

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/104213>

**Copyright and reuse:**

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

THE BRITISH LIBRARY

BRITISH THESIS SERVICE

TITLE

ANALYSIS OF THE PROMOTER REGION OF THE  
XENOPUS BOREALIS N-CADHERIN GENE.

AUTHOR

P M  
WEBBER

DEGREE

Ph.D

AWARDING  
BODY

Warwick University

DATE

1993

THESIS  
NUMBER

DX182708

THIS THESIS HAS BEEN MICROFILMED EXACTLY AS RECEIVED

The quality of this reproduction is dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction. Some pages may have indistinct print, especially if the original papers were poorly produced or if awarding body sent an inferior copy. If pages are missing, please contact the awarding body which granted the degree.

Previously copyrighted materials (journals articles, published texts etc.) are not filmed.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no information derived from it may be published without the author's prior written consent.

Reproduction of this thesis, other than as permitted under the United Kingdom Copyright Designs and Patents Act 1988, or under specific agreement with the copyright holder, is prohibited.

Analysis of the promoter region of the  
*Xenopus borealis* N-Cadherin gene

P.M. Webber (M.A. Cantab.)

A thesis submitted for the degree of  
Doctor of Philosophy

University of Warwick

M.R.C. Animal Molecular Genetics Group

Department of Biological Sciences

University of Warwick

Gibbet Hill Road

Coventry

CV4 7AL

England

August 1993

# NUMEROUS ORIGINALS IN COLOUR



## Table of Contents

	Page
Acknowledgements	iii
Declaration	iv
Summary of Thesis	v
Abbreviations	vi
Prologue	viii
 <u>Chapters:</u>	
I. Utility of the <i>Xenopus</i> system	1
II. Cell adhesion molecules and N-CAM	24
III. The Cadherin gene family	54
IV. Gene regulation and promoter structure	77
V. Materials and methods	98
VI. Preliminary tests and library screen	119
VII. Theoretical analysis of clones	149
VIII. Experimental analysis of transcriptional start site	177
IX. $\beta$ -galactosidase clones and micro-injection of constructs	210
X. Discussion and further experiments	271
Appendix A	282
Appendix B	290
Bibliography	298

### Acknowledgement

I would like to thank all members of DB whose advice I have sought over the past 4 years, especially for giving it so readily. In particular, I would like to thank Bob Old and Darren Smith for their sagacious comments, Surinder ... for being Surinder and, of course, Liz for her boundless energy and enthusiasm.

My thanks also to Helen James for help with the antibody library screening and to Jean Westerman in the animal house for ensuring my fingers did not become rat-fodder.

This work was supported by the Medical Research Council, my parents and Lucy.

### Declaration

I hereby declare that all the results contained herein were independently obtained by myself unless specifically stated otherwise.

I particularly acknowledge the assistance of Dr. E.A Jones who performed all the micro-injection of the DNA constructs into *Xenopus* embryos.

All information sources are acknowledged.

None of the work contained herein has been used in any previous application for any degree.

P.M.Webber

### Summary of Thesis

A *Xenopus borealis* genomic library was screened with the 5'-end of the *Xenopus laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990). Four groups of clones were isolated that differed in restriction-enzyme digestion patterns.

The sequencing of one of these clones, 3-9/4.8BS/pBS, has identified regions of DNA highly homologous to the *X.laevis* N-Cadherin gene. Accordingly, it is believed that the clone 3-9/4.8BS/pBS contains the 5'-end and promoter region of the *X.borealis* N-Cadherin gene.

A sequence analysis of this region has shown it to be GC-rich and has revealed consensus TATA-box, CCAAT-box and Sp1 binding sites. Other possible transcription-factor binding-sites have also been identified, as well as the first exon/intron boundary.

A series of promoter-deletions were fused to the bacterial  $\beta$ -galactosidase gene and micro-injected into *X.laevis* embryos. The  $\beta$ -galactosidase staining patterns of whole embryos has visually shown that 1.3Kb of genomic DNA upstream of the translational start-site is sufficient to direct neural-specific transcription of this reporter gene.

### Abbreviations

Kb	kilobases
bp(s)	base pair(s)
nt(s)	nucleotide(s)
DNA	deoxyribonucleic acid
ssDNA	single-stranded DNA
dsDNA	double-stranded DNA
dNTP	deoxynucleoside triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
rATP	adenosine triphosphate
rUTP	uridine triphosphate
RNA	ribonucleic acid
mRNA	messenger RNA
hnRNA	heteronuclear RNA
snRNA	small nuclear RNA
tRNA	transfer RNA
pBS	pBluescript
RPA	RNAse protection assay
PCR	polymerase chain reaction
PAGE	polyacrylamide gel electrophoresis

pfu	plaque forming unit
mAb	monoclonal antibody
mM	milli-molar
SDS	sodium dodecyl sulphate
SSC	salt/sodium citrate solution
CIAP	calf intestinal alkaline phosphatase
CAT	chloramphenicol acetyl-transferase
EtOH	ethanol
EDTA	ethylene diamine tetra-acetic acid
DEPC	diethyl pyrocarbonate
LMP	low melting point

## Prologue

Imagine that 1930's man is transported into the future and is presented with a microcomputer. With the aid of a cathode ray tube and probes, he may read the voltage and frequencies between defined places on the circuit board; he may even discover regular patterns of results. Through his analysis of the computer's components, he may discover some of the materials that they are constructed from. From his knowledge of electronics, he may make some inferences as to the possible routes of construction and mechanics of the workings of some parts of this device. But he would be totally ignorant of the ways that the components of the silicon chips perform their functions and of the code used in the interactions between the silicon chips.

1990's man has a similar problem. Although he has mastered the complexities of micro- and super-computers, he is faced with a biological "black-box", comprising up to  $10^{12}$  components. He has used similar electrical probes and made various measurements. He has defined different areas of the "black-box" that perform different functions; he can even stimulate defined responses from set areas. He has made some inroads into its construction. But he still has a lot to learn about the code used and the interactions between the areas of the "black-box". The "black-box" is, of course, the vertebrate brain.

The knowledge currently gained only underlines the enormity of the task in hand. These  $10^{12}$  neurons may differ, *inter alia*, in the

connections they make and receive, their signalling capacities and their responsiveness to signals, and the neurotransmitters used; and to cap it all, these characteristics are not fixed!

Faced with a problem of this magnitude, a logical place to start is where things are at their simplest - at the beginning.

The mechanics of early neural development are now well defined in a number of organisms, although knowledge of the underlying biochemical processes lags somewhat behind. The formation of the nervous system in particular has been intensely studied due to the importance of this event in the life of any organism.

Of equal importance to the proteins that mediate these processes are the control mechanisms that direct and coordinate their actions.

The aim of this thesis, therefore, is to contribute to the increasing body of knowledge on neural gene regulation, and in particular to the study of the regulation of the *X borealis* N-Cadherin gene, taking advantage of the opportunities that the *Xenopus* system affords to follow gene expression directly.

## Chapter I

### Utility of the *Xenopus* system

1.1 *Xenopus* - an experimental system for studying vertebrate development

1.2 Early *Xenopus* development

- (i) Introduction
- (ii) Embryonic induction
- (iii) Mesoderm induction

1.3 Formation of the nervous system

- (i) Introduction
- (ii) Morphological changes associated with neural induction
- (iii) Experimental assays for neural competence and signalling capacity
- (iv) Neural competence
- (v) Neural inducing signals and current models of neural induction
- (vi) Regional patterning of the neural tissues

1.4 Expression of endogenous and exogenous DNA and RNA in *Xenopus*

- (i) Expression of endogenous RNA
- (ii) Expression of exogenous micro-injected DNA

1.5 Conclusions

### 1.1 Xenopus - an experimental system for studying vertebrate development

*Drosophila* researchers have over half-a-century's knowledge of genetics to draw upon. In mouse, researchers have a close mammalian model of human development to study. Why do we need to study the South African long-clawed toad, *Xenopus laevis*?

The main advantages of *Xenopus* centre around its eggs and embryos. These are over 2000 times larger than human eggs (making them readily manipulable) and, due to their external development, they can be observed directly through all stages of development. This latter point has facilitated the production of an accurate stage series (NIEUWKOOP and FABER, 1956) and enables any artificial perturbation of the development of the embryo to be easily analysed. (All stages referred to herein are as defined by NIEUWKOOP and FABER, 1956; and are illustrated in Appendix A).

*Xenopus* are easy to maintain in the laboratory and can readily be induced (through the use of hormone injections) to produce thousands of eggs at any time of the year. These may be fertilised through natural means or artificially, through *in vitro* fertilisation, to produce synchronously developing embryos. Individual tissues are easily isolated by micro-dissection.

The large size of both eggs and oocytes facilitates the micro-injection of exogenous DNA and RNA; *Xenopus* eggs have also been used as an artificial *in vitro* translation system (HAMES and HIGGINS,

1984).

The relatively long reproductive cycle of *Xenopus* (three years), however, makes genetic studies impractical.

There exists therefore, in *Xenopus*, a system through which the processes of early vertebrate development may readily be studied.

## 1.2 Early *Xenopus* development

### 1.2(i) Introduction

It is impossible to summarise 70 years of "classical" experiments and the results from "modern" molecular biological techniques in one short chapter. Given below, therefore, is an outline of the types of techniques that have been used to study the mechanisms of amphibian development and also a summary of the current models that have been proposed to explain these events.

Traditionally, early amphibian development has been studied by following the fates of embryonic tissues that have been stained with vital dyes (VOGT, 1929). More recently, injectable lineage tracers have been used to produce fate maps of a higher resolution (DALE and SLACK, 1987). The resolution of both of these methods is inherently limited, however, due to the mixing of cells during the complex morphological changes associated with gastrulation (defined below).

The advent of modern molecular biological tools such as antibodies and gene markers now allows the identification of tissues and the differentiation-state of cells with far great accuracy than previously

achievable with general histological stains.

Extensive use is also made of micro-surgical transplantation techniques; these are discussed below.

#### 1.2(ii) Embryonic induction

A major theme underlying early amphibian development is the interaction between one (inducing) tissue and one (responding) tissue which results in the responding tissue changing its direction of differentiation (GURDON, 1987).

The origin of this concept of embryonic induction is often attributed to Spemann (1901) and Lewis (1904) who established that in certain species of *Rana* the formation of the lens from ectoderm was induced by the underlying optic lobe of the brain. Classical embryonic induction, however, was first demonstrated in the famous experiments of Spemann and Mangold (1924). In these experiments, the dorsal lip of the blastopore (the "Organiser") of the unpigmented newt *Triturus cristatus* was transplanted to the ventral region of a pigmented *Triturus taeniatulus*, where it dorsalised surrounding ventral mesoderm. This dorsalised mesoderm then induced the pigmented host cells to form a secondary axis complete with secondary nervous system. Even now, nearly 70 years later, the molecular mechanisms of this dorsalising effect and neural induction are incompletely understood.

The newly-laid frog egg is already polarised by the pigmented and unpigmented animal and vegetal poles (respectively) and also internally with respect to the yolk granules and maternal RNAs. The

dorso-ventral axis is defined by the sperm-entry point (defining ventral; Figure I.1). Thus before cleavage starts, this primary axis is established.

#### 1.2(iii) Mesoderm induction

Mesoderm induction in *Xenopus* occurs at the 32 cell stage (stage 6). As shown in Figure I.1(a), some animal pole cells at the animal/vegetal interface are induced by vegetal pole cells to form a third cell type, mesoderm.

Mesoderm induction may be followed by the identification of mesodermal derivatives such as muscle, kidney and notochord through the use of histological stains or gene markers. Due to the lack of discrete regional mesodermal markers, the anterior-posterior polarity of the mesoderm has largely been assayed through indirect means, i.e. through transplantation and observable markers in the nervous system (reviewed SLACK and TANNAHILL, 1992).

Although many issues remain to be resolved, the induction and specification of mesodermal tissues appears to be dependent upon the coordinated action of at least three signals (Figure I.2). Separate signals from the ventro-vegetal and dorso-vegetal regions specify two areas of mesoderm, termed M3 and Organiser. (The dorso-vegetal signalling region is also known as the "Nieuwkoop Centre", after its discoverer). A third signal, from the Organiser, then specifies the fate of the mesoderm (M3) still further: mesodermal cells on the dorsal side of the embryo are induced to become notochord and muscle:

Figure I.1

Mesodermal and neural induction in *Xenopus*

External view from side

Transverse sections

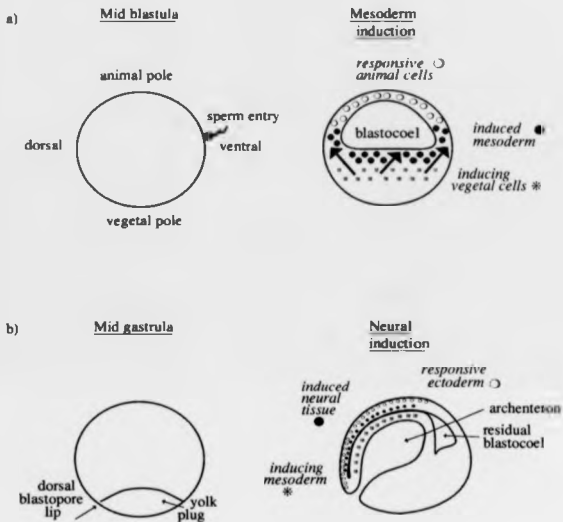
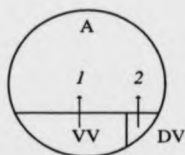
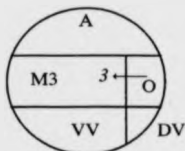


Figure 1.2

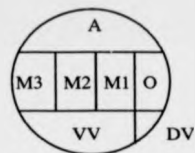
Three-signal model of mesoderm induction



1. Two signals are released from the vegetal hemisphere: a ventral-vegetal (VV) signal (1); and a dorso-vegetal (DV) signal (2).



2. The ventro-vegetal signal (1) converts the equatorial region above it to ventral mesoderm (M3). The dorso-vegetal signal induces the organiser (O).



3. The organiser then sends out a third signal (3) which converts the ventral mesoderm to different types of lateral mesoderm.

(Adapted from KIMELMAN *et al.*, 1992)

mesodermal cells on the ventral side become blood and mesenchyme.

The actions of putative signalling molecules have been tested by either incubating embryonic fragments in their presence or micro-injecting mRNAs coding for such molecules into individual blastomeres. The current candidates for signals 1. and 2. (the vegetal pole signals) are FGF (fibroblast growth factor) and activin (a member of the TGF $\beta$  family). It appears that these two molecules alone are sufficient to generate all mesodermal tissues (reviewed WOODLAND, 1993). Candidates for signal 3., which specifies the dorso-ventral patterning within the mesoderm, have also been identified by micro-injecting synthetic RNA into eggs or embryos: these include *Xwn1-8* (SMITH and HARLAND, 1991), *noggin* (SMITH and HARLAND, 1992) and Bone Morphogenetic Protein (BMP, a member of the TGF $\beta$  family, DALE *et al.*, 1992).

After approximately 9 hours of development following the formation of mesoderm, the *Xenopus* embryo undergoes complex morphogenetic movements, termed gastrulation. During this process the embryo involutes through the blastopore, converting an essentially simple embryo into a complex tri-layered structure with ectoderm covering the involuted mesoderm and endoderm. One result of this is that the newly-induced mesoderm underlies the ectoderm. This rearrangement is essential for correct development and allows the next important steps to occur.

### 1.3 Formation of the nervous system

#### 1.3(i) Introduction

As stated above, gastrulation results in mesodermal tissue underlying the external ectoderm. The mesodermal tissues that underlie the dorsal ectoderm are the notochord and the precursors to the somites (muscle-blocks).

The notochord is a stiff rod of vacuolated mesodermal cells that underlies the most dorsal region of the ectoderm. It provides support for the embryo prior to the formation of a skeleton. The somites flank the notochord on both left and right sides of the embryo (Figure 1.3(1)).

Neural induction is believed to commence at the mid-gastrula stage (stage 10). An undefined neuralising signal passes to responsive ectoderm resulting in the responsive ectoderm undertaking a neural pathway of development (Figure 1.1(b)).

#### 1.3(ii) Morphological changes associated with neural induction

Upon receipt of the neuralising signal, ectodermal cells undergo gross morphological changes that result in the formation and raising of the neural folds, dropping of the neural plate, and ultimately the pinching off and closing of the neural tube as a separate tissue between the ectoderm and mesodermal derivatives, as shown in Figure 1.3 (HAUSEN and RIEBESELL, 1991).

Two-dimensional computer models have shown that the apical restriction of cells (for example by rings of actin microfilaments, at

Figure 1.3a

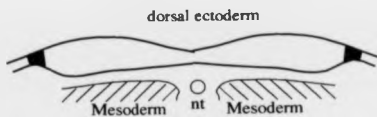
The diagram shows how, after neural induction, the ectoderm folds up to form the neural tube and neural crest cells (black). This process occurs along the length of the dorsal side of the embryo.

1. At stage 10 the neural signal first passes from the dorsal mesoderm (which includes the notochord, nt) to the overlying ectoderm. (For clarity, these mesodermal tissues are only shown in 1).
2. The extent of induced neural tissue is delineated by the neural folds (shown in black), which ultimately will become the neural crest cells (precursors to the peripheral nervous system).
3. Cell intrinsic and cell-extrinsic events conspire to bring the neural folds together.
4. The neural folds eventually meet and the epidermis fuses over the neural tube.
5. The neural tube closes and the neural crest cells migrate to their designated positions in the embryo.

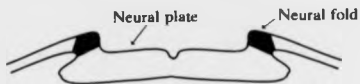
Figure 1.3b

Diagrammatic representation of neural-tube formation  
in amphibians

1.



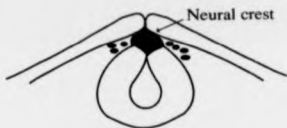
2.



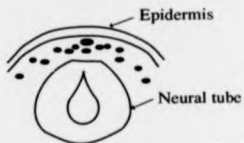
3.



4.



5.



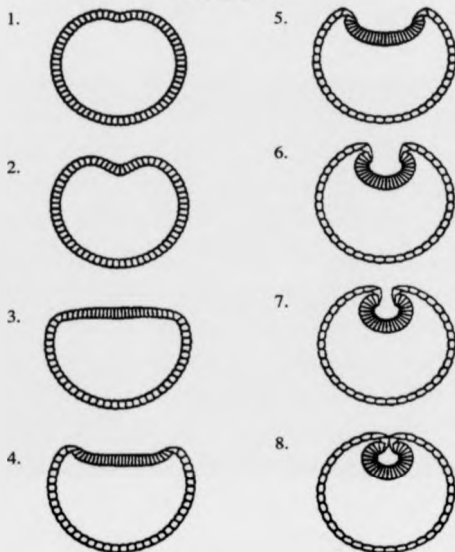
constant volume) can result in wedge-shaped cells, the alignment of which could form tubes i.e. the neural tube (Figure 1.4). Firm evidence has been provided for the involvement of actin microfilaments and microtubules in cell constriction and elongation respectively (SCHOENWOLF and POWERS, 1987; SCHOENWOLF *et al.*, 1988). In three dimensions, however, such apical restriction would result in cone-shaped cells, whose alignment would lead to the formation of vesicles. Whilst the simplicity of such models certainly has appeal, they cannot yet fully explain all the morphological events; in addition, their predictions do not fully parallel histological observations.

In chick, electron micrograph pictures have shown that the cell morphology and the cell movements that accompany neurulation are variable and dependent upon the position of the cells along the length of the neural tube. In contrast to the simple tube-formation model mentioned above, the closure of the neural tube appears to be brought about by bending of the neural plate at median and dorso-lateral hinge-points. The movement of neurepithelial cell nuclei (which are positioned at the widest point of the cells) also may play some role in neural plate bending. (For a comprehensive review see SCHOENWOLF and SMITH, 1990).

Whilst the morphological movements that accompany neural induction are now well documented, knowledge of the biochemical processes responsible for these events is still far from complete.

Figure 1.4

Computer-generated model of neural tube formation  
in amphibians



This computer model is based on simple changes in cell morphology i.e. from cuboidal to wedge-like. It can be seen that such transitions, in selected cells, can readily transform a sheet of cells into an enclosed tube.

Whilst this model holds fast in two dimensions, these simple changes are not sufficient to explain the three-dimensional cell movements. Histological observations also show more complex patterns of cell morphology.

### 1.3(iii) Experimental assays for neural competence and signalling capacity

Molecular studies on neural induction have revolved around two central issues. The first is the competence of the ectoderm, i.e. its ability to respond to a neural-inducing signal. The second is the nature of this neural-inducing signal and its effect on competent tissues. These issues have largely been studied through the use of micro-surgical techniques which are used to transplant dissected tissues to ectopic sites (PHILLIPS, 1991).

In the "einsteck" method, a tissue is transplanted into the ventral blastocoel cavity of a host embryo. The production of a second dorsal axis is used as an assay to show that the transplanted tissue has the ability to induce overlying ectoderm to differentiate into neural tissue (MANGOLD, 1933).

Sandwich cultures of ectoderm wrapped around another tissue may also be used to assay for neural competence and the presence of neural inducers emanating from the wrapped tissue.

If pre-gastrula embryos are cultured in high salt conditions, the normal involution of the mesoderm into the ventral cavity is inhibited. The resulting exo-gastrulated embryo (exogastrula) therefore provides a situation in which the mesoderm is only connected to the ectoderm via a thin isthmus of tissue. This system is particularly suited to the study of the effects of potential inducing signals propagating through the plane of the tissue from mesoderm to ectoderm.

Tissues may also be labelled with fluorescent dyes or artificially aged (i.e. cultured independently) prior to incorporation into host embryos.

The above "classical" techniques (many of which date back to the 1920's) have now been supplemented by the use of gene markers. A number of such markers are known in *Xenopus* including mAbs (NCAM - JACOBSON and RUTISHAUSER, 1986; 2G9 - JONES and WOODLAND, 1989) and cDNA probes (NCAM - KINTNER and MELTON, 1987). The identification of homeobox genes in *Xenopus* has lead to the discovery of a multitude of additional gene markers (reviewed SLACK and TANNAHILL, 1992). In common with the *Drosophila* and mouse homeobox (HOX) gene complexes, the *Xenopus* HOX genes appear to share a correlation between gene order on the chromosome and expression patterns along the neural tube (i.e. 5'-posterior, 3'-anterior). The genes that comprise these complexes therefore provide valuable markers for discrete areas of the brain and spinal cord.

### 1.3(iv) Neural competence

Results from experiments such as those described above indicate that ectoderm loses its competence to form neural tissue by mid-gastrula. It has also been found that the timing of this loss of competence varies with respect to the region of the embryo: competence is lost first in the posterior ventral regions and then subsequently in the dorsal anterior regions. Many of the known neural inducing tissues (e.g. involuted

dorsal mesoderm, blastopore lip), however, maintain their inductive power long after the ectoderm has lost its competence.

It is also accepted that heterogeneity in the neural competence of the *Xenopus* ectoderm is generated at least in part by cell autonomous differences between cleavage stage blastomeres (GALLAGHER *et al.*, 1991).

### 1.3(v) Neural inducing signals and current models of neural induction

Discussions on the subject of neural induction are dominated by a debate over whether the inductive signal is sent from the dorsal lip of the blastopore a) anteriorly in a "cis" fashion through the plane of ectodermal cells; or b) in "trans" fashion from the dorsal mesoderm up to the overlying ectoderm (reviewed GILBERT and SAXEN, 1993). Evidence for both mechanisms has been provided; both pathways probably play some role in neural induction.

In favour of the "trans" pathway, Gerhart *et al.* (1989) have shown that when mesodermal invagination is inhibited, no dorsal axis is formed. Hemmati-Brivanlou *et al.* (1990) have also demonstrated that anterior notochord (a mesoderm derivative) is needed for the induction of the anterior-most neural markers.

The list of molecules known to possess these neuralising qualities includes: oleic, linoleic and nucleic acids; ether extracts of adult newt; natural and artificial steroid hormones; and dead embryonic intestine and epidermis! The lack of any unifying characteristics within this group suggests that a more general effect, such as a distinct change in

pH, may be responsible for the neuralising signal. Variations in the dorsal and ventral expression of two isoforms of protein kinase C have been reported (OTTE *et al.*, 1991); these may be responsible for differentially regulating the cAMP-pathway in these two regions.

In favour of the "cis" pathway, experiments by Dixon and Kintner (1989) have shown that neural specific gene markers are expressed in *Xenopus* exogastrulae (where the dorsal mesoderm has not involuted to lie under the ectoderm). Other neural-specific gene markers have been shown to be expressed in the ectoderm of sandwich assays in which dorsal mesoderm is prevented from contacting ectoderm (DONIACH *et al.*, 1992).

Neural induction, therefore, appears to be a multi-stage process involving cell autonomous events, cis-active signals from the dorsal lip of the blastopore and trans-active signals from the dorsal mesoderm.

Candidate genes that possess potential characteristics of the Organiser signal (i.e correct spatial and temporal localisation) include *gooseoid* (a homeobox-containing gene that resembles *Drosophila bicoid* and *gooseberry*; CHO *et al.*, 1991), *Xenopus forkhead* (DIRKSEN and JAMRICH, 1992), *Xlim-1* (TAIRA *et al.*, 1992) and *Pintallaris* (RUIZ I ALTABA and JESSELL, 1992). It is most likely that no single gene product specifies the Organiser signal.

#### 1.3(vi) Regional patterning of the neural tissues

In the absence of further inductive signals, neural tissue adopts an

anterior specification. Transplantation experiments of the type described above, however, have provided some evidence for a further signal that arises from the posterior mesoderm. The gradient of this signal along the anterior-posterior axis provides positional information for the receptive neural tissues that facilitates the regionalisation of the neural tissue into forebrain, hindbrain and spinal cord.

From parallels in the *Drosophila* system, it has been suggested that this anterior-posterior axis is achieved by the expression not of a single signal but by a combinatorial code homeo-box-containing genes.

The role of retinoic acid in the establishment of anterior-posterior polarity has also been much investigated: although the effects of its misexpression are legion, its precise role in normal embryos, if any, remains unclear.

Classical and contemporary models appear to be converging on a two-gradient model in which an initial induction step establishes anterior neural structures in the dorsal ectoderm, followed by the imposition of a gradient from the posterior mesoderm that facilitates regional patterning of the neural tissues along the anterior-posterior axis.

#### 1.4 Expression of endogenous and exogenous DNA and RNA in *Xenopus*

##### 1.4(i) Expression of endogenous RNA

The ease with which *Xenopus* eggs/embryos may be micro-injected

with exogenous RNA (together with the other attributes of this system mentioned in Section I.1) has meant that many workers have studied the expression patterns of endogenous RNAs in this vertebrate (VIZE *et al.*, 1991). The fate of exogenously introduced DNA/RNA, however, is less well understood.

Transcription of the endogenous *Xenopus* genome is known to commence only after 12 rounds of cell division at the so-called Mid-Blastula Transition (MBT, stage 8.5). Before this stage, it was widely held that no transcription occurred; and that translation of stored maternal RNAs and stable proteins supplied the embryos requirements. Isotopic labelling experiments, amongst others, have now demonstrated that some synthesis of heterogeneous mRNA-like RNA does occur from the early cleavage stages (SHIOKAWA, 1991). It has now been proposed that RNA synthesis in *Xenopus* follows at least three phases, defined by the RNA polymerases involved (Figure I.5).

#### 1.4(ii) Expression of exogenous (micro-injected) DNA

There remains much debate in the literature over the degree of expression to be expected from different forms of micro-injected DNA (i.e. linear, circular, or concatenated), and also the degree of tissue-specificity achievable.

There is some consensus that circular-injected DNA is not well replicated and is rapidly lost by dilution; and that during cell division linear-injected DNA concatamerises to form complexes that are

Figure 1.5a

Changes in approximate rates of RNA synthesis per cell  
in *Xenopus* embryos

The graph shows the change in relative RNA synthesis rates in the phases pre-MBT, MBT and post-MBT (MBT, mid-blastula transition, stage 8.5).

These three phases are characterised by predominant activities of RNA polymerases II, III and I respectively.

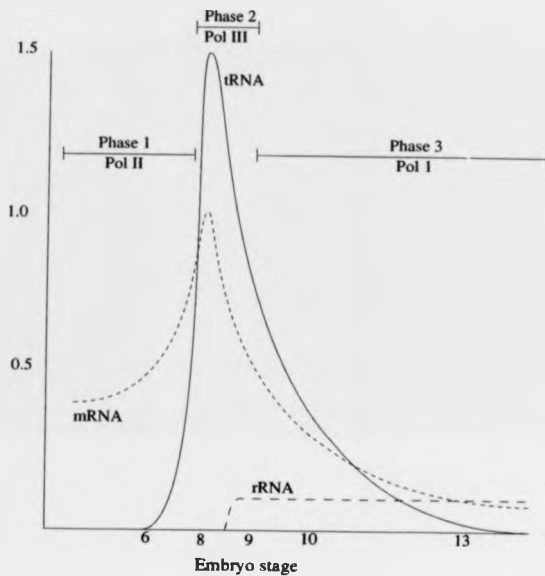
The Y-scale units are approximate rate of synthesis (pg/cell/hour) and are based upon radioactivity incorporated into the different RNA species.

"mRNA" represents heterogeneous mRNA-like RNA

"tRNA" represents 4S RNA

"rRNA" represents ribosomal RNA

Figure 1.5b



(Based on SHIOKAWA, 1991)

extensively replicated by the blastula stage (FU *et al.*, 1989, WILSON *et al.*, 1986). These concatamerised complexes gradually disappear by the tadpole stage, leaving a residual level of expression from DNA transcripts that are believed to have stably integrated into the *Xenopus* genome (estimated to be 5-10%, ETKIN and PEARMAN, 1987).

A number of anomalous results have been reported that highlight the unpredictability of these experiments. Shiokawa *et al.* (1990) report the micro-injection of an actin-chloramphenicol acetyl transferase (CAT) clone which was expressed before MBT in linear form, but only (correctly) after neurula stage when in circular form. Fu *et al.* (1989) have reported that a linearised CAT gene was expressed in early embryos irrespective of whether a (viral) promoter sequence was attached or not.

Given the long history of experiments on amphibia, it is surprising that it was only relatively recently that examples of the **correct** developmental regulation of exogenous DNA have been reported. Krieg has shown correct temporal expression of injected GS17, a gene expressed between the mid-blastula transition and the mid-gastrula stage (KRIEG and MELTON, 1985). The correct localization of a *X.borealis* cardiac actin gene product has also been reported (WILSON *et al.*, 1986). In general, however, tissue-specific gene expression of micro-injected DNA/RNA appears to be rare.

In conclusion therefore, although the expression of some exogenous DNA prior to MBT has been shown, the patterns and fate of circular

and linear DNAs largely defined, and correct regulation of some exogenous DNAs demonstrated, some doubts still remain about the fate of these exogenously introduced DNAs. It appears that such experiments should therefore be approached with careful controls and interpreted with caution.

### 1.5 Conclusions

The *Xenopus* system and its close amphibian relatives have now been the subject of research for the best part of a century. Over that time, the morphological changes that occur during early development have been accurately documented using staining histological techniques. Molecular biological techniques are now available to confirm, contradict or extend the previous hypotheses. The sheer complexity of these developmental mechanisms, however, makes unravelling the underlying biochemical processes a slow and difficult endeavour.

One group of molecules that are thought to play a major role in tissue and organ formation are members of the cell adhesion molecule (CAM) families. These proteins are discussed in Chapter II and III; their regulation is discussed in Chapter IV.

## Chapter II

### Neural Cell-Adhesion Molecule, NCAM.

#### 2.1 Introduction

- (i) History of cell adhesion molecules
- (ii) Early experiments
- (iii) Cell adhesion molecules classes

#### 2.2 Isolation of NCAM protein

#### 2.3 Structure of NCAM protein

- (i) Immunological studies
- (ii) Glycosylation pattern of NCAM
- (iii) Post-translational modifications of NCAM protein

#### 2.4 Expression of the NCAM protein

#### 2.5 Induction experiments

#### 2.6 Expression sequences of NCAM protein

#### 2.7 Structural organisation of NCAM isoforms

- (i) Splicing of NCAM RNA
- (ii) Exon usage
- (iii) Conservation of intron/exon boundaries
- (iv) *Notch* homology in the cytoplasmic domain
- (v) The muscle-specific exon, MSD1
- (vi) A secreted form of NCAM
- (vii) Differential splicing mechanisms
- (viii) Transcriptional termination

2.8 NCAM expression in vertebrate muscle

2.9 Muscle development in *Xenopus*

2.10 Conclusions

## 2.1 Introduction

### 2.1(i) History of cell adhesion molecules

One of the striking characteristics of multi-cellular organisms is the high degree of order shown by the organism's constituent cells. After the initial rapid cell divisions following fertilisation, the differentiation of embryonic cells produces a wide variety of cell types. A high degree of organization is therefore involved to ensure that this initially-heterogeneous group of cells come together in the well defined and ordered manner that goes to make a functional organism.

But in spite of the high level of cellular complexity exhibited by the multi-cellular organisms, it would not be necessarily true to assume that equally complex processes are necessary to achieve this end.

### 2.1(ii) Early experiments

In the early years of this century Wilson (1907) uncovered the ability of spongi to autoassemble themselves. He described how separate species of spongi would reconstitute the individual species if the cells were artificially dissociated and then remixed. Despite this early start, and extensive study since, the biochemical nature of this ability is not yet fully understood.

In the middle of this century Holtfreter (1939) pioneered the study of self-assembling cells, through his work on amphibia. He found that dissociated cells from embryonic amphibia had the ability to differentially segregate and form tissue-like structures.

He reasoned that in a dissociated mass, cells would collide through

random motion but only cells of the same type would form productive and lasting bonds. He (correctly) attributed this ability of like-cells to adhere to one another to a selective adhesive property of the outer surface of cells.

It has taken until the end of this century for scientists to be within reach of understanding this mechanism of selective cell adhesion.

The current model of selective cell adhesion utilises the (relatively recent) concepts of differential gene expression, protein modification and protein targeting to that results in a defined array of glycoproteins in the outer cell membrane. It is through these glycoproteins that a cell 'advertises' its cellular phenotype and hence its readiness to form productive inter-cellular contacts.

#### 2.1(iii) Cell adhesion molecule classes

A number of cell adhesion molecules (CAMs) have now been identified that fall into two main classes, the classes being defined by their mechanisms of action: a calcium-dependent mechanism; and a calcium-independent mechanism.

The calcium-independent class consists of the Immunoglobulin (Ig) superfamily and the Integrins; the calcium-dependent class is characterised by the Cadherins.

The Neural Cell Adhesion Molecule, NCAM, is a well-studied example of a calcium-independent CAM. The discovery of this molecule has provided a testing-ground for many of the speculative theories that have risen since the discovery of CAMs. It remains the

ground mark with which other CAMs are compared. This protein and the corresponding DNA sequence is discussed in Chapter II.

The Cadherins are a rapidly expanding group of calcium-dependent transmembrane CAMs that share particularly high homologies in their cytoplasmic domains. The Cadherin family is discussed in Chapter III, with particular emphasis on the neural cadherin, N-Cadherin.

## 2.2 Isolation of NCAM protein

The first discovered, and currently best-characterized, cell adhesion molecule is the Neural Cell-Adhesion Molecule, NCAM. NCAM expression is mainly restricted to neural and muscle tissues, where it may play a part in neurogenesis and myogenesis. The single NCAM gene comprises 19 exons (in chick) and spans 120kb (in mouse), and is differentially spliced to produce a variety of different sizes and forms of the protein including secretory and trans-membrane isoforms. The extra-cellular domain of the protein is characterized by vast sugar complexes that appear to play some role in regulating cell-cell binding.

The NCAM protein is a good example of the way gene-regulation and post-translational modification are efficiently used to produce a variety of different molecules from a single linear sequence of DNA. The results obtained from the study of NCAM have set an important precedent for what we can expect from studies of other cell-adhesion molecules.

NCAM was initially isolated from chick neural tissue (THIERY *et al.*, 1977), hence its name. Antibodies to chick brain NCAM were originally produced using an antibody neutralization assay, a technique first applied to the study of slime mould adhesion (GERISCH and MALCHOW, 1976). The basic technique is outlined below in Figure II.1.

NCAM specific monoclonal antibodies (mAbs) were produced from the small amount of antigen identified in the binding assay; the mAbs were then used to isolate larger amounts of NCAM antigen using immuno-affinity columns. Since NCAM protein accounts for approximately 1% of total chick brain cell-surface protein such affinity column procedures have been able to produce large amounts of the protein. This has enabled NCAM's properties to be studied in many ligand-binding assays and has also facilitated the biochemical analysis of the protein itself.

## 2.3 Structure of the NCAM protein

### 2.3(i) Immunological studies

Western blots using monoclonal and polyclonal antibodies reveal a complex array of species- and tissue-specific bands that reflect the protein's numerous forms and glycosylation states. Removal of sugar residues from brain-derived proteins, with endoglycosidases or neuraminidases, clarifies the situation to reveal three main protein isoforms of approximately 180, 140 and 120KDa, as illustrated in Figure II.2. A similar pattern of isoforms is produced from muscle.

Figure II.1a

Identification of anti-NCAM antibodies by

Antibody-Neutralization assay

- (1) Antisera were raised, in rabbits, against chick retinal cells; (2) the chick retinal proteins were also fractionated on polyacrylamide gels.
- (3) Addition of the antisera to cultures of chick retinal cells resulted in the disaggregation of these cells due to blocking of the binding sites by antibodies.
- (4) If the correct protein fractions from the chick retinal cells were added (i.e. the fraction containing the cell-adhesion molecule that the antisera was against), these bound to the antisera thus allowing the cell-bound adhesion molecules to initiate cell aggregation once more.

Once the gel fraction containing the cell-adhesion molecule had been defined, this fraction was first used to produce a polyclonal sera of greater specificity and subsequently to produce monoclonal antibodies.

In this way, the chick NCAM protein was identified.

Figure II.1b

Identification of anti-NCAM antibodies by Antibody-Neutralization assay

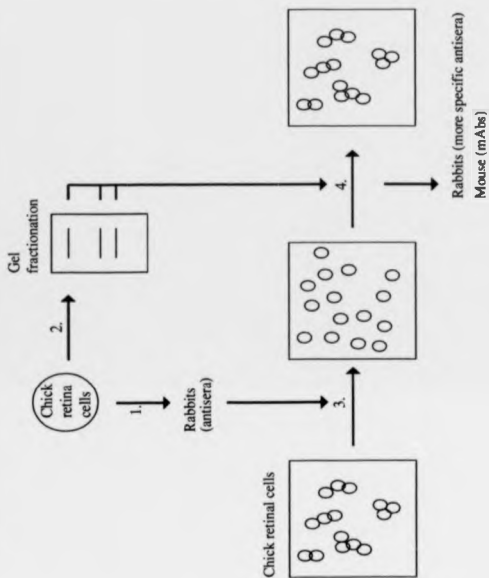
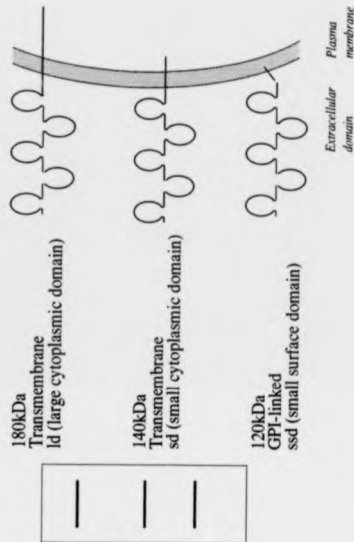


Figure II.2

Structure of the three main NCAM protein isoforms.



The above figure illustrates the three main NCAM isoforms that are resolved by SDS-PAGE and their predicted cellular positions in relation to the plasma membrane.

The complete DNA coding sequence and amino-acid sequence of NCAM are now known (Table II.1). A combination of biochemical studies and amino-acid predictions based on DNA sequence produces the following picture:

### 2.3(ii) Glycosylation pattern of NCAM

As described below, one level of diversity in NCAM protein is achieved by differential splicing of exons to form the 3 main isoforms. The mechanisms underlying the choice of isoform are not understood. The next level of diversity, that of the degree of glycosylation, has attracted much research and some basic principles do appear to be emerging.

All NCAM proteins are characterised by large sugar complexes based on  $\alpha$ -2,8-linked neuraminic acid units (polysialic acid or PSA). These negatively-charged sugar complexes vary from less than 10% (wt/wt) to up to 30% and are attached to asparagine-linked oligosaccharides in the extra-cellular domain. There appears to be a general trend with respect to the amount of PSA present on NCAM: embryonic cells have more PSA than adult cells, the so called "E-A conversion". But it is the volume that these complexes take up that is possibly the most important factor: HPLC gel filtration experiments have produced values for the excluded volume at greater than 670kDa! These sugar complexes are not thought to be involved in binding but it can readily be appreciated that the presence of such large macromolecules on the surfaces of cells may have a profound steric

Table II.1

Comparison of protein and cDNA sizes for the main NCAM isoforms

<u>Form/Species</u>	<u>Protein (kDa)/cDNA (kbp)</u>				
	<u>Chick</u>	<u>Mouse</u>	<u>Human</u> (brain) (muscle)		<u>Xenopus</u>
<u>Transmembrane</u>	180/7.0	180/7.4	180	145	180
<u>Transmembrane</u>	140/6.2	140/6.7	180		140
<u>GPI-linked</u>	120/6.0	120/5.2	120	155	120
	120/4.2	120/2.9		125	
<u>No. of exons</u>	19	7			
<u>Chromosome no.</u>		9	11	11	

Protein sizes are de-sialylated forms; GPI=glyco-phosphotidyl inositol

References: Chick - MURRAY *et al.*, (1986), HEMPERLY *et al.*, (1986); Mouse - SMALL *et al.*, (1987), BARBAS *et al.*, (1988), D'EUSTACHIO *et al.*, (1985); Human - (brain) NGUYEN *et al.*, (1986), GOWER *et al.*, (1988), (muscle) COVAULT and SANES, (1986) BARTON *et al.*, (1988); *Xenopus* - KINTNER (1987)

inhibitory effect on the ability of cell-adhesion molecules to initiate adhesion.

Experiments using membrane vesicles displaying NCAM on their surface showed that such vesicles aggregated together to a greater extent if the PSA was removed (SADOUL *et al.*, 1983). Such experiments provoked simple theories in which differential expression of low-PSA NCAM or removal of PSA from NCAM would promote binding.

Subsequent experiments have complicated the issue by showing that NCAM is only one of the molecules that plays a part in cell-cell and cell-substratum adhesion, but its influence via its PSA moiety may be significant.

F11 rat sensory ganglion/mouse neuroblastoma hybrid cells expressing NCAM and L1 were used in cell-substrate attachment and cell-aggregation assays (ACHESON *et al.*, 1991). These cells bind to laminin substrates and aggregate together weakly. Removal of PSA with endo-N results in an increased rate of cell attachment to a laminin substrate and cell aggregation.

In a cell aggregation assay, the addition of anti-NCAM antibodies to F11 cells prevented the previously weak cell aggregation; after PSA was removed from the cells, the aggregation rate was much higher. Aggregation could be substantially reduced, however, by the addition of anti-L1 antibodies. These experiments demonstrate that it is NCAM-PSA that is inhibiting the mainly L1-mediated cell adhesion.

They also illuminate a possible regulatory mechanism whereby changes in NCAM PSA content can have a profound effect on the ability of other ligands to successfully initiate cell-cell or cell-substratum adhesion. In *in vivo* terms, small changes in the relative adhesiveness of the neurite bundles with respect to the substratum could be reflected in nerve cells changing from a fasciculated pattern of growth to growth away from the neurite bundle.

A "threshold effect", in which small increases in the amount of NCAM expressed can produce disproportionately large increases in cell-cell adhesion has been reported (DOHERTY *et al.*, 1990, 1991).

#### 2.3(iii) Post-translational modification of the NCAM protein

In addition to the glycosylation described above, NCAM is also known to be phosphorylated in the carboxy-terminus and to possess asparagine-linked carbohydrates (SORKIN *et al.*, 1984).

#### 2.4 Expression of the NCAM protein

The importance of *Xenopus* as a biological tool for the elucidation of inductive and morphogenetic processes in vertebrates has already been mentioned in Chapter I. The vast body of knowledge accumulated on mesoderm and neural induction, gastrulation, myogenesis and neurulation alone make *Xenopus* the organism of choice for the studying of a potentially important morphogenetic orchestrator such as NCAM.

This chapter continues with a brief analysis of the major experiments on the expression of NCAM (with particular relevance to *Xenopus*),

trends that appear to be emerging with respect to NCAM function and how the initial theories are standing up in the light of an ever-increasing body of experimental data. Discussion of the promoter sequence and regulation of NCAM is included in Chapter IV. The reader is referred to the papers cited below for more specific details.

As mentioned above, NCAM was initially isolated and studied in chick (THIERY *et al.*, 1977). In early immuno-histological studies NCAM was first found at the blastoderm stage before becoming restricted to the neuroectoderm with a brief expression on mesodermal derivatives (THIERY *et al.*, 1982). Subsequent studies placed NCAM on all three primary germ layers at one time or another (CROSSIN *et al.*, 1985). The transient appearance and disappearance of NCAM from premigratory neural crest cells lead various groups to postulate theories regarding the possible major role for NCAM in coordinating neural development. These theories are discussed below.

In *Xenopus*, immuno-histological studies have placed the first appearance of the NCAM protein at neural plate stages (14/15) on the neuroectoderm, underlying chordamesoderm and adjacent somitic mesoderm (BALAK *et al.*, 1987). Expression persists on all neuroectoderm cells throughout neurulation, subsequently becoming concentrated on the dorso- and ventro-lateral margins of the spinal cord.

The notocord expresses NCAM between stages 14 and 18 during which time the neural groove forms. NCAM expression ceases before

the notocord cells reorientate and become vacuolar (stage 20-24).

At the neural plate stage the somitic mesoderm weakly expresses NCAM as do the developing somites. The somites cease expressing NCAM after formation except where they are innervated at neuromuscular junctions; these junctions subsequently become strongly NCAM expressing.

As in chick, neural crest cells follow a sequence of positive/negative/positive expression, the negative stage being the migratory stage in between lying adjacent to the neural tube and forming cranial/spinal ganglia.

NCAM expression has also been investigated in a number of "classical" molecular/neurophysiological systems: agar spikes doped with NCAM and inserted into *X.laevis* embryo optic tectum have been shown to disrupt the pattern and precision of the retino-tectal projection (FRASER *et al.*, 1984, FRASER *et al.*, 1988); in the developing chick hindlimb, NCAM expression patterns on nerve- and muscle-tissue has been reported (TOSNEY *et al.*, 1986); NCAM antibodies have been shown to inhibit neurite outgrowth from chick retinal ganglion cells (DRAZBA and LEMMON, 1990); a number of unique NCAM isoforms have been isolated from the frog, *R. catesbeiana*, in a study of its olfactory pathways (KEY and AKESON, 1991); NCAM appears to be only one of the several cell adhesion molecules that are required to mediate neuron-myotube adhesion (BIXBY *et al.*, 1987); and unusual NCAM isoforms have

been reported to be present in the mouse neural tube-defective *Spotch* mutant (MOASE and TRASLER, 1991).

## 2.5 Induction experiments

As described in Chapter I, the process of nervous system formation is initiated by an inductive interaction between mesoderm and ectoderm. The classical neural induction protocol of Spemann and Mangold in which the dorsal lip of a blastopore is excised and combined with excised animal cap tissue was also described. In this procedure the dorsal lip of the blastopore exerts its neuralising influence directly on the animal cap tissue producing tissues that have been identified as neural by histological means.

A number of groups have attempted to define the part, if any, that NCAM plays in this inductive process (JACOBSON and RUTISHAUSER, 1986, KINTNER and MELTON, 1987).

In order to be able to visualize an inductive process, it is necessary to have knowledge of a marker that distinguishes the induced cells from the rest of the population. Previous to NCAM, only markers that disappeared from ectoderm cell upon neural induction were known (AKERS *et al.*, 1986; JONAS, 1989)); no positive markers had been identified.

In a study using polyclonal antibodies against purified *X.laevis* NCAM, Jacobson showed that combining stage 10 animal caps and dorsal blastopore lips for more than 18 hours resulted in strong NCAM expression in the animal cap tissue; culturing either tissue

alone produced a consistently negative result. By isolating animal caps at progressively later stages he determined the stage at which animal cap was committed to becoming NCAM-positive; this was between stages 10.5 and 10.75.

Jacobson combined the facts that NCAM expression only occurred as a result of the inductive process and that the inductive process was known to produce neural tissue; his (not unreasonable) conclusion was that NCAM expression was a relatively early indicator of neural induction.

It must be remembered that Balak did not detect NCAM in *X.laevis* embryos until stage 14 and Jacobson's tissue combinations did not express NCAM until 18 hours after stage 10; by this time neural induction has already occurred. This rules out the possibility that NCAM is involved in the mechanism of neural induction.

Kintner, having isolated the cDNA for *X.laevis* NCAM, has used RNA probes to analyse the distribution of NCAM RNA quantitatively (using an RNase protection assay) and histologically (by *in situ* hybridization). His conclusions are consistent with those of Balak: he first detected NCAM mRNA at stage 10-12, a few hours before the protein. One surprising result was his failure to detect NCAM expression in the myotomes; in birds and mammals NCAM is expressed in embryonic muscle (KINTNER and MELTON, 1987, and references therein).

Kintner also described the expression of NCAM in exogastrulae

(embryos which have been stripped of their vitelline membranes and cultured in high salt, see Chapter I). These aberrant embryos, a result of incorrect gastrulation, consist of an ectodermal sack connected through a stalk of tissue to mesoderm and endoderm. Since the ectoderm does not underlie the mesoderm, it is not induced into neural tissue. In accordance with the above experiments, NCAM was only found to be expressed in the immediate vicinity of the mesoderm/ectoderm junction.

Given that NCAM gene expression follows very closely after neural induction and may be a direct consequence of it, the obvious next question is this: are the morphogenetic movements of the neural tissues regulated by NCAM expression?

This question was addressed by Kintner (KINTNER, 1988). *In vitro* transcribed NCAM RNA was microinjected into *Xenopus* embryos, resulting in high expression of NCAM on both induced and non-induced ectodermal cells. If correct NCAM expression is necessary for correct neural tube formation then aberrant NCAM expression would be predicted to produce abnormal nervous systems. Alternatively, if NCAM is not playing a regulatory role in neural development then aberrant expression of NCAM should not significantly affect neurogenesis. The latter scenario was found to be true: although some defects in the epidermis and somitic mesoderm were reported, neural tube development was essentially normal.

