1. INTRODUCTION

Development to multicellularity in *Dictyostelium discoideum* is achieved by the aggregation of formerly individual identical cells. Cells within the aggregates initiate alternate pathways of differentiation that results in the appearance of two distinct cell types and morphogenesis (Gross, 1995). The cell differentiation aspect of development is achieved mainly by selective gene expression, a process controlled chiefly at the level of transcription. The complexity of the process may be understood when both *cis*- and *trans*-acting factors, as well as the required signals, are well defined. Having knowledge of interactions between these factors in various model systems may help to characterize general mechanisms required for activating gene expression and thus provide insight into evolutionarily conserved features.

The cellular slime mold, *Dictyostelium discoideum*, is an excellent eukaryotic model system for investigating the events involved in cell differentiation, eukaryotic development, and aging. If a homogeneous population of amoebae is starved, it enters into a developmental process resulting in a morphologically distinct structure consisting of two different cell types. The fact that the organism possesses multicellular characteristics similar to higher organisms, yet clearly separates growth and differentiation as two distinct events establishes this organism as a useful model system.
Glycogen phosphorylase 2 (*gp-2*) is a developmentally regulated gene that is first expressed during the transition from growth to differentiation in *Dictyostelium discoideum.* The gene product of *gp-2* catalyzes the degradation of glycogen into glucose monomers. These monomers are then used to produce the structural end products of terminal differentiation. Based on the *gp-2* gene's temporal expression, as well as its critical role in the development of this organism, it is likely that studying the transcriptional role of this gene may yield insight into the genetic switch responsible for the transition from cellular division to differentiation. In addition, due to the AT-rich nature of the *gp-2* promoter, determining the *cis*-acting elements of this gene may prove

to be very informative in that it may reveal novel mechanisms required to achieve selective gene expression under the constraint of this aberrant base composition.

BACKGROUND AND LITERATURE REVIEW

1.1 The life cycle of *Dictyostelium discoideum* :

The life cycle of *Dictyostelium discoideum* is simple and short, consisting of few developmental stages and only two different cell types **(Figure 1.1).** These properties make it a useful organism for following genetic and biochemical events. When there is enough food, the slime mold exists as single-celled vegetative amoebae, which feed on the bacteria of decaying organic materials. The amoebae cells are genetically, biochemically and morphologically identical (Loomis, 1982). Once food becomes scarce, the cells are triggered to differentiate. Eight hours after the onset of starvation, cellular aggregation is seen. Each amoebae secretes adenosine 3, 5 -monophosphate (cAMP) that chemotaxically lead the cells to aggregate. The signal is put out periodically, once every six minutes. Because secreted cAMP only affects the cells within a certain distance (6 to 10 cell diameters from the signaling cell), aggregates of approximately 100,000 cells form by 10 h. This multicellular mound goes on to form the final fruiting body. The mound has two distinct cell types, prespore cells comprise about 80% of the mound, while prestalk cells account for the remaining 20% of the cells. At this point, subsequent cell movements cause the prestalk cells to shape a tip on top of the mound. The tip elongates and then bends until the structure lies flat to form the slug. The slug has prestalk cells at the anterior and prespore cells at the posterior end. The slug provides aggregated amoebae with the ability to move to a more desirable place, using light, heat or both as guide for migration. Also, the slug can be considered

as an intermediate stage for the activation of developmental programs that make up the mature fruiting body (Loomis, 1982). In fact, biochemical and genetic events in amoebae are initiated just before the slug is formed. After the slug forms, culmination occurs and cell sorting continues, with prestalk and prespore cells exchanging places. Differentiation into mature spore and stalk cells proceeds. While stalk cells are programmed to die, spore cells can remain viable despite unsuitable conditions, such as starvation and changes in temperature. As soon as the conditions become favorable, the spore cells germinate into single-celled amoebae (Kessin *et al*., 1992). In the laboratory, differentiation is completed in 24 h at 23 C.

1.2. Developmental Effects of Morphogenic Signals: cyclic AMP (cAMP), Differentiation Inducing Factor (DIF), adenosine and ammonia:

cAMP is reported to be a significant signal for aggregation as well as for activation and repression of some genes during development. Starvation causes the activation of the adenylate cyclase enzyme that converts ATP into cAMP. Within 4 h of starvation, cells begin to secrete cAMP. cAMP is sensed by its specific receptor on the cell surface, and the information is passed into the cell to activate a mechanism that will be used to accomplish necessary actions, such as chemotaxis or specific gene expression.

The cells gain the ability to move toward cAMP and then to produce more cAMP after interaction with other cells. Consequently, increased cAMP production attracts more amoebae cells to aggregate into the center. cAMP appears to induce all cell types, however, prestalk cells need DIF during later stages of development. In order for

prespore cells to complete development, a continuous level of extracellular cAMP is required.

DIF changes the expression of genes that are important in stalk formation. Specifically, it activates the genes responsible for generating stalk cells, and inhibits spore cell-specific genes, even at very low concentrations (0.1-1nM) (Williams *et al*., 1991). It has also been reported that DIF seems to be involved in regulating cAMP levels.

Ammonia is produced during development and appears to inhibit the formation of the culminate. The mechanism is unknown; however, it has been shown that high concentrations of ammonia reduce the levels of intracellular cAMP. Similarly, adenosine also seems to play an antagonistic role to cAMP during aggregation. It has also been suggested that cGMP and intracellular pH should be considered when analyzing the interactions that trigger gene expression.

1.3. GLYCOGEN METABOLISM

Glycogen metabolism is an important biochemical pathway in differentiation in *Dictyostelium discoideum* **(Figure 1.2)**. Considerable speculation has focused on glycogen metabolism as a criterion for differentiation in *Dictyostelium discoideum* (Wright, 1973; Hames *et al.*, 1972). When the cells are forced to differentiate because of unsuitable conditions, first an increase (with glycogen synthase activity) and then a decrease (with glycogen phosphorylase I and II activities) in glycogen level is observed. Degradation of glycogen delivers a source of stored energy and glucose monomers that are used to form stalk and spore cells (Gustafson and Wright, 1972). The energy for differentiation is mainly provided by protein degradation. Glucose is primarily used to make the cell wall (cellulose) of spore and stalk cells. It has been suggested that cells containing higher amounts of glycogen are more likely to form spore cells, while cells

containing lower amounts of glycogen form stalk cells (Inouyr and Takeuchi, 1982). In early development, glycogen is the only polysaccharide in *Dictyostelium* cells (White and Sussman, 1963).

Glycogen phosphorylase is responsible for the breakdown of glycogen into glucose monomers. Two isozymes of this enzyme exist in *Dictyostelium discoideum*, gp-1 and gp-2. The gp-1 and gp-2 enzymes are encoded by two different genes and expressed at different stages of the life cycle (Rutherford *et al.*, 1992; Sucic, 1992; Yin, 1993). The *gp-1* and *gp-2* coding regions are 52% identical, but 5' upstream non-coding regions are distinct. The gp-1 isoform is active in early stages, provides the energy necessary for development and is 5'-AMP dependent. The *gp-2* activity, on the other hand, is expressed later, increases in expression throughout development and is independent of 5'-AMP.

The levels of *gp-2* expression are affected by various morphogen factors; specifically cAMP, DIF, ammonium and adenosine. The roles of these four diffusible molecules in *gp-2* expression have been analyzed (Yin *et al*., 1993). According to the findings: cAMP and DIF induce *gp-2* individually. DIF requires cell-cell contact for induction of *gp-2*, while cAMP does not. cAMP alone induces *gp-2* more than cAMP and DIF together, suggesting that DIF inhibits cAMP induction. NH_3 inhibits both DIF and cAMP induction, while adenosine inhibits DIF, but not cAMP induction.

gp-2 mRNA first appears 8 h after starvation during aggregation and is maximal at culmination. The protein is detected in both prespore and prestalk cells. In prespore cells, gp-2 is active during early stages of development, whereas in prestalk cells, gp-2 activity is found only later in development (Sucic *et al*, 1993). The product of *gp-2* is used to generate the raw materials necessary to build the structures required for terminal differentiation (Sucic *et al*., 1993; Rogers *et al*., 1994).

1.4. TRANSCRIPTIONAL REGULATION OF *GP-2* EXPRESSION:

The *gp-2* promoter is 88% AT-rich, which is characteristic of non-coding sequences in *Dictyostelium discoideum.* These AT-rich regions of the promoter are interrupted by seven repeated sequences: two C boxes (ACCCACT), two TAG boxes (TAAAAATGGA) and three TA boxes (TAATTATAA). The locations of these boxes are shown in **Figure 1.3**.

Deletion analysis and cAMP inducibility experiments have shown that these repeated regions are crucial to the expression of *gp-2.* When the boxes are removed by 5' deletion, a decrease in luciferase reporter gene activity is detected (Rutherford *et al*., 1997; Sucic et al., 1992). Specifically, deletion of the upstream TAG box resulted in a dramatic drop in expression. Continued deletion of the promoter through the 3' C box resulted in nearly a complete loss of expression **(Figure 1.4)**.

