Engineering Yeast Cytosine Deaminase for Improved Efficacy in Cancer Gene Directed Enzyme Prodrug Therapy

Aaron Korkegian

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2007

Program Authorized to Offer Degree:
Molecular and Cellular Biology

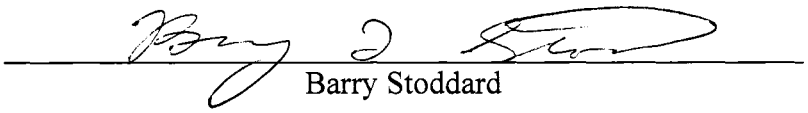
University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Aaron Korkegian

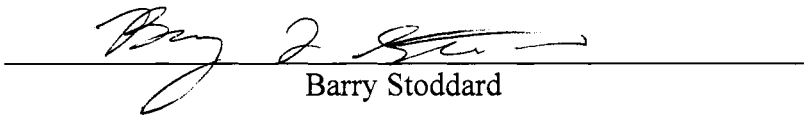
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:



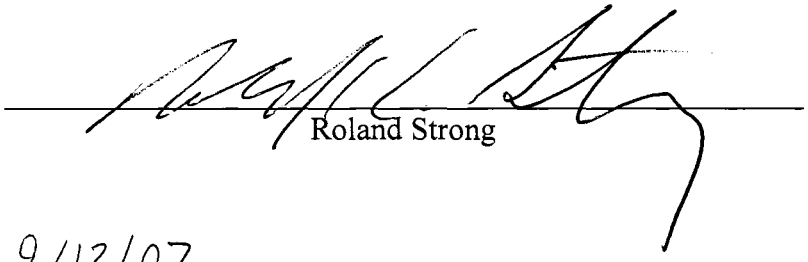
Barry Stoddard

Reading Committee:



Barry Stoddard

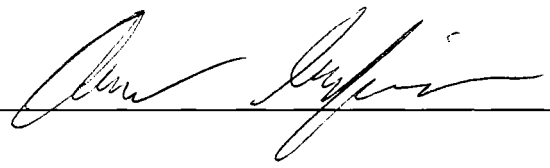
David Baker



Roland Strong

Date: 9/13/07

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 9/17/07

University of Washington

Abstract

Engineering Yeast Cytosine Deaminase for Improved Efficacy in Cancer Gene Directed Enzyme Prodrug Therapy

Aaron Korkegian

Chair of the Supervisory Committee:
Professor Barry Stoddard
Basic Sciences

Conventional treatment of cancer typically involves some combination of surgery coupled with radio- or chemotherapy. Prognosis in cases of tumor metastasis or inoperable solid-tumors remains grim as cancer cells are often refractory to the DNA damage caused by these treatments due to the loss of cell cycle regulatory proteins during oncogenesis. Higher dosages of either radiation or chemotherapeutic agents can compensate, however, lack of cancer cell specificity and low therapeutic index put strict limitations on dosages that can be administered safely.

Growing understanding of oncogenesis and molecular biology has lead to the development of vectors capable of targeting and transducing tumors with high specificity. One of the most successful applications of this technology has been a form of targeted chemotherapy called gene-directed enzyme/prodrug therapy (GDEPT), in which tumor cells are transduced with a suicide gene encoding a non-endogenous enzyme that later metabolically converts a systemically administered prodrug into a potent cytotoxin locally within the tumor. These enzymes are often non-ideal due to

both lack of stability and low catalytic efficiency. Protein engineering methods could be used in order to tailor the pharmacokinetic profiles of these enzymes for improved therapeutic use.

Our work focused on the re-engineering the enzyme yeast cytosine deaminase for improved efficacy in GDEPT. Two distinct enzyme engineering strategies were employed: regiospecific random mutagenesis was used in order to select for mutants with increased sensitization to the prodrug 5-fluorocytosine while computational design was used to select thermostabilizing mutations. Both approaches lead to independent mutations that conferred improved pharmacokinetics in *in vivo* mouse tumor xenograph models. Additional study revealed that in both cases the biochemical cause of improvement appeared to be thermostabilization of the enzyme and not an improvement in catalytic efficiency.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express thanks to the Molecular Cellular Biology Department for providing me with a home from which to conduct my study and specifically Michelle Karantsavalos at the Fred Hutchinson Cancer Research Center for going out of her way to help me deal with the graduate school administration. I'd like to express my deepest appreciation to Barry Stoddard for being a wonderful PI, for being supportive of me through good times and bad, for being open to my ideas and for setting up some great collaborations with experts within my field. I would also like to thank and acknowledge members of the Stoddard lab who helped me with my research: Greg Ireton for helping get started on the yCD project; Django Sussman for all his help getting me started with x-ray crystallography and protein expression; Betty Shen for her help with x-ray crystallography; and Jill Boldouc for her dedication to maintaining the lab and for her witty sarcasm which kept me smiling.

I'd like to thank my collaborators Dr. David Baker and Dr. Margaret Black for all their input on aspects of my computational design and gene directed enzyme prodrug therapy projects respectively. I'd like to acknowledge members of the Baker lab who had a direct hand in helping me with RosettaDesign and protein engineering concepts: Brian Kuhlman, the original designer of RosettaDesign, for getting me started with the program; Gautam Dantas for being both a friend and an excellent source of knowledge and experience; Alex Watters for help with use of the CD spec

and being there to bounce ideas off of; and David Callender for his help with the technical aspects of lab instrumentation. I'd also like to thank members of the Black lab, Candice Willmon, Tiffany Stolworthy, Andress Ardiani and Jennifer Cundiff for their dedicated work on the 5-fluorocytosine sensitivity assays both in cancer cell lines and in mouse tumor xenograph models.

Thank you to all of those who provided feedback and expertise on aspects of my projects and just graduate school in general. To the members of my committee Dr. Bill Parson and Dr. Roland Strong who in addition to my PI and collaborators took the time out of their busy schedules to review my project aims and help mold the direction that my project eventually took. To Roland Strong specifically for taking a lot of time to help me draft my first paper. To Dr. Rachel Klevit for her guidance through my TAship and her advice and genuine concern over the path of my graduate career.

Finally I'd like to thank and acknowledge my family, friends and to my loving girlfriend Ashley Sherrid for their support without which I probably would not have made it past my second year.

DEDICATION

I would like to dedicate this dissertation and the work therein to my mother, Rae Armantrout, whose love and guidance has helped me persevere in my pursuit of knowledge and a scientific career. Her current battle with cancer has underscored for me the importance of work being done at the Fred Hutchinson Cancer Research Center and institutions like it.

Chapter One: Introduction

Cancer is not a single disease but rather a variety of somatic genetic diseases brought on by mutations or epigenetic changes, often within oncogenes or tumor suppressor genes. These genetic changes can be caused by almost any form of DNA damage or mutation and it is their effect rather than their cause that defines them as cancer. Gain or loss of function, or disregulation of these genes cause the breakdown of cell cycle control as well as the failure of DNA repair machinery that lead to the cancer hallmarks of uncontrolled cell growth and accumulation of genetic mutations. The inherent variation and heterogeneity within cancer cell populations spawned during carcinogenesis makes it very difficult to develop treatments or drugs with high enough specificity to the highly variable cancer cell population to be therapeutically viable.

Conventional treatments for cancer involve some combination of surgery, radiotherapy and chemotherapy. Surgery provides a way of physically removing all of the cancerous cells in a single step, however this option is only viable in the cases of accessible localized tumors that have not metastasized. There is also an inherent risk to the patient with any form of invasive procedure. Since the physical resection of a tumor is unlikely to remove every last cancer cell, in cases that surgery is an option the patient is typically treated with some form of radio- or chemotherapy in order to reduce the chance of recurrence.

Radiotherapy provides a non-invasive way of eradicating a tumor regardless of its accessibility by causing cell death due to DNA damage. As radiation is potentially harmful to all tissues within the body, radiotherapy can only be given in a localized manner and therefore cannot be used to effectively treat metastases. Since the toxicity is due to DNA damage, cells that are actively replicating are much more susceptible to radiation induced death than quiescent cells giving radiotherapy a small amount of specificity. Although solid tumors often contain replicating cell populations they also contain quiescent cells as well as hypoxic cells within pockets of necrosis, both of which are highly resistant to radiation. Although higher dosages of radiation may surmount this resistance, the lack of specificity introduces the danger of causing secondary malignancy or massive cell death to otherwise healthy tissue.

Chemotherapy provides a way of treating metastases with the systemic delivery of a drug that is cytotoxic to actively replicating cells. This ability to treat both solid tumors and metastases has made chemotherapy a staple in cancer therapy. Unfortunately the non-specificity of chemotherapy drugs coupled with systemic delivery leads to a high degree of adverse effects that limits the dosage that can be safely administered to a patient. In the case of solid tumors poor or unequal vascularization and regions of necrosis can lead to poor uptake of the drug and intratumoral concentrations too low to deliver a therapeutic effect[1]. In addition, the heterogeneity of cancer cells may lead to variance in susceptibility to a given chemotherapeutic drug, leading to resistance.

One of the major downfalls of these conventional cancer treatments is their lack of specificity and as a result their low therapeutic index. Therefore, methods that increase anticancer specificity of chemotherapy agents are highly desirable. At least two strategies for achieving this goal are under active investigation: the development of drugs that specifically target unique cancer genotypes, and the development of targeting strategies that specifically sensitize cancer cells to the effects of chemotherapy agents that otherwise display undesirable toxicities to normal tissues.

Strategies that target specific cancer geneotypes for drug development rely on an understanding of the causal relationship between mutations in that cancer line and the onset of oncogenesis. In the case of chronic myelogenous leukemia (CML) it was discovered that a reciprocal translocation event between the long arms of chromosomes 9 and 22 leads to the fusion of the *c-abl* oncogene on chromosome 9 with sequence from the breakpoint cluster region (bcr) on chromosome 22 resulting in a chimeric *bcr-abl* gene[2]. *bcr-abl* encodes for an expressed 210-kDa protein that acts as a constitutive tyrosine kinase leading to the activation of numerous signaling pathways and the onset of oncogenesis. Screens of drug libraries looking for tyrosine kinase inhibition led to the discovery of the drug STI571, also known as Gleevec, which demonstrated potent inhibition of the Abl tyrosine kinase. Phase I and II clinical trials using Gleevec as a therapy against CML began in 1998 have shown the drug to be very effective[3,4]. The success of Gleevec has made it a model for the treatment of cancer by targeting cancer line specific hyperactive oncogenes[5]. The

drawback to this protein targeted drug development approach is that successful treatment is often limited to one specific type of cancer.

A second promising approach for improving tumor selectivity of drugs is the use of gene therapy vectors as a way of sensitizing cancer cells to their cytotoxic effects. In a variation of gene therapy known as gene-directed enzyme prodrug therapy (GDEPT), or colloquially, suicide gene therapy, a vector is used to transduce a target tissue with a transgene that encodes an exogenous enzyme, which performs a catalytic reaction non-endogenous to the host. Subsequent treatment involves the systemic administering a non-toxic prodrug that is converted by the enzyme locally into an extremely potent cytotoxin within the target tissue. Since transduction efficiency is unlikely to be 100% due to the accessibility and heterogeneity of the cancer cells, to produce a therapeutic effect the cytotoxin must also cause cell death in neighboring cells, in what is known as a bystander effect. By metabolically coupling the production of the cytotoxic drug to the cancer cells targeted for death a feedback mechanism is created that limits the production of drug to only what is necessary to kill the target cells. This mechanism increases the intratumoral concentration of drug while limiting the dosage to only what is required to kill the target cells thereby minimizing harm to nearby healthy tissue and maximizing the therapeutic index.

The promise of GDEPT for cancer therapy has made it the most studied and well-funded form of gene therapy against cancer. Its efficacy as a form of treatment was first tested in 1986 by Moolten *et. al.* who demonstrated that transfection of a

cancer cell line with herpes simplex virus-thymidine kinase and subsequent treatment with the prodrug ganciclovir led to targeted cell death with a high therapeutic index[6]. Since 1986 considerable study and funding has gone into developing GDEPT for use in cancer therapy with around 25 different enzyme/prodrug combinations currently under investigation as of 2007[7] and over 75 registered clinical trials utilizing GDEPT for treatment of cancer as of 2005[8]. Examples of vectors and enzyme/prodrug systems currently being investigated for use in cancer GDEPT are reviewed extensively in chapter two.

With the development of improved targeted vectors against cancer cells, the efficacy of cancer GDEPT increasingly depends on the effectiveness of enzyme/prodrug system used. Enzymes used for GDEPT catalyze non-native reactions in non-native conditions often in the presence of high concentrations of competitive substrate which all act to severely limiting their effectiveness for therapy *in vivo*. This is seen in even the most successful enzyme/prodrug systems such as herpes simplex virus-thymidine kinase (HSV-TK) coupled with ganciclovir (GCV) and yeast cytosine deaminase (yCD) coupled with 5-fluorocytosine (5-FC)[9-12].

Both HSV-TK and yCD convert a non-toxic nucleoside analogue into a potent cytotoxic nucleoside analogue whose cytotoxic effect is elicited both by misincorporation into DNA leading to chain termination and by inhibition of enzymes involved in DNA replication and repair. For therapeutic effect both enzymes are required to catalytically convert their non-native prodrug substrate in the presence of

their native substrate (cytosine for yCD and thymidine for HSV-TK). Competition by these endogenous substrates therefore reduces the conversion rate of the prodrug to the cytotoxic form considerably. In addition yCD has been shown to be unstable in human physiological conditions[13], degrading rapidly leading to a reduction in therapeutic effect. Altering these enzymes in a way that they would be more kinetically selective for the prodrug substrate or more stable in human physiological conditions could improve the therapeutic effect of these treatments.

Protein engineering methods offer us tools to custom tailor these enzymes to perform better within their therapeutic niche. For example, investigators at Washington State University (the Margaret Black lab) improved the properties of HSV-TK by screening enzymes produced from library of sequences where the active site region of the gene had been subjected to random mutagenesis. The enzymes were screened for enhanced conversion of GCV compared to thymidine, the enzyme's native substrate. From the screen they identified mutants whose catalytic activities were 500-fold higher for GCV relative to thymidine when compared to wild-type enzyme[14]. Subsequent tests and clinical trials have shown substantial improvement in the efficacy of the engineered enzyme[11,15]. An in depth review of protein engineering and its potential use for improving therapeutic enzymes is provided in chapter three.

My graduate work has centered on using protein-engineering techniques to enhance the efficacy of yeast cytosine deaminase for use in cancer gene directed enzyme prodrug therapy. Two approaches to design were taken. The first approach,

described in detail in chapter four, was carried out in collaboration with the Baker lab where I used the RosettaDesign program developed in their lab along with my own methodology in order to pick out potentially thermostabilizing mutations within the enzyme without disrupting the enzyme's activity. The second approach, described in detail in chapter five, was carried out in collaboration with the Black lab where we used the same regiospecific random mutagenesis screening technique used with HSV-TK in order to identify mutations in yCD that conferred higher sensitivity to the prodrug 5-fluorocytosine. The results and implications of these two studies are discussed in detail within chapters four, five and six.

Chapter Two: Gene Directed Enzyme Prodrug Therapy

INTRODUCTION

Gene-directed enzyme prodrug therapy (GDEPT), also known as suicide gene therapy, is a two-step treatment (Figure 2.1). The first step involves the tissue-specific delivery of a "suicide gene" that encodes an enzyme with catalytic function exogenous to the patient. Gene delivery is achieved by either systemic or local administration of a vector that can transduce the target tissue with high specificity. In the second step of treatment a nontoxic prodrug is administered systemically. The prodrug is then converted metabolically into a potent cytotoxin locally in the target tissue by the non-endogenous enzyme. This localization of the cytotoxin allows for much higher concentrations of active drug that could be achieved by systemic delivery and therefore a higher therapeutic effect.

Since transduction of the target tissue is unlikely to be one-hundred percent, cytotoxins are chosen that can kill not only the cell in which they are metabolically produced but also induce death of neighboring cells in a manner known as a bystander effect. Subsequent immune response to the cell death adds to the therapeutic effect. The success of a GDEPT treatment is dependent on a number of variables: the choice of vector used to deliver the suicide gene encoding the enzyme; the enzyme/prodrug system chosen; and the subsequent bystander effect.

VECTORS USED IN GDEPT

The success of any cancer GDEPT relies on the ability to target the expression of the therapeutic suicide gene selectively within tumor cells. As our knowledge of viral pathogenesis and oncogenesis expands, so does our ability to manipulate viral vectors for tumoral-targeted transgene expression. A wide variety of viral vectors have been developed for use in gene therapy and choosing the appropriate one for use in cancer GDEPT is dependent on features inherent both in the treatment and in the target. In the end vectors are chosen based on three traits: their ability to transduce or transfect tumor cells with high efficiency; their specificity for tumor cells over normal cells or their therapeutic index; and their inherent toxicity and the safety of their administration to human patients.

Efficient gene transfer is vital for the efficacy of any GDEPT treatment and as a result, vectors are chosen that are capable of transducing or transfecting the largest population of tumor cells possible. As human cancers tend to be a heterogeneous mix of dividing and non-dividing cells, vectors used for cancer therapy should preferably be capable of transducing either replicating or quiescent cells. As GDEPT only requires the transient expression of the suicide gene, stable integration of viral DNA into the cellular genome is not required. As a result, vectors whose genetic material remains episomal are viable and often preferred as they reduce the chance of insertional mutagenesis.

Specific cell targeting can vastly raise the efficacy of any treatment by increasing its therapeutic index. Targeting of cancer cells can be achieved by a variety of methods, including the addition or modification of vector coat proteins to recognize tumor-specific cell surface receptors (cell targeting), addition of tumor-specific promoters to the transgene that allow for high levels of expression within cancer cells but not in normal cells (transcriptional targeting)[16], or, in the case of biological vectors, modification to inhibit replication in normal tissue but allow it in cancer cells (replication-selective targeting).

As with any human therapy, issues of safety are paramount in choosing an appropriate vector for gene delivery. Immunogenecity of the vector should be as low as possible to avoid both reduced therapeutic efficacy due to clearing of the vector as well as toxic shock induced by a systemic humoral immune response. Once target cells have been transduced the expression of the transgene should be strong but short in duration. Although transient gene expression is a problem for gene replacement therapies, in the case of GDEPT short-term expression is sufficient for cell killing and preferable for reasons of safety. Replicating vectors can increase transduction efficiencies but often at the cost of specificity and at the risk of systemic toxicity.

Vectors currently used for cancer gene therapy often share some but not all of these traits and choice of vector will often come down to the specific type of cancer and whether the vector can be introduced intratumorally or systemically. The types of vectors used for GDEPT can be broken down into two classes: viral-vectors

(adenoviral, retroviral, herpes simplex viral, and adeno-associated viral vectors) and non-viral vectors (bacterial and liposomal based vectors). These vectors and their use in GDEPT are discussed in detail in the following sections.

Viral Vectors

Viral vectors have been the vectors of choice in the majority of clinical applications of cancer GDEPT due to both their natural evolved ability to efficiently transduce cells with actively expressed transgenes in a selective manner and the relative ease in which they can be mass produced[17]. When using viral vectors for gene delivery, GDEPT systems are sometimes referred to as Viral-Directed Enzyme Prodrug Therapy or VDEPT[16]. Viral vectors under investigation for use in cancer GDEPT have included adenoviruses, retroviruses, herpes simplex virus, adeno-associated virus (AAV)[18-20], polio virus and Epstein-Barr Virus (EBV)[21]. By far the most studied have been the adenoviral and retroviral vectors.

Adenoviral Vectors

Adenoviral vectors are widely used for gene therapies and are the most commonly used in GDEPT as they can be produced in high titers and can infect a large variety of cell types. Adenoviral vectors have a long history of therapeutic use and were used in the treatment of cancer as early as 1956 by Huebner *et. al.* who used unmodified replication-competent adenovirus to treat patients with cervical

carcinoma[22]. Wild-type adenoviruses consist of a linear double-stranded DNA genome contained within a capsid consisting of the proteins *hexon*, *fiber* and *penton*. *Hexon* is purely structural, *fiber* is involved in cell targeting by binding to the common cell surface receptor called the coxsacki and adenovirus receptor (CAR) and *penton* binds the $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrins leading to cellular internalization.

Upon wild-type adenoviral infection the adenoviral genome is transferred to the nucleus where it remains episomal. Proteins expressed from the E1A region of the adenoviral genome bind and inhibit the function of the cell regulatory protein pRB leading to cell entry into S phase[23]. Cellular apoptotic response is prevented by expression of a p53 inhibitor from the E1B region and FasL and TNF inhibitors from the E3 region. Several additional proteins expressed from the E3 region are involved in evasion of the immune response via inhibition of MHC-class I expression. During infection host cell protein synthesis is then terminated and cellular enzymes are co-opted for viral DNA replication and viral protein synthesis. Virons are assembled in the nucleus until cell death is triggered by expression of the adenoviral death protein (ADP) leading to virus release[24].