## 2.6 Expression sequences of NCAM

Edelman has been the main protagonist of theories relating to the combinatorial expression of CAMs and the major effects that they mediate. His group have painstakingly analysed and catalogued the expression of both NCAM and LCAM in chick (CROSSIN *et al.*, 1985) and subsequently in *Xenopus* (LEVI *et al.*, 1987) by immunological means. Their conclusions may (but see below) have profound effects on our understanding of how morphogenetic movements of tissues are effected.

His main finding was that boundaries between different tissues (particularly just after differentiation) were marked by expression of different combinations of CAMs.

The principle is illustrated here with reference to the otic and optic vesicles. These structures are formed by an inductive mechanism not dissimilar to that of neural induction: ectoderm is induced by the underlying tissue to thicken and then invaginate to produce the respective vesicles. Edelman found that whereas initially the ectoderm expresses both N- and LCAM, upon induction NCAM is lost from the invaginating ectoderm and becomes wholly localized to the ganglia that innervate the nascent vesicle. LCAM expression remains on the epithelia only.

In another example, the neural crest cells, that initially lie adjacent to the developing neural tube, lose their NCAM before migrating to various locations to form part of the PNS. Once at their final destination, NCAM is again expressed.

Edelman postulated a set of rules that seem to be followed by a number primary and secondary inductive processes regarding the expression of CAMs on the inducing, induced and subsequently-differentiated tissues. These rules fitted in with his Morphoregulator Hypothesis - a grand scheme, in which CAM interactions would lead to the production of morphogens that acted on adjacent cell complexes, altering the expression of their CAMs and tissue-specific proteins. Thus CAMs would have a direct role both in initiating and propagating a cascade of inductive processes.

A couple of points need to be made here:

- 1) Whilst great care was taken to use antibodies that bound to all of the (known) three main NCAM isoforms, great care must be taken in the interpretation of any negative result in such an antibody assay (or indeed any antibody-based experiment). A small variation in any part of the protein molecule (e.g. the extra-cellular sugar content) could easily remove/alter the antibody's epitope rendering antibody binding impossible; large variations in the sugar content are known. The use of polyclonal antibodies helps to reduce this problem but sialic acid residues are known to be highly immunogenic and produce a restricted range of immuno-dominant epitopes. These experiments therefore need to be followed up with rigorous RNA assays. These would provide a firmer basis to any claim

that the protein sequentially **appears/disappears** as opposed to just alters its form somewhat.

Therefore, the data as yet only shows that modulation of the form of the CAM proteins occurs, but this in itself is significant.

- 2) As discussed above, experiments by Kintner appear to rule out a direct role for NCAM in the primary mechanism of neural induction. No comment can be made as to whether other CAMs could not play the regulatory roles proposed by Edelman.

## 2.7 Structural organisation of NCAM isoforms

### 2.7(i) Splicing of the NCAM message

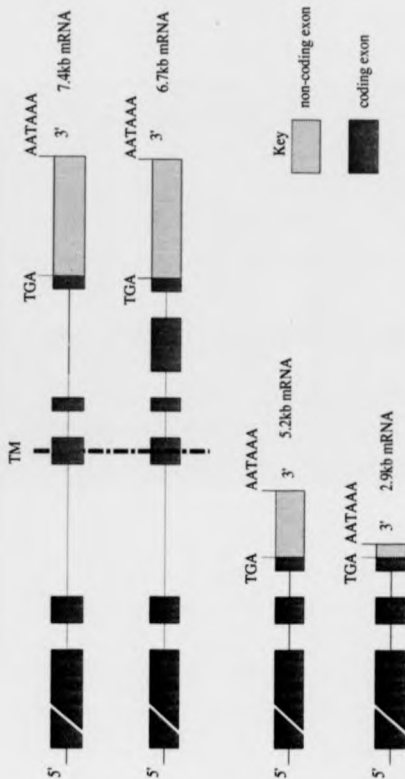
In mouse, human and chick, the only species to have genomic sequence published so far (Krieg has *Xenopus*, pers. comm.), all of the NCAM isoforms are derived from a single gene. This gene is known to be consist of at least 19 exons in chick (OWENS *et al.*, 1987) and spans over 120kb in mouse (GENNARINI *et al.*, 1986). Table II.1 illustrates the current state of knowledge regarding the protein and corresponding cDNA sizes.

### 2.7(ii) Use of different exons

The four main NCAM isoforms are all derived from one gene. Figure II.3 illustrates the way in which the different exons are spliced in order to produce these different isoforms.

Figure II.3

Use of differential splicing to produce the main mouse NCAM isoforms.



The above figure illustrates the differential splicing in the 3' region of the mouse NCAM gene. Note that the mouse GPI linked mRNAs (5.2 and 2.9kb) differ only in the length of the 3'-untranslated regions; both 3'-ends are derived from the same exon by the use of two different poly A+ addition sites (BARBAS *et al.*, 1988).

#### 2.7(iii) Conservation of intron/exon sites

Comparisons between the available mouse (only 3' six exons published) and chick genomic DNAs show that the positions of the splice sites are highly conserved (BARBAS *et al.*, 1988). Although the sequence homology varies between exons, this is also overall very high.

#### 2.7(iv) Notch homology in the cytoplasmic region

A number of short (5-7 amino acids) peptide sequences that are present in the cytoplasmic domain of mouse NCAM have high homology to sequences in the cytoplasmic domain of the *Drosophila* Notch protein (WHARTON *et al.*, 1985); no homologies were found in the extra-cellular domain. Notch is involved in cell fate determination, but the significance of these peptides is unclear. The role, in general, of the cytoplasmic domain is also unclear.

#### 2.7(v) The muscle-specific exon, MSD1

The first case of alternative splicing in the human NCAM extra-cellular domain has been reported by Walsh's group (DICKSON *et al.*, 1987). An additional 105bp was found in mRNA expressed preferentially in differentiated skeletal muscle myotubes; this results in an extra 35 amino-acids in the membrane proximal region of the GPI-linked proteins. Due to its tissue-specific expression it was given the name muscle-specific domain 1 (MSD1). This short block is known to comprise at least three exons MSD1a, b, and c comprising 15, 48 and 42bp respectively. MSD1a is known to be more than one

exon but finding this 15bp in >12kb is proving surprisingly difficult.

The role of MSD1 is as yet unresolved but it is known to have structural characteristics similar to the hinge region of immunoglobulin (WALSH,1988). It has been postulated that the presence of the MSD1 domain near the attachment site of NCAM to the plasma membrane may modulate the flexibility or conformation of this region of the protein. MSD1 is one of the sites in NCAM for O-linked glycosylation (WALSH *et al.*, 1989); this may affect the adhesive properties of the cell through steric hindrance, in a similar way to sialic acid variation (see below).

#### 2.7(vi) A secreted form of NCAM - NCAMSEC

Walsh's group have also discovered another unusual human NCAM exon, NCAM-SEC, the expression of which results in the production of a secreted form of NCAM (GOWER *et al.*, 1988). This secreted NCAM isoform (115kDa) appears to be a variant of the GPI-linked NCAM isoform (125kDa) that lacks the 3' amino-acids for GPI-attachment to the plasma-membrane.

The NCAM-SEC exon lies between exons 12 and 13 (with respect to the chick exons) and immediately 3' to MSD1. Unlike MSD1 it is not restricted to muscle tissue; it is found in both muscle and neural tissues. NCAM-SEC contains an in-frame stop codon after the extra-cellular domain coding sequences and thus prevents expression of any 3' transmembrane/cytoplasmic domains. Transfection of this cDNA

into tissue culture cells results in the accumulation of vesicular cytoplasmically-localized products and secretion of a 115kDa NCAM product into the media.

A number of groups had previously reported the presence of NCAM in the extra-cellular medium of tissue-culture cells (RUTISHAUSER *et al.*, 1976, COLE and GLASER, 1986). Previously it had been unresolved whether this soluble form of NCAM was produced by enzymatic release of the GPI-linked isoform or whether it was a genuinely secreted molecule. Although Walsh's group appear to have settled this point, the role of this soluble form of NCAM is still unclear.

The evidence, upon which to postulate a role for soluble NCAM, is as yet fragmentary. (It has been found at later stages of myogenesis (DICKSON *et al.*, 1987) and in denervated muscles (COVAULT and SANES, 1986)). It can easily be envisaged how soluble NCAM could act in competition against cell-bound NCAM for adhesive sites; therefore, in this role it would act as an adhesion modulator.

#### 2.7(vii) Differential splicing mechanisms

The mechanisms that control differential splicing and transcriptional termination in eukaryotes are not very well understood to say the least. But it is through these two mechanisms that eukaryotes manage to conjure up a diverse range of proteins from just one stretch of DNA. As shown above, NCAM is just one example of how these mechanisms are utilized to produce range of proteins at different times and places during cellular differentiation.

Members of Goridis' group (BARBAS *et al.*, 1988), in common with all others who have looked, failed to find any convincing evidence to explain the differential selection of mouse NCAM exons. Inverted repeats were found to span one exon that could form hair-pin loops and hence stabilize excision products, but the stabilization energy was not considered significant enough to influence exon choice.

The availability of mouse clonal cell lines that demonstrate a switch from one NCAM isoform to another may provide an ideal system in which to study this problem.

#### 2.7(viii) Transcriptional termination

The factors determining the choice of transcriptional termination sites is also poorly defined. It has been shown that transcription can continue for kilobases after the end of the mRNA (BIRNSTIEL *et al.*, 1985). An analysis of the DNA surrounding the two poly A sites in the mouse 2.9/5.2kb mRNAs (described above) again found no clues as to the nature of this selection process.

#### 2.8 NCAM expression in vertebrate muscle

In most vertebrates, the process of myogenesis involves the fusion of mononuclear myoblasts to form polynuclear myotubes, these subsequently forming muscle fibres. A change in expression of NCAM isoforms during this process has been reported (COVAULT and SANES, 1986; KNUDSEN *et al.*, 1990) and also quantitated (MOORE *et al.*, 1987).

Figure 11.4 summarizes the above references. This figure also

Figure II.4

Changes in NCAM expression patterns during mouse myoblast fusion.

Myoblast - myoblast fusion -> Myotube

mRNA/protein sizes:

6.7kb mRNA (145kDa)	5.2kb mRNA (155kDa)
	2.9Kb mRNA (125kDa)

Isoform:

Transmembrane	GPI-linked (includes MSD1)
---------------	----------------------------

Mouse muscle cell lines:

G8-1	4x increase in NCAM staining
C2	15x increase in NCAM staining (even staining over entire cell; expression precedes fusion)

The above figure summarizes the predominant NCAM mRNA, protein-size and isoform changes that occur during the transition from myoblast to myotube. These changes in NCAM protein expression have been quantified using mouse muscle cell lines.

(COVAULT and SANES, 1986; KNUDSEN *et al.*, 1990; MOORE *et al.*, 1987).

illustrates the NCAM isoform transition during *in vitro* myogenesis. It has been suggested that rapid cell-cell release could be effected by the cleavage of this GPI-linked NCAM isoform (HE *et al.*, 1986). There is no evidence, however, that this "escape-mechanism" is actually used.

As stated above, the rapid rise in NCAM expression precedes myoblast fusion but fusion is not dependent on NCAM expression. Knudsen *et al.* (above) have showed that anti-NCAM antibodies significantly reduced the rate of myoblast fusion but did not affect the final extent of fusion. They postulated that other cell adhesion molecules (in particular calcium-dependent ones) also have a role to play in this process. The co-localization of NCAM and N-Cadherin (a calcium-dependent CAM) has since been demonstrated in avian skeletal myoblasts (SOLER and KNUDSEN, 1991).

As development proceeds *in vivo*, NCAM expression on muscle cells is gradually lost but remains at neuromuscular junctions, highlighting a possible role for NCAM in innervation. NCAM expression is also induced at the sarcolemma when the tissue is subjected to experimental abuse (e.g. crushing) or certain pathological diseases (COVAULT and SANES, 1986).

### 2.9 Muscle development in *X.laevis*

Muscle development in *X.laevis* proceeds by more than one route: muscles surviving through metamorphosis (e.g. intermandibular muscle) develop as described above; embryonic muscles such as the myotomes are produced from large mononuclear myocytes which

become fully functional in the absence of cell fusion. Later in development these myocytes do fuse with satellite cells; ultimately they are reabsorbed at metamorphosis. Expression of NCAM in these two different muscle-types has been reported (KAY *et al.*, 1988).

As described above *X.laevis* somites are unusual amongst vertebrates in that they do not appear to express NCAM. Kay *et al.* examined the intermandibular muscle for NCAM expression to determine if the developmental origin of the muscle tissue affected NCAM expression or whether *X.laevis* embryonic muscle was uniformly NCAM-negative. They found the latter to be true: both myotome and intermandibular muscle were negative in the embryo.

However, post-embryonic skeletal muscle does show similar patterns of NCAM expression to other mammalian and avian systems.

Therefore, in the *X.laevis* embryo at least, NCAM does not play a role in muscle development; other cell adhesion molecules must therefore be involved.

## 2.10 Conclusions

The above chapter has contained a concise description of the NCAM protein and corresponding DNA sequences. It is clear that much still remains to be discovered about the differing roles of the three main isoforms and the mechanisms that govern their distribution. The role of PSA in NCAM-mediated binding does appear to be important but, again, this role is not completely resolved.

Whilst speculative theories are abundant, the hard evidence only

points to NCAM playing a contributory role in neural and muscle development, not so much playing second fiddle, but as one of a group of molecules whose actions and interactions result in the desired tissue development. In retrospect, given NCAM's abundance and the fact that it was the first well-characterised cell-adhesion molecule, it would have been extremely fortuitous to discover the secret of neural and muscle development from the study of one molecule. It is now obvious that in order to achieve this goal, it is necessary to study the expression of further CAMs, the regulation of these molecules and how they interact with one another.

## Chapter III

### The Cadherin gene family

#### 3.1 Introduction

#### 3.2 Identification and definition of Cadherins

- (i)  $\text{Ca}^{2+}$ -dependent and  $\text{Ca}^{2+}$ -independent CAMs
- (ii) Isolation of the first Cadherin protein
- (iii) Isolation of N-Cadherin and P-Cadherin

#### 3.3 Structure of Cadherins

#### 3.4 Cadherin multigene family

- (i) Cadherin family of proteins
- (ii) Expression of N-Cadherin
- (iii) Ectopic expression of N-Cadherin in *Xenopus* embryos

#### 3.5 Extra-cellular domain

- (i) Theoretical binding mechanisms
- (ii) Cell aggregation experiments
- (iii) L-cell transfection experiments
- (iv) E/P-Cadherin DNA chimeras
- (v) Monoclonal antibodies to amino-terminal region
- (vi) Site directed mutagenesis of possible binding region
- (vii) Glycosylation of extracellular domain
- (viii) Role of calcium ions in binding

#### 3.6 Intra-cellular domain

- (i) Function of cytoplasmic domain
- (ii) Association of Cadherins with catenins

#### 3.7 Role of Cadherins in morphogenesis

- (i) Examples of the dynamic regulation of Cadherins
- (ii) General theories of CAM-regulated morphogenesis

### 3.8 Role of Cadherins in tumorigenesis

- (i) Role of CAMs as tumour suppressors
- (ii) Human rectal carcinoma gene - DCC
- (iii) *Drosophila* tumour suppressor gene - *fat*

### 3.9 Importance of the study of Cadherins

### 3.10 Conclusions

### 3.1 Introduction

The historical aspects of the study of cell-adhesion and the division of CAMs into two main classes of cell-adhesion molecules (calcium-independent and calcium-dependent) has been discussed above in Chapter I. The role of NCAM, as an example of the former class of molecules, has been reviewed in Chapter II. This chapter therefore comprises a review of the rapidly-expanding class of calcium-dependent cell adhesion molecules - the Cadherins.

These transmembrane CAMs have now been found on a wide variety of tissues and hence have evoked much speculation about the possible role that they might play in the morphogenesis of the tissues in which they are found. The roles of the extra-cellular domains, in ligand binding, and the possible role of the conserved intra-cellular domain, in cell-signalling and/or cytoskeleton binding, are yet to be completely determined.

The main protagonist in this area of study is Masatoshi Takeichi and his group at Kyoto University in Japan. The reader is referred to his many papers and reviews for additional information (TAKEICHI 1987, 1988, 1990, 1991).

### 3.2 Identification and definition of Cadherins

#### 3.2 (i) $\text{Ca}^{2+}$ -dependent and $\text{Ca}^{2+}$ -independent CAMs

The fact that  $\text{Ca}^{2+}$  ions give a degree of protection against degradation by proteases to  $\text{Ca}^{2+}$ -dependent CAMs has been used as a way of distinguishing the two main CAM classes.  $\text{Ca}^{2+}$ -dependent and  $\text{Ca}^{2+}$ -independent cell-cell adhesion systems may be selected for by the differential use of  $\text{Ca}^{2+}$  and trypsin according to the following cell treatment regimes:

Treatment:High trypsin +  $\text{Ca}^{2+}$ Low trypsin -  $\text{Ca}^{2+}$ High trypsin -  $\text{Ca}^{2+}$ Selects for: $\text{Ca}^{2+}$ -dependent CAMs $\text{Ca}^{2+}$ -independent CAMs

no CAMs

### 3.2 (ii) Isolation of the first Cadherin protein

The actual procedure used to identify the cadherin proteins was similar to that described above for NCAM, (i.e. the use of antibody neutralisation assays or "Fab strategy") and is illustrated in Figure III.1. Antisera to F9 teratocarcinoma cells raised in rabbits was shown to inhibit F9 cell aggregation. This aggregation was then shown to be specific to  $\text{Ca}^{2+}$ -dependent CAMs by use of the above regimes. Trypsin digestion in the absence of calcium was shown to release a soluble factor into the medium which was subsequently shown to be part of a 124KDa glycoprotein (YOSHIDA and TAKEICHI, 1982)

A monoclonal antibody, ECCD-1, was later employed to demonstrate similar cell-aggregation inhibiting effects (YOSHIDA *et al.*, 1984); this mAb also recognizes a 124KDa protein confirming this protein's role in cell-adhesion. This protein was subsequently called "Cadherin".

Similar proteins have now been identified in other organisms: uvomorulin (mouse, PEYRIERAS *et al.*, 1983), Cell-Cam 120/80 (human, DAMSKY *et al.*, 1983), Arc-1 (dog, BEHRENS *et al.*, 1985) and L-CAM (chick, GALLIN *et al.*, 1983). These molecules all share similar molecular masses,  $\text{Ca}^{2+}$ -sensitivities and tissue distributions (epithelial cells in a variety of embryonic and adult tissues), which strongly suggested that they are identical or inter-species homologues.

### 3.2 (iii) Isolation of N-Cadherin and P-Cadherin

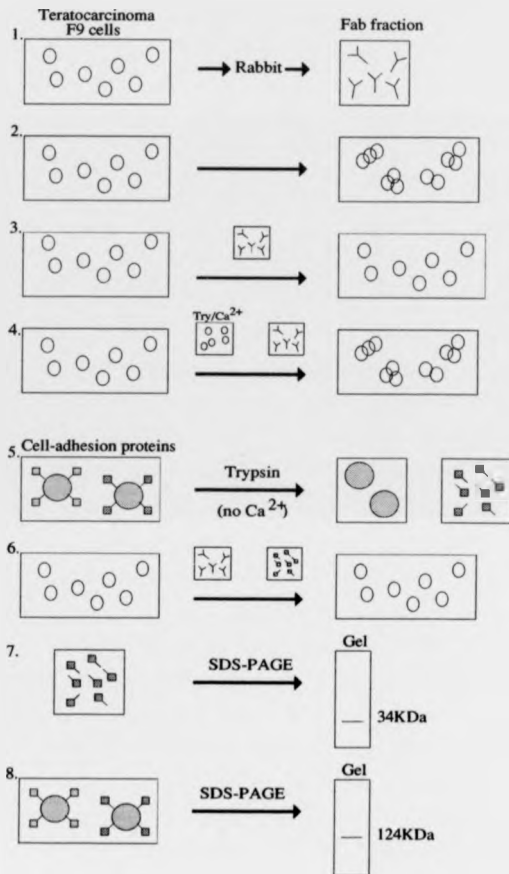
Takeichi's group then went on to identify mAbs to  $\text{Ca}^{2+}$ -dependent CAMs in mouse and chicken brains (NCD-1, HATTA *et al.*, 1985; NCD-2, HATTA and

Identification of  $\text{Ca}^{2+}$ -dependent cell-adhesion molecules

1. Fab fragments were prepared from sera obtained by injecting teratocarcinoma F9 cells into rabbits. Such sera would be expected to contain antibodies to cell-adhesion molecules (CAMs).
2. Since F9 cells express CAMs, these cells normally aggregate into clumps.
3. The addition of Fab fragments prevents the aggregation of the F9 cells, due to inhibition of the CAMs.
4. Cells that have been treated with trypsin/ $\text{Ca}^{2+}$  will fully absorb the aggregation-inhibiting activity of the Fab fragments. Since this treatment leaves  $\text{Ca}^{2+}$ -dependent CAMs unaffected, it assumed that the Fab fragment is specific to  $\text{Ca}^{2+}$ -dependent CAMs.
5. Treatment of the F9 cells with trypsin in the absence of  $\text{Ca}^{2+}$  releases a soluble factor from the CAMs into the supernatant.
6. This soluble factor can compete with the Fab fragment to prevent the inhibitory effect of the Fab fragment on aggregation of F9 cells.
7. Concentration of this soluble factor has shown it to be a 34KDa protein by immuno blotting.
8. A 124KDa cell-surface glycoprotein can compete with this 34KDa protein in immunoprecipitation; these proteins therefore share an epitope recognised by the Fab fragment.

It was therefore concluded that this 124KDa glycoprotein is a component of the  $\text{Ca}^{2+}$ -dependent cell-adhesion system of the teratocarcinoma cells.

Figure III.1b



TAKEICHI, 1986) and mouse placenta (PCD-1, NOSE and TAKEICHI, 1986). The antigens recognised by these mAbs also have similar molecular masses and  $\text{Ca}^{2+}$ -sensitivities to those listed above but have distinct tissue distributions. Therefore, whilst these molecules are immunologically distinct, they clearly form a part of the Cadherin family. The subsequent cloning of the cDNAs of these molecules has verified this fact.

### 3.3 Structure of Cadherin proteins

A sequence comparison of E-, N- and P-cadherin and L-CAM (chick E-Cadherin) has shown that all have a primary structure of 723-748 amino-acids comprising a putative signal polypeptide, putative precursor region, and a highly hydrophobic region that is likely to be a membrane-spanning domain. This putative primary structure is shown in Figure III.2; an inter-species assessment of the amino-acid homologies is shown in Figure III.3.

Amino-acid sequence conservation is highest in the cytoplasmic domain and in the N-terminal region of the extra-cellular domain, indicating the functional importance of these regions.

### 3.4 Cadherin multigene family

#### 3.4(i) Cadherin family of proteins

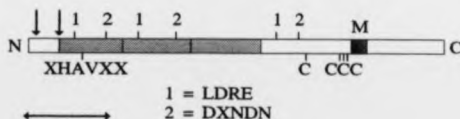
DNA sequencing of the cadherin cDNAs has now clarified the relationship between the multitude of proteins that were previously classified solely by mAb epitopes. The current Cadherin family of published cDNAs is given in Figure III.4, together with recent genomic DNA data.

The reader is referred to the individual references for further, more detailed, information. The expression of N-Cadherin is, however, of obvious relevance this work; the expression of this molecule will be therefore now be discussed, followed by the results of ectopic N-Cadherin expression experiments.

#### 3.4(ii) Expression of N-Cadherin

Figure III.2

Putative primary structure of Cadherins



The diagram illustrates the features that are common to all vertebrate Cadherins known to date.

Key:

N= amino-terminal;

C= carboxy terminal;

M is the putative membrane spanning region;

Shaded regions are internally repeated domains;

1 and 2 are major repeated amino-acid sequences;

Bipolar arrow indicates the amino-terminal 113 amino-acids involved in specificity of binding;

Vertical arrows show the position of mAb epitopes that inhibit Cadherin binding;

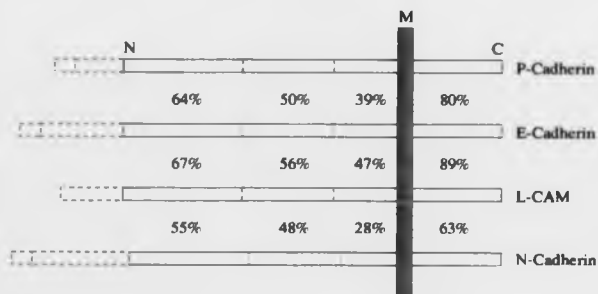
C's are conserved cysteine residues;

XHAVXX is an important amino-acid sequence in which substitutions of the X residues alters binding specificity of E-Cadherin.

(Based on TAKEICHI, 1990)

**Figure III.3**

**Amino-acid sequence conservation among Cadherin proteins**



**Key:**

N= putative amino-terminus; C= putative carboxy-terminus;

M= putative membrane spanning region.

The primary structures of the mouse P- and E-Cadherin and chick LCAM (E-Cadherin) and N-Cadherin are shown. Percentage amino-acid similarities are given between adjacent proteins for three regions of the extra-cellular domain and for the cytoplasmic domain.

The putative precursor regions are shown in dotted lines at the N-terminus.

The figure illustrates the high degree of sequence conservation in the cytoplasmic domain and, to a lesser extent, in the N-terminal region of the extra-cellular domain.

(Based on TAKEICHI, 1988)

Figure III.4

Published Cadherin Sequences

cDNA	Species:	Derivative tissue:	Reference:
N-Cadherin	Chick <i>Xenopus</i> <i>Xenopus</i> Mouse	Neural	HATTA <i>et al.</i> , (1988) GINSBERG <i>et al.</i> , (1991) DETRICK <i>et al.</i> , (1990) MIYATANI <i>et al.</i> , (1989)
E-Cadherin	Chick (LCAM) Mouse (uvomorulin) + Mouse (E-Cadherin)	Epidermis	GALLIN <i>et al.</i> , (1987) RINGWALD <i>et al.</i> , (1987) NAGAFUCHI <i>et al.</i> , (1987)
P-Cadherin	Mouse	Placenta	NOSE and TAKEICHI, (1986)
R-Cadherin	Chick	Retina	INUZUKA <i>et al.</i> , (1991)
EP-Cadherin	<i>Xenopus</i>	Egg/kidney	GINSBERG <i>et al.</i> , (1991)
PVA (Pemphigus vulgaris antigen)	Human	Squamous epithelia	AMAGAI <i>et al.</i> , (1991)
XB-Cadherin	<i>Xenopus</i>	Blastomere	HERZBERG <i>et al.</i> , (1991)
M-Cadherin	Mouse	Muscle	DONALIES <i>et al.</i> , (1991)
T-Cadherin	Chick	Truncated	RANSCHT <i>et al.</i> , (1991)
U-Cadherin	<i>Xenopus</i>	Ubiquitous	ANGRES <i>et al.</i> , (1991)
Desmocollin	Cow	Epidermis	MECHANIC <i>et al.</i> , (1991)

Genomic sequences

N-Cadherin	Mouse	MIYATANI <i>et al.</i> , (1992)
	Human	WALLIS and WALSH, (1992)
P-Cadherin	Mouse	HATTA <i>et al.</i> , (1991)
L-CAM/ B-Cadherin	Chick	SORKIN <i>et al.</i> , (1991)

Expression patterns of N-Cadherin in early *X.laevis* embryos have been described by at least four groups (DETRICK *et al.*, 1990; FUJIMORI *et al.*, 1990; GINSBERG *et al.*, 1991; and SIMONNEAU *et al.*, 1992). (Due to the close genetic relationship between *X.laevis* and *X.borealis*, the expression patterns in *X.borealis* would be assumed to be very similar if not identical to those as published above). Expression of N-Cadherin in chick has been extensively published by members of Takeichi's group (see TAKEICHI, 1988, for a review).

Experiments by members of Kintner's group (DETRICK *et al.*, 1990) show that the N-Cadherin mRNA is approximately 4.2Kb, is not detectable by Northern blot in egg or early gastrulae (stage 10), but can be detected in early neurulae (stage 14) and early tadpoles (stage 20) by Northern blot. RNase protection assays (RPAs) have been used to determine expression patterns at a greater resolution. Kintner's group first detected mRNA transcripts at stage 10-12 using RPAs; expression increased thereafter up to at least stage 20.

The probe used for the RPAs was taken from the least-conserved "pre" region of the N-Cadherin cDNA to minimise cross-Cadherin hybridisation. No N-Cadherin expression was detected in early endoderm or ectoderm (late blastula) in the absence of any mesoderm/neural induction. However, the dorsal blastopore lip region (the "organiser"), when isolated and cultured to early neurula stage, expressed significant levels of N-Cadherin.

During the early stages of neural development (see Section 1.3), N-Cadherin expression is high in the neural plate; expression also continues into the derivative tissues, i.e. neural tube and tadpole brain (although Ginsberg *et al.*, surprisingly, failed to detect N-Cadherin in tadpole brain by Northern blot).

Some mesoderm tissues also express significant levels of N-Cadherin. The notocord and pronephros strongly express N-Cadherin; most other mesodermal

derivatives express low amounts; but the somites are said to be virtually negative (*in situ* hybridisation, SIMMONEAU *et al.*, 1992).

Isolated ectoderm was also assayed for N-Cadherin expression before and after contact with a neural-inducing agent. Kintner's group used the Hensen's node from chick as the inducing agent since the endogenous neural inducer, dorsal mesoderm, expresses N-Cadherin itself. Ectodermal cap/Hensen's node combinations were cultured together for various times: after 2 hours (stage 10.5 equivalent) no N-Cadherin mRNA was detectable; but there was expression of N-Cadherin after 4 hours (stage 13 equivalent). Ectodermal caps cultured alone showed no such expression. These experiments showed that ectoderm is capable of being induced by a neuralising agent to produce N-Cadherin mRNA and that such expression is co-ordinate with the morphogenetic movements associated with neurulation.

#### 3.4(iii) Ectopic expression of N-Cadherin in *Xenopus* embryos

The misexpression of N-Cadherin has been reported to produce gross morphological defects in *X.laevis* embryos (DETRICK *et al.*, 1990; FUJIMORI *et al.*, 1990). Synthetic N-Cadherin RNA was injected into early stage embryos; the embryos were then scored for physiological defects at various stages thereafter. Thickening, clumping and fusion of cell layers were reported as well as a striking rift in the ectoderm (exposing the underlying tissues) that was often observed during gastrulation.

As the authors themselves acknowledge, such experiments must be interpreted with extreme caution since the conditions imposed upon the embryos may be far removed from those normally experienced. This places a severe restriction upon the usefulness of such experiments. (After all, if you pour glue into a clock, the clock will stop; but this tells you nothing about how the clock works!). It must be noted, however, that many RNAs **can** be injected

at similar concentrations and have no detrimental effect on development).

### 3.5 Binding activity of extracellular domain

#### 3.5(i) Theoretical binding mechanisms

Theoretically, binding of cells expressing Cadherins could be achieved by a) the interaction of one Cadherin molecule with another Cadherin molecule; b) the interaction of one Cadherin molecule with another species of molecule (e.g. an integrin); or c) a combination of the above. The current thinking is that Cadherin binding is through homophilic interactions (i.e. choice (a)) with other Cadherin molecules of the same class, e.g. N-Cadherin with N-Cadherin. All the experimental evidence points in this direction.

#### 3.5(ii) Differential cell aggregation experiments

The classical differential cell aggregation experiments performed with sponges (Section 2.1(ii)) have now been reproduced using L-cells transfected with E-, P- and N-Cadherin (TAKEICHI *et al.*, 1981). Takeichi's group have shown that cells expressing one particular Cadherin preferentially aggregate with cells expressing similar adhesion molecules. It was therefore, not surprisingly, concluded that it was the extra-cellular domain of the protein that defines the particular binding specificity of that particular Cadherin.

#### 3.5(iii) L-cell transfection experiments

Homophilic binding activity of Cadherins has been demonstrated by transfecting L-cells with E-Cadherin (NAGAFUCHI *et al.*, 1987). L-cells have very low endogenous levels of E-Cadherin and hence do not naturally form tight intercellular connections in monolayer cultures. Upon transfection with E-Cadherin, however, the transfectants expressing E-Cadherin acquired  $\text{Ca}^{2+}$ -dependent aggregation activity that could be quantitatively correlated with the amount of E-Cadherin expressed. Subsequent experiments with N- and P-Cadherin have provided similar results (HATTA *et al.*, 1988).

Control experiments combining transfected L-cells with untransfected L-cells do not show this aggregating activity. It therefore can be concluded that homophilic binding must be occurring between Cadherin proteins, and that the expression of a single class of Cadherin is sufficient to induce cell-cell aggregation.

A series of different experimental approaches have successively defined with greater resolution the regions of the protein that are involved both in protein-protein binding and those regions that confer specificity upon that binding. These experiments include the production of chimeric cadherin molecules, the use of monoclonal antibodies with defined epitopes and site-directed mutagenesis.

#### 3.5(iv) E/P-Cadherin chimeras

Transfection of L-cells with E/P-Cadherin chimeric DNA constructs has shown that the replacement of the amino-terminal 113 amino-acids of the Cadherin protein is sufficient to alter the binding specificity of the expressed protein. In the published experiments, specificity was changed from that characteristic of E-Cadherin to that characteristic of P-Cadherin (NOSE *et al.*, 1990).

#### 3.5(v) Monoclonal antibodies to amino-terminal end of the protein

The mAbs PCD-1 and NCD-2 (*ibid.*) inhibit P- and N-Cadherin mediated aggregation respectively. They are also known to bind near the amino-terminus of the extracellular domain of the Cadherin protein; more specifically, both mAbs bind to the 31st amino-acid from the amino-terminal end (TAKEICHI, 1990).

#### 3.5(vi) Site directed mutagenesis of possible binding region

The amino-terminal 113 amino-acids is 65% conserved between E- and P-Cadherin, but in a non-random manner, i.e. stretches of conserved amino-acids

are interrupted by stretches of non-conserved amino-acids. It may be assumed that the non-conserved amino-acids are the ones that define that particular binding specificity of that molecule.

The amino-acid sequence HAV (initially identified by sequence comparison with influenza strain A haemagglutinin) has been termed a "Cell Adhesion Recognition" (CAR) sequence, due to the ability of synthetic peptides containing this sequence to inhibit the Cadherin-mediated compaction of mouse embryos (BLASCHUK *et al.*, 1990). Site-directed mutagenesis was used to investigate adjacent amino-acids whose substitution might lead to altered binding specificity. Substitution of the E-Cadherin amino-acids with those found in P-Cadherin lead to an alteration of the binding specificity of the expressed protein from that characteristic of E-Cadherin to a molecule that bound both E- and P-Cadherin. This region is therefore important in determining binding specificity but obviously other regions are also involved.

The above experiments therefore have started to define the regions/sites involved both in Cadherin-Cadherin binding and those that confer specificity upon this binding. It is obvious that full details of this cell-aggregation mechanism will be available in the near future.

### 3.5(vii) Glycosylation of extracellular domain

The role of sugar side-chains in the regulation of NCAM binding was discussed above (Section 2.3). In this molecule, the developmental regulation of polysialic acid (PSA) has been shown to have profound effects on the adhesiveness of cell expressing NCAM. No such effect has been shown for any Cadherins.

A number of possible glycosylation sites have been identified by amino-acid sequence analysis and Cadherins are known to be glycosylated to some degree. Inhibition of glycosylation by tunicamycin, however, has been shown to have

no effect on E-Cadherin binding activity (SHIRAYOSHI *et al.*, 1986).

### 3.5(viii) Role of $\text{Ca}^{2+}$ ions in binding

By definition, Cadherins require the presence of  $\text{Ca}^{2+}$  ions for cell-cell aggregation. Furthermore,  $\text{Ca}^{2+}$  ions protects Cadherins from protease digestion. The experiments given below have shone some light on the possible role of these  $\text{Ca}^{2+}$  ions in Cadherin regulation.

The E-Cadherin 84KDa (intermediate extra-cellular) trypsin digestion product was found to bind  $^{45}\text{Ca}^{2+}$  when immobilised on nitrocellulose paper, demonstrating that it is this extracellular domain of the Cadherin protein that binds the  $\text{Ca}^{2+}$  ions.

Some mAbs that bind to Cadherin proteins have been identified that only bind to their epitopes in the presence of  $\text{Ca}^{2+}$  ions. This suggests that the presence of  $\text{Ca}^{2+}$  ions induces a conformational change in the Cadherin protein.

The amino-acid analysis discussed above reveals a number of repeated motifs and conserved regions, such as the DXNDN motif. None of these sequences are recognized to be  $\text{Ca}^{2+}$  ion binding motifs, however, the presence of such repeated sequences obviously makes them prime candidates for such binding.

The current view of the Cadherin protein is illustrated in Figure III.5.

### 3.6 Intra-cellular domain

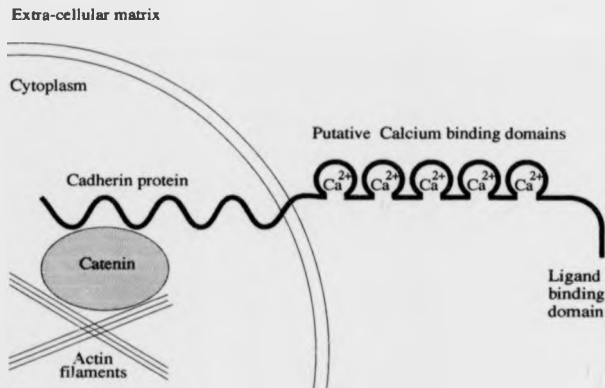
#### 3.6(1) Function of cytoplasmic domain

The intra-cellular domains of the Cadherin proteins are well conserved between the different classes and species; Cadherins are also known to be bound to the cytoskeleton. Deletion experiments have attempted to determine the important regions within this domain that dictate protein function.

A series of stepped DNA-deletions of this cytoplasmic domain were constructed which left unaffected both the transmembrane and extra-cellular domains (NAGAFUCHI and TAKEICHI, 1988). These deletion constructs

Figure III.5

Proposed structure of Cadherin proteins



The diagram illustrates the proposed transmembrane structure of the Cadherin proteins and the possible interactions that they make with the cortical actin bundles via catenin proteins.

Cadherins possess up to five putative  $\text{Ca}^{2+}$ -binding domains in the extra-cellular (EC) portion of the protein. The 5'-amino-terminal region of the protein is believed to be responsible for specificity of ligand binding.

were then transfected into L-cells and their Cadherin-mediated cell-cell aggregation ability assayed. It was found that deletion of the COOH-terminal region eliminated both attachment to the cytoskeleton and external binding activity. From these experiments it appears that the COOH-terminal region is not only necessary for some sort of anchorage to the cytoskeleton but that such anchorage is also necessary for the binding activity of the extra-cellular domain. This result may be interpreted as follows: an individual Cadherin molecule does not in itself have the 'strength' to mediate cell-cell binding: it needs to be attached to an underlying framework, i.e. the cytoskeleton (although no such attachment is necessary for NCAM). Alternative theories in which the attachment of the cytoplasmic domain to the cytoskeleton sends a signal to the extra-cellular domain (to be "receptive to binding") have also been postulated.

Evidence for such binding to the cytoskeleton comes from two main sources. Firstly, double-staining of antibodies to Cadherins and cortical actin bundles shows that both are co-localized at cell-cell junctions. Secondly, extraction of cells with non-ionic detergents appears to show some form of structural association between Cadherins and cortical actin bundles. (Not all Cadherins are attached to the cytoskeleton as a fraction of Cadherin proteins are soluble in non-ionic detergents).

### 3.6(ii) Association of Cadherins with catenins

Wild-type Cadherins are known to be localised at cell-cell junctions and also at zonula adherens - an intercellular junctional complex that contains cortical actin fibres, and also other intra-cellular proteins such as vinculin, radixin and  $\alpha$ -actinin. E-Cadherin is known to immunoprecipitate with specific intra-cellular proteins -  $\alpha$ ,  $\beta$ , and  $\gamma$ -catenin proteins (NELSON *et al.*, 1990). The Cadherin's catenin-binding sites are the same as those sites important for actin-

binding and full cell-cell binding activity. This suggests that catenin proteins act as links between Cadherins and the cytoskeleton, and that such association is necessary for full Cadherin function.

### 3.7 Cadherins in morphogenesis

#### 3.7(i) Examples of dynamic regulation of Cadherins

Cadherins are not cell-specific molecules. The expression of Cadherins on the surface of cells of a particular tissue is dynamically regulated, in that Cadherin expression patterns change as the cells differentiate. A number of examples of such dynamic expression are known; some are given in Figure III.6.

#### 3.7(ii) General theories of CAM-regulated morphogenesis

The examples in Figure III.6 of temporal variations in Cadherin expression provide illustrations as to how the separation of two cell populations from one precursor cell population may be achieved. It can readily be seen how the loss or variation of Cadherin expression from an initially Cadherin-homogeneous tissue, and the subsequent expression of different Cadherins in subpopulations of that tissue, may lead to the division of that tissue into two distinct tissues. Obviously this type of mechanism would not be necessarily unique to the Cadherin family.

Since most tissues express more than one CAM, a specific degree of cell-cell binding may also be regulated by quantitatively controlling the number and type of CAMs expressed. A low degree of adhesiveness may be appropriate for a neuron following a set path defined by a CAM-expressing substrate (so that the neuron is not irreversibly bound to the substrate), whereas the terminal differentiation of a tissue may require the irreversible binding together of its constituent cells, i.e. high CAM expression.

Thus by relatively simple mechanisms, both the specificity and the degree of cell-cell binding may be effectively controlled.

Figure III.6

Examples of dynamic Cadherin expression  
in the morphogenesis of tissues

<u>System</u>	<u>'Pre' expression</u>	<u>Change</u>	<u>'Post' expression</u>
1. Chick embryo (blastula stage)	Epiblast cells LCAM <sup>+</sup> N-Cadherin <sup>-</sup>	Gastrulation	Mesodermal cells LCAM <sup>-</sup> N-Cadherin <sup>+</sup>
2. Neural tube formation	Neural plate cells LCAM <sup>+</sup> N-Cadherin <sup>-</sup>	Neurulation	Neural tube LCAM <sup>-</sup> N-Cadherin <sup>+</sup>
3. PNS formation	Neural crest cells LCAM <sup>+</sup>	Migration	Neural crest cells LCAM <sup>-</sup>

The examples referred to above have now been established as "classical" examples of differential Cadherin expression (TAKEICHI, 1990). The "+" and "-" refer to whether the protein is expressed or not.

The disappearance or appearance of the above CAMs, at these important times during tissue-morphogenesis, has evoked much speculation about the role these proteins may play in these developmental events.

Note:

LCAM is also known as E-Cadherin.

### 3.8 Role of Cadherins in tumorigenesis

#### 3.8(i) Role of CAMs as tumour suppressors

The malignant phenotype is often associated with a loss of cell-cell or cell-basement membrane contact, resulting in metastasis and the establishment of ectopic malignant tissues. It is not difficult to appreciate how the disruption of CAMs could play a major role in this process.

The role of tumour-suppressor genes is now as well recognised in the prevention of neoplasias as the role of oncogenes in their initiation. A number of tumour-suppressor genes have now been identified that, once sequenced, have revealed significant homology to CAMs, in particular NCAM and the Cadherin gene family.

The recent Cadherin literature has been overwhelmed with references to the potential involvement of Cadherins in tumorigenesis (e.g. genital cancers, INOUE *et al.*, 1992; prostate cancer, ISAACS *et al.*, 1992; lung cancer, RYGAARD *et al.*, 1992).

The majority of the literature refers to the loss or reduction in expression of E-Cadherin in various cancers, leading to loss of epithelial differentiation and gain of invasiveness: phosphorylation of the Cadherin- $\beta$ -catenin complex may also play a part in this process (BEHRENS *et al.*, 1993). Transfection of transformed tissue culture cell lines with Cadherin cDNAs has been shown to reduce the invasiveness of those cells, often resulting in partially differentiated tumours instead of the previous fully undifferentiated tumours (VLEMINCKX *et al.*, 1991; FRIXEN *et al.*, 1991).

A number of newly-identified genes have been shown to have homology to known CAMs, including DCC and *fat* (see below).

#### 3.8(ii) Human rectal carcinoma gene - DCC

Deletions in human chromosome 18q are found in more than 70% of colorectal

cancers, implying the presence of a possible tumour-suppressor gene in this region. The gene DCC (Deleted in Colorectal Carcinoma) has been located in this region and has significant homology to NCAM, in particular to the Ig domains; it also has fibronectin-type repeats (FEARON *et al.*, 1990).

### 3.8(iii) *Drosophila* tumour suppressor gene - *fat*

The cDNA coding for a giant 5000 amino-acid protein has been identified in *Drosophila* by screening a *Drosophila* genomic library with PCR probes homologous to the Cadherin extra-cellular domain and then using the genomic clones obtained to screen a *Drosophila* cDNA library (MAHONEY *et al.*, 1991). The protein, called *fat*, comprises 34 repeats of a 100 amino-acid Cadherin-like domain, four EGF-type repeats, a transmembrane domain and a cytoplasmic domain (that has no homology to the corresponding Cadherin cytoplasmic domain).

This is the first Cadherin-like protein that has been identified in a non-vertebrate and is considerably larger than any known Cadherin protein identified to date (most are approximately 730 amino-acids). Recessive (loss of function) mutations in this gene leads to excessive cell proliferation in the imaginal discs; this gene has therefore been termed a tumour suppressor.

### 3.9 Importance of the study of Cadherins

"What genes control Cadherin expression? The differential expression of multiple Cadherins in development must be under strict control of regulatory genes. It is most important to identify such genes to understand the genetic mechanisms of morphogenesis." (TAKEICHI, 1988)

### 3.10 Conclusions

This chapter has described a class of transmembrane proteins that initially had three members: Epidermal-, Placental- and Neural-Cadherin. These proteins were known to mediate the  $\text{Ca}^{2+}$ -dependent aggregation of cells. It is now

known that these proteins are far from tissue specific and that differential expression, both temporally and spatially, of combinations of these proteins may have major roles to play in the very 'construction' of the animal form.

This group of proteins has recently expanded to encompass at least eleven close relatives and many other proteins having significant homology to Cadherin protein domains.

The recent interest in CAMs, Cadherins in particular, as tumour suppressors will ensure that this group of proteins will continue to be the subject of intense study.

## Chapter IV

### Eukaryotic gene regulation and genomic DNA structure

#### 4.1 Introduction

#### 4.2 Eukaryotic gene regulation

#### 4.3 Transcriptional control

##### (i) Promoter structure

##### (ii) Neural-specific promoter elements

##### (iii) Transcription factors

#### 4.4 NCAM promoter

##### (i) Introduction

##### (ii) NCAM promoter clones

##### (iii) Mouse NCAM promoter

##### (a) Transcriptional initiation

##### (b) Poly hetero-nucleotide regions

##### (c) Deletion constructs

##### (d) Protein footprints

#### 4.5 Cadherin promoters and genomic DNA structure

##### (i) N-Cadherin genomic DNA structure

##### (ii) P-Cadherin genomic DNA structure

##### (iii) L-CAM/K-CAM genomic DNA structures

##### (iv) E-Cadherin promoter

#### 4.6 Conclusions

#### 4.1 Introduction

The ability of the cell to manufacture a wide variety of proteins having numerous specific functions is of little use without effective control over the temporal and spatial expression patterns of those proteins. As was seen in Section 3.4(iii), the misexpression of just one protein (N-Cadherin) during neurulation can have a profound detrimental effect on the development of that organism. Control mechanisms therefore must exist to finely coordinate the expression of a protein with the time and place where it is required.

This chapter will therefore look into the organisation of eukaryotic genes with specific reference to the promoter and genomic DNA structures of NCAM and the known Cadherins.

#### 4.2 Eukaryotic gene regulation

The steps involved in the expression of a protein in a eukaryotic cell are now known in some detail: these include, *inter alia*, the binding of transcription factors to DNA control sequences, transcription of the DNA by RNA polymerases, splicing of the hnRNA, export of the mRNA to the cytoplasm, translation of the mRNA into protein and post-translation modifications. Theoretically, any one of these steps could be the subject of a control mechanism (and most are, in one organism or another, LATCHMAN, 1990).

It is now generally accepted that, whilst regulation at other levels are known, the fundamental level at which control of protein expression is exercised is that of gene transcription. The remainder of this chapter will therefore concentrate on transcriptional control mechanisms. However, one additional mechanism is worthy of comment at this stage and that is post-translational modification.

It is well known that in eukaryotic organisms chemical modifications are often made to the amino-acids. Examples of such chemical modifications are the phosphorylation, acylation and sulphation of amino-acids; the purpose of

these are the subject of much study. NCAM is a paradigm of a protein to which sugar side chains are added (see Section 2.3). Through alterations of the proteins such as these, the function/configuration of the protein may be altered and the protein may therefore be regulated on a continuous and ongoing basis.

#### 4.3 Transcriptional control

##### 4.3(i) Promoter structure

The standard elements of a eukaryotic promoter (e.g. TATA box, CCAAT box, enhancer sequences) are now well accepted although it must be noted that not all eukaryotic promoters include a TATA box (e.g. 'house-keeping' genes lack TATA boxes and are transcribed at a basal rate in most cells).

The TATA or Hogness-Goldberg box is usually embedded in a region of relatively low AT content (NUSSINOV *et al.*, 1986); a plot of the distribution of ATA + TAT triplets normally shows a peak around the TATA box and often a secondary peak around -275bp. The CCAAT box is less well defined. It has been found that even the presence of both of these elements is insufficient to categorically define the transcriptional initiation site of a eukaryotic gene. This underlines the importance of gene-specific transcription factors in aiding the initiation of transcription by RNA polymerases.

The position of enhancer sequences is variable. Enhancers have been reported to have activating effects upstream (often at kilobase distances), downstream or in the spliced introns of the gene. Enhancers often appear to act as 'magnets' for DNA-binding transcription factors that, once attached to the DNA in question, 'migrate' along the DNA helix until a suitable binding site is found. It must be remembered that the binding of transcription factors is under control of stochastic processes and so any process that attracts these factors and direct them to the desired sites will increase the likelihood of transcriptional initiation. Other possible mechanisms of enhancer action include the "looping-

out" hypothesis: transcription factors remain bound to the enhancer elements but interact directly with the transcription complex through the "looping out" of the intervening DNA.

In addition to these relatively common DNA sequence motifs, a number of other motifs are known (see Section 4.3(iii)). These motifs are often characterised by homo-purine or homo-pyrimidine sequences (or combinations of these) that have the potential of forming H-DNA or Z-DNA helices; or palindromic sequences that may bind dimeric transcription factors.