Site directed mutagenesis of the downstream C box has generated a mutant where CCC is changed to AGT and one or more nucleotides in the upstream A-rich region have been deleted (McCaffery *et al*., manuscript in preparation). Also, TAG box mutagenesis has been completed. Both types of mutants result in the loss of reporter gene expression during development. These two boxes have also been found to be cAMP-responsive *cis*-acting elements.

Evidence from other genes in *Dictyostelium discoideum* implicates the importance of GC-rich promoter elements in gene expression and specifically in cAMP induction (Powell-Coffman *et al*., 1994). The deletion and mutational analyses of the *gp-2* promoter are consistent with this result.

DNase I hypersensitive analysis of active (first finger) and inactive (amoebae) *gp-2* promoters have revealed several hypersensitive sites. The hypersensitive sites map approximately –250bp, -290bp, -350bp, -445bp and –505bp. One of these sites contains C box regions. During cell proliferation, when *gp-2* is not active, this region produces no

DNase I hypersensitive sites, however, this same C box becomes hypersensitive when *gp-2* is active (Favis *et al*., submitted).

Given all these data, it is highly possible that C box(es) and TAG box(es) are regulatory promoter elements of the *gp-2* gene. Footprint data of these specific sequences will therefore be vital in identifying the precise sequence elements involved in gene regulation. The data that will be obtained can then be used to purify the transcription factor(s) involved in the expression of *gp-2*.

1.5. Replication protein A in *Dictyostelium discoideum* (DdRPA):

Gel shift assays have revealed that TAG box and C box regions of the promoter act as sites for a specific DNA binding activity. Non-specific competitors, poly dA:dT and poly dI:dC, were unable to compete with these regions for binding, while other sequences were easily competed (Rutherford *et al*., submitted). Our laboratory identified the binding activity as Replication Protein A (referred to as DdRPA) in *Dictyostelium discoideum* (Rutherford *et al*., submitted). Partial purification was accomplished by affinity chromotography. Amino acid sequencing and subsequent DNA sequencing of the cloned gene confirmed the identification of the binding activity as the *Dictyostelium* homologue of RPA.

DdRPA produces a 0.40 Rf gel shift band in undifferentiated cells, and a 0.32 Rf band in differentiated cells. We have determined that the 0.40 Rf band corresponds to a form of DdRPA consisting of 62 kDa, 35 kDa and 18 kDa subunits, whereas the 0.32 Rf corresponds to a form of the protein comprised of 84 kDa, 35 kDa, and 18 kDa subunits. Southwestern analysis revealed that the biggest subunits are the DNA binding domains. The gene encoding the large subunit (84 kDa) was cloned and characterized (Rutherford *et al*., submitted).

Northern blot data indicates that DdRPA expression level was constant throughout development and the size of mRNA of all stages is the same. This size corresponds to the 84 kDa form. So, the 62 and 84 kDa are the products of the same gene. The biggest subunit of DdRPA is transcribed as 84 kDa and then undergoes posttranslational modifications to become 62 kDa form (Rutherford *et al.*, submitted).

Replication Protein A from different organisms has been found to have functional roles in DNA replication, repair and recombination, as well as in the regulation of transcription. These RPA's have 3 subunits, a feature that is evolutionarily conserved. Some of the subunits are thought to be involved in interactions with other proteins. It has been proposed that RPA can act as a replication protein in undifferentiated cells and as a transcription factor during the differentiation process in *Dictyostelium discoideum*. Determining the DNA binding properties of these two forms of DdRPA on the *gp-2* promoter may provide insight into how this transition may be accomplished.

2. MATERIAL AND METHODS

2.1. CELL CULTURE AND HARVEST

Dictyostelium discoideum strain AX-3K was grown in liquid HL-5 medium (60 mM glucose, 1% Oxoid peptone, 0.5% Oxoid yeast extract, 2mM Na_2HPO_4 and 3 mM KH_2PO_4 at 21^oC on a rotary shaker at 130 rpm until $5x10^6$ cells were generated.

Vegetative amoebae cells grown to log phase were placed in differentiation buffer (20 mM N-morpholinoethanesulfonic acid (pH 6.8), 0.2 mM $CaCl_2$, 2 mM $MgSO_4$).

For protein purification, thirty 2 liter-flasks containing 500 ml HL5 with 10^7 cells/ml were harvested using a flow through centrifuge and the pellet was washed with 1 liter of differentiation buffer. The cells were then transferred to flasks containing 500 ml MES and placed on a rotary shaker (New Brunswick model 626) for 6hr or O/N, at 130 rpm at 21^oC.

For nuclear extract preparation, amoebae were either harvested immediately or transferred to MES-LPS agar plates and allowed to develop. Cells were harvested after 8 hr, 16 hr and 21 hr of development (Rogers *et al*., 1992).

2.2. NUCLEAR EXTRACT PREPARATION

Approximately $5x10^6$ – $1x10^7$ harvested cells (as explained above) were transferred onto MES-LPS buffer agar plates. Cells were incubated at 22^oC and allowed to differentiate. Each developmental time point was harvested again and spun at 3000xg for 2 min. The supernatant was poured off and the pellet was resuspended in 5 cell volumes of Buffer A (10 mM Tris Cl pH 7.9, 5 mM $MgCl_2$, 1 μg/ml aprotinin, 0.1 μg/ml chymostatin, 1 μg/ml pepstatin, 1 μg/ml antipain, 100 μg/ml TAME, 25 μg/ml TPCK, 10 μg/ml TLCK, 1

μg/ml leupeptin, 100 μg/ml PMSF). NP40 was added to a final concentration of 2%. The cell solution was mixed, clumps were broken with a dounce homogenizer and then passed through a 21-G needle and evaluated under a microscope for successful breakage of the cells. The lysate was then centrifuged at 10,000 x g for 10 min and the supernatant (cytoplasmic fraction) was aliquoted into eppendorf tubes and stored at –80^{o}C. The pellet was next washed twice by homogenizing with a dounce in one original cell volume of Buffer A. The resuspended material was passed through the needle and centrifuged as above. The supernatant from each wash (nuclear wash 1 and nuclear wash 2) was aliquoted and stored at –80^{o}C. The remaining pellet was then weighed and resuspended in (0.5 w/v). NaCl was added to a final concentration of 0.3 M. The resuspension was homogenized again as above, gently agitated at 4^{o}C for 30 min and pelleted for 10 min at 10,000 x g. All of the steps above were done on ice. The final supernatant was aliquoted and stored at –80^{o}C for further use. The protein concentration for each nuclear extract was determined using the Bradford assay.

2.3. PROTEIN (DdRPA) EXTRACT PREPARATION:

Starved cells (6h or O/N) were harvested, resuspended in 1 liter of milli-Q water, divided into two and centrifuged for 5 min at 1,300 x g. The cells were then ruptured by adding 5 volumes of Buffer X (25 mM Tris (pH 7.9), 0.5 mM EDTA, 150 mM EDTA, 150 mM sucrose, 2% Nonidet P40) and incubated on ice for 15 min. The lysate was centrifuged and the resulting extract was stored at - 80^{o}C.

2.3.1. DEAE Sephacel Batch Assay

The cell extract was thawed at 37^{o}C. Phenylmethylsulfunyl fluoride (PMSF) was added to the concentration of 10 μg/ml, the mixture was stirred at 4^{o}C for 15 min and centrifuged at 20,000 x g for 30 min. The supernatant was passed through glass wool and combined with DEAE Sephacel resin (Pharmacia, 1 g resin/ 6 ml packed cells) in a 1 liter bottle. After shaking 60 min at room temperature at 150 rpm, the suspension was poured

into a column and washed with 20mM Tris buffer (pH 7.9). The resulting fractions were assayed for DNA binding activity using a gel shift assay and the 3' C probe.

2.3.2. DNA Affinity chromatography

Preparation and coupling the DNA target sequence to CNBr Sepharose-4B was performed as previously described (Kampang, 1995). Briefly, 25bp double stranded oligonucleotide containing the 3' C box and surrounding sequences (5'GATCCAGTGCAAACTCACCCACT CACAAT3' (the 3' C probe) was used as DNA to isolate binding proteins on an affinity column. The oligonucleotide strands were annealed and concatemerized (using a non-promoter GATC sequence added to the 5' end) and coupled to CNBr Sepharose-4B (Kampang, 1995).

2.4. GEL SHIFT ASSAY:

The 3' C probe was labeled using a fill-in reaction with alpha-^{32}P dATP. The probe was separated from unincorporated nucleotides by using a Biospin 6 Column (Biorad). Reaction mixture, containing the labeled probe (30,000 cpm), 2 µl of extract and 7.5 µl reaction buffer (22 mM Tris-HCl, K_2HPO_4/KH_2PO_4, Imidazole, or MES containing 8.5% glycerol, 43 mM NaCl, 4.4 mM $MgCl_2$, 4.4 mM EDTA, 2.2 mM dithiothreitol, 4.4% Nonidet P40) was incubated at room temperature for 20 min. After adding 2 µl of loading dye (250 mM Tris-HCl (pH 7.9) 50% glycerol, 4 mg/ml bromphenol blue, 4 mg/ml xylene cyanol), the gel shift assay was performed using 5-6% polyacrylamide gels in TBE buffer. The gel was dried and autoradiographed for 10-20 hours.