Engineering adenoviruses for use as cancer therapy vectors has involved both the creation of non-replicating and replication-selective adenoviruses. Non-replicating adenoviral vectors have been created by knocking out the E1 and/or E3 region of the native virus with the expression cassette. By eliminating viral replication the chance of non-tissue specific spread of the virus is minimized. Since these vectors are

incapable of replication, large viral loads and accurate targeting are essential. Targeting of these adenoviral vectors to cancer cells has been achieved utilizing both cell specific targeting as well as transcriptional targeting as is extensively reviewed[17,25,26].

Replication-selective adenoviruses have been engineered to replicate preferentially within cancer cells. Selective amplification of the vector within the tumor leads to much higher transduction efficiencies and the oncolytic action of the vector itself can add to the therapeutic effect[27]. One approach used to create oncolytic viral vectors has been to tie expression of gene regions critical for viral replication to a tumor-specific promoter. In adenoviral vectors, for example, the E1A region has been modified with tumor-specific promoters to generate vectors only capable of replication within cancer cells.

Another approach has been to complement the loss-of-function of cell regulatory proteins seen in cancers with loss-of-function mutants within adenoviruses. The deletion or mutation of viral genes that act to inhibit cellular regulatory proteins prevent the adenovirus from replicating within healthy cells that normally contain those proteins but allow them to replicate within cells that don't. One example is a mutation in the E1A conserved region 2 (E1A-CR2) which encodes a protein that binds and inhibits the tumor suppressor cell regulatory protein pRB leading to S phase induction[28,29]. In the E1A-CR2 mutant *dl*922/947 viral replication is impeded in quiescent cells but not within tumor cells where the G1-S checkpoint has been lost[30].

Another example is the deletion of a conserved region found within the E1B polygene encoding a 55-kD protein involved in binding the p53 tumor suppressor for the prevention of p53 mediated apoptosis[27,31]. Loss of function of this protein in the E1B-55kD deletion mutant *dl*1520 (Onyx-015) creates adenoviral vectors that trigger early p53 mediated apoptosis in normal cells thus preventing the replication and spread of the virus[27]. The majority of human cancers have genetic defects that lead to the loss of p53[32] and therefore the Onyx-015 vector is capable of replicating within cancer cells to the point of lysis and viral spread.

To date Onyx-015 has been the only replication-selective oncolytic virus to be used in phase I and II clinical trials both alone and with therapeutic genes such as those used in GDEPT. Onyx-015 used in virotherapy, without the addition of therapeutic transgenes, has been injected at high titers ($2x10^{13}$ particles) either intratumorally, intraparotenially, intraarterially and intravenously demonstrating low toxicity with some patients exhibiting flu-like symptoms[33-35]. To date GDEPT modified Onyx-015 has only been tested in clinical trials with intratumoral injections where toxicity was also low[10,11,36].

Several problems and safety issues with regard to adenoviral vectors need to be considered. Although in the vast majority of infections the adenoviral genome remains episomal, in some cases the viral genetic material can be incorporated into the host genome. Therefore side-effects resulting from insertional mutagenesis in non-target tissues should still be considered potential risk. For use in GDEPT this risk is

minimized by the fact that cells transduced with the suicide gene should be killed by the subsequent administration of the prodrug.

Eliminating or modifying the ability of the adenovirus to constitutively replicate is an important safety feature. However, changes to adenoviruses to remove their ability to replicate are often deletions of relatively small regions of the viral genome that could potentially be reconstituted if the vector transduces a cell with a wild-type adenoviral infection. This is especially true of replication-selective oncolytic adenoviruses such as ONYX-015 where only relatively small regions of the wild-type adenoviral genome have been removed. Such an event could have the potential to create infectious virus particles containing the therapeutic gene, which could potentially spread.

Another issue with adenoviral vectors is due to their inherent immunogenecity. Administration of large titers of viral particles systemically has been shown to occasionally lead to toxic shock. In one non-GDEPT clinical trial, administration of an adenoviral vector led to a fatality when the virus triggered activation of the patient's innate immunity, leading to systemic inflammation and multi-organ failure[37]. Although death due to immune response to adenoviral vectors is rare, the immune response to the vectors is known to reduce transduction rates and therapeutic efficacy, especially in the case of repeated treatment.

Adenoviral particles delivered systemically tend to accumulate in the liver often leading to nonspecific infection and transduction of liver cells. In the case of

GDEPT subsequent treatment with prodrug can then lead to liver damage. Therefore the amount of adenoviral vector administered must be controlled to minimize damage to the liver or potential toxic shock.

Retroviral Vectors

Retroviral vectors are the most widely used viral vectors in current gene therapy trials. They are characterized by their reverse-transcribing of their RNA genome into double stranded DNA which is then incorporated into the host genome via insertion. This stable insertion of therapeutic genes into the host genome allows for long-term expression that is favored in some types of gene therapy. They are capable of carrying foreign genes of ~8kB and can be generated in large enough titres for efficient gene transfer[38] and are immunologically "silent".

The most commonly used retroviral vectors are the murine leukemia virus (MuLv) and lentiviruses (of which HIV is a member). Of these two MuLv has been in more common usage with a large amount of clinical trials and data. To date lentiviruses for GDEPT have not yet been clinically tested. MuLv vectors can only transduce actively replicating cells whereas lentiviruses naturally do not require cell replication for infection. The inability of MuLv based vectors to transduce non-replicating cells has been seen as both a serious drawback for its general application to cancer GDEPT as well as beneficial for application to brain tumors where the inability to transduce quiescent cells confers additional specificity. Examples of VDEPT using

retroviral vectors include the delivery of the nitroreductase/CB1954 system into colorectal and pancreatic cancer cell lines utilizing a retroviral vector[39,40].

Non-Viral Vectors

Safety issues with the use of viral vectors have inspired research in non-viral based gene delivery systems. Although these systems currently lack the transduction efficiencies of their viral counterparts many of them hold the promise of increased specificity and safety.

Bacterial Vectors

Salmonella's natural tropism for cancer has made it a target for therapeutic development[41,42]. Observations have shown that *Salmonella* accumulates in tumors at thousand-fold higher numbers than in normal tissue across a broad range of tumor types. Much like replication-selective adenoviruses *Salmonella* alone has demonstrated tumor suppression. Investigation into arming attenuated strains of these tumor-targeting forms of *Salmonella* with suicide genes for GDEPT therapy has recently been tested using the HSV-TK/GCV system.

Liposomal Vectors

Liposomal vectors are non-biological gene delivery systems constructed *in vitro* by the combination of mixtures of cationic lipids and nucleic acids, which self-

assemble to form a protective lipid membrane envelope bound to therapeutic nucleic acids in a lipoplex. These lipoplex's have been shown to be competent, transfecting cells by a form of nonspecific endocytosis. The benefit of such systems is that they are non-immunogenic however, to date, low transfection efficiency and lack of cell specificity have limited their use in GDEPT to preclinical tests.

BYSTANDER EFFECT

Even with the most successful vectors under the best of circumstances, transduction efficiencies will never reach 100%. Experimental observation has shown that with current vectors transduction rates are closer to around 10-15% *in vivo*[43]. Therefore, for GDEPT to be effective in cancer treatment the active form of the prodrug needs to cause cell death not only in transduced cells but in nearby cells as well in what is known as a bystander effect (BE)[6,44]. Via the bystander effect treatment can be effective with much lower transduction efficiency. Using the prodrug 5-fluorocytosine, whose cytotoxic form 5-FU is freely diffusible and has a demonstrated strong BE, tumor regression has been observed after treatment when as little as 4% of the tumor cells were expressing the GDEPT enzyme cytosine deaminase[45]. As a result prodrugs for use in GDEPT are chosen for the ability of their drug form to mount an active BE after conversion to their active form in transduced cells.

The bystander effect is caused by three major factors: the transfer of the cytotoxic drug to neighboring cells; the triggering of apoptosis in neighboring cells due to transfer of apoptotic factors; and the body's natural immune response to the cytotoxin induced cell death. Many drugs, such as ganciclovir-triphosphate (the highly-charged activated form of the prodrug ganciclovir), require cell to cell contact to illicit a BE and their transfer has been shown to be dependent on the formation of gap junctions[46]. Some drugs require active transport out of the cell or the formation of apoptotic vesicles while others can transfer to neighboring cells by passive diffusion.

Studies comparing the efficacy of GDEPT systems *in vivo* in immunodeficient and immunocompetent animals have shown that the bystander effect is augmented by the immune system[47-51]. Cell death within the tumor brought on by the course of the treatment can elicit an immune response that can lead to the rejection of cancer cells even in areas distant from the tumor itself[52,53]. Studies have also shown that in immunocompetent animals GDEPT treatment in animal cancer models can lead to immunity; causing the rejection of subsequently implanted tumors of the same cancer type[54,55].

ENZYME/PRODRUG SYSTEMS USED IN GDEPT

Once an appropriate vector for gene transfer is found the success of the therapy depends on the enzyme/prodrug system chosen. To be considered for clinical use for

cancer therapy, both the GDEPT enzyme and prodrug have to meet certain requirements. A cancer GDEPT enzyme should either be exogenous to humans or be a human protein that is not expressed or expressed at very low concentration within normal tissues. The enzyme should carry out a catalytic reaction that is not present within normal tissue and that efficiently converts a prodrug into an active cytotoxic drug. For therapeutic effect the enzyme should be effectively expressed within the target cancer cells and exhibit a high kcat and low K_m for the prodrug under human physiological conditions. Ideally the enzyme should be non-immunogenic in order to prevent clearing from the body before conversion of the prodrug can lead to a bystander effect (BE).

A GDEPT prodrug should be a chemically inert, stable, non-immunogenic non-toxic molecule that is efficiently turned over into a potent cytotoxin by and only by the suicide gene enzyme introduced selectively into the target tissue. Since delivery of a therapeutic prodrug is usually systemic, the prodrug needs to be nontoxic enough to be given at high dose and stable enough in physiological conditions to travel throughout the body without degrading or being activated before reaching the target tissue. The differential in cytotoxicity between the inert prodrug and active drug forms should be as high as possible while still being safe to administer as therapy with a 100-fold differential considered necessary to see a therapeutic effect.

Both the prodrug and the cytotoxic drug should be able to travel into and through cells either by diffusion or active transport in order to reach the target tissue

and mount a BE. The chemical and biological half-lives of the released cytotoxin should also be taken into consideration. If the half-life of the cytotoxin is too long it can spread systemically via diffusion and cause cellular damage away from the target site. If the half-life is too short the cytotoxin might not elicit a suitable BE for therapy. Based on standard diffusion rates the optimum half-life of the drug is usually considered to be about one minute[56].

A wide variety of enzyme/prodrug combinations are currently being studied for use in cancer GDEPT[7]. The majority of GDEPT prodrugs have been derived from compounds that have already been used clinically in the treatment of cancer or other illness since the pharmacological, pharmacokinetic, dosage and safety parameters of these drugs are well established. In most cases the prodrug's cytotoxic form causes damage to DNA or DNA replication enzymes via misincorporation during replication or cross-linking. Described below are some of the most prominent GDEPT enzyme/prodrug combinations that have reached the stage of clinical.

Herpes Simplex Virus Thymidine Kinase (HSV-TK) and Ganciclovir (GCV)

To date the most extensively studied GDEPT system is herpes simplex virus thymidine kinase coupled with the prodrug ganciclovir. The first experimental evidence for the usefulness of GDEPT for cancer therapy was provided by Moolten *et. al.* in 1986 utilizing the HSV-TK/GCV system[6]. The first clinical trial using GDEPT utilized HSV-TK and was approved in 1991 and carried out by Freeman *et. al.*[57]. As

of 2005 82 of 95 cancer GDEPT clinical trials utilized the HSV-TK/GCV system to treat a wide variety of cancers including prostate[58], mesothelioma[59], glioblastoma multiforme[60,61] and many more[8]. Of the 25 or more cancer GDEPT systems currently being investigated only HSV-TK/GCV has advanced to the level of phase III clinical trials[8,62].

Herpes simplex virus thymidine kinase (HSV-TK) is a 90kDa homodimeric nucleoside kinase that acts as part of a thymine salvage pathway, catalyzing the transfer of the γ-phosphate of ATP to the 5'-hydroxyl of deoxythymidine (dT) to form deoxythymidine-5'-monophosphate (dTMP) and ADP. HSV-TK's activity is not specific to dT, however and unlike its human homologue it is capable of phosphorylating a wide-variety of nucleosides including other pyrimidine and guanosine analogs.

One non-natural substrate of HSV-TK is the guanosine analogue ganciclovir (GCV), which is converted to its nucleoside monophosphate by HSV-TK and subsequently to a diphosphate by guanylate kinase and then to a triphosphate by nucleoside diphosphokinase (Figure 2.2). While GCV is relatively non-toxic, GCV-TP is a potent cytotoxin, competing with dGTP for incorporation into elongating DNA during cell division causing inhibition of the DNA polymerase and single strand breaks. Cells transfected with HSV-TK and treated with GCV were shown to die by p53-independent apoptosis likely triggered by DNA damage[63]. Since GCV is a poor substrate for the mammalian nucleoside monophosphate kinases, it can be used safely

as a drug and is currently marketed as an anti-viral. HSV-TK's ability to phosphorylate GCV at considerably higher rates than the mammalian homologue plus GCV's established use as a drug have made the HSV-TK/GCV combination one of the most promising candidates for enzyme/prodrug therapy.

Despite success the HSV-TK/GCV GDEPT system does have some notable flaws. The activated drug form of GCV, GCV-TP, is highly charged and therefore incapable of diffusing through membranes. As a result the bystander effect of GCV-TP is dependent on both cell to cell contact and the presence of gap junctions[46]. Due to the high rate of mutation and heterogeneous population of most tumors, cancer cells can often lack gap junctional proteins thus limiting the therapeutic effect. Despite this drawback, preclinical trials have shown that the HSV-TK/GCV system can elicit a therapeutic effect when only around 10% of the tumor cell population was expressing HSV-TK[44,64].

In addition, the catalytic efficiency of HSV-TK to the prodrug GCV is quite low relative to its native substrate deoxythymidine (dT). Measurement of K_m and kcat shows a substantial loss in catalytic efficiency for the non-native substrates, mostly due to increased K_m of GCV (48μM)[65] compared to dT (0.2μM)[66-68]. The presence of dT in the cell and other nucleoside substrates of the enzyme such as deoxycytidine will therefore compete with GCV, lessening the effects of any treatment. This effect can be seen in current trials where very high doses of GCV are needed to see any ablation of the tumors transfected with HSV-TK. Although relatively non-toxic

compared to its triphosphate, high doses of GCV can lead to toxic effects, especially in the bone marrow and immunosuppression. In humans the dosage of GCV considered safe is around 10 mg/kg/day[64].

Cytosine Deaminase (CD) and 5-Fluorocytosine (5-FC)

Following HSV-TK/GCV the cytosine deaminase(CD)/5-fluorocytosine(5-FC) system is the most studied and prevalent GDEPT system. Like ganciclovir, the prodrug 5-FC has had a prior history in therapy where it has been used as an antifungal drug. In addition the cytotoxic product of the CD/5-FC reaction, 5-fluorouracil (5-FU), is a commonly used cancer chemotherapy drug. The fact that the pharmacokinetic and pharmacological properties of both the prodrug 5-FC and the active cytotoxic drug 5-FU had already been extensively studied led to the rapid development of the CD/5-FC system for use in clinical cancer GDEPT. The first study utilizing the CD/5-FC system for use in GDEPT was carried out in 1992 by Mullen *et. al.* who transferred the bacterial cytosine deaminase gene to mammalian cells and tested their sensitivity to 5-FC[69]. As of 2005 there were ten registered clinical trials utilizing the CD/5-FC GDEPT system for cancer therapy[8].

Cytosine deaminase is naturally present in both bacteria (bCD) and yeast (yCD) where it acts as part of a nucleotide salvage pathway converting cytosine to uracil and ammonia. Both enzymes can also convert the non-toxic nucleoside analogue 5-fluorocytosine (5-FC) into 5-fluorouracil (5-FU) which is then converted

by endogenous enzymes into the cytotoxic drugs 5-fluorodeoxyuradine-5'-monophosphate (5-FdUMP) and 5-fluorouradine-5'-triphosphate (5-FUTP) (Figure 2.2). 5-FdUMP is as an irreversible inhibitor of thymidylate synthase and its presence leads to the inhibition of DNA synthesis by preventing the formation of thymidylate, a precursor for thymidine triphosphate[70]. 5-FUTP is misincorporated into RNA leading to chain termination and inhibition of protein synthesis[71].

One of the benefits of the CD/5-FC system is that the active form of the drug, 5-FU, is diffusible through membranes and thus generates a strong bystander effect without the requirement of cell contact or gap junctions[72]. Studies of CD/5-FC for use in cancer therapy have shown tumor regression with transduction efficiencies as low as 4%[45].

Comparisons between bacterial and yeast cytosine deaminase have revealed that although both are capable of converting the prodrug to the active drug, yCD, catalytic efficiency against the prodrug relative to its natural substrate cytosine is much higher than bCD[13]. This catalytic advantage has been shown to translate to a therapeutic advantage when yCD is substituted for bCD in GDEPT applications[11,13,73]. Despite its kinetic advantages, yCD is unstable in physiological conditions, which has a negative effect on its therapeutic efficacy[13].

Although humans have no homologue to either yeast or bacterial cytosine deaminase, the presence of naturally occurring CD in the intestinal flora can lead to

side-effects upon administration of the prodrug 5-FC. A safe maximum dosage of 5-FC is considered to be around 500 mg/kg/day.

Nitroreductase (NTR) and CB1954

The development of the nitroreductase/CB1954 system started with the discovery of the prodrug CB1954 by screening a drug compound library for activity against rat Walker 256 carcinoma cancer cell line[74]. Although CB1954 demonstrated effective cell kill with high therapeutic index in rats direct translation of the drug to human trials was ineffective with dose-limited toxicity reached before any tumor regression was seen. Later it was determined that the cytotoxicity of this compound was generated by conversion into a DNA alkylating agent by the enzyme DT-diaphorase[75]. Interestingly the homologous enzyme in humans is a poor catalyst of the conversion of the CB1954 prodrug to the active species despite 85% identity to the rat protein. It was then discovered that the *E. coli* enzyme nitroreductase (NTR) was capable of converting the prodrug CB1954 to its active form efficiently (Figure 2.4) which led to the formulation of the NTR/CB1954 enzyme/prodrug system[76].

The first *in vitro* test of the NTR/CB1954 system was published in 1995 by Bridgewater *et. al.* who demonstrated the successful selective cell killing of mouse fibroblast cells transduced with a retrovirus containing a *E. coli* NTR transgene upon exposure to CB1954[77]. After proof of principle was established the NTR/CB1954 system was tested again *in vitro* against human colorectal, ovarian and pancreatic

cancer cell lines where a large increase in sensitization to CB1954 was demonstrated[39,40,78]. The first clinical trial of NTR/CB1954 GDEPT was carried out by Chung-Faye *et. al.* in 2001 in order to look at the pharmacokinetic properties of the therapy in humans[79]. Many examples of *in vitro* and *in vivo* preclinical and clinical investigations of the NTR/CB1954 cancer GDEPT system have been reviewed[8,80,81].

Cytochrome P450 (CYP) and Cyclophosphamide (CP)

The cytochrome P450s (CYPs) are a family of hemeprotein monoxygenases involved in the metabolism of drugs, pollutants and other xenobiotics. The rat liver cytochrome P450 enzyme CYP2B1 was shown to catalyze the hydroxylation of the oxazophorine prodrug cyclophosphamide (CP) as well as its isomer iphosphamide (IP). The 4-hydroxy metabolites then spontaneously decompose through a β-elimination reaction to yield both acrolein and the phosphoramide mustard in equal molar amounts[82] (Figure 2.5). Phosphoramide mustard acts as a cell-cycle independent alkylating agent capable of forming cross-links in DNA.

Preclinical animal trials using the CYP2B1/CP enzyme/prodrug model for cancer GDEPT demonstrated tumor regression with low global toxicity[83,84]. Success in these investigations led to the conduction of early phase I/II clinical trials using CYP2B1/CP for the treatment of 14 patients with inoperable pancreatic cancer. The trials showed several fold increases in survival relative to records of the progression of similar cancer[85]. One issue with the use of cytochrome P450s for GDEPT cancer

therapy is the presence of CYPs in humans. Although the human homolgue to rat CYP2B1 (CYP2B6) is also capable of converting CP to its active form it does so at a much lower catalytic rate[86]. Trials have registered minimal liver toxicity likely due to the low catalytic activity of CYP2B6 coupled with its low level of expression and the short half-life of the activated form of CP.