As stated above, TATA-boxes are not always found in eukaryotic promoters. Such promoters are often found involved in "house-keeping" genes, i.e. genes that are active in most cells and that perform non tissue-specific functions (e.g. glucose metabolism).

It has been reported that the 5'-ends of vertebrate genes are often associated with an increase in (G+C) content, in particular with an increase in the frequency of the dinucleotide CpG (BIRD, 1986). Such regions can be identified experimentally by the presence of "HTF islands", i.e. a multiplicity of CCGG sequences that act as substrates for the restriction enzyme HpaII, producing a number of small DNA fragments (HpaII Tiny Fragments). According to Bird (*ibid.*), there is no relationship between the presence or absence of TATA boxes and CpG-rich 5'-DNA regions.

#### 4.3(ii) Neural-specific promoter elements

It has been widely hoped that promoter-DNA sequence comparisons would readily illuminate common sequence motifs that could be implicated as transcription factor binding sites. While a number of such DNA-motifs have come to light (see FAISST and MEYER, 1992, for a compilation of vertebrate-encoded transcription factors and their binding sites), neural-specific DNA sequences are still relatively elusive.

A number of neural promoters have been studied, including those of neuron-specific enolase (SAKIMURA *et al.*, 1987), SCG10 (WUENSCHHELL *et al.*, 1990), dystrophin (BOYCE *et al.*, 1991), nerve growth factor (ZHENG and HEINRICH, 1988), myelin basic protein (TAMURA *et al.*, 1990) and brain creatine kinase (HOBSON *et al.*, 1990). In the majority of published papers, however, the only identified transcription-factor binding-sites are TATA, CCAAT and GC-rich regions harbouring potential Sp1-binding sites.

A neural-specific "identifier sequence" was reported to be present in a number of neural genes (SUTCLIFFE *et al.*, 1984); this was later correctly identified as being a small mRNA for rat brain myelin proteolipid protein (MILNER *et al.*, 1985).

The comparison of rat GAP-43, type II Na<sup>+</sup>-channel, peripherin and SCG10 gene promoters (all neural-specific genes) has, however, lead to the identification of a 7bp element common to all of these promoters (CCAGGAG), with additional high homology in some of the flanking sequences (NEDIVI *et al.*, 1992). It can only be hoped that such neural-specific motifs can be elaborated upon and extended to other species.

In an analysis of the promoter regions of neural genes in vertebrates and invertebrates (BATLEY, 1992), at least one of a small number of transcription factor binding sites were found to be present in most of these genes: E2F, a vertebrate factor that complexes with the retinoblastoma gene product (BAGCHI *et al.*, 1990); GATA factor, thought to play a regulatory role in chick brain and T-cell specific expression (YAMAMOTO *et al.*, 1990); *MyoD*, found in proliferating myoblasts and differentiated myotubules (MURRE *et al.*, 1989); and NF- $\kappa$ B, a relative of the *rel*-oncogene (MAJELLO *et al.*, 1990). Some of these transcription factors and others are mentioned below in the discussion of the N-Cadherin promoter DNA sequence (Section 7.5).

#### 4.3(iii) Transcription factors

As mentioned above, the presence of suitable DNA elements (e.g. TATA/CCAAT boxes) and RNA polymerases is not sufficient to initiate the transcription of most eukaryotic genes. The binding of tissue- and temporal-specific factors is necessary in order to activate the transcription complexes. It has been found that some factors are capable of binding to more than one DNA-binding site; additionally, DNA-binding sites are known that can be bound by more than one factor. Thus there will exist competition between factors for DNA-binding sites that will add an extra tier of complexity to the regulatory mechanisms.

Included in the definition of transcription factors are not only proteins, but also other protein complexes that may bind (directly or indirectly) to the promoter elements. Often the status of intra-cellular metabolites or the presence of extra-cellular hormones is communicated to the transcriptional machinery by the binding of the metabolite/hormone to a protein receptor, thus activating the receptor and allowing the receptor-complex to bind to the appropriate DNA element.

Whilst sequence comparisons of eukaryotic promoters have led to the identification of the common DNA sequence elements described above, the transcription factors that bind to and regulate the transcription of these promoters require more complex experimental procedures for their identification.

As a first step (after the identification of a putative promoter sequence), gel retardation and DNA-footprinting studies can be used to define the DNA sequences that are bound by the activating proteins. The attachment of various lengths of the putative promoter to readily assayable reporter genes (such as chloramphenicol acetyl transferase or CAT,  $\beta$ -galactosidase and luciferase) is

also used to identify important parts of the promoter that confer tissue-specificity, temporal specificity or high/low levels of promoter activity.

The study of oncogenes has lead to the identification of a number of transcription factors and contributed to the unification of the studies of cancer and gene regulation (BRADSHAW and PRENTIS, 1987).

Many transcription factors have now been sufficiently purified to enable identification and assignment to families based on DNA-sequence binding similarities. Such transcription factors are extensively reviewed elsewhere, for eukaryotes (JOHNSON and MCKNIGHT, 1989), mammals (MITCHELL and TJIAN, 1989), in the brain (HE and ROSENFELD, 1991) in *Xenopus* (WOLFFE, 1991). A list of the main transcription-factor DNA-binding motifs is given in Table IV.1.

Transcriptional regulation is described further below by reference to the NCAM and E-Cadherin promoters.

#### 4.4 NCAM promoter

##### 4.4(i) Introduction

As discussed in Chapter II, the expression of NCAM is tightly defined in terms of both the protein's tissue distribution and the developmental stages during which it is expressed. During embryonic development, NCAM is expressed on tissue derivatives of all three germ layers. In tissues involved in embryonic induction, however, NCAM is often seen transiently, such expression having sharply defined borders. This has prompted speculation that NCAM may be an initiator or major participant in such induction processes.

The NCAM protein is known to be encoded by a single gene, the multiple protein isoforms being produced by the use of alternative exon combinations and of different polyA+ addition sites.

##### 4.4(ii) NCAM promoter clones

Table IV.1

Classification of DNA-binding proteins based on  
protein motifs

Motif:	Reference:
1. Homeodomain	SCOTT <i>et al.</i> , (1989)
2. POU domain	ROSENFELD, (1991)
3. Paired box	KESSEL and GRUSS, (1990)
4. Nuclear-receptor-type zinc finger	BEATO, (1989)
5. TFIIIA-type zinc finger	KLUG and RHODES, (1987)
6. Leucine zipper	LANDSCHULZ <i>et al.</i> , (1988)
7. Helix-loop-helix	MURRE <i>et al.</i> , (1989)
8. "ets" homology	KARIM <i>et al.</i> , (1990)
9. "rel-NF- $\kappa$ B-dorsal" homology	KIERAN <i>et al.</i> , (1990)
10. "SRF-MCM" homology	PASSMORE <i>et al.</i> , (1988)
11. "Fork head-NF3 $\alpha$ homology	LAI <i>et al.</i> , (1991)
12. "HMG" homology	GUBBAY <i>et al.</i> , (1990)

The above classes of DNA-binding proteins are based on conserved amino-acid motifs found within each member of the class. The three-dimensional structures of some of these motifs are known or predicted (e.g. helix-loop-helix).

Some DNA-binding proteins do not conform to one of these classes; further classes are therefore likely to be discovered in due course.

In order to study the transcriptional regulation of this gene, the NCAM genomic DNA sequences have been cloned in mouse (HIRSCH *et al.*, 1990), rat (CHEN *et al.*, 1990) and human (BARTON *et al.*, 1990).

All three cloned promoters can be characterised by the following features: no functional TATA- or CCAAT-boxes; presence of potential Sp1-binding sites; multiple transcriptional initiation sites; a (G+A)-rich region; and an (A+T)-rich region. To avoid undue repetition, only the mouse NCAM promoter will be discussed in detail here.

#### 4.4(iii) Mouse NCAM promoter

The mouse NCAM genomic DNA sequences were obtained by screening a mouse genomic DNA library with a 5' fragment of the mouse cDNA (*ibid.*). Transcriptional initiation sites were defined by primer extension and S1-nuclease protection assays. Regions containing promoter activity were identified initially by promoter-CAT constructs; gel-retardation and DNA-footprinting were subsequently used to define the transcription-factor binding sites with higher resolution. The results of these experiments are summarised in Figure IV.1.

#### 4.4(iii) Transcriptional initiation

As is often the case in genes lacking TATA-boxes, more than one transcription-initiation site was identified. Two sites were seen, at -193bp and -336bp (numbering relative to translational start site). Use of these two sites was not tissue-dependent as the same results were obtained using embryonic or adult brain, C2-muscle and N2A-neuroblastoma cell lines. The only TATA-like box in the first 1Kb upstream of the translational start site lies at -994bp and hence is regarded as being too great a distance away from the above-defined transcriptional start-sites to act as a usual promoter element.

#### 4.4(iii) Poly hetero-nucleotide regions

Figure IV.1a

NCAM promoter organisation and deletion constructs

Eight footprints were mapped within the first 1Kb upstream of the translational start site (+1). Two main transcriptional start sites were mapped to -336 and -193. The (A+T)-rich and (GGA)<sub>n</sub>-domains are also shown.

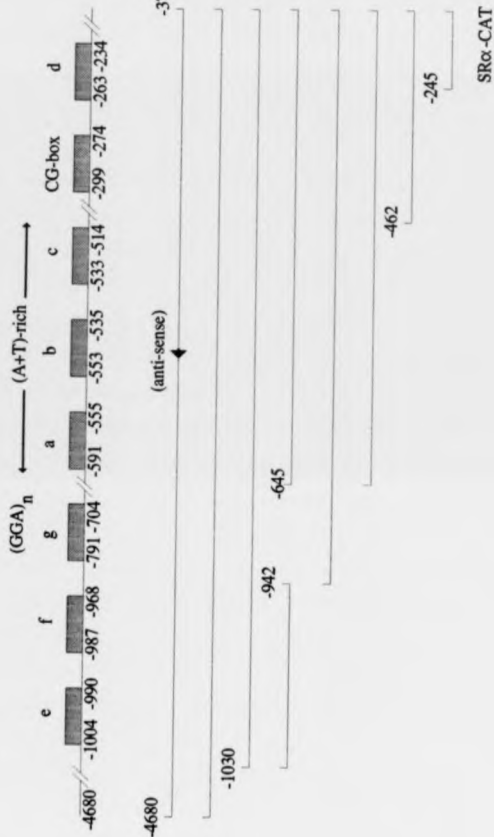
Deletion constructs fused to the CAT -gene were transfected into N2A-neuroblastoma cells and L-cells (as a negative control). The average %-promoter activities are shown on the right, relative to a composite SV40-HTLV-I promoter (SR $\alpha$ -CAT) which was used as a positive control.

From this data it can be seen that the strongest promoter activity is provided by elements between -645 and -245 but these elements lack the capacity to provide tissue specificity. DNA regions upstream of these elements act to reduce this activity and impose tissue-specific constraints, so that expression in (non-specific) L-cells is reduced relative to N2A cells.

(Data from HIRSCH *et al.*, 1990)

Figure IV.1b

NCAM promoter organisation and promoter activity of deletion constructs



In between the two transcriptional start-sites, spanning -299bp to -274bp, lies a GC-rich region having putative Sp1 binding site homology. Three other regions of this promoter are also dominated by specific nucleotide pairs: -1004bp to -970bp by A/T; -813bp to -671bp by A/G (including a (GGA)<sub>15</sub> stretch); and -591bp to -519bp by A/T. Such poly-heteronucleotide domains are known to affect the conformation of the DNA helix, although the significance of this fact (with respect to the effect that it has on transcription factor binding) has not yet been resolved. It can readily be appreciated, however, that an alteration of DNA helix conformation could provide a protein-recognisable domain and/or bring together two previously-distant protein-DNA complexes whose interaction could trigger a transcription-initiation event.

#### 4.4(iii) Deletion constructs

A series of promoter-CAT deletion constructs were made and transfected into both neuroblastoma-N2A cells (which express NCAM) and L-cells (which do not). Variations in CAT-plasmid transfection efficiencies were corrected for by co-transfection with the pCH110 ( $\beta$ -galactosidase) plasmid and normalising CAT expression against  $\beta$ -galactosidase expression.

These experiments illuminated the presence of DNA regions that provided positive and negative contributions to protein specificity and overall promoter activity. The results are illustrated and discussed in Figure IV.1.

#### 4.4(iiiid) Protein footprints

Eight protein-complex footprints have been defined for this promoter region by gel-retardation assays and DNaseI footprinting; most of these map to these repeated-nucleotide regions (see Figure IV.1).

"Footprint g", which maps to the (GGA)<sub>n</sub> region, contains a silencer or repressor element. Loss of this domain results in the loss of tissue-specificity as seen in the promoter-CAT experiments.

"Footprint b" contains an (A+T)-rich element (ATTATTA) that possesses sequence homology to the central core of most of the *Antennapedia* homeodomain protein binding sites. Certain *Drosophila* homeodomain proteins (*Antp*, *zen* but not *en*) have been shown to bind, *in vitro*, to this mouse NCAM element (HIRSCH *et al.*, 1991) suggesting that the NCAM gene may be subject to the control by homeodomain proteins.

It is hoped that, before too long, the factors responsible for the other footprints will also be identified, providing a valuable picture of the mechanisms controlling this gene.

#### 4.5 Cadherin genomic DNA organisation

##### 4.5(i) Mouse N-Cadherin gene organisation

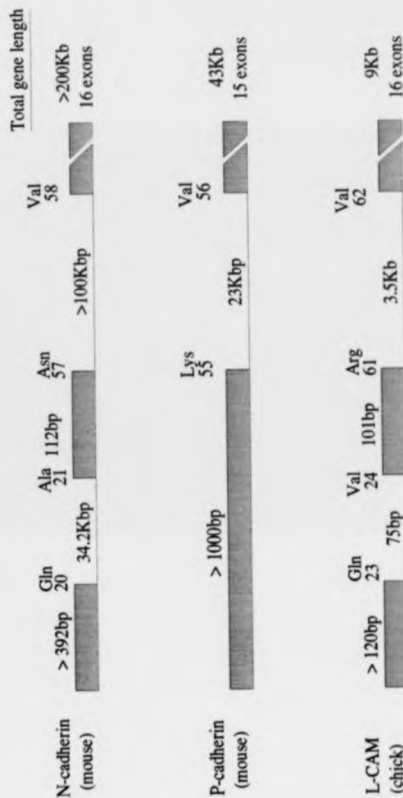
The intron/exon structure and chromosomal localization of the mouse N-Cadherin gene has recently been reported by Takeichi's group (MIYATANI *et al.*, 1992). This gene comprises 16 exons that span >200kb.

A comparison of the intron/exon boundaries of this gene with other known Cadherins (see below) has shown that these boundaries are remarkably conserved in all the Cadherin genes published to date (the one exception being P-Cadherin, whose first exon comprises the first+second exon of the others). The 5'-ends of these genes are compared in Figure IV.2. From a comparison of these boundaries and the protein domains, it can be stated that the exons do not correspond to the division of the protein into recognisable domains.

During the cloning of the N-Cadherin genomic DNA, two clones were found that corresponded to exon 16 but had different restriction-enzyme digest patterns. One clone overlapped with exon 15; the other also had 100% homology to the coding sequence and 99% homology in the 3'-untranslated region, but could not be mapped relative to the rest of the N-Cadherin genomic DNA. It would appear therefore the second exon 16 is a duplication of the first.

Figure IV.2

Comparison of the 5'-ends of three Cadherin genes



5'-ends of first exons based on cDNA sequence comparison. Exons are shown as boxed elements. Amino-acid residues and corresponding numbers are given at each intron-exon boundary. (Data from MIYATANI *et al.*, (1992), HATTA *et al.*, (1991), SORKIN *et al.*, 1988).

and a very recent one at that. This second exon 16 may provide an explanation for the fact that three bands are normally seen on a N-Cadherin Northern blot - 5.3, 4.3 and 3.5Kb, the middle band being the major one. Further experimental evidence is required to confirm this speculation.

#### 4.5(ii) Mouse P-Cadherin gene organisation

Takeichi's group have also published the genomic organisation of the mouse P-Cadherin gene (HATTA *et al.*, 1991). This paper parallels the N-Cadherin paper (*ibid.*) in that the intron/exon boundaries and chromosomal localization of P-Cadherin are reported.

In contrast to the N-Cadherin gene (which is on chromosome 18), P-Cadherin is tightly linked to E-Cadherin on chromosome 8. The linkage of Cadherin genes is a point that is returned to below.

As stated above, the intron/exon boundaries of P-Cadherin are highly conserved except for the first P-Cadherin exon comprising the first two exons of the other Cadherins. The authors also make the interesting comment that the first intron of the P-Cadherin gene (23Kb) appeared to possess enhancer activity.

#### 4.5(iii) Mouse L-CAM and K-CAM gene structures

The genomic structure of the mouse L-CAM gene has been reported (SORKIN *et al.*, 1991). Surprisingly, an analysis of the region 5' to exon 1 of the L-CAM gene has revealed the presence of a second Cadherin, termed K-Cadherin (believed to correspond to B-Cadherin). Only 700bp separate the polyA-addition site of the K-CAM gene and the translational start-site of the L-CAM gene. Although these two genes have distinct specificities and significant DNA sequence changes, the physical closeness and high degree of intron/exon boundary conservation between these two genes makes it very likely that they arose by gene duplication.

The intervening 700bp between these two genes contains no TATA- or CCAAT-boxes which could act as possible promoter elements for the L-CAM gene. CAT and  $\beta$ -galactosidase constructs attached to this 700bp also failed to show any sign of promoter activity when transfected into the appropriate tissue-culture cells. Based on this, and inconclusive primer extension assays, the authors stated that they believed the promoter of the L-CAM gene to be further upstream than the 3'-end of the K-CAM gene. (It is interesting to note that the authors reported that primer extension experiments failed to produce any viable result. They attributed this possibly to a high GC content in the region of interest.)

#### 4.5(iv) E-Cadherin promoter

The epithelial CAM, E-Cadherin is one of the most intensely studied of the Cadherins due to its possible role in determining the invasiveness of carcinomas. In order to study the regulation of this molecule, Behrens *et al.*, have cloned the mouse E-Cadherin promoter (BEHRENS *et al.*, 1991). Through the use of deletion/CAT constructs, DNase I footprinting and gel retardation assays two regions of the promoter have been identified as being involved in the regulation of this gene (see Figure IV.3).

The single transcriptional start site was defined by primer extension as being 127nts upstream of the translational start site. No consensus TATA box was identified, although the sequence spanning the transcriptional start site was noted as having homology to other "initiator sequences" described in other TATA-less promoters. A CAAT box was present at -65bp; (quoted base pairs are relative to the transcriptional start site). The promoter region also comprises a GC-rich region at -58 to -25bp, harbouring a putative Sp1 binding site; and a 12bp palindromic sequence bounded by 4bp inverted repeats at -86 to -75, termed "E-pal". This latter sequence has homology to keratin gene promoter

Figure IV.3a

E-Cadherin promoter  
and the activities of promoter-deletion/CAT constructs

The structure of the mouse E-Cadherin promoter is shown illustrating the positions of the E-pal and GC-boxes.

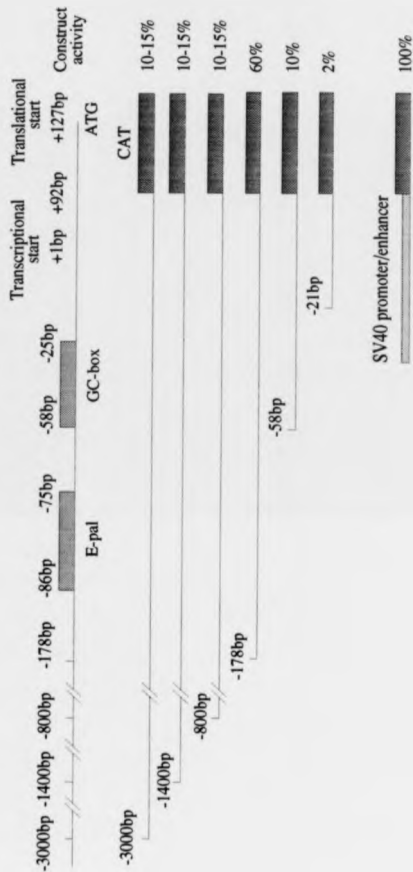
The Figure also shows the construction of six promoter-deletion constructs that were fused to the CAT reporter gene and transfected into epithelial cells. CAT activity was normalised relative to an SV40-promoter/enhancer-CAT construct.

These results show the repressive effects of elements 5' to -178bp and the positive effects of the E-pal and GC-boxes. Experiments illustrating the tissue-specific properties of the E-pal element are referred to in the main text.

(Data from BEHRENS *et al.*, 1991)

Figure IV.3b

E-Cadherin promoter and activities of promoter-deletion/CAT constructs



sequences in *Xenopus* and human.

Deletion constructs were used to determine the roles played by these elements in the transcription of a cloned CAT gene. Relative to the expression of a SV40 promoter/enhancer construct (arbitrarily defined as 100%), the 3Kb of DNA 5' to the transcriptional start site had only 10-15% activity (see Figure IV.3). A set of deletion constructs, all having the same 3'-end (at +92bp), illuminated the presence of a negative silencer region 5' to -178bp; and that the E-pal element (and to a lesser extent the GC-box) both had positive promoter activity. Gel retardation and DNA footprinting assays both confirmed the presence of protein complexes covering both (independently) E-pal and GC-box elements.

Cloning variable numbers of E-pal elements upstream of an SV40 promoter conferred epithelial-cell specific activity in a dose-dependent manner: transfection of this construct into epithelial cells lead to a 2-20 fold stimulation; transfection into fibroblasts and smooth muscle cells lead to a 2-5 fold decrease (relative to the SV40 promoter alone).

The E-Cadherin promoter therefore conforms with the currently accepted modular view of promoters, in that the promoter comprises a number of positive and negative regulatory sequences which combine to confer correct temporal and spatial expression on the gene.

#### 4.6 Conclusions

The eukaryotic promoter has shown itself to be a complex region of DNA where multi-factor complexes jostle for key regulatory sites. The result is the suppression or activation of the cell's transcriptional machinery, depending on the factors present or absent.

Characterization of the NCAM promoter (which regulates the transcription of a molecule having a similar specificity to N-Cadherin) is at an advanced stage, with the identification of eight protein-binding sites, one of which possibly

being the site of binding of a homeobox protein. The further identification of the transcription factors involved in the regulation of this gene can only serve to enhance our knowledge of the control mechanisms underlying neural transcription in general. Armed with such knowledge, defects in neural regulatory processes may be better understood and as well as hopefully illuminating the ways that such defects may be corrected.

E-Cadherin is the only Cadherin whose promoter sequence has been published to date. Like the NCAM promoter, the E-Cadherin promoter also lacks a TATA-box. A number of DNA elements that may be involved in the regulation of the gene have been identified.

Although the Cadherins as a family show remarkable coding sequence conservation, which even extends to the intron/exon boundaries, this conservation would not be expected to continue in the promoter regions given the diverse expression patterns of the Cadherin proteins.

The study of proteins that form part of a family or that have restricted tissue distributions can provide information, the usefulness of which extends far beyond that specific protein. Through a comparison with the other members of the Cadherin family, the study of N-Cadherin may provide additional information regarding the history and divergence of this family. By a comparison with molecules such as NCAM, the study of the N-Cadherin may also be useful to further add to the general body of knowledge on neural regulatory mechanisms.

By using a system such as *Xenopus*, which is highly amenable to the developmental biologist, such regulatory processes may be followed both quantitatively and qualitatively i.e. promoter activity may be measured numerically and followed visually. The following study therefore has aimed to

take full advantage of the *Xenopus* system in starting to elucidate the processes involved in the regulation of the *X.borealis* N-Cadherin gene.

## Chapter V

### Materials and methods

- 5.1 General remarks
- 5.2 Restriction-enzyme digests
- 5.3 Phenol/chloroform-extraction and EtOH-precipitation
- 5.4 Dephosphorylation of DNA ends
- 5.5 Blunting of restriction-enzyme digested DNA ends with Klenow I
- 5.6 Agarose-gel electrophoresis
- 5.7 LMP-agarose gel electrophoresis and isolation of DNA bands
- 5.8 Ligations of DNA fragments using T4 DNA ligase
- 5.9 Preparation of competent bacterial cells
- 5.10 Transformation of competent bacterial cells with DNA plasmids
- 5.11 Small-scale plasmid minipreps
- 5.12 Large-scale DNA plasmid preparations
- 5.13 End-labelling of DNA fragments with Klenow I and  $^{32}\text{P}$ -dNTPs
- 5.14 Isolation and cleaning of oligonucleotides
- 5.15 End-labelling of oligonucleotides with T4 polynucleotide kinase
- 5.16 Primer extension of labelled oligonucleotides
- 5.17 Polyacrylamide-gel electrophoresis
- 5.18 RNA extraction from *Xenopus* embryos
- 5.19 *In vitro* transcription of  $\alpha$ - $^{32}\text{P}$ -labelled RNA
- 5.20 RNase protection assays
- 5.21 Determination of %-label incorporated into DNA/RNA probes
- 5.22 Sephadex-column chromatography

- 5.23 Nick-translation of DNA fragments
- 5.24 Hybridisation of DNA probes to nitrocellulose filters
- 5.25 Preparation of plating bacteria
- 5.26 EMBL3 genomic library screening
- 5.27 Production of lambda-phage plate lysate
- 5.28 Lambda phage DNA minipreps
- 5.29 Single-stranded DNA (ssDNA) preps
- 5.30 Single-stranded DNA sequencing
- 5.31 Southern blots
- 5.32 Production of promoter deletion/ $\beta$ -galactosidase clones
- 5.33 Linearisation of DNA templates for micro-injection into *Xenopus*
- 5.34 Hormone-stimulated production of *Xenopus* eggs and *in vitro* fertilisation
- 5.35 Injection of DNA clones into *Xenopus* embryos
- 5.36 Staining of *Xenopus* for  $\beta$ -galactosidase expression

### 5.1 General Remarks.

The "standard methods" referred to below are obtainable from SAMBROOK *et al.* (1989), PERBAL (1988) or other similar readily-available laboratory manuals. Suppliers of reagents used are not given unless one particular manufacturer was preferred, using the quantities of enzyme recommended by the manufacturers. Centrifuging of Eppendorf tubes was always done at 12,000rpm on a standard Eppendorf bench centrifuge.

Unless otherwise stated, for plasmid work the bacteria used was BB4 (supF58, supE44, hsdR514 ( $r_k^-$ ,  $m_k^-$ ), galK2, galT22, trpR55, metB1, tonA, lambda-, D(arg-lac)U169 [F, proAB, lacIQZDM15, Tn10( $tet^R$ )], Stratagene, 1987) which was always grown in the presence of tetracycline (10µg/ml); ampicillin was also added (100µg/ml) when the bacteria was transformed with pBluescript. For EMBL3 library screening the *E.Coli* strain K803 (hsdR<sup>-</sup> hsdM<sub>k</sub><sup>-</sup> gal<sup>-</sup> met<sup>-</sup> supE) was used.

For RNA, work all laboratory-made RNA solutions were either treated with diethyl-pyrocabonate (DEPC) or made with DEPC-distilled water; glassware was baked at 250°C overnight and all reagents isolated from main laboratory stocks.

Extensive precautions were always taken when dealing with any radionucleotides.

### 5.2 Restriction-enzyme digests

Restriction-enzymes were used as per manufacturers' instructions.

using the buffers supplied. Complete digestion was always verified on an agarose-TBE gel (see below) by ensuring the complete absence of starting plasmid bands (run in an adjacent lane). In the case of double-digests, unless both enzymes worked with 100% efficiency in the same buffer, the digests were carried out consecutively with a phenol/chloroform-extraction/EtOH-precipitation step (see below) in between each digestion. For the second digest, digestion of the original plasmid (with the second enzyme) was included as a control, to monitor the extent of the second digestion.

#### 5.3 Phenol/chloroform-extraction and ethanol-precipitation

Phenol/chloroform-extraction was carried out as per standard methods. Ethanol (EtOH) precipitation was carried out using 1/10 vol. 3M sodium acetate pH 6.5 for DNA precipitations and 3M sodium acetate pH 5.2 for RNA precipitations unless otherwise stated. After centrifugation, pellets were washed with 80% EtOH and desiccated briefly.

#### 5.4 Dephosphorylation of DNA ends

Calf intestinal alkaline-phosphatase (CIAP) was used to remove terminal phosphates from restriction-enzyme digested DNA ends. The Zn/Mg/Tris buffer and reaction conditions were used as in SAMBROOK *et al.* (1987).

#### 5.5 Blunting of restriction-enzyme digested DNA ends with Klenow I

Restriction-enzyme digests were performed as above. The DNA was then phenol/chloroform-extracted and EtOH-precipitated; the DNA

was resuspended in 20.5µl distilled water. To this was added 2.5µl Nick Translation Buffer (see below), 1µl dNTP mix (dA, dC, dG, dTTP at 2mM each, in distilled water), and 1µl Klenow I (6U/µl, BRL). The reaction was allowed to proceed for 30 minutes at room temperature and then 1µl EDTA pH 8.0 and 74µl TE pH 7.5 were added to stop the reaction. (TE used in all cases was 10mM Tris/HCl, 1mM EDTA pH 7.6). The solution was then phenol/chloroform-extracted and EtOH-precipitated with 0.5 vol. 7.5M ammonium acetate pH 7.5.

#### 5.6 Agarose-gel electrophoresis

Restriction-enzyme digested DNA was fractionated on 0.5-2% agarose-TBE gels using standard methods. Two main gel-sizes were used: minigels (10 x 8cm, Pharmacia) were used for routine monitoring of restriction-enzyme digest progress; Maxigels (14.5 x 19.5cm, Pharmacia) were used for high-resolution band-size determination and for Southern blots. Restriction-enzyme digested lambda-phage markers were used on all gels.

#### 5.7 LMP-agarose gels and isolation of DNA bands

Low melting point (LMP)-agarose gels were prepared and run using standard methods but the current was always limited to 50mA to avoid melting the gel. Bands were visualized for the minimum period possible on a UV-light box (to avoid damage to the DNA) and the smallest slice of gel that incorporated the DNA band of interest was taken into an Eppendorf tube. The tube was weighed and the volume

of gel (1 vol.) estimated (based on 1g=1ml). 4 vols. distilled water were added and the tube was incubated at 65°C for 5 minutes to melt the gel. An equal volume of phenol was then added, the tube was vortexed for 1/2 minute and then incubated on ice for 10 minutes. The tube was then centrifuged, and phenol/chloroform-extracted twice before EtOH precipitation, usually with 10µg tRNA carrier.

#### 5.8 Ligations of DNA fragments using T4 DNA Ligase

DNA and vector (50-100ng each) were combined in 6.5µl distilled water and heated for 5 minutes at 45°C, and then 5 minutes at 0°C (to dissociate any sticky ends). 1µl 10mM rATP, 2µl 5x T4 DNA Ligase Buffer (5x = 250mM Tris/HCl pH 7.6, 50mM MgCl<sub>2</sub>, 5mM DTT, 25% PEG-8000) and 0.5µl T4 DNA Ligase (BRL, 2 Weiss U/µl) were then added. Inter-molecular ligations were incubated at 4°C for 20 hours; intra-molecular ligations were incubated at 37°C for 6 hours without any PEG-8000 in the Ligase Buffer.

#### 5.9 Preparation of competent bacterial cells

Overnight cultures were seeded from 15% glycerol stocks or single bacterial colonies into 10ml LB (SAMBROOK *et al.*, 1989) + antibiotics. These were shaken overnight at 200rpm, 37°C, in a rotary incubator. Next morning, 100µl of overnight culture was added to 10ml fresh LB + antibiotics and shaken at 300rpm for 120 minutes. At the end of this time the OD<sub>600</sub> was approx. 0.3 (as previously determined by OD<sub>600</sub> v cell number titration experiments for BB4 bacterium). Bacterial cultures were then chilled on ice for 10 minutes,

and centrifuged at 200rpm for 5 minutes before the supernatant was decanted. Pellets were resuspended in 5ml ice-cold 100mM  $\text{CaCl}_2$  and incubated on ice for a further 30 minutes. Cultures were then centrifuged as before and resuspended in 1ml 100mM  $\text{CaCl}_2$  and stored at 4°C until use (invariably the same day).

#### 5.10 Transformation of competent bacterial cells with DNA plasmids

Competent cells were prepared as above. 1/2 ligation mix or 50ng plasmid were diluted into 25 $\mu\text{l}$  sterile distilled water (Tube A). 3 $\mu\text{l}$  of this was diluted into 27 $\mu\text{l}$  sterile distilled water into a second tube (B). Likewise, 3 $\mu\text{l}$  of Tube B was diluted into 27 $\mu\text{l}$  sterile distilled water for Tube C. (In this way, 1/10 and 1/100 dilutions of the original DNA solution were made). To each tube was added 200 $\mu\text{l}$  competent bacteria. Each tube was incubated on ice for 40 minutes before heat-shocking at 43°C for 2 minutes, incubating on ice for 5 minutes and then spreading out on LB-agar-antibiotic plates. Controls included cut vector minus insert, circular plasmid and untransformed-bacteria on antibiotic-resistance selecting plates. The ratio of [cut vector +plus insert]:[cut vector minus insert] determined the number of DNA preps subsequently made, which was usually between 4 and 12 for each clone.

#### 5.11 Small-scale plasmid minipreps

Transformed-bacterial cultures were seeded from glycerol stocks or from single colonies into 10ml 2xYT (SAMBROOK *et al.*, 1989) + antibiotics and shaken overnight at 200rpm. The following morning

15% glycerol stocks were taken from any new cultures (850µl culture + 150µl glycerol, mixed and stored -70°C). 1.5ml fractions were taken into 1.5ml Eppendorf tubes and centrifuged; the supernatant was discarded. The pellet was resuspended in 200µl ice-cold STET (8% sucrose, 0.5% Triton X-100, 50mM EDTA, 10mM Tris/HCl pH 8.0) before 10µl 10mg/ml lysozyme/distilled water was gently mixed in by inversion. This was incubated at 4°C for 10 minutes, boiled at 100°C for 40 seconds and then centrifuged for 10 minutes. The bacterial DNA/cell-debris 'blob' was removed with a toothpick before precipitation with 200µl isopropanol + 20µl 3M Na acetate pH 6.5. After centrifugation, the DNA was resuspended in 200µl TE, phenol/chloroform-extracted twice and then EtOH-precipitated with 100µl 7.5M ammonium acetate + 500µl EtOH. Routinely, 2x 1.5ml cultures produced approximately 2µg plasmid DNA.

#### 5.12 Large-scale DNA preparations

Large-scale DNA preparations were made using the lysozyme/SDS/NaCl method followed by centrifugation through CsCl gradients and subsequent dialysis into TE (SAMBROOK *et al.*, 1989). This 'gentle' method was used because it could produce high-molecular weight plasmids possessing the minimum of nicks and loss of supercoiling.

#### 5.13 End-labelling of DNA fragments with Klenow I and $\alpha$ -<sup>32</sup>P-dNTPs

This method was used to label recessed 3'-termini of restriction-enzyme digested DNA fragments; it was also used for labelling DNA

markers. A restriction-enzyme digest was carried out as above in a 30µl reaction mix. 2µl  $\alpha$ -<sup>32</sup>P-dGTP or  $\alpha$ -<sup>32</sup>P-dCTP (together with the remaining cold dNTPs at 100µM final concentration, as necessary) and 1µl Klenow I (6U/µl, BRL) were added. The reaction was incubated at room temperature for 20 minutes before separation of unincorporated counts by Sephadex-column chromatography. The %-labelled dNTP incorporation of each 0.5ml fraction was determined as below.

#### 5.14 Isolation and cleaning of oligonucleotides

Synthetic oligonucleotides were made on a Applied Biosystems-oligonucleotide synthesizer using standard methods and supplied in a dessicated form. Oligonucleotides were cleaned by butanol extraction before being isolated by UV-shadowing from 15% denaturing polyacrylamide gels (SAMBROOK *et al.*, 1989).

#### 5.15 End-labelling of oligonucleotides with T4 Polynucleotide Kinase

The following were combined in a 0.5ml Eppendorf: 1µl oligo (10 pmoles/µl), 2µl 10x T4 Kinase Buffer (10x = 0.5M Tris/HCl pH 7.6, 0.1M MgCl<sub>2</sub>, 50mM DTT, 1mM spermidine.HCl, 1mM EDTA), 2µl  $\gamma$ -<sup>32</sup>P-rATP (40µCi), 14µl distilled water and 1µl T4 DNA Polynucleotide Kinase (6U/µl, BRL). The reaction was incubated at 37°C for 45 minutes, 65°C for 10 minutes and then 1µl 0.5M EDTA pH 8.0 and was added. The terminated reaction was run down a Sephadex-G-50-150 column (see below) and eluted with TE. 0.5ml fractions were taken, each being assessed for %-incorporation by the

method below. 1/10 vol. sodium acetate pH 6.5 and 10 $\mu$ g tRNA were added to the selected fractions before each was individually phenol/chloroform-extracted and EtOH-precipitated.

#### 5.16 Primer extension of labelled oligonucleotides

Oligonucleotides were prepared and labelled as above. Labelled oligonucleotide (50,000cpm) and RNA were combined in an Eppendorf tube and EtOH-precipitated together. After centrifugation, washing in 80% EtOH, recentrifugation and brief dessication, the oligo/RNA mix was resuspended in 10 $\mu$ l Aqueous Hybridisation Buffer (0.4M NaCl, 10mM PIPES pH 6.4, 0.5mM EDTA) and sealed into a glass-capillary tube. After denaturation at 85°C for 10 minutes, the reaction mix was incubated in a water-bath at the desired temperature overnight (usually 40-65°C). In the morning, the capillary-tube contents were expelled into 42.5 $\mu$ l Aqueous Hybridisation Buffer and 110 $\mu$ l EtOH added; this was stored at -20°C for 30 minutes. After centrifugation and 80% EtOH wash, the pellet was resuspended in 20 $\mu$ l RTase Buffer (50mM Tris/HCl pH 8.3, 60mM KCl, 10mM MgCl<sub>2</sub>, 1mM dA, dC, dG, dTTP, 1mM DTT, 50 $\mu$ g/ml actinomycin D, 1 $\mu$ l human placental RNase inhibitor (10U/ $\mu$ l), 0.5 $\mu$ l AMV reverse transcriptase (Life Sciences Inc, 17.4U/ $\mu$ l). The reaction was incubated at 42°C for 2 hours. 1 $\mu$ l 0.5M EDTA pH 8.0 and 1 $\mu$ l 200 $\mu$ g/ml RNase A were then added and the reaction incubated at room temperature for 15 minutes. 80 $\mu$ l TE was added followed by phenol/chloroform-extraction and EtOH-precipitation

with 5µg tRNA. Pellets were resuspended in 3µl Formamide Loading Buffer (formamide, 0.1% xylene cyanol, 0.1% bromophenol blue, 10mM EDTA pH 8.0), and run on 5-8% acrylamide, 7M urea, 0.5x TBE gels after denaturation of the DNA by boiling at 95°C for 5 minutes.

Labelled oligonucleotide + synthetic sense-strand RNA/embryo RNA was run as a positive control.

#### 5.17 Polyacrylamide gel electrophoresis

Polyacrylamide-gels were run as per standard methods. Gel reagent stocks were filtered prior to use; gels were routinely pre-run for 15-30 minutes prior to use. After running, gels were fixed in 10% acetic acid/10% EtOH/distilled water and then dried-down prior to exposure at -70°C with X-ray film (Fuji Medical X-ray Film RX). Labelled DNA markers were always run: either <sup>32</sup>P-labelled-pBR322/HpaII or <sup>35</sup>S-labelled DNA sequence were used.

#### 5.18 RNA extraction from *Xenopus* embryos

20 healthy embryos (of a distinct stage, but always below stage 30) were selected and washed with fresh 1/10 BarthX (1x = 88mM NaCl, 24mM NaHCO<sub>3</sub>, 15mM Tris/HCl pH7.5, 0.33mM Ca(NO<sub>3</sub>)<sub>2</sub>, 0.41mM CaCl<sub>2</sub>, 0.8mM MgSO<sub>4</sub>) in a 1.5ml Eppendorf tube. The supernatant was decanted and 0.8ml Buffer A (50mM Tris/HCl pH 7.6, 50mM NaCl, 10mM EDTA, 0.5% SDS) at 37°C was added and then 10µl Proteinase K (20mg/ml, freshly made). Embryos were homogenised by rapid pipetting with a 1.5ml Gilson pipette-tip. The homogenate

was incubated for 60 minutes at 37°C, split into 2 tubes, phenol/chloroform-extracted and then EtOH-precipitated with 0.3M sodium acetate pH 5.2. After centrifugation and an 80% EtOH-wash, each pellet was resuspended in 400µl TE. 400µl ice-cold 8M LiCl (RNase-free) was then immediately added. After a 3-12 hour precipitation at -20°C, the pellet was centrifuged as before, washed, and resuspended in 400µl TE. A 15µl fraction was taken for gel-electrophoresis and RNA concentration determination. 1ml EtOH was added to the remainder and then it was stored -70°C until required.

The LiCl treatment removed a large amount of contaminating DNA and glycoprotein. The RNA, when run on agarose-TBE gels, routinely produced discrete strong rRNA and weaker tRNA/5SRNA bands.

For embryos more advanced than stage 30 and for *Xenopus* brain, the RNA extraction procedure of Conn (1989) was used. This guanidine.HCl-method is incompatible with the yolky, early-stage embryos.

#### 5.19 *In vitro* transcription of $\alpha$ -<sup>32</sup>P-labelled RNA

The protocol used was a revised version (Paul Krieg, pers. comm.) of Krieg and Melton (1989). Plasmid DNA was always derived from CsCl-gradient purified DNA preparations. Particular attention was paid to the requirement for the complete linearization of the DNA template and the avoidance of leaving 3'-overhanging DNA ends. *In vitro* transcriptions were performed with  $\alpha$ -<sup>32</sup>P-UTP (60µCi/reaction), no 'cold' UTP, and T3/T7 RNA polymerase (50U/µl, BRL). Prior to

the addition of the polymerase, the reaction was cooled to 4°C for 5 minutes; polymerase was then added and the reaction continued at 4°C for 90 minutes. The reaction was DNAsed, an aliquot was taken for %-incorporation-determination and then the full-length probe was isolated from a 5% acrylamide, 8.3M urea, 1x TBE gel. Elution of probe from the gel was into 0.3M sodium acetate pH 5.2, 1mM EDTA, 0.1% SDS; sufficient elution of probe from the isolated gel-slice took place in just 45 minutes.

Sense-RNA *in vitro* transcriptions were made using similar protocols but with unlabelled UTP. These reactions were DNAsed, but no attempt was made to purify the full-length transcript.

#### 5.20 RNase protection assays (RPAs)

This protocol was also taken from Krieg and Melton (1989). Gel-isolated labelled-RNA probe was added to target/control RNA and were EtOH-precipitated together. After centrifugation/wash, they were resuspended in 30µl Hybridisation Buffer (80% formamide, 0.4M NaCl, 40mM PIPES pH 6.4, 1mM EDTA) in 1.5ml Eppendorfs. After heating to 85°C for 10 minutes, the RNA/probe were allowed to hybridise overnight in a 45°C water bath. Following hybridisation, 300µl RNase Digestion Buffer [300mM NaCl, 10mM Tris/HCl pH 7.5, 5mM EDTA, 10µg/ml RNase A, 200U/ml RNase T1 (Boehringer Mannheim)] was added; digestion was carried out for 30 minutes at 37°C. 15µl 10% SDS and 2.5µl 20mg/ml Proteinase K were added and the reaction was incubated for a further 15 minutes at

37°C. The reaction was then phenol/chloroform-extracted and EtOH-precipitated. After centrifugation, the pellet was resuspended in 5µl Formamide Loading Buffer (see above) and run on an 8.3M urea, 1x TBE, 5-8% polyacrylamide-gel.

In most experiments the corresponding sense-RNA (+probe) was added as an extra control to mark the length of the completely-protected cDNA fragment; (the probe-alone also contains transcribed vector sequences and so is larger than the cDNA). This sense-RNA was usually added against a background of embryo RNA to control for adverse substances in the RNA prep that might have prevented the assay from working correctly.

#### 5.21 Determination of %-label-incorporated into DNA/RNA probes

The disodium hydrogen phosphate ( $\text{Na}_2\text{HPO}_4$ ) precipitation method was used (SAMBROOK *et al.*, 1989). A sample of the labelling reaction was diluted 1/100 in TE and four 1µl fractions spotted onto squares of Whatman DE-81 (DEAE) paper. Two squares were kept as unwashed controls; the other two were washed consecutively for 2 minutes each in: 0.15M  $\text{Na}_2\text{HPO}_4$  (5x), distilled water (1x), 50% acetone/50% EtOH (1x). The filters were then counted on a standard liquid-phase scintillation counter. The %-incorporation was determined from the formula  $[\text{counts after washes}]/[\text{counts before washes}] \times 100$ . Typical incorporation values of radionucleotides into RNA probes were 70 - 80%.

#### 5.22 Sephadex column chromatography

Sephadex G-50-150 (Sigma) was used in some cases for the separation of unlabelled radionucleotides from those incorporated into DNA. Columns were run in a 10ml glass pipette and eluted with TE. 3 $\mu$ l xylene cyanol (XC) was added to 'hot' samples prior to loading: (XC comigrates with the unincorporated radionucleotides and therefore marked their position). 0.5ml fractions were taken; all were analysed for %-incorporation as above.

#### 5.23 Nick-translation of DNA fragments

The following were combined in a 1.5ml Eppendorf tube: 2 $\mu$ l 10x Nick Translation Buffer (10x = 0.5M Tris/HCl pH 7.8, 50mM  $\beta$ -mercaptoethanol, 50mM MgCl<sub>2</sub>, 0.5mg/ml BSA (#5, Sigma)), dNTP Mix (dA, dC, dTTP at 1mM each), 3 $\mu$ l  $\alpha$ -<sup>32</sup>P-dGTP, 1 $\mu$ l DNA Polymerase I (2U/ $\mu$ l, BRL), 1 $\mu$ l DNase I (20ng/ml), 50-100ng DNA, and distilled water to 20 $\mu$ l. The reaction was incubated at 14-16°C for 2-3 hours prior to phenol/chloroform-extraction and fractionation down a Sephadex G-50-150 column.

#### 5.24 Hybridisation of DNA probes to nitrocellulose filters

Nitrocellulose filters (Hybond C, Sigma) were rehydrated in 6x SSC. Filters were incubated in Prehybridisation Buffer (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% BSA (#5, Sigma), 6x SSC, 0.5% SDS, 50% deionised formamide, 100 $\mu$ g/ml salmon sperm DNA, in distilled water) for 2-4 hours at 42°C. Filters were then incubated in Hybridisation Buffer (same as Prehybridisation Buffer except with the addition of 10mM EDTA and labelled probe) overnight at 42°C. In the

morning, the solution was decanted and the filters washed in Prewash Solution (2x SSC, 0.5% SDS at room temperature, 2x 5 minutes each) prior to the main higher-stringency washes as detailed in the Results chapters.

#### 5.25 Preparation of plating bacteria

A culture was seeded from a single bacterial colony or glycerol-stock of *E.Coli* strain K803 into LB + 10mM MgSO<sub>4</sub>. This was grown at 37°C, 200rpm overnight. It was then centrifuged at 2000rpm for 5 minutes and the pellet resuspended in 4ml 10mM MgSO<sub>4</sub> and stored at 4°C until required, invariably the same day.

#### 5.26 EMBL3 genomic library screening

1.5% agar/LB + 10mM MgSO<sub>4</sub> 9.5x9.5cm plates were made and stored at 37°C immediately before use. 100µl library dilution (EMBL3 phage in SM) and 100µl plating bacteria (*E.Coli* strain K803, see above) were combined and incubated at 37°C for 20 minutes. They were then added to 4.5ml molten (45°C) 0.7% agarose/LB, mixed and poured out onto the 1.5% plates prepared above. Plates were incubated, inverted, at 37°C overnight. The following morning, duplicate 9x9cm nitrocellulose filters (Sigma, Hybond C) were laid onto the above plates for 1 minute; orientation marks were made on the filter and plate with a sterile syringe needle. The filters were then removed and laid (plate-contact side up) onto Whatman 3MM filter paper soaked successively in a) Denaturation Buffer (1.5M NaCl, 0.5M NaOH) for 1.5 minutes, b) Neutralisation Buffer (0.5M Tris/HCl

pH 7.4, 1.5M NaCl) for 6 minutes and then c) 2x SSC (0.3M NaCl, 30mM sodium citrate pH 7.0) for 6 minutes. Filters were then air-dried and baked at 80°C for 2 hours in a vacuum dessicator.

Filters were probed with nick-translated DNA and hybridised as above; details of washing conditions are given in the Results.

Following autoradiography of the filters, the positive clones were identified and plugs taken from the agar plates and stored in 1ml SM + 0.1% chloroform at 4°C.

#### 5.27 Production of lambda-phage plate lysate

This method was used for both R408 helper-phage production (for DNA sequencing) and as an intermediate step in the library-screening process. Phage were plated out to confluence (20,000 p.f.u./9cm round plate) on bacterial lawns of *E.Coli* K803 bacteria on LB/agarose plates and grown overnight. 5ml SM was added to each plate and the plate was gently shaken for 1-2 hours at 4°C. The SM was then collected and a further 1ml SM added; this extra SM was then added to the first 5ml. 100µl chloroform was added, the solution vortexed briefly, centrifuged for 10 minutes at 4°C and then the supernatant was collected and stored at 4°C. It was subsequently titrated by serial dilution.

#### 5.28 Lambda-phage DNA minipreps

Lambda-phage plate lysate was produced as above. RNase and DNase (1µg/ml each) were added to the supernatant; it was then incubated for 30 minutes at 37°C. An equal volume of 2M NaCl/20%

PEG-6000/SM was added and incubated at 4°C for 1 hour. The solution was centrifuged and the supernatant discarded. The pellet was resuspended in 0.5ml SM; 5µl 10% SDS and 5µl 0.5M EDTA pH8.0 were then added. The solution was incubated at 68°C for 15 minutes, after which time it was phenol-extracted, phenol/chloroform-extracted and the DNA precipitated with an equal volume of isopropanol.

#### 5.29 Single-stranded DNA (ssDNA) preps

ssDNA preps were produced using the protocol of Stratagene (1989a). This method uses R408 helper-phage to aid the packaging-deficient pBluescript to produce single-stranded DNA. The DNA pellet from a 4ml culture was resuspended in 5µl TE; 1µl was used per sequencing reaction.