2.5. FOOTPRINT

2.5.1. NADIR (NEW AMPLIFIED APEX-DERIVED IN-VITRO REACTION)(PRIMER EXTENSION FOOTPRINTING)

2.5.1.1. TEMPLATE PREPARATION

Gp-2 construct P2124 (Yin, 1993) was prepared using Qiagen Plasmid Maxi Prep Kit (Qiagen, 1992). This 5.6 kb construct is contains all of the upstream region of the *gp-2*. 5 μg of the plasmid preparation was linearized using *BamHI*, which cuts exclusively in the vector sequence. Following incubation at 37^{o}C for 3 h, the reaction mixture was diluted to the concentration of 10 ng/μl with milliQ water and stored at –20^{o}C for further use.

2.5.1.2. PRIMER PREPARATION

To examine footprints covering the C box regions, 9001100 primer (5'GCTATTTTTTCACAGATACC3') was used. To visualize the binding on the complementary strand and the TAG primer(5'TAAAAATGGATAAACTAAAAATGGA 3') was used. For the TAG box region, the 9001100 primer was used to view the region on the non-coding strand. The N primer(5'AGAACAAATAAATTATTGATTGTGTT3') and 5'C COMP primer (5'TCCAACAATAGTGGGTAGAA3') were used to view the TAG box region on the coding strand (see **Figure 2.1** for locations of the primers).

5 μl of 10 μM primer was end-labeled (Ausebel *et al*., 1995) using 100 uCi γ-^{32}P-ATP (6000 C_i/mmol γ-^{32}P-ATP, NEN Dupont). Unincorporated nucleotides were eliminated using a Biospin6 column (Biorad). The specific activity of the primers was 1-2 x 10^6 CPM/pmol.

2.5.1.3. IN-VITRO REACTION

50 ng linearized P2124 template, 2 μg poly dI: dC or poly dA: dT (Pharmacia), purified DdRPA or nuclear extract and Buffer B (10mM bis-Tris-Cl, 2.5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM EDTA, 200 mM KCl, 100 μg/ml BSA, 2 μg/ml herring sperm DNA; titrate to pH 7.0) (Ausubel 1995) was added to a final volume of 195 μl then incubated at

room temperature for 20 min. 10 mU DNaseI (Worthington) was diluted 1:1000 in 50 mM Tris-Cl pH 7.2, 10 mM $MgSO_4$, 1 mM DTT, 50 % glycerol. 5 µl of this enzyme was used to digest DNA at room temperature for 8 min. Following the digestion, 700 µl of stop solution (92% ethanol, 0.7 M NH_4SO_4, 5 mg yeast t-RNA) was added, the reaction was vigorously vortexed and put into a dry ice/ethanol bath. The resuspension was centrifuged and the pellet was washed with 80% ethanol to remove the salt. The pellet was then resuspended in 20 µl annealing mix (APEX buffer (10 mM $MgCl_2$, 50 mM KCl, 25 mM Tris-Cl, pH 7.2) and 10^6 CPM of labeled primer) then covered with mineral oil and placed in a Perkin Elmer Cetus thermocycler. The DNA was denatured at 94°C for 5 min and slowly cooled to 78°C over 30 min before 80 µl extension mix (APEX buffer, 300 µM dNTP, 1 U Taq polymerase (Boehringer Mannheim)) was added directly into the annealing mixture. Following this, 20 rounds of primer extension were performed (94°C for 1 min, 84°C in 10 sec, 52°C in 3 min, 52°C for 1 min, and 78°C for 1 min). In some cases, Tm's were changed based on the primer used. The primer extension product was re-suspended in 350 µl stop solution at –20°C for at least 1 h and then centrifuged at 18,000 x g. The pellet was washed with 80% ethanol and re-suspended in 11 µl formamide loading dye (Ausubel, 1995). A 6% denaturing polyacrylamide gel was pre-run for 20-25 min at 2000V and after the samples were heated for 5 min at 90°C, equal amounts of radioactivity were loaded into each well. The fragments were separated for approximately 2 h at the same voltage. The gel was then dried and subjected to autoradiography overnight with one or two intensifying screens based on the level of radioactivity.

For the other templates (460660K-9001100 PCR product, 4600660-9001100 cloned in Bluescript product and OS/XXM luciferase constructs), approximately the same number of DNA molecules were used as template for these reactions. DNaseI conditions were determined to be optimum at 5 min for the first and second templates and 15 min for OS/XXM luciferase constructs (McCaffery, manuscript in preparation).

2.5.2. CONVENTIONAL FOOTPRINT STRATEGY (ENDLABELING FOOTPRINT TECHNIQUE)

The footprint protocol as described in Current Protocols (Ausubel *et al.*, 1995) was employed with minor modifications as described below.

2.5.2.1. TEMPLATE PREPARATION:

Genomic DNA was isolated from amoebae cells (Nellen *et al.*, 1987). A Bluescript clone containing 453 bp of the promoter with all boxes (-726bp to -273bp) was prepared using Qiagen Plasmid Maxi Prep Kit (Qiagen, 1992). Appropriate primers were used with this template to produce the desired fragment using the polymerase chain reaction (see PCR). The products were checked with agarose gel electrophoresis and purified with the Gene Clean System (Bio 101, Vista, CA). The yield was measured with the EtBr dot quantitative analysis (Ausebel, 1995).

2.5.2.2. PRIMER PREPARATION

Primers of N, 9001100, 460660 (5' AAATCCATAAAGTCCAAAAT 3'), TAG, TAG COMP (5' TCCATTTTTAGTTTATCCATTTTTA 3') were used (see **Figure 2.1**). In order to ensure only one 5' end could be radioactively endlabeled, the 5' ends of the appropriate primers as first phosphorylated with cold rATP using T4 polynucleotide kinase. One kinased and one regular primer were used for PCR reaction (see **Figure 2.2**). Success of phosphorylation was checked using the exonuclease assay.

2.5.2.3. POLYMERASE CHAIN REACTION (PCR)

Optimum PCR conditions for each primer set were determined. The amplification reaction contained; 30 pmol kinased primer, 20 pmol regular primer, 5 µl 10 X PCR buffer (Ausubel, 1995), 0.08 mM dNTP, 2.5 U Taq polymerase in a final volume of 50

µl. The following program was run using Perkin Elmer thermocycler; 2.30 min at 94^{o}C; 0.30 min at 94^{o}C, 0.30 min at 52^{o}C, 1 min at 72^{o}C, for 34 cycles, and 10 min at 72^{o}C.

For the N-9001100K primer pair, all conditions were the same except 6.5 µl 10X PCR buffer was used and Tm of 48^{o}C was applied.

The 5' end of the purified (see 2.5.2.1) PCR product (180-300 ng/total) was radioactively endlabeled as described in 2.5.1.2. The specific activity was measured using a scintillation counter and was generally 100,000-200,000 CPM/µl.

2.5.2.4. IN-VITRO REACTION

A cocktail containing 2.5-4 ng PCR product (radiolabeled at one end), 2 µg poly dI:dC or poly dA:dT (Pharmacia) and Buffer B (10 mM bis-Tris-Cl, 2.5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM EDTA, 200 mM KCl, 100 µg/ml BSA, 2 µg/ml herring sperm DNA; pH 7.0) (Ausubel 1995)) was prepared. To the cocktail, purified DdRPA or nuclear extract was added and allowed to incubate at room temperature for 20 min. The final volume was 200 µl. 0.01U DNaseI (Worthington) was added for partial digestion of the template. The reaction was stopped by adding 700 µl of stop solution (92% ethanol, 0.7 M NH_4SO_4, 5 mg yeast t-RNA), vortexing vigorously and putting into a dry ice/ethanol bath. Following the precipitation, the resuspension was centrifuged and the pellet was washed with 80% ethanol and resuspended in 6 µl formamide loading dye. 1 µl of this solution was used to measure the specific activity as determined by a scintillation counter. A 6% denaturing polyacrylamide gel was pre-run for 20-25 min at 2000 V, while the samples were denatured for 20-30 min at 90^{o}C. An equal amount of radioactivity was loaded into each lane of the gel. Separation and autoradiography was performed as described before (2.5.1.3).

2.5.3. MARKER GENERATION

The Maxam Gilbert reaction was used to generate a G ladder from the footprinting probes (Maxam and Gilbert, 1977, 1980; Rubin and Schmid, 1980; Ambrose and Pless, 1987). Dideoxy sequencing reactions (Sanger *et al*., 1977) were applied using the same primer used for footprint analyses.

3. RESULTS

3.1. OPTIMIZING THE CONDITIONS FOR FOOTPRINT ASSAYS

As explained earlier, the *gp-2* promoter and most of the other promoters of *Dictyostelium discoideum* are highly AT-rich. This property makes it difficult to obtain data using conventional footprint methodologies, because these methods require appropriate restriction sites near the region of interest or sequence-specific PCR primers with compatible Tm values (Ausubel *et al*., 1995). In fact, previous attempts to produce definitive footprints in this organism in other laboratories had been unsuccessful (C. L. Rutherford, personal communication).