Additional Enzyme/Prodrug Systems

Many more enzyme/prodrug systems are under investigation for potential cancer therapy in preclinical trials. Portsmouth *et. al.* reviewed the state of development of cancer suicide gene therapies in 2007 citing 25 examples of enzyme/prodrug systems currently under investigation[7]. As with the HSV-TK/GCV, CD/5-FC, NTR/CB1954 and CYP450/CP systems, in most cases enzyme/prodrug systems designed or chosen for use in cancer GDEPT act by causing some form of DNA damage either by misincorporation or cross-linking.

IMPROVING THE EFFICACY OF CANCER GDEPT

Limitations in current vector based gene delivery and enzyme/prodrug systems have reduced the potential therapeutic effect as seen in clinical trials. Techniques currently under investigation to improve the efficacy of cancer GDEPT include using multiple suicide genes, combining radiosensitizing GDEPT prodrugs with radiotherapy, co-transduction of suicide genes with downstream metabolic enzymes,

co-transduction of suicide genes with cytokines to bolster the immune response, and improving the biophysical properties of GDEPT enzymes.

Double-Suicide Gene Therapy

The effectiveness of GDEPT therapy can be increased tremendously by the combination of multiple enzyme/prodrug systems into a single treatment The most successful clinical trial to date using a GDEPT system utilized a mix of HSV-TK/GCV and CD/5-FC in what is known as double suicide gene therapy[9-11,36,87]. Studies testing the efficacy of HSV-TK/GCV and CD/5-FC therapies alone and in combination have shown considerable improvement and synergy between the two therapies when they are used together. This synergy is thought to arise from the inhibition of thymidylate synthase by 5-FdUTP leads to an allosteric effect on the nucleoside pool regulation enzyme ribonucleotide reductase (RNR), which is regulated by the production of thymidine nucleotides. This leads to a reduction in the amount of guanosine-triphosphate, which acts as a competitive inhibitor of the cytotoxic effect of the HSV-TK/GCV cytotoxic drug form GCV-TP. Tumors resistant to either HSV-TK/GCV or CD/5-FC alone were shown to completely regress when treated by the combination therapy.

Radiosensitization by GDEPT

DNA damage caused by radiation acts synergistically with DNA damaging cytotoxic drugs to lead to considerably more damage and cell death. The combination of radiotherapy with GDEPT has been shown to act synergistically demonstrating that treatment with GDEPT can radiosensitize tumors so that concurrent radiotherapy has a much larger effect[9-11] The HSV-TK systems potential to radiosensitize cells was demonstrated first by Kim *et. al.*, 1994 using the prodrug BvdUrd not GCV[88]. The chemotherapy drug 5-FU is known to be radiosensitizing in cancer cells as shown by Bruso *et. al.*, 1990[89] and treatment with CD/5-FC has a similar radiosensitizing effect.

Co-transduction with Downstream Effectors

Another method to increase the efficacy of enzyme/prodrug therapy has been to include additional genes encoding for proteins that aid in either the conversion of the prodrug to the active drug, the dissemination of the active drug to nearby cells or the potency of the bystander effect.

The CD/5-FC system relies on downstream endogenous enzymes to convert the 5-FU product of the cytosine deaminase reaction to the active drugs 5-FdUMP and 5-FUTP. The first step in this pathway requires the enzyme uracil phosphoribosyl-transferase (UPRT) to add a ribosyl sugar to the uracil nucleobase. Although this enzyme is naturally found in humans, levels of this enzyme naturally occurring may

be a limiting step in the conversion of 5-FC to active drug[90]. Co-transduction of CD with UPRT has shown to increase the efficacy of therapy[91-93].

Other investigators have used contransduction with effectors as a means of bolstering the bystander effect of GDEPT systems. The HSV-TK/GCV system relies on the presence of gap junctions for its bystander effect since the highly charged triphosphorylated drug form GCV-TP is incapable of diffusion across membranes. In an attempt to increase the strength of the bystander effect, the HSV-TK suicide gene was co-transduced with the gap junctional protein connexin[94]. Another approach was to enhance the immune response by co-transduction suicide genes with cytokines and thereby increasing both the local and distal bystander effect[95]. Studies of co-transduction of cytosine deaminase[96] or herpes simplex virus thymidine kinase[97] with interleukin-2 showed some enhancement of tumor growth inhibition.

Improving GDEPT Enzymes

Another way to improve the efficacy of GDEPT would be to increase the efficiency of prodrug-drug conversion by finding or engineering improved enzymes. One way to do this is to simply find enzymes in nature that show improved catalytic properties to the conversion of prodrug relative to natural substrate. This was the case in the discovery of yeast cytosine deaminase (yCD) which showed considerably higher specificity for the prodrug compared to the previously used enzyme bacterial cytosine deaminase (bCD)[13,98]. Another approach would be to use protein design methodology

in order to tailor some of the currently used enzymes for GDEPT in order to improve their pharmacokinetic and pharmacological properties.

The use of protein engineering techniques to improve GDEPT enzymes has been discussed within recent review articles[99,100]. Examples include the use of directed evolution to alter the substrate specificity of HSV-TK[65,100-102] and the use of rational structure based design to identify mutants to increase catalytic efficiency of *E. coli* nitroreductase[103]. With the exception of work contained within this dissertation (chapters four and five), there are no examples of computational protein design being used to improve the efficacy of enzymes for use in human therapy.

Table 2.1 Enzyme/Prodrug Systems for Cancer Therapy. The released drug is defined as the catalytic product of the reaction between the nonendogenous suicide gene enzyme and the prodrug. The active drug is defined as the final cytotoxic form of the drug, which may be formed only after additional catalytic steps performed by endogenous enzymes. The original publication is the first appearance of this enzyme/prodrug combination for use in GDEPT in the literature.

Enzyme	Prodrug	Released Drug	Active Drug	Mode of Toxicity	Enzyme Origin	Clinical Trials	Original Publication
Thymadine Kinase (TK)	GCV, ACV, BVDU	GCV-MP, ACV-MP, BVDU-MP	GCV-TP, ACV-TP, BVDU-TP	Inhibition of DNA Polymerase, DNA strand breaks	Herpes Simplex Virus	Phase I/II/III	Moolten *et. al.* 1986
Cytosine Deaminase (CD)	5-FC	5-FU	5-FdUMP, 5-FdUTP, 5-FUTP	Inhibition of thymidilate synthase, DNA and RNA strand breaks	*E. coli, S. cerevisiae*	Phase I/II	Mullen *et. al.* 1992
Nitroreductase (NTR)	CB1954	5-(aziridin-1-yl)-4-hydroxylamino-2-nitrobenzamide	5-(aziridin-1-yl)-4-acetoxyamino-2-nitrobenzamide	DNA cross-links	*E. coli*	Phase I	Knox *et. al.* 1988
Cytochrome P450 (CYP 450)	cyclophosphamide (CPA)	4-hydroxy-CPA	phosphoramide mustard	DNA cross-links, protein alkylation	Human	Phase I	Clark and Waxman, 1989? Chang *et. al.* 1993?
Carboxypeptidase G2 (CPG2)	CMDA	4-[(2-chloroethyl)(2-mesyloxyethyl)amino]benzoic acid	4-[(2-chloroethyl)(2-mesyloxyethyl)amino]benzoic acid	DNA cross-links	pseudomonas	preclinical, *in vitro and in vivo*	Bagshawe *et. al.*, 1988
Horseradish peroxidase	indole-3-acetic acid	radical cation	3-methylene-2-oxidole (MOI)	DNA damage, thiol depletion, lipid oxidation	horseradish plant	preclinical, *in vitro*	Greco *et. al.*, 2000
Deoxycytidine kinase (dCK)	ara-C,dFdC (gemcitabine)	araC-MP, dFdC-MP	araC-TP, dFdC-TP	Inhibition of DNA synthesis	Human	preclinical	Manome *et. al.* 1996
Purine nucleoside phosphorylase (PNP)	MeP-dR, F-araA	MeP, F-Ade	ATP analogues	Inhibition of RNA and protein synthesis	*E. coli*	preclinical	Sorscher *et. al.* 1994

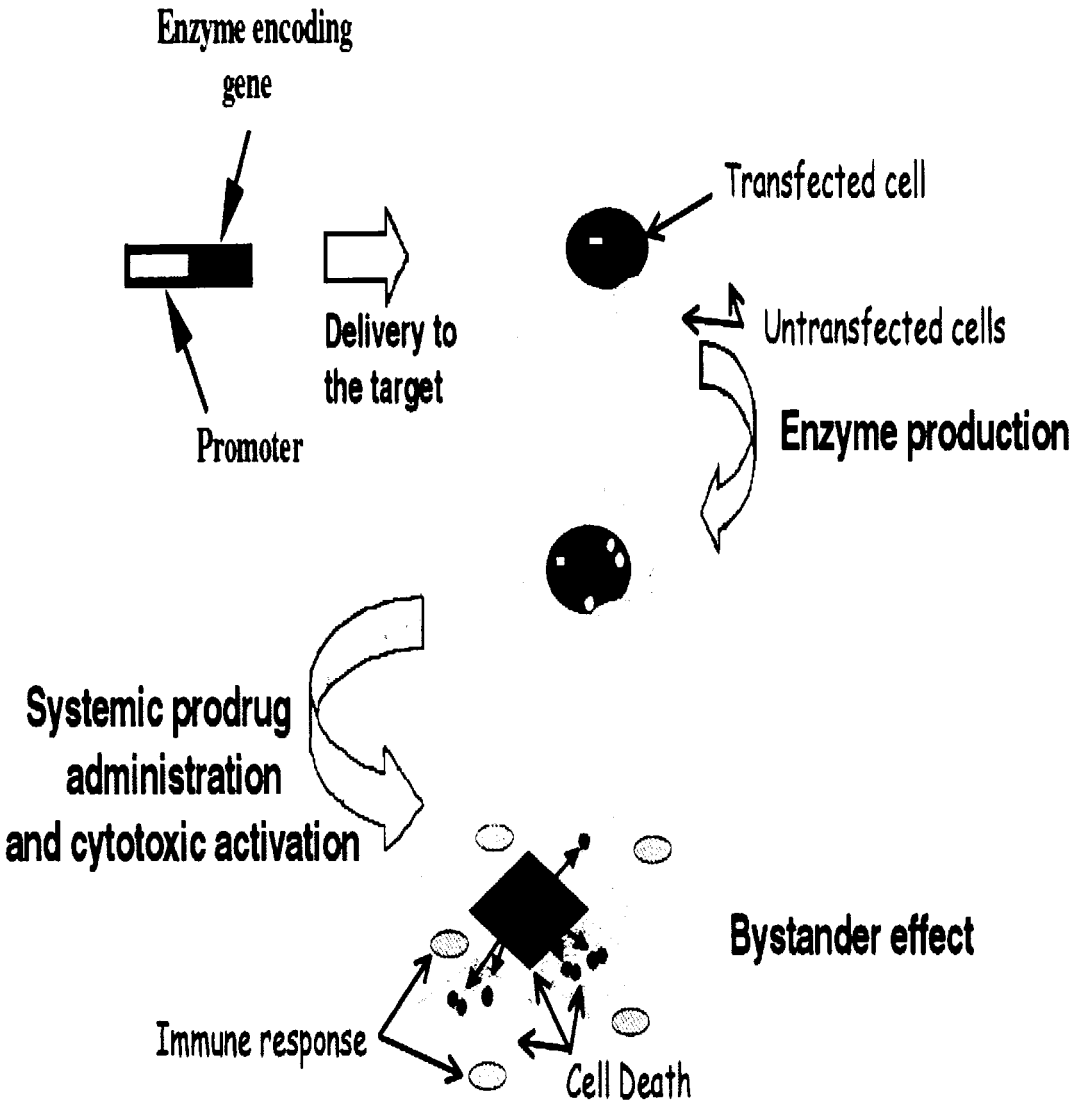


Figure 2.1 Gene-Directed Enzyme/Prodrug Therapy (from *Greco and Dachs, 2001).* GDEPT involves the delivery of a gene encoding an enzyme with non-endogenous catalytic function into a target tissue. Administration of a nontoxic prodrug that can be converted into a potent cytotoxin by the non-endogenous enzyme leads to cell death in the transduced cells as well as neighboring cells in what is known as the bystander effect.

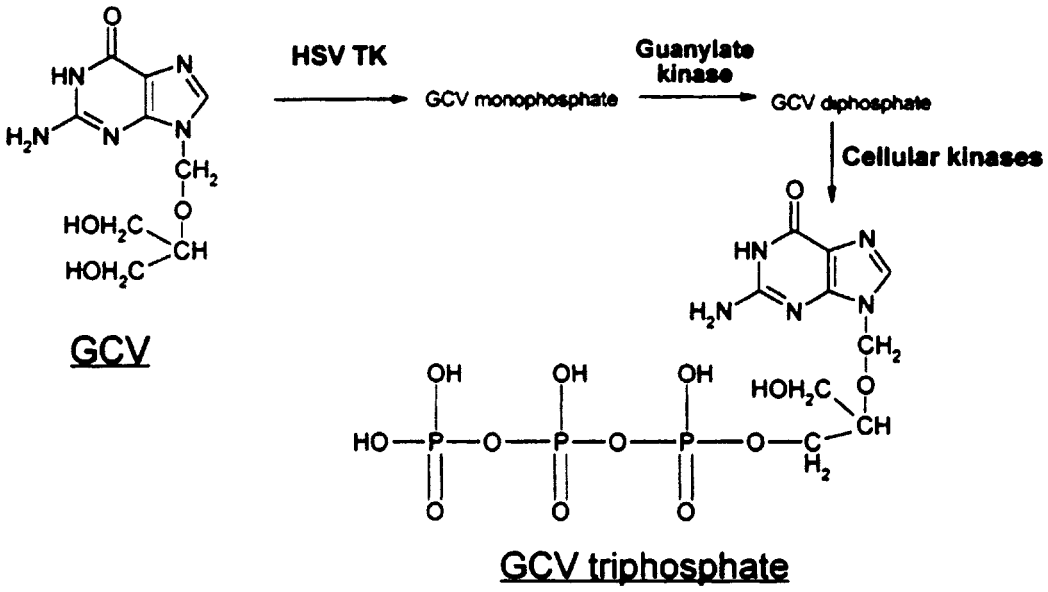


Figure 2.2 HSV-TK/Ganciclovir Prodrug Activation Pathway (from *Greco and Dachs, 2001*). Herpes simplex virus thymidine kinase phosphorylates the nucleoside analogue prodrug ganciclovir (GCV) to form GCV-monophosphate which is subsequently converted to the diphosphate by guanylate kinase and to the active cytotoxic drug GCV-triphosphate by additional cellular kinases. GCV-triphosphates acts as an inhibitor of DNA polymerase and misincorporation into DNA leads to chain termination and strand breaks.

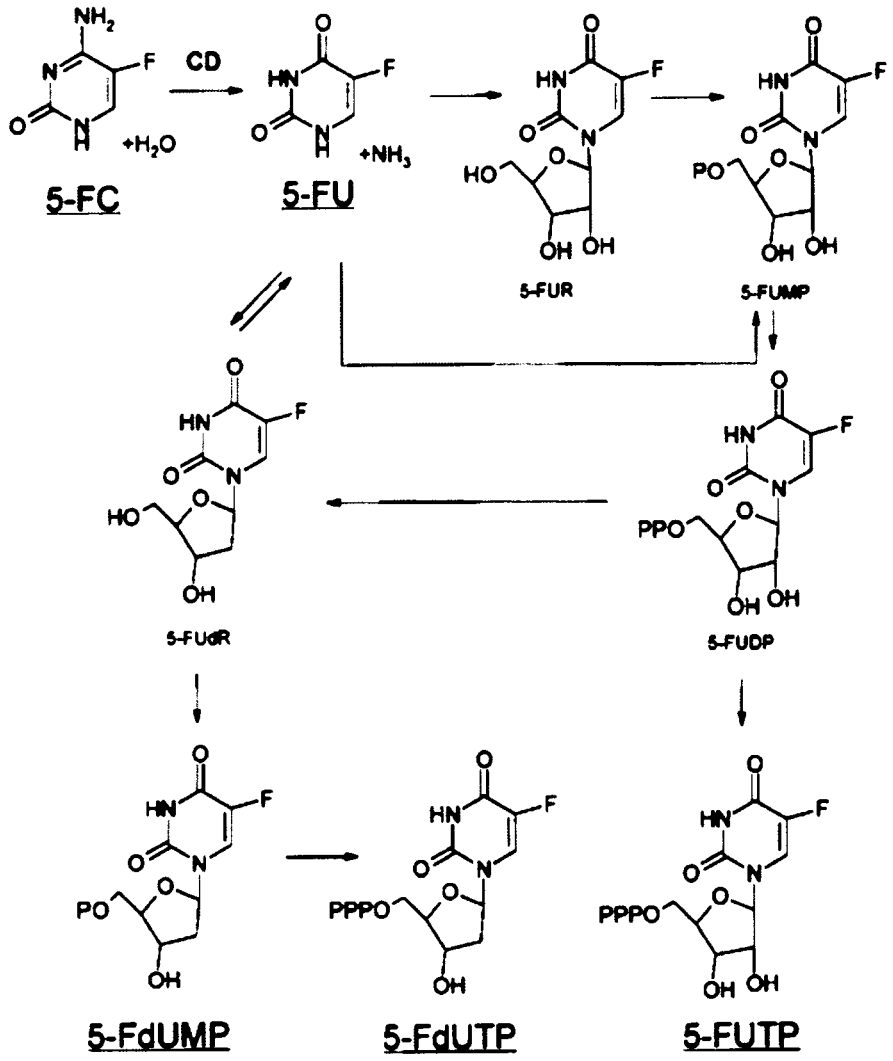


Figure 2.3 Cytosine Deaminase/5-Fluorocytosine Prodrug Activation Pathway (from *Greco and Dachs*, 2001). After conversion of the prodrug 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FC) by cytosine deaminase the endogenous cellular enzyme converts 5-FU to 5-fluorodeoxyuradine monophosphate (5-FdUMP), which acts as a potent inhibitor of thymidylate synthase, and into 5-fluorodeoxyuradine triphosphate (5-FdUTP) and 5-fluorouradine triphosphate (5-FUTP) whose incorporation into DNA or RNA respectively leads to chain termination.

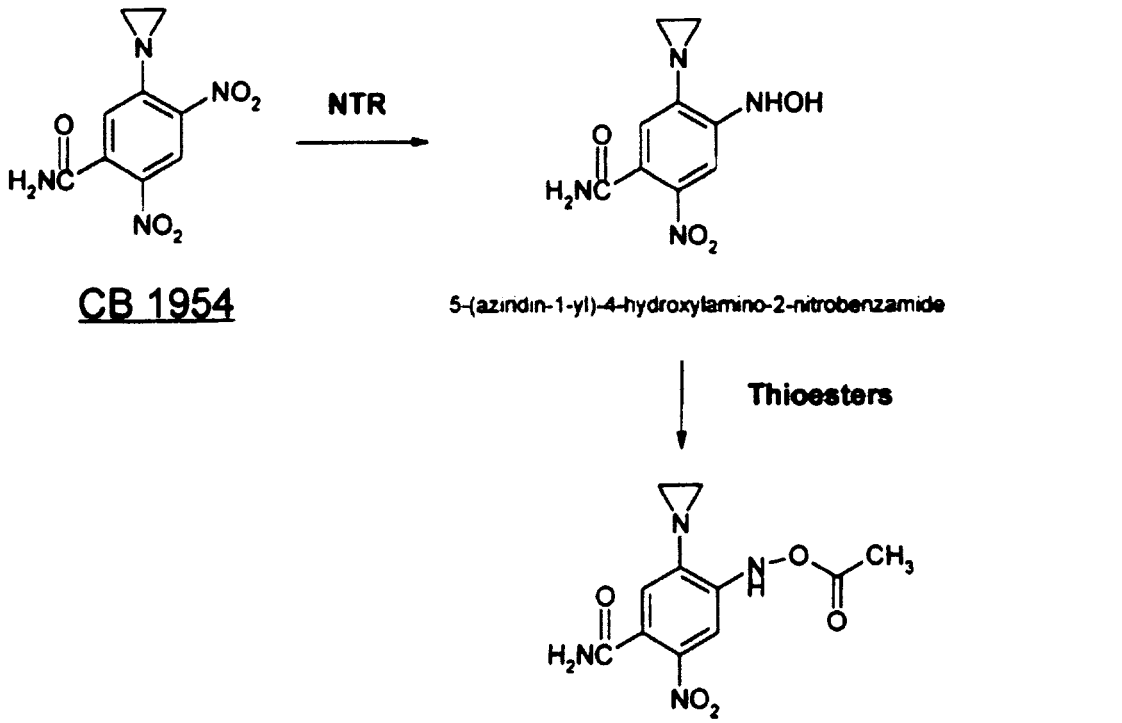


Figure 2.4 Nitroreductase/CB1954 Prodrug Activation pathway (from *Greco and Dachs, 2001*). Nitroreductase catalyzes the aerobic reduction of the prodrug CB1954 to 5-(aziridin-1-yl)-4-hydroxylamino-2-nitrobenzamide. This derivitive is activated in the presence of thioesters, such as coenzyme A, into the cytotoxic drug 5-(aziridin-1-yl)-4-acetoxyamino-2-nitrobenzamide, which reacts with DNA to form crosslinks.