#### 5.30 Single-stranded DNA sequencing

ssDNA sequencing was performed by the Sanger method of dideoxy sequencing according to the protocol of Amersham (1984). M13 primers or synthetic oligonucleotides were used on pBS clone templates. <sup>35</sup>S-dATP was incorporated into the sequencing reactions; these were electrophoresed on 7.3M urea, 1x TBE, 5% polyacrylamide-gels.

#### 5.31 Southern blots

Agarose "Maxigels" were run as above. After photographing the gel, it was treated as follows: a) 20 minutes 0.25M HCl, b) 3x 15 minutes Denaturation Buffer, and c) 3x 15 minutes Neutralisation Buffer (buffer details are given under EMBL3 library screening above). In

between each of these steps the gel was rinsed twice in distilled water. The gel was then blotted using standard procedures: a nitrocellulose filter was used (Hybond C, Amersham); the eluant was 20x SSC; and elution was overnight. After blotting the filter was rinsed in 6x SSC, air-dried and baked at 80°C for 2 hours in a vacuum oven.

#### 5.32 Production of promoter deletion/ $\beta$ -galactosidase clones

The vector used in the  $\beta$ -galactosidase expression studies was pCaSpeR-AUG- $\beta$ gal (THUMMEL *et al.*, 1988). All clones were excised from pBluescript (pBS) clones described in Chapter IX and inserted into the EcoRI-BamHI-KpnI polylinker site of pCaSpeR-AUG- $\beta$ gal, using some of the pBS polylinker if necessary.

Precise details of the cloning steps are given in Chapter IX. After ligation into pCaSpeR-AUG- $\beta$ gal the clones were transformed into the bacterium BB4 and large-scale DNA preparations performed as above.

#### 5.33 Linearization of templates for micro-injection

10 $\mu$ g of plasmid was digested with the appropriate enzyme/buffer; complete digestion was verified by the total absence of circular-plasmid bands on agarose-gel electrophoresis. The DNA was then phenol/chloroform-extracted, chloroform-extracted and then ethanol-precipitated twice with sodium acetate. After precipitation the pellet was washed with 80% ethanol and resuspended in TE at 40 $\mu$ g/ml.

#### 5.34 Hormone-stimulated production of *Xenopus* eggs and *in vitro* fertilisation

The hormone-stimulated production of *Xenopus laevis* and *borealis*

eggs was performed as per standard procedures. A combination of 'natural' and 'artificial' (i.e. by removal of testes, followed by *in vitro* fertilisation) matings were used.

#### 5.35 Injection of DNA clones into *Xenopus laevis* embryos

*Xenopus laevis* embryos were produced as described above and dejellied with 2% cysteine-HCl pH 8.0 (NaOH). Two-cell embryos were then transferred into 5% Ficoll in 1/10 BarthX (see above) to remove liquid from the egg-cell membrane space thus facilitating microinjection. 20nl of DNA at 40µg/ml was injected into one cell of each two-cell *Xenopus laevis* embryo. Embryos were incubated until cleavage had proceeded and then transferred back to 1/10 BarthX to allow normal gastrulation. Incubation was subsequently performed at 14-20°C up to the desired stage of development.

#### 5.36 Staining of *Xenopus* embryos for $\beta$ -galactosidase expression

The method used was recommended by Richard Harland, University of California at Berkley (SANES *et al.*, 1986) that had been used on transgenic mice. All staining steps were performed in 1x3cm glass bottles. Embryos of the desired stage were washed in ice-cold-PBS and then fixed in Fix (2% formaldehyde, 0.2% glutaraldehyde, 0.02% NP40, 0.01% sodium deoxycholate, PBS) for 40 minutes on ice.

After rinsing twice in ice-cold PBS, Xgal Stain (5mM  $K_4Fe(CN)_6$ , 5mM  $K_3Fe(CN)_6$ , 1mg/ml X-gal, 2mM  $MgCl_2$ , PBS, made fresh) was added and the embryos gently shaken for 48 hours at room temperature in foil-covered bottles. They were then

rinsed twice in PBS and fixed in MEMFA (0.1M MOPS pH 7.4, 2mM EGTA, 1mM  $\text{MgSO}_4$ , 3.7% formaldehyde, distilled water) for 2 hours at room temperature, before being bleached (70% MeOH/30%  $\text{H}_2\text{O}_2$ ) for 5-12 hours, and then fixed in 100% MeOH for 2x 5 minutes. Embryos were then cleared in Murray's (2 benzyl benzoate:1 benzyl alcohol) and photographed. If storage of the embryos was necessary, they were kept in 100% MeOH in the dark.

## Chapter VI

### Preliminary tests and library screen

#### 6.1 Introduction to results chapters

#### 6.2. Origin of N-Cadherin cDNA

#### 6.3 Southern blot of *X.borealis* genomic DNA with *X.laevis*

##### N-Cadherin cDNA

#### 6.4 RNase protection assays

(i) Confirmation of lack of heterogeneity at 5'-end  
of cDNA

(ii) Use of elongation factor EF-1 $\alpha$  as a positive control

(iii) *X.laevis* N-Cadherin cDNA BamHI-PvuII fragment

(iv) *X.laevis* N-Cadherin cDNA EcoRI-BamHI fragment

#### 6.5 Genomic screen of *X.borealis* genomic DNA library with *X.laevis*

##### N-Cadherin cDNA

(i) Choice of library

(ii) Library construction

(iii) Library screen

#### 6.6 Mapping of putative positive clones

(i) Southern blot of putative positives

(ii) Mapping of positive clones 3- 6, 9, 12,  
and 21/EMBL3

#### 6.7 Conclusions

## 6.1 Introduction to results chapters

The case for the exploration of the N-Cadherin promoter region has been made in the previous introductory chapters. It has been justified on the basis of contributing to our knowledge of neural promoters and the regulation of  $\text{Ca}^{2+}$ -dependent cell adhesion molecules in general, and also to the study of the regulation of the N-Cadherin molecule in particular. Chapters VI-IX therefore contain the results of this study.

Chapter VI contains the details of the initial tests performed on the *X.laevis* N-Cadherin cDNA prior to its use in the screening of a *X.borealis* genomic DNA library. The isolation of the genomic clones and the cloning of the DNA fragments that hybridised to the *X.laevis* N-Cadherin cDNA are also described.

In Chapter VII, the sequence data from the genomic clones is presented and then analysed from a theoretical aspect.

Chapter VIII provides experimental data to support the predictions made in Chapter VII.

In Chapter IX, the construction of promoter/deletion- $\beta$ -galactosidase clones is described together with the results of the micro-injection of these clones into *Xenopus* embryos. Photographs of these injected embryos are presented and also an interpretation of the  $\beta$ -galactosidase expression patterns obtained.

Unless otherwise stated, all experimental procedures were carried out as detailed in Chapter V.

## 6.2 Origin of *X.laevis* N-Cadherin cDNA

The *Xenopus laevis* N-Cadherin cDNA was originally obtained by members of Kintner's group (DETRICK *et al.*, 1990) as described in Chapter III. Briefly, the chick N-Cadherin cDNA was identified as a result of antibody-neutralisation assays. This was then used to screen a *X.laevis* stage 17 (early neurula)  $\lambda$ gt10 library, the resulting clones being identified as the *X. laevis* homologues of chick N-Cadherin by DNA sequence comparison.

The complete *X.laevis* N-Cadherin cDNA (approx 4.0kb) was very kindly provided by Chris Kintner (Salk Institute, San Diego), ready-cloned into the EcoRI site of the vector SP72 (Promega). On receipt, this plasmid was transformed into the bacteria MC1061 and a large-scale DNA preparation carried out. An extensive restriction-enzyme fragment analysis of the plasmid was then performed to confirm the identity of the plasmid against the published sequence. A restriction-enzyme map of the *X.laevis* N-Cadherin cDNA is given in Figure VI.1. This figure also illustrates the main cDNA restriction-enzyme fragments used in this study.

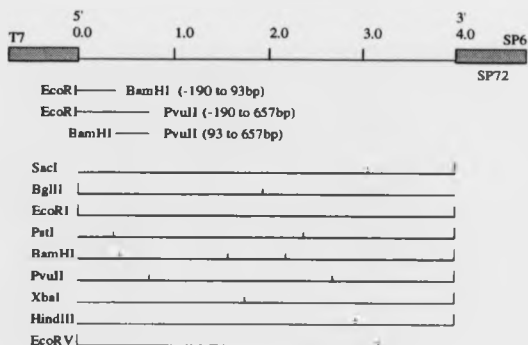
## 6.3 Southern blot of *X.borealis* genomic DNA with *X.laevis* N-Cadherin cDNA

In order to confirm the presence of the N-Cadherin gene in the genome of *X.borealis*, the *X.laevis* cDNA was used to probe *X.borealis* genomic DNA in a Southern blot.

8 $\mu$ g *X.borealis* DNA was digested to completion with EcoRI and

**Figure VI.1**

**Restriction-enzyme digest map of *X. laevis* N-Cadherin cDNA**



This clone was kindly donated by Chris Kintner (Salk Institute, San Diego). Upon receipt, the above restriction-enzyme digests were performed and the results verified against the published DNA sequence data (DETRICK *et al.*, 1990).

The vector is SP72 (Promega); the insert is cloned into the EcoRI site. The scale is in kilobases. The fragments used in this study are illustrated in bold and numbered relative to the translational start-site (+1).

PvuII. The digested DNA was run on a 1% agarose-TBE Maxigel together with lambda markers and N-Cadherin cDNA digested with PstI+EcoRI to act as a positive control.  $8 \times 10^{-3} \mu\text{g}$  and  $8 \times 10^{-7} \mu\text{g}$  digested cDNA were run; the former was a positive control for the hybridization reaction, the latter being representative of the amount of DNA expected from a single copy gene in  $8 \mu\text{g}$  total DNA. The gel was run, in triplicate, at 40mA for 16 hours.

The Southern blot and subsequent hybridisation with  $^{32}\text{P}$ -dGTP labelled nick-translated *X.laevis* N-Cadherin cDNA (total, gel-isolated) were carried out as described in Chapter V. Filters were washed for 3x30 minutes at a)  $55^{\circ}\text{C}$ , 2xSSC, 0.5%SDS, b)  $55^{\circ}\text{C}$ , 0.1xSSC, 0.5%SDS, and c)  $60^{\circ}\text{C}$ , 0.1xSSC, 0.5%SDS and autoradiographed for 2.5 days at  $-70^{\circ}\text{C}$ . The gel and corresponding autoradiograph are shown in Figures VI.2 and VI.3.

The highest stringency wash (c) was a compromise between preventing cross-reaction of the probe with other Cadherins in the *X.borealis* genome, whilst still preserving hybridisation using the inter-species probe. The result shows several genomic bands particularly in the EcoRI track (Lane 10), thus confirming the presence of at least one copy of the N-Cadherin gene in the pseudo-tetraploid *X.borealis* genome.

#### 6.4 RNase protection assays

##### 6.4(i) Confirmation of lack of heterogeneity at 5'-end of cDNA

Differential splicing is known to be common within other neural/

Figure VI 2a

Agarose gel of *X.borealis* genomic DNA  
restriction-enzyme digested with PvuII and EcoRI

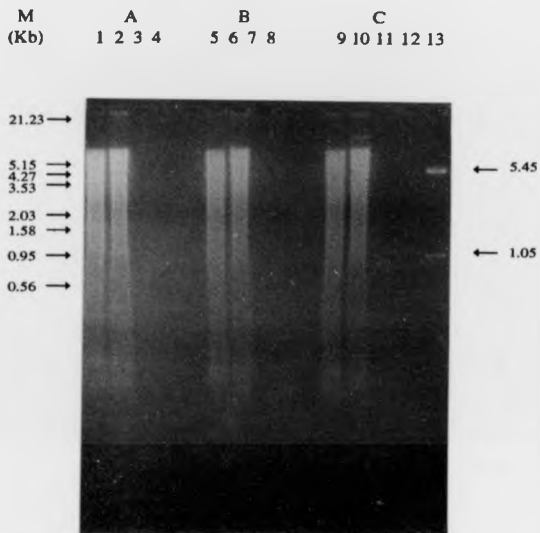
- Lanes 1, 5 and 9: 8  $\mu$ g *X.borealis* genomic DNA cut with PvuII  
Lanes 2, 6 and 10: 8  $\mu$ g *X.borealis* genomic DNA cut with EcoRI
- Lanes 3, 7 and 11: 10<sup>-3</sup>  $\mu$ g N-Cadherin cDNA cut PstI+EcoRI  
Lanes 4, 8 and 12: 10<sup>-7</sup>  $\mu$ g N-Cadherin cDNA cut PstI+EcoRI
- Lane 13: 10<sup>-2</sup>  $\mu$ g N-Cadherin cDNA cut HindIII

Markers (M): Lambda plasmid cut EcoRI+HindIII

Gel: 1%-agarose TBE Maxigel

The N-Cadherin cDNA is that of Detrick *et al.* (1990) as described in  
Figure VI.1.

Figure VI.2b



This gel was blotted and probed with a DNA fragment from the *X.laevis* N-Cadherin cDNA; this data is shown in Figure VI.3.

**Figure VI.3a**

Southern blot of agarose gel of *X.borealis* genomic DNA  
restriction-enzyme digested with PvuII and EcoRI,  
probed with *X.laevis* N-Cadherin cDNA

Lanes 1, 5 and 9: 8  $\mu$ g *X.borealis* genomic DNA cut with PvuII

Lanes 2, 6 and 10: 8  $\mu$ g *X.borealis* genomic DNA cut with EcoRI

Lanes 3, 7 and 11: 10<sup>-3</sup>  $\mu$ g *X.laevis* N-Cadherin cDNA cut  
PstI+EcoRI

Lanes 4, 8 and 12: 10<sup>-7</sup>  $\mu$ g *X.laevis* N-Cadherin cDNA cut  
PstI+EcoRI

Lane 13: 10<sup>-2</sup>  $\mu$ g *X.laevis* N-Cadherin cDNA cut HindIII

Markers (M): Lambda plasmid cut EcoRI+HindIII

Probe: *X.laevis* N-Cadherin cDNA (approx. 4kb)

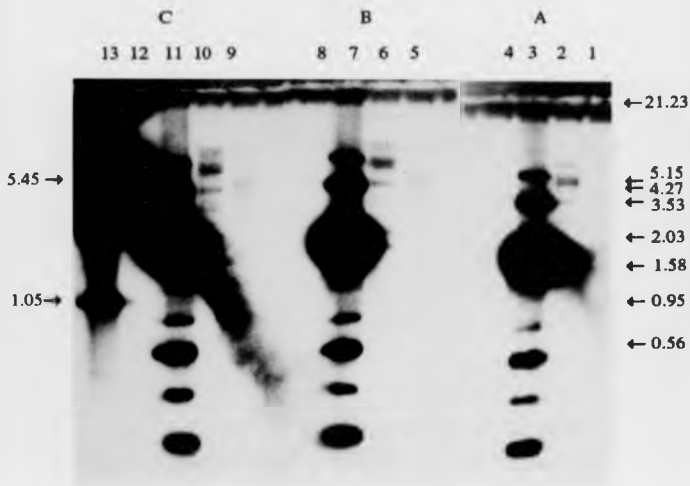
Wash stringencies:

A	55°C, 2x SSC, 0.5% SDS
B	55°C, 0.1x SSC, 0.5% SDS
C	60°C, 0.1x SSC, 0.5% SDS

3 washes each for 30 minutes

The Southern blot was autoradiographed for 2.5 days at -70°C, with one screen.

Figure VI.3b



The Southern blot shows several genomic bands particularly in the EcoRI lane (10), thus confirming the presence of at least one copy of the N-Cadherin gene in the *X. borealis* pseudo-tetraploid genome.

muscle genes, as discussed in the introductory chapters. An RNase Protection Assay (RPA) was therefore performed on *X. laevis* embryo RNA to verify that there was no heterogeneity in the 5'-end of the N-Cadherin mRNA, at least in the part that corresponded to the cDNA. It must be remembered that there was no guarantee that the cDNA would extend to the transcriptional start-site. Effects such as GC-rich regions producing hairpins, or repeats of one particular dNTP producing a local shortage of that dNTP, can produce stalling and fall-off of the reverse-transcriptase molecule. Therefore, although the cDNA contains 182bp upstream of the translational start site, it can not be assumed that the start of the cDNA coincides exactly with the transcriptional start site.

The cDNA could have been used to screen a cDNA-library in order to isolate more-5' DNA sequence. There was, however, no indication that this was necessary or any guarantee of success.

#### 6.4(ii) Use of elongation factor EF-1 $\alpha$ as a positive control

The elongation factor, EF-1 $\alpha$  (KRIEG *et al.*, 1989), was used initially as a positive control to establish the *in vitro* transcription / RPA procedure; as an 'abundant' message, it was readily detectable. It also was used initially to verify the integrity of the RNA extracted from *Xenopus* embryos.

The EF-1 $\alpha$  clone used (G1EF) was obtained from Paul Krieg (University of Texas at Austin, USA) and contains a 378bp PstI-SacI fragment of the *X. laevis* cDNA clone in pGEM1. For RPAs, the clone

was linearised with HindIII and transcribed with T7 RNA polymerase. As can be seen in Figure VI.4, one main strong band is seen with *X. laevis*; two lower-size bands are seen with *X. borealis*.

6.4(iii) *X. laevis* N-Cadherin cDNA BamHI-PvuII fragment

The fragment used by Kintner in his RNase protection assays (to analyse tissue and temporal expression) was a 564bp BamHI-PvuII fragment (93 to 657 nts, illustrated in Figure VI.1) that encodes primarily the "pre" region of the N-Cadherin protein (DETRICK *et al.*, 1990). Therefore, this fragment was used as a known positive control to establish the RPA reaction conditions for N-Cadherin probes on *X. laevis* RNA.

The BamHI-PvuII fragment was gel-isolated from the N-Cadherin cDNA/SP72 clone. pBS KS+ was cut with SacI, blunted, cut with BamHI and then CIAPed, the vector being phenol extracted and EtOH-precipitated between each step. (The SacI/blunting procedure produced a blunt end to ligate the PvuII end to - PvuII being a blunt-end producing enzyme and there being no PvuII site in the pBS polylinker). The BamHI-PvuII fragment was then cloned into the above-prepared pBS vector, transformed into BB4 and minipreped.

Positive clones were verified by restriction-enzyme digests and then large-scale DNA preparations carried out. ssDNA preps were made from the large-scale DNA clone and sequenced from pBS into the PvuII end of the insert to double-check the cloning procedure.

Labelled anti-sense RNA probe was made by linearizing the BamHI-

Figure VI.4a

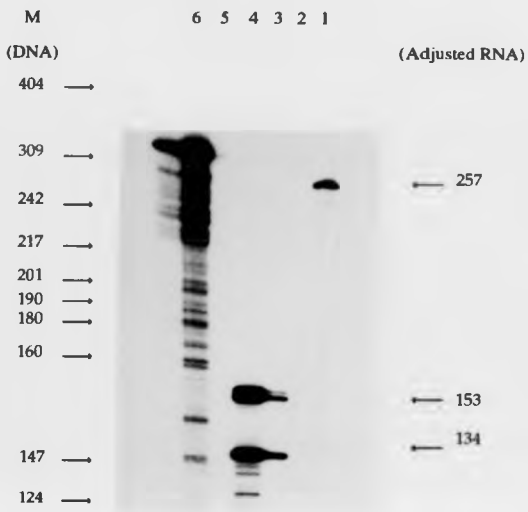
RNAse Protection Assay:

EF-1 $\alpha$  probe on *X.laevis* brain RNA and *X.borealis* embryo RNA

- Lane 1: *X.laevis* adult brain RNA, 1 brain (10 $\mu$ g)  
Lane 2: *X.borealis* embryo RNA (stage 19-27) - 25 $\mu$ g  
Lane 3: " - 2.5 $\mu$ g  
Lane 4: " - 0.25 $\mu$ g  
Lane 5: Control - 10 $\mu$ g tRNA with RNAse A  
Lane 6: Control - 10 $\mu$ g tRNA without RNAse A

To each of the above was added 500cpm EF-1 $\alpha$  probe. Hybridisation and digestion conditons were as described in Chapter V.

**Figure VI.4b**



The autoradiograph shows that the EF-1 $\alpha$  probe readily detects homologous RNA in *X.laevis* brain (Lane 1) and *X.borealis* embryo (Lanes 3 and 4) preparations.

PvuII/pBS clone with BamHI and transcribing with T7 RNA polymerase; control sense-RNA was made by linearizing with PvuII (which cuts down-stream of the pBS polylinker) and transcribing with T3 RNA polymerase.

This experiment also compared two different RNA extraction procedures: the first involved homogenizing embryos directly into phenol, phenol/chloroform extracting and EtOH-precipitating (the "Phenol" method); the second method was the Proteinase K method detailed in Chapter V (the "Proteinase K" method). The results are shown in Figure VI.5.

In addition to confirming that the cDNA fragment is completely protected (compare lanes 4 and 10, although a smaller band of 348bp is seen), the experiment demonstrated the superior RNA extraction properties of the "Proteinase K" method over the "Phenol" method. The "Proteinase K" method was therefore used in all subsequent experiments, and reliably produced clean, un-degraded RNA (as judged by RNA-agarose gel electrophoresis).

#### 6.4(iv) *X.laevis* N-Cadherin cDNA EcoRI-BamHI fragment

The EcoRI-BamHI fragment at the extreme 5' end of the *X.laevis* N-Cadherin cDNA (-190 to +93nts, Figure VI.1) was ligated directly into pBS KS+, transformed into the bacterium TG<sub>2</sub> and subsequently minipreps were made. After restriction-enzyme analysis to verify the correct insert, a large-scale DNA preparation was carried out, ssDNA preps were made and the insert confirmed again by DNA sequencing.

Figure VI.5a

RNAse Protection Assay:

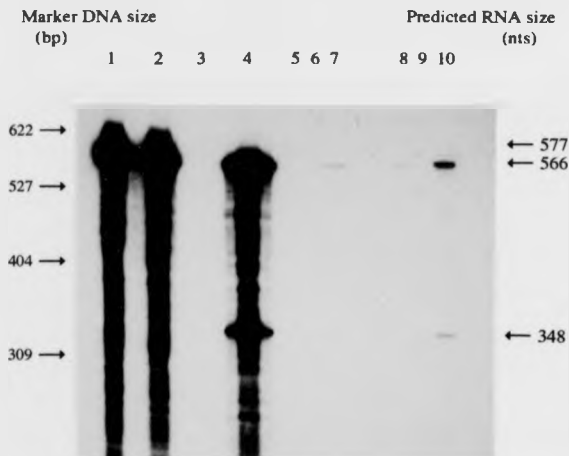
*X.laevis* N-Cadherin BamHI-PvuII fragment on *X.laevis* embryos

Lane 1:	Control tRNA - RNase (+ overnight hybridisation)
Lane 2:	Control tRNA - RNase (- overnight hybridisation)
Lane 3:	Control tRNA + RNase
Lane 4:	Control tRNA + sense RNA + RNase
Lane 5:	<i>X.laevis</i> embryo RNA (stage 17-22) 3µg (Phenol)
Lane 6:	" 9µg "
Lane 7:	" 15µg "
Lane 8:	" 3µg (Proteinase K)
Lane 9:	" 9µg "
Lane 10:	" 15µg "

To each of the above was added 50,000cpm *X.laevis* N-Cadherin BamHI-PvuII *in vitro* transcribed RNA probe (fragment as shown in Figure VI.1); hybridisation and digestion conditions were as described in Chapter V.

"Phenol" and "Proteinase K" refer to the method of extracting the embryo RNA; full details are given in the text.

Figure VL5b



The above gel shows the complete protection of the *X. laevis* N-Cadherin BamHI-PvuII fragment when used to probe *X. laevis* stage 17-22 embryo RNA.

It also illustrates the superior RNA extraction properties of the "Proteinase K" method (Lanes 8-10) over the "Phenol" method (Lanes 5-7).

To make the anti-sense probe, the EcoRI-BamHI/pBS clone was linearised with EcoRI and transcribed with T7 RNA polymerase. The sense-RNA was made by linearizing with BamHI and transcribing with T3 RNA polymerase. The experiment compared *X.laevis* unfertilised egg, stage 8 and stage 18 RNA. The results are shown in Figure VI.6.

This experiment confirmed a number of points:

1. The same temporal expression pattern is seen with this EcoRI-BamHI fragment as the published (DETRICK *et al.*, 1990) more 3' BamHI-PvuII fragment: prior to MBT (stage 10.5) no expression is detectable; expression is strong in mid-neurula (stage 18). These results are consistent with a molecule supposedly involved in neural development.
2. An amount of RNA between 3 and 15µg (1-5 embryos) would give a readily detectable signal using the *X.laevis* cDNA probe on *X.laevis* embryo RNA.
3. No heterogeneity is seen in the length of the protected transcript. Differential splicing at the 5'-end does therefore not appear to occur within the regions tested.

This *X.laevis* probe was not used on *X.borealis* RNA for the following reason: RNase A will cut ssRNA 3' to pyrimidine residues; RNase T1 will cut 3' to guanine residues. Therefore, the mismatch of any one RNA base will result in the cutting of the probe at that point. As can be seen from the subsequently-derived *X.borealis* DNA clone

Figure VI.6a

RNase Protection Assay:

*X.laevis* N-Cadherin EcoRI-BamHI fragment on *X.laevis* embryos

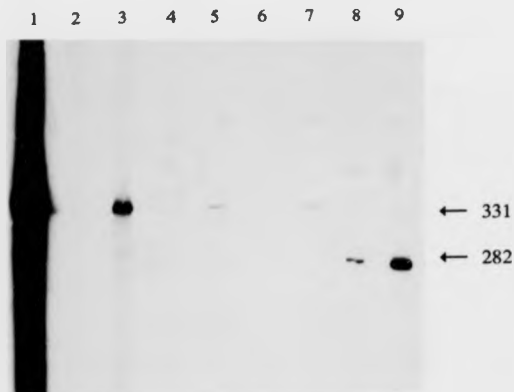
Lane 1:	Control tRNA - RNase	
Lane 2:	Control tRNA + RNase	
Lane 3:	Control tRNA with sense synthetic RNA + RNase	
Lane 4:	<i>X.laevis</i> embryo RNA egg, 3µg	
Lane 5:	"	" 15µg
Lane 6:	"	stage 8, 3µg
Lane 7:	"	" 15µg
Lane 8:	"	stage 18, 3µg
Lane 9:	"	" 15µg

Markers: DNA sequence

Probe: 50,000cpm *X.laevis* N-Cadherin EcoRI-BamHI fragment (as shown in Figure VI.1) was added to all reactions.

Hybridisation and digestion conditions were as described in Chapter V.

Figure VI.6b



The above gel shows that a full-length protected band (282nts) is only seen with stage 18 embryos and not in the egg or at stage 8.

Only background (undigested probe, 331nts) is present in the egg and stage 8 lanes.

sequence, the sequence conservation between *X.laevis* and *X.borealis* is high but far from 100% (Figure VII.6); hence the omission of this experiment is justified.

#### 6.5 Genomic screen of *X.borealis* genomic DNA library with *X.laevis* N-Cadherin cDNA

##### 6.5(i) Choice of library

A *X.borealis* genomic library was obtained from Clive Wilson (WILSON *et al.*, 1986). The choice of *Xenopus* library was primarily dictated by availability. In retrospect it would have probably been better to have used a *X.laevis* genomic library but only due to the far greater availability of *X.laevis* embryos within our laboratory. It must be said that a clone obtained from a *X.laevis* genomic library would probably have a higher sequence homology with the *X.laevis* cDNA probe, but it would not necessarily be the identical sequence because of the pseudo-tetraploid nature of the *Xenopus* genome. Even if the genomic homologue of the cDNA was obtained, as stated above, it could not be assumed that the cDNA extended all the way to the transcriptional start site; the 5' 'uncharted' DNA would still have to be mapped in the same way that a *X.borealis* clone would be.

On the plus side, the use of a *X.borealis* clone has allowed promoter constructs, micro-injected into the larger *X.laevis* embryos, to be differentiated from the endogenous (*X.laevis*) message.

##### 6.5(ii) Library construction

Genomic DNA had been extracted from an adult *X.borealis* by

cardiac puncture after injection with heparin. This DNA had been partially digested with Sau3A and 15kb fragments (average) selected and cloned into the BamHI site of EMBL3; greater than 98% of recombinants were said to have inserts. The library had been used previously to obtain clones for actin gene sequences (WILSON *et al.*, 1986).

The library had been stored as a phage lysate at 4°C. Serial dilutions of the library were first plated out onto a bacterial lawn of K803 to establish the current titre.

#### 6.5(iii) Library screen

The above-mentioned *X. borealis* library was screened with fragments of the *X. laevis* N-Cadherin cDNA according to the following regime. To avoid losing any putative-positives from any stage of the screening, all positive colonies, and adjacent colonies in ambiguous cases, were taken at each stage. Included in each stage of the library screening was a Southern-blot filter of restriction-enzyme digested *X. laevis* N-Cadherin cDNA to act as a positive control for each hybridization reaction.

Conditions for screening of *X.borealis* genomic library

1st screen:

Probe: N-Cadherin 854bp EcoRI-PvuII fragment  
(-197 to 657nts)  
24 x 2 (duplicate) plates  
15,000 p.f.u. per plate (i.e.  $3.6 \times 10^5$  p.f.u.total)  
Washes: 55°C, 2x SSC, 0.1% SDS  
26 putative positives taken

2nd screen:

Probe: N-Cadherin 300bp EcoRI-BamHI fragment  
(-197 to 93nts)  
26 x 2 plates  
500 p.f.u. per plate  
Washes: 55°C, 2x SSC, 0.1% SDS  
27 putative positives taken

3rd screen:

Probe: N-Cadherin 300bp EcoRI-BamHI fragment  
(-197 to 93nts)  
27 x 2 plates  
100 p.f.u. per plate  
Washes: 60°C, 2x SSC, 0.1% SDS  
32 putative positives taken

## 6.6 Mapping of putative positive clones

### 6.6(i) Southern blot of putative positives

Phage preps were performed on each of the 32 plaque-pure plugs taken from the 3rd-round screen. Each was subjected to an EcoRI+BamHI restriction-enzyme digest and run out on an 0.6% agarose-TBE Maxigel. The gel was Southern-blotted and the filter hybridised with a <sup>32</sup>P-labelled nick-translated *X.laevis* N-Cadherin 300bp EcoRI-BamHI fragment (Figure VI.1). The filter was washed consecutively at a) 60°C, 2x SSC, 0.1% SDS, b) 65°C, 2x SSC, 0.1% SDS and then c) 65°C, 0.1x SSC, 0.1% SDS, the filter being exposed at -70°C for 36 hours in between each wash (data not shown).

The resulting gel and autoradiograph revealed four distinct restriction-enzyme digest patterns; clones 3- 6, 9, 12 and 21/EMBL3 were chosen as representative of each of these four patterns (3- refers to the third-round screen).

### 6.6(ii) Mapping of positive clones 3- 6, 9, 12, and 21/EMBL3

The production of the *X.borealis* genomic EMBL3 library had involved the ligation of Sau3A fragments of genomic DNA into the BamHI site of EMBL3 (Section 6.5(ii)). The resultant DNA therefore reads: (EMBL3) Sall-BamHI/Sau3A-genomic DNA insert-Sau3A/BamHI-Sall (EMBL3). Hence Sall is the only unique site in EMBL3 that spans the insert, except for the 1/4 occasions that the BamHI site is reconstituted after Sau3A ligation. Given the average insert size of >20kb, it could not be assumed that the insert would be liberated

cleanly by Sall or BamHI, and this was indeed the case: restriction-enzyme digest with Sall and/or BamHI produced a variety of fragments, from which the insert size was determined.

Each of the above clones was then restriction-enzyme digested with BamHI+Sall, run on a 0.5% agarose-TBE gel and Southern blotted/probed as before. Equal amounts of each insert (given its size) were used to ensure the resulting autoradiograph would be comparably-quantitative with respect to the strength of signal produced by each band (the strength of the signal being a combined measure of the degree and extent of homology between probe and insert). The agarose-gel and Southern blot are shown in Figures VI.7 and VI.8. The resultant filters were washed at 65°C, 2x SSC, 0.1% SDS for 3x 30 minutes.

It was found that 3- 9/EMBL3 and 3- 21/EMBL3 clones gave the strongest signal when probed with <sup>32</sup>P-labelled *X.laevis* N-Cadherin EcoRI-BamHI cDNA fragment and so restriction-enzyme maps of these two clones were compiled (Figures VI.9 and VI.10).

The fragments within 3-9/EMBL3 and 3-21/EMBL3 that hybridised to the *X.laevis* N-Cadherin cDNA were subcloned into pBluescript KS+ and restriction-enzyme mapped further. Again the clones were Southern blotted to determine, on a finer scale, the position of the fragment that hybridised to the 5'-end of the *X.laevis* N-Cadherin cDNA. Once these fragments were identified, they were isolated and recloned for sequencing. The sequence data and analysis is given in

**Figure VI.7a**

**Agarose gel of restriction-enzyme digested EMBL3 clones.**

Lane 1: Control N-Cadherin cDNA/SP72 XhoI digested

Lane 2: Lambda markers (EcoRI+HindIII)

Lane 3: Sall digest of 3-21/EMBL3

Lane 4: " 3-12/EMBL3

Lane 5: " 3-9/EMBL3

Lane 6: " 3-6/EMBL3

Lane 7: Lambda markers (HindIII)

Lane 8: BamHI+Sall digest of 3-21/EMBL3

Lane 9: " 3-12/EMBL3

Lane 10: " 3-9/EMBL3

Lane 11: " 3-6/EMBL3

Lane 12: Lambda markers (BamHI)

Lane 13: BamHI digest of 3-21/EMBL3

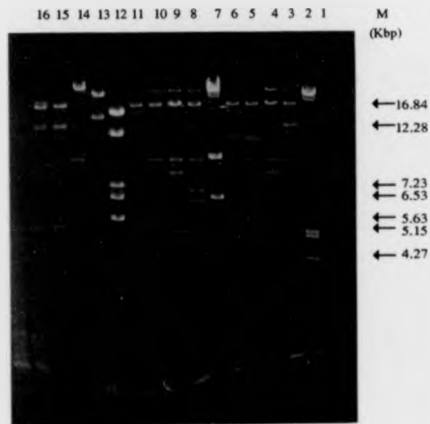
Lane 14: " 3-12/EMBL3

Lane 15: " 3-9/EMBL3

Lane 16: " 3-6/EMBL3

M: Compilation of the lambda markers, shown in Kbs.

Figure VI.7b



The genomic clones isolated from the library screen all fell into one of four restriction-enzyme digestion patterns, as exemplified by clones 3-6, 9, 12 and 21/EMBL3.

Figure VI.8 shows a Southern blot of the above gel, probed with a fragment from the *X.laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990).

**Figure VI.8a**

**Southern-blot of restriction-enzyme digested EMBL3 clones**

**Lane 1:** Control N-Cadherin cDNA/SP72 (XhoI digested)

**Lane 2:** Lambda markers (EcoRI+HindIII)

**Lane 3:** Sall digest of 3-21/EMBL

**Lane 4:** " 3-12/EMBL3

**Lane 5:** " 3-9/EMBL3

**Lane 6:** " 3-6/EMBL3

**Lane 7:** Lambda markers (HindIII)

**Lane 8:** BamHI+Sall digest of 3-21/EMBL3

**Lane 9:** " 3-12/EMBL3

**Lane 10:** " 3-9/EMBL3

**Lane 11:** " 3-6/EMBL3

**Lane 12:** Lambda markers (BamHI)

**Lane 13:** BamHI digest of 3-21/EMBL3

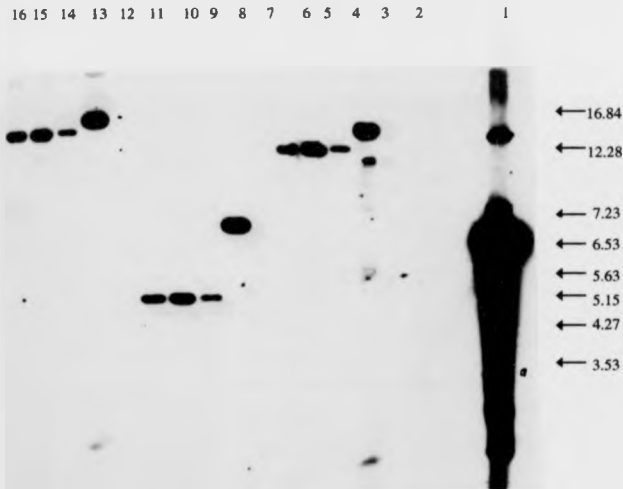
**Lane 14:** " 3-12/EMBL3

**Lane 15:** " 3-9/EMBL3

**Lane 16:** " 3-6/EMBL3

**Markers:** Compilation of above lambda markers, shown in Kbp

Figure VI.8b

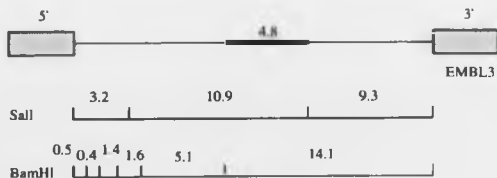


Southern blot of Figure VI.7 agarose gel. The four restriction-enzyme digested genomic clones were probed with the 5'-end of the *X. laevis* N-Cadherin cDNA (EcoRI-BamHI fragment) in order to map the regions of homology. The filter was washed at 65°C, 0.1xSSC, 0.1% SDS and exposed at -70°C for 36 hours.

Restriction-enzyme maps for 3-9/EMBL3 and 3-21/EMBL3, based on this data, are shown in Figures VI.9 and 10.

**Figure VI.9**

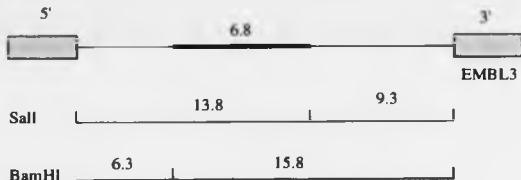
**Restriction-enzyme map of 3-9/EMBL3 genomic clone**



The bold 4.8Kb BamHI-Sall fragment was found to hybridise to the 5'-end of the *X.laevis* N-Cadherin cDNA. It was therefore gel-isolated and cloned into pBluescript to become clone 3-9/4.8BS/pBS. (See "Note" on page 216a.)

**Figure VI.10**

**Restriction-enzyme map of 3-21/EMBL3 genomic clone**



The bold 6.8Kb BamHI-Sall fragment was found to hybridise to the 5'-end of the *X.laevis* N-Cadherin cDNA. It was therefore gel-isolated and cloned into pBluescript to become clone 3-21/6.8BS/pBS.

## Chapter VII.

### 6.7 Conclusions

This chapter has described how the 5'-end of the *X.laevis* N-Cadherin cDNA was used to screen a *X.borealis* genomic library in order to isolate the corresponding *X.borealis* genomic clones.

The clones isolated from the genomic library fell into one of four groups defined by restriction-enzyme digest patterns.

After Southern blotting with the *X.laevis* N-Cadherin cDNA, two of these clones showed strong signals after high-stringency washes. These clones, 3-9/EMBL3 and 3-21/EMBL3, were mapped and the DNA regions of homology with the *X.laevis* N-Cadherin cDNA subcloned into pBluescript to form 3-9/4.8BS/pBS and 3-21/6.8BS/pBS respectively.

## Chapter VII

### Analysis of isolated *X.borealis* genomic DNA sequences

#### 7.1 Introduction

#### 7.2 Sequencing of genomic clones

(i) 3-21/6.8BS/pBS (6.8Kb BamHI-Sall fragment)

(ii) 3-9/4.8BS/pBS (4.8Kb BamHI-Sall fragment)

#### 7.3 Alignment of *X.laevis* N-Cadherin cDNA and 3-9/4.8BS/pBS *X.borealis* genomic DNA sequences

#### 7.4 Inter-species comparison of N-Cadherin amino-acid sequences

#### 7.5 DNA-binding protein motifs present in genomic DNA

#### 7.6 Frequency of occurrence of TAT and ATA triplets in the 3-9/4.8BS/pBS *X.borealis* N-Cadherin promoter region

#### 7.7 Analysis of CG/GC ratio in *X.borealis* promoter region

#### 7.8 Analysis of translational start-site

#### 7.9 Potential secondary structure upstream of the translational start-site

#### 7.10 Intron/exon structure of *X.borealis* genomic DNA sequence

#### 7.11 Conclusions

### 7.1 Introduction

This chapter provides a theoretical analysis of the sequence data derived from the genomic clones isolated in Chapter VI (Figures VI.9 and VI.10). Clone 3-21/6.8BS/pBS was found to contain only a small region of homology to the *X.laevis* N-Cadherin genomic DNA sequence. Longer regions of homology to the *X.laevis* N-Cadherin cDNA, and also 1.3Kb DNA upstream of this, were found in clone 3-9/4.8BS/pBS. The work therefore centred around the clone 3-9/4.8BS/pBS, in the hope of finding the *X.borealis* transcriptional start-site and part of the promoter within this 1.3Kb.

Unless otherwise stated, the *X.laevis* N-Cadherin cDNA referred to below is that of Detrick *et al.* (1990).

### 7.2 Sequencing of genomic DNA clones

#### 7.2(i) 3-21/6.8BS/pBS (6.8Kb BamHI-SaII fragment)

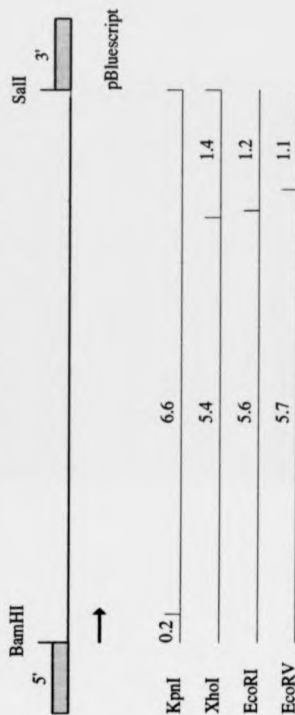
The restriction-enzyme map of this clone is given in Figure VII.1. The *X.laevis* N-Cadherin cDNA probe was found to hybridise to the 0.2Kb KpnI-BamHI fragment but not to the 6.6Kb KpnI-KpnI fragment. Therefore this clone was sequenced directly from the pBS M13 primer, at the BamHI end. The sequence is given in Figure VII.2, aligned (manually) against the sequence of the *X.laevis* N-Cadherin cDNA probe.

Comparison of this sequence with the *X.laevis* cDNA probe sequence showed that this clone had significant homology to regions within the untranslated leader region of the cDNA, but sequences 5' to

Figure VIII.1

*X. borealis* N-Cadherin genomic DNA clone: 3-21/6.8BS/pBS

Restriction-enzyme map and sequencing strategy



Restriction map of 3-21/6.8BS/pBS clone in pBluescript. Numbers on the map refer to kilobase distances as determined by restriction-enzyme digests. The bold arrow shows the part of the clone that hybridised to the 5'-end of the *X. laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990); this section was therefore sequenced (Figure VII.2).

Comparison between *X. laevis* N-Cadherin cDNA and *X. borealis* N-Cadherin genomic sequences (Clone 3-21/6.8BS/pBS)

Lower sequence: *X.borealis* N-Cadherin genomic DNA

```

183                                     X.laevis
GCGGAGCACAGGATTCCTCTGGAAATCAGCCCT GCCTTGTG TTTG CTC
      ** *****
      CCT GGGAAATCAGCCGGAGCCCTGGTGCTTTGGTGGATTCTC
                                     X.borealis

-182

-134                                     X.laevis
CGTACGTGCCCATGCCCG CT CCC CGGCACCTGCCACTGCTGCTGCTGAGTGT
      *****
      A TGCCCGGTTCCGACTACCTACTGCTACTG GCTGCTACAG GT TCT
                                     X.borealis

-141

-82                                     X.laevis
GTAAGGCACGACTGTATGTGCTGCTGCTGCTCGTCAATGTTTCATCTCCAGTCCAAGC
      **** *
      GTAA CCAGGACAG ATGAGC CGGCGGCTCGTCATTGTTTCATCTCCAGTACAAGC
                                     X.borealis

-93

-23                                     X.laevis
CTC CGCCGACCCCAAC AGCATCACCATGTGCCGGAAACAGC
      ***
      CAGTGCTGCTCCCCCACTCCTCACAGCAGCAGCAGCACCATGTGCCGGAG
                                     X.borealis

-40

```

These sequences were aligned and spaces were introduced manually so as to achieve maximum homology. The numbering is relative to the translational start site (underlined) of *X. laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990).

The high degree of homology shows that the genomic clone 3-21/6.8BS/pBS contains DNA from one allele of the *X. borealis* N-Cadherin gene.

this (i.e. putative promoter) were not present in this clone. Promoter sequences would therefore be expected to be present in the more-5' 6.3Kb BamHI-SalI fragment (Figure VI.10). No further experimental work was performed on this clone.

#### 7.2(ii) 3-9/4.8BS/pBS (4.8Kb BamHI-SalI fragment)

This clone has been extensively restriction-enzyme mapped (Figure VII.3), particularly at the 5'-end which has been found to correspond to the 5'-end of the *X.laevis* N-Cadherin cDNA.

All of the individual restriction-enzyme fragments within the 5' 1.6Kb of this clone have been gel isolated and subcloned directly, in both orientations, into pBS. They have all been sequenced at least twice in both orientations. Many other clones that span restriction-enzyme sites have also been made and sequenced to verify both the contiguity of adjacent clones and that no small restriction-enzyme fragments have been lost. The sequencing strategy is outlined in Figure VII.4.

It was also necessary to break the 383bp BamHI-SmaI fragment (-1281 to -888, Figure VII.4) into smaller overlapping Sau3A and HpaII fragments in order to obtain more accurate sequence of this region. This was done by gel-isolating this 383bp fragment, digesting with Sau3A and HpaII (individually), and cloning the resulting DNA fragments into BamHI-digested or SmaI-digested pBS respectively (in the case of the HpaII fragments, these were blunted first with Klenow I). The sequence of this 383bp BamHI-SmaI region has not yet been

Figure VII.3a

*X.borealis* N-Cadherin genomic clone: 3-9/4.8BS/pBS

Restriction-enzyme map of 5'-end

The restriction-enzyme map of the extreme 5'-end of the *X.borealis* genomic clone 3-9/4.8BS/pBS is shown with distances marked in base-pairs relative to the translational start site (defined by homology with the *X.laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990)).

This region was found to contain the putative promoter of the *X.borealis* N-Cadherin gene; it therefore formed the basis of the herein-described study.

Figure VII.3b

*X. borealis* N-Cadherin genomic clone: 3-9/4.8BS/pBS

Restriction-enzyme map of 5'-end

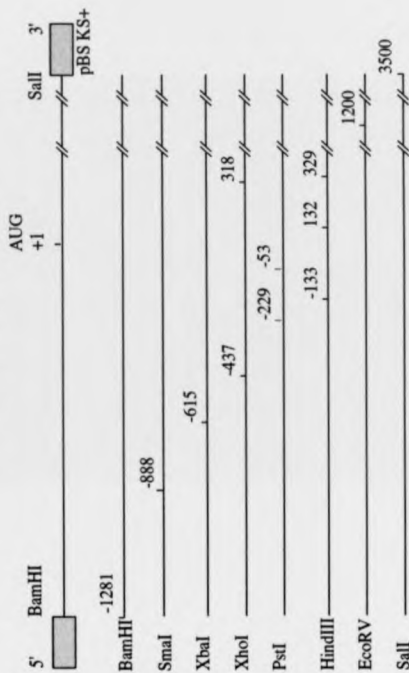
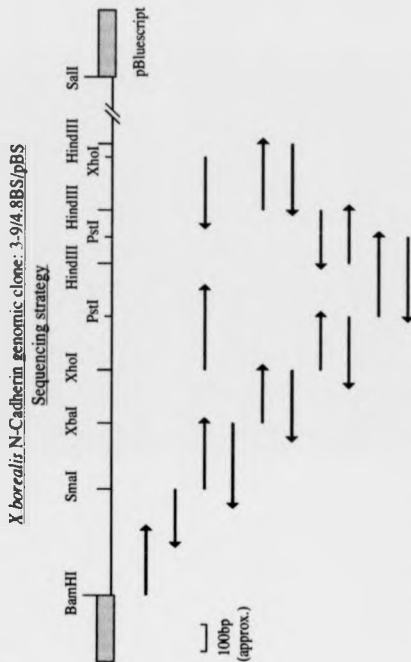


Figure VII.4



A series of overlapping clones were produced and DNA-sequenced according to the above Figure. The majority of clones were sequenced at least twice in both orientations. The BamHI-SalI clone has been divided into Sau3A and HpaII fragments to obtain more accurate sequence of this region.

verified to the same high degree of accuracy as the rest of this clone; further work is still in progress. With this caveat, the full sequence is given in Figure VII.5. The numbering of the sequence is based on the translational start-site (1 = ATG), as defined by DNA sequence alignment with the *X.laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990).

### 7.3 Alignment of *X.laevis* N-Cadherin cDNA and

#### 3-9/4.8BS/pBS *X.borealis* genomic DNA sequences

The DNA sequences of the *X.laevis* N-Cadherin cDNA and *X.borealis* 3-9/4.8BS/pBS are aligned in Figure VII.6. In the first 260bp the degree of homology is 68% (matches/[mismatches + spaces]); it is obviously much higher over shorter stretches. The degree of homology changes abruptly after the AGGT motif at +70; the possibility of this being the site of an exon/intron junction is discussed below. However, it is interesting to speculate why, if this is an exon/intron junction, a reasonable degree of homology still exists after this point between the cDNA and genomic DNA sequences.