A new approach has recently been implemented in our laboratory to circumvent the difficulties of analyzing the promoters of organisms with unusual base composition (Favis, unpublished). This technique combines the concepts of conventional *in vitro* and *in vivo* footprint techniques. It is faster, more economical and most importantly, more sensitive than conventional techniques; that is, even small amounts of protein can give visible footprints. Using this method, significantly larger templates can be assayed for protein binding, thus, increasing the specificity of DNA-protein interactions by increasing the abundance of nonspecific binding sites. The technique depends on multiple rounds of primer extension that results in a linear amplification of the targeted template. Because only one specific primer is required to visualize the binding, the region of interest can be more readily targeted for inspection at high resolution on a sequencing gel. Also, the technique significantly decreases the possibility of non-specific DNA-protein interactions, as relatively high amount of non-specific competitors can be used.

In order to meet the requirements of both techniques, certain factors needed to be optimized, specifically: 1) the method of nuclear extraction; 2) the purification of DdRPA; 3) the generation of appropriate template; and 4) the conditions for sufficient DNase I digestion.

A number of nuclear extraction methods have been employed in our laboratory. The activity of these extracts was checked primarily by gel shift assays and southwestern analyses by other members of the laboratory, although Bradford assays were also used to determine the total protein content. In general, extracts yielded an equivalent amount of total protein (~1ug/ul), however, first finger stage nuclear extracts yielded ten-fold more protein. The extracts yielding the best activity were used in footprint experiments.

For DdRPA, the binding activity of the concentrated purified protein was checked with gel shift assay using the 3'C probe (Figure 3.1). This method was thought to be more relevant, as it provided information about the active protein present in the preparations.

The conventional footprint technique relied on endlabeled DNA to serve as template. A number of probes were amplified using the polymerase chain reaction (see materials and methods). For a probe to be useful, the 5' end of the fragment must be very close to the targeted region of interest. Also the PCR product must be full-length and pure, otherwise aborted PCR fragments could produce bands on the sequencing gel that might obscure possible footprints. Generally, G-C rich regions of the promoter were used to generate PCR primers. Several primer pairs proved useful in producing the appropriate probes (data not shown) for analysis of the TAG and C box regions (460660-9001100, N-9001100, TAG-9001100, 460660-TAGcomp, 5'C comp-460660, 5'C-9001100). By selectively phosphorylating one primer in an amplification reaction, the unmodified primer could serve as the site for endlabeling, thus allowing one strand of DNA to be examined at a time. Some of these probes have been used in this manner for footprint analysis and all probes may prove useful for other purposes, such as gel shift assays.

For the new technique, the templates were prepared directly from Qiagen maxi-prep plasmid preparations. The templates were linearized by restriction digest and diluted to pertinent concentrations for use. The ability to use cloned DNA is another advantage of this technique. Due to the extreme AT-rich composition of the *gp-2* promoter, subcloning appropriate regions for footprint analysis was problematic, as convenient restriction sites were unavailable. Although PCR circumvents this dilemma, using cloned DNA eliminates the possibility of PCR-generated mutations. The template that was used most frequently was p2124 (5.6kb) (see materials and methods). In addition, the 460660-9001100 PCR product (453bp) and a cloned version of this product (3.4 kb) were also used to determine the effect of template length on footprints (*i.e.*, does increased bending and looping enhanced by increased template length affect the footprints detected?). Finally, one attempt was made to analyze the TAG box footprint by comparing wild type and mutated luciferase reporter constructs that have been shown to affect expression *in vivo* (data not shown). These templates were provided from site-directed mutagenesis experiments (McCaffery *et al.*, unpublished).

DNase I digestion conditions were standardized for all templates such that there would be one cleavage event per strand. This results in the formation of a ladder where each band is separated by one nucleotide. This has been done mostly on naked DNA, but in some cases proteins were added to observe the effect of protein on the overall efficiency of DNase I digestion. One such experiment using the new technique and both the 460660-9001100 PCR product and its cloned counterpart as templates is shown in Figure 3.2. As seen in the figure, there are no low molecular weight fragments produced when there is no digestion. As the digestion time is increased, the probes are cleaved partially and moved to the middle region of the gel, enabling the regions of interest to be observed. The best partial digestion conditions were chosen for footprint experiments; specifically, 5 min for the two templates (460660-9001100 PCR product and its cloned counterpart), 6 min in non-protein reactions, 8 min in protein containing reactions for p2124 and 13 min in non-protein and 15 min in protein containing reactions for OS and XXM luciferase constructs.

3.2. FOOTPRINTS OF REPLICATION PROTEIN A (DdRPA):

In the beginning, the immediate aim of the project was to ascertain the DNA binding ability of the purified replication protein A (DdRPA) and to see if there was a distinction between the binding affinities of undifferentiated and differentiated forms of the protein, and if this protein acted as a transcription factor that functioned by regulating the expression of the *gp-2* gene. Footprint analysis was one of the means to test this hypothesis by challenging the binding ability of DdRPA through the use of competitors.

The data shows that DdRPA can bind to some nucleotides in the C box region, the TAG box region, TA box-1 (-608bp) and upstream regions (-636bp). Thus, DdRPA behaves like a general DNA binding protein under these conditions, binding to many regions.

In terms of the difference between footprints generated by DdRPA from undifferentiated and differentiated cells, undifferentiated DdRPA bound to the TAG box region, whereas differentiated DdRPA did not.

All footprint experiments involving DdRPA produced a footprint covering the region (CAGTGC at –347 bp relative to the translation start site) between the two C boxes. This may imply that DdRPA prefers to bind to this region, rather than directly to the C boxes. The explanation for this may be that these nucleotides, not the 3'C box, were responsible for binding the protein during purification using affinity chromatography and the 3'C probe. Nevertheless, later experiments with more active protein showed that DdRPA binding could expand through the 5'C box region.

Both footprint techniques were employed using different templates (*e.g.*, 460660K-9001100, N-9001100K, 9001100K-460660 and p2124) in an effort to assess the effect of template length on protein binding and to view the regions of interest on both strands of

the DNA template. Also, two different nonspecific competitors (poly dI:dC and poly dA:dT) were used to interpret the nature of the binding interaction. For example, if the protein bound a specific site when poly dI:dC is used as competitor, but the footprint can no longer be detected when poly dA:dT is used, this would indicate that protein binding is facilitated by interactions with A and/or T nucleotides in the binding site.

3.2.1. C BOX REGIONS FOOTPRINTS

3.2.1.1. ENDLABELING FOOTPRINT TECHNIQUE

Using this technique, the footprints detected were small in size, sometimes affecting only a single band. This may be due to an insufficient amount of active purified protein or simply a failure of the technique for this specific promoter. However, even these small footprints can be considered valid because they may reflect the binding sites to which the protein binds tightly.

In Figure 3.3, using increasing times of DNase I digestion (2, 5 and 10 min), clear but small footprints are shown covering the C box region. The regions of protection are especially evident between the C boxes. Another footprint can be seen upstream of the 5'C box, covering stretches of T's (-413 bp). In this experiment, only DdRPA from undifferentiated cells was used with poly dI:dC as nonspecific competitor.

Using DdRPA from differentiated cells, footprints were detected on essentially the same regions. Experiments performed using poly dA:dT and DdRPA from differentiated cells (Figure 3.4, lanes 5 and 6) and dI:dC (Figure 3.5, lanes 1 and 2) appeared to give the same result.

3.2.1.2. PRIMER EXTENSION FOOTPRINTING

After seeing the success of the primer extension technique using crude nuclear extracts, an attempt was made to repeat the DdRPA footprint again using the primer extension technique. Using p2124 as template and poly dA:dT as competitor (Figure 3.6.), the footprint resembled those generated with the endlabeling approach. The region containing and surrounding the C box region was protected with both forms of DdRPA. A footprint on the TAG box region was detected with DdRPA from undifferentiated cells (Figure 3.6., lane 2).

3.2.2. TAG BOXES FOOTPRINT

Interestingly, the only footprint difference observed between DdRPA from differentiated and undifferentiated cells was on the TAG box region. Undifferentiated DdRPA generated a footprint that appeared to bind between the boxes, whereas the differentiated form did not (Figure 3.7)

3.2.3. OTHER FOOTPRINTS WITH DdRPA

Using 9001100K-460660 as the template to target TA box-1 and upstream regions, several footprints were detected. The footprints covered TA box-1, the region between –577 and –572bp (TGTGT) and further upstream (-636bp and around). The result was the same whether poly dA:dT or poly dI:dC was used as competitor (Figure 3.8. and Figure 3.9, respectively). Doubling the template concentration did not eliminate some of these footprints (Figure 3.9, lanes 8,9).

All of the data explained thus far is summarized in Table A. In summary, under these conditions DdRPA binds several sites on the *gp-2* promoter. However, it is possible that DdRPA may be involved in protein-protein interactions that could augment its specificity. This statement is based on the results involving crude nuclear extract in the following sections.

3.3. FOOTPRINTS OF NUCLEAR EXTRACT PROTEINS DURING DEVELOPMENT:

Because the *gp-2* gene is temporally expressed, nuclear extracts from key stages (where *gp-2* expression differs) were prepared and tested for developmentally regulated footprints. The stages of interest were amoeba, aggregation (8h), first finger (16h) and culmination (21-22h) (see sections 1.1, 1.2 for information of the change in expression between the stages). Both footprint techniques were employed again using different templates and competitors as described in the previous section.