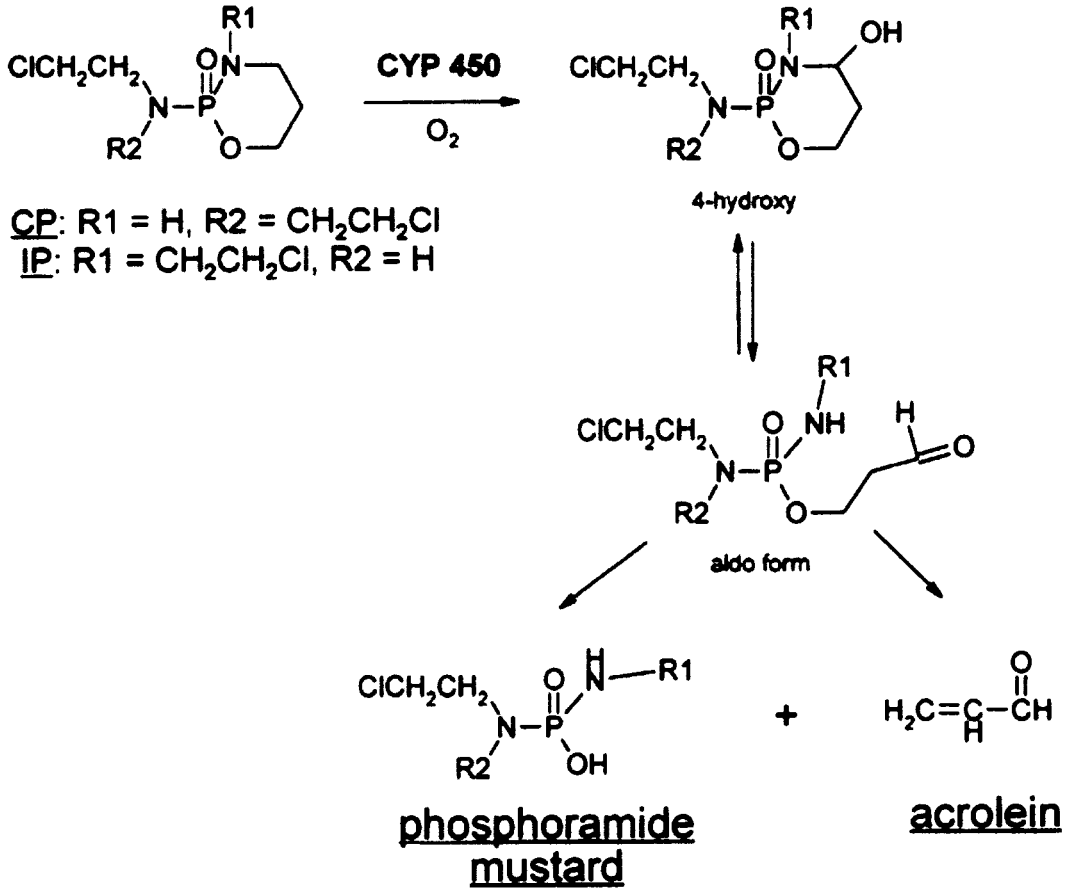


Figure 2.5 Cytochrome P450/Cyclophosphamide Prodrug Activation Pathway (from *Greco and Dachs, 2001*). Cytochrome P450 (CPY 450) catalyzes the conversion of the prodrugs cyclophospamide (CP) and its isomer iphosphamide (IP) into 4-hydroxy-cyclophosphamide or 4-hydroxy-iphosphamide respectively via a hydroxylation reaction. These 4-hydroxy compounds form an equilibrium with the open ring aldol form which spontaneously and irreversibly decomposes into acrolein and a phosphoramide mustard. The phosphoramide mustard acts as the active form of the cytotoxic drug causing DNA damage via alkalation and crosslinking. Acrolein is a biproduct of the reaction and can cause side-effects at high dosages.

Chapter Three: Protein Engineering

INTRODUCTION

The biophysical properties of proteins used for industry, molecular biology or therapeutics are often non-ideal for the task that they are asked to perform. The protein engineering field is devoted to developing methodologies for the optimization of proteins for any given process, increasing the effectiveness of currently used proteins as well as potentially providing new tools by improving proteins that are currently too unstable or lack the functionality necessary to be useful. The two most prominent design methodologies used for protein design are directed evolution and rational design. Directed evolution methods, at their core, do not require any structural knowledge of the protein being designed. Rather they employ evolutionary processes such as random mutagenesis, recombination and selection in order to select and amplify desired traits. Rational design methods, on the other hand, rely on structural knowledge of the target protein and knowledge of protein structure/function in order to predict potentially beneficial mutations. The field of computational protein design combines our growing knowledge of protein structure/function with the exponential growth of computational power in order to automate ration design and apply it to vast numbers of potential sequences and conformations for a given protein structure. In this chapter we look at both methodologies, how they work, their benefits and drawbacks and what they have accomplished.

DIRECTED EVOLUTION

The majority of protein engineering efforts to date have utilized the process of directed evolution, which mimics Darwinian evolution's process of non-random selection from random mutation in order to identify beneficial mutations. In directed evolution protocols, the random variation seen in nature is artificially created in the form of vast libraries of randomly or semi-randomly mutated protein gene sequences. These sequence libraries can be used to generate large combinatorial libraries by DNA shuffling, a method analogous to homologous recombination, whereby sequences are fragmented and reassembled based on homology. Mutant libraries are then surveyed for variants with improved function, either by high-throughput screening or a selection scheme. The best variant may then be used as a parent sequence for a new round of mutation, recombination and screening or selection.

Screening methods work by testing each individual mutant for improved function, while selection methods work by the creation of an artificial environment in which only beneficial mutations expressing the desired trait will survive. The gene encoding for the best protein variant from this screen or selection then becomes the parent sequence for a new round of mutation and surveying. This process is repeated until a desired goal is achieved or until no more improvement is seen.

The cost of directed evolution is a function of the size of the library generated, the expense of materials used in screening and how many rounds of design are run. Larger libraries and more rounds don't guarantee success since even a huge mutant

library of 10^9 sequences represents a miniscule fraction of 20^{100} (1.3×10^{130}) potential sequences for even a small 100 amino acid protein. Still, this cost and risk is often outweighed by the fact that directed evolution protocols allow for the successful design of proteins without the need for any biophysical or structural knowledge of the design target.

Directed evolution protocols have been used successfully to improve the biophysical properties of both enzymes[104,105] and binding proteins. For the improvement of enzymes, directed evolution techniques have been used to shift enantioselectivity[106,107], alter substrate specificity[108,109], improve catalytic efficiency[110,111], increase thermostability[112-114], increase stability in organic solvents[115], increase solubility[116] and improve heterologous expression[117].

Generating a Library

The first step in any directed evolution protocol is the generation of a library of mutants to screen. Large combinatorial libraries of mutated genes were made possible by the development of PCR based random mutagenesis[118] and gene-shuffling techniques[119,120] analogous to the evolutionary processes of natural mutation and sexual recombination respectively. Although larger sequence libraries have a greater chance of identifying beneficial mutations library size is limited by technology, cost, and feasibility with the desired screen. Even the largest libraries to date, consisting of

some 10^9 mutants, represent only a tiny sliver of the 20^{100} or 1.3×10^{130} possible sequences in a modest 100 amino acid protein.

Knowledge of the proteins sequence structure/function relationships can be used to narrow the search and increase the likelihood of success by focusing mutations regiospecifically at the site of protein function. This specific sequence-to-structure/function information about the target protein could come from any number of sources including multiple sequence alignments with proteins of known function, structural homology modeling or direct structure information from a crystal or NMR structure. By limiting random mutation to the relatively few amino acids directly involved in an enzyme's catalytic function or a binding protein's surface, libraries can be generated that represent a much larger slice of potential sequence space. For example, random mutagenesis of a 10 amino acid sequence representing a catalytic cleft would require a library size of 20^{10} or about 10^{13} possible sequences to fully saturate available sequence space.

Error-Prone PCR

Error-prone PCR is a commonly used technique for introducing random mutations throughout a gene. These random mutations are introduced by running the polymerase chain reaction under conditions that reduce the fidelity of replication[118,121]. The natural error rate of *Taq* polymerase is about 0.001-0.02% per nucleotide which is too low to introduce the number of genetic mutations desired for rounds of directed

evolution. To increase this error rate error-prone PCR methods must change either the PCR reaction conditions or the polymerase itself in order to further reduce fidelity. In 1989 Leung *et al* demonstrated that an error rate of around 2% could be achieved by increasing the concentration of $MgCl_2$, addition of $MnCl_2$, increasing the concentration of dNTPs, increasing the concentration of polymerase and increasing the DNA extension used in PCR[121]. Unfortunately this method results in an excess of A to G and T to C conversions. Cadwell *et al* demonstrated that this GC bias can be overcome by simply unbalancing the dNTP pool, raising the concentration of dCTP and dTTP 5 fold relative to dGTP and dATP. Their modified protocol demonstrated a mutation rate of about 0.7% with no GC or AT bias[118].

DNA Shuffling

DNA shuffling involves the random fragmentation of large libraries of mutant genes followed by their reassembly in a way that mimics homologous recombination[119,120]. This results in a "shuffling" of all the nucleotide polymorphisms within the mutant library to form recombinant gene sequences. This combinatorial process significantly broadens the potential sequence space sampled by the mutant library when compared to the gradual accumulation of point mutants in methods such as error-prone PCR.

Random fragmentation of the mutant sequence library is carried out by digestion with DNase I. These gene fragments are then reassembled by PCR where

the fragments act as both the primers and the template leading to recombination of genes within the pool by the incorporation of fragments from one sequence into another based on their homology (Figure 3.1). The process of recombination introduces new mutations through copying errors at a rate of about 0.7%, similar to that seen in error-prone PCR[119]. Additional variation can be introduced by spiking the mixture with synthetic oligos containing both regions of gene homology and varying sequence[120]. Any extraneous neutral mutations that have no effect on the selection process can then be removed by backcrossing pools of mutants with excess of wild-type DNA fragments[119]. Using multiple related DNA sequences for DNA shuffling has been shown to vastly accelerated the directed evolution of a target protein[122].

Regiospecific Random Mutagenesis

Regiospecific random mutagenesis is the process by which random mutations are generated in a specific area of a protein sequence. To be effective, structural knowledge of the protein being designed is needed in order to identify functionally important residues. To target specific residues for mutation the protein gene sequence of interest is first cloned into an expression vector. Two unique restriction sites flanking the region to be randomly mutagenized (R1 and R2) are identified or engineered into the sequence. A loss of function mutation is then introduced within the design region to reduce background during the screening process. A library of randomly mutagenized oligonucleotides, which span the target region and contain the

restriction site R1, is then generated. A DNA primer containing a second restriction site (R2) is then used along with DNA polymerase to generate a double stranded DNA oligos with flanking restriction sites on either end of the random mutagenized section. These oligos are then cleaved with restriction enzymes R1 and R2 and then ligated into a similarly restricted expression vector containing the protein gene sequence (Figure 3.2). This produces a library of expression vectors containing the protein gene sequence of interest with a specific region being randomly mutagenized.

Selection Methods

After generating a library of mutants for a given gene a selection or screening method is needed to find any improved variants from the vast pool of sequences. Selections methodologies use artificially created environmental conditions in order to confer a survival advantage to genes that express the desired trait. Competition between these genes within a restrictive environment results in the enrichment of genes coding for proteins that express the desired trait. Selection methods can be split into two categories: *in vivo* and *in vitro*

in vivo selection methods create environments that link the survival of a microorganism transformed with the mutant library to the desired function being selected for from that library. Under such conditions any mutant genes that lead to a loss of function of the designed protein immediately perish. The surviving microorganisms, which contain the transgenes encoding for a functional protein, are

then allowed to compete under restrictive conditions in order to select for the best variant. Very large mutant libraries can be screened by *in vivo* selection since their size is limited only by transformation efficiencies of the microorganism.

This basic *in vivo* selection protocol can be used to select for any biophysical property crucial for the function of any protein whose function can be made essential for the survival of a microorganism. Examples include improving the catalytic activity of enzymes, increasing binding affinity, and improve the thermostability of a protein[123]. Of course not all proteins express a phenotypic difference that can be exploited for selection *in vivo*. In those cases an *in vitro* selection method may be used.

In vitro selection methods link genotype to phenotype via a non-biological process. The most common *in vitro* selection methods are surface display techniques, such as phage[124], cell-surface[125] and ribosome display[126], in which the mutant sequence library is expressed in such a way that the individual protein variants created are tethered, either indirectly or directly, to their own gene. These tethered proteins can then be screened as a population for the desired trait. In phage display, the sequences from the mutant library are inserted into a viral phage genome at a site encoding a coat protein. Upon viral assembly the resulting fusion between the viral coat protein and one of the protein variants forms the viral capsid with the protein variant displayed on the outside of the phage. Each phage displays one protein variant from the library and inside each phage is the genetic code for that particular protein

variant. Phage display methods are commonly used to select for improved binding affinity by simply flowing phage over the surface coated in immobilized ligand or substrate. Any phage that bind can then be harvested and sequenced.

A variation on phage display technique, known as Proside (proteins stabilized by directed evolution), has been used to find stabilized protein variants from large mutant libraries. Rather than relying on protein function as a method of screening, Proside relies on the assumption that conformational stability and proteolytic resistance are correlated[127]. Protein variants are expressed on the surface of viruses containing their genetic code as a linker between the capsid envelope and a filament crucial for virulence[128]. This virus library is then subjected to proteolysis and then allowed to infect cells. Protein variants that are unstable will be proteolytically cleaved cutting the filament from the virus preventing subsequent infection. As a result only viruses encoding stable versions of the design protein will be capable of infection. The genes for these stable variants can then be excised from infected cells and sequenced to determine what mutations are present.

Screening Methods

In the case where genotype cannot be translated to a selectable phenotype the mutant library will need to be screened. Where selection looks at whole populations of mutants screening protocols involve testing each mutant variant individually for

improvement. Even with a relatively small mutational library screening is labor intensive and requires the use of high-throughput techniques.

The majority of screening methodologies involve tying the design protein function (genotype) to a change in optical absorption or fluorescence (phenotype). This phenotype can then be screened by a high-throughput microplate spectrophotometer or by fluorescence-activated cell sorting (FACS). For example an enzyme's catalytic function can be screened in this manner by coupling the conversion of substrate to product to a spectrophotometric or fluorescent change. This can be used to screen for either increased catalytic efficiency, by rate of substrate turnover, or for increased thermostability, by measuring residual catalytic function of enzymes exposed to elevated temperatures[129].

COMPUTATIONAL PROTEIN DESIGN

Although directed evolution has been successful in optimizing a wide variety of protein biophysical properties, the requirement of an effective screening process and the cost associated with high-throughput screening techniques often makes such approaches unfeasible. Computational structure-based design methods provide an alternate approach to protein engineering and in the past several years methods for computational protein design have improved significantly[130-134].

Computational protein design programs operate by attempting to predict the optimal sequence of amino acids for a given protein fold. All computational protein

design programs are comprised of two essential elements: an energy function[135,136] derived from mathematical models, empirical data or both that is used to evaluate the fitness of a given structure; and a search algorithm, a method of rapidly sampling the vastness of sequence space to identify the sequence with the highest fitness based on the energy function. All design programs use structural data, usually derived from a crystal structure, as a starting point or template for design. The variety seen in different computational design programs comes from what goes into the energy function and what search algorithm is employed.

Programs can often be run on a standard desktop computer in a period of minutes to a few hours and the only cost is the cost associated with materials used to generate and test the relatively small list of suggested mutations. The promise of computational design therefore is to provide a cheap, fast, viable alternative to directed-evolution that may open up new engineering opportunities that were closed before due to lack of a reasonable selection method or prohibitively expensive screening methods.

To date computational design has been used successfully to thermostabilize proteins[137-141], redesign binding pockets[142-147], create a novel protein fold[148], turn a receptor into an enzyme[149], increase the catalytic efficiency of an enzyme[150,151] and alter the specificity of an enzyme[151]. Our work, described in chapter four, was the first example of computational design being used to thermostabilize an enzyme[152].

Energy Functions

Computational protein design programs use scoring or energy functions in order to evaluate the fitness of a given sequence for a given structure. These energy functions are comprised of a collection of mathematical and/or statistically derived terms that when combined form a computational representation of the interactions between all the atoms in the protein and surrounding solvent. The energy function must also balance accuracy with low computational cost in order to produce successful predictions in a reasonable amount of time. Much of the computational expense is saved by keeping the protein backbone fixed, modeling solvent as a continuous dielectric rather than explicit molecules, and representing amino-acid side-chains as a statistically-based collection of discrete conformations known as rotamers[153] rather than simulating motion. In addition the terms used in energy functions must be simple enough to run quickly on a standard desktop computer while still being valid representations of interactions occurring within a protein.

The terms used in energy functions can be split into three distinct groups: bonded, non-bonded and solvation terms. Bonded energy terms describe the interactions between atoms that are covalently bonded to one another. These terms serve to constrain bond lengths and bond angles as well as represent the potential energy associated with the torsional rotation of bonded atoms. Non-bonded terms describe the interactions between atoms that are not directly bonded to each other. They include the modeling of the hydrophobic van der Waals attraction and repulsion,

electrostatic interactions and hydrogen bonds. Solvent energy terms model the proteins interaction with its environment, typically that of water. Although water can be modeled explicitly, as in molecular dynamic simulations, in computational protein design addition of thousands of explicit solvent molecules is too computationally expensive and therefore implicit solvent models are used instead. These models treat the solvent as a single continuum with a set dielectric constant.

The terms used in the energy function can be modeled either using mathematical functions derived from quantum calculations (molecular mechanics) or using statistical inferences from the collection of known protein structures within the protein structure database (knowledge-based).

Molecular mechanic potential energy functions are mathematical descriptions of molecular interactions. Examples include: the Lennard-Jones potential, used to model the attractive and repulsive components of the van der Waals force; Coulomb's law, used to model electrostatic interactions; and dipole-dipole potentials used to model hydrogen bonding.

Knowledge-based potential energy functions use statistical data parsed from solved protein structures in order to determine the probability of seeing a particular structural element. These probabilities are converted into potential energies by use of the Boltzmann equation, $\Delta G = -RT\ln(p_{obs}/p_{exp})$, where p_{obs}/p_{exp} is the ratio of the probability of observing a particular structural element in the structural database versus the probability of observing that structural element by chance[154-156]. These

empirical potential energy functions are used to model structural behavior where an accurate mathematical model of the behavior does not exist or is too computationally complex to be used feasibly for protein design. As an example Kortemme and Morozov developed a statistically derived potential energy function to model hydrogen bonding which was used in the Baker lab's RosettaDesign program[157,158].

The completed energy function is comprised of a combination of these terms. Each term within the energy function is assigned a weight, which is then either calibrated by training the energy function or manually adjusted. Training typically involves testing the program's ability to recover wild-type amino acid sequence information based solely on backbone coordinates[159]. Weights may also be adjusted manually in order to change the impact each term has on sequence prediction. A common example of adjusting the weights is to "soften" the repulsive component of the Lennard-Jones potential. This allows the acceptance of small steric clashes by the energy function which often occur when rotameric representation of side-chains are used. These clashes can latter be alleviated by minimization steps which essentially model the ability of the protein core to accommodate a different side-chain.

Search Algorithms

If you were to graph the energy function values for every possible solution in sequence and torsional space for a given protein structure you would form an *N*-dimensional energy landscape of hills and valleys much like the three dimensional

representations shown in Figure 3.3. On this landscape, valleys represent a decrease in energy and thus an increase in sequence/structure fitness while hills represent an increase in energy and thus a decrease in fitness. The deepest valley is known as the global energy minimum, which corresponds to the sequence representing the best possible solution for a given structure based on the energy function. Since evaluating the energy function for every single possible sequence or backbone orientation would be impossible computational protein design programs instead employ a search algorithms as a means of exploring this energy landscape. Search algorithms can either be run to select for a given trait by finding the global energy minimum in what is known as positive design, or they can be run to select against a given trait by finding hilltops in the energy landscape in what is known as negative design. These search algorithms can be split into two basic types: sampling algorithms and pruning algorithms.