### 7.4 Inter-species comparison of N-Cadherin amino-acid sequences

In Figure VII.7, the amino-acid sequences of all the currently published N-Cadherin genes are aligned with 3-9/4.8BS/pBS *X.borealis* genomic DNA sequence. In addition to the *X.laevis* cDNA used above as the probe to obtain the genomic sequence (DETRICK *et al.*, 1990), a further *X.laevis* cDNA sequence has been published by an Israeli group (GINSBERG *et al.*, 1991). It can be seen that the

Figure VII.5a

*X. borealis* N-Cadherin genomic sequence  
 Clone 3-9/4.8BS/pBS (5'-end)

				BamHI	-1270	-1260
				GGAT	CCTCTAAGAC	ATGCTTTTGA
-1250	-1240	-1230		-1220	-1210	-1200
GGCCTCCAG	ACAGTTAGGG	GGTCAGCAGA	GGTTGCATT	TCCAACCAAGT	ATTGCTCCCTG	
-1190	-1180	-1170	-1160	-1150	-1140	
TTCTTGTGTT	ACCCTATCAG	ACACTTCAGG	TACAGTAAGC	CATTAGTGT	TAAGGCGCCA	
-1130	-1120	-1110	-1100	-1090	-1080	
CATAGGGGTA	GTACAGTCTT	GATCTGACCT	GATGGATAAG	GAAAGTGGGG	CATGATCTGA	
-1070	-1060	-1050	-1040	-1030	-1020	
CGCANTGGCC	CCCGGAAGCC	TGTTTCAGAG	ATACCGGGAT	CGGCCAACTC	AAACTTGCCA	
-1010	-1000	-990	-980	-970	-960	
GCTNTGCAC	CGGCGTGCCN	GTCNTTGTGA	GTGGCTGAAT	AGCAAGAGTA	CGCTCGTCCA	
-950	-940	-930	-920	-910	-900	
CCTGGGTGGG	CTTGCCCTCA	GATGTCAATA	AAGTTATAAG	CGTGACACCA	CTTAAGCAAG	
-890	-880	-870	-860	-850	-840	
GGGGTGGGG	CCGGGTACC	TGGGGGGCAA	CGGTCTATCT	CCGGTTTCAG	AACCATATTA	
-830	-820	-810	-800	-790	-780	
AGGATCACCC	ATTAGCAGCG	TGGGGGCTGG	GTCCATATTC	NATATTTTCC	CCATGACTTC	
-770	-760	-750	-740	-730	-720	
CTCAAGTATG	GCNATGTCAG	CCGGTGGGGG	NAAATATACA	CAAACCAAGT	TAATTTAACTG	
-710	-700	-690	-680	-670	-660	
TTGTGGAAG	GAACATTGCA	TAACAATATA	TCTACCAAGT	CGATCAGTAT	GTATCTCCAC	
-650	-640	-630	-620	-610	-600	
TATGTTGAGA	TTTATTTTTT	TACTCACTAG	GATTGCCACT	CCTCTAGAGT	AGTTTGAGAG	
-590	-580	-570	-560	-550	-540	
CTCCGCATGA	TACACTCTTG	CAACCCACGC	TTTTTTCAAA	GCTAGTATGC	GGCCACCTGT	
-530	-520	-510	-500	-490	-480	
AAGGTGTGTT	TCCTGAAGGA	GAACCAAGTG	TGGGCGGTAT	TTTTTGAGAT	ATTGGAATAC	
-470	-460	-450	-440	-430	-420	
CAGGCCTCTT	TTGTGCTGAA	AAGGTTTACC	TTGGCTTATA	CTCGAGTCAG	TATAAACCAG	
-410	-400	-390	-380	-370	-360	
GCATGTCCAA	AGTGTGGTTT	GCGGGGCGCG	ATGCTTCGCT	GTGACAGTAT	CTAAACTGAC	
-350	-340	-330	-320	-310	-300	
TGGGGTGTTT	CGCATCTCTG	GGCTCTACTT	CCTGTCTACT	GGACCAAAGA	ATGTACTTGC	
-290	-280	-270	-260	-250	-240	
GCAAGTAGGT	GTGCACGGGA	TTGTGAGGGC	ATGAACAAGC	ACAGCACAGC	AGTGGGAAGG	
-230	-220	-210	-200	-190	-180	
CCCTGATACT	GCAGCCCTAG	TCCGACCTTG	TTGATACCGA	TCCGTTGCCG	TGNAGCATTG	
-170	-160	-150	-140	-130	-120	
AGGGACTCCT	CCCGTTAGCT	GGGGGCAAAAC	AGGTCTATCA	CCTGAAGCTT	TGTGTTTGGT	
-110	-100	-90	-80	-70	-60	
GAAGTCCCAT	TGCCCATGCC	CCGCTCCCTG	GGCAGTGGGC	ACCCGTGCTGC	TGCCTCTGTT	
-50	-40	-30	-20	-10	0	
ATATCTGCAG	CTCACCATTG	TTCATCTACA	GTCCAAGCCC	CCCACACTTC	ACAGCAGCAC	

Figure VII.5b

Start	10	20	30	40	50	60
ATGTGCTCTGA	AAGAGCCCTT	CCTACTACAA	ACTGCCCTCA	GCATTATAGT	GGCCCTGATG	
70	80	90	100	110	120	
ATGCACCAAG	TAATGTCCGG	AGAGGCCCCG	GGGCTTGGCT	CAGGCTACAT	GGCATTTTAG	
130	140	150	160	170	180	
CTAAAACTAA	<u>AGCTTATTGT</u>	GGTGATAAT	GTAAAAATTGT	AACTAAAATA	GTATTTCAT	
190	200	210	220	230	240	
ACTGTATATA	TAGGAAAGTA	TTTAATAACA	TGCTATAGTC	TTGACTTGTA	GATTAAAAACA	
250	260	270	280	290	300	
CTGAAACAAG	AAACCATTTT	ACGGTATATG	GGTTTTACAG	CATGTAGCTG	CTGTATTTC	
310	320	330				
CACCCTAGGC	<u>TTATGCTCGA</u>	<u>GTCAATAAGC</u>	TT			

DNA sequencing was performed according to the strategy on Figure VII.4 and by the single-stranded dideoxy method. Numbering is relative to the translational start site (start) defined by DNA sequence comparison with the *X.laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990).

Figure VII.6a

Comparison between *X. laevis* N-Cadherin cDNA  
and *X. borealis* N-Cadherin genomic sequences  
(Clone 3-9/4.8BS/pBS)

Upper sequence: *X. laevis* N-Cadherin cDNA

Lower sequence: *X. borealis* N-Cadherin genomic DNA

```

-183                                     X.laevis
GCGGAGCACAGGATTCTCTGGAAATCAGCCCT   GCCTTGTGTTTG   CTCCGT
* * * * *                               * * * * *
GGGC AA ACAGG   TCT   ATCA   CCTGAAGCTTTGTGTTTGGTGAAGTCC C
-156                                     X.borealis

-131                                     X.laevis
ACGTGCCCCATGCCCCGCTCCC   CGGCACCTGCCACTGCTGCTGCTGAGTGTGTAAGGCAC
* * * * *                               * * * * *
AT TGCCCCATGCCCCGCTCCCTGGCAGTGGGCAC   CTTGCTGCTGCCTCTGT
-111                                     X.borealis

-73                                     X.laevis
GACTGTATGTGCTGCTGCTGCTCGTCATTGTTTCATCTCCAGTCCAAGCCTCCGCCGAC C
* * * * *                               * * * * *
TATAT   CTGCAGCTCACCATTGTTTCATCTACAGTCCAAGC   CCCCCGACAC
-60                                     X.borealis

-12                                     X.laevis
CCACAGCATCACCATGTGCCGAAACAGCCCTTCCTGTACCGACTCTACTCGGCATCCT
***** * * * * * * * * * * * * * * * * * * * * * * * * *
TCACAGCAGCAC   ATGTGTCTGAAAGAGCCCTTCCTACTACAACTGCCCTCAGCATTAT
-11                                     X.borealis

48                                     X.laevis
AGCGGCCCTGATGCTGCAGCAGGGACCAGTT   GAAGCATT   CGGGGGATCCAGATTATG
* * * * *                               * * * * *
AGTGGCCCTGATGATGCACCAAGTAA   TGTCCGAG   AGGCCCCGGGGCTTG   GCTCA G
48                                     X.borealis
      exon/intron

```

Figure VII.6b

```

105                                     X.laevis
CAAGACA GG ATTTCTTG AGGA T   GTTTATCACGCCAGTGTCTACAG AAG TG
*** ** * * * * * * * * * * * * * * * * * * * * * * * *
GCT ACATGGCATTTTAGCTAAAACATAAGCTTATTGTG   GTGTATAATGTAAATTG
104                                     X.borealis

155                                     X.laevis
TACATGAA GGGCAGCCGCTT C   TAAATGT GA   TGTTTACTGACTGTGGTACTG
** * * * * * * * * * * * * * * * * * * * * * * * *
TAACTAAAAATAGTATTTGCATACCTGTATATATAGGAAAGTATTTAATAACA TGCTA TA
159                                     X.borealis

206                                     X.laevis
ATAGACGGATACAGTATGAAACCACT AACCCAACAGA   TTTT CGGATTGATGGTGA
* * * * * * * * * * * * * * * * * * * * * * * *
GTCTTGACTTGTAGATTAAAAC ACTGAAAC AAGAAACCATTTTACGG TATATGGGTT
218                                     X.borealis

261                                     X.laevis
317
TGGAATTGTGTTTGCCCTCAAGAACCTTTGACATTT CTCCA GAGC AGGCAG AATTTT
* * * * * * * * * * * * * * * * * * * * * * * *
TTACAGCATGT AGCTGCTGTA TTTCTCACCTAGGCTT ATGCTCGAGTCAATAAGCTT
275                                     X.borealis
332

```

Sequences are aligned and spaces introduced manually so as to provide maximum homology. Numbering is relative to the translational start site (underlined) of *X.laevis* N-Cadherin cDNA (DETRICK *et al.*, 1990).

The proposed exon/first intron junction is also marked (+70).

Total homology between 5'-ends and proposed exon/intron junction is 68%; obviously, over shorter stretches it is considerably higher.

Figure VIII.7

Inter-species comparison of N-Cadherin amino-acid sequences from translational start site

<i>X. borealis</i>	(3-9/EMBL3)	M C L K E	P F L L Q T	A L S I I V A L M M H
<i>X. laevis</i>	(DETRICK <i>et al.</i> , 1990)	M C R K Q	P F L L P T L	L G I L A A L M L Q
<i>X. laevis</i>	(GINSBERG <i>et al.</i> , 1991)	M C R K E	P F L L P T	A L C I L A A L V L H
Mouse	(MIYATANI <i>et al.</i> , 1989)	M C R I A G G R G T L L P		L L A A L L Q
Chick	(HATTA <i>et al.</i> , 1988)	M C R I A G T P P R I L P P L A L M L L A A L		Q

The amino-acid differences between *X. borealis* and *X. laevis* (DETRICK *et al.*, 1990) are detailed below (5' to 3'):*X. borealis*

L = leucine (non-polar) -  
 E = glutamic acid (negatively charged) -  
 Q = glutamine (polar, uncharged) -  
 A = alanine (non-polar) -  
 S = serine (polar, uncharged) -  
 I = isoleucine (non-polar) -  
 V = valine (non-polar) -  
 M = methionine (non-polar) -  
 H = histidine (positively charged) -

*X. laevis*

R = arginine (positively charged)  
 Q = glutamine (polar, uncharged)  
 P = proline (non-polar)  
 L = leucine (non-polar)  
 G = glycine (polar, uncharged) [GINSBERG has C]  
 L = leucine (non-polar)  
 A = alanine (non-polar)  
 L = leucine (non-polar)  
 Q = glutamine (polar, uncharged) [GINSBERG has H]

It is interesting to note that the more 3' changes do not alter the charge/polarity of the amino-acids (see above note for the last change). These amino-acids form the signal-sequence, which does not contribute to the mature protein.

*X.borealis* genomic amino-acid sequence has a higher degree of homology to this latter Israeli sequence than to the original probe. It is also interesting to note that the amino-acid differences detailed between the *X.borealis* and *X.laevis* sequences fall into two groups: the three 5'-changes alter the charge/polarity of the amino-acid; the larger 3'-group of changes do not. This appears to indicate that a greater selective pressure is put upon the more-3' amino-acids not to alter their charge/polarity.

Given that the degree of homology between members of the Cadherin family (as a whole) is low at the 5'-end of the published cDNAs, and taking into account the above DNA and amino-acid evidence, it can be stated with reasonable confidence that the 3-9/4.8BS/pBS clone contains genomic DNA sequence from the *X.borealis* N-Cadherin gene.

#### 7.5 DNA-binding protein motifs present upstream of translational start

The most obvious sequence motif in the putative promoter region of the 3-9/4.8BS/pBS clone is the TATAAA element at -429bp. This element conforms exactly with the consensus TATA A/T A A/T, and is ideally placed to act as the site of transcriptional initiation. This would place the start of the mRNA at between -404bp and -396bp, according to presently accepted models. Further possible DNA-binding sites are given in Figure VII.8; obviously, these sites require confirmation by, at the very least, DNA-footprinting studies and gel-

Figure VII.8

Known DNA sequence motifs present in *X. borealis* N-Cadherin promoter region

(Clone 3-9/4.8BS/pBS)

Motif	Consensus	<i>X. borealis</i> sequence	Location	Reference
1. TATA box	TATA A/T A A/T	TATAA	-429 (100%)	
2. Sp1	GGGCGG	GGGCG	-508 (100%)	
3. CCAAT box	CCAAT	(CCAAT)	-485 (100%)	
4. GATA box	T/A GATA A/G	(AGATAG)	-861 (100%)	YAMAMOTO <i>et al.</i> , (1990)
5. Homeodomain	CPyPyNATTA T/G C/T	aCCCAATTAGC	-833 (90%)	ODENWALD <i>et al.</i> , (1987)
6. EF2	TTTT G/C G/C CG G/C	TTTTCCCa	-795 (89%)	BAGCHI <i>et al.</i> , (1990)
7. Homeodomain	CPyPyNATTA T/G C/T	gTTAAATTAaC	-731 (80%)	ODENWALD <i>et al.</i> , (1987)
8. MyoD	CAACTGAC	CAACgGc	-872 (75%)	MURRE <i>et al.</i> , (1989)
9. V-erbA	GTGTCAAAGGTCA	GgGgCAAcGGTCt	-876 (69%)	DE VERNEUIL <i>et al.</i> , (1990)

Bracketed sequences refer to non-coding DNA strand. Sequence locations refer to 5'-ends of motifs as shown in Figure VII.5. Percentage figures after sequence locations are %-homologies with consensus sequences. Lower case implies mismatched dNTPs.

shift analyses.

#### 7.6 Frequency of occurrence of TAT and ATA triplets

in the 3-9/4.8BS/pBS *X.borealis* N-Cadherin promoter region

As discussed in Chapter IV, TATA-boxes are known to be placed in regions of relatively low AT content. Therefore, to provide further contextual evidence that the TATA-box is the site of transcriptional initiation, the frequency of the DNA triplets TAT+ATA was plotted for the putative promoter region (Figure VII.9). The graph obtained exactly mimics that of Nussinov (NUSSINOV *et al.*, 1986) in the positions of both of the main peaks: one peak lies approximately 275bp 5' to the transcriptional start site (-675 on Figure VII.9); the second peak coincides with the TATA-box.

Therefore, as both the TATA-box sequence and context conform exactly with generally-accepted models, it is reasonable to assume that transcriptional start occurs between 22 and 28bp 3' to the end of the TATA-box, that is between -404bp and -396bp relative to the translational start site. The RNA-based assays detailed in Chapter VIII provide some evidence to confirm this statement.

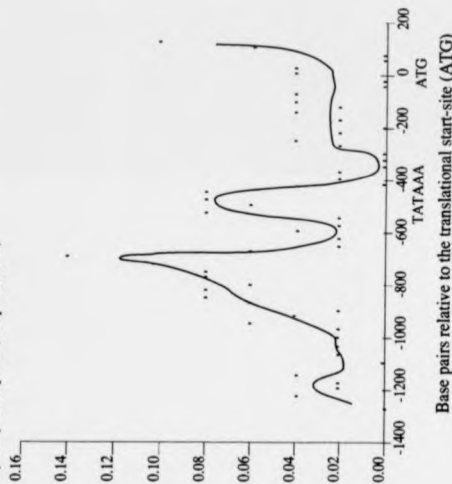
#### 7.7 Analysis of CG/GC value in *X.borealis* 'promoter' region

As discussed in Chapter IV, eukaryotic promoter regions have often been found to have an abnormal ratio of the dinucleotides CG and GC. In the majority of eukaryotic DNA the value of CG/GC is approximately 0.2 (BIRD, 1986). In the NCAM promoter this ratio rises to an average of 0.63 in the 1.4Kb upstream of the translational

Figure VII.9

Graph of occurrence of ATATAT triplets in *X. borealis* N-Cadherin promoter region (Clone 3-9/4.8BS/pBS)

Frequency of triplets (50bp window)



The frequency of ATA+TAT triplets in a 50bp window is plotted against the position of the mid-point of that window for the genomic clone 3-9/4.8BS/pBS.

In accordance with Nussinov *et al.* (1986), the main peaks occur around the area of the TATA box (at -425bp) and also approximately 275bp 5' to the TATA box (at -700bp).

start-site. An analysis of the CG/GC value in the putative N-Cadherin promoter region (clone 3-9/4.8BS/pBS) was undertaken. The results are shown in tabular form in Figure VII.10 and plotted graphically in Figure VII.11. These figures show that the N-Cadherin 'promoter' also has a increased CG/GC value: the average over the 1.3Kb upstream of the translational start-site is 0.60.

#### 7.8 Analysis of translational start-sites

The translational start-site of the published *X.laevis* N-Cadherin sequences was defined by inter-species sequence alignment in a similar manner to that given in Figure VII.7. In the absence of any corroborating experimental evidence, this does not define the start unambiguously particularly since both *X.borealis* genomic clones and the *X.laevis* cDNA of DETRICK *et al.* (1990) contain AUGs upstream of the proposed translational start-site. Further details are given in Figure VII.12.

In the case of the *X.borealis* clone 3-9/4.8BS/pBS there are four upstream AUGs, the first three of which terminate at in-frame stop codons relatively shortly after translational initiation. The fourth, at -104, is in frame with the proposed translational start-site and, if anything, is in a slightly better context than the AUG at +1bp. Out of the three AUGs detailed for the *X.laevis* cDNA (DETRICK *et al.*, 1990) the proposed AUG at +1bp appears to be in the best context. Upstream AUGs are known to be negative regulators of translation: after translating these short peptides, the ribosome complex has

Figure VII.10

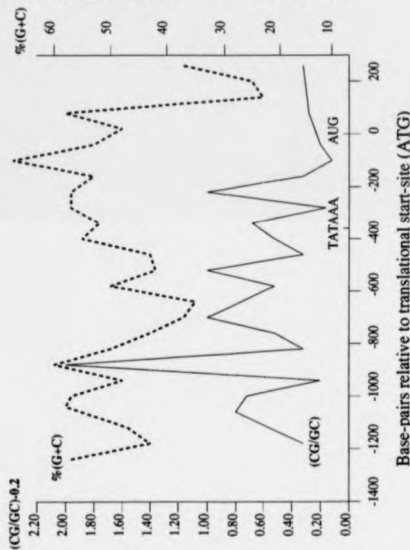
Analysis of CG/GC content of *X.borealis* N-Cadherin promoter region  
(Clone 3-9/4.8BS/pBS)

	<u>Base pair Range</u>		<u>%(G+C)</u>	<u>CG/GC</u>		<u>(CG/GC)-0.2</u>
1.	-1259	-1200	57	0/4	-	-
2.	-1199	-1140	43	1/3	0.33	0.13
3.	-1139	-1080	47	0/1	-	-
4.	-1079	-1020	58	4/5	0.80	0.60
5.	-1019	-960	57	5/7	0.71	0.51
5.	-959	-900	48	1/4	0.25	0.05
6.	-899	-840	60	4/2	2.00	1.80
7.	-839	-780	50	1/3	0.33	0.13
8.	-779	-720	43	1/2	0.50	0.30
9.	-719	-660	37	1/1	1.00	0.80
10.	-659	-600	35	0/2	-	-
11.	-599	-540	50	3/6	0.50	0.30
12.	-539	-480	42	1/1	1.00	0.80
13.	-479	-420	43	1/3	0.33	0.13
14.	-419	-360	55	3/6	0.50	0.30
15.	-359	-300	52	2/3	0.67	0.47
16.	-299	-240	57	1/7	0.14	-0.06
17.	-239	-180	57	4/4	1.00	0.80
18.	-179	-120	53	1/3	0.33	0.13
19.	-119	-60	67	1/8	0.13	-0.07
20.	-59	0	53	1/5	0.20	0.00
21.	1	60	48	0/4	-	-
22.	61	120	58	2/7	0.29	0.09
23.	121	180	23	0/2	-	-
24.	181	240	25	0/1	-	-
25.	241	300	37	1/3	0.33	0.13

The above data shows the %(G+C) content (average is 40% for normal DNA) and the deviation from the normal ratio of CG/GC (i.e. 0.2) for 60bp windows in the genomic clone 3-9/4.8BS/pBS. This data is also plotted graphically in Figure VII.11.

Figure VII.11

Graph of  $\%(G+C)$  and  $CG/GC$  ratio for *X.borealis* N-Cadherin promoter region  
(Clone 3-9/4.8BS/pBS)



The  $\%(G+C)$  content (data from Figure VII.10) is plotted against its position in the genomic DNA relative to the translational start-site (ATG). The average  $\%(G+C)$  for eukaryotic DNA in general is 40%; it can be seen that the majority of this region of DNA exceeds this level.

Also plotted is the ratio of  $CG/GC$  dinucleotides. In eukaryotic DNA in general, the average  $CG/GC$  ratio is 0.2. It can be seen that this region of DNA greatly exceeds the average.

Figure VII.12a

Analysis of translational start-sites in *X.laevis* N-Cadherin cDNA  
and *X.borealis* genomic clones

The adjacent figure illustrates the variation in translational start-site sequences in two of the *X.borealis* genomic clones and two published *X.laevis* cDNAs. These are compared against the consensus translational start-site as defined by Kozak (1986).

The *X.laevis* cDNA translational start site was defined purely by inter-species DNA sequence alignment with other N-Cadherin homologues.

Three short open reading frames (ORFs) are present in the *X.borealis* 3-9/4.8BS/pBS genomic clone, upstream of two possible initiation sites. Without firm transcriptional start data, it is not possible to determine whether these ORFs are transcribed. (It is not impossible that the region containing these ORFs is contained within an intron).

Figure VII. 2b

Analysis of translational start sites in *X.laevis* N-Cadherin cDNA and *X.borealis* genomic clones

	AUG position	Most conserved bases
Eukaryotic mRNA (KOZAK, 1986)		<div> <div>CC</div> <div>A</div> <div>C</div> <div>C</div> <div>AUG</div> <div>G</div> <div>80% A</div> <div>18% G</div> <div>40% G</div> </div>
<i>X.borealis</i> 3-9/4.8BS/pBS	[-389 (AUG) -368 (UAA)]	
"	[-309 (AUG) -294 (UAG)]	
"	[-269 (AUG) -236 (UGA)]	
"	-104	UG C C C AUG C
"	I	AG C A C AUG U
<i>X.borealis</i> 3-21/6.8BS/pBS	-80	GA C A G AUG A
"	I	GC A C C AUG U
<i>X.laevis</i> cDNA (DETRICK <i>et al.</i> , 1990)	-123	UG C C C AUG C
"	-66	AC U G U AUG U
"	I	UC A C C AUG U
<i>X.laevis</i> cDNA (GINSBERG <i>et al.</i> , 1991)	I	GC A C C AUG U

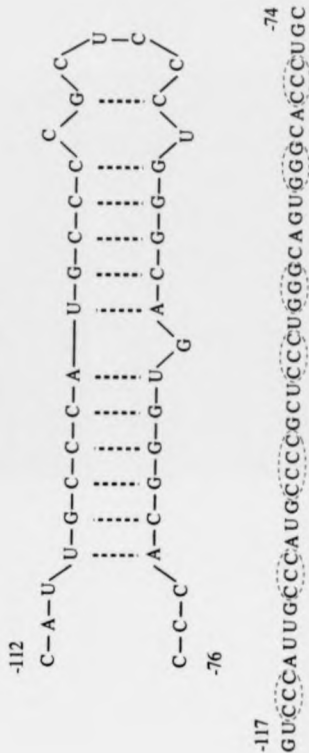
difficulty reinitiating, and hence the main protein does not get produced (e.g. *lck* oncogene, MARTH *et al.*, 1988; yeast GCN4, FINK, 1986). In the absence of any experimental data, it is not possible to speculate further on the effect that these upstream AUGs have on translation of the protein.

#### 7.9 Potential secondary structure upstream of the translational start-site

The difficulty experienced in mapping the region upstream of the proposed translational start-site by RNA-based assays (see Chapter VIII and also the similar problems reported by SORKIN *et al.*, 1991 and NEDIVI *et al.*, 1992), lead to speculation about whether the transcription of this region might produce RNA that is rich in secondary structure. The formation of any hairpin loops would affect any laboratory assay that utilised ssRNA as part of its procedure and also would have an effect on translation (by making the ssRNA less available to the ribosome). On examination, it was found that this region is rich in repeated bases, particularly triplets of rC and rG. Computer-generated matrix plots ("Diagon" plots) have revealed a number of possible hairpin loop structures in the region between the putative translational and transcriptional start-sites; an example is given in Figure VII.13. It can easily be seen how different combinations of the  $rC_n$  and  $rG_n$  groups circled in Figure VII.13 could base-pair to form a number of different loop structures. It is shown in Chapter VIII that control RNase protection assays are possible using

Figure VII.13

Diagram of potential secondary structure in RNA upstream of translational start site  
(Clone 3-9/4.8BS/pBS)



This figure illustrates the possible secondary structures that can be formed in the region of DNA upstream of the translational start-site (+1). From Figure VII.10, it can be seen that the region between -60bp and -119bp is 67% (G+C). There therefore exists ample scope for the formation of relatively stable secondary structures in the RNA. The top diagram illustrates a possible hair-pin loop structure; the lower diagram shows the multiplicity of poly-C and poly-G tracts in this region. The relevance of these structures to RNA-based assays is discussed in Section 8.16.

an excess of synthetic RNA: but the assay conditions (hybridisation in 80% formamide) are obviously not representative of those conditions found *in vivo*. Therefore, whether the length of the loops and the plurality of possible structures would have any relevance under *in vivo* conditions is uncertain.

#### 7.10 Intron/exon structure of *X. borealis* genomic sequence

As mentioned above, the degree of DNA sequence homology between the *X. borealis* genomic clone and *X. laevis* cDNA is nearly absolute between the translational start-site and an AGGT sequence at +70bp. This AGGT sequence, and the immediately adjacent bases, conform well to the known consensus donor splice site of an exon/intron junction. Also, as shown in Figure VII.14, stop codons appear in all three reading frames within 78bp of this junction and the sequence 3' to the AGGT is characterised by simple repeating motifs and single-nucleotide stretches. It would be reasonable to assume therefore that this AGGT sequence marks the boundary of the first intron. There is a possible consensus acceptor splice site at +293bp but the sequence 3' to this does not appear to be homologous with the cDNA to anywhere near the same extent as the base-pairs between +1bp and +70bp.

The position of this first intron is confirmed by sequence comparison with other published Cadherin intron/exon boundaries (see Figure IV.2). This boundary occurs in mouse N-Cadherin at Gln-20, in chick L-CAM at Gln-23, and in the herein defined *X. borealis* N-Cadherin sequence at Gln-23.

Figure VII.14

Analysis of stop codons downstream of translational start  
(Clone 3-9/4.8BS/pBS)

Start

ATG TGT CTG AAA GAG CCC TTC CTA CTA CAA ACT GCC CTC AGC  
(2)

EXON/intron

ATT ATA GTG GCC CTG ATG ATG CAC CAG gta atg tcc gga gag  
(2) (2) (2) 2

gcc ccg ggg ctt ggc tca ggc tac atg gca ttt tag cta aaa  
1 2

cta aag ctt att gtg gtg tat aat gta aaa ttg taa cta aaa  
2 3 2 1 2

tag tat ttg cat act gta tat ata gga aag tat tta ata aca  
1 2 2 2

tgc tat agt ctt gac ttg tag att aaa aca ctg aaa caa gaa  
3 3 1 3 2

acc att tta cgg tat atg ggt ttt aca gca tgt agc tgc tgt  
3

att tct cac cct agg ctt atg ctc gag tca ata agc tt  
3 2

The above listing details the positions of stop codons (TAA, TGA, and TAG) down-stream of the translational start-site of the *X.borealis* N-Cadherin genomic DNA sequence (clone 3-9/4.8BS/pBS). The numbers below the DNA triplets refer to the reading frame of the stop codon, i.e. 1 means it is in the correct reading frame; 2 and 3 are both out of frame with respect to the translational start-site.

It can be seen that, within 50 bases of the putative exon/intron junction, an in-frame stop codon exists (tag); stop codons in all three reading frames lie within 78 bases of this junction.

The splice site conforms well to the standard consensus sequence:

	exon	intron
Consensus:	C/A A G	G T A/G A G T
Actual:	C A G	G T A A T G

(MOUNT, 1982)

### 7.11 Conclusions

It has been shown in this chapter, by DNA and amino-acid sequence alignment, that the 3-9/4.8BS/pBS clone appears to contain DNA sequence that is the *X.borealis* homologue of the N-Cadherin gene. Theoretical analyses of this sequence also appear to define a number of possible protein-DNA binding sites, the transcriptional and translational start-sites and the position of the first exon/intron boundary.

## Chapter VIII

### Confirmation of transcriptional start-site:

#### Experimental evidence

#### 8.1 Introduction

#### 8.2 RNase protection assays

- (i) Introduction
- (ii) *X.borealis* histone H4 clone
- (iii) *X.borealis* genomic clone: HindIII-HindIII
- (iv) *X.borealis* genomic clone: PstI-PstI
- (v) *X.borealis* genomic clone: XbaI-Pst
- (vi) *X.borealis* genomic clone: XbaI-HindIII
- (vii) RNase protection assay conclusions

#### 8.3 Primer extensions

- (i) Introduction
- (ii) *X.borealis* histone H1 oligo
- (iii) *X.laevis* N-Cadherin oligo: NCAD5
- (iv) *X.borealis* genomic oligo: NCADJ1 oligo
- (v) *X.borealis* genomic oligos: NCADB2 and NCADZ5
- (vi) Primer extension conclusions

#### 8.4 A PCR-based method for the determination of transcriptional start-site

#### 8.5 Conclusions

## 8.1 Introduction

In this chapter experimental evidence is provided in an attempt to confirm that the TATA-box described in Chapter VII is in fact used in the initiation of transcription of the *X. borealis* N-Cadherin mRNA. Two RNA-based procedures are described: an RNase Protection Assay (RPA) that utilises a labelled anti-sense RNA probe to hybridise to the mRNA; and a Primer Extension assay that utilises a labelled anti-sense DNA oligonucleotide that binds to the mRNA and then is extended to the 5'-end of the mRNA by avian reverse transcriptase. Reference is also made to a PCR technique whose component steps were established but whose procedure was not completed.

Whilst the control experiments used to establish these procedures provided no problems, obtaining reproducible data from the *Xenopus borealis* RNA proved a lot more difficult.

## 8.2 RNase protection assays (RPAs)

### 8.2(i) Introduction

The use of these procedures and necessary controls has been discussed already in Chapter VI.

The rationale for using a RNase protection assay to determine the transcriptional start site is as follows:

1. From DNA sequence comparison of the *X. laevis* cDNA and *X. borealis* genomic clones, a region of *X. borealis* genomic DNA could be unambiguously classified as corresponding to mRNA (see Figure VII.6), be it protein-coding or untranslated leader sequence.
2. From this known site, overlapping, increasingly 5', RNA probes could be made from this genomic DNA and used to determine the extent of continuous

transcription (i.e. mRNA) 5' to this site. Transcription would be interrupted only by the transcriptional start-site or by a 3' intron splice site; it was hoped that the former of these would be found.

There is a limit to the resolution attainable with this method as shorter probes may bind non-specifically to mRNAs of other related genes, if sufficient homology is present, giving false results; results obtained from longer probes are therefore more significant.

A further point, that is often overlooked but that needs to be recognised, is with regard to the use of RNA probes on denaturing-PAGE gels using DNA markers. It is reported (SAMBROOK *et al.*, 1989) that the mobility of RNA and DNA on denaturing-PAGE may differ by up to 10%; this difference is reduced when running the gels at higher voltage/power. Care must therefore be taken when sizing the protected RNA bands against DNA markers. In the experiments detailed below, adjustment for this difference was made by comparing the known sizes of the RNA probe and sense-RNA/probe control with the DNA markers and then working out the necessary adjustment coefficient. Protected RNA bands could then be sized from a graph of log (DNA nucleotides) v distance-moved of the DNA markers. At the conditions used, this value was found to be in the range of 0 - 6%. RNA markers can be made, but the greater ease with which DNA markers are made and the greater stability of DNA markers has meant that DNA markers are routinely used.

The DNA markers used were HaeIII-digested pBR322, HpaII-digested pBR322 or DNA sequence; often a combination of these was used for greater accuracy. The experimental details, with regard to the performance of the assays, are given in Chapter V.

The clones used for the *X.borealis* RPAs and the construction of each are detailed below and in Figure VIII.1. The nucleotide sequence numbers refer to the DNA sequence as given in Figure VII.5.

#### 8.2(ii) *X.borealis* histone H4 clone

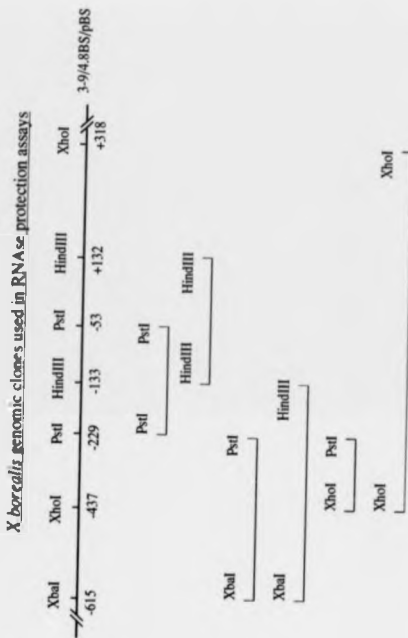
This clone was used as a positive control for the RNA extraction procedure to ensure the integrity of the extracted *X.borealis* embryonic RNA; details of the clone and the gel are given in Figure VIII.2. The clone comprises a 430bp *X.borealis* histone H4 cDNA and T3 RNA polymerase promoter site that have been inserted into the EcoRI-HindIII site of pBluescript SK+. RNA probe was made by linearising the clone with EcoRI and transcribing with T7 RNA polymerase. Two control sense-RNAs were made by linearising with HindIII and transcribing with T3 RNA polymerase; two protected bands are thus seen in the control-sense-RNA track. The gel illustrates that the RNA extraction procedure used is entirely adequate to provide full-length mRNA.

#### 8.2(iii) *X.borealis* genomic clone: HindIII-HindIII

As shown in Chapter VI, this RNase protection procedure was shown to work effectively using the *X.laevis* N-Cadherin cDNA/*X.laevis* RNA and the following reaction conditions: 15µg total RNA, 50,000cpm of labelled probe and a 2 day exposure of the resultant gel. These conditions were therefore the obvious initial ones to use with the *X.borealis* RNA probes/*X.borealis* embryonic RNA.

An attempt was made to establish the RPA reaction conditions with *X.borealis* embryonic RNA using the HindIII-HindIII fragment, this being one that, theoretically (based on the data in Chapter VII), should be completely protected. The HindIII-HindIII fragment (-133bp to +132bp) was cloned

Figure VIII.1



The figure shows a section of the map of genomic clone 3.9/4.8BS/pBS together with the fragments that were used in RNAse protection assays on *X. borealis* RNA as described in this Chapter.

Figure VIII.2a

RNAse Protection Assay gel

*X.borealis* histone H4 clone



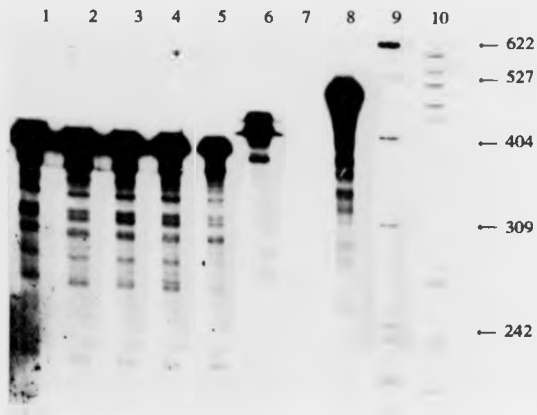
Track no:

1. Stage 10
2. Stage 12
3. Stage 18
4. Stage 8
5. Stage 20
6. Control: synthetic RNA/tRNA
7. Control: Probe + RNAse
8. Control: Probe - RNAse
9. Markers: HpaII digested pBR322
10. Markers: HaeIII digested pBR322

All above embryonic RNAs refer to 5µg *X.borealis* total RNA

Amount of probe used = 20,000cpm

Figure VIII.2b



The above gel demonstrates that both the *X.borealis* RNA extraction procedure and the RNase protection assay procedures were working sufficiently to detect the *X.borealis* histone H4 RNA.

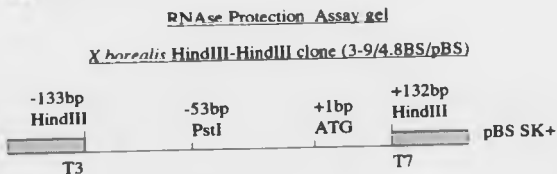
directly into pBluescript SK+ and, after DNA sequencing to check the correct insert, it was linearised with EcoRI and then transcribed with T7 RNA polymerase to make the labelled probe. The control sense-RNA was made by linearising with EcoRI and transcribing with T3 RNA polymerase.

The *X. laevis* reaction conditions described above produced no result when using the HindIII-HindIII probe with *X. borealis* stage 22 total RNA. The range of variables was therefore expanded using an array of conditions: 50µg, 25µg, 10µg (stage 22, total RNA) v 50,000cpm, 25,000cpm, 10,000cpm (labelled probe) was tried (Figure VIII.3). This gave a high background of probe degradation products in all embryo-RNA tracks, decreasing in strength as the amount of labelled probe was reduced. A further array of variables produced a similar result of high background in all tracks (stage 10 (20µg and 40µg), stage 22 (20µg and 40µg), stage 22 (0.2µg and 2µg polyA+) RNA v 10,000cpm, 50,000cpm, and 100,000cpm probe, data not shown).

The failure of this clone to produce a result could be due to one or more of a number of factors. The range of experiments performed and controls used rule out the majority of them however.

- a) Degradation of the RNA. The integrity of the RNA was verified by using the histone H4 clone as a positive control for each new RNA preparation (although the abundance of histone H4 is greater than that of N-Cadherin).
- b) Presence of factors within the RNA preparation that prevent the assay from working. Both the use of the histone H4 clone and the use of synthetic RNA/embryo RNA control rule out this possibility.
- c) The probe forming an internal hairpin loop. As shown in Figure VII.9, this region of DNA can form a number of hair-pin loop structures. The synthetic

Figure VIII.3a

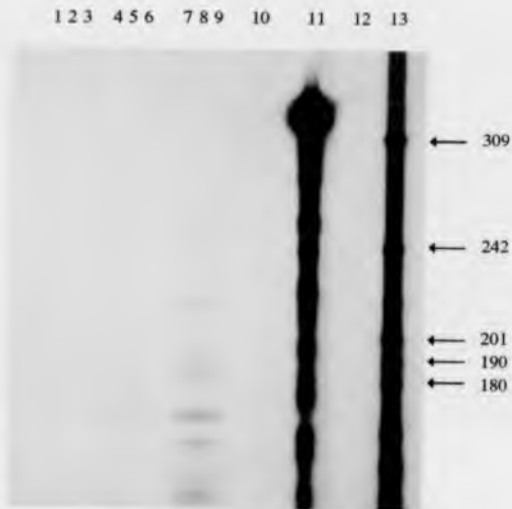


Track no:

<u>Total RNA (ug)</u>	<u>Probe (cpm)</u>
1. 10	10
2. 25	10
3. 50	10
4. 10	25
5. 25	25
6. 50	25
7. 10	50
8. 25	50
9. 50	50
10. Control: Probe + RNase A	
11. Control: Probe - RNase A	
12. Control: Probe + RNase T1	
13. Markers: HpaII digested pBR322	

All embryonic RNA used was *X. borealis* stage 22.

Figure VIII 3b



The above gel illustrates the failure of the HindIII-HindIII probe from the 3-9/4.8BS/pBS *X. borealis* clone to detect homologous RNA from *X. borealis* embryo RNA, despite the fact that a wide variety of probe and RNA concentrations were tested.

RNA control shows, however, that some binding of probe and RNA may occur under the RPA conditions used. The control does not show whether any, or a significant proportion, of the probe is unavailable for binding to the mRNA; it only shows that some probe is still available in single-stranded form to bind to the synthetic message. The synthetic message may be present (in this control) in excess over the endogenous N-Cadherin mRNA.

Another alternative which cannot be controlled for is the possibility that the majority of the mRNA itself is in the form a hair-pin loop and so is not available for the probe to bind to.

d) Binding to non-specific RNAs. If regions of the HindIII-HindIII clone show sufficient homology to other RNAs (in particular to other Cadherin RNAs), non-specific binding may occur and so 'soak-up' a significant proportion of the available probe, depleting the pool of probe available to bind to the desired sense homologue.

e) The assay is not sensitive enough to detect the desired mRNA. The RPA is a very sensitive technique as it involves the use of a continuously-labelled, high-specific activity probe and is considered to be more than twenty times more sensitive than Northern blots. As shown above, the RPA was found to work adequately in *X.laevis* with 3-15 $\mu$ g total RNA, and 50,000cpm probe. If conditions were found that differed markedly from the *X.laevis* conditions this would imply a widely different abundance of the *X.borealis* message. This would be highly unlikely if the herein identified DNA sequence is indeed the true homologue of the *X.laevis* N-Cadherin gene.

f) Activity of DNase and RNase. Incomplete digestion of the *in vitro* transcribed DNA template may result in the RNA probe hybridising to this

template thus giving high backgrounds. Lack of RNase activity of the RNase A and/or RNase T1 would also produce high backgrounds. Both of these effects are controlled for in the "Probe + RNase A" control.

g) The HindIII-HindIII clone does not form part of the mRNA. This must remain an option given the lack of any viable result.

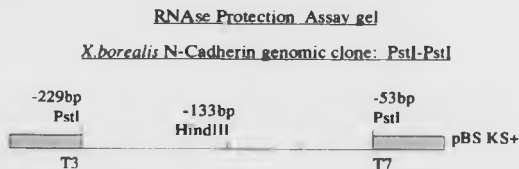
#### 8.2(iv) *X.borealis* genomic clone: PstI-Pst I

This clone was produced by cloning the PstI-PstI fragment (-229bp to -53bp) directly into pBluescript KS+. Probe was made, after the relevant checks on the insert, by linearising the clone with EcoRI and transcribing with T7 RNA polymerase. The control sense-RNA was made by cloning the XhoI-XhoI fragment (-437bp to +318bp) into pBluescript SK+, linearising with EcoRI and transcribing with T7 RNA polymerase.

Similar experiments were performed with this clone as with the above HindIII-HindIII clone in order to try to find reaction conditions that would produce an interpretable result. Variations in the amount of RNA, amount of probe and embryo stage produced negative results except for a fully-protected band seen in the egg (Figure VIII.4).

As opposed to other experimental systems such as Northern or Southern blotting, the presence of a protected band in RNase protection assays means that the homology between probe RNA and mRNA is almost absolute. As discussed above (Section 6.4(iv)), the presence of just one mismatched RNA base may result in the endonucleolytic action of one of the RNases used at that point in the RNA. The presence of this band in egg is therefore highly significant in terms of DNA homology. The egg RNA was initially added as a (supposedly) negative control; (this RNA being readily available following the

Figure VIII.4a



Track no:

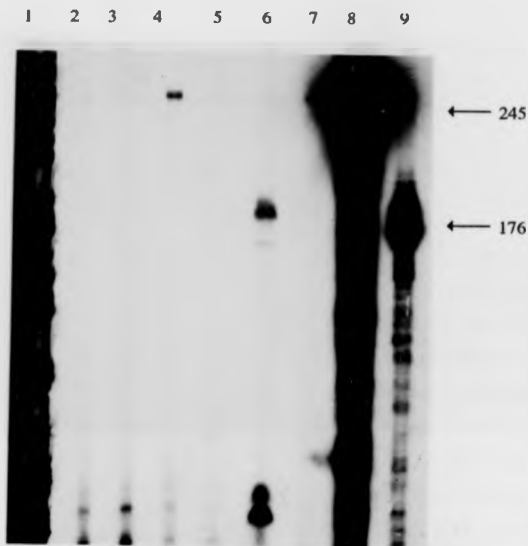
1. Markers: DNA sequence
2. *X.borealis* stage 37
3. *X.borealis* stage 23
4. *X.borealis* stage 20
5. *X.borealis* stage 15
6. *X.borealis* egg
7. Control: Probe + RNase A
8. Control: Probe - RNase A
9. Control: Probe + synthetic RNA (XhoI-XhoI fragment)

Probe amount:  $2.7 \times 10^5$  cpm

RNA amount: 20µg total

Figure VIII.4b

(Predicted RNA sizes)



The probe used in this gel was the *X. borealis* N-Cadherin genomic clone PstI-PstI (-229 to -53, from 3-9/4.8BS/pBS); a range of embryonic tissues were assayed.

Although the probe consistently detected a signal from egg, no other tissues produced signals. A range of other RNA and probe amounts were also tested, the results of each being the same as above.

failure of a *Xenopus* natural-mating). No nervous-system-specific RNA would be expected to be seen in the egg and so the question remains as to whether the bands are truly due to N-Cadherin or some other highly homologous (at least at the extreme 5' end) gene. It is interesting to note that in a comparison of other Cadherins published to date, the 5'-end of the mRNA contains sequences that are amongst the least conserved between Cadherins. Obviously further investigation, in the form of a stage series/distribution analysis and the use of more-3' RNA probes, is necessary to evaluate the significance of this result.

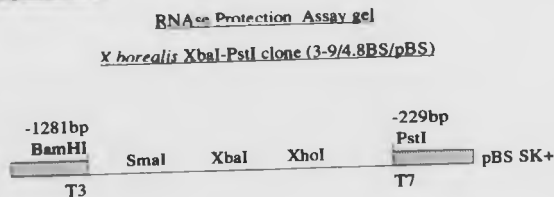
The failure to find any bands in any other embryonic stages may be due to one of the options discussed above in Section 8.4.

#### 8.2(v) *X.borealis* genomic clone: XbaI-PstI

This clone spans the putative TATA site. The PstI-BamHI fragment (-229bp to -1281bp) was cloned directly into pBluescript SK+. The clone was linearised with XbaI and transcribed with T7 RNA polymerase. The sense RNA was made from the same clone but linearised with EcoRI and transcribed with T3 RNA polymerase.

According to the published expression sequence of the *X.laevis* N-Cadherin gene (DETRICK *et al.*, 1990), expression of N-Cadherin is first seen (in RPAs) at stage 12, increases up to stage 20 and is specific to neural and muscle tissues. In Figure VIII.5, expression is just seen weakly at stage 8 (which had been included as a negative control, based on Detrick's results), strongly at stage 12 and at stage 25; expression of the same size band is also seen in a stage 32 tadpole brain. A band of this size maps to approximately the TATA box, but only on the assumption that the 3'-end of the protected band extends to the Pst site (-53bp); this proviso cannot be proven from this gel

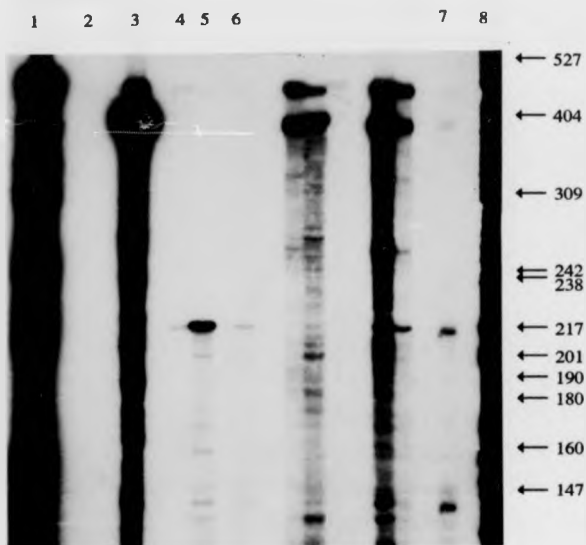
**Figure VIII.5a**



**Track no:**

1. Control: Probe + RNase A
2. Control: Probe - RNase A
3. Control: Probe + synthetic RNA
4. *X. borealis* stage 8
5. *X. borealis* stage 12
6. *X. borealis* stage 25
7. *X. borealis* stage 32 tadpole brain
8. Markers: HpaII digested pBR322

Figure VIII.5b



The XbaI-PstI probe protects a 202nt band (corrected RNA size) in stages 8, 12 and 25 *X. borealis* embryos and more significantly, in tadpole brain. If the 3'-end of this protected band maps to the PstI site, then the other end of the band would map just 3' to the TATA-box.

alone (but see below).

#### 8.2(vi) *X.borealis* genomic clone: XbaI-HindIII

This clone was made as above but then linearised with PstI. The PstI-PstI (-229bp to -53bp) fragment was then ligated into the clone and the necessary checks performed. The clone was then linearised with HindIII, liberating the HindIII fragment (between -133bp and the pBS polylinker site) and then religated. To make the probe, the clone was linearised with XbaI and transcribed with T7 RNA polymerase. The results are given in Figure VIII.6. Bands of 216nts are seen in the stage 8 and 12 tracks; the bands in the stage 18, 20 and 25 tracks are believed to be probe degradation products as similar length bands are seen in the control tracks.

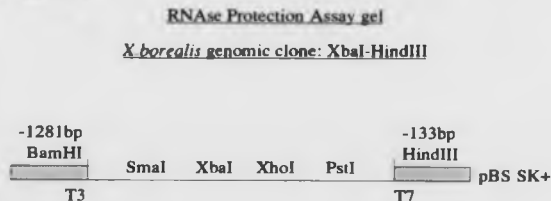
The result that this gel produces is very useful when used in conjunction with the XbaI-PstI gel result. As both the above two clones have a common 5'-end, any longer bands seen on the latter gel must be extended at the 3'-end. The fact that a longer band is seen on the XbaI-HindIII gel indicates that the 3'-end of the protected RNA from the XbaI-PstI clone must be PstI site, thus confirming the 5'-end of protection near the TATA box.

The failure of the PstI-HindIII region to be completely protected ties in with the failure of the PstI-PstI clone to produce a viable result (in tissues other than egg). It may imply the presence of the start of an intron (at approx -210bp) or may be due to one of the factors discussed above in Section 8.2(iii).