Because of the increased sensitivity of the new approach, large footprints were revealed on the C boxes and TAG boxes. These footprints were shown to be developmentally regulated. Aggregation and first finger extracts generated DNase I protected regions, whereas amoeba and culmination did not produce the same kind of footprints.

3.3.1. C BOX REGIONS FOOTPRINT

3.3.1.1. ENDLABELING FOOTPRINT METHOD

The conventional footprint technique with nuclear extract gave a similar type of footprint as it did with DdRPA. Using amoeba and slug (17h) extract and 460660K-9001100 as template, footprints were observed with slug but not with amoeba. These footprints map to several regions starting at –338 bp, and progressing through the 5'C box region (shown in detail in Figure 3.5, lanes 3,4). This experiment used poly dI:dC as competitor, however, poly dA:dT resulted in the same regions of protection (Figure 3.4, lanes 1,2,3,4).

Using the same approach, extracts from different developmental time points were also assayed. In Figure 3.10, extracts from aggregation, first finger and culmination were reacted with the 460660K-9001100 template and poly dI:dC was used as a nonspecific competitor. The same pattern of protection was observed. In this experiment, no difference between the extracts was detected.

3.3.1.2. PRIMER EXTENSION FOOTPRINTING METHOD

Using the same extracts as described in the previous section and the p2124 template, developmentally regulated footprints were observed. The protected region extended over a very large area, stretching from the transcription start site (-286 bp) to around –426 bp with the 3'C and 5'C boxes affected the most. Figure 3.11 shows a developmentally regulated footprint. The 9001100 primer was used in order to visualize the C box region. The TAG box region can also be observed using this primer (see TAG box region footprint in section 3.2.2.2.). Extracts from amoeba cells grown at low-density did not generate the same footprint as did aggregation, first finger and culmination extracts. In comparing the no-protein lane with amoebae low-density extract and culmination extract lanes, these extracts can be considered to have given some small footprints. The footprints of aggregation and first finger extracts are expanded and different.

When extracts derived from amoebae grown at high density were used, footprints similar to those produced by aggregation and first finger extracts were detected. This is because *gp-2* is induced when amoebae cells reach high density. One such result is shown in Figure 3.12. The competitor used was poly dI:dC. Except for culmination, all other extracts produced footprints on the regions mentioned above.

In another experiment using p2124 as template and poly dI:dC as competitor, first finger extract was assayed using increasing amounts of extract from 1 ul to 8.5 ul. The same footprints were observed. Increasing the protein concentration in the reactions did not cause an expansion of protection (Figure 3.13).

In some cases, culmination stage generated more expansive footprints (Figures 3.15 and 3.16). These experiments used p2124 as template, 9001100 primer and poly dA:dT.

3.3.2. TAG BOX REGIONS FOOTPRINTS

3.3.2.1. ENDLABELING FOOTPRINT TECHNIQUE

Using N-9001100K as the template and poly dI:dC as competitor, nuclear extracts from amoebae, 13h (early slug) and 17h (late slug) were assayed. A footprint on the TAG box region was observed for all stages examined (Figure 3.14), but 13h and 17h footprints were larger than amoebae footprints. The site of greatest protection lies between the two TAG boxes (-447 to –451; TAAAC). This region seems to be the center of the interaction. In amoebae (lane 2), only this region was protected and may involve the edges of both TAG boxes, however, the entire region of both boxes seem to be protected using 13h and 17h extracts.

3.3.2.2. PRIMER EXTENSION FOOTPRINTING METHOD

Different competitor conditions did not influence the footprint results except in one situation. Interestingly, this exception turned out to be on the TAG box region. As explained in 3.3.1.2, the C box regions were always protected, regardless of what competitor was used. For the TAG boxes, however, a new footprint was observed when poly dA:dT was employed as the competitor (Figures 3.15, 3.16 and 3.11). This region was not protected if poly dI:dC was used (Figures.3.11, 3.12 and 3.13). Moreover, the footprint was present only with first finger extract.

This result can be very important if we consider that *gp-2* expression peaks at first finger stage. It may be a starting point that will lead us to new interesting information about the expression of the gene. Because only poly dI:dC competed for the binding activity, it is likely that the protein interacts with GC regions. The TAG box region contains 1 C and 4

G's (TAAAAATGGA TAAAC TAAAAATGGA). If we consider the AT-rich situation in this organism, a protein preferring to bind GC regions should have significance. To be precise, this protein may provide the specificity needed for selectively expressing *gp-2* (see discussion for more explanation).

In almost all attempts, aggregation and first finger footprints gave similar footprints in the presence of poly dI:dC. In Figure 3.11, the N primer was employed using p2124 as template to view the TAG boxes in the downstream direction. As seen in this figure, there is no difference between these two stages. The 9001100 primer gave the same results (Figure 3.12, 3.13). However, when poly dA:dT is used as competitor (Figures 3.15, 3.16 and 3.11), a footprint corresponding to the TAG boxes appeared with first finger extract. The difference between Figure 3.15 and Figure 3.16 is the pattern presented in the no-protein lanes. Specifically, a different banding pattern was observed upstream of the TAG boxes, possibly due to small inconsistencies between the time of digestion or the concentration of DNase I used between experiments. However, the protection was consistent with the two experiments.

3.3.3. OTHER FOOTPRINTS WITH NUCLEAR EXTRACTS

Using the endlabeling technique, TA box-1 and the surrounding regions were targeted for inspection by using the template 9001100K-460660. Only amoeba, 13h (early slug) and 17h (late slug) extracts were applied in this case. The same footprints were observed (as explained in section 3.2.3) in the presence of either nonspecific competitor poly dA:dT or poly dI:dC (Figures 3.8 and 3.9, respectively).

The data for the footprints on the C box and TAG box regions are summarized in Table B.

4. DISCUSSION

Because *gp-2* plays a vital role in differentiation, and because the promoter of this gene possesses a profound AT-rich bias, understanding the expression of *gp-2* may not only uncover novel mechanisms for cell differentiation, but it may also give insight into mechanisms for generating specificity in gene expression. *Dictyostelium discoideum* is considered an intermediate stage between single celled and multiple celled organisms, between the plant and animal kingdoms. Comprehending how gene expression is achieved may, therefore, provide significant clues to understanding transcription as an evolutionary phenomenon.

In order to study gene expression, the first required step is to identify the *cis*-acting elements of the gene. Thus, this thesis has been a part of this effort, by determining the binding affinity of the C and TAG box regions using various protein sources as described. Once this is accomplished, the data can then be used to isolate crucial transcription factors of potentially global significance for cell differentiation.

ARE C BOX AND TAG BOX REGIONS *CIS*-ACTING ELEMENTS OF THE *GP-2* GENE IN *DICTYOSTELIUM DISCOIDEUM* ?

It has been reported that the ACCCACT sequence element acts as regulatory sites in the *Dictyostelium* prespore genes SP40**,** SP70, SP96 and Dp87. In our laboratory, 5' deletion analysis (Sucic *et al.,* 1993) of the promoter has shown: 1) deletion of the first TAG box and about 20 nucleotides upstream resulted in a 100-fold decrease in the expression of luciferase reporter gene; 2) deletion of the C box region resulted in another 100-fold decrease in expression (see Figure 1.4). Site directed mutagenesis (McCaffery et al., manuscript in preparation) has also shown that these two regions are transcriptionally relevant. Mutation of the C boxes resulted in a substantial drop in expression of the

reporter gene and an inability to respond to cAMP. Similarly, TAG box mutation resulted in a significant decrease in luciferase activity, but less than that produced by the C box mutations.

Gel shift experiments using probes containing the C and/or TAG boxes showed that proteins can bind these regions with sufficient specificity for them to be considered potential *cis*-elements (Rutherford *et al.*, 1997). Indeed, DNase I hypersensitive analyses indicated that the region containing the C and TAG boxes becomes bound by *trans*-acting factors upon differentiation (Favis *et al.*, submitted).

Finally, footprint data confirms the importance of these boxes (explained specifically in Results and below). The C and TAG box regions are, therefore, likely to be *cis*-acting elements involved in regulating the *gp-2* gene in *Dictyostelium discoideum.*

THE RELIABILTY OF THE NEW APPROACH (New Apex-Derived In-vitro Reaction)

DNase I can not efficiently cut AT-rich sequences (Drew *et al.*, 1984). This enzyme cuts in the 12 A^{o} minor groove of B-DNA (Gross *et al.*, 1988). AT-rich sequence possess a much narrower minor groove, hence this enzyme will be inhibited from cutting these sites efficiently. This inability of DNase I to cleave highly biased tracts of DNA can cause misinterpretations of footprint patterns, since it is difficult to distinguish between non-occupied AT-rich regions in no-protein reactions and occupied sites in reactions containing protein. The new technique eliminates this inability with the application of primer extension. That is, even though AT-rich sequences are cut poorly in both techniques, the new technique will be more successful in terms of detecting resultant fragments from these sites. This is because several cycles of primer extension will boost the signal given by these poorly cut regions. The conventional method, on the other hand, is not able to overcome the disadvantage as it directly reflects the result of the

original number of cleaved end-labeled fragments. Endlabeled fragments yield a very weak signal on these sites.