Sampling algorithms sample sequence space semi-randomly making random changes that are then evaluated by the energy function and selected based on their apparent fitness. In the case of protein design these changes can be changes in amino-acid type, change in side-chain torsions (rotamers) or changes in backbone torsions. These algorithms are capable of rapid searching; however, there is no guarantee that their sampling will find the global energy minimum based on the energy function since their search is not exhaustive. Confirmation of a global energy minimum can be obtained, however, by testing to see if multiple independent runs converge to the same

solution. Search algorithms that fall into this category include Monte Carlo and genetic algorithms.

Monte Carlo search algorithms iterate to a random position within the protein structure and make a random change, which could be the change of a side-chain or the rotation of a torsion angle. The change is then evaluated by the energy function and either kept or discarded based on a simple rule. The most commonly used rule is the Metropolis criterion[160] by which the change is accepted if the energy is lower and either rejected or accepted if the energy is higher based on a Boltzmann function. Accepting lower energy confirmations is akin to sliding down into the bottom of a valley in the energy landscape while accepting higher energy confirmations is akin to jumping up the hill of a valley. These jumps allow for the escape from local energy minima in order to continue searching for the global minimum.

Genetic algorithms use an evolutionary approach to select favorable mutations in a population of sequences using the energy function as a measure of fitness[161]. The algorithm starts with a population consisting of N random sequences each with a score based on the energy function. In its first iteration these sequences are then randomly mutated, re-scored and ranked based on their score. The top M are then chosen for mating where M is a variable optimized by the user. Mated sequences produce offspring that inherent sequence information from the multiple parent sequences with equal probability. These offspring then compete by selecting C sequences from the offspring population, scoring them based on the energy function and allowing the top

scoring sequence to pass to the next generation. This process is repeated *N* times to produce a new population which begins the next round of mating. This entire cycle of mutation, mating, and competition is repeated with each round improving the fitness of the sequence based on the energy function until an equilibrium is reached.

The second types of search engines are pruning algorithms, which emulate an exhaustive search by simplifying sequence and structural space. This is achieved by the systematic removal of bad side-chain conformations and combinations until a single solution remains. Assuming the pruning algorithm converges to a single solution this solution is guaranteed to be the global energy minimum solution for that structure based on the energy function. The drawbacks to their use is that they require that the amino acid side chains are represented as discrete rotamers and that the energy function used is written as a sum of individual and pairwise energy terms. The most commonly used pruning criteria for computational protein design has been the Dead End Elimination (DEE) theorum[162-164].

DEE algorithms operate by pruning sequence space based on the pairwise comparison of the energies of discrete side-chain rotamers[164]. If rotamer *A* at positon *i* (i_A) is higher in energy than rotamer *B* at the same position *i* (i_B) then i_A is removed from the set of possible solutions. This elimination procedure also includes pairs of rotamers where if i_A+j_B when combined are higher energy than the combined rotamers i_C+j_D then the pair i_A+j_B are removed from the solution set although the singe rotamers

i_A and j_B may still appear individually. This elimination process continues until the solution set converges to the global minimum energy sequence.

Example of a Computational Design Program: the RosettaDesign Program

One particularly successful computational design program has been the RosettaDesign program developed in the Baker lab at the University of Washington. RosettaDesign uses a stochastic Metropolis Monte Carlo search algorithm[161,165] to rapidly sample sequence space. The program starts with a fixed-backbone template derived from a crystal structure where all the side-chains within the design region have been stripped and replaced by random amino acids in one of many potential orientations, known as rotamers, from the Dunbrack rotamer library[153]. The replacement of the native sequence with a random one ensures that the only structural bias in the input is from the backbone and from the side-chains outside of the designed area. The fitness of this sequence based on the structure is then evaluated by the energy function, which returns a numeric energy value.

The energy function for the RosettaDesign program[159] is a linear combination of nine terms. Two terms represent the attraction (E_{atr}) and repulsion (E_{rep}) between atoms modeled using a standard Lennard-Jones 12-6 potential. Van der waals radii are extrapolated by fitting the Lennard-Jones potential to the distribution of distances observed between atom types in the PDB with well depths provided by the CHARMM19 parameter set[166]. Another term represents the internal strain of a

particular rotamer or clashes between that rotamer and its backbone (E_{intra}). A Lazaridis-Karplus implicit solvation model[167] is used to compute solvation energies (E_{solv}), excluding the intrinsic amino acid solvation term which is incorporated into the amino acid reference energy. Two terms representing the backbone-dependent free energies are determined empirically from probabilities based on how often a particular amino acid (E_{aa}) or a particular rotamer from the Dunbrack rotamer library (E_{dun}) appears in the PDB given the particular phi-psi angles of the backbone at that position. Another term approximates electrostatic interactions (E_{pair}) in proteins derived empirically from PDB statistics based on the probability of seeing two amino-acids close to each other in space given their local environment[168]. An orientation dependent hydrogen bonding potential (E_{hbnd}) derived empirically by examining hydrogen bond geometries in the PDB based on their local environment[157]. Finally reference values (E_{ref}) for each amino acid are included that approximate the free energies of each amino acid in the denatured state. Each term is coupled with a weight that has been optimized along with the reference energies based on an original training set of 30 proteins[159] and on experimental feedback. Summing the individual terms modified by their weights and subtracting the reference energy of the amino acid chosen at that position yields a final calculated energy for that residue position.

CONCLUSIONS

Both directed evolution and computational protein design have been used successfully in the engineering of a wide variety of protein biophysical properties. Each method has its own strengths and weaknesses. Directed evolution may be used without any structural knowledge of the protein design target but is restricted by the cost and difficulty of running accurate and efficient screening or selection methods. Computational protein design methods, however, can screen vast numbers of variants via an *in silico* energy function and search algorithm in a very short period of time with no associated cost though their reliance on both structural knowledge of the design target and our imperfect understanding of protein structure/function relationships has limited their application.

Many of the strengths and weaknesses of these two protein engineering strategies balance one another. Directed evolution methods may find beneficial mutations that would be difficult to predict based on structural data and our current understanding of protein biochemistry, while computational design techniques can search a much larger and more diverse section sequence space for a given structure. Using both methodologies on a single design problem could vastly increase the chance of success while increasing our understanding of protein design[169].

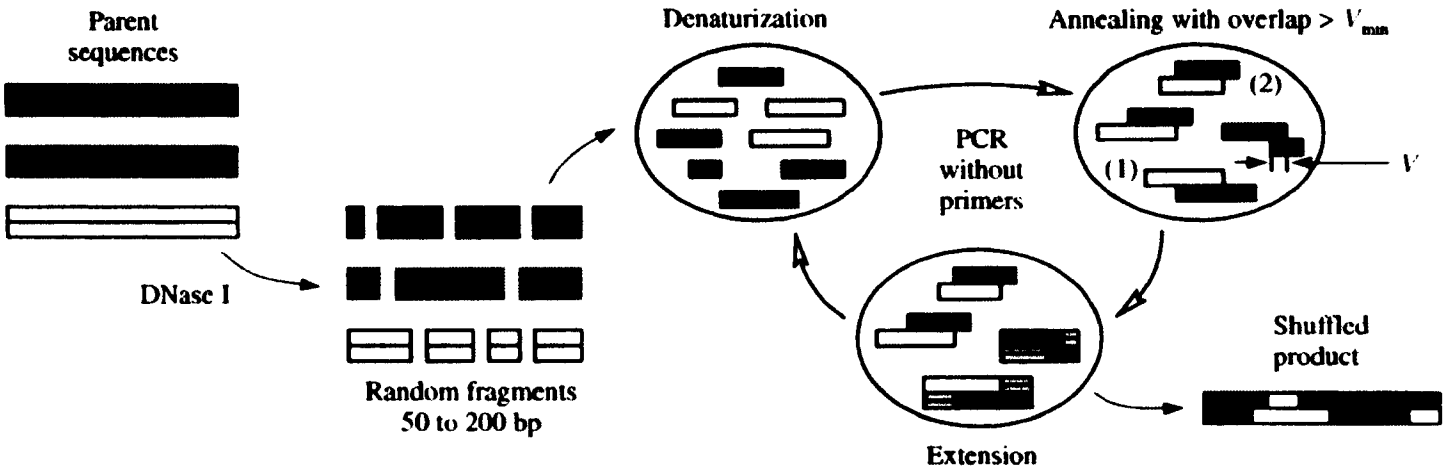


Figure 3.1 DNA Shuffling (adapted from Moore and Maranas, 2000[161]). DNA shuffling is a three step process. First a pool of sequences homologous to one another is digested into fragments by DNase I. Next these fragments are denatured by heat into single strands. Finally these fragments are reannealed by cooling self-priming their own extension by DNA polymerase. These leads to a final shuffled gene product comprised of randomized fragments from the original pool of homologous genes.

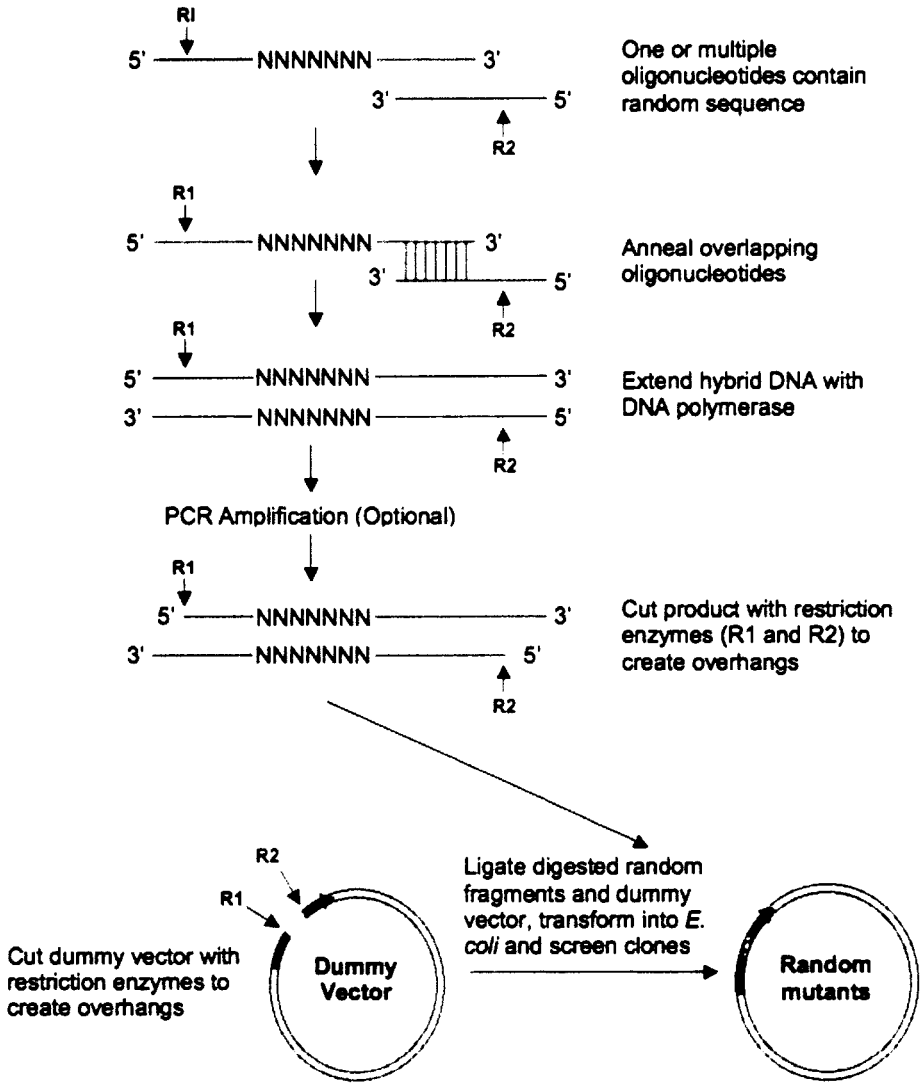


Figure 3.2 Regiospecific Random Mutagenesis (adapted from Kurtz and Black, 2004)[171]. Regiospecific random mutagenesis is a method used for randomly mutagenizing a specific region of a gene. First oligos are created with some sequence homology to the gene region of interest as well as regions of randomly sequence. These oligos are then primed with primer containing a restriction site R2 to form double stranded DNA with 5' R1 and R2 restriction sites. These double stranded DNA fragments are then digested using the R1 and R2 restricition enzymes and ligated into an expression vector with the homologous sequence removed by the same digestion.

Figure 3.3 Potential Energy Landscapes. The energy landscape for a given protein can be thought of as *N*-dimensional landscape representing the potential energy on one axis and all possible sequences and conformations. Above is a simplified three-dimensional version of a potential energy landscape where N = 3. The global minimum energy in these landscapes are represented by the deepest valley while other valleys represent metastable local minima. Computational protein design search algorithms are used to locate the global minimum energy for all of sequence or conformational space for a given protein structure.

Chapter Four: Computational Thermostabilization of an Enzyme

SUMMARY

Thermostabilizing an enzyme while maintaining its activity for industrial or biomedical applications can be difficult with traditional selection methods. We describe a rapid computational approach that identified three mutations within a model enzyme that produced a 10°C increase in apparent melting temperature T_m and a 30-fold increase in half-life at 50°C, with no reduction in catalytic efficiency. The effects of the mutations were synergistic, giving an increase in excess of the sum of their individual effects. The redesigned enzyme induced an increased, temperature-dependent bacterial growth rate under conditions that required its activity, thereby coupling molecular and metabolic engineering.

INTRODUCTION

Enzymes are the most efficient catalysts of chemical reactions known, enhancing reaction rates by as much as 23 orders of magnitude[172,173]. However, there has been little evolutionary pressure for them to become more thermostable than is required by their native environment. Many studies indicate that enzymes (like most proteins) exhibit closely balanced free energy profiles for folding and unfolding, thereby allowing functionally important dynamic motions and appropriate degradation

in vivo[174]. However, in a laboratory or industrial setting, this lack of thermostability can lead to undesirable loss of activity[175].

The physical principles of protein folding that result in a balance of stability and flexibility, while maintaining function, are not perfectly understood and have been difficult to exploit for the development of thermostabilized enzymes[175]. For hyperthermophiles, selective pressures have generated proteins with denaturation temperatures upwards of 110°C[176]. Their proteins exhibit topologies and stabilizing interactions similar to those from mesophilic and thermophilic organisms[177,178], leading to diverse hypotheses regarding their relative behaviors[179]. However, a key mechanism for thermostabilization appears to be the optimization of interactions between amino acids within a protein's core[176], complementing computational design methods that optimize a sequence for a given fold[130,131,135,136,180].

The thermostabilization of an enzyme presents additional challenges for computational protein design methods, because the active-site substrate geometry and the molecular dynamic behavior during an enzymatic reaction often appear fine-tuned for maximum catalytic efficiency[173,174]. Therefore, the design method must be capable of predicting thermostabilizing mutations within a given fold while minimizing any shift in the backbone that might structurally disrupt the active-site structure or quench its flexibility.

In the past several years, methods for computational protein structure prediction and design have improved substantially[130-132]. Recently, computational

design has been used successfully in thermostabilizing noncatalytic proteins[137-140], redesigning binding pockets[142-146], creating a protein fold[148], and designing catalytic activity into a bacterial receptor[149]. We use the program RosettaDesign[159], which uses an energy function for evaluating the fitness of a particular sequence for a given fold and a Metropolis Monte Carlo search algorithm for sampling sequence space. The program requires a backbone structure as input and generates sequences that have the lowest energy for that fold.

We picked the homodimeric hydrolase enzyme yeast cytosine deaminase (yCD), which converts cytosine to uracil, as a target for computational thermostabilization. yCD was chosen because its high-resolution crystal structure is available[181], its catalytic mechanism is well characterized[181], it is thermolabile[13,182], and it has potential use in antitumor suicide gene applications[13,100,181,183]. As do many commercially useful enzymes, yCD displays irreversible unfolding behavior at high temperatures (presumably because of aggregation) rather than the more simple, fully reversible behavior common among model systems for the study of protein folding. The problems inherent in engineering such catalysts have been recently reviewed[175]. We used computational redesign to predict a series of point mutations in the enzyme core that might lead to thermostabilization of the enzyme without losing catalytic efficiency. We then prepared a series of designed enzyme variants and determined their folded thermostability, catalytic behavior, ability to complement metabolic cytosine deaminase activity, and three-dimensional crystal structures.

RESULTS

Our general computational strategy was largely unchanged from that described by Kuhlman and Baker[159]. An energy function evaluated target sequences threaded onto a template backbone[135,136,159,167]. Sequence space was searched with an iterative Metropolis Monte Carlo procedure, starting with a random sequence, replacing a single amino acid rotamer with a rotamer from the Dunbrack backbone-dependent rotamer library[153], and reevaluating the energy. Sequences with lower energy were automatically adopted, whereas sequences with higher energy were accepted with a probability based on the Metropolis criterion in order to prevent trapping in a local energy minimum.

All residues directly involved in catalysis, those located within 4 Å of the active site, and those involved in the dimer interface were held fixed (Figure 4.1A). The remaining 65 residues of the 153-residue monomer were included in the redesign, allowing them to be changed to any amino acid except cysteine. Thirty-three of the 65 residues subjected to redesign (49%) remained wild-type, a result similar to those of prior applications[140,159]. Sixteen of the point mutations suggested by the program were located on the surface of the protein and were not pursued, whereas the remainder were in the core. The core mutations could be further subdivided into two localized clusters of interacting residues, as well as four additional isolated point mutations (Figure 4.1B).

Site-directed mutagenesis was used to generate each of the two complete

clusters of point mutations and the four individual mutations described above. Cluster 1, consisting of nine simultaneous mutations packed between an helix and several ß-strands (including replacement of a buried salt-bridge), aggregated at concentrations above 0.4 mg/mL and was not characterized further. Cluster 2, consisting of four mutations packed between two helices, remained soluble when concentrated to 20 mg/mL. Individual mutations from this cluster revealed that A23L and I140L were key to the thermostabilization of the enzyme and were included in the final construct described below. Of the remaining four individual mutations, one (V108I) was also incorporated in the thermostabilized triple-mutant enzyme, whereas the remaining three (W10T, T67E, and E69L) were not as well behaved and were not characterized further. Both the double mutant (A23L/I140L) and the final triple mutant (A23L/I140L/V108I) were well behaved during expression and purification, more thermostable than the wild-type enzyme, and fully active (Figure 4.1B).

We performed thermal denaturation experiments on all constructs using circular dichroism (CD) spectroscopy (Figure 4.2A). Wild-type yCD and the redesigned mutants displayed largely reversible unfolding behavior over the range of temperatures examined; however, at higher temperatures, they unfolded irreversibly (data not shown). We quantified the thermal stability of yCD and the mutant constructs by deriving an apparent melting temperature (T_m) from the CD-unfolding curves. This value for the wild-type enzyme was determined to be 52°C. The isolated single mutations A23L, I140L, and V108I each slightly thermostabilized the enzyme,

increasing the apparent T_m by 2°C. However, simultaneous incorporation of all three mutations increased apparent T_m to 62°C, 10°C higher than that of the wild type. Therefore, combination of individual point mutations in a single construct produced a synergistic effect beyond their individual contributions. This result is not simply due to the formation of contacts between redesigned residues, because residue 108 was physically separated from residues 23 and 140.

The kinetic behavior of the wild-type enzyme and the double and triple mutants was measured at 22°C to determine the effects of the mutations (Table 4.1 and Figure 4.3), as were their relative activities as a function of temperature (Figure 4.4). At 22°C, the wild-type enzyme displays a turnover (k_{cat}) of 160 mol (mol enzyme)$^{-1}$ s^{-1} and a Michaelis constant K_m of 1.98 mM, and the double and triple mutants displayed a slightly reduced maximum rate V_{max} coupled with a reduction in the K_m. The catalytic efficiency of the enzyme mutants (expressed as the ratio k_{cat}/K_m) was unchanged relative to the wild-type enzyme. The overall temperature activity profile was broadened, for the redesigned enzyme, with near–wild-type activity retained at lower temperatures and higher activity above 50°C.

The preservation of overall catalytic efficiency (achieved by reducing both k_{cat} and K_m, rather than by maximizing overall velocity) and the unusual change in shape and breadth of the enzyme's thermal profile might suggest that the computational redesign generated mutations that natural or directed evolution pathways might not select, except perhaps as intermediate species. Therefore, computational strategies for

thermostabilization might offer a bonus of selecting mutations that differ in these properties, as compared to selection or redesign experiments based on natural selection.

In order to visualize the time-dependent decay of activity at elevated temperatures, wild-type yCD and the double and triple mutants were incubated at 50°C, and the decrease in their relative activity was monitored over time (Figure 4.2B). The wild-type enzyme showed a rapid loss of activity at 50°C, with a half-life of 4 hours. The double mutant displayed a half-life of 21 hours, whereas the triple mutant had a half-life at 50°C of 117 hours (a 30-fold increase over that of the wild type).