#### 8.2(vii) RNAse protection assay: conclusions

It has been found not possible to follow the rationale laid out at the start of this chapter for the RPAs due to the failure of the PstI-PstI and HindIII-HindIII fragments to produce viable results. The results from the XbaI-PstI and

Figure VIII.6a

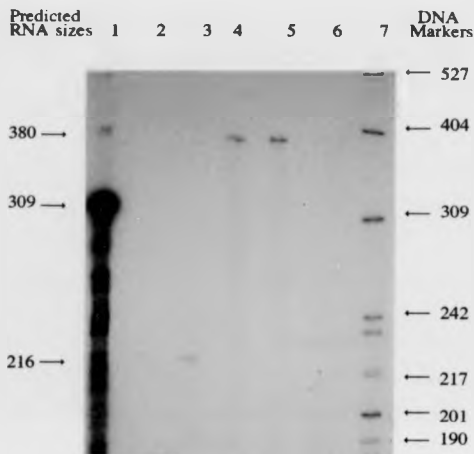


Track no:

1. Control: synthetic RNA/*X. laevis* egg
2. *X. borealis* stage 8
3. *X. borealis* stage 12
4. *X. borealis* stage 18
5. *X. borealis* stage 20
6. *X. borealis* stage 25
7. Markers: HpaII digested pBR322

The additional controls (Probe - RNase A and Probe + RNase A) were run on a separate gel (not shown); the bands produced were entirely as expected.

Figure VIII.6b



The above gel shows that the XbaI-HindIII probe protects a band of approximately 216nts in stage 12 *X.borealis* embryo RNA and also weakly in stage 8. In the stage 18, 20 and 25 *X.borealis* embryo RNA tracks, the only bands seen appear to correspond to probe degradation products (as seen in the synthetic-control track, 1.)

XbaI-HindIII fragments, when combined, point to the approximate region of the TATA box as either the site of an intron/exon junction or the start of transcription. The inaccuracies inherent in sizing these relatively large probes prevent any fine-scale dissection of the results based on this data alone.

The results presented above represent only the results that were consistently reproducible. RPAs were attempted using a number of other clones (see Figure VIII.1) in an attempt to supplement the above results in terms of resolution and, in particular, to carry out further stage series and controls. Further experimental data is certainly required if any conclusions are to be put on a firmer scientific footing.

### 8.3 Primer extension assays

#### 8.3(i) Introduction

Four oligos were made for this study: one *X.laevis* oligo was made from the published *X.laevis* N-Cadherin cDNA sequence and three *X.borealis* oligos were made from the *X.borealis* genomic DNA region that had high homology to the *X.laevis* N-Cadherin cDNA (and was therefore assumed to be mRNA-coding).

The rationale for synthesizing three *X.borealis* oligos was that the results obtained with one oligo could be confirmed by the results from the other oligos if the obtained extended bands were at the same distance apart as the positions of the oligos themselves in the genomic DNA. The extended bands could then be said to be specific to the DNA sequences of all three oligos.

The oligos were carefully chosen so as to avoid excessive G+C content, the presence of simple repetitive motifs, hair-pin loops or of similar sequences to those found in other parts of the gene (as far as the sequence data was

available). High G+C content oligos can bind non-specifically and relatively strongly to RNA regions that also have high G+C contents. It can also easily be seen how the presence of simple repetitive sequences or potential hair-pin loop structures would be undesirable. The probes used are given in Figure VIII.7, and highlighted where they occur in the relevant DNA sequence.

Since the primer extensions were carried out in low salt aqueous buffer, the melting temperature of the oligos ( $T_m$ ) could be approximated from the Wallace Rule:  $T_m = 4(G+C) + 2(A+T) ^\circ C$ . This value was not relied upon, however, but was verified by actual binding studies at various temperatures between oligo and synthetic RNA/embryo RNA background (Figure VIII.8). These experiments gave a value for  $T_m$  at which specific binding of the oligo would be expected to decrease; it therefore gave an additional control with which to judge whether the binding of the oligo was specific or not.

The experimental procedure used for the primer extension assays is described in Chapter V. The integrity of the RNA was established by the RPA procedure using the histone H4 clone described in Section 8.2(ii).

It must be remembered that the end product of the primer extension assay is a labelled single-stranded piece of DNA; this may be sized directly against labelled DNA markers on a denaturing-PAGE gel.

### 8.3(ii) *X.borealis* histone H1 oligo

A *X.borealis* histone H1 oligo was used in establishing the primer extension protocol (BAGENAL, 1990). It served to confirm the correct operation of the protocol and also as further verification of the integrity of the RNA.

The oligo used was a 17mer that spans the translational start site and produces an extension product of 44nts. The details of the oligo and the

# Figure VIII.7

## *X. laevis* N-Cadherin and genomic DNA oligos used in primer extension assays

### *X. laevis* N-Cadherin oligo

-180                      -170                      -160                      -150                      -140  
CGG GAGCACAGGA TTCTCTGGA AATCAGCCCT GCCTTGTGTT  
 NCAD5

The numbering is based on DETRICK *et al.* (1990), relative to the translational start-site.

### *X. borealis* genomic oligos

-110                      -100                      -90                      -80  
 GAAGTCCCAT TGCCCATGCC CCGCTCCCTG GGCAGTGGG  
 -70                      -60                      <sup>PstI</sup> -50                      -40  
 ACCCTGCTGC TGCCTCTGTT ATATCTGCAG CTCACCATG  
 NCADB2

-30                      -20                      -10                      0  
 TTCATCTACA GTCCAAGCCC CCGACACTC ACAGCAGCAC  
 NCADJ1

Start              10                      20                      30                      40  
ATGTGTCTGA AAGAGCCCTT CCTACTACAA ACTGCCCTCA  
 NCADZ5

The numbering is relative to the translational start-site, as given in Figure VII.5.

Figure VIII.8a

$T_m$  determination of oligonucleotides used in primer extension assays

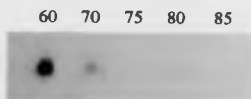
The melting temperature ( $T_m$ ) was determined experimentally for each of the oligonucleotides used in the primer extension assays (except the histone H1 control, for which the  $T_m$  had already been established).

The hybridisation step of the primer extension assay was performed at various temperatures using synthetic RNA to establish the temperature above which binding was significantly reduced. The results obtained with these controls could then be used to establish the binding characteristics of each oligonucleotide to homologous RNA.

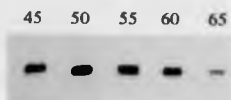
The results obtained with corresponding assays using embryonic RNA could then be compared with these control results to determine whether the binding seen was indeed specific. The temperatures given are in °C.

Figure VIII.8b

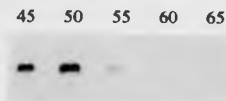
*X. laevis* N-Cadherin oligo: NCAD5



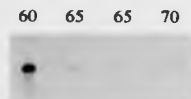
*X. borealis* genomic oligo: NCADJ1



*X. borealis* genomic oligo: NCADB2



*X. borealis* genomic oligo: NCADZ5



resultant gel are given in Figure VIII.9.

### 8.3(iii) *X.laevis* N-Cadherin oligo: NCAD5

As with the above RPA experiments, the initial embryonic experiments using this assay were performed using a system that, it was hoped, would act as a positive control for the assay. An oligo was therefore made at the extreme 5'-end of the *X.laevis* N-Cadherin cDNA, as described in Figure VIII.7, for use on *X.laevis* embryo RNA.

The *X.laevis* N-Cadherin cDNA obtained from Chris Kintner (see Chapter VI; DETRICK *et al.*, 1990) was 4.0Kb long, whereas published Northern blot data put the full length mRNA at 4.3kb. It was therefore initially assumed that the cDNA was missing 0.3kb at the 5' end (due to possible premature termination of the reverse transcriptase); upon reflection, this was not a valid assumption. Since the initial transcription of the cDNA (from mRNA) was primed by an oligo dT primer, there is no guarantee that this primer would hybridise to the very end of the mRNA, as such 3' ends of mRNA are known to be poly-A rich over a length of RNA that possibly extends for hundreds of nucleotides. Consequently, the missing 0.3kb could easily reside at the 3'-end of the cDNA.

It is entirely possible therefore that the 5'-end of the cDNA is coincident with the 5'-end of the mRNA; an oligo made to this 5'-end would therefore have no 'extending' to do. An oligo to sequences, say, -100bp from the 5'-end should have been made as an initial step; this would have confirmed, or otherwise, the position of the 5'-end. If the 5'-end of the mRNA did reside further upstream from the known sequence, then the oligo made (to the extreme 5'-end of the cDNA) would have been a logical second step.

**Figure VIII.9a**

**Primer extension assay gel:**

*X borealis* histone H1 oligonucleotide

Sequence of *X. borealis* histone H1 oligonucleotide:

TTCAGCCATTAGCAAATCTCTGTCGTTTAAACAAAGACAAACAC  
5' XbH1a 3'

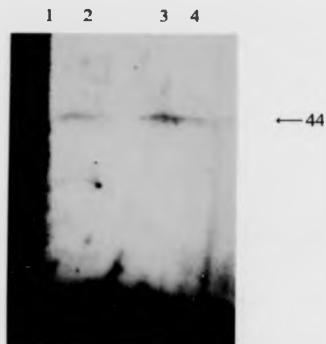
The underlined sequence corresponds to the oligonucleotide termed XbH1a (BAGENAL, 1990); it spans the translational start site (overlined).

**Track no:**

1. Markers: DNA sequence
2. *X.borealis* egg, 10µg
3. *X.borealis* stage 37, 10µg
4. *X.borealis* stage 22, 10µg

**All hybridisations at 37°C overnight.**

Figure VIII.9b



The detection of a 44nt band in all embryo RNA tracks shows that both the primer extension assay and RNA extraction procedure were working to a degree necessary to detect this semi-abundant mRNA.

The results obtained using the NCAD5 oligo, not surprisingly, were negative (data not given).

#### 8.3(iv) *X. borealis* genomic oligo: NCADJ1

Figure VIII.10 show the results obtained with this oligo. A single band of 351nts is seen in stage 8, 18 and 23; it has also been shown to be present at stages 20 and 25 (data not shown). If there are no intervening introns, this would place transcriptional start at -349bp (based on Figure VII.5 numbering).

#### 8.3(v) *X. borealis* genomic oligos: NCADB2 and NCADZ5

These oligos consistently failed to produce any results when used under the same conditions as the NCADJ1 oligo.

#### 8.3(vi) Primer extension assays: conclusions

As with the RPA experiments, these experiments also fall short of being conclusive and therefore need supplementing, at the minimum, with additional stage controls before any firm inferences can be drawn. Ideally, the assays using either NCADB2 or NCADZ5 would have produced a band at the appropriate distance (shorter or longer in relation to NCADJ1) which would have confirmed that the binding of NCADJ1 was indeed specific.

Many of the reasons discussed in Section 8.4 are relevant here as to the question of why a negative result was obtained with these two oligos. In particular, should either oligo have a high homology to a non-specific RNA, then this RNA could 'soak-up' a significant proportion of oligo, thus reducing the pool of oligo available to bind to the N-Cadherin mRNA; situations such as this cannot be foreseen.

#### 8.4 A PCR-based method for the determination of transcriptional start-site

An attempt was initiated to combine the primer extension assay with PCR in

Figure VIII.10a

Primer extension assay gel:

*X.borealis* genomic oligonucleotides NCADJ1 and NCADB2

Track no:

1. Markers: HpaII digested pBR322

NCADJ1

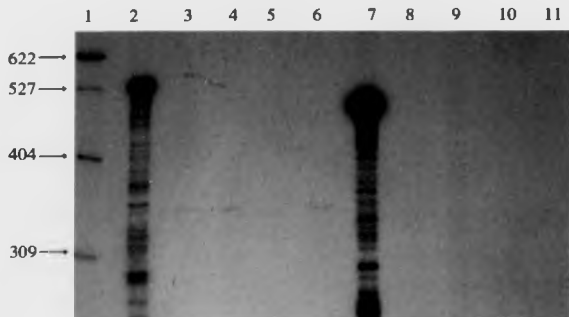
2. Control: synthetic sense RNA  
3. *X.borealis* stage 25, 50µg, 60°C  
4. *X.borealis* stage 25, 10µg, 60°C  
5. *X.borealis* stage 20, 50µg, 60°C  
6. *X.borealis* stage 20, 10µg, 60°C

NCADB2

7. Control: synthetic sense RNA  
8. *X.borealis* stage 25, 50µg, 50°C  
9. *X.borealis* stage 25, 10µg, 50°C  
10. *X.borealis* stage 20, 50µg, 50°C  
11. *X.borealis* stage 20, 10µg, 50°C

The control sense RNA used was the XhoI-XhoI (-437bp to +318) cloned into pBluescript; the temperature referred to is the temperature of hybridisation of probe to template RNA.

Figure VIII.10b



The bands seen in the NCADJ1 lanes are approximately 341nts in length. In the absence of any introns in the genomic DNA in the region between the translational and transcriptional start-sites, this would map the transcriptional start-site to approximately 70bp downstream of the TATA-box.

Additional oligonucleotides have now been made to confirm the specificity of this result and at a greater resolution.

order to produce sufficient primer extension product to be able to sequence. Each individual step of this novel assay was adapted from known procedures and then fully worked out using control reagents, but insufficient time prevented performance of the complete procedure.

The rationale is given below, whereas a more complete experimental method is given in Appendix B and illustrated in Figure B.1.

- a) Hybridisation of oligo to mRNA and extension with reverse transcriptase;
- b) Separation of unused oligo from extended product on PAGE;
- c) Poly-dG tailing of the extended product using terminal transferase and dGTP;
- d) PCR using a second oligo and oligo-dC as primers;
- e) Isolation and sequencing of product using third oligo.

The advantage of using this technique is that sequence data is obtained directly that, by a comparison with the genomic DNA, would illuminate transcriptional start and the presence of any introns in one step (assuming that the genomic DNA sequence extended beyond any introns found).

It was envisaged that the primer extension would be done with the 3'-most oligo, PCR would be done with the second oligo, and the sequencing would be done with the 5'-most oligo. The sequenced product would therefore be specific for all three oligos ensuring a high degree of specificity. The constraints of time, however, intervened to defeat this goal.

## 8.5 Conclusions

Although the preliminary results obtained with the RNase protection and primer extension assays are encouraging, the objectives discussed in the preambles to both RPA and primer extension assays have not yet been

translated into conclusive results capable of answering the main question posed by this chapter.

By combining the results from the XbaI-PstI and XbaI-HindIII RNase protection assays, it has been shown that the region of DNA from approximately -200bp (relative to the translational start-site) to the TATA box forms part of the mRNA population. The NCADJ1 oligo has also been shown to extend by 351nts from -20bp which, in the absence of any introns, would map transcriptional start just downstream of the TATA box.

The appearance of a protected band in the RPAs in the *X.borealis* egg remains an anomaly that requires further investigation. The appearance of bands in stage 8 embryos tracks may be due to inaccuracies in staging pre-gastrula embryos as neural induction commences shortly afterwards (and indeed, many people believe that neural induction starts prior to gastrulation).

In conclusion, the above results clearly give positive pointers towards the fact that the region of the TATA box is likely to be where transcriptional initiation occurs but further data is required to confirm this absolutely and with greater resolution. In this regard, additional primer extension experiments are underway using oligos at distances 100-200bp 3' to the TATA box. It is hoped that these experiments or the PCR assay will resolve this issue.

## Chapter IX

### Expression of *X.borealis* N-Cadherin

#### promoter/ $\beta$ -galactosidase constructs in *Xenopus* embryos

#### 9.1 Introduction

- (i) Promoter assays
- (ii)  $\beta$ -galactosidase expression vectors used
- (iii) Micro-injection of pCaSpeR-AUG- $\beta$ gal constructs

into *Xenopus* embryos

#### 9.2 Construction of pCaSpeR-AUG- $\beta$ gal promoter constructs

- (i) BPP/ $\beta$ gal
- (ii) BbPP/ $\beta$ gal
- (iii)  $\alpha$ PB/ $\beta$ gal
- (iv) BPH/ $\beta$ gal
- (v) BPP(Eco-)/pBS
- (vi) XbaPP/ $\beta$ gal
- (vii) XhoPP/ $\beta$ gal
- (viii) PP/ $\beta$ gal
- (ix) BP/ $\beta$ gal
- (x) SmaPP/ $\beta$ gal
- (xi) pCaSpeR-AUG and pCaSpeR- $\beta$ gal

#### 9.3 Controls for the expression of micro-injected plasmids into *Xenopus* embryos

- (i) Form of micro-injected DNA

(ii) Quantity of micro-injected DNA

(iii) Localization of micro-injected DNA

9.4 Expression patterns of promoter deletion/ $\beta$ -galactosidase constructs in *Xenopus laevis*

9.5 Explant and inductive sandwich experiments

9.6 Transcriptional start-site used by promoter-deletion/ $\beta$ gal constructs

9.7 Conclusions

## 9.1 Introduction

### 9.1(i) Promoter assays

The fusion of promoter deletions to the CAT gene is a well established and popular procedure for the analysis of promoter DNA sequences. The CAT assay is a simple chromatographic process that results in easily-quantifiable data. The disadvantage of this method is that the resolution of the areas of CAT protein expression are limited by the extent to which the expressing tissues can be dissected or isolated. In *Xenopus*, dissections can be made at the level of some single organs or tissues, but these can be very time-consuming to perform; smaller divisions rapidly become impractical.

An alternative, or rather a complement, to such assays are those in which reporter gene expression patterns are visualised directly. Whilst this approach does not result in quantitative data, it does provide a wealth of information on the fine-scale expression patterns of the reporter gene.

The advantages of utilising the *Xenopus* system in such studies are legion and have been discussed in Chapter I.

The  $\beta$ -galactosidase staining procedure was recommended by Richard Harland (pers. comm.) and had previously been successful in mice (Sanes *et al.*, 1986).

### 9.1(ii) $\beta$ -galactosidase expression vectors used

The vector used in the  $\beta$ -gal. expression studies was pCaSpeR-AUG- $\beta$ gal (THUMMEL *et al.*, 1988). This vector contains coding sequences

from the *E.Coli*  $\beta$ -galactosidase and *Drosophila white* genes, and part of the pUC vector bounded by P-elements; their relative orientations are shown on Figure IX.1.

The  $\beta$ -galactosidase gene in this vector is bounded at the 5'-end by a short polylinker (EcoRI, BamHI, KpnI) and *Drosophila* alcohol dehydrogenase (Adh) AUG translational start site; the 3'-end has a SV40 polyA-tail, giving enhanced mRNA stability. The presence of the Adh AUG obviates the need for in-frame splicing of promoter sequences into the polylinker site.

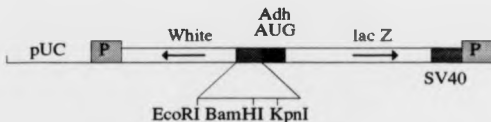
This vector has been used extensively for enhancer-trapping (OKANE and GEHRING, 1987) in *Drosophila*. P-element-mediated integration of the vector into the *Drosophila* genome results in the *white-eye* phenotype and  $\beta$ -galactosidase expression only if the integration site is adjacent to enhancer/promoter sequences. The resulting  $\beta$ -galactosidase expression pattern is taken to be analogous to that of the gene controlled by the adjacent enhancer/promoter sequences.

All constructs were cloned into the EcoRI-BamHI-KpnI polylinker (as detailed below), transformed into the *E.Coli* bacterium strain BB4, and then a large-scale DNA preparation was performed using the CsCl-gradient method. Most vectors were injected into embryos in the linearised form, these vectors having been cut 5' to the inserted N-Cadherin sequences.

In retrospect, the choice of vector was fortuitous. It has since been

Figure IX.1

pCaSpeR-AUG- $\beta$ gal ( $\beta$ -galactosidase vector)



The pCaSpeR-AUG- $\beta$ gal vector comprises the *Drosophila white* and *E.Coli lacZ* genes, together with the *Drosophila* alcohol dehydrogenase (*Adh*) gene translational start-site, AUG (THUMMEL *et al.*, 1988). The *Adh* AUG is in frame with, and directs transcription into, the *E.Coli lacZ* gene. These DNA sequences are bounded by P-element ends (P) and inserted into the pUC vector.

The sense orientation of the *white* and *lacZ* genes are indicated above by the arrows.

The majority of *X.borealis* N-Cadherin genomic DNA constructs were inserted into the BamHI-KpnI sites of the polylinker and linearised with EcoRI.

learned that the choice of vector in particular plays an important role in the level of non-specific expression: pUC and pBR322 vectors are said to give acceptable backgrounds but pGEM vectors do not (T.J. Mohun, NIMR Mill Hill, pers. comm.).

#### 9.1(iii) Microinjection of pCaSpeR-AUG- $\beta$ gal into *Xenopus* embryos.

Initial microinjection trials with this vector alone (i.e. minus any insert) in *X. laevis* embryos gave negligible  $\beta$ -galactosidase background expression. This indicated that there was no undirected expression from the Adh AUG from any source within the embryos. Promising results were seen with N-Cadherin promoter sequences inserted into the polylinker; a deletion series of the N-Cadherin promoter was therefore constructed in this vector (Figure IX.2).

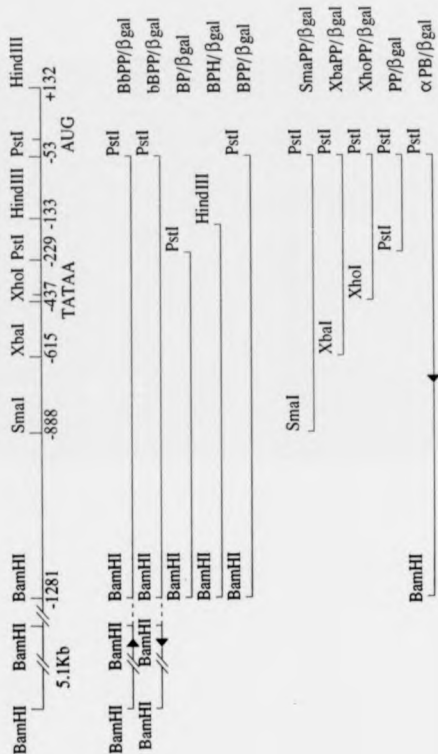
All micro-injection experiments were performed into *X.laevis* embryos as these embryos are larger than those of *X.borealis*; they were also more readily available in this laboratory. Similar results (and possibly more specific) would be expected from micro-injection into *X.borealis* embryos.

After the analysis of a number of necessary controls, a range of promoter deletion constructs were made as described below. These were inserted into a pUC-based vector upstream of a  $\beta$ -galactosidase reporter gene and micro-injected into *X.laevis* embryos. I gratefully acknowledge the assistance of Dr. E.A. Jones in performing the micro-injections of the DNA constructs.

#### 9.2 Construction of pCaSpeR-AUG- $\beta$ gal promoter clones

Figure IX.2

Construction of *X. borealis* N-Cadherin promoter-deletion/ $\beta$ -galactosidase clones



(See "Note" on page 216a.)

Note:

The 5.1Kb BamHI-BamHI fragment is not shown joined directly to the BamHI (-1281) - HindIII (+132) fragment for the following reason:

The 0.5, 0.4, 1.4 and 1.6Kb fragments shown at the 5'-end of the BamHI map in Figure VI.9 (page 147) could not be unambiguously assigned the positions shown due to the small size of these fragments.

It is therefore possible that the 5.1Kb BamHI-BamHI fragment is not contiguous with the BamHI-HindIII fragment, i.e. these latter fragments may be separated by one of the small fragments mentioned above.

In all the cases mentioned below "βgal" refers to the pCaSpeR-AUG-βgal vector as defined by Thummel *et al.* (ibid.). The genomic fragments referred to below (by base pair numbers) are detailed in Chapter VII (Figure VII.3).

All βgal clones were both single and double-digested to check for the presence of only one copy of the correct insert before large-scale DNA preparations were made by the CsCl gradient method.

#### 9.2(i) BPP/βgal

BamHI-PstI (-1281bp to -229bp) and PstI-PstI (-229bp to -53bp) fragments were gel-isolated from the 3-9/4.8BS/pBS clone. The BamHI-PstI fragment was cloned directly into pBS SK+ and then linearised with PstI. The PstI-PstI fragment was then ligated into this clone. The correct orientation of the PstI-PstI fragment in the resulting minipreps (termed BPP/pBS) was checked by DNA sequencing from the pBS M13 primer site into the PstI-PstI fragment.

The resulting BamHI-PstI-PstI genomic sequence (-1281bp to -53bp) was removed from pBS by digestion with BamHI+KpnI, gel-isolated, and cloned into the BamHI-KpnI site of the β-gal vector. This vector could be linearised with BamHI.

#### 9.2(ii) BbPP/βgal and bBPP/βgal

The 5.1Kb BamHI-BamHI fragment was gel isolated from the 3-9/EMBL3 clone (Figure VI.9) and ligated into the BamHI site of the above BPP/β-gal clone. The orientation of the BamHI-BamHI fragment was determined by restriction digests using EcoRI. Clones

with both orientations were used (BbPP/ $\beta$ gal and bBPP/ $\beta$ gal). These clones could not be linearised at the extreme 5'-end of the BamHI-BamHI fragment as this fragment contains an EcoRI site. Linearization with EcoRI produces a clone with either 1.3 or 3.8Kb of genomic DNA, depending on the orientation of the 5.1Kb clone.

#### 9.2(iii) $\alpha$ PB/ $\beta$ gal

BPP/pBS (Section 9.4(i)) was cut with BamHI+EcoRI, the liberated genomic sequence (-1281bp to -53bp) gel-isolated and cloned into the EcoRI-BamHI site of the  $\beta$ gal vector. This clone could be linearised with EcoRI.

This manipulation places the genomic sequence 3'-5' (i.e. in reverse orientation) with respect to the  $\beta$ -galactosidase gene. It was therefore used as a negative control (in addition to vector without insert).

#### 9.2(iv) BPH/ $\beta$ gal

BPP/pBS (Section 9.4(i)) was digested with HindIII and the resulting bands resolved on a low-melting point agarose gel. The larger (main plasmid plus BamHI-PstI-HindIII fragment) band was gel purified, religated, transformed into BB4 and minipreped. In this way the genomic sequences between the HindIII site (at -133bp) and the 3' pBS HindIII site were removed.

Restriction digests were performed on a number of minipreps to verify the loss of the HindIII fragment. Positive clones were also DNA sequenced from the pBS M13 primer across the HindIII site into the genomic sequence to double-check.

The genomic sequence (-1281bp to -133bp) was then removed from pBS by digestion with BamHI+KpnI, gel-isolated, and cloned into the BamHI-KpnI site of  $\beta$ -gal. This clone could be linearised with EcoRI.

#### 9.2(v) BPP(Eco-)/pBS

BPP/pBS (Section 9.4(i)) was cut EcoRI, blunted with Klenow I, and religated to remove the EcoRI site from the 3' pBS polylinker. This clone (termed BPP(Eco-)/pBS) was then sequenced from the pBS M13 primer site to verify that the EcoRI site was removed. This clone was the basis for subsequent clones that needed to be linearised at the 5' end by EcoRI.

#### 9.2(vi) XbaPP/ $\beta$ gal

BPP(Eco-)/pBS was cut with XbaI, blunted with Klenow I and then cut with KpnI; the genomic DNA was then gel-isolated. pCaSpeR-AUG- $\beta$ gal was cut BamHI, blunted with Klenow I and then cut with KpnI; the main plasmid was then gel-isolated. The XbaI (blunted)-KpnI fragment genomic DNA fragment (-615 to -53) was then ligated into the pre-prepared  $\beta$ gal vector. This clone could be linearised with EcoRI.

#### 9.2(vii) XhoPP/ $\beta$ gal

BPP(Eco-)/pBS was cut with XhoI and the liberated genomic fragment cloned directly into pBS KS+. The orientation of the XhoI fragment was determined by DNA sequencing. This clone was then digested with BamHI+KpnI, the genomic DNA fragment (-437 to -53) was gel isolated and cloned into [BamHI+KpnI]-cut  $\beta$ gal. The

orientation of the genomic DNA was such that it read 5' to 3' into the  $\beta$ gal coding sequence. This clone could be linearised with EcoRI.

#### 9.2(viii) PP/ $\beta$ gal

Genomic DNA fragment PstI-PstI (-229 to -53) was cloned directly into pBS; its orientation was determined by DNA sequencing. This clone was digested EcoRI+BamHI to liberate the genomic DNA. The genomic fragment (-229 to -53) was gel-isolated and then cloned into [BamHI+EcoRI]-cut  $\beta$ gal vector. This clone could be linearised with EcoRI. The orientation of the genomic DNA was such that it read 5' to 3' into the  $\beta$ -galactosidase coding sequence.

I thank S.A. Burbidge for his assistance in cloning the PstI-PstI genomic fragment into the  $\beta$ gal vector.

#### 9.2(ix) BP/ $\beta$ gal

BPP(Eco-)/pBS was cut with PstI to liberate the PstI-PstI (-229 to -53) fragment; the main plasmid band was gel-isolated, and religated. This clone was then cut BamHI+KpnI to release the genomic DNA, the genomic fragment (-1281 to -229) was gel-isolated and then cloned into [BamHI+EcoRI]-cut  $\beta$ gal vector. This clone could be linearised with EcoRI.

#### 9.2(x) SmaPP/ $\beta$ gal

BPP(Eco-)/pBS was cut with SmaI+KpnI; the genomic band was gel-isolated.  $\beta$ gal was cut with BamHI, blunted and then cut with KpnI. The genomic fragment (-888 to -53) was then cloned into the pre-prepared  $\beta$ gal vector. This clone could be linearised with EcoRI.

#### 9.2(xi) pCaSpeR-AUG-βgal and pCaSpeR-βgal

These vectors (with no genomic clone inserts) were also injected to act as controls. pCaSpeR-βgal is similar to pCaSpeR-AUG-βgal but has no AUG translational start-site (THUMMEL *et al.*, 1988). Both could be linearised with EcoRI.

#### 9.3 Controls for the expression of micro-injected plasmids into *Xenopus* embryos.

##### 9.3(i) Form of micro-injected DNA

Micro-injection experiments performed by other groups have failed to show conclusively the form of DNA (i.e. linear or circular) that yields the best expression of the plasmid (see Chapter I). The ambiguous results in the literature indicate that micro-injection experiments should be preceded by controls to investigate this point. The preliminary experiments therefore involved injecting a number of constructs independently in both linear and circular forms.

Four clones were used: αPB/βgal (reverse orientation promoter fragment) and pCaSpeR-AUG-βgal, as controls; and BPP/βgal and XbaPP/βgal, as promoter deletions. Micro-injections were performed into one cell of a two-cell embryo and analysed at stages 10-10.5 and 16. The results are shown in Figures IX.3 -IX.6 and tabulated in Table IX.1. Uninjected controls are shown in Figure IX.10.

Both of the control clones produced very low levels of β-galactosidase expression demonstrating:

- (1) that there was very low levels of endogenous β-

Figure IX.3a

Comparison between  $\beta$ -galactosidase expression patterns obtained  
from circular and linearised micro-injected DNAs

pCaSpeR-AUG- $\beta$ -gal - circular DNA

Stage 10-10.5



Stage 16



Figure IX.3b

pCaSpeR-AUG- $\beta$ -gal - linearised DNA

Stage 10-10.5



Stage 16

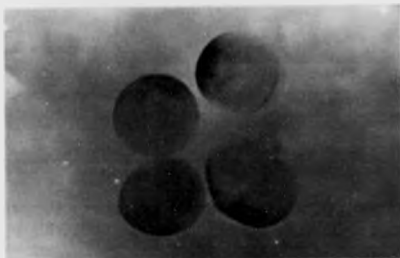


Figure IX.4

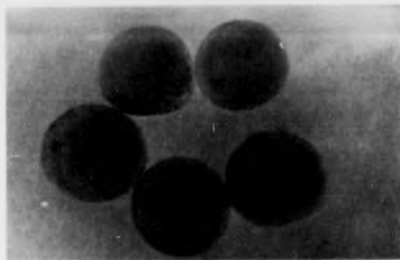
Comparison between  $\beta$ -galactosidase expression patterns obtained  
from circular and linearised micro-injected DNAs

$\alpha$ PB/ $\beta$ gal - circular DNA

Stage 10-10.5



Stage 16



(See Figure IX.11 for linearised clones)

Figure IX.5a

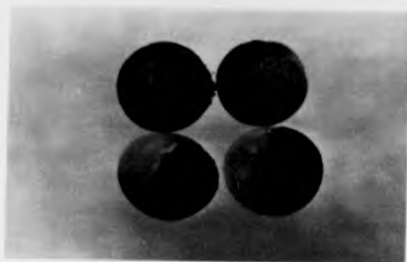
Comparison between  $\beta$ -galactosidase expression patterns obtained  
from circular and linearised micro-injected DNAs

BPP/ $\beta$ gal - circular DNA

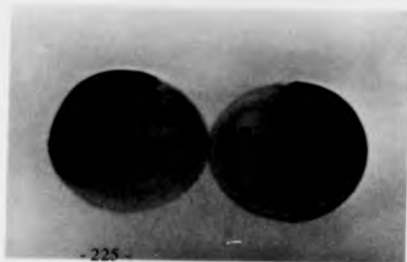
Stage 10-10.5



Stage 16



Stage 16 (magnified)



Stage 10-10.5



Stage 16



Stage 16 (magnified)



Figure IX.6a

Comparison between  $\beta$ -galactosidase expression patterns obtained from circular and linearised micro-injected DNAs

XbaPP/βgal - circular DNA

Stage 10-10.5



Stage 16



Stage 16



Stage 10-10.5



Stage 16



Stage 16 (magnified)

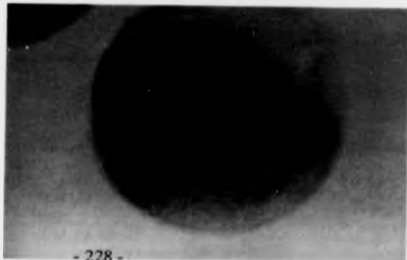


Table IX.1

Summary of  $\beta$ -galactosidase expression patterns obtained from circular and linearised micro-injected DNAs

Construct	Form	Stage analysed	Number scored	Level of expression A B C D	Specificity
pCaSpeR-AUG-Bgal	Linear	10-10.5	5	0 0 0 5	Negative
pCaSpeR-AUG-Bgal	Linear	16	4	0 0 0 4	Negative
pCaSpeR-AUG-Bgal	Circular	10-10.5	4	0 0 0 4	Negative
pCaSpeR-AUG-Bgal	Circular	16	4	0 0 0 4	Negative
$\alpha$ PB/ $\beta$ gal	Linear	10-10.5	5	0 0 0 5	Negative
$\alpha$ PB/ $\beta$ gal	Linear	16	4	0 0 0 4	Negative
$\alpha$ PB/ $\beta$ gal	Circular	10-10.5	4	0 0 0 4	Negative
$\alpha$ PB/ $\beta$ gal	Circular	16	4	0 0 0 4	Negative
BPP/ $\beta$ gal	Linear	10-10.5	4	0 4 0 0	3/4 expression on dorsal side
BPP/ $\beta$ gal	Linear	16	4	0 4 0 0	3/4 very good dorsal expression
BPP/ $\beta$ gal	Circular	10-10.5	5	5 0 0 0	Expression in 1 half of embryo: non-sp.
BPP/ $\beta$ gal	Circular	16	4	4 0 0 0	Lots of expression in NS and epidermis
XbaPP/ $\beta$ gal	Linear	10-10.5	4	0 0 4 0	3/4 dorsal expression
XbaPP/ $\beta$ gal	Linear	16	5	0 4 0 1	Mostly in dorsal tissues
XbaPP/ $\beta$ gal	Circular	10-10.5	4	0 4 0 0	All dorsal expression
XbaPP/ $\beta$ gal	Circular	16	5	0 5 0 0	All dorsal expression

All injections into one cell of a two-cell *X. laevis* embryo. First cleavage division is down dorsal-ventral axis, separating left and right sides of embryo. non-sp.=non-specific expression; NS=nervous system.

Level of expression Key: A=Strong stain; B=weak stain; C=small number of isolated stained cells; D=no stain

- galactosidase expression in *Xenopus*; and  
(2) that there was negligible internally-initiated expression  
from the pCaSpeR-AUG- $\beta$ gal vector.

Since the reverse-orientation promoter fragment clone ( $\alpha$ PB/ $\beta$ gal) was negative, the expression patterns seen with the BPP/ $\beta$ gal and XbaPP/ $\beta$ gal can be said to be specific to those fragments of promoter DNA used.

A qualitative analysis of the results obtained with the BPP/ $\beta$ gal (and others, not shown) showed that circular DNAs produced stronger expression signals but in less defined patterns; linear DNAs produced more restricted and tissue-specific expression patterns.

In the light of this experiment, most of the subsequent micro-injection experiments were performed with linearised DNAs (if a suitable restriction site allowed the linearization of the DNA).

#### 9.3(ii) Quantity of micro-injected DNA.

A large amount of micro-injected DNA is known to be toxic to the developing embryo. An excess of DNA could also possibly result in non-specific expression patterns being obtained. A dilution series was therefore performed using three clones to establish the non-toxic range of DNA concentrations and whether specificity was enhanced by using lower DNA concentrations.

20nl of 40 $\mu$ g/ml DNA was chosen as the standard upper concentration. This value had been used successfully in many previous micro-injection experiments in this laboratory and had been found to

produce consistently reproducible results. DNA amounts higher than 20nl 40µg/ml are known to be toxic. (Other laboratories have found that a concentration 10-fold less than this amount yields acceptable results with some DNAs; Paul Krieg, globin gene, pers. comm.). Dilutions of this 40µg/ml solution were made in TE, keeping a constant micro-injection volume of 20nl. The results are shown in Figures IX.7 to IX.9 and tabulated in Table IX.2.

It can be seen that by diluting the DNA concentration below 40µg/ml, expression is quickly lost with no effect on β-galactosidase specificity. Negligible expression is seen with 4µg/ml. All subsequent experiments therefore were performed with 20nl 40µg/ml DNA.

#### 9.3(iii) Localization of micro-injected DNAs

Expression patterns of exogenous DNAs in *Xenopus* are usually mosaic (i.e. not all cells express the plasmid) as invariably not all cells will receive the plasmid due to the uneven partitioning of the plasmid between cells during cell division. The micro-injections themselves were also only into one cell of a two-cell embryo. Experiments such as these are therefore always open to the criticism that the expression patterns seen only reflect the spatial distribution of the micro-injected plasmid. In many previous micro-injection experiments this suggestion has been refuted by the demonstration that the plasmid is present in the majority of tissues. Southern blots have now been performed with the clones used in this study. These have shown that the plasmids are represented approximately equally in all tissues tested (S.A. Burbidge,

Figure IX.7a

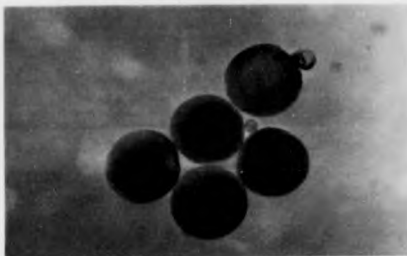
Effect of dilution of DNA concentration on expression  
of promoter/ $\beta$ -galactosidase constructs

BbPP/ $\beta$ gal - stage 11

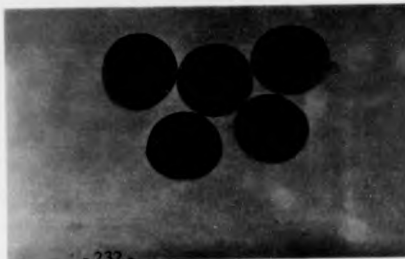
40 $\mu$ g/ml



8 $\mu$ g/ml



4 $\mu$ g/ml



40 $\mu$ g/ml



8 $\mu$ g/ml



4 $\mu$ g/ml



Figure IX.8a

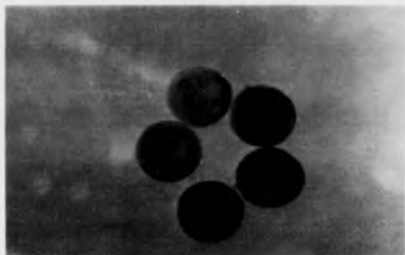
Effect of dilution of DNA concentration on expression  
of promoter/ $\beta$ -galactosidase constructs

XbaPP/ $\beta$ gal - stage 11

40 $\mu$ g/ml



8 $\mu$ g/ml



4 $\mu$ g/ml

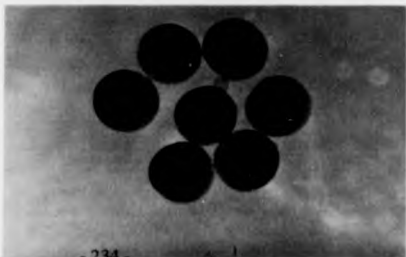
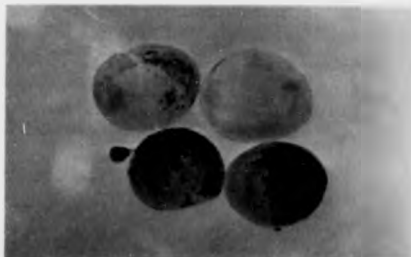


Figure IX.8b

XbaPP/βgal - stage 20

40μg/ml



8μg/ml



4μg/ml

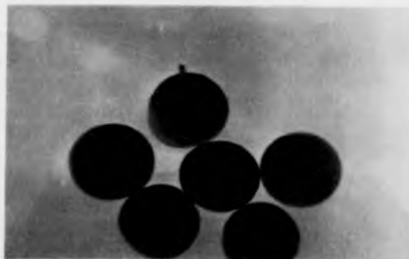


Figure IX 9a

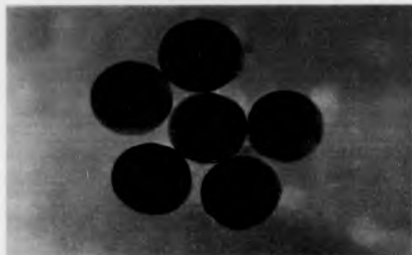
Effect of dilution of DNA concentration on expression  
of promoter/ $\beta$ -galactosidase constructs

XhoPP/ $\beta$ gal - stage 11

40 $\mu$ g/ml



8 $\mu$ g/ml



4 $\mu$ g/ml



Figure IX.9b

XhoPP/βgal - stage 20

8μg/ml



4μg/ml



Table IX.2

Summary of effect of variation in micro-injected DNA concentration  
on  $\beta$ -galactosidase expression patterns in *X. laevis* embryos

Construct	Dilution ( $\mu$ g/ml)	Stage tested	Number of embryos	Number positive	% positive
BbPP/ $\beta$ gal	40	11	5	5	100
BbPP/ $\beta$ gal	40	20	1	1	100
BbPP/ $\beta$ gal	8	11	5	1	20
BbPP/ $\beta$ gal	8	20	5	1	20
BbPP/ $\beta$ gal	4	11	5	0	0
BbPP/ $\beta$ gal	4	20	5	1	20
XbaPP/ $\beta$ gal	40	11	5	5	100
XbaPP/ $\beta$ gal	40	20	9	9	100
XbaPP/ $\beta$ gal	8	11	5	3	60
XbaPP/ $\beta$ gal	8	20	5	0	0
XbaPP/ $\beta$ gal	4	11	7	0	0
XbaPP/ $\beta$ gal	4	20	5	0	0
XhoPP/ $\beta$ gal	40	11	6	6	100
XhoPP/ $\beta$ gal	40	20	1	1	100
XhoPP/ $\beta$ gal	8	11	6	2	33
XhoPP/ $\beta$ gal	8	20	3	1 <sup>a</sup>	33
XhoPP/ $\beta$ gal	4	11	6	2 <sup>b</sup>	33
XhoPP/ $\beta$ gal	4	20	5	0	0

<sup>a</sup> injected DNA "blebbed" out

<sup>b</sup> very small number of staining cells

Although the numbers in each individual experiment are small, an overall trend can clearly be seen. Dilution of the DNA concentration quickly titres out the activity of the micro-injected DNA. There was also no indication that specificity changed as the DNA concentration was altered.

pers. comm.).

#### 9.4 $\beta$ -galactosidase expression patterns in whole embryos.

##### 9.4(i) Introduction

A wide range of promoter/ $\beta$ gal constructs were micro-injected into *X.laevis* embryos. These constructs are shown in Figure IX.2. Embryos were fixed at various stages of development and stained for  $\beta$ -galactosidase expression (as described in Chapter V). Little expression was seen in most embryos prior to MBT (stage 8.5). This could be due either to control elements within the promoter sequences suppressing transcription before this time or to the fact that RNA polymerases become most active after this time (Figure 1.5).

The stained embryos are shown in Figures IX.10 - IX.18.

These photographs are representative of the expression patterns obtained from a large number of experiments; it is obviously impracticable to include vast numbers of each construct at each stage in a study such as this. Table IX.3, therefore, tabulates the data obtained from a greater number of embryos. The data is, by necessity, qualitative but valuable results may be obtained by the comparison of clones that differ in defined amounts of promoter DNA.

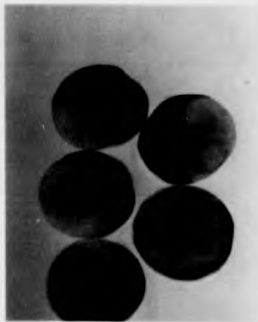
Although the expression patterns seen, even with the longer lengths of promoter DNA present, are not always entirely restricted to one or a small number of tissues, the tissue-specificities achieved are significantly better than those achieved by other groups (Paul Krieg, pers. comm.). Contamination from using the same micro-injection

Figure IX.10

Expression patterns of promoter/ $\beta$ -galactosidase constructs in *Xenopus laevis* embryos

Control: Uninjected embryos

Stage 10



Stage 17



Stage 23



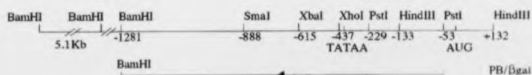
Stage 38



**Figure IX.11**

Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos

Negative control:  $\alpha$ PB/ $\beta$ gal



Stage 10



Stage 15

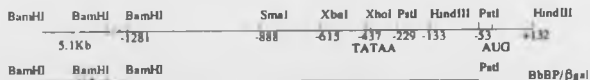


Stage 35

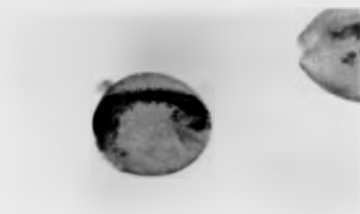


**Figure IX.12a**

**Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos  
BbPP/ $\beta$ gal (circular)**



**Stage 10**



**Stage 15 (magnified)**



**Stage 16**

Figure IX.12b

Stage 32

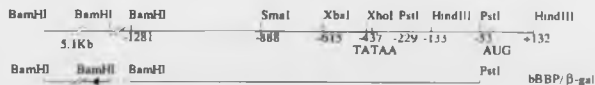


Stage 32 (magnified)



**Figure IX.13**

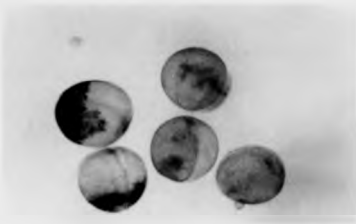
**Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos  
bBPP/ $\beta$ gal (circular)**



**Stage 10**



**Stage 15**

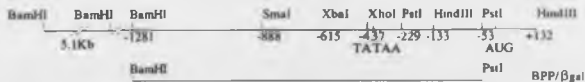


**Stage 35**



**Figure IX.14**

**Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos  
BPP/ $\beta$ gal (linearised)**



Stage 10



Stage 15

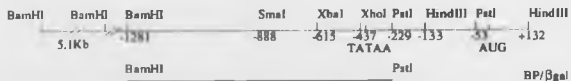


Stage 35

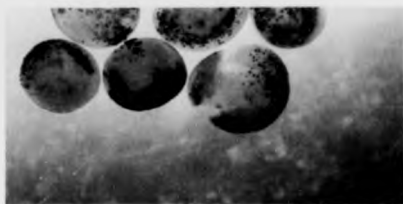


Figure IX.15a

Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos  
BP/ $\beta$ gal (linearised)



Stage 10



Stage 20



Stage 26

Figure IX.15b

Stage 33



Stage 37

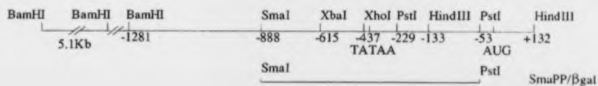


Stage 40

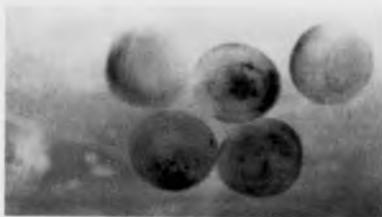


**Figure IX.16**

**Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos  
SmaPP/ $\beta$ gal (linearised)**



**Stage 10**



**Stage 20**

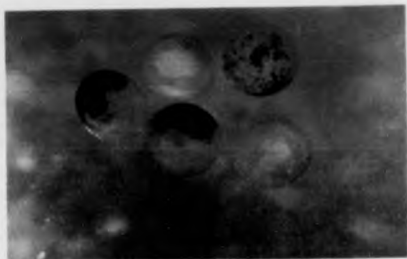
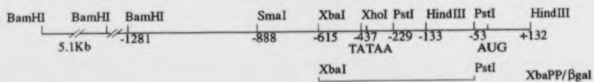


**Stage 26**

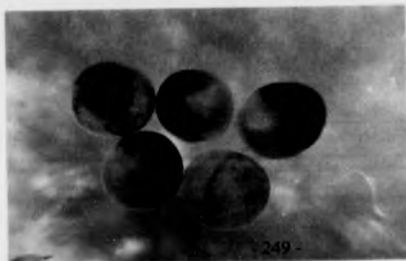


Figure IX.17a

Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos  
XbaPP/ $\beta$ gal (linearised)



Stage 11



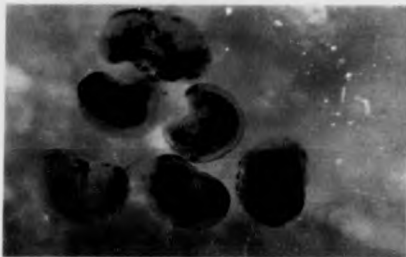
Stage 18

Figure IX.17b

Stage 21



Stage 23

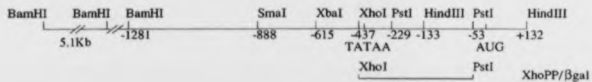


Stage 30



Figure IX.18a

Expression patterns of promoter/ $\beta$ -galactosidase constructs  
in *Xenopus laevis* embryos  
XhoPP/ $\beta$ gal (linearised)



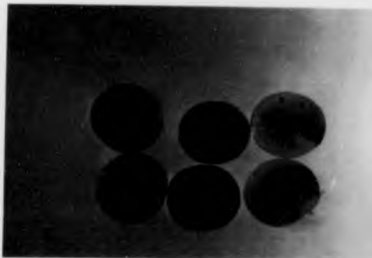
Stage 10.5



Stage 19

Figure IX.18b

Stage 21



Stage 23



Stage 30



Table IX. 3a

Summary of  $\beta$ -galactosidase expression patterns in micro-injected *X. laevis* embryos with  $\beta$ gal constructs

Construct	Stage analysed	Number scored	% of embryos with stain present in:			
			Epidermis	Gut	Somite	NS + notochord
BbPP/ $\beta$ gal	10.5	10	70	0	-	-
BbPP/ $\beta$ gal	12	9	77	0	0	33
BbPP/ $\beta$ gal	17-20	8	63	0	0	88
BbPP/ $\beta$ gal	25	8	75	12	50	87
BbPP/ $\beta$ gal	32	7	0	14	71	43
bBPP/ $\beta$ gal	10.5	10	80	0	-	-
bBPP/ $\beta$ gal	12	9	89	0	0	22
bBPP/ $\beta$ gal	17-20	9	44	0	0	55
bBPP/ $\beta$ gal	25	4	50	0	0	100
bBPP/ $\beta$ gal	35	4	0	50	0	100
BPP/ $\beta$ gal	10.5	10	80	10	-	-
BPP/ $\beta$ gal	12	10	40	0	0	0
BPP/ $\beta$ gal	17-20	9	88	0	0	44
BPP/ $\beta$ gal	23-25	8	75	0	0	75
BPP/ $\beta$ gal	32	4	50	0	25	75

Table IX. 3b

Summary of  $\beta$ -galactosidase expression patterns in micro-injected *X. laevis* embryos with  $\beta$ gal constructs

Construct	Stage analysed	Number scored	Epidermis	% of embryos with stain present in: Gut	Somite	NS + notochord
BP/ $\beta$ gal	10.5	5	0	0	-	-
BP/ $\beta$ gal	20	6	100	0	0	100
BP/ $\beta$ gal	27	7	43	14	100	100
BP/ $\beta$ gal	33	5	80	20	80	40
BP/ $\beta$ gal	37	4	25	0	75	0
XbaPP/ $\beta$ gal	10.5	5	60	0	-	-
XbaPP/ $\beta$ gal	17-20	4	100	0	0	0
XbaPP/ $\beta$ gal	23	5	100	0	0	0
XbaPP/ $\beta$ gal	30	5	100	0	40	40
XhoPP/ $\beta$ gal	10.5	5	100	40	-	-
XhoPP/ $\beta$ gal	17-20	5	20	40	20	0
XhoPP/ $\beta$ gal	21-23	9	100	88	66	44
XhoPP/ $\beta$ gal	30	3	100	100	100	0

needle in all experiments may account for a small amount of background. It must also be remembered that the clones used were derived from *X.borealis* genomic DNA and the host embryos were *X.laevis*.