One question can, however, be asked: how correct and reliable are the results obtained from this technique?

First, all footprints resulting from either the end-labeling technique or primer extension footprinting map approximately to the same regions. Certainly, the new technique yielded more expanded footprints covering much larger areas due to the reasons cited above (also see material and methods). However, it can be pointed out again that the sum of the results is consistent with the two techniques. The footprints generated by the end-labeling technique may represent the sites that are the center of interactions, to which proteins bind more tightly.

Second, the agreement of the two techniques can be confirmed by comparison to data obtained from a different organism, yeast. The promoters of this organism contain a more balanced base composition. Therefore, both or either approaches can be successfully applied in this organism. DMS (Dimethyl sulfate) footprint results in this organism were similar to each other using either indirect labeling (Church *et al.*, 1984) or primer extension (Axelrod *et al.*, 1989).

Third, large footprints have been reported in both *in vivo* and *in vitro* analyses in bacteria, plants, insects and mammals (Brooks *et al*., 1991; Gilmore *et al*., 1990, Hauffe *et al.*, 1991; Strautch *et al.*, 1989). The proteins causing these footprints have been found to be multimers of single protein and/or multi-protein complexes.

Finally, when the coding region of *gp-2* is used as a control in primer extension footprinting, no footprints was detected (Favis, unpublished data). Also, as explained above, the data from all other experiments support the footprint results.

Therefore, the footprint results are unlikely to be artifacts generated by the primer extension technique. It is likely that the new technique divulges the real *cis*-acting elements.

THE POSSIBLE ROLE OF REPLICATION PROTEIN A (DdRPA)

We hypothesize that DdRPA is involved in DNA replication during cell proliferation and in transcriptional regulation during cell differentiation in *Dictyostelium discoideum*. The protein containing a 62 kDa subunit may act in replication, while the 84kDa isoform may play a role in transcription during differentiation. The DNA binding function of DdRPA resides in this large subunit of the protein (Wen *et al.*, submitted).

In gel shift analyses, DdRPA can bind to the TAG and C box regions, discussed above as likely *cis*-elements of the *gp-2* gene. While purifying this protein, a pH dependent *in vitro* conversion was discovered. At pH 7.0 and 7.5, the 84 kDa was observed. When the pH was decreased to 6.0, this subunit was shifted to 62 kDa form and the change was irreversible (Rutherford *et al*., submitted). It is likely that the 84 kDa subunit was cleaved by an unidentified protease in a pH dependent manner.

Other studies have suggested that the pH of the cytoplasm increases upon cell differentiation of *Dictyostelium discoideum*. Perhaps a general increase in cytoplasmic pH inhibits pH-dependent protease. DdRPA containing the 84 kDa subunit may then be translocated into to the nucleus where it can act as a transcription factor.

Western data indicated that DdRPA becomes enriched in the nucleus of differentiated cells (Favis, unpublished data). In addition, using southwestern analyses, no DdRPA was found in the nucleus of undifferentiated cells (amoebae), but the nucleus of differentiated cells (slug) contained DdRPA. Therefore, this data implies that the 84 kDa subunit contains a signal responsible for translocating the protein to the nucleus, whereas the 62 kDa subunit does not. It is likely that the additional 22kDa (84kDa-62kDa) provides the necessary signal to pass through the nuclear pore complex. Perhaps DdRPA containing

the 62 kDa subunit can only enter the nucleus following breakdown of the nuclear envelope during prophase of mitosis and then acts as a replication protein.

The footprint results of this thesis reveal that DdRPA can bind to the C and TAG box regions. There is no dramatic difference between the binding of the two forms to the C boxes; however, a difference in binding to the TAG box region was discovered. The undifferentiated form (62 kDa) produced a footprint, while the differentiated form (84 kDa) did not.

I propose that the C box and TAG boxes are the sites that undifferentiated DdRPA binds to achieve replication in proliferating cells. However, differentiated DdRPA may bind to the C box regions and not the TAG box region in the process of transcription during cell differentiation. There is possibly another factor that competes for the TAG box region during transcription (see nuclear extract footprint, also southwestern data from Favis *et al.*, submitted).

With regard to the small footprints obtained from DdRPA, it is likely that purified DdRPA can not produce large footprints. Protein-protein interactions may be required to generate sufficient specificity in binding. Because regions surrounding the C box regions also footprint, this is a likely possibility. In fact, there is evidence from other organisms that the second subunit (35kDa) of DdRPA is involved in protein-protein interactions. Thus far, the function of the smallest subunit (18kDa) remains uncharacterized in all systems in which it has been studied.

One last item to consider is that the level of DdRPA message is constant throughout development (Wen *et al.*, 1997). Thus, it is possible that in the process of differentiation, the change in intracellular pH results in a modified isoform of a protein that is already present in the cell and playing a role in DNA replication. This protein (DdRPA) can then be re-employed to participate in a different task, cell differentiation (see model in Figure 4.1).

Future experiments that need to be done to clarify the role of DdRPA involve ectopic expression of wild type of mutant forms of the protein, and/or knocking out the gene, and overexpressing the protein in the organism.

NUCLEAR EXTRACT FOOTPRINTS

As described in the Results section, developmentally regulated footprints have been observed spanning the C box and TAG box regions. These results are summarized in Table B. Specifically, the region from the transcription start site to the TAG box region was protected using high density amoebae extracts, aggregation, first finger and sometimes culmination extracts. The pattern of the protection was the same using either poly dI:dC or poly dA:dT as competitor. Only in one case was a difference in DNase I protection observed. When poly dA:dT and first finger extract are used in a footprint reaction, a footprint covering the region of the TAG boxes is observed. This protein is very likely a GC binding protein, as no footprint was detected on the same region using poly dI:dC as the competitor. This protein will be referred as Xc in the proposed model presented in the next section.

PROPOSED MODEL FOR EXPRESSION OF THE *GP-2* GENE

Based on the results from this thesis and others obtained in our laboratory, a model has been proposed to further investigate the control of *gp-2* expression. The large footprints detected may imply the involvement of several protein-protein and protein-DNA interactions. These interactions are necessary in order to overcome the lack of specificity provided the by AT-rich situation. Specifically, this may prevent the wrong genes from being expressed during development by increasing the requirements for specificity. DdRPA is assumed as one of the possible transcription factors and will be referred as TF-1.

Another protein has recently been identified as binding to the region from –394 to –357 bp (contains 5’ C box). This protein produces 0.39 Rf band on gel shift analysis and is

different from DdRPA with respect to size and subunit composition (Rutherford, unpublished). The protein showed enough specificity on the region described to be a candidate for a transcription factor. It migrates as a 58kDa peptide on SDS-PAGE and forms a dimer under non-denaturing conditions. We will assume this is a transcription factor and refer to it as TF-2.

Gel shift assays have revealed some peptides binding to A and/or T tracts in the promoter. One of these binds between two C boxes, 10 adenine residues from –348 to –357. We think this protein probably enhances the bending of DNA at this region, because oligo(dA) tracts have been implicated in causing bending in DNA (Wu *et al.*, 1984). I will name this protein factor Xa. Another protein, Xb, binds to the region of stretches of T's from –413 to –435bp. If DNA bends at the region of A's with factor Xa, it may cause a loop to form, bringing the two C boxes into contact.

Southwestern analyses using several different probes have detected six proteins that bind specifically to various regions of the promoter (Favis *et al.*, submitted). Four of the proteins are apparent during differentiation: 83 kDa, 77 kDa, 92 kDa and 62 kDa. The probe containing the transcription start site detected the 83 kDa, 77 kDa and 92 kDa binding activities. Western data suggests that the 92 kDa protein is the G box binding protein (GBF), a known transcription factor in the organism. A probe containing the two C boxes detected the 62kDa activity, which may possibly be DdRPA. Lastly, a probe containing the TAG boxes and some upstream sequence detected the 77 kDa and 83 kDa proteins. It is interesting that the 83 and 77 kDa proteins can bind the two probes that contain suspected regions of importance, specifically the transcription start site and the TAG boxes. It should be noted that this data may not reveal all of the proteins binding to the regions, because the technique will only detect the proteins that retain the binding activity upon renaturation as monomers.

It has been shown that extracellular cAMP induces *gp-2* expression (Rutherford *et al.*, 1992). Without cAMP, starved cells express low levels of *gp-2*. In the promoter, the region from –247 to –317 and the region from –461 to –357 are required for cAMP

responsive transcription (Sucic *et al.*, 1993). Also, site-directed mutagenesis of the C Box and TAG box regions have then found to be cAMP responsive (McCaffery, manuscript in preparation).