In order to determine the effects of the mutations in vivo, a strain of *Escherichia coli* dependent on cytosine deaminase function for uracil synthesis was engineered and transformed with both wild-type and mutant yCD reading frames. Doubling times were then measured at 30°C and 37°C on minimal media lacking uracil (Figure 4.5). The thermostabilized mutant construct induced slightly accelerated growth relative to the wild-type enzyme at 30°C and a clear acceleration at 37°C. This suggests that the properties of the engineered variants (a reduced K_M and thermostabilization) measured in vitro correlate with improved enzyme flux in vivo under growth conditions limited by the activity of the enzyme.

The crystal structures of both the double and triple mutants were solved to 1.9 Å and 1.7 Å, respectively. The interpretation of density around the redesigned regions

of the protein core (in unbiased omit maps) was unambiguous (Figure 4.6A). The root mean square deviation values comparing the wild-type enzyme and both constructs were under 0.5 Å on all common carbons and under 0.8 Å on all common atoms. Thus, the redesign and subsequent incorporation of point mutations in the enzyme core had a negligible effect on overall structure of the enzyme, including the active site (Figure 4.6B). The redesigned, mutated residues all appear to pack more tightly in the enzyme core, with more surface area in contact with neighboring residues without altering the nearby side chain rotamers or backbone conformation. Approximately 70 $Å^2$ of additional buried surface area is incorporated as a result of the three mutations [based on an analysis of residue-by-residue packing, using the program NACCESS[184]]. The A23L/I140L double mutation increased the amount of hydrophobic packing against a neighboring tyrosine ring (Figure 4.6C), and the addition of V108I in the triple mutant added an additional methyl group to fill a cavity (Figure 4.6D).

DISCUSSION

The stabilized triple mutant was pieced together from part of a cluster of mutations predicted by the program and another single mutation predicted in a separate part of the core. Although the degree of thermostabilization produced by these mutations was relatively modest (an increase for T_m of 2°C for the first change and 4°C for each subsequent mutation), there is no obvious reason why additional mutations predicted by the program could not be iteratively incorporated into the enzyme core,

resulting in a panel of catalysts that display sequential increases in thermal stability.

Not all mutations predicted by the program were equally thermostabilizing. Redesigns involving incorporation or alteration of polar or charged residues in the core (such as replacement of a buried salt-bridge in cluster 1 and individual mutations T67E, E69L, and W10T) were less successful than mutations involving substitution of one hydrophobic side chain for another. These latter mutations were predicted and observed to fill cavities within the core with additional van der Waals packing interactions. In future design efforts, selecting mutations of this type in silico may be most successful. Furthermore, modeling of interactions involving buried polar and charged side chains in the enzyme core is an area for future development in computational redesign algorithms.

MATERIALS AND METHODS

Computational Design Method

RosettaDesign was used as described in Kuhlman and Baker, 2000. The program uses an energy function to evaluate target sequences threaded onto a template backbone[135,136,159,167]. Sequence space is searched with an iterative Metropolis Monte Carlo procedure starting with a random sequence, replacing a single amino acid rotamer with a rotamer from the Dunbrack backbone dependent rotamer library[153] and reevaluating the energy using the energy function. Sequences with lower energy are automatically kept while sequences with higher energy are accepted with a probability

based on the Metropolis criterion in order to prevent trapping in a local energy minimum. Multiple independent runs are done to confirm convergence onto a global minimum energy sequence.

Design was carried out in three rounds. The first two rounds were similar to that described in Dantas *et. al.*, 2003 with the exception that residues subject to redesign were limited to those not located in or nearby the active site or dimer interface; incorporation of cysteine was avoided. The 1.14 Å crystal structure of yCD bound to the mechanism-based inhibitor dihydropyridine (pdb code 1p6o) was used as a structural template with residues 3-5 and 198-205, bound inhibitor, waters and metals removed. Round 1 of the design allowed residues 10-28, 37-49, 67-70, 81-87, 105, 107-109, 128-136 and 138-146 to change (Figure 4.1A) and residues on both monomers of the dimer were allowed to change independently during the design. The search used a limited side chain conformational database that contained rotameric models varied around the first chi angle only in order to save computational expense. 50 independent runs were done generating 100 sequences (50 for each monomer). Round 2 limited the search to amino acids that were picked during the 50 runs in round one at the individual residue positions. This limitation allowed a larger rotamer library including chi-2 angle rotations to be used. Again both monomers were allowed to change independently and 100 sequences were generated. A third round of design was necessary because RosettaDesign produced two different sequences for the monomers in the dimeric structure, possibly due to some asymmetry in the template

structure itself (all atom rmsd between monomers of 0.74 Å). To determine which sequence was better overall the sequences of the monomers were evaluated in the dimeric form with the same sequence used for both monomers. The sequence that produced the lowest energy was then chosen (Figure 4.1C). The mutations that successfully result in thermostabilization appeared to generally correspond to those positions that increase the size of a hydrophobic residue without changing its chemical characteristics.

Protein Expression, Purification and Mutagenesis

Protein expression and purification of yCD was carried out as previously described[181] with the exception that buffer exchange was carried out with Biorad pre-packed Econo-Pac 10DG buffer exchange columns rather than overnight dialysis. The protein was expressed from a pET15b vector with a thrombin cleavable 6-His tag within BL21-RIL cells. All kinetics and thermostability experiments were carried out with fresh, unfrozen protein stored at 4 degrees C for no longer than 2 weeks due to observed adverse effects of freezing on protein stability (data not shown). Mutations were made with the Stratagen QuickChange ® Site-Directed Mutagenesis Kit. Mutated DNA was transformed into XL-10 Ultracompetant cells and a Qiagen miniprep kit was used to purify the plasmid from a culture of the cells. All sequences were confirmed using the dye-termination method (data not shown).

Circular Dichroism Measurement and Thermal Denaturation Experiments

CD data were collected on an Aviv 62A DS spectrometer as described in Dantas *et. al.*, 2003[140]. Wavelength scans were run from 200-260 nm to determine the folded state of the protein, the ratio of concentration to signal strength and the wavelength where signal strength was at its highest (Figure 4.2A). Temperature melts were run with 8-12uM protein in a 2 mm pathlength cuvette from 10° to 98° C in 2° increments with temperature regulated by a Peltier device. Sample temperature was allowed to equilibrate for 30 seconds before measurement. Chemical melts were run with 1.2uM protein in a 1 cm pathlength cuvette from 0 to 4M guanidine hydrochloride stepping in 0.2M increments with 0.1M increments in the transition range. In both temperature and chemical melts samples were allowed 30 seconds to equilibrate and signal was collected and averaged over 30 seconds. Denaturation was recorded as a change in elipticity over temperature. Apparent T_m's were determined by curve-fitting (Figure 4.2A).

Chemical denaturation experiments measuring CD signal at 220 nm over a range of guanidine hydrochloride concentrations were performed, however the irreversible unfolding of wild-type yCD and the mutant constructs in these experiments prevented an accurate determination of the change in free energy of folding (data not shown).

Activity Assays

The conversion of cytosine to uracil by yCD was measured by UV spectroscopy by monitoring change in absorbance at 286 nM. The protein was diluted to 2 uM in 50 mM Tris pH 7.5 and mixed 1:1 with a range of 9 cytosine concentrations from 0.2 - 1mM in the same buffer. Absorbance at 286 nm was measured every 5 seconds until baseline was reached with the first reading taken 5 seconds after mixing. Measurements were taken in quadruplicate and averaged to reduce error. Initial velocity was calculated as a function of the initial slope by curve-fitting the resulting plot, taking the derivative and extrapolating back to time zero. K_m, V_{max} and k_{cat} of wild-type yCD and mutant constructs were determined from a double reciprocal (Lineweaver-Burke) plot of the resulting data and the catalytic efficiency k_{cat}/K_m was calculated from these values (Table 4.1, Figure 4.3).

The thermal activity profiles for the wild-type and the triple mutant were created by measuring enzyme activity from 5 to 60 degrees at a fixed substrate concentration of 0.70 mM (Figure 4.4). Data was collected and analyzed as above with temperature controlled by a Peltier device.

Decay of enzyme activity over time at 50° C for wild-type and the double and triple mutants was monitored by incubating 30uM protein stock in a thermocycler at 50° C and measuring activity at a series of time-points (Figure 4.2B). Activity was measured in the same manner as above with the exception that runs were done in triplicate.

Complementation and Bacterial Growth Assays

Strain GIA38 of *Escherichia coli*, deficient both in CD activity and in *de novo* uracil biosynthesis (*thr*⁻ *dad*B3 *leu*B6 *fhu*A21 *cod*A1 *lac*Y1 *tsx*-95 *gln*V44(AS) l⁻ *pyr*F101 *his*-108 *arg*G6 *ilv*A634 *thi*-1 *deo*C1 *glt*-15) was obtained from the *E. coli* Genetic Stock Center (CGSC #5594). This strain was lysogenized with DE3 according to the instructions from Novagen (Madison, WI). The derived strain GIA39(DE3) was used in the genetic complementation bacterial growth curve assays (Figure 4.5). The strain was transformed with an inducible pET-15b expression vector containing either wild-type or triple-mutant yCD and grown in media that is selective for CD activity: 1.96g yeast synthetic dropout without uracil, 0.1mM $CaCl_2$, 0.24mM (or 0.267mg/ml) cytosine, 111ml M9 cocktail (100ml 10 x M9 salts (30g KH_2PO_4, 67.8g Na_2HPO_4, 5g NaCl, 10g NH_4Cl per liter), 1ml 1M $MgSO_4$, 5ml 20% glucose per liter) and 50μg/ml carbenicillin per liter. As a control, growth of the same strain transformed with a null pET-15b vector was also measured and found to be negligible. Growth curves were measured in several independent experiments for each transformed strain at both 30° and 37° C.

Crystal Growth and Structure Determination

Crystals were grown using methodology from Ireton[185]. After two days crystals appeared in the drops, but were often twinned. These crystals were then used

to streak seed into the clear lower concentration drops where higher quality crystals eventually grew. Crystals were looped out and transferred briefly to a cryobuffer containing the motherliquor plus 25% DMSO and then flash-frozen in liquid nitrogen. Crystals diffracted to 1.9 Å for the double mutant and 1.7 Å for the triple mutant (Table 4.2).

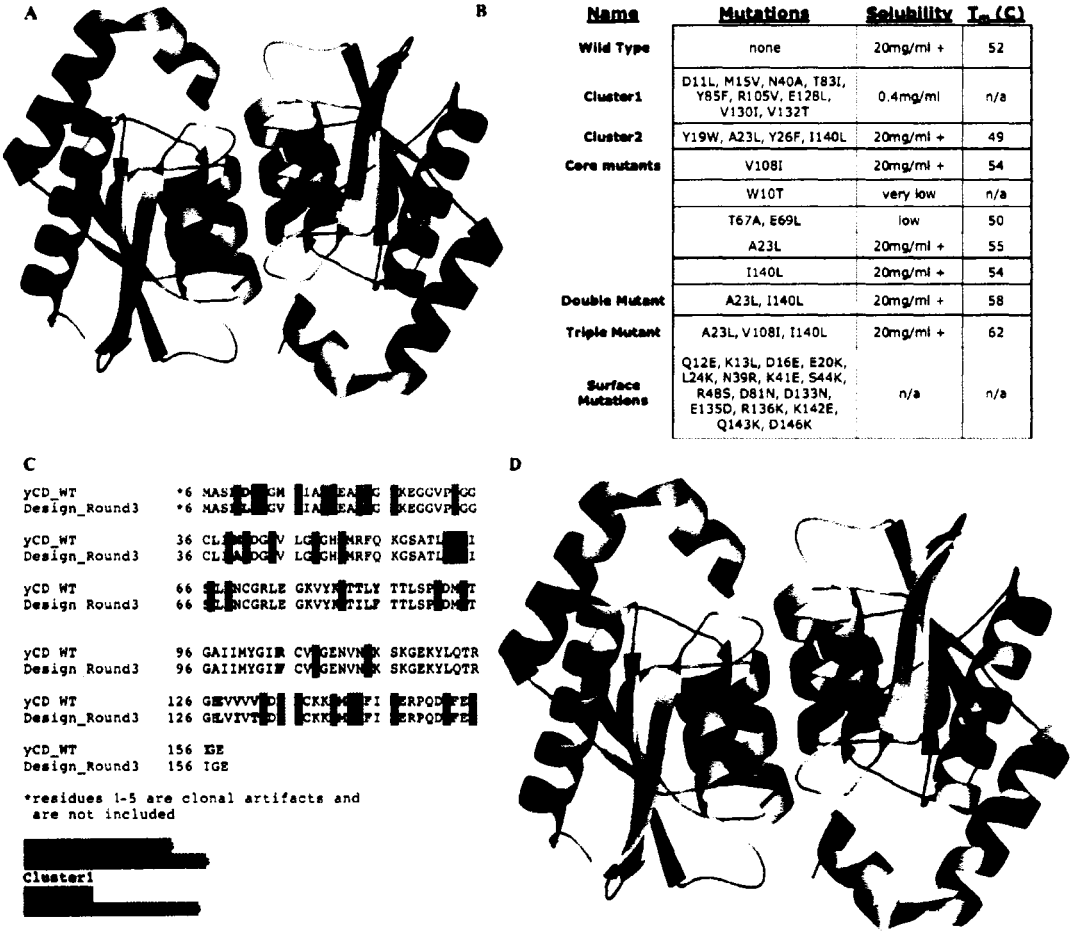
All data were collected on an RAXIS-IV area detector using a rotating anode generator. Both crystals diffracted strongly but exhibited a faint minor twinned lattice, which did not appear to affect data quality or subsequent structure refinement behavior. Data was indexed and scaled using the DENZO/SCALEPACK program suite[186]. The initial structures were generated by molecular replacement using the program EPMR[187] using the wild-type yCD crystal structure[181] with the mutated residues and any residues within a 4 Å shell of those residues truncated to alanine. Refinement was carried out in rounds using the program CNS[188,189] with a random 10% of data withheld for cross-validation. The double mutant crystal structure was refined to a R_{work}/R_{free} of 17.7/21.4 and the triple mutant structure was refined to a R_{work}/R_{free} of 16.7/19.4 (Table 4.2). Electron density maps and models were visualized in XtalView[190] (Figure 4.6A) while cartoon representations of the structure were generated using Pymol[191] (Figure 4.6B-D). Coordinates and diffraction data for the double and triple redesign mutants have been deposited in the PDB database (ID codes 1YSD and 1YSB, respectively).

Table 4.1 Kinetics Measurements Wild-Type and Triple Yeast Cytosine Deaminase. Both the double and triple mutants displayed a slightly reduced V_{max}, coupled with a reduction in the Michaelis constant K_m. The catalytic efficiency of the various enzyme mutants (expressed as the ratio k_{cat}/K_m) was unchanged relative to the wild-type enzyme. M prod, molar concentration of product; enz, enzyme.

	Wild type	Double mutant	Triple mutant
K_m (mM)	1.98	1.50	1.33
V_{max} (M prod s^{-1})	0.00016	0.00012	0.00011
k_{cat} (M prod M enz^{-1} s^{-1})	160	120	110
k_{cat}/K_m (M enz^{-1} s^{-1})	80800	80000	82700

Table 4.2 Crystallographic Statistics for Double and Triple Mutant Yeast Cytosine Deaminase Structures

		Double Mutant A23L / I140L	Triple Mutant A23L / V108I / I140L
Data Collection			
Space Group		P2(1)	P2(1)
Unit Cell Dimensions		a=39.4 Å b = 54.8 Å c = 68.4 Å β = 105.7°	a=39.5 Å b = 54.6 Å c = 68.1 Å β = 105.3°
Resolution		1.9	1.7
Completeness (%) (outer shell 0.1 Å)		99.3 (96.3)	93.9 (86.1)
Rmerge (%)		4.3	4.7
Average I/o		38.8 (12.5)	19 (9.5)
Number of Observations		75,950	62,386
Number of Reflections (hkl)		22,096	29,093
Redundancy		3.4	2.2
Refinement			
Resolution		50-1.9	50-1.7
Reflections		21965	28869
Test set (10%)		2182	2874
Rcryst (%)		17.7	16.7
Rfree (%)		21.4	19.4
Rms deviations	Bonds	0.00564	0.003914
	Angles	1.16343	1.10045
Protein atoms		2448	2450
Water molecules		318	398
Bound metals		5	5
Average B factors	Protein	19.5	17.1



Name	Mutations	Solubility	T_m (C)
Wild Type	none	20mg/ml +	52
Cluster1	D11L, M15V, N40A, T83I, Y85F, R105V, E128L, V130I, V132T	0.4mg/ml	n/a
Cluster2	Y19W, A23L, Y26F, I140L	20mg/ml +	49
Core mutants	V108I	20mg/ml +	54
	W10T	very low	n/a
	T67A, E69L	low	50
	A23L	20mg/ml +	55
	I140L	20mg/ml +	54
Double Mutant	A23L, I140L	20mg/ml +	58
Triple Mutant	A23L, V108I, I140L	20mg/ml +	62
Surface Mutations	Q12E, K13L, D16E, E20K, L24K, N39R, K41E, S44K, R48S, D81N, D133N, E135D, R136K, K142E, Q143K, D146K	n/a	n/a

Figure 4.1 Scope of Computational Design (A) a cartoon representation of yCD with active site residues highlighted in purple and residues involved in the initial design highlighted in blue. **(B)** A table of all the constructs and mutations dealt with in this paper along with their solubility and T_m if measured. **(C)** A sequence alignment between yCD WT and the final 3rd round design with 33 mutations highlighted red for surface mutations, green for single mutations within the core, yellow for residues involved in Cluster1, cyan for residues involved in Cluster2 and pink for residues involved in the active site. **(D)** A cartoon representation of yCD, the color-coding matches that in the sequence alignment.

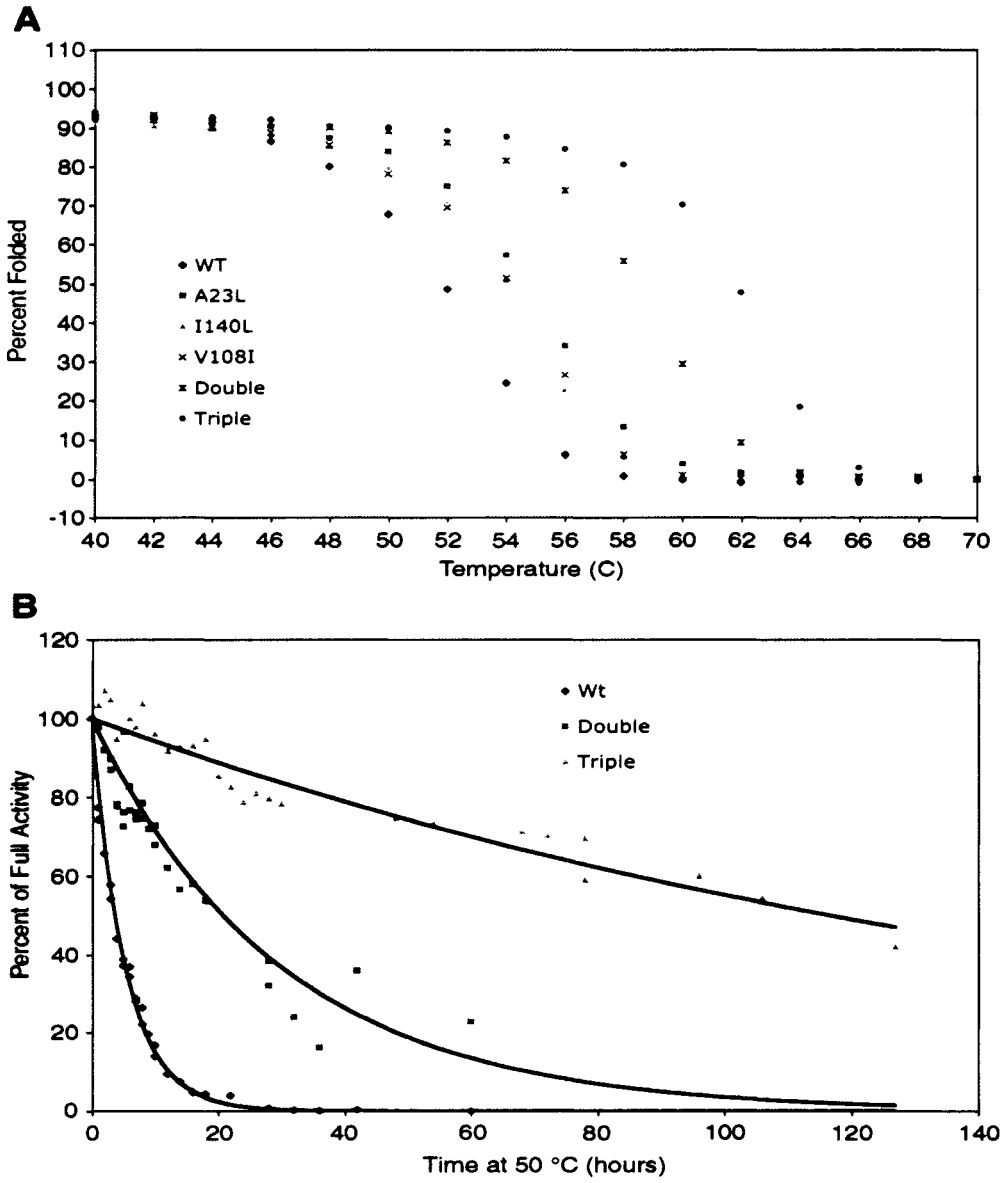


Figure 4.2 Thermal Denaturation and Activity Half-Life Measurements. (**A**) Temperature melt measuring the change in signal at 220 nm over a range of temperatures. All constructs show a folded baseline followed by a sigmodial two-state transition to an unfolded baseline. Only data from 40° to 70° are shown; at lower temperatures, the baseline plateaus corresponded to an assignment of 100% folded protein. (**B**) Activity decay at 50°C. Wild-type yCD and the double and triple mutant constructs were incubated at 50°C, and their activity was measured over time. The resulting curves gave half-lives for the enzymes at 50°C of 4 hours for the wild type (WT), 21 hours for the double mutant, and 117 hours for the triple mutant.

