoeae</i>	--FYAVKEAV	FPFIFKFPQVD	TILNPEMRST	GEVMGVGRSF	GEAYYK
<i>E. coli</i>	--YYSVKEVV	LPFNKFPGVD	PLLGPENRST	GEVMGVGRSF	AEAFK
<i>P. aeruginosa</i>	--FFSVKEAV	FPPAKFPQVD	PILGPEMKST	GEVMGVGDSF	AEAFK
<i>P. stutzeri</i>	--FFSVKEAV	FPPAKFPQVD	TILGPEMKST	GEVMGVGDTF	AEAFK
<i>B. subtilis</i> (Arg)	--GVAVKFPV	FSSHAIQDQD	VKLGPENKST	GEGMCVAYDS	NSALKK
<i>B. subtilis</i> (Pyr)	--GVFVKAPV	FSFAKLRRVD	ITLGPENKST	GEVMGKDSTL	EKALYK
<i>B. caldolyticus</i> (Pyr)	--GVYVKVPV	FSFAKLRRVD	ISLGPENKST	GEVIGKDVTF	EKALYK
<i>S. cerevisiae</i> (Arg)	YDYVATKVPQ	FSFTRLGAD	PFLGVEMAST	GEVASFGRDL	IESYWT
<i>S. cerevisiae</i> (Pyr)	DDYVAVKVPQ	FSFTRLGAD	PVLGVEMAST	GEVATFGHSK	YEAYLK
<i>D. discoideum</i> (Pyr)	INYVGKVPQ	FSFIRLKGAD	PVLGVEMAST	GEVACFGNTR	EEAYVK
<i>P. falciparum</i>	LEYTAVKAPI	FSFNRLHGSD	CILGVEMKST	GEVACFGLNK	YEALLK
<i>B. bovis</i>	IDYVAVKVPV	FSFHRLSPSH	PVVGVDKST	GEVVGFGANK	YEALLK
<i>T. cutaneum</i> (Arg)	RDYVAIKVPQ	FSWTRLPGAD	PFLGVEMAST	GEVASFGKDI	YDAYWA
<i>R. catesbieana</i> (CPSI)	ADYVGKAPM	FSWPRLRGAD	PVLKCEMAST	GEVACFGQNV	YSAFLK
<i>S. acanthias</i> (CPSIII)	TEYVGKAPM	FSWPRLRDAD	PILRCMAST	GEVACFGPNI	YSAFLK
Hamster (Pyr)	SGVVGKVPQ	FSFSRLGAD	VVLGVEMTST	GEVAGFGESR	CEAYLK
Human (CPSI)	VDYVAIKAPM	FSWPRLRDAD	PILRCMAST	GEVACFGEGI	HTAFLK

FIGURE A3. Phylogenetic tree constructed from alignments of deduced amino acid sequences of amidotransferase genes/domains of CPS, using the Neighbour-joining distance matrix method. For organisms which contain two CPS enzymes, the sequence is identified by the letters “Arg” (for arginine-specific CPSII), “Pyr” (for pyrimidine-specific CPSII), “CPSI”, and “CPSIII”.