Given all this data, the scenario can be described as follows (see model in Figure 4.1): DdRPA is always synthesized with an 84 kDa large subunit that is cleaved to 62 kDa when it acts as a replication protein in proliferating cells. When differentiation commences, the 84 kDa subunit is not cleaved (as explained above) and can act as TF-1 in differentiating cells by binding to the region from –371 to –323bp containing the 3' C box. TF-2 binds to the region from –357 to -394 containing the 5' C box. With the involvement of DNA bending factor Xa, the interactions between these 3 factors stabilizes a loop. This situation will be referred to as the "first triangle". Following the formation of this first triangle, the *gp-2* gene can be expressed at some sub-maximal level. This may be the situation in aggregating cells. As differentiation continues, the other factor (Xc, a GC binding activity) binds specifically to the TAG box region and perhaps to the transcription start site (however this activity may be a different protein). Upon binding of the other T binding protein (Xb), interactions between the proteins cause a second triangle to form that eventually results in a peak in *gp-2* expression. The regions that are bound by Xc, TF-1 and TF-2 are cAMP responsive. Interaction between these two triangles can be speculated to accelerate transcription. In deletion analysis, when the region containing one of the TAG boxes was deleted a 100-fold decrease was seen in the expression of luciferase (deletion 32, -449bp). However, the remaining region could continue to support expression of the reporter gene. In other words, the gene could be transcribed even though the TAG boxes were removed. This may imply that the TAG boxes act as an enhancer to increase the expression of the gene in first finger cells. (If the TAG boxes are removed, the second triangle is disrupted, but first one is still intact to function).

Site directed mutagenesis of the *gp-2* promoter showed that that mutating the C boxes resulted in a greater loss of expression of the luciferase gene than did mutating TAG boxes. This may be viewed as evidence supporting the model explained above. The First

triangle situation is required to express the gene and the second triangle to enhance expression. The protein binding the region may be of significance, because it is seen in a stage-dependent manner and bound to the TAG boxes.

In summary, if a loop is formed by the assembly of the first triangle, there is expression of *gp-2* (*i.e.*, loop but no enhancer; loop not complete). On the other hand, when Xc and Xd proteins participate, the second triangle situation arises and perhaps the interaction between the two triangles completes the loop in a cAMP-responsive manner. If there is no loop formed, the gene will be expressed at very low levels.

Figure 4.2 shows the interactions of the potential *trans*-acting factors during different stages of development. In amoebae, there is no differentiating signal, hence there are no factors and *gp-2* is not transcribed. I hypothesize Xa and Xb proteins are present in these cells, and because they are A and T binding proteins, they may not be considered to provide the specificity individually because of the high possibility of binding nonspecifically in other regions of the genome due to the AT-rich situation. They will, on the other hand, be involved in transcription to aid in the formation of the loop. In aggregation cells, cAMP is synthesized upon starvation and cell-cell contacts are made; cells turn on and off crucial genes as they respond to cAMP and other signals. TF-1 is synthesized and postranslationally modified to act as a transcription factor and is directed into the nucleus. These factors form the loop described and *gp-2* is expressed. In first finger stage cells, *gp-2* expression continues; however, two additional members (Xc and Xd) participate in the situation to increase expression. In culminating cells, the interactions between proteins are removed to obstruct formation of the triangles and the loop is destroyed. Then, *gp-2* expression is lost.

According to this model, the problem of the lack of specificity can be resolved by providing; 1) specific GC binding proteins (TF-1 [DdRPA], TF-2, Xc and Xd) and 2) a looping formation assembled by the proteins Xa and Xb. In this situation, GC binding proteins will create the loop with the help of A, T and AT binding proteins.

Future studies will involve purification of these proteins and mutational analyses in order to test this hypothesis.

LITERATURE CITED

Aerts, R.J., Durston, A.J. and Moolenaar, W.H., 1985, Cytoplasmic pH and the regulation of the *Dictyostelium* cell cycle, *Cell*, 43, 653-657.

Aerts, R.J., de Wit, R.J.W. and van Lookeren Campagne, M.M., 1987, Cyclic AMP induces a transient alkanization in *Dictyostelium, FEBS Lett.*, 220, 366-370.

Axelrod, J.D. and Majors, J. 1989. An improved method for photofootprinting yeast genes in vivo using Taq polymerase. *Nucl. Acids Res*. 17, 171-183.

Ausubel F.M., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A., and K. Struhl, 1995 ed. Current Protocols in Molecular Biology 2nd ed Vol. 1,2 and 3. John Wiley& Sons, Inc.

Brickey, D.A., Naranan, V., Sucic, J.F. and Rutherford, C.L., 1990, Regulation of the two forms of glycogen phosphorylase by cAMP and its analogs in *Dictyostelium discoideum, Mol. Cell. Biochem.*, 97, 17-33.

Brickey, D.A. and C.L. Rutherford. (1987). The effect of cAMP and its analogs on the regulation of the two forms of glycogen phosphorylase in *Dictyostelium discoideum*. *Fed. Proc.* 462014.

Brickey, D.A., *et al.,* (1990). Regulation of the two forms of glycogen phosphorylase by cAMP and its analogs in *Dictyostelium discoideum*. Mol. Cell. Biochem. 9717-9733.

Brooks, A.R., Blackhart, B.D., Haubold, K. and Levy-Wilson, B. 1991. Characterization of tissue-specific enhancer elements in the second intron of the human apolipoprotein B gene. *J. Biol. Chem.* 266: 7848-7859

Browitz, M., Senear, D.F. and Kingston, R.E. (1995) DNase I footprint analysis of protein DNA binding. In Current protocols in Molecular Biology (F.A. Ausubel, R. Brent, R.E., Kingston, *Dictyostelium discoideum* . Moore, J.G., Seidman,J.A. Smith, And K. Struhl, eds.) pp. 12.4.1-12.4.6 John Wiley & Sons, Newyork

Cartwright, I.L. and Kelly, S.E. 1991. Probing the nature of chromosomal DNA-protein contacts by in vivo footprinting. *Biotechniques* 11: 188-203.

Church, G.M. and Gilbert, W. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81:

1991-1995.

Cloutier, M.J. and Rutherford, C.L., 1987, Glycogen phosphorylase in *Dictyostelium*. Developmental regulation of two forms and their physical and kinetic properties, J. Biol. Chem., 262, 9486-9493

Cohen, S.M., Knecht, D., Lodish, H.F. and Loomis, W.F. 1986. DNA sequence required for expression of a *Dictyostelium* actin gene. *EMBO J* 5: 3361-3366.

Drew, H.R. and Travers, A.A. 1984. DNA structural variations in the E. coli *tyrT* promoter. *Cell* 37: 491-502.

Echols, H. 1986. Multiple DNA-protein interactions governing high-precision DNA transactions. *Science* 233: 1050-1056.

Favis, R.L., J. Littman, and C.L. Rutherford. cAMP Induction of glycogen phosphorylase 2 is Regulated by both De-repression and Activation by a C box Factor. Manuscript in Preparation. .

Garreau, H. and Williams, J.G. 1983. Two nuclear DNA binding proteins of *Dictyostelium discoideum* with a high affinity for poly(dA)-poly(dT). *Nucl. Acids Res.* 11: 8473-8484.

Gilmore, D.S., Dietz, D.J., Elgin, S.C.R. 1990. UV cross-linking identifies four polypeptides that require the TATA box to bind to the *Drosophila* hsp70 promoter. *Molecular and Cellular Biology* 10: 4233-4238.

Gralla, J.D. 1985. Rapid "footprinting" on supercoiled DNA. *Proc. Natl. Acad. Sci. USA* 82: 3078-3081.

Gross, D.S. and Garrard, W.T. 1988. Nuclease hypersensitive sites in chromatin. *Ann. Rev. Biochem.* 57: 159-97.

Gross J.D., 1994, Developmental decisions in *Dictyostelium discoideum*. Am. Soc. Microbiology, vol 58, no3, p.330-351

Grosschedl, R. 1995. Higher-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr. Opin. Cell Biol.* 7: 362-370.

Gustafson, G.L. and B.E. Wrigth. 1972, Analysis of approaches used in studying differentiation of the cellular slime mold. *Crit. Rev. Microbiol.,* 1:453-478

Harwood, A.J, A.E. Early, K.A. Jermyn, and J. Williams. 1992, Unexpected localization of cells expressing a prespore marker of *Dictyostelium discoideum*. *Differention* 46:713

Hauffe, K.D., Paszkowski, U., Schulze-Lefert, P., Hahlbrock, K., Dangl, J.L. and Douglas, C.J. 1991. A parsley 4CL-1 prmoter fragment specifies complex expression patterns in transgenic tobacco. *Plant Cell* 3: 435-443.

Hjorth, A.L., Pears, C., Williams, J.G. and Firtel, R.A. 1990. A developmentally regulated trans-acting factor recognizes dissmiliar G/C-rich elements controlling a class of cAMP-inducible *Dictyostelium* genes. *Genes & Dev.* 4: 419-432.

Hori, R. and Firtel, R.A. 1994. Identification and characterization of multiple A/T-rich cis acting elements that control expression from Dcityostelium actin promoters: the *Dictyostelium* actin upstream activating sequence confers growth phase Expression and has enhancer-like properties. *Nucl. Acids Res.* 22: 5099-5111.

Kadonaga, J.T.a.T., R., 1986, Affinity purification of sequence-specific DNA binding proteins., *Proc. Natl. Acad. Sci.* U.S.A., 83, 5889-5893.

Latchman, D.S. 1993, Transcription Factors, A Practical Approach. New York, Oxford University Press Inc.

Kampang, P. 1995. Purification of the proteins that bind to the gp-2 promoter in *Dictyostelium discoideum*. In Virginia Polytechnic Institue and State University.