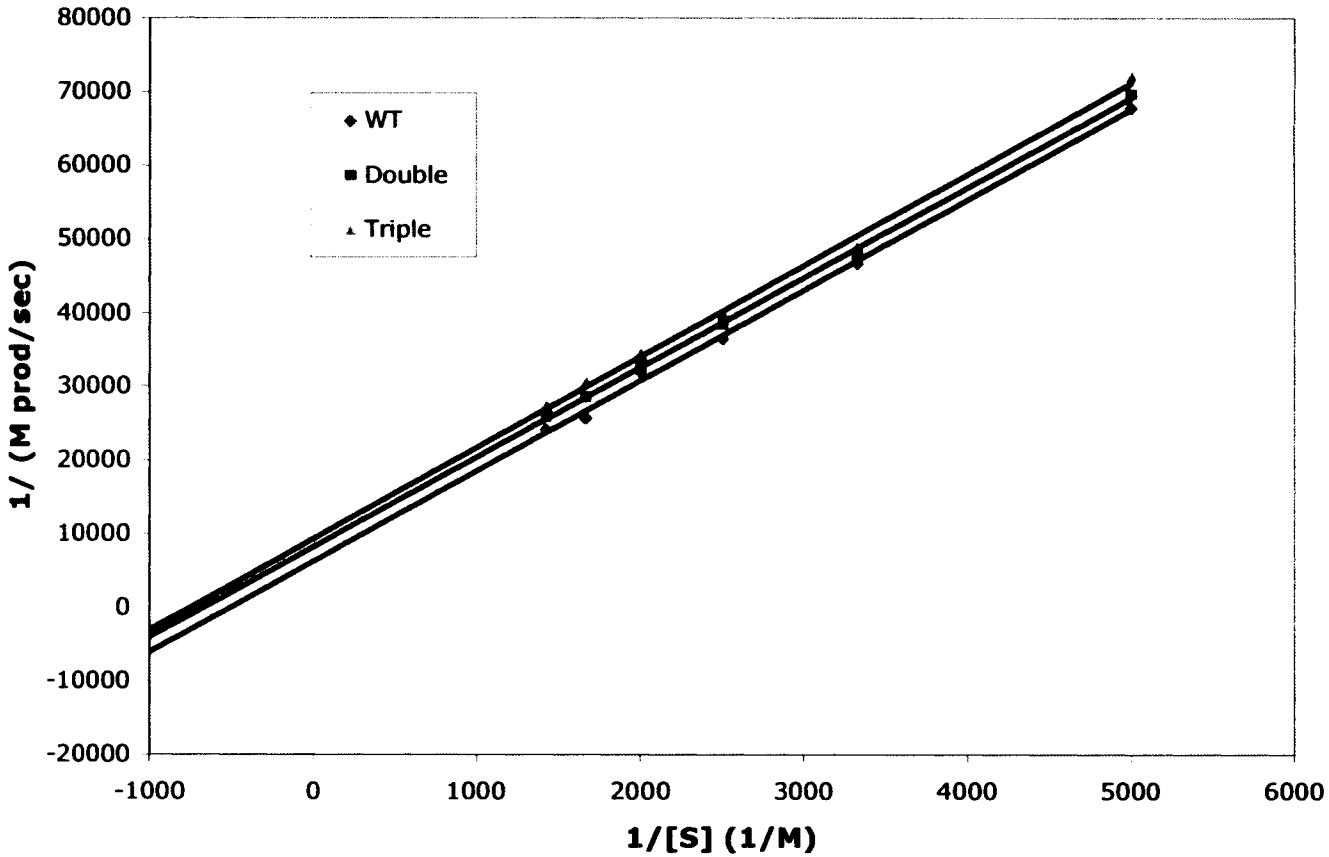


Figure 4.3 Yeast Cytosine Deaminase Enzyme Variants. The kinetic parameters k_{cat} and K_m of yCD wild-type and the double and triple mutants were measured as described in methods, and are plotted here in double reciprocal (Lineweaver-Burke) format. The turnover rate (k_{cat}) of the mutants displays slight quenching, counterbalanced by a slight reduction in apparent substrate affinity (K_m), yielding an overall catalytic efficiency (k_{cat}/K_m) that was very close to wild-type.

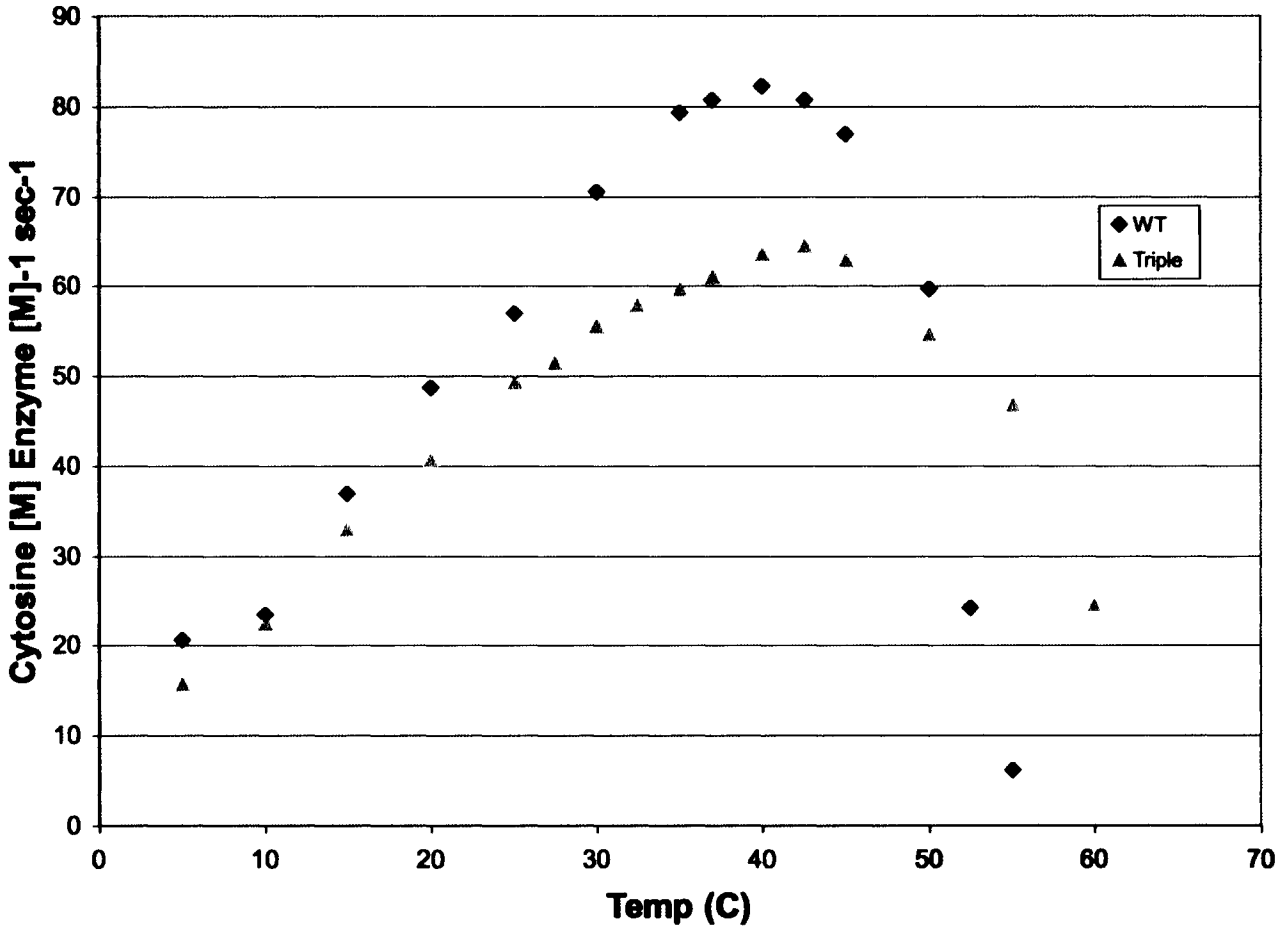


Figure 4.4 Yeast Cytosine Deaminase Wild-Type and Triple Mutant Thermal Activity Profile. Production of uracil product at variable temperatures in a single timepoint quenching experiment under saturating concentrations of cytosine, as described in the methods section.

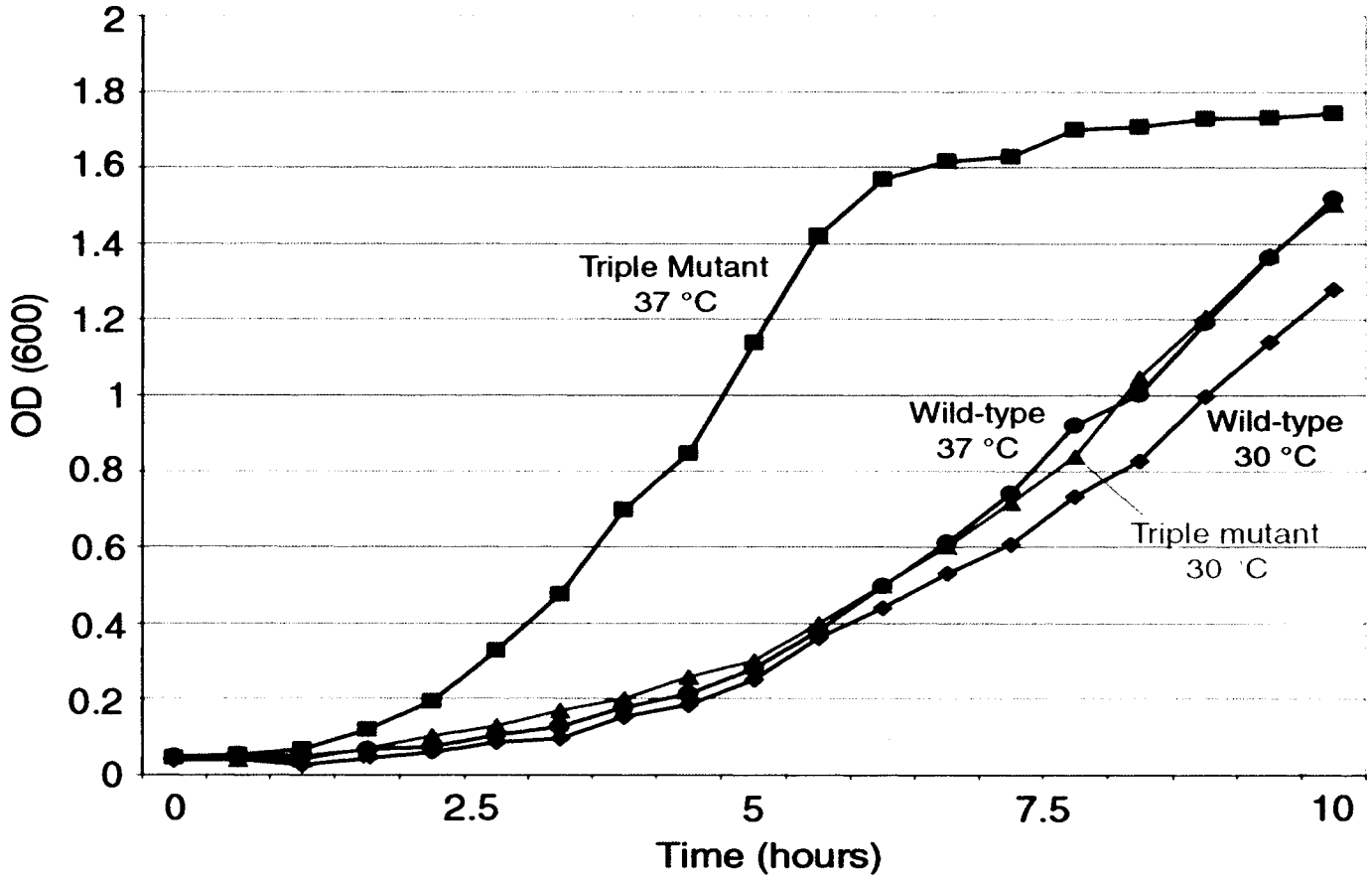


Figure 4.5 Assay for *in vivo* Metabolic Growth Phenotype. Bacterial growth curves in media conditions requiring cytosine deaminase activity for generation of uracil. Both wild-type and reengineered mutants of yCD complement the bacterial activity; the thermostabilized enzyme variant displayed a slight increase in growth rate at 30°C and a clear increase at 37°C. OD (600), optical density at 600 nm.

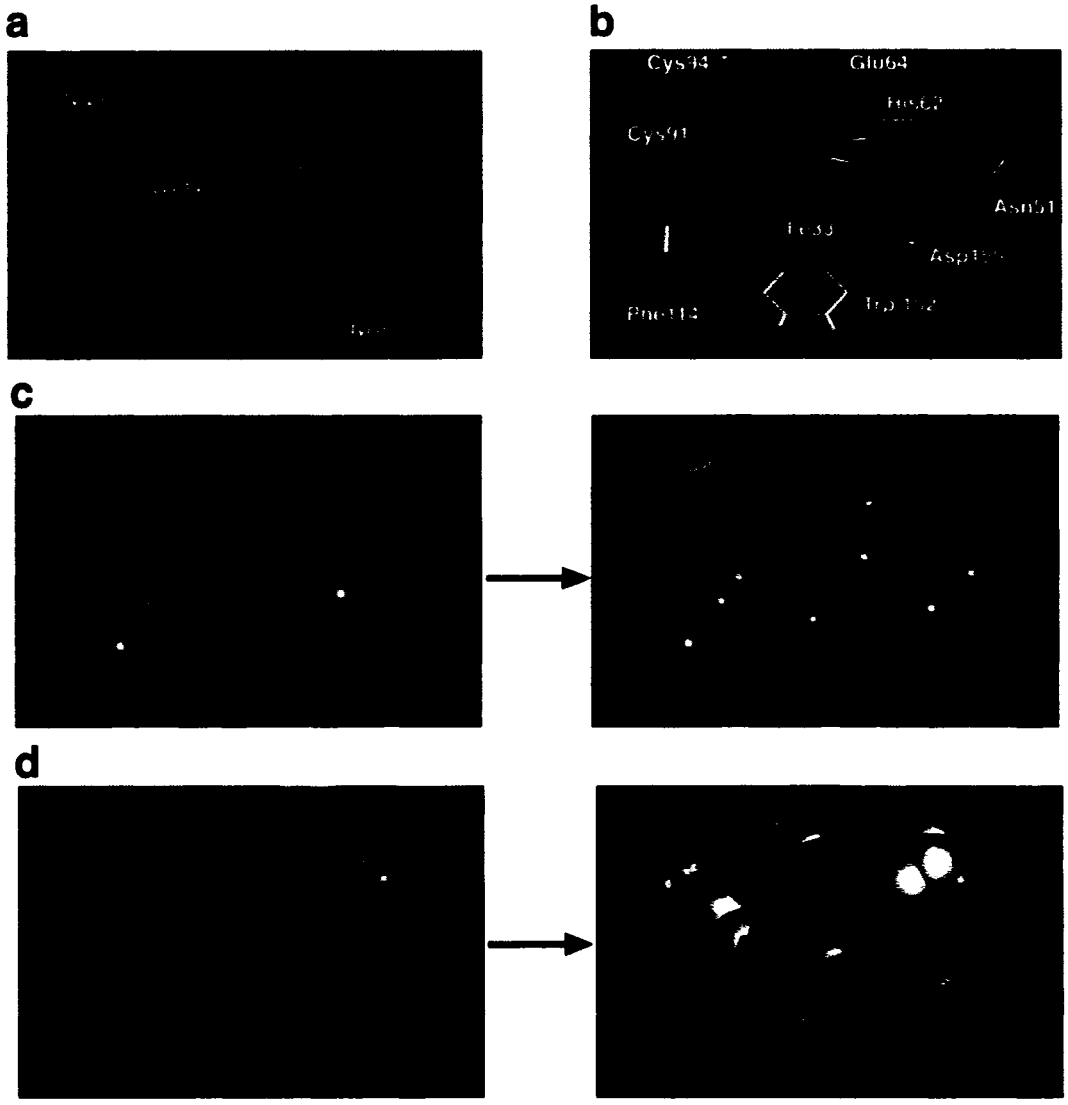


Figure 4.6 Structural Analyses (**A**) $2F_o$-F_c simularted annealing omit map of the final protein 'triple mutant' construct centered on the A23L / I140L mutations. (**B**) Overlay of the yCD active site residues with wild-type green and the triple mutant yellow. (**C**) Van der Waals representation of residues Y19, A23, Y26, and I140 in the wild-type yCD crystal structure (left) and the same representation and orientation for the mutant construct with A23L and I140L mutations (right). (**D**) Van der Waals radii representation of the area around V108 in the wild-type structure (left) and a similar representation of the area around the V108I mutation in the triple mutant crystal structure.

Chapter Five: Thermostabilized Yeast Cytosine Deaminase Variants Derived by Random Mutagenesis or Computational Design Increase Sensitivity to 5-Fluorocytosine For Prodrug Gene Therapy

SUMMARY

Prodrug gene therapy (PGT) is a treatment strategy in which tumor cells are transfected with a 'suicide' gene that encodes a metabolic enzyme capable of converting a nontoxic prodrug into a potent cytotoxin. One of the most promising PGT enzymes is cytosine deaminase (CD), a microbial salvage enzyme that converts cytosine to uracil. CD also converts 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU), an inhibitor of DNA synthesis and RNA function. Over 150 studies of cytosine deaminase-mediated PGT applications have been reported since 2000, all using wild-type enzymes. However, various forms of cytosine deaminase are limited by inefficient turnover of 5-FC and/or limited thermostability.

In a previous study we stabilized yeast cytosine deaminase (yCD) by repacking of its hydrophobic core at several positions distant from the active site[152]. Here we report that random mutagenesis of residues in and near the active site, followed by selection for enhanced sensitization to 5-FC, also yields a thermostabilized enzyme variant (yCD-D92E). The new mutation is located at the enzyme's dimer interface, indicating that independent mutational pathways can lead to stabilization. This result implies that protein stability is the main limiting property for yCD's performance in

prodrug gene therapy. Combinations of redesigned and selected substitutions further increase the enzyme's resistance to thermal denaturation. However, while each independently derived set of mutations significantly improves the enzyme's performance in PGT assays both in cell culture and in animal models, their combination is deleterious. Characterization of these constructs demonstrates that this might be due to a 'threshold' effect, a limiting thermostability beyond which enzymatic activity is reduced, restricting prodrug activation.

INTRODUCTION

The pyrimidine salvage pathway enzyme cytosine deaminase (CD; EC 3.5.4.1) is responsible for converting the nucleobase cytosine to uracil and ammonia. This activity is found primarily in microbes[192], including *Saccharomyces cerevisiae* and *Esherichia coli*, and has arisen at least twice, using completely separate protein folds[181,193]. In addition to cytosine, cytosine deaminase catalyzes the conversion of 5-fluorocytosine (5-FC) to the potent chemotherapeutic drug, 5-fluorouracil (5-FU). Thus, the combination of CD enzyme activity and 5-FC as its substrate forms the basis of a potential anti-tumor gene therapy, where CD plays the role of a 'suicide gene'[16].

In suicide gene therapy applications, the gene for cytosine deaminase is introduced into cancer cells, followed by systemic administrations of the prodrug 5-FC. Following deamination by CD, 5-FU is converted by cellular enzymes to 5FdUMP, an irreversible inhibitor of thymidylate synthase (TS). Inhibition of TS

blocks dTTP production and prevents DNA synthesis[194-196]. The CD/5-FC system has been used in numerous animal models and is currently being evaluated in clinical trials for solid tumors. One limitation to this approach is the poor transfection efficiency of current vector delivery systems. Consequently, high 5-FC doses must be administered to achieve therapeutic value and are associated with unwanted side effects suggested to be a result of the generation of 5-FU by intestinal bacteria[197].