#### 9.4(ii) Analysis of expression patterns

As stated above, both uninjected and  $\alpha$ PB/ $\beta$ gal clones (Figures IX.10 and IX.11) produced negligible backgrounds of  $\beta$ -galactosidase expression.

Clone BbPP/ $\beta$ gal (Figure IX.12, having 5.1+1.3kb of promoter sequence) produced the most spatially and temporally restricted patterns of  $\beta$ -galactosidase expression. At stage 10, relatively weak expression may be seen in a few clumps of cells. At stage 15, discrete staining is seen along the neural folds and presumptive neural tissues. Staining persists into the spinal cord, notochord and somites in later stage embryos. This pattern of expression is in full agreement with that seen by *in situ* analysis (although  $\beta$ -galactosidase staining in the somites does appear to be earlier and stronger than that reported; SIMMONEAU *et al.*, 1992) and is consistent with the RNase protection assays (DETRICK *et al.*, 1990). At early stages, bBPP/ $\beta$ gal produces less-specific expression than BbPP/ $\beta$ gal. At later stages, staining is clearly more specific to the brain and spinal cord areas, although expression is more 'spotted' than BbPP/ $\beta$ gal.

It must be noted that clones BbPP/ $\beta$ gal and bBPP/ $\beta$ gal were two clones that were micro-injected in circular form due to the lack of

suitable restriction-enzyme sites for linearisation: (the 5' BamHI-BamHI fragment contains an EcoRI site). In Section 9.3(i), it was shown that micro-injected circular DNAs produced stronger, but less specific  $\beta$ -galactosidase expression patterns than the corresponding linearised DNAs. Therefore if it was possible to linearise the above-mentioned clones, weaker, but even more specific expression patterns would be expected.

Clone BPP/ $\beta$ gal produces very similar staining patterns to bBPP/ $\beta$ gal which may indicate that this latter clone contains the 5.1Kb BamHI-BamHI fragment in the reverse (antisense) orientation and that this 5.1Kb fragment contains orientation-specific promoter elements.

A comparison between BPP/ $\beta$ gal and BP/ $\beta$ gal (Figures IX.14 and IX.15, respectively) shows the effect of deleting 176bp of 5'-untranslated RNA. This deletion results in less staining in the nervous tissues and somites, and a distinctly more 'speckled' expression pattern overall which extends to gut and ventral tissues. This appears to indicate that DNA sequence elements 3' to the (proposed) transcriptional start site play a role in the regulation of protein expression.

The expression patterns obtained from clone SmaPP/ $\beta$ gal also have a 'speckled' appearance. This pattern is still nervous system and somite based although there is distinct staining in the gut and intestine regions.

Clone XbaPP/ $\beta$ gal contains only 120bp upstream of the TATA box.

Expression is more variable in these embryos than seen with the above clones: some embryos are negative for nervous system tissues and somites; others have distinct staining in these tissues. Prior to stage 30, however, the majority of embryos have large amounts of non-specific staining in addition to that mentioned above. Greater staining is also seen at stage 11.

Clone XhoPP/ $\beta$ gal contains 10bp upstream of the TATA box. Although there is limited expression prior to stage 10.5, expression subsequent to this is strong but non-specific; there is possibly a bias against nervous system and somite tissues. Whilst a lack of specificity would be expected (if the TATA box does indeed define the transcriptional start site), the level of expression is surprising. This clone, in effect, contains only the TATA box, the region immediately downstream and approximately 360bp of 5'-untranslated sequence. This result provides further evidence for the presence of regulatory sequences downstream of the TATA box.

These results clearly show varying degrees of temporal and, more particularly, spatial patterns of expression. Table IX.3 catalogues the changes in specificity from BbPP/ $\beta$ gal having expression in nervous system/notochord/epidermis (with later expression in somites), through to the complete lack of specificity seen in XhoPP/ $\beta$ gal.

The primary conclusion from these above results is that a reduction of specificity correlates well with a reduction in the length of the promoter DNA present.

## 9.5 Explant and inductive sandwich experiments

### 9.5(i) Explants experiments

Explant experiments were performed in which micro-injected embryos were dissected with mounted eyebrow-hairs at stage 9 (blastula) into animal (A), vegetal (V) and equatorial (E) regions and then cultured in isolation until the equivalent of stage 30. These experiments were designed to show whether animal, vegetal or equatorial regions have the capacity, in isolation, to differentiate into tissues that express the promoter/ $\beta$ gal constructs. In such dissections, animal pole region explants will form ciliated epidermis; equatorial tissue explants will form mesoderm, nervous system and some epidermis; and the vegetal pole region explants will form endoderm. The experimental procedure is outlined in Figure IX.19.

### 9.5(ii) Induction experiments

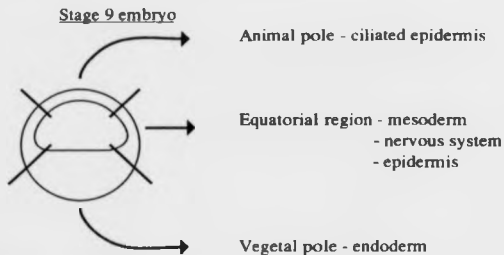
These experiments are designed to show whether induced ectoderm expresses the promoter-deletion/ $\beta$ gal constructs. A comparison of the animal pole explants (above) with these inductive 'sandwiches' illustrates the effect that the neuralising influence has on the developing embryo. The procedure is shown in Figure IX.19.

The dorsal lip of a blastopore (the "Organiser") from an (uninjected) stage 10 embryo is sandwiched between two epidermal caps from (injected) stage 9 embryos. Within 30 minutes the inductive "sandwich" heals to form a "hamburger" shape; and within 60 minutes rounds-up to form a ball.

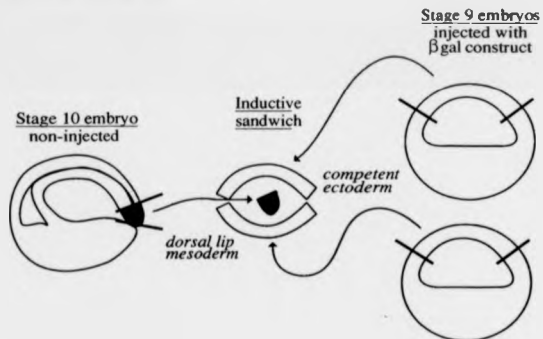
Figure IX.19

Dissection of *Xenopus* embryos for explants  
and inductive sandwiches

Explant dissections



Inductive sandwich dissections



If the ectodermal host cells are induced to form neural tissues (by the neuralising influence of the dorsal lip of the blastopore), the embryo will form a smooth round shape extended in one direction by the developing notochord. If no induction occurs, the embryos will undertake a wrinkled, crunulated appearance.

#### 9.5(iii) Results of explant and inductive sandwich experiments

The results of these experiments are shown in Figures IX.20 to IX.23; these experiments and others are tabulated in Table IX.4.

From Figures IX.20 (uninjected controls) and IX.21 (PP/ $\beta$ gal, 5'-untranslated leader sequence), it can be seen that there is negligible background of  $\beta$ -galactosidase expression in the control explants.

Clone BbPP/ $\beta$ gal (Figure IX.22, having 5.1+1.3kb of promoter sequence) shows high levels of  $\beta$ -galactosidase expression in both equatorial and inductive sandwich explants (i.e. mesoderm/nervous system/some epidermis). This correlates well with the expression patterns seen in whole embryos. Little expression is seen in either animal or vegetal explants.

Clone XbaPP/ $\beta$ gal (Figure IX.23) contains only 120bp DNA upstream of the TATA box. Strong expression is still remarkably specific to mesoderm, nervous system and some epidermal tissues as shown by expression mainly in the equatorial and inductive sandwich explants.

Table IX.4 documents further controls, in particular, results obtained with a metallothionine/thymidine kinase promoter/lacZ construct

Figure IX.20

Expression patterns of promoter $\beta$ -galactosidase constructs in *Xenopus laevis* embryo explants  
Control: Uninjected embryos

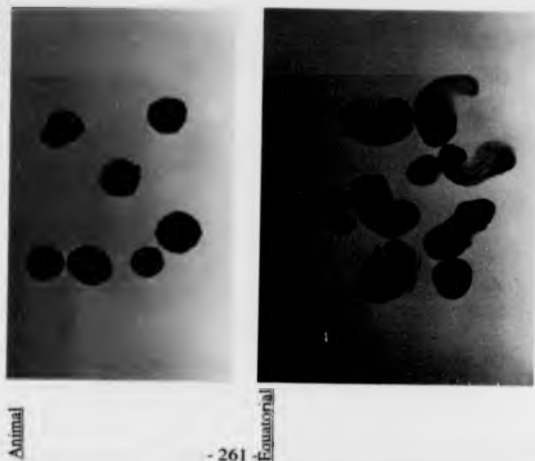


Figure IX.21

Expression patterns of promoter $\beta$ -galactosidase constructs in *Xenopus laevis* embryo explants  
PP $\beta$ gal (linearised)

Animal



Equatorial



Vegetal



Figure IX.22

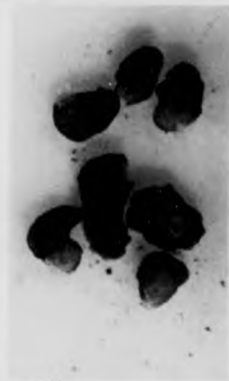
Expression patterns of promoter/ $\beta$ -galactosidase constructs in *Xenopus laevis* embryo explants

BbFP/ $\beta$ gal (circular)

Animal



Dorsal  
mesoderm  
+ ectoderm



Vegetal



Equatorial



Figure IX.23

Expression patterns of promoter $\beta$ -galactosidase constructs in *Xenopus laevis* embryo explants  
XbaPP $\beta$ gal (linearised)

Animal



Dorsal  
mesoderm  
+ ectoderm



Equatorial



Vegetal



Table IX.4a

Summary of  $\beta$ -galactosidase expression patterns seen in explants of micro-injected *Xenopus* embryos

Construct	Explant/ graft	Number analysed	Level of expression <sup>a</sup>			Morphology of stain <sup>b</sup>
			A	B	C	
PP/ $\beta$ gal	animal	8	0	0	0	100
	vegetal	9	0	0	0	100
	equatorial	13	0	0	0	100
	dors. mes. + ect.	ND	-	-	-	-
$\alpha$ PB/ $\beta$ gal	animal	8	0	0	0	100
	vegetal	6	0	0	0	100
	equatorial	10	0	0	0	100
	dors. mes. + ect.	6	0	0	0	100
BbPP/ $\beta$ gal	animal	27	4	0	11	85
	vegetal	11	0	0	36	64
	equatorial	34	76	0	6	18
	dors. mes. + ect.	7	72	0	0	28
BP/ $\beta$ gal	animal	16	31	0	0	69
	vegetal	6	0	0	50	50
	equatorial	14	0	86	0	14
	dors. mes. + ect.	3	0	0	0	100

Table IX.4b

XbaPP/βgal	animal	15	0	0	20	80	ectodermal
	vegetal	5	0	0	100	0	endodermal
	equatorial	15	67	0	33	0	axial
	dors. mes.+ ect.	4	100	0	0	0	induced
MTKlacZ	animal	15	47	0	13	40	ectodermal
	vegetal	7	57	0	0	43	endodermal
	equatorial	17	94	0	0	6	axial
	dors. mes. + ect.	5	80	0	0	20	induced

Key:Level of expression<sup>a</sup> (percentage of number analysed)

A= strong stain

B= weak stain

C= small number of isolated stained cells

D= no stain

ND = not done

Morphology of stain<sup>b</sup>

"ectodermal" = showing staining characteristic of epidermal cells on outside of explant

"vegetal" = staining in large yolk cells characteristic of endoderm

"axial" = staining mainly in axial structures e.g. notocord and somites

"induced" = sandwiches adopt the extended morphology of induced tissues i.e. containing mesoderm and neural tissues (confirmatory data from histology not shown).

("Minimal TK" (Herpes Simplex Virus) promoter; PCR product corresponding to a 225bp region upstream of the translation start-site of the TK gene; Clive Mason, pers. comm.). These results show that the injected DNAs are not spatially excluded from any of the embryonic tissues.

#### 9.6 Transcriptional start-site used by promoter-deletion/ $\beta$ gal constructs.

##### 9.6(i) Introduction

It is recognised that reporter-gene experiments such as these can be criticised on the grounds that the transcriptional start site(s) used in these artificial constructs may not be the same as that (those) used in the endogenous transcripts; the results seen may therefore not be a true representation of endogenous embryos.

In this regard, the following points must be noted:

- a) Both circular and linear forms of the pCaSpeR-AUG- $\beta$ gal plasmids gave negligible expression patterns when micro-injected into *Xenopus* embryos. This would indicate that there was very little undirected initiation of transcription from within this plasmid.
- b) The majority of promoter-deletion/ $\beta$ gal constructs were linearised immediately 5' to the *X.borealis* N-Cadherin genomic DNA. Once micro-injected into the embryos, however, the extent of concatamerisation (resulting in head to tail multimers) of these constructs is unclear. The effect that such contiguous vector/genomic DNA combinations may have on altering the transcriptional start-sites

is uncertain.

c) It may be the case that DNA sequences more-5' to those provided in these promoter/ $\beta$ gal constructs are required to correctly direct initiation of transcription.

#### 9.6(ii) Experimental details

A number of attempts were therefore made to try to establish the transcriptional start sites being used in these promoter-deletion/ $\beta$ gal constructs. These suffered, inevitably, from the same difficulties experienced in trying to establish the endogenous transcriptional start-site (Chapter VIII).

An oligonucleotide that spanned the Adh AUG was synthesised (5'-CAAAGTAAACGACATGGTGAC-3', -6nts to +15nts, BENYAJATI *et al.*, 1981). The precise choice of oligo was influenced by factors discussed in Section 8.9. Primer extension experiments were performed on RNA extracted from micro-injected embryos but no conclusive results were obtained (data not given).

#### 9.7 Conclusions

The expression patterns seen from this series of *X.borealis* N-Cadherin promoter-deletion/ $\beta$ gal constructs have graphically illustrated the utility of the *Xenopus* system in the analysis of a vertebrate promoter.

Although the majority of results given above are of a qualitative nature, the following conclusions can be drawn for the clones used:

a) circular micro-injected DNAs provide stronger but less-specific

expression patterns with the  $\beta$ gal constructs used than the corresponding linearised constructs;

b) 20nl of 40 $\mu$ g/ml construct was the optimum amount of DNA to be micro-injected;

c) the pCaSpeR-AUG- $\beta$ gal and  $\alpha$ PB/ $\beta$ gal (reverse-orientation N-Cadherin promoter) gives negligible backgrounds of  $\beta$ -galactosidase expression;

c) the upstream 5.1Kb BamHI-BamHI *X.borealis* N-Cadherin genomic DNA fragment provides control elements that act in an orientation-specific manner and that mimic the expression patterns of N-Cadherin seen previously by antibody staining and RNase protection assays;

d) the region of DNA between (the proposed) transcriptional and translational start-sites contains sequences important for the regulation of the N-Cadherin gene;

e) clone XbaPP/ $\beta$ gal (containing approx. 200bp upstream of the TATA-box) directs the majority of transcription to nervous system and somites in some early tadpole stage embryos (stage 30);

f) clone XhoPP/ $\beta$ gal, having only 10bp 5' to the TATA-box (but including 380bp 3' to it) produces strong but non-specific  $\beta$ -galactosidase expression throughout the micro-injected embryos; and that

g) explant experiments demonstrate that the micro-injected DNA is not preferentially localised to any specific tissue and that reduction of the amount of DNA contained within the  $\beta$ gal constructs produces

gradually less nervous system/mesoderm-specific expression.

The results obtained are consistent with the established model of the eukaryotic promoter having a number of defined DNA control elements (not necessarily all 5' to the transcriptional start-site) that combine to provide correct regulation of the downstream gene.

## Chapter X

### Discussion and conclusions

#### 10.1 The Cadherin family of CAMs

#### 10.2 *X.borealis* N-Cadherin promoter region

#### 10.3 N-Cadherin and other CAMs

#### 10.4 Promoter-deletion/ $\beta$ -galactosidase constructs

#### 10.5 Future areas of research

##### (i) Introduction

##### (ii) Definition of transcriptional start-sites

##### (iii) Genomic organisation of *X.borealis* N-Cadherin gene

##### (iv) $\beta$ -galactosidase assays of dissected tissues

##### (v) Identification of transcription-factor binding sites

#### 10.6 Concluding remarks

### 10.1 The Cadherin family of CAMs

Interest in the Cadherin family of proteins has increased markedly in the last few years. This is due in part to the increase in number of members of this family (Figure IV.4) but more particularly to the recognition of the importance of the study of these molecules to a wide range of molecular biological fields. Studies of the Cadherins may have profound bearing on our understanding of early development and tissue morphogenesis, metastatic transformation and on the role of the cytoskeleton in the binding of transmembrane proteins; each of these has been discussed in Chapter III.

A further area of interest, on which the study of the Cadherins may provide illumination, is the way that tissue-specificity of gene expression is achieved. Little appears to be known about the the divergence of promoters of members of the same gene family. Although all Cadherin family members presumably stem from the same ancestral gene and selective pressure has maintained high levels of homology in some regions of the protein, the promoters of the different Cadherin genes can be expected to be markedly divergent, reflecting the varied tissue-specificities and expression patterns of each.

### 10.2 *X.borealis* N-Cadherin promoter region

The *X.borealis* N-Cadherin promoter appears to fall neatly into the classically-accepted mould of eukaryotic promoters, having a conserved TATA-box, CCAAT-box and Sp1 binding site. A number

of other potential transcription-factor binding sites, including a homeodomain and *MyoD* binding site, have also been identified on the basis of sequence similarity to known motifs. The promoter region is GC-rich and has an above-average CpG content. The first exon-intron junction has been identified and comparisons with other Cadherins show that this boundary is well conserved in all published Cadherins (except P-Cadherin which lacks the first intron).

The TATA-box is placed 425bp 5' to the translational start-site and, contextually, is well placed to direct the initiation of transcription (Figure VII.5). This is supported by experimental data provided by RNase protection assays and primer extension experiments (Chapter VIII). The unambiguous definition, by experimental means, of the transcriptional start-site has been hindered (it is believed) by the GC-richness of region of DNA between TATA-box and translational start-site.

A recent paper by Simmoneau *et al.* (1992) has added weight to the assertion that the herein identified TATA box defines the start of transcription. Two *X.laevis* N-Cadherin cDNA clones were isolated that appear to correspond to those previously isolated by Detrick *et al.* (1990) and Ginsberg *et al.* (1991). Both of these Simmoneau *et al.* clones extend approximately 370nts 5' from the translational start-site. A comparison of one of these clones and the genomic DNA sequence from 3-9/4.8BS/pBS is given in Figure X.1. The genomic sequence shows significant homology to these cDNAs; the 5'-end of the cDNAs

**Figure X.1**

Comparison between *X. borealis* N-Cadherin  
genomic clone (3-9/4.8BS/pBS) and  
*X. laevis* N-Cadherin cDNA (SIMONNEAU *et al.*, 1992)

Upper sequence: *X. borealis* genomic clone (3-9/4.8BS/pBS)

Lower sequence: *X. laevis* N-Cadherin cDNA

(SIMONNEAU *et al.*, 1992)

	- 420	- 410	- 400
	<u>TATAAA</u> CCAG	GCATGTCCAA	AGTGTGGTTT
- 390	- 380	- 370	- 360
GCGGGGCCGC	ATGCCTGCGT	GTGACAGTAT	CTAAACTGAC
			****
			**
		G	CTAATA GAG
- 353	- 345	- 337	- 327
TG GGGTG	TTCC GCAT	CCTG T GGC	TCTACTTCCT
*****	***	***	*****
ACTGAGGGCT	CTCCATTCAA	TCTCCTCGGC	TTTACCTGCA
- 319	- 311	- 302	- 292
G TCTACTG	GACC AAAG	AATGTACT T	GCGCAAGTAG
*****	***	*****	*****
GCATCTGCCG	CAGCGACTCT	CAC TCCTGT	G G ATT
- 282	- 273	- 263	- 253
GTGTGCACGG	GATTGTGAG	GGCATGAACA	AGCACAGCAC
***	***	***	***
GTGTCCTCTG	CTACTC GCG	GTGTTTACC	GTTACCCCCC
- 243	- 233	- 223	- 213
AGCAGTGGGA	AGGCCCTGAT	ACTGCAGCCC	TAGTCCGACC
***	***	***	**
GCCTTG A A	T AT A		AG
- 203	- 199	- 190	- 185
TTGTTGATAC	CGA T	CG GTGCCGG	TG NAG
***	**	** *****	** **
TTGG GAGAG	GAAAGGCTG	CGAGTGCCAA	CCCTGATAGC
- 175	- 166	- 158	- 151
CATTGAGGGA	CTCCTCC CG	TTAGCTGG GG	GCAAA
*****	*****	** *****	*****
CATTGAGGGA	CTTGTCTGCG	GGGTGACTGC	GAAC TGCAAA

Upper sequence: *X borealis* genomic clone (3-9/4.8BS/pBS)  
 Lower sequence: *X laevis* N-Cadherin cDNA  
 (SIMONNEAU *et al.*, 1992)

-144	-139	-129	-119
CAGG TC T	ATCAC	CTGAAGCTTT	GTGTTTGGTG
*** ** *	*****	* ** *	*****
GAGGATCCCT	GGGAAATCAG	CCAGAGGCTC	GTGTTTGCTT
-109	-99	-89	-79
AAGTCCCATT	GCCCATGCCC	CGCTCCCTGG	GCAGTGGGCA
*** ** *	*****	* ** *	*****
CAGTGC CGCT	GCCTATGCCC	CTCTCCCGG	ACACCC TGCC
-71	-62	-60	-57
CC C TGCTG	CTGCCTCT G	TT	ATA
* ** *	* ** *	**	***
GCTCGTGCCG	CCGC TAGAG	CTTGTGTAAC	CGGGACAATA
-47	-37	-27	-18
TCTGCAGCTC	ACCATTGTTC	ATCTACAGTC	CAAGCC CCC
* ** *	*****	*****	* ** *
TCAGCGGCTC	CTCATTGTTC	ATCT CAAGC	CATCGCTCCC
-9	+1		
C GACACTCA	CAGCAGCACA	TG	
* ** *	*****	**	
CCGATCCT	CACA	TG	

Simonneau *et al.* (1992) have isolated two *X.laevis* N-Cadherin cDNA clones from a cDNA library. These clones are said to show high homology to the Detrick *et al.* (1990) sequence and the Ginsberg *et al.* (1991) sequence respectively, but extend further in a 5'-direction. The two isolated clones are highly homologous to one another and both have the same 5'-end as shown above.

This Figure compares the herein isolated genomic DNA with one of the Simonneau *et al.* cDNA sequences ("Clone 8"). The extent of homology continues up to 52bp downstream from the TATAAA box.

The numbering is relative to the translational start-site of the *X.borealis* genomic sequence.

This provides additional support for the proposition that the *X.borealis* TATA box at -424 is directly involved in transcriptional initiation.

lies approximately 50bp 3' to the TATA box in the genomic DNA. This evidence adds yet further weight to the proposed utility of the TATA box.

### 10.3 N-Cadherin and other CAMs

Whereas mouse E-Cadherin and *X.laevis* N-Cadherin share approximately 60% homology in the cytoplasmic region of the protein, a comparison between the promoter regions of the mouse E-Cadherin and *X.borealis* N-Cadherin genes shows little sequence conservation. (A recent paper by Shimamura and Takeichi (1992) shows that E-Cadherin is actually expressed in mouse brain, although only transiently during early brain development). This divergence extends to the fact that one promoter possesses a TATA box, the other does not. It is interesting to note that the NCAM promoter also lacks a functional TATA-box. All three promoters, however, lie in GC-rich regions.

Potential transcription-factor binding sites have been identified in each of these promoters, on the basis of known sequence motifs, DNase I-footprinting, gel-retardation or a combination of these. Despite the fact that a number of promoters active in neural tissues have been isolated, the DNA sequence motifs to which neural transcription-factors bind have remained largely elusive. This may be a testament to the large number of such transcription-factors involved in neural gene regulation.

#### 10.4 Promoter-deletion/ $\beta$ -galactosidase constructs

It is so often the case in the field of molecular biology that inferences must be made about sub-cellular processes on the basis of a number obtained from a machine or a "blob" on an autoradiograph. Whilst any doubts about the utility of such data are clearly unfounded, such results cannot be a **substitute** for the direct observation of these sub-cellular processes at the level of a single cell or group of cells. Qualitative information and quantitative data must not be viewed, however, as opposite sides of the same coin: they are complementary to one another and the possession of both may lead to synergistic results.

In this study, the construction of a series of *X.borealis* N-Cadherin promoter deletions have been described, together with the cloning of these deletions into the pUC-based *Drosophila* pCaSpeR-AUG- $\beta$ gal plasmid (THUMMEL *et al.*, 1988). Control experiments have shown that this plasmid produces a negligible background of  $\beta$ -galactosidase expression in *X.borealis* embryos. A number of further controls were performed in order to establish the most appropriate form and concentration of DNA to be micro-injected.

The expression pattern of the BbPP/ $\beta$ gal clone (containing 5.1+1.3kb of DNA 5' to the translational start-site) has been shown to be largely restricted to the developing nervous system and mesodermal tissues and corresponds well with the previously published expression patterns of the N-Cadherin protein and RNA (SIMONNEAU *et al.*,

1992; DETRICK *et al.*, 1990).

Reduction of the amount of 5'-promoter DNA attached to the pCaSpeR-AUG- $\beta$ gal plasmid results in less spatially restricted expression. Clone XhoPP/ $\beta$ gal (comprising 10bp 5' and 120bp 3' to the TATA box) surprisingly shows strong but non-specific expression throughout the embryo.

From a comparison of the expression patterns between these constructs, it can be inferred that the distal 5.1kb BamHI-BamHI fragment contains important regulatory elements that act in an orientation-specific manner; and that the region of DNA immediately 3' to the TATA-box also plays a regulatory role.

Explants of dissected *X.laevis* embryos that have been micro-injected with the  $\beta$ gal constructs have confirmed that 1kb of N-Cadherin promoter DNA (5' to the TATA-box) is sufficient to direct the majority of  $\beta$ -galactosidase expression to neural and mesodermal tissues; and that expression of these constructs is also seen in ectoderm that has experienced the inductive influence of late gastrula dorsal mesoderm.

#### 10.5 Future areas of research

##### 10.5(i) Introduction

There will always be a limit to the amount that one researcher can be achieve in a three-year period. In an open ended project such as this, each new avenue of research readily branches into many smaller ones, but it is not possible to explore all paths or even to do full justice to

more than a few. Thus there always exists much scope for further study.

It is gratifying to see that Glaxo Group Research have recognised the value of this research and are providing sponsorship for its continuation. With their financial backing (and the use of their facilities) it is hoped that progress will be rapid.

Detailed below, therefore, are some of the immediate aims to be recommended to those wishing to further this work (many of which have already been initiated).

#### 10.5(ii) Identification of transcriptional start-site

The greatest frustration of this study has been the failure to unambiguously identify the transcriptional start site(s). In common with other groups (Sorkin *et al.*, 1991; Nedivi *et al.*, 1992), primer-extension and RNase protection assays have failed to provide conclusive results, probably due to the formation of secondary structures in the RNA in the high GC-content regions immediately 5' to the translational start-site. These problems were overcome in Nedivi *et al.* (*ibid.*) by the use of PCR to amplify the small number of full-length primer extension cDNA products. A similar procedure has been initiated here (Appendix C) but not yet completed.

Oligonucleotides nearer to the TATA-box than those described here have also now been made and are being used in an attempt to define, with greater resolution, the site or sites of transcriptional initiation.

Should similar problems be experienced as herein described, it is

recommended that the PCR protocol be followed in order to amplify low frequency full-length mRNA transcripts (see Appendix B and NEDIVI *et al.* (*ibid.*)). The advantage of such a PCR approach is that the products may be sequenced directly using the PCR oligos.

Further scope also exists for resolving the anomalous results obtained with the RPA assay (in detecting RNA transcripts in *X.borealis* egg).

#### 10.5(iii) Genomic organisation of the *X.borealis* N-Cadherin gene.

Only the first exon of the *X.borealis* N-Cadherin gene has currently been identified. Within the isolated genomic clone 3-9/EMBL3, 12.5Kb of DNA is as yet unmapped with regard to protein-coding regions and the intron/exon organisation.

#### 10.5(iv) $\beta$ -galactosidase assays of dissected tissues

An analysis of the  $\beta$ -galactosidase expression in injected *X.laevis* embryos through the use of a colorimetric assay of dissected tissues is already well advanced. Although less aesthetically appealing than the stained-embryo photographs, this data will contribute a valuable quantitative aspect to the expression studies.

#### 10.5(v) Identification of transcription-factor binding sites

Gel-retardation and DNA-footprinting studies can be used to identify promoter DNA regions bound by transcription factors. A number of potential sites have been identified by DNA sequence inspection and it will be interesting to see how many of these will be supported by experimental data.

### 10.5 Concluding remarks

The ultimate goal of this research has always been the identification of neural transcription factors and to gain an insight into the mechanisms of neural gene regulation. With the identification of this N-Cadherin promoter, it is hoped that this goal may be one step nearer.

## Appendix A

### Stage series of *Xenopus laevis* embryos

The *Xenopus laevis* and *Xenopus borealis* embryo stages referred to herein are defined according to the widely accepted classification of Nieuwkoop and Faber (1967). The diagrams reproduced here, showing embryo stages 1-45, are taken from this work.

#### Note

The embryos illustrated below are *X.laevis*; *X.borealis* embryos are smaller.

All the photographs shown in Chapter IX are of *X.laevis* embryos.



St. 1 — Animal view (An.)  
Magnification 24x



St. 1 — Dorsal view (Dors.)



St. 1 — Ventral view (Ventr.)



St. 2 — Ventr.



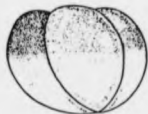
St. 2 — An.



St. 2 — An.



St. 3 — An.



St. 3 — Dorsolateral view (Dors.-lat.)



St. 4 — An.



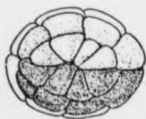
St. 5 — Dors.



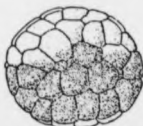
St. 5 — An.



St. 4 — Dors.-lat.



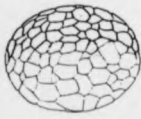
St. 6 — An.



St. 6½ — An.



St. 6½ — Dors.



St. 7 — Dors.



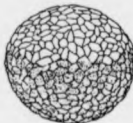
St. 7 — An.



St. 6½ — Ventr.



St. 7 — Ventr.



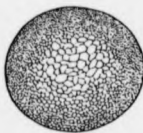
St. 8 — An.



St. 8 — Dors.



St. 8 — Ventr.



St. 9 — Vegetative view  
Dorsal side above (Veg.)



St. 10 — Veg.



Two forms  
St. 10 1/2 — Veg.



St. 11 — Veg.



St. 12 — Veg.



St. 11 1/2 — Veg.



St. 12 1/2 — Post.-dors.



St. 13 — Posterio-dorsal view (Post.-dors.)



St. 14 — Post.-dors.



St. 15 — Post.-dors.



St. 14 — Lateral view (l. side) (Lat.)



St. 15 — Anterior view (Ant.)



St. 16 — Ant.



St. 16 — Post.-dors.



St. 17 — Dors.



St. 17 — Ant.



St. 20 — Dors.



St. 19 — Ant.



St. 18 — Ant.



St. 19 — Dors.



St. 20 — Ant.



St. 21 — Dors.



St. 22 — Dors.



St. 21 — Ant.



St. 22 — Lat.



St. 23 — Lat.



St. 23 — Dors.



St. 24 — Dors.



St. 24 — Lat.  
(Small individual)



St. 25 — Dors.



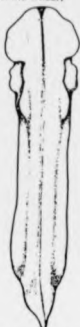
St. 25 — Lat.



St. 26 — Dors.



St. 27 — Lat.



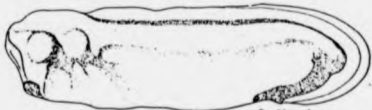
St. 27 — Dors.



St. 26 — Lat.



St. 26 — Dors.



St. 28 — Lat.





Idem Magn. 12x



St. 29.30 — Lat.



St. 31 — Lat.



St. 32 — Lat.



St. 33.34 — Lat.



St. 35.36 — Lat.



St. 42 — Lat.



St. 41 — Lat.



St. 41 — Ventr.



St. 40 — Lat.  
(Small ind.)



St. 39 — Lat.  
(Small ind.)



St. 37.38 — Lat.



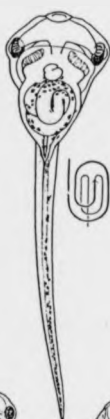
St. 43 - Lat.



St. 43 - Ventr.



St. 44  
Ventr.



St. 45  
Ventr.



St. 45 - Lat.



St. 42 Ventr.



St. 46 - Lat.



St. 46 - Ventr.



Idem Magn. 8x



## Appendix B

### PCR-based protocol to determine transcriptional start-site

#### B.1 Introduction

#### B.2 Separation of oligonucleotides from primer extension products

##### (i) Test procedures

##### (ii) Experimental conditions used

#### B.3 Poly-dG tailing of extended products with terminal transferase

##### (i) Use of terminal deoxynucleotidyl transferase (TdT)

##### (ii) Experimental conditions

#### B.4 PCR of TdT-tailed products

#### B.5 Conclusions

## B.1 Introduction

An attempt was initiated to combine the primer extension procedure with PCR in order to amplify the small number of full-length extension products that (it was hoped) would be present after a primer extension reaction. The rationale for this procedure is illustrated in Figure B.1.

This procedure involves the use of terminal deoxynucleotidyl transferase (TdT) to add an oligo-dG tail onto the 3'-end of the cDNA resulting from a primer extension reaction and then PCR-amplifying the product using the primer extension oligo and an oligo-dC primer. If successful, this technique would overcome the deficiencies of both primer extension and RNase protection assay in that it would provide not only a figure for the distance between the known oligo and transcriptional start, but also, by sequencing the PCR-product, the actual DNA sequence. A simple comparison could then be made with the genomic DNA sequence which would highlight the presence of any introns and illuminate the transcriptional start-site to within a few nucleotides.

## B.2 Separation of oligonucleotides from primer extension products

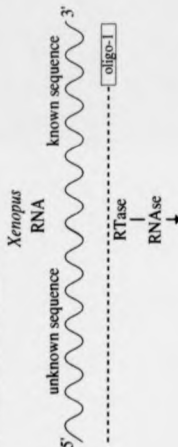
### B.2(i) Test procedures

The largest single obstacle to overcome in using this technique is the removal of unextended oligo after the primer extension step. After the primer-extension step, oligo will be present in excess over the extended product and, if not removed, will compete with the extended

**Figure B.1**

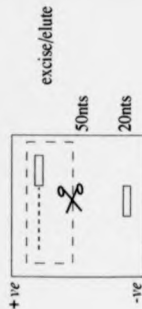
**Protocol for the PCR-based identification of transcriptional start-site**

**1. Primer extension reaction**

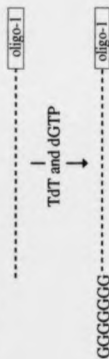


+ [oligo-1]

**2. Separation of extension product from excess oligo on an acrylamide gel**



**3. TdT-tailing with oligo-dG**



**4. PCR using oligo-2 and oligo-dC**



**5. Sequence using oligo-3**



product in the tailing reaction, resulting in few tailed extension products. The method of Belyavsky *et al.* (1989) uses an agarose gel to resolve oligo from extended product. The oligo/extended product mix is run out on the agarose gel, the low molecular weight end (including oligo) is run off, the polarity of the electrodes are then reversed and the DNA products are then condensed into a small gel slice near the original wells.

For the situation herein described, where the expected extension product was less than 0.5kb, an acrylamide gel was considered more appropriate to resolve oligo from extension products. The gel conditions and elution procedure were established using <sup>35</sup>S-labelled DNA sequence reactions. DNA sequence was run out on various (4-12) %-acrylamide gels and exposed overnight. It was found that a 5% acrylamide gel gave the best resolution of oligo (20mers) and ssDNA products larger than 50nts, while still retaining the DNA extension products in a relatively small area. Orange G (10nts), bromophenol blue (35nts) and xylene cyanol (130nts) markers were used to clearly define the tracks and as molecular weight markers (numbers refer to the 5% acrylamide gel used).

This acrylamide-gel based procedure has the advantage over the procedure of Belyavsky *et al.* (*ibid.*) in that no reverse-running of the gel is required; the DNA products are also eluted from a much thinner gel slice (1mm as opposed to 4.5mm).

After elution and EtOH-precipitation, the excised test DNAs were

run on a further acrylamide gel to confirm that all DNA molecular weights were eluted with equal efficiency and to verify that the DNA was not degraded by this procedure. The recovery of DNA from such acrylamide slices was also determined, by scintillation counter, to be in the region of 60-70%.

#### B.2(ii) Experimental conditions used

The primer extension and RNA hydrolysis procedures are described in the Methods (Chapter V). The reaction products were run on a pre-warmed 5% acrylamide, 1xTBE, 7M urea gel until the bromophenol blue marker was at the bottom of the gel. The area above the xylene cyanol marker (130nts) was excised and extended products eluted from this gel slice into 200 $\mu$ l 0.3M sodium acetate pH6.5, 2mM EDTA in a 0.5ml siliconised Eppendorf tube for 1-3 hours on a rotary shaker at 37°C. After this time the buffer was replaced with a further 200 $\mu$ l and the elution repeated. 2 vols EtOH was then added to the combined eluted fractions and the DNA precipitated overnight with glycogen.

#### B.3 Poly-dG tailing of extended products with terminal transferase.

##### B.3(i) Use of Terminal deoxynucleotidyl transferase (TdT)

Terminal deoxynucleotidyl transferase (TdT, Gibco BRL, 10-20U/ $\mu$ l) will add supplied dNTPs to the 3'-end of a ssDNA template or overhanging 3'-ends of dsDNAs (NELSON and BRUTLAG, 1979; ROYCHOUDHURY and WU, 1980). Therefore, the addition of TdT + one dNTP to a ssDNA primer-extension product will result in the 3'-

end of the primer-extension product acquiring an oligo-dNTP tail; the length of the tail is dependent on the buffer conditions and time of reaction.

It was decided to tail with poly-dG or poly-dC to avoid any problems with the endogenous poly-A tails on the ends of the mRNAs. An poly-dC primer was available in the laboratory and so this dictated that the cDNA would be tailed with poly-dG.

The manganese-based TdT buffer of Deng and Wu (1983) was compared with the cobalt-based buffer of Belyavsky *et al.* (1989). Test tailing reactions were carried out using  $\alpha$ - $^{32}$ P-dGTP and then subsequently determining the %-counts incorporated using the DEAE-sodium phosphate method. Using this labelled dGTP reaction, 18% labelling occurred with the manganese buffer whilst only 1% occurred with the cobalt buffer.

The activity of the TdT was verified by tailing a  $\gamma$ - $^{32}$ P-dGTP labelled oligo and resolving the products on a 5%-acrylamide gel.

#### B.3(ii) Experimental conditions

A 10 $\mu$ l reaction mix was performed using the buffer of Deng and Wu (1983) i.e. 100mM sodium cacodylate pH7.1, 2mM MnCl<sub>2</sub>, 0.1mM DTT. Incubation was for 60 minutes at 37°C. The reaction mix was then phenol/chloroform-extracted, chloroform-extracted and EtOH-precipitated.

#### B.4 PCR of TdT-tailed products

The *X.borealis* N-Cadherin genomic DNA oligos used in the PCR

reactions are described in Chapter VIII (Figure VIII.7). The poly-dC oligo was termed TRT5 and had the sequence CTGCAGATT(C<sub>17</sub>). The poly-dC stretch is preceded in this oligonucleotide by PstI and EcoRI sites to facilitate cloning of the PCR products.

PCR was performed using standard protocols (Taq polymerase 1, Promega) according to the following regime: 94°C 1.5 minutes, 60°C 1 minute, 72°C 1.5 minutes; 25 cycles.

The PCR protocol was established using 3-9/4.8BS/pBS restriction-enzyme digested with BamHI and poly-dG tailed as above. PCR was performed with TRT5 and NCADJ1 oligonucleotides and resulted in a DNA band of the correct size.

### B.5 Conclusions

Whilst the individual steps for this PCR-based amplification of the 5'-end of the cDNA have been established, the constraints of time prevented the completion of the full procedure. Theoretically this approach should yield the desired result.

The variable amount of full-length cDNA produced after the primer extension step (if any) produces difficulties in controlling the tailing reaction. The tailing reaction would need to be allowed to proceed for a number of different times (followed by PCR of each) in order to ensure that the optimum length poly-dG tail for the subsequent PCR step was obtained.

Other PCR-based procedures for determining the 5'-end sequences of cDNAs are available including the head-to-head ligation method of

Hofmann and Brian (1991) and the rapid amplification of cDNA ends (RACE) procedure of Frohman and Martin (1989).

## Bibliography

- ACHESON, A., SUNSHINE, J.L., RUTISHAUSER, U. (1991)  
NCAM polysialic acid can regulate both cell-cell and cell-substratum interactions.  
*J. Cell Biol.* 114 (1), 143-153
- AISSANI, B., BERNARDI, G. (1991)  
CpG islands: Features and distribution in the genomes of vertebrates.  
*Gene* 106, 173-183
- AKERS, R.M., PHILLIPS, C.R., WESSELS, N.K. (1986)  
Expression of an epidermal antigen used to study tissue induction in the early *Xenopus laevis* embryo.  
*Science* 231, 613-616
- ALBERS, B. (1987)  
Competence as the main factor determining the size of the neural plate.  
*Develop., Growth and Diff.* 29, 535-545
- ALLSOPP, T.E., MOSS, D.J. (1989)  
A developmentally regulated chicken neuronal protein associated with the cortical cytoskeleton.  
*J. Neuroscience* 9 (1), 13-24
- AMAGAI, M., KLAUS-KOVTUN, V., STANLEY, J.R. (1991)  
Autoantibodies against a novel epithelial cadherin in *Pemphigus vulgaris*, a disease of cell adhesion.  
*Cell* 67, 869-877
- AMERSHAM (1984)  
M13 cloning and sequencing handbook.  
Amersham International Plc, Amersham UK.
- ANGRES, B., MULLER, A.H.J., KELLERMANN, J., HAUSEN, P. (1991)  
Differential expression of two cadherins in *Xenopus laevis*.  
*Development* 111, 829-844
- BAGCHI, S., RAYCHAUDHURI, P., NEVINS, J.R. (1990)  
Adenovirus E1A proteins can dissociate heteromeric complexes involving E2F transcription-factor: a neural mechanism for E1A trans-activation.  
*Cell* 62, 659-669

BAGENAL, E. (1990)

A study of the histone genes in *Xenopus borealis*.

PhD thesis, University of Warwick, Coventry, England

BALAK, K., JACOBSON, M., SUNSHINE, J. (1987)

Neural cell adhesion molecule expression in *Xenopus* embryos.

Devl. Biol. 119, 540-550

BARBAS, J.A., CHAIX, J.-C., STEINMETZ, M., GORIDIS, C. (1988)

Differential splicing and alternative polyadenylation generates distinct NCAM transcripts and proteins in the mouse.

EMBOJ 7 (3), 625-632

BARD, J.B.L. (1988)

A traction-based mechanism for somitogenesis in the chick.

Roux's Arch. Dev. Biol. 197 (8), 513-517

BARTON, C.H., DICKSON, J.G., GOWER, H.J., ROWETT, L.H., PUTT, W.,

ELSON, V.L., MOORE, S.E., GORIDIS, C., WALSH, F.S. (1988)

Complete sequence and *in vitro* expression of a tissue-specific phosphatidyl inositol-linked NCAM isoform.

Development 104, 164-173

BARTON, C.H., MANN, D.A., WALSH, F.S. (1990)

Characterization of the human N-CAM promoter.

Biochemical Journal 268 (1), 161-168

BATLEY, I. (1992)

The analysis of the promoter regions of neural-specific genes in vertebrates and invertebrates.

3rd Year Project, Department of Biological Sciences, University of Warwick, Coventry, England

BEATO, M. (1989)

Gene regulation by steroid hormones.

Cell 56, 335-344

BEHRENS, J., LOWRICK, O., KLEIN-HITPASS, L., BIRCHMEIER, W. (1991)

The E-Cadherin promoter: Functional analysis of a GC-rich region and an epithelial cell-specific palindromic regulatory element.

Proc. Natl. Acad. Soc. USA 88, 11495-11499

BEHRENS, J., VAKAET, L., FRIIS, R., WINTERHAEGER, E., VANROY, F., MAREEL, M.M., BIRCHMEIER, W. (1993)

Loss of epithelial differentiation and gain of invasiveness correlates with tyrosine phosphorylation of the E-Cadherin-beta-catenin complex in cells transformed with a temperature-sensitive *v-src* gene.

J. Cell. Biol. 120, 757-766

BEHRENS, J., BIRCHMEIER, W., GOODMAN, S.L., IMHOF, B.A. (1985)

Cell Biol. 101, 1307-1315

BELYAVSKY, A., VINOGRADOVA, T., RAJEWSKY, K. (1989)

PCR-based cDNA library construction: general cDNA libraries at the level of a few cells.

Nucleic Acid Res. 17 (8) 2919-2932

BENYAJATI, C., PLACE, A.R., POWERS, D.A., SOFER, W. (1981)

Alcohol dehydrogenase gene of *Drosophila melanogaster*: relationship of intervening sequences to functional domains of protein.

Proc. Natl. Acad. Sci. USA 78 (5), 2717-2721

BIRD, A.P. (1986)

CpG-rich islands and the function of DNA methylation.

Nature 321, 209-213

BIRNSTIEL, M.L., BUSSLINGER, M., STRUB, K. (1985)

Cell 41, 349-359

BIXBY, J.L., PRATT, R.S., LILIEN, J., REICHARDT, L.F. (1987)

Neurite outgrowth on muscle cell surfaces involves extracellular matrix receptors as well as  $Ca^{2+}$ -dependent and -independent cell adhesion molecules.

Proc. Natl. Acad. Sci. USA 84, 2555-2559

BIXBY, J.L., LILIEN, J., REICHARDT, L.F. (1988)

Identification of the major proteins that promote neuronal process outgrowth on Schwann cells.

J. Cell Biol. 107 (1), 353-361

BLASCHUK, O.W., SULLIVAN, R., DAVID, S., POULIOT, Y. (1990)

Identification of a Cadherin cell adhesion recognition sequence.

Development 139, 227-229

BOYCE, F.M., BEGGS, A.H., FEENER, C., KUNKEL, L.M. (1991)  
Dystrophin is transcribed in brain from a distant upstream promoter.  
Proc. Natl. Acad. Sci. USA 88, 1276-1280

BRADSHAW, R.A., PRENTIS, S. (Ed.) (1987)  
Oncogenes and growth factors.  
Elsevier Science Publishers.