Kelly, S.E. and Cartwright, I.L. 1989. Genomic footprinting by an amplified primer extension (APEX) technique. *J. Cell. Biochem. Supp.* 13E: 278.

Kimmel, A.R. and Firtel, R.A. 1982. The Organization and Expression of the *Dictyostelium* Genome. The Development of *Dictyostelium discoideum*. Loomis, W.F. 233-324. Academic Press, New York.

Konijn, T.M., van de Meene, J.G.C., Bonner, J.T. and Barkley, D.S., 1967, The acrasin activity of adenosine-3',5'-cyclic phosphate, Proc. Natl. Acad. Sci. USA, 58, 1152 1154.

Lodish, H.F., Blumberg, D.D., Chisholm, R., Chung, S., Coloma, A., Landfear, S., Barklis, E., Lefebvre, P., Zuker, C. and Mangiarotti, G. 1982. Control of Gene Expression. The Development of *Dictystelium discoideum*. Loomis, W.F. 325-352. Academic Press, New York.

Loomis, W.F. 1982 The Development of *Dictyostelium discoideum*, Academic Press, New York

Loomis, W.F., 1988, Signals that regulate differentiation in *Dictyostelium*, 25-30. *In* ISI atlas of science: Immunology., Inst. Scientific Inform.: Philadelphia. p. 25-30.

Longhese, M.P., Plevani, P., and Lucchini, G., 1994, Replication Factor-A is Required *In Vivo* for DNA Replication, Repair and Recombination., *Mol. Cell. Biol*., 14, Dec, 7884-7890.

Marshall, R., D. Sargent, and B.E. Wright 1970, Glycogen turnover in *Dictyostelium discoideum. Biochemistry*, 9:3087-3094

Naranan, V., Sucic, J., Brickey, D.A. and Rutherford, C.L., 1988, The relationship between two forms of glycogen phosphorylase in *Dictyostelium discoideum*, *Differentiation*, 38, 1-10

Qiagen 1992 Qiagen plasmid handbook for plasmid maxi kti, and plasmid mega kit. Diagen GmbH, Qiagen Inc. 14-17.

Raper, K.B., 1935, *Dictyostelium discoideum*, a new species of slime mold from decaying forest leaves, *J. Agr. Res*., 50, 135-147.

Rogers, P.V., Luo Shun, Sucic, J.F. and Rutherford, C.L. 1992. Characterization and cloning of glycogen phosphorylase 1 from *Dictyostelium discoideum*. *Biochimica et Biophysica Acta.* 1129: 262-272.

Rogers, P.V., Sucic, J.F., Yin, Y.Z. and Rutherford, C.L., 1994, Disruption of glycogen phosphorylase gene expression in *Dictyostelium* - evidence for altered glycogen metabolism and developmental coregulation of the gene products, *Differentiation,* 56, 1-12.

Rutherford, C.L., Naranan, V., Brickey, D. and Cloutier, M.J., 1986, Identification of two forms of glycogen phosphorylase during cellular differentiation of *Dictyostelium discoideum*, *Fed. Proc*., 45, 1511.

Rutherford, C.L., Naranan, V., Brickey, D.A., Sucic, J.F., Rogers, P.V. and Selmin, O., 1988, Glycogen phosphorylase in *Dictyostelium discoideum*: demonstration of two developmentally regulated forms, purification to homogeneity, immunochemical analysis cAMP induction, in vitro translation, and molecular cloning, *Dev. Genet*., 9, 469-482.

Rutherford, C.L., Peery, R.B., Sucic, J.F., Yin, Y., Rogers, P.V., Luo, S. and Selmin, O. 1992. Cloning, structural analysis and expression of the glycogen phosphorylase-2 gene in *Dictyostelium*. *J. Biol. Chem.* 267: 2294-2302.

Rutherford, C.L., *et al.*, 1982, Cellular pattern formation: *Dictyostelium discoideum* as a system for a biochemical approach. *TIBS*. 7108-111.

Rutherford, C.L., Selmin, O. and Peters-Weigel, S. 1997. Temporal regulation of the *Dictyostelium* glycogen phosphorylase 2 gene. *Biochem. Biophys. Acta*. 1351: 111-125

Saluz, H. and Jost, J-P. 1989. A simple high-resolution procedure to study DNA methylation and in vivo DNA-protein interactions on a single-copy gene level in higher eukaryotes. *Proc. Natl. Acad. Sci. USA* 86: 2602-2606.

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

Schnitzler, G.R., Fischer, W.H and Firtel, R.A. 1994. Cloning and characterization of the G-box binding factor, an essential component of the developmental switch between early and late development in *Dictyostelium*. *Genes & Dev.* 8: 502-514.

Seeman, N.C., Rosenberg, J.M. and Rich, A. 1976. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Nat. Acad. Sci. USA* 73: 804-808.

Sucic, J.F., O. Selmin, and C.L. Rutherford. 1993, Regulation of the *Dictyostelium* glycogen phosphorylase 2 gene by cyclic AMP. *Dev. Genet*. 14313-322.

Strauch, MA, Spiegelman, GB, Perego, M, Johnson, WC, Burbulys, D and Hoch, JA. 1989. The transition state transcription regulator *abrB* of *Bacillus subtilis* is a DNA binding protein. *EMBO J* 8: 1615-1621.

Sucic, J. F., Selmin, O. and Rutherford, C. L. 1993. Regulation of the *Dictyostelium* glycogen phosphorylase 2 gene by cyclic AMP. *Dev. Genet.* 14: 313-322.

Sussmann, M. and R. Brackenbury. 1976, Biochemical and molecular genetics aspects of cellular slime mold development. *Ann. Rev. Plant Physiol*., 27:229-265

Van Duijn, B. and Inouye, K., 1991, Regulation of movement speed by intracellular pH during *Dictyostelium discoideum* chemotaxis, *Proc. Natl. Acad. Sci*. USA, 88, 4951-4955.

Wahls, W.P., Swenson, G. and Moore, P.D. 1991. Two hypervariable minisatellite DNA binding proteins. *Nucl. Acids Res.* 19: 3269-3274.

Wen, X., Khampang, P. and Rutherford, C.L. 1997. The glycogen phosphorylase-2 promoter binding protein in *Dictyostelium* is replication protein A. Submitted to *J. Biol. Chem.*

Werner, M.H., Gronenborn, A.M. and Clore, G.M. 1996. Intercalation, DNA kinking and the control of transcription. *Science* 271: 778-784.

Williamson, B.D., Favis, R., Birckey, D.A. and Rutherford, C.L. 1996. Isolation and characterization of the glycogen synthase gene in *Dictyostelium discoideum*. *Dev. Genetics* 19: 350-364.

Wright, B.E., *et al.*, 1973. Glycogen metabolism during differentiation in *Dictyostelium discoideum. Ann. N. Y. Acad. Sci*. 21051-63.

Wu, H.M. and Crothers, D.M. 1984. The locus of sequence-directed and protein-induced DNA bending. *Nature* 308: 509-513.

Yin, Y.H., P.V. Rogers, and C.L. Rutherford. 1994. Dual regulation of the glycogen phosphorylase 2 gene of *Dictyostelium discoideum*: The effects of DIF-1, cAMP, NH3 and adenosine. *Development* 1201169-1178.

.

CURRICULUM VITAE
December 1997

BEKIR COL

Biology Department
2029 Derring Hall
Virginia Polytechnic University & State University
Blacksburg, VA 24061
Telephone: Home (540) 953 0743
Office (540) 231 8940
e-mail: bcol@vt.edu

Personal Data:

Date of Birth : March 6, 1972
Place of Birth: Mugla, Turkey
Marital status : Single

Education:

B.Sc. 1989-1993
Hacettepe University, Science Faculty, Department of Biology, Ankara, Turkey

M.S. Candidate in Molecular & Cellular Biology Program in Biology Department
1996-present
Virginia Polytechnic Institute & State University, (VPI&SU), USA.

Working & teaching experience:

General Biology Laboratory I, Graduate teaching assistant at Virginia Tech 1997

Presentations:

Departmental Seminar, Oct 1997. Footprint analysis of the transcriptional control of glycogen phosphorylase 2 gene in *Dictyostelium discoideum*

Terminal Seminar, Dec 1997 same title

Abstract:

Virginia Academy of Science, 1997; Recognition site DNA screening of lambda GT-11 expression libraries of various developmental stages of *Dictyostelium* . C. Chanchoa, B.Col, R. Favis and C.L. Rutherford, Dept. of Biology Va. Polytechnic Ins. State Univ.

13th Annual Research symposium Virginia Tech., 1997, Screening of lambda GT-11 expression libraries of *Dictyostelium discoideum* at amoeba, 8h and 12h of development with recognition site DNA. Chanpen Chanchoa, Bekir Col, Reyna Favis, Charles L. Rutherford

Honors:

Passed the exam given by Department of Biology at Hacettepe university with the highest grade to be accepted into molecular biology program.

Ranked as "4th" after a national exam in selected 300 selected biology graduate students to qualify for a scholarship.

Awarded a full scholarship by Ministry of Education of Turkish Government for M.S. and Ph.D studies in abroad

Reference:

Dr. Charles L. Rutherford
Professor of Biology
VPI&SU
Blacksburg, Va, 24061
Tel: 540 231 5349 (office)