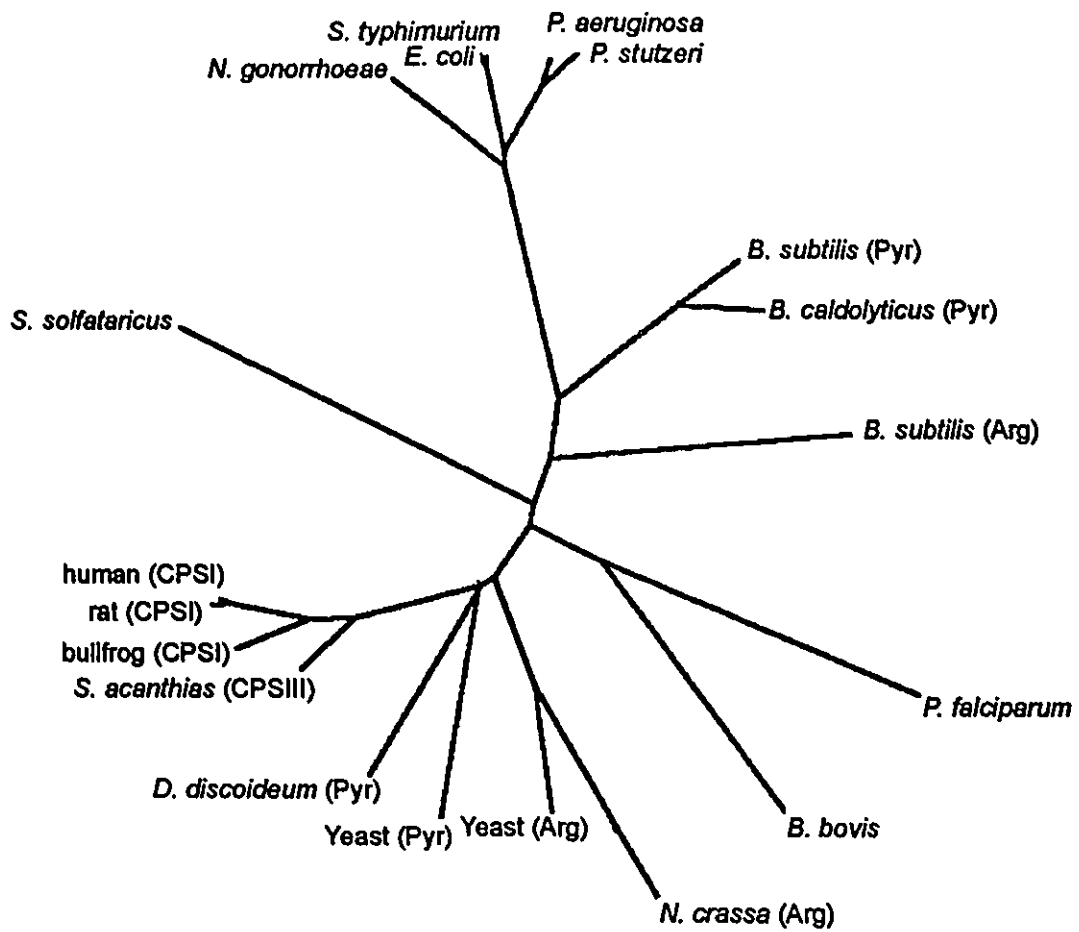


FIGURE A4. Phylogenetic tree constructed from alignments of deduced amino acid sequences of amidotransferase genes/domains of CPS, using the method of maximum parsimony. For organisms which contain two CPS enzymes, the sequence is identified by the letters “Arg” (for arginine-specific CPSII), “Pyr” (for pyrimidine-specific CPSII), “CPSI”, and “CPSIII”.

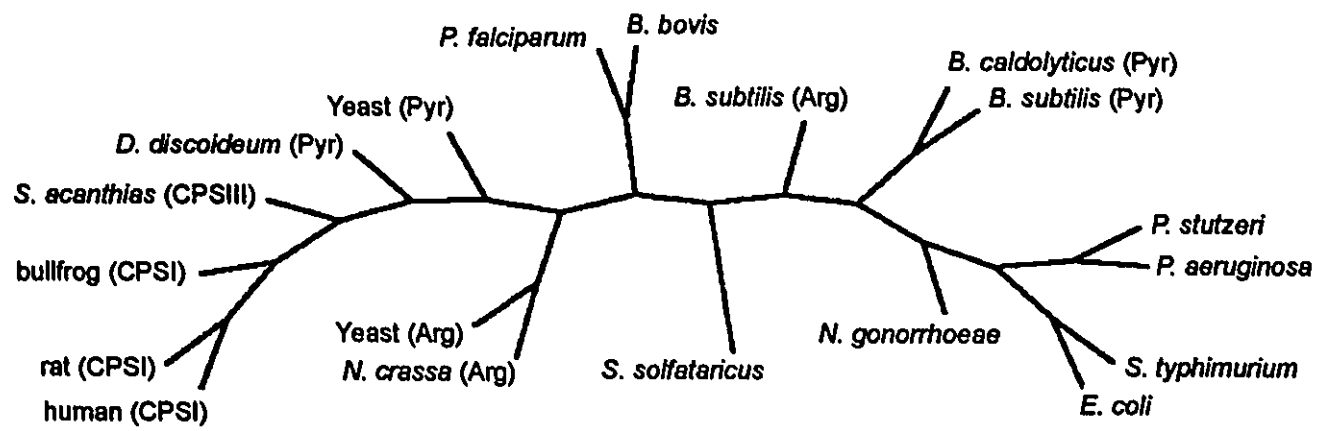


FIGURE A5. Phylogenetic tree constructed from alignments of deduced amino acid sequences of synthetase genes/domains of CPS, using the Neighbour-joining distance matrix method. For organisms which contain two CPS enzymes, the sequence is identified by the letters "Arg" (for arginine-specific CPSII), "Pyr" (for pyrimidine-specific CPSII), "CPSI", and "CPSIII".