BRIVANLOU, A.H., HARLAND, R.M. (1989)  
Expression of an *engrailed*-related protein is induced in the anterior neural  
ectoderm of early *Xenopus* embryos.  
Development 106, 611-617

CHEN, A., REYES, A., AKESON, R. (1990)  
Transcription initiation sites and structural organisation of the extreme 5'  
region of the rat neural cell adhesion molecule gene.  
Mol. Cell. Biol. 10 (7), 3314-3324

CHO, K.W.Y., BLUMBERG, B., STEINBEISSER, H., DE ROBERTIS, E.  
(1991)  
Cell 67, 1111-1120

COLE, G.J., GLASER, L. (1986)  
A heparin-binding domain from N-CAM is involved in neural cell-  
substratum adhesion.  
J. Cell Biol. 102, 403-412

CONN, P.M. (1989)  
Methods in neurosciences - Volume 1.  
Gene Probes, Academic Press

COVAULT, J., SANES, J.R. (1986)  
Distribution of NCAM in synaptic and extrasynaptic portions of developing  
and adult skeletal muscle.  
J. Cell Biol. 102, 716-730

CRITTENDEN, S.L., RUTISHAUSER, U., LILIEN, J. (1988)  
Identification of two structural types of calcium-dependent adhesion  
molecules in the chicken embryo.  
Proc. Natl. Acad. Sci. USA 85 (10), 3464-3468

CROSSIN, K.L., CHUONG, C-M., EDELMAN, G.M. (1985)  
Expression sequences of cell adhesion molecules.  
Proc. Natl. Acad. Sci. USA 82, 6942-6946

CUNNINGHAM, B.A., HEMPERLY, J.J., MURRAY, B.A., PREDIGER, E.A., BRACKENBURY, R., EDELMAN, G.M. (1987)  
Neural cell adhesion molecule: Structure, immunoglobulin-like-domains, cell surface modulation and alternative RNA splicing.  
Science 236, 799-806

DALE, L., HOWES, G., PRICE, B.M.J., SMITH, J.C. (1992)  
Bone morphogenetic protein 4: a ventralising factor in early *Xenopus* development.  
Development 115, 573-585

DALE, L., SLACK, J.M.W. (1987)  
Fate map of the 32 cell stage of *Xenopus laevis*.  
Development 99, 527-551

DAMSKY, C.H., RICHA, J., SOLTER, D., KNUDSEN, K., BUCK, C.A. (1983)  
Identification and purification of a cell surface glycoprotein mediating intercellular adhesion in embryonic and adult tissues.  
Cell 34, 455-466

DENG, G., WU, R. (1983)  
Terminal transferase: Use in the tailing of DNA and for *in vitro* mutagenesis.  
Methods in Enzymology 100, 96-116

DETRICK, R.J., DICKEY, D., KINTNER, C.R. (1990)  
The effects of N-Cadherin misexpression on morphogenesis in *Xenopus* embryos.  
Neuron 4, 493-506

D'EUSTACHIO, P., OWENS, G.C., EDELMAN, G.M., CUNNINGHAM, B.A. (1985)  
Chromosomal location of the gene encoding the neural cell adhesion molecule (NCAM) in the mouse.  
Proc. Natl. Acad. Sci. USA 82, 7631-7635

DE VERNEUIL, H., METZGER, D. (1990)  
The lack of transcriptional activation of the v-erbA oncogene is in part due to a mutation present in the DNA binding domain of the protein.  
Nucleic Acid Res. 18, 4489-4497

DICKSON, G., GOWER, H.J., BARTON, C.H., PRENTICE, H.M., ELSOM, V.L., MOORE, S.E., COX, R.D., QUINN, C., PUTT, W., WALSH,

F.S. (1987)

Human muscle neural cell adhesion molecule (NCAM): identification of a muscle-specific sequence in the extracellular domain.

Cell 50, 1119-1130

DICKSON, J.G., PRENTICE, H.M., KENIMER, J.G., WALSH, F.S. (1986)

Identification and characterization of neuron-specific and developmentally regulated gene transcripts in the chick embryo spinal cord.

J. Neurochemistry 46, (3), 787-793

DICKSON, G., PRENTICE, H., JULIEN, J.P., FERRARI, G., WALSH, F.S. (1986)

Nerve growth factor activates Thy-1 and neurofilament gene transcription in rat PC12 cells.

EMBOJ 5, (13), 3449-3453

DIRKSON, M.L., JAMRICH, M. (1992)

Genes Dev. 6, 599-608

DIXON, J.E., KINTNER, C.R. (1989)

Development 106, 749-757

DOHERTY, P., FRUNS, M., SEATON, P., DICKSON, G., BARTON, C.H., SEARS, T.A., WALSH, F.S. (1990)

A threshold effect of the major isoforms of NCAM on neurite outgrowth.

Nature (London) 343, 464-466

DOHERTY, P., ROWETT, L.H., MOORE, S.E., MANN, D.A., WALSH, F.S. (1991)

Neurite outgrowth in response to transfected N-CAM and N-Cadherin reveals fundamental differences in neuronal responsiveness to CAMs.

Neuron 6, 247-258

DONALDSON, M., CRAMER, M., RINGWALD, M., STARZINSKI-POWITZ, A. (1991)

Expression of M-Cadherin, a member of the cadherin multigene family, correlates with differentiation of skeletal muscle cells.

Proc. Natl. Acad. Sci. USA 88, 8024-8028

DONIACH, T., PHILLIPS, C.R., GERHART, J.C. (1992)

Science 257, 542-545

DRAZBA, J., LEMMON, V. (1990)

The role of cell adhesion molecules in neurite outgrowth on Muller cells.  
Dev. Biol. 138, 82-93

ETKIN, L.D., PEARMAN, B. (1987)

Distribution, expression and germ-line transmission of exogenous DNA sequences following microinjection into *Xenopus laevis* eggs.  
Development 99, 15-23

FAISST S., MEYER, S. (1992)

Compilation of vertebrate-encoded transcription factors.  
Nucleic Acid Res. 20 (1), 3-26

FEARON, E.R., CHO, K.R., NIGRO, J.M., KERN, S.E., SIMONS, J.W.,  
RUPPERT, J.M., HAMILTON, S.R., PREISINGER, A.C., THOMAS, G.,  
KINZLER, K.W., VOGELSTEIN, B. (1990)

Identification of a chromosome 18q gene that is altered in colorectal cancers.  
Science 247, 49-56

FINK, G.R. (1986)

Translational control of transcription in eukaryotes.  
Cell 45, 155-156

FORSS-PETTER, S., DANIELSON, P., CATSICAS, S., BATTENBERG,  
E., PRICE, J., NERENBERG, M., SUTCLIFFE, J.G. (1990)

Transgenic mice expressing beta-galactosidase in mature neurons under neuron-specific enolase promoter control.  
Neuron, 187-197

FORSS-PETTER, S., DANIELSON, P., SUTCLIFFE, J.G. (1986)

Neuron-specific enolase: Complete structure of rat mRNA, multiple transcriptional start-sites, and evidence suggesting post-transcriptional control.  
J. Neuroscience Res. 16, 141-156

FRASER, S.E., CARHART, M.S., MURRAY, B.A., CHUONG, C-M.,  
EDELMAN, G.M. (1988)

Alterations in the *Xenopus* retinotectal projection by antibodies to *Xenopus* N-CAM.  
Developmental Biology 129, 217-230

FRASER, S.E., MURRAY, B.A., CHUONG, C-M., EDELMAN, G.M.  
(1984)

Alterations of the retinotectal map in *Xenopus* by antibodies to neural cell adhesion molecules.

Proc. Natl. Acad. Sci. USA 81, 4222-4226

FRIXEN, U.W., BEHRENS, J., SACHS, M., EBERLE, G., VOSS, B.,  
WARDA, A., LOCHNER, D., BIRCHMEIER, W. (1991)

E-Cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells.

J. Cell Biol. 113, 173-185

FROHMAN, M.A., MARTIN, G.R. (1989)

Rapid amplification of cDNA ends using nested primers.

Technique 1 (3), 165-170

FU, Y., SATO, K., HOSOKAWA, K., SHIOKAWA, K. (1990)

Expression of circular plasmids which contain bacterial chloramphenicol acetyltransferase gene connected to the promoter of polypeptide IX gene of human Adenovirus type 12 in oocytes, eggs and embryos of *Xenopus laevis*.  
Zoological science 7, 195-200

FU, L., YE, R., BROWDER, L.W., JOHNSTON, R.N. (1991)

Translational potentiation of mRNA with secondary structure in *Xenopus*.  
Science 251, 807-810

FU, Y., HOSOKAWA, K., SHIOKAWA, K. (1989)

Expression of circular and linearized bacterial chloramphenicol acetyl transferase genes with or without viral promoters after injection into fertilized eggs, unfertilized eggs and oocytes of *Xenopus laevis*.

Roux's Arch. Dev. Biol. 198, 148-156

FUJIMORI, T., MIYATANI, S., TAKEICHI, M. (1990)

Ectopic expression of N-Cadherin perturbs histogenesis in *Xenopus* embryos.  
Development 110, 97-104

GAHMBERG, C.G., ANDERSSON, L.C. (1982)

Role of sialic acid in the mobility of membrane proteins containing O-linked oligosaccharides on polyacrylamide gel electrophoresis in SDS.

Eur. J. Biochem. 122, 581-586

GALLAGHER, B.C., HAINSKI, A.M., MOODY, S.A. (1991)

Autonomous differentiation of dorsal axial structures from an animal cap cleavage stage blastomere in *Xenopus*.

Development 112, 1103-1114

- GALLIN, W.J., SORKIN, B.C., EDELMAN, G.M., CUNNINGHAM, B.A. (1987)  
Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule, L-CAM.  
Proc. Natl. Acad. Sci. USA 84, 2808-2812
- GENNARRINI, G., HIRSCH, M.R., HE, H.T., HIRN, M., FINNE, J., GORIDIS, C. (1986)  
Differential expression of mouse N-CAM mRNA species during brain development and in neural cell lines.  
J. Neurosci. 6, 1983-1990
- GERHART, J., DANILCHIK, M., DONIACH, T., ROBERTS, S., ROWNING, B., STEWART, R. (1989)  
Development (Suppl.) 107, 37-52
- GERISCH, G., MALCHOW, D. (1976)  
Adv. Cyclic Nucleotide Res. 7, 49
- GILBERT, S.F., SAXEN, L. (1993)  
Spemann's organizer: models and molecules.  
Mechs. Develop. 41, 73-89
- GINSBERG, D., DESIMONE, D., GEIGER, B. (1991)  
Expression of a novel cadherin (EP-cadherin) in unfertilized eggs and early *Xenopus* embryos.  
Development 111, 315-325
- GORIDIS, C., HIRN, M., SANTONI, M.-J., GENNARRINI, G., DEAGOSTINI-BAZIN, H., JORDAN, B.R., KIEFER, M., STEINMETZ, M. (1985)  
Isolation of mouse NCAM-related cDNA: detection and cloning using monoclonal antibodies.  
EMBOJ 4 (3), 631-635
- GOWER, H.J., BARTON, C.H., ELSOM, V.L., THOMPSON, J., MOORE, S.E., DICKSON, G., WALSH, F.S. (1988)  
Alternative splicing generates a secreted form of N-CAM in muscle and brain.  
Cell 55, 955-964
- GOWER, H.J., MOORE, S.E., DICKSON, G., ELSOM, V.L., NAYAK, R., WALSH, F.S. (1989)

Cloning and characterization of a myoblast cell surface antigen defined by 24.1D5 monoclonal antibody.  
Development 105, 723-731

GRANIOU, M., CHANCONIE, M., ZILLER, C. (1988)  
A monoclonal antibody recognizing a common antigen on neurons and fibroblasts in chicken and quail.  
Cell Differentiation 24, 179-190

GUBBAY, J., COLLIGNON, J., KOOPMAN, P., CAPEL, B.,  
ECONOMOU, A., MUNSTERBERG, A., VIVIAN, N., GOODFELLOW, P.,  
LOVELL-BADGE, R. (1990)  
A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes.  
Nature (London) 346, 245-250

GURDON, J.B. (1987)  
Embryonic induction - molecular prospects.  
Development 99, 285-306

HAMES, B.D., HIGGINS, S.J. (1984)  
Transcription and Translation - a practical approach.  
Edited by Hames and Rickwood, IRL Press Oxford, Washington DC.

HATTA, K., OKADA, T.S., TAKEICHI, M. (1985)  
A monoclonal antibody disrupting calcium-dependent cell-cell adhesion of brain tissues: Possible role of its target antigen in animal pattern formation.  
Proc. Natl. Acad. Sci. USA 82, 2789-2793

HATTA, K., TAKEICHI, M. (1986)  
Expression of N-Cadherin adhesion molecules associated with early morphogenetic events in chick development.  
Nature (London) 320, 447-449

HATTA, K., TAKAGI, S., FUJISAWA, H., TAKEICHI, M. (1987)  
Spatial and temporal expression pattern of N-Cadherin cell adhesion molecules correlated with morphogenetic processes of chicken embryos.  
Dev. Biol. 120, 215-227

HATTA, K., NOSE, A., NAGAFUCHI, A., TAKEICHI, M. (1988)  
Cloning and expression of cDNA encoding a neural calcium dependent cell adhesion molecule: Its identity in the Cadherin gene family.  
J. Cell Biol. 106, 873-881

- HATTA, M., MIYATANI, S., COPELAND, N.G., GILBERT, D.J., JENKINS, N.A., TAKEICHI, M. (1991)  
Genomic organization and chromosomal mapping of the mouse P-Cadherin gene.  
Nucleic Acid Res. 19 (16), 4437-4441
- HAUSEN, P., RIEBESELL, M. (1991)  
The early development of *Xenopus laevis*: an atlas of histology.  
Published Springer-Verlag.
- HE, H.-T., BARBET, J., CHAIX, J.-C., GORIDIS, C. (1986)  
Phosphatidylinositol is involved in the membrane attachment of N-CAM-120, the smallest component of the neural cell adhesion molecule.  
EMBOJ 5 (10), 2489-2494
- HE, X., ROSENFELD, M.G. (1991)  
Mechanisms of complex transcriptional regulation: Implications for brain development.  
Neuron 7, 183-196
- HEMMATI-BRIVANLOU, A., STEWART, R.M., HARLAND, R.M. (1990)  
Science 250, 800-802
- HEMPERLEY, J.J., EDELMAN, G.M., CUNNINGHAM, B.A. (1986)  
cDNA clones of the neural cell adhesion molecule (NCAM) lacking a membrane spanning region consistent with evidence for membrane attachment via a phosphatidyl inositol intermediate.  
Proc. Natl. Acad. Sci. USA 83, 9822-9826
- HERZBERG, F., WILDERMUTH, V., WEDLICH, D. (1991)  
Expression of XBCad, a novel cadherin, during oogenesis and early development of *Xenopus*.  
Mechanisms of Development 35, 33-42
- HIRSCH, M.R., GAUGLER, L., DEAGOSTINI-BAZIN, H., BALLY-CUIF, L., GORIDIS, C. (1990)  
Identification of positive and negative regulatory elements governing cell-type-specific expression of the neural cell adhesion molecule gene.  
Mol. Cell. Biol. 10 (5), 1959-1968
- HIRSCH, M.R., VALARCHE, I., DEAGOSTINI-BAZIN, H., PERNELLE, C., JOLIOT, A., GORIDIS, C. (1991)  
An upstream element of the NCAM promoter contains a binding-site for homeodomains.

HOBSON, G.M., MOLLOY, G.R., BENFIELD, P.A. (1990)

Identification of *cis*-acting regulatory elements in the promoter region of the rat brain creatine kinase gene.

Mol. Cell. Biol. 10 (12), 6533-6543

HOFMANN, M.A., BRIAN, D.A. (1991)

A PCR-enhanced method for determining the 5' end sequence of mRNAs.

PCR Meths. Apps. 1 (1), 43-45

HOFFMAN, S., SORKIN, B.C., WHITE, P.C., BRACKENBURY, R.,

MAILHAMMER, R., RUTISHAUSER, U., CUNNINGHAM, B.A.,

EDELMAN, G.E. (1982)

Chemical characterization of a neural cell adhesion molecule purified from embryonic brain membranes.

J. Biol. Chem 257, 7720-7729

HOLTFRETER, J. (1939)

Gewebeaffinität, ein Mittel der embryonalen Formbildung.

Arch. Exp. Zellforsch. Besonders Gewebezücht. 23, 169-209

INOUE, M., OGAWA, H., MIYATA, M., SHIOZAKI, H., TANIZAWA, O.

(1992)

Expression of E-Cadherin in normal, benign, and malignant-tissues of female genital organs.

Am. J. Clin. Path. 98, 76-80

INUZUKA, H., MIYATANI, S., TAKEICHI, M. (1991)

R-Cadherin: A novel Ca-2+-dependent cell-cell adhesion molecule expressed in the retina.

Neuron 7, 69-79

ISAACS, J.T., TAKEICHI, M., DEBRUYNE, F.M.J., SCHALKEN, J.A.

(1992)

Decreased expression of E-Cadherin in the progression of rat prostatic-cancer.

Cancer Res. 52, 2916-2922

ITO, K., KUBOTA, H.Y. (1989)

Expression of neural antigens in normal *Xenopus* embryos and induced explants.

Development, Growth and Differentiation 31 (6), 563-571

JACOBSON, M. (1984)

Cell lineage analyses of neural induction: Origins of cells forming the induced nervous system.

Dev. Biol. 110, 102-113

JACOBSON, M., RUTISHAUSER, U. (1986)

Induction of neural cell adhesion molecule (NCAM) in *Xenopus* embryos.

Dev. Biol. 116, 524-531

JOHNSON, P.F., MCKNIGHT, S.L. (1989)

Eukaryotic transcriptional regulatory proteins.

Ann. Rev. Biochem. 58, 799-839

JONAS, E.A., SNAPE, A.M., SARGENT, T.D. (1989)

Transcriptional regulation of a *Xenopus* embryonic epidermal keratin gene.

Development 106, 399-405

JONES, E.A., WOODLAND, H.R. (1989)

Spatial aspects of neural induction in *Xenopus laevis*.

Development 107, 785-791

KARIM, F.D., URNESS, L.D., THUMMEL, C.S., KLEMSY, M.J.,

MCKERCHER, S.R., CELADA, A., VANBEVEREN, C., MAKI, R.A.,

GUNTHER, C.V., NYE, J.A., GRAVES, B.J. (1990)

The ETS domain: a new DNA-binding motif that recognises a purine-rich core DNA sequence.

Genes Dev. 4, 1451-1453

KAY, B.K., SCHWARTZ, L.M., UTISHAUSER, U., QUI, T.H., PENG,

H.P. (1988)

Patterns of N-CAM expression during myogenesis in *Xenopus laevis*.

Development 103, 463-471

KESSEL, M., GRUSS, P. (1990)

Murine developmental control genes.

Science 249, 374-379

KEY, B., AKESON, R.A. (1991)

Delineation of olfactory pathways in the frog nervous system by unique glycoconjugates and NCAM glycoforms.

Neuron 6, 381-396

KIERAN, M., BLANK, V., LOGEAT, F., VANDEKERCKHOVE, J.,  
LOTTSPICH, F., LE BAIL, O., URBAN, M.B., KOURILSKY, P.,  
BAEUEERLE, P.A., ISRAEL, A. (1990)

The DNA-binding subunit of NF- $\kappa$ B is identical to factor KBF-1 and  
homologous to the *rel* oncogene product.  
Cell 62, 1007-1018

KIMELMAN, D., CHRISTIAN, J.L., MOON, R.T. (1992)

Synergistic principles of development: overlapping patterning systems in  
*Xenopus* mesoderm induction.  
Development 116, 1-9

KINTNER, C.R., MELTON, D.A. (1987)

Expression of *Xenopus* N-CAM RNA in ectoderm is an early response to  
neural induction.  
Development 99, 311-325

KINTNER, C.R. (1988)

Effects of altered expression of the neural cell adhesion molecule, N-CAM,  
on early neural development in *Xenopus* embryos.  
Neuron 1, 545-555

KLUG, A., RHODES, D. (1987)

"Zinc fingers": a novel protein motif for nucleic acid recognition.  
Trends Biochem. Sci. 12, 464-469

KNUDSEN, K.A., MCELWEE, S.A., MYERS, L. (1990)

A role for the neural cell adhesion molecule, NCAM, in myoblast interaction  
during myogenesis.  
Dev. Biol. 138, 159-168

KORNER, M., RATTNER, A., MAUXION, F., SEN, R., CITRI, Y. (1989)

A brain-specific transcription activator.  
Neuron 3, 563-572

KOZAK, M. (1986)

Point mutations define a sequence flanking the AUG initiator codon that  
modulates translation by eukaryotic ribosomes.  
Cell 44, 283-292

KRIEG, P.A., MELTON, D.A. (1987)

*In vitro* RNA synthesis with SP6 RNA polymerase.  
Meths. Enzymology 155, 397-415

- KRIEG, P.A., MELTON, D.A. (1985)  
Developmental regulation of a gastrula-specific gene injected into fertilised *Xenopus* eggs.  
EMBOJ 4, 3463-3471
- KRIEG, P., VARNUM, S., WORMINGTON, W., MELTON, D.A. (1989)  
The mRNA encoding the elongation factor-1 $\alpha$  is a major transcript at the mid-blastula transition.  
Dev. Biol. 133, 93-100
- LAGUNOWICH, L.A., GRUNWALD, G.B. (1989)  
Expression of calcium-dependent cell adhesion proteins during ocular development: a biochemical, histochemical and functional analysis.  
Dev. Biol. 135 (1), 158-171
- LAI, E., PREZIOSO, V.R., TAO, W., CHEN, W.S., DARNELL, J.E. (1991)  
Hepatocyte nuclear factor 3 $\alpha$  belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork-head*.  
Genes Dev. 5, 416-427
- LAND, H., GREZ, M., HAUSER, H., LINDENMAIER, W., SCHUTZ, G. (1983)  
Synthesis of ds-cDNA involving addition of dCMP tails to allow cloning of 5'-terminal mRNA sequences.  
Methods in Enzymology 100, 285-292
- LANDER, A.D. (1989)  
Understanding the molecules of neural cell contacts: emerging patterns of structure and function.  
Trends Neuroscience 12 (5) 189-195
- LANDSCHULZ, W.H., JOHNSON, P.F., MCKNIGHT, S.L. (1988)  
The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins.  
Science 240, 1759-1764
- LATCHMAN, D.S. (1990)  
Gene regulation: a eukaryotic perspective.  
Published Unwin Hyman Ltd., London.
- LEVI, G., CROSSIN, K.L., EDELMAN, G.M. (1987)  
Expression sequences and distribution of two primary cell adhesion molecules during embryonic development of *Xenopus laevis*.

J. Cell Biol. 105, 2359-2372

LEWIS, W.H. (1904)

Experimental studies on the development of the eye in Amphibia. I.  
On the origin of the lens. *Rana palustris*.  
Amer. J. Anat. 3, 505-536

LINDEMANN, D., BOCK, E. (1989)

Cell adhesion molecules in neural development.  
Dev. Neurosci. (Basel) 11 (3), 149-173

MAHONEY, P.A., WEBER, U., ONOFRECHUK, P., BIESSMANN, H.,  
BRYANT, P.J., GOODMAN, C.S. (1991)

The *fat* tumor suppressor gene in *Drosophila* encodes a novel member of the  
Cadherin gene superfamily.  
Cell 67, 853-868

MAJELLO, B., ARCONI, R., TONIATTI, C., CILIBERTO, G. (1990)

Constitutive and IL-6-induced nuclear factors that interact with the human C-  
reactive protein promoter.  
EMBOJ 9, 457-465

MANGOLD, O. (1933)

Über die Induktionfähigkeit der verschiedenen Bezirke der Neurula von  
Urodelen.  
Naturwissenschaften 21, 761-766

MANIATIS, T., GOODBOURN, S., FISCHER, J.A. (1987)

Regulation of inducible and tissue-specific gene expression.  
Science 236, 1237-1245

MANN, D.A., BARTON, C.H., WALSH, F.S. (1990)

Characterization of a regulatory region within the human neural cell-adhesion  
molecule gene.  
Biochemical Soc. Trans. 18 (3), 410-412

MANSOURI, A., SPURR, N., GOODFELLOW, P.N., KEMLER, R. (1988)

Characterization and chromosomal localization of the gene encoding the  
human cell adhesion molecule uvomorulin.  
Differentiation 38, 67-61

MARTH, J.D., OVERELL, R.W., MEIER, K.E., KREBS, E.G.,  
PERLMUTTER, R.M. (1988)

Translational activation of the *lck* oncogene.  
*Nature (London)* 332, 171-173

MATSUNAGA, M., HATTA, K., NAGAFUCHI, A., TAKEICHI, M. (1988)  
Guidance of optic nerve fibres by N-Cadherin adhesion molecules.  
*Nature (London)* 334, 62-64

MATSUNAGA, M., HATTA, K., TAKEICHI, M. (1988)  
Role of N-Cadherin cell adhesion molecules in the histogenesis of neural retina.  
*Neuron* 1, 289-295

MAUE, R.A., KRANER, S.D., GOODMAN, R.H., MANDEL, G. (1990)  
Neuron-specific expression of the rat brain type II sodium channel gene is directed by upstream regulatory elements.  
*Neuron* 4, 223-231

MECHANIC, S., RAYNOR, K., HILL, J.E., COWIN, P. (1991)  
Desmocollins form a distinct subset of the cadherin family of cell adhesion molecules.  
*Proc. Natl. Acad. Sci. USA* 88, 4476-4480

MILNER, R.J., BLOOM, F.E., LAI, C., LERNER, R.A., SUTCLIFFE, J.G. (1984)  
Brain-specific genes have identifier sequences in their introns.  
*Proc. Natl. Acad. Sci. USA* 81, 713-717

MILNER, R.J., LAI, C., NAVE, K.A., LENOIR, D., OGATA, J., SUTCLIFFE, J.G. (1985)  
Nucleotide sequence of two mRNAs for rat brain myelin proteolipid protein.  
*Cell* 42, 931-939

MITCHELL, P.J., TJIAN, R. (1989)  
Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.  
*Science* 245, 371-378

MIYATANI, S., SHIMAMURA, K., HATTA, M., NAGAFUCHI, A., NOSE, A., MATSUNAGA, M., HATTA, K., TAKEICHI, M. (1989)  
Neural cadherin: Role in selective cell-cell adhesion.  
*Science* 245, 631-635

MIYATANI, S., COPELAND, N.G., GILBERT, D.J., JENKINS, N.A.,  
TAKEICHI, M. (1992)  
Genomic structure and chromosomal mapping of the mouse N-Cadherin  
gene.  
Proc. Natl. Acad. Sci. USA 89, 8443-8447

MOASE, C.E., TRASLER, D.G. (1991)  
NCAM alterations in *spotch* neural tube defect mouse embryos.  
Development 113, 1049-1058

MOORE, S.E., THOMPSON, J., KIRKNESS, V., DICKSON, J.G., WALSH,  
F.S. (1987)  
Skeletal muscle neural cell adhesion molecule (N-CAM): Changes in protein  
and mRNA species during myogenesis of muscle cell lines.  
J. Cell Biol. 105, 1377-1386

MOUNT (1982)  
A catalogue of splice junction sequences.  
Nucleic Acid Res. 10, 459-472

MURRAY, B.A., HEMPERLEY, J.J., GALLIN, W.J., MACGREGOR, J.S.,  
EDELMAN, G.M., CUNNINGHAM, B.A. (1984)  
Isolation of cDNA clones for the chicken neural cell adhesion molecule  
(NCAM).  
Proc. Natl. Acad. Sci. USA 81, 5584-5588

MURRAY, B.A., HEMPERLEY, J.J., PREDIGER, E.A., EDELMAN, G.M.,  
CUNNINGHAM, B.A. (1986)  
Alternatively spliced mRNAs code for different polypeptide chains of the  
chicken neural cell adhesion molecule (NCAM).  
J. Cell Biol. 102, 189-193

MURRE, C., MCCAWE, P.S., BALTIMORE, D. (1989)  
A new DNA-binding and dimerisation motif in immunoglobulin enhancer  
binding, *daughterless*, *MyoD*, and *myc* proteins.  
Cell 56, 777-783

NAGAFUCHI, A., SHIRAYOSHI, Y., OKAZAKI, K., YASUDA, K.,  
TAKEICHI, M. (1987)  
Transformation of cell adhesion properties by exogenously introduced E-  
Cadherin cDNA.  
Nature (London) 329, 341-343

NAGAFUCHI, A., TAKEICHI, M. (1988)

Cell binding function of E-Cadherin is regulated by the cytoplasmic domain.  
EMBOJ 7, 3679-3684

NEDIVI, E., BASI, G.S., AKEY, I.V., SKENE, J.H.P. (1992)

A neural-specific GAP-43 core promoter located between unusual DNA elements that interact to regulate its activity.  
J. Neuroscience 12 (3), 691-704

NELSON, T., BRUTLAG, D. (1979)

Addition of homopolymers to the 3'-ends of duplex DNA with terminal transferase.  
Methods in Enzymology 68, 41-50

NELSON, W.J., SHORE, E.M., WANG, A.Z., HAMMERTON W.W. (1990)

J. Cell Biol. 110, 349

NEUGEBAUER, K.M., TOMASELLI, K.J., LILIEN, J., REICHARDT, L.F. (1988)

N-Cadherin, NCAM, and integrins promote neurite outgrowth on astrocytes in vitro.  
J. Cell Biol. 107 (3), 1177-1187

NGUYEN, C., MATTEI, M.-G., MATTEI, J.-F., SANTONI, M.-J., GORIDIS, C., JORDAN, B.R. (1986)

Localisation of human NCAM gene to band q23 of chromosome 11: the third gene coding for a cell interaction molecule mapped to the distal portion of the long arm of chromosome 11.  
J. Cell Biol. 102, 711-715

NIEUWKOOP, P.D., FABER, J. (1967)

Normal table of *Xenopus laevis* (Daudin)  
2nd Edition, Amsterdam, North Holland Publishing Co.

NOSE, A., TAKEICHI, M., (1986)

A novel cadherin cell adhesion molecule: its expression patterns associated with implantation and organogenesis of mouse embryos.  
J. Cell Biol. 103, 2649-2658

NOSE, A., TSUJI, K., TAKEICHI, M. (1990)

Localisation of specificity determining sites in Cadherin cell adhesion molecules.  
Cell 61, 147-155

NUSSINOV, R., OWENS, J., MAIZEL, J.V. (1986)

Sequence signals in eukaryotic upstream regions.

Biochim. Biophys. Acta 866, 109-119

ODENWALD, W.F., TAYLOR, C.F., PALMER-HILL, F.J., FRIEDRICH,

V., TANI, M., LAZZARINI, R.A. (1987)

Expression of a homeodomain protein in non contact-inhibited cultural cells and post mitotic neurons.

Genes Dev. 1, 482-496

O'KANE, C.J., GEHRING, W.J. (1987)

Detection *in situ* of genomic regulatory sequences in *Drosophila*.

Proc. Natl. Acad. Sci. USA 84, 9123-9127

OTTE, A.P., KRAMER, I.J.M., DURSTON, A.J. (1991)

Science 251, 570-573

OWENS, G.C., EDELMAN, G.M., CUNNINGHAM, B.A. (1987)

Organisation of the neural cell adhesion molecule (NCAM) gene: Alternative exon usage as the basis for different membrane associated domains.

Proc. Natl. Acad. Sci. USA 84, 294-298

PASSMORE, S., MAINE, G.T., EIBLE, R., CHRIST, C., TYE, B.K. (1988)

A *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of *MATa* cells.

J. Mol. Biol. 204, 593-606

PELTOMAKI, P., SISTONEN, P., MECKLIN, J., PYLKKANEN, L.,

JARVINEN, H., SIMONS, J.W., CHO, K.R., VOGELSTEIN, B., DE LA

CHAPELLE, A. (1991)

Evidence supporting exclusion of the DCC gene and a portion of chromosome 18q as the locus for susceptibility to hereditary nonpolyposis colorectal carcinoma in five kindreds.

Cancer research 51, 4135-4140

PERBAL, B. (1988)

A practical guide to molecular cloning.

2nd Ed., Wiley and Sons

PEYRIERAS, N., HYAFIL, F., LOUVARD, D., PLOEGH, H.L., JACOB, F.

(1983)

Uvomorulin: a non-integral membrane protein of early mouse embryos.

Proc. Natl. Acad. Sci. USA 80, 6274-6277

PHILLIPS, C.R. (1991)

Neural induction.

Meths. Cell Biol. 36, 329-436

RANSCHT, B., DOURS-ZIMMERMANN, M.T. (1991)

T-Cadherin, a novel cadherin cell adhesion molecule in the nervous system lacks the conserved cytoplasmic region.

Neuron 7, 391-402

RICHTER, K., GRUNZ, H., DAWID, I.B. (1988)

Gene expression in the embryonic nervous system of *Xenopus laevis*.

Proc. Natl. Acad. Sci. USA 85, 8086-8090

RINGWALD, M., SCHUH, R., VESTWEBER, D., EISTETTER, H.,

LOTTSPEICH, F., ENGEL, J., DOLZ, R., JAHNIG, F., EPPLEN, J.,

MAYER, S., MULLER, C., KEMLER, R. (1987)

The structure of cell adhesion molecule uvomorulin. Insights into the molecular mechanism of Ca-2+-dependent cell adhesion.

EMBOJ 6 (12), 3647-3653

ROSENFELD, M.G. (1991)

POU-domain transcription factors: pou-er-ful developmental regulators.

Genes & Development 5, 897-907

ROYCHOUDHURY, R., WU, R. (1980)

Terminal transferase-catalysed addition of nucleotides to the 3'-termini of DNA.

Methods in Enzymology 65, 43-49

RUIZ I ALTABA, A., JESSELL, T. (1992)

Development 116, 81-93

RUIZ I ALTABA, A., MELTON, D.A. (1989)

Bimodal and graded expression of the *Xenopus* homeobox gene *Xhox3* during embryonic development.

Development 106, 173-183

RUSSEL, M., KIDD, S., KELLY, M.R. (1986)

An improved filamentous helper phage for generating single-stranded plasmid DNA.

Gene 45, 333-338

RUTISHAUSER, U., THIERY J.-P., BRACKENBURY, R., SELA, B.-A.,  
EDELMAN, G.M. (1976)

Mechanisms of adhesion among cells from neural tissues of the chick  
embryo.

Proc. Natl. Acad. Sci. USA 73, 577-581

RYGAARD, K., MOLLER, C., BOCK, E., SPANGTHOMSEN, M. (1992)

Expression of Cadherin and NCAM in small-cell lung-cancer cell-lines and  
xenografts.

Brit. J. Canc. 65, 573-577

SADOUL, R., HIRN, M., DEAGOSTINI-BAZIN, H., ROUGON, G.,  
GORIDIS, C. (1983)

Adult and embryonic mouse neural cell adhesion molecules have different  
binding properties.

Nature (London) 304, 347-349

SAKIMURA, K., KUSHIYA, E., TAKAHASHI, Y., SUZUKI, Y. (1987)

The structure and expression of neuron-specific enolase.

Gene 60, 103-113

SAMBROOK, J., FRITSCH, E.F., MANIATIS, T. (1989)

Molecular cloning: A laboratory manual

2nd Ed., Cold Spring Harbour Laboratory Press

SANES, J.R., RUBENSTEIN, J.L.R., NICOLAS, J.F. (1986)

Use of a recombinant retrovirus to study post-implantation cell lineage in  
mouse embryos.

EMBOJ 5 (12), 3133-3142

SCHOENWOLF, G.C., POWERS, M.L. (1987)

Shaping of the chick neuroepithelium during primary and secondary  
neurulation: Role of cell elongation.

Anat. Rec. 218, 182-195

SCHOENWOLF, G.C., FOLSOM, D., MOE, A. (1988)

A re-examination of the role of microfilaments in neurulation in the chick  
embryo.

Anat. Rec. 220, 87-102

SCHOENWOLF, G.C., SMITH, J.L. (1990)

Mechanisms of neurulation: traditional viewpoint and recent advances.

Development 109, 243-270

SCOTT, M.P., TAMKUN, J.W., HARTZELL, G.W. (1989)

The structure and function of the homeodomain.

Biochim. Biophys. Acta 989, 25-48

SHIMAMURA, K., TAKEICHI, M. (1992)

Local and transient expression of E-Cadherin involved in mouse embryonic brain morphogenesis.

Development 116, 1011-1019

SHIOKAWA, K., YOSHIDA, M., FUKAMACHI, H., FU, Y., TASHIRO,

K., SAMESHIMA, M. (1992)

Cytological studies of large nucleus-like structures formed by exogenously-injected linear and circular DNAs in fertilised eggs of *Xenopus laevis*.

Develop., Growth and Differ. 34 (1), 79-90

SHIOKAWA, K. (1991)

Gene expression from endogenous and exogenously-introduced DNAs in early embryogenesis of *Xenopus laevis*.

Develop., Growth and Differ. 33(1), 1-8

SHIOKAWA, K., YAMANA, K., FU, Y., ATSUCHI, Y., HOSOKAWA, K.

(1990)

Expression of exogenously introduced bacterial chloramphenicol acetyl transferase genes in *Xenopus laevis* embryos before the mid blastula transition.

Roux's Arch. Dev. Biol. 198, 322-329

SHIOKAWA, K., FU, Y., NAKAKURA, N., TASHIRO, K., SAMESHIMA,

M. (1989)

Effects of the injection of exogenous DNAs on gene expression in early embryos and coenocytic egg cells of *Xenopus laevis*.

Roux's Arch. Dev. Biol. 198, 78-84

SHIMAMURA, K., TAKEICHI, M. (1992)

Local and transient expression of E-Cadherin in mouse embryonic brain development.

Development 116, 1011-1019

SHIRAYOSHI, Y., HATTA, K., HOSODA, M., TSUNASAWA, S.,

SAKIYAMA, F., TAKEICHI, M. (1986)

Cadherin cell adhesion molecules with distinct binding specificities share a common structure.

EMBOJ 5 (10), 2485-2488

- SIMONNEAU, L., BRODERS, F., THIERY, J.P. (1992)  
N-Cadherin transcripts in *Xenopus laevis* from early tailbud to tadpole.  
Dev. Dynamics 194, 247-260
- SLACK, J.M.W., TANNAHILL, D. (1992)  
Mechanisms of anteroposterior axis specification in vertebrates. Lessons  
from amphibians.  
Development 114, 285-302
- SMALL, S.J., SHULL, G.E., SANTONI, M.J., AKESON, R. (1987)  
Identification of a cDNA clone that contains the complete coding sequence  
for a 140kDa rat NCAM polypeptide.  
J. Cell Biol. 105, 2335-2345
- SMITH, W.C., HARLAND, R.M. (1992)  
Expression cloning of *noggin*, a new dorsalising factor localised to the  
Spemann organiser in *Xenopus* embryos.  
Cell 70, 829-840
- SMITH, W.C., HARLAND, R.M. (1991)  
Injected Xwn1-8 RNA acts early in *Xenopus* embryos to promote formation  
of a vegetal dorsalising centre.  
Cell 67, 753-765
- SMITH, J.C., SLACK, J.M.W. (1984)  
Dorsalization and neural induction properties of the organizer in *Xenopus  
laevis*.  
J. Embryol. Exp. Morph. 78, 279-317
- SOLER, A.P., KNUDSEN, K.A. (1991)  
Colocalisation of NCAM and N-Cadherin in avian skeletal myoblasts.  
Dev. Biol. 148, 389-392
- SORKIN, B.C., HOFFMAN, S., EDELMAN, G.M., CUNNINGHAM, B.A.  
(1984)  
Sulfation and phosphorylation of the neural cell adhesion molecule, NCAM.  
Science 225, 1476-1478
- SORKIN, B.C., HEMPERLY, J.J., EDELMAN, G.M., CUNNINGHAM,  
B.A. (1988)  
Proc. Natl. Acad. Sci. USA 85, 7617-7621

SORKIN, B.C., GALLIN, W.J., EDELMAN, G.M., CUNNINGHAM, B.A.  
(1991)

Genes for two calcium-dependent cell adhesion molecules have similar structures and are arranged in tandem in the chicken genome.

Proc. Natl. Acad. Sci. USA 88, 11545-11549

SPEMANN, H. (1901)

Über Korrelationen in der Entwicklung des Auges.

Verh. Anat. Ges. 15 Vers. Bonn (Anat. Anz 15), 61-79

SPEMANN, H., MANGOLD, H. (1924)

Über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren.

Arch. mikrosk. Anat. EntwMech. 100, 599-638

STRATAGENE (1987)

Predigested Lambda ZAP/EcoRI: Cloning kit instruction manual.

Stratagene, 11099 North Torrey Pines Road, La Jolla, California. USA

STRATAGENE (1989a)

Recovery of single-stranded DNA from cells containing pBluescript plasmids.

Stratagene, 11099 North Torrey Pines Road, La Jolla, California. USA

STRATAGENE (1989b)

pBluescript II exo-mung DNA sequencing system: Instruction manual.

Stratagene, 11099 North Torrey Pines Road, La Jolla, California. USA

SUTCLIFFE, J.G., MILNER R.J., GOTTESFELD, J.M., LERNER, R.A.  
(1984)

Identifier sequences are transcribed specifically in brain.

Nature (London) 308, 237-241

TAIRA, M., JAMRICH, M., GOOD, P.J., DAWID, I.B. (1992)

Genes Dev. 6, 356-366

TAKEICHI, M., ATSUMI, T., YOSHIDA, C., UNO, K., OKADA, T.S.  
(1981)

Selective adhesion of embryonal carcinoma cells and differentiated cells by  $Ca^{2+}$ -dependent sites.

Dev. Biol. 87, 340-350

TAKEICHI, M. (1987)

Cadherins: a molecular family essential for selective cell-cell adhesion and animal morphogenesis.

Trends Genetics 3 (8), 213-216

TAKEICHI, M. (1988)

The Cadherins: cell-cell adhesion molecules controlling animal morphogenesis.

Development 102, 639-655

TAKEICHI, M. (1990)

Cadherins: A molecular family important in selective cell-cell adhesion.

Ann. Rev. Biochem. 59, 237-252

TAKEICHI, M. (1991)

Cadherin cell adhesion receptors as a morphogenetic regulator.

Science 251, 1451-1455

TAMURA, T., SUMITA, K., HIROSE, S., MIKOSHIBA, K. (1990)

Core promoter of the mouse myelin basic protein gene governs brain-specific transcription *in vitro*.

EMBOJ 9 (10), 3101-3108

THIERY, J-P., DUBAND, J-L., RUTISHAUSER, U., EDELMAN, G.M. (1982)

Cell adhesion molecules in early chicken embryogenesis.

Proc. Natl. Acad. Sci. USA 79, 6737-6741

THIERY, J-P., BRACKENBURY, R., RUTISHAUSER, U., EDELMAN, G.M. (1977)

J. Biol. Chem. 252, 6841-6845

THUMMEL, C.S., BOULET, A.M., LIPSHITZ, H.D. (1988)

Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection.

Gene 74, 445-456

TOSNEY, K.W., WATANABE, M., LANDMESSER, L., RUTISHAUSER, U. (1986)

The distribution of NCAM in the chick hindlimb during axon outgrowth and synaptogenesis.

Developmental Biology 114, 437-452

SORKIN, B.C., GALLIN, W.J., EDELMAN, G.M., CUNNINGHAM, B.A.  
(1991)

Genes for two calcium-dependent cell adhesion molecules have similar structures and are arranged in tandem in the chicken genome.  
Proc. Natl. Acad. Sci. USA 88, 11545-11549

SPEMANN, H. (1901)

Über Korrelationen in der Entwicklung des Auges.  
Verh. Anat. Ges. 15 Vers. Bonn (Anat. Anz 15), 61-79

SPEMANN, H., MANGOLD, H. (1924)

Über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren.  
Arch. mikrosk. Anat. EntwMech. 100, 599-638

STRATAGENE (1987)

Predigested Lambda ZAP/EcoRI: Cloning kit instruction manual.  
Stratagene, 11099 North Torrey Pines Road, La Jolla, California. USA

STRATAGENE (1989a)

Recovery of single-stranded DNA from cells containing pBluescript plasmids.  
Stratagene, 11099 North Torrey Pines Road, La Jolla, California. USA

STRATAGENE (1989b)

pBluescript II exo-mung DNA sequencing system: Instruction manual.  
Stratagene, 11099 North Torrey Pines Road, La Jolla, California. USA

SUTCLIFFE, J.G., MILNER R.J., GOTTESFELD, J.M., LERNER, R.A.  
(1984)

Identifier sequences are transcribed specifically in brain.  
Nature (London) 308, 237-241

TAIRA, M., JAMRICH, M., GOOD, P.J., DAWID, I.B. (1992)  
Genes Dev. 6, 356-366

TAKEICHI, M., ATSUMI, T., YOSHIDA, C., UNO, K., OKADA, T.S.  
(1981)

Selective adhesion of embryonal carcinoma cells and differentiated cells by  $Ca^{2+}$ -dependent sites.  
Dev. Biol. 87, 340-350

TAKEICHI, M. (1987)

Cadherins: a molecular family essential for selective cell-cell adhesion and animal morphogenesis.

Trends Genetics 3 (8), 213-216

TAKEICHI, M. (1988)

The Cadherins: cell-cell adhesion molecules controlling animal morphogenesis.

Development 102, 639-655

TAKEICHI, M. (1990)

Cadherins: A molecular family important in selective cell-cell adhesion.

Ann. Rev. Biochem. 59, 237-252

TAKEICHI, M. (1991)

Cadherin cell adhesion receptors as a morphogenetic regulator.

Science 251, 1451-1455

TAMURA, T., SUMITA, K., HIROSE, S., MIKOSHIBA, K. (1990)

Core promoter of the mouse myelin basic protein gene governs brain-specific transcription *in vitro*.

EMBOJ 9 (10), 3101-3108

THIERY, J-P., DUBAND, J-L., RUTISHAUSER, U., EDELMAN, G.M. (1982)

Cell adhesion molecules in early chicken embryogenesis.

Proc. Natl. Acad. Sci. USA 79, 6737-6741

THIERY, J-P., BRACKENBURY, R., RUTISHAUSER, U., EDELMAN, G.M. (1977)

J. Biol. Chem. 252, 6841-6845

THUMMEL, C.S., BOULET, A.M., LIPSHITZ, H.D. (1988)

Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection.

Gene 74, 445-456

TOSNEY, K.W., WATANABE, M., LANDMESSER, L., RUTISHAUSER, U. (1986)

The distribution of NCAM in the chick hindlimb during axon outgrowth and synaptogenesis.

Developmental Biology 114, 437-452

- TREISMAN, J., HARRIS, E., DESPLAN, C. (1991)  
The paired-box encodes a second DNA-binding domain in the Paired homeodomain protein.  
*Genes and Development* 5, 594-604
- VIZE, P.D., MELTON, D.A., HEMMATI-BRIVANLOU, A., HARLAND, R.M. (1991)  
Assays for gene function in developing *Xenopus* embryos.  
*Meths. Cell Biol.* 36, 367-387
- VLEMINCKX, K., VAKAET, L., MAREEL, M., FIER, W., VAN ROY F. (1991)  
Genetic manipulation of E-Cadherin expression by epithelial tumor cells reveals an invasion suppressor role.  
*Cell* 66, 107-199
- VOGT, W. (1929)  
Gestaltungsanalyse am Amphibienkeim mit örtlicher Vitalfärbung. II Teil. Gastrulation und Mesodermbildung bei Urodelen und Anuren.  
*Wilhelm Roux's Arch. EntwMech. Org.* 120, 384-706
- WALLIS, J., WALSH, F.S. (1992)  
Structure of the genes encoding the neural cell-adhesion molecules NCAM and N-Cadherin.  
*Biochem. Soc. Trans.* 20 (3), 656-658
- WALSH, F.S. (1988)  
The NCAM gene is a complex transcriptional unit.  
*Neurochem. Int.* 12, 262-267
- WALSH, F.S., PAREKH, R.B., MOORE, S.E., DICKSON, G., BARTON, C.H., GOWER, H.J., DWEK, R.A., RADEMACHER, T.W. (1989)  
Tissue specific O-linked glycosylation of the neural cell adhesion molecule (N-CAM)  
*Development* 105, 803-811
- WALSH, F.S., BARTON, C.H., PUTT, W., MOORE, S.E., KELSELL, D., SPURR, N., GOODFELLOW, P.N. (1990)  
N-Cadherin gene maps to human chromosome-18 and is not linked to the E-Cadherin gene.  
*J. Neurochemistry* 55 (3), 805-812

WHARTON, K.A., JOHANSEN, K.M., XU, T., ARTAVANIS-TSAKONAS, S. (1985)  
Cell 43, 567-581

WHEELOCK, M.J., KNUDSEN, K.A. (1991)  
N-Cadherin-associated proteins in chicken muscle.  
Differentiation 46, 35-42

WILSON, H.V. (1907)  
J. Exp. Zool. 5, 245-258

WILSON, C., CROSS, G.S., WOODLAND, H.R. (1986)  
Tissue-specific expression of actin genes injected into *Xenopus* embryos.  
Cell 47, 589-599

WOLFFE, A.P. (1991)  
*Xenopus* transcription factors: key molecules in the developmental regulation  
of differential gene expression.  
Biochemical Journal 278, 313-324

WOODLAND, H.R. (1993)  
Identifying the three signals.  
Current Biology 3 (1), 27-29

WUENSCHALL, C.W., MORI, N., ANDERSON, D.J. (1990)  
Analysis of SCG10 gene expression in transgenic mice reveals that neural  
specificity is achieved through selective derepression.  
Neuron 4, 595-602

YAMADA, T. (1990)  
Regulations in the induction of the organized neural system in amphibian  
embryos.  
Development 110, 653-659

YAMAMOTO, M., KO, L.J., LEONARD, M.W., BEUG, H., ORKIN, S.H.,  
ENGEL, J.D. (1990)  
Activity and tissue-specific expression of the transcription-factor NF-E1  
multigene family.  
Genes & Dev. 4, 1650-1662

YOSHIDA, C., TAKEICHI, M. (1982)  
Teratocarcinoma cell adhesion: Identification of a cell surface protein  
involved in calcium-dependent cell adhesion.

Cell 28, 217-224

YOSHIDA-NORO, C., SUZUKI, N., TAKEICHI, M. (1984)

Molecular nature of the calcium-dependent cell-cell adhesion system in mouse teratocarcinoma and embryonic cells studied with a monoclonal antibody.

Dev. Biol. 101, 19-27

ZHENG, M., HEINRICH, G. (1988)

Structural and functional analysis of the promoter region of the nerve growth factor gene.

Mol. Brain Res. 3, 133-140

The guidelines followed for the presentation of this research thesis were those of the University of Warwick, as revised 17 October 1989.

THE BRITISH LIBRARY

BRITISH THESIS SERVICE

TITLE ANALYSIS OF THE PROMOTER REGION OF THE  
XENOPUS BOREALIS N-CADHERIN GENE

AUTHOR P M  
WEBBER

DEGREE Ph D

AWARDING BODY Warwick University

DATE 1993

THESIS DX182708  
NUMBER

THIS THESIS HAS BEEN MICROFILMED EXACTLY AS RECEIVED

The quality of this reproduction is dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction. Some pages may have indistinct print, especially if the original papers were poorly produced or if awarding body sent an inferior copy. If pages are missing, please contact the awarding body which granted the degree.

Previously copyrighted materials (journals articles, published texts etc.) are not filmed.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that it's copyright rests with its author and that no information derived from it may be published without the author's prior written consent.

Reproduction of this thesis, other than as permitted under the United Kingdom Copyright Designs and Patents Act 1988, or under specific agreement with the copyright holder, is prohibited.

**DX**

**182708**