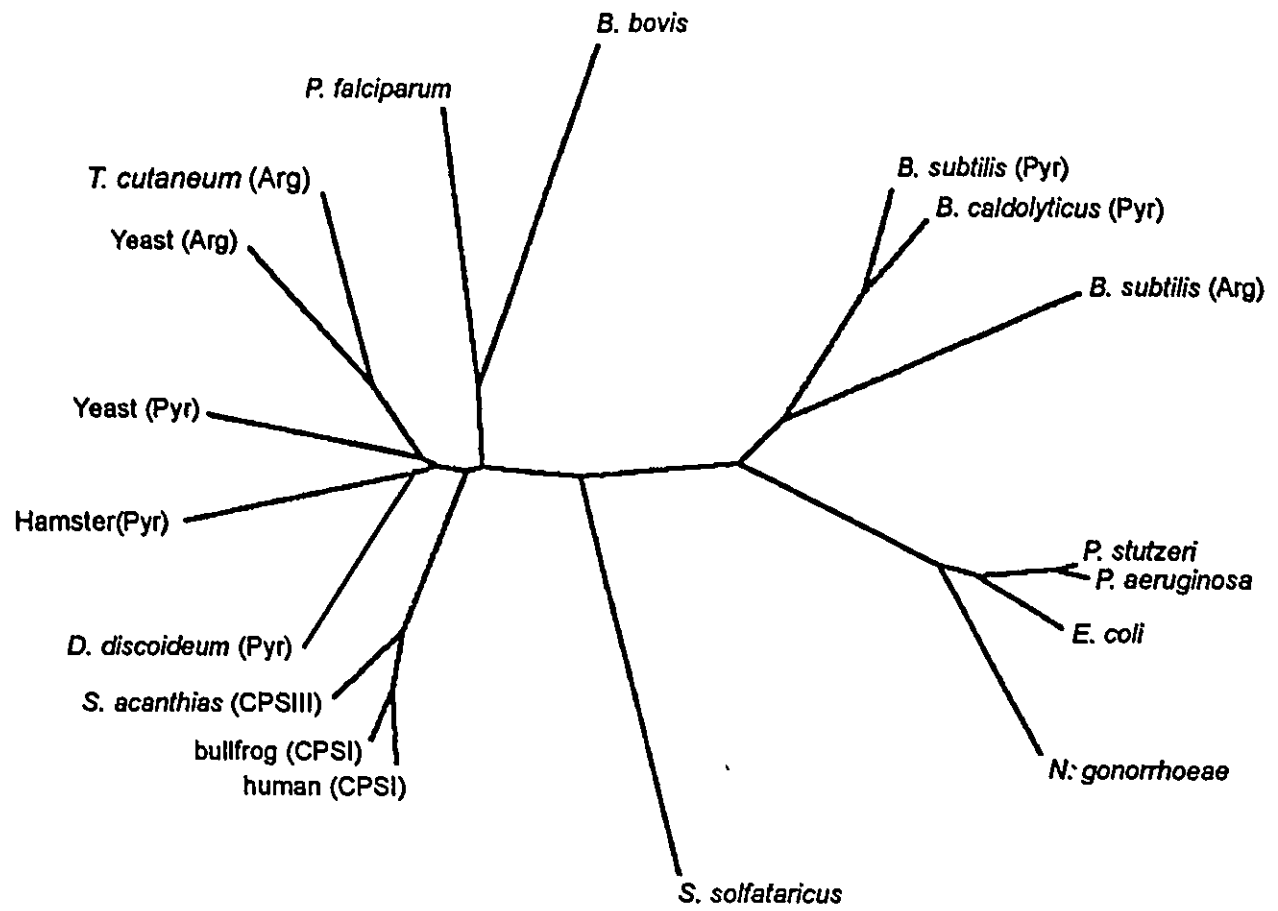


FIGURE A6. Phylogenetic tree constructed from alignments of deduced amino acid sequences of synthetase genes/domains of CPS, using the method of maximum parsimony. For organisms which contain two CPS enzymes, the sequence is identified by the letters "Arg" (for arginine-specific CPSII), "Pyr" (for pyrimidine-specific CPSII), "CPSI", and "CPSIII".