Key to suicide gene therapy is the phenomenon known as the bystander effect in which non-transfected neighboring cells are killed through the transfer of antimetabolites from CD expressing cells in close proximity. A strong bystander effect has been associated with CD and 5-FC because 5-FU is a small, uncharged molecule capable of non-facilitated diffusion through cellular membranes[72,198-201]. Unlike other suicide gene therapy systems such as herpes simplex virus thymidine kinase (HSV-TK) and ganciclovir (GCV) that rely on transfer of metabolites through gap junctions, the CD/5-FC bystander effect is not dependent upon cell-to-cell contact[198].

Another advantage of the CD/5-FC combination is that 5-FU has radiosensitizing properties[73,202]. Since it is unlikely that treatment with gene therapy would be the only course of action in patients, radiosensitizing effects can augment treatment regimens. Several groups have reported *in vivo* results with CD to have a significant bystander effect[73,203] at clinically relevant 5-FC doses and radiation regimens[73,204,205].

Two completely separate forms of CD have evolved in nature, and both are being studied in anti-tumor gene therapy investigations. Yeast CD (yCD) belongs to the amidohydrolase protein fold family and shares homology with bacterial and eukaryotic cytidine deaminases[181]. The enzyme is assembled into a homodimer comprised of 17.5 kDa subunits that contain a catalytic zinc ion. In contrast, bacterial CD (bCD) belongs to the alpha-beta 'TIM' barrel fold family and most closely resembles human adenosine deaminase[193]. In *E. coli*, the enzyme is assembled into a hexamer of 60 kDa subunits that contain a catalytic iron.

Although product release from the yeast cytosine deaminase is rate-limiting[206], yCD has been observed to display superior kinetic properties towards 5-FC (corresponding to a 22-fold lower K_m for the prodrug) and slightly improved efficacy for treating tumors in mice than bCD[13]. However, wild-type yCD is relatively thermolabile as compared to bCD, a property that may limit its performance in therapeutic applications[13,182].

Using computational protein engineering, we previously created a series of mutant yCD variants with increased thermal stability[152]. A mutant with two substitutions (A23L/I1140L or 'yCD-double') and a subsequent, final redesigned mutant (A23L/V108I/I140L or 'yCD-triple') display 5- and 30-fold increases in their half-lives at 50°C, respectively, with preservation of wild-type catalytic efficiencies. The latter construct displays improved complementation of cytosine deaminase activity at elevated temperatures in bacteria.

In the present study, we have randomly mutagenized residues in and near the wild-type yCD active site, and have used positive and negative genetic complementation strategies to identity mutations that confer increased 5-FC sensitivity (i.e., that lead to toxicity at the lowest possible concentrations of 5-FC). The resulting mutations were analyzed alone and in combination with those previously engineered in the protein core for their effect on (1) prodrug sensitivity in mammalian cells and in a mouse xenograft tumor model, (2) substrate specificity, (3) kinetic efficiency, and (4) enzyme stability. The crystal structure of the mutant from the screen with the strongest effect on 5-FC sensitivity was determined. The effects of various individual and combined mutations in this study demonstrate that improved efficacy by yCD is largely a function of increased thermostability, which can be derived through multiple pathways of protein mutations. Furthermore, combinations of mutations that are individually effective in both enzyme stabilization and in increasing 5-FC sensitivity appears to indicate a 'threshold' beyond which improved performance in PGT assays is lost--possibly due to kinetic inhibition as a result of quenching the dynamic flexibility of the enzyme and its active site. Separately, however, the computationally engineered yCD-triple and the genetically selected yCD-D92E variant each provide enhanced 5-FC sensitivity to cells and therefore may serve as improved candidates for future suicide gene therapy studies.

MATERIALS AND METHODS

Materials

Oligonucleotides used to mutate and sequence yCD were obtained from either Sigma-Proligo (St. Louis, MO) or Integrated DNA Technologies (Coralville, IA). Restriction enzymes were obtained from New England Biolabs (Beverly, MA). The plasmid pETHT:yCD expressing His-tagged yCD was constructed as described in Ireton *et al.* (2003)[181]. DNA purification was done using several kits (Wizard PCR prep kits Promega (Madison, WI), HiSpeed Plasmid Mini Kit from Qiagen (Valencia, CA) and StrataPrep EF Plasmid Midikit from Stratagene (La Jolla, CA)). Alamar Blue was purchased from Serotec Limited (Oxford, UK). All cell culture reagents were purchased from Gibco (Carlsbad, CA). All other reagents were purchased from Sigma (St. Louis, MO) unless otherwise noted.

Bacterial strains

Escherichia coli GIA39, a strain deficient in cytosine deaminase and orotidine 5'-phosphate decarboxylase, was obtained from the *E. coli* Genetic Stock Center (CGSC #5594: *thr*$^-$ *dad*B3 *fhu*A21 *cod*A1 *lac*Y1 *tsk*-95 *gln*V44(AS) λ^- *pyr*F101 *his*-108 *arg*G6 *ilv*A634 *thi*-1 *deo*C1 *glt*-15). *E. coli* GIA39 was lysogenized with DE3 according to the manufacturer's instructions (Novagen, Madison, WI). The derived strain, GIA39(DE3), was used in the genetic complementation assays for cytosine deaminase activity. *E. coli* strain CJ236 (F^+LAM$^-$, *ung*-1, *rel*A1, *dut*-1, *spo*T1, *thi*-1)

was used to produce single-stranded DNA for site-directed mutagenesis procedures. The *E. coli* strain NM522 (F^+ *lac*I^q Δ(*lac*Z)-M15 *pro*A^+B^+/*sup*E *thi*Δ(*lac-pro*AB) Δ(*hsd*MS-*mcr*B)5($r_k^-m_k^-$McrBC$^-$)) and *E. coli* strain XL1-Blue (F'::Tn10 *proA*$^+$*B*$^+$ *lacI*^{q} D(*lacZ*) *M15/recA1 endA1 gyrA*96 (Nal^r) *thi hsdR17* ($r_K^-m_K^+$) *supE44 relA1 lac*) were used as recipients for certain cloning procedures. *E. coli* BL21-RIL (Novagen, Madison, WI) was used for large scale protein purification.

Construction of the yCD Regiospecific Random Library

Preparation of the Randomized Insert

The basic protocol for regio-specific random mutagenesis was performed as outlined in Kurtz and Black (2004)[171]. The following oligonucleotides were designed to synthesize a 139bp dsDNA fragment, that spans 12 codons (T83, L84, Y85, T86, L88, S89, D92, M93, T95, G96 and I98) that were randomized at 21% at the positions indicated in bold in the MB224 template, below.

MB223 (56mer),

5' GTGAGATCTCCACTTTGGAAAACTGTGGGAGATTAGAGGGCAAA GTGTACAAAGAT 3'

MB224 (95mer),

5' CCGACAA(CACAGCGTG)GAATACCATACATGATGATGGCACCTG TACACATGTCGCATGGAGACAGCGTCGTATACAAAGTGGTATCTTT GTACAC 3'

MB225 (18mer), 5' GTG(AGATCT)CCACTTTGG 3'

MB226 (17mer), 5' CCGACAACACAGCGTGG 3'

*Dra*III and *Bgl*II restriction sites noted in the parenthesis in MB224 and MB225, respectively, were used for cloning purposes. Briefly, the139bp dsDNA fragment was synthesized by annealing 50pmol each of MB223 and MB224 with 10 x annealing buffer (70mM Tris-HCl at pH 7.5, 60mM $MgCl_2$, 200mM NaCl) in a final volume of 50μL at 95°C for 5min, 65°C for 20min, and room temperature for 10min. Next, the annealed product was extended with the Klenow fragment of *E. coli* DNA polymerase in an 80μL reaction mixture consisting of the 40μL annealed product, 4μL 10 x annealing buffer, 5.6μL 10mM dNTPs, 1.6μL 0.1M DTT and 4.8μL Klenow (5U/μL) at 37°C for 30min, 65°C for 10min and room temperature for 10min.

Amplification of the Random Insert

A master mix was prepared consisting of 110μL 10 x PCR buffer (200 mM Tris-HCl at pH 8.3, 250mM KCl, 15mM $MgCl_2$, 0.5% Tween-20), 200pmol MB225 and MB226 each, 4.4μL 10mg/mL BSA, 5.5μL 10mM dNTPs and 4.4μL *Taq* polymerase (5U/μL). The extended product (16pmol or 5μL) and 27.8μL of the master mix were mixed to a final volume of 200μL and split into four tubes containing 50μL each. The 50μL mixtures were subjected to amplification using an Eppendorf Thermocycler by 30 cycles with 1 cycle at 94°C for 1min, 34°C for 2min, and

followed with a 7min extension at 72°C. Amplification of the 139bp insert was confirmed by gel electrophoresis.

Construction of Recombinant yCD Variants

To construct the vector carrying the inactive or dummy gene, the *Dra*III and *Bgl*II sites in the vector backbone were removed by site-directed mutagenesis using the following oligonucleotides: MB277 (21mer), 5' CGATGG CCCAATACGTGAACC 3' to remove the *Dra*III site and MB307 (22mer), 5' CGGGATCGC GATCGCGGGCAGC 3' to remove the *Bgl*II site. To reduce the presence of wild-type yCD in the selection, the yCD gene was inactivated by restriction with *Acc*I followed by extension with the Klenow fragment and religation. The resulting inactivated yCD was designated as "pETHT:yCD-dummy." Following digestion with *Bgl*II and *Dra*III, the gel purified insert was cloned into the vector digested with the same restriction enzymes.

Transformation of E. coli GIA39(DE3) and Positive Selection

Approximately 3.5μL of the ligated product was electroporated into 40μL of electrocompetent *E. coli* GIA39(DE3) and then shaken at 37°C for 1hr in 1 mL of SOC medium (3g Bactopeptone, 2.5g yeast extract, 1M NaCl, 1M KCl, 5mM $MgSO_4$, 5mM $MgCl_2$ and 1.8% glucose per liter). The transformation mixture was concentrated by pelleting, resuspended in 100μL of 0.9% NaCl and plated at various

volumes onto 2 x YT rich medium, uracil and cytosine minimal media plates supplemented with 50 μg/mL carbenicillin. Growth on cytosine minimal medium requires the presence of a functional yCD, while 2 x YT and uracil minimal media were used as positive controls. Uracil minimal medium (500mL) was prepared from 0.36g yeast synthetic dropout without leucine, 50mL 10 x M9 salts (15g KH_2PO_4, 33.9g anhydrous $NaHPO_4$, 2.5g NaCl, 5.0g NH_4Cl), 1mM $MgSO_4$, 2.5mL, 20% glucose, 0.1mM $CaCl_2$, 1mL 2% leucine, and 7.5g Bactoagar. Cytosine minimal medium (500mL) was prepared from 1mM cytosine, 0.96g yeast synthetic dropout without uracil, 8.5g Bactoagar, 50mL 10 x M9 salts, 1mM $MgSO_4$, 2.5mL 20% glucose, and 0.1mM $CaCl_2$. The 2 x YT plates were incubated at 37°C overnight, and the uracil and cytosine plates were incubated at 37°C for approximately 36hr. The number of transformants on the 2 x YT plate allows estimation of the library size. Transformants from the cytosine plates were picked and restreaked onto fresh cytosine plates to confirm the phenotype.

Selection of 5-FC Sensitive Mutants

Genetic complementation of yCD was done as previously established for bacterial CD[207]. To determine the ability of the mutants to confer 5-FC sensitivity, the functional variants determined by the positive selection described above were streaked onto cytosine plates supplemented with 5-FC at 10μg/mL, a sublethal dose for wild-

type yCD, and incubated for approximately 36hr at 37°C. Colonies unable to grow on the 5-FC plates were selected from the control plates and subjected to additional rounds of negative selection at decreasing 5-FC concentrations (5, 2, 1 and 0.5μg/mL). Plasmid DNA of the yCD variants was isolated and the randomized region sequenced using the T7 terminator primer (5'TATGCTAGTTATTGCTCAG 3') at the core sequencing laboratory at Washington State University.

Construction of Mammalian Expression Vectors

Computationally designed thermostabilized yCD genes described in Korkegian *et al.* (2005)[152] (pET15b:yCD-A23L/V108L (yCD-double) and pET15b:yCD-A23L/V108L/I140L (yCD-triple)) were sub-cloned into the mammalian expression vector, pCDNA6/*myc*-His B (Invitrogen, Carlsbad, CA). The yCD-double and yCD-triple genes were subcloned into pCDNA6/*myc*-His B digested with *Eco*RV and *Xho*I as *Nco*I (blunt-ended)/*Xho*I fragments. After restriction enzyme verification, DNA sequencing analysis confirmed the presence of the yCD-double and yCD-triple genes.

Site-directed mutagenesis was performed to introduce the mutations derived from the regio-specific random mutagenesis to the yCD-triple mutant using the QuikChange Site-directed Mutagenesis Kit from Stratagene (La Jolla, CA) according to the manufacturer's protocol. Three pairs of oligonucleotides containing the D92E, M93L or I98L substitutions and a silent mutation to remove a restriction site for screening purposes were synthesized by Integrated DNA Technologies (Coralville,

IA). In the mutagenic oligonucleotides the bolded nucleotide indicates the regio-specific random substitution and the underlined nucleotide indicates the removal of the restriction site.

D92E, loss of Afl*III site*

MB374 (31mer), 5' CGCTGTCTCCATGCGAAATGTGTACAGGTGC 3'

MB375 (31mer), 5' CGACCTGTACACATTTCGCATGGAGACAGCG 3'

M93L, loss of Afl*III site*

MB376 (31mer), 5' CGCTGTCTCCATGCGACCTGTGTACAGGTGC 3'

MB377 (31mer), 5' GCACCTGTACACAGGTCGCATGGAGACAGCG 3'

I98L, loss of BanI site

MB378 (34mer), 5' GCGACATGTGTACAGGAGCCCTCATCATGTATGG 3'

MB379 (34mer), 5' CCATACATGATGAGGGCTCCTGTACACATGTCGC 3'

After restriction enzyme verification, DNA sequencing analysis was performed to confirm the presence of the correct mutation.

5-FC Sensitivity Assays

One μg of each DNA (pCDNA (vector alone), pCDNA:yCD, pCDNA:yCD-D92E, pCDNA:yCD-M93L, pCDNA:yCD-I98L, pCDNA:yCD-double, pCDNA:yCD-triple;, pCDNA:yCD-triple/D92E; pCDNA:yCD-triple/M93L;

pCDNA:yCD-triple/I98L) was used to transfect 1 x 10^5 rat C6 glioma cells by lipofection using FuGENE 6 transfection reagent (Roche Diagnostics, Penzberg, Germany) at a 3:1 ratio according to the manufacturer's directions. Immunoblots were performed to assess protein expression. Briefly, pools of transfectants were harvested and resuspended at 100,000 cells/μL in lysis buffer (for 2mL: 2μL 1M DTT, 20μL 1M HEPES, 40μL Nonidet P40 (Roche Diagnostics, Pernzberg, Germany) 2μL $MgAc_2$, H_2O to final volume). The resuspended pellets were incubated on ice for 20min and subjected to centrifugation at 4°C for 20min to pellet debris. Heat denatured samples (10μL per well) were subjected electrophoresis on a 15% SDS containing polyacrylamide gel, transferred to a nitrocellulose membrane and blocked with 3% gelatin in Tris-buffered saline. The membrane was probed with rabbit polyclonal yCD antiserum (gift from Dr. Alnawaz Rehemtulla, U. Michigan, Ann Arbor, MI) followed by goat anti-rabbit AP-conjugated antiserum. The blot was developed using the AP Conjugate Substrate Kit (Bio-Rad, Hercules, CA). To determine the cytotoxicity of 5-FC, pools of transfectants were transferred to 96 well microtiter plates at an initial density of 250 cells per well in DMEM plus supplements[65]. After cell adherence overnight, 5-FC (0-10mM) was added in sets of eight wells for each concentration tested. The plates were incubated for 6 days at 37°C in 5% CO_2 at which time the redox-indicator dye Alamar Blue was added. Cell survival was determined several hours later as according to the manufacturer's instructions and the data were plotted with a standard error of the mean bar. At least three replicates were performed.

Xenograft Tumor Model

Pools of C6 glioma cells stably transfected with pCDNA, pCDNA:yCD, pCDNA:yCD-D92E, or pCDNA:yCD-triple (0.5×10^6 cells in 200μL of phosphate buffer saline (PBS) at pH 7.2) were injected subcutaneously into 5- to 6- week-old female nude mice (n = 5 for each group) (BALB/cAnNCr-nu/nu; National Cancer Institute, Fredrick, MD, USA). When the tumors reached 3-4mm (day 0), PBS or 5-FC (500mg/kg) was administered by intraperitoneal injection once a day for 18 consecutive days. Starting at day 0, the tumor volume was monitored using caliper measurement (length, width, and height) every other day. Tumor volume was calculated using the formula: $4/3\pi((\text{widthxlengthxheight})/2)$. Tumor volume was plotted and analyzed for statistical significance using Student's T-test.

Protein Expression and Purification

Protein expression and purification of yCD was carried out as previously described (Ireton *et al.*, 2003)[181], with the exception that buffer exchange was carried out with Biorad pre-packed Econo-Pac 10DG buffer exchange columns rather than overnight dialysis. Protein expression from pET15b:yCD (BL21-RIL) yields yCD with a thrombin cleavable 6-His tag fused to the N-terminus. All kinetics and thermostability experiments were carried out with fresh, unfrozen protein stored at 4°C for 2 weeks or less.

Activity Assays

The conversion of cytosine to uracil by yCD was measured spectrophotometrically by monitoring change in absorbance at 286nM while the conversion of 5-FC to 5-FU was monitored at 238nm. The protein was diluted to 2μM in 50mM Tris-Cl (pH 7.5) and mixed 1:1 with a range of nine cytosine concentrations from 0.2 – 1mM in the same buffer. Absorbance at 238nm was measured every 5sec until baseline was reached with the first reading taken 5sec after mixing. Measurements were taken in quadruplicate and averaged to reduce error. Initial velocity was calculated as a function of the initial slope by curve-fitting the resulting plot, taking the derivative and extrapolating back to time zero. K_m and k_{cat} values of wild-type yCD and mutant constructs were determined from a double reciprocal (Lineweaver-Burk) plot of the resulting data and the catalytic efficiency k_{cat}/K_m was calculated from these values.

Circular Dichroism Measurement and Thermal Denaturation Experiments

Circular dichroism data were collected on an Aviv 62A DS spectrometer as described in Dantas *et al.* (2003)[140]. Wavelength scans were run from 200-260nm to determine the folded state of the protein, the ratio of concentration to signal strength and the wavelength where signal strength was at its highest. Temperature melts were run with 8-12μM protein in a 2mm pathlength cuvette from 10° to 98°C in 2° increments with temperature regulated by a Peltier device. Sample temperature was

allowed to equilibrate for 30sec before measurement and signal was collected and averaged over 30sec. Denaturation was recorded as a change in elipticity over temperature. Apparent T_ms were determined by curve-fitting.

Crystallization and Structure Determination

Crystals were grown using methodology described in Ireton *et al.* (2003)[181]. After two days crystals appeared in the drops, but were often twinned. These crystals were then used to streak seed into the clear lower concentration drops where higher quality crystals eventually grew. Crystals were looped out and soaked for 20min in a mother liquor solution containing 2-hydroxypyrimidine concentrated 1.2:1 relative to protein. After soaking the crystals were immediately transferred briefly to a cryobuffer containing the 2-hydroxypyrimidine mother liquor plus 25% DMSO and then flash-frozen in liquid nitrogen. All data were collected from a single crystal on the 5.0.1 beamline at the ALS syncrotron to 1.7A. Data were indexed and scaled using the DENZO/SCALEPACK program suite[186]. R_{merge} values for the higher resolution scales were poor so the data was re-indexed and scaled at 2.3Å. The initial structures were generated by molecular replacement using the program EPMR[187] using the wild-type yCD crystal structure with the mutated residues and any residues within a 4Å shell of those residues truncated to alanine. Refinement was carried out in rounds using the program CNS[188,189] with a random 10% of data withheld for cross-validation. The yCD-triple/D92E mutant crystal structure was refined at 2.3Å to a

R_{work}/Rf_{ree} of 20.4/25.7. Electron density maps and models were visualized in XtalView[190] while cartoon representations of the structure were generated using PyMOL[191]. Coordinates and diffraction data for yCD-triple/D92E structure have been deposited in the PDB database (PDB 2o3k).

RESULTS

Identification of Mutants in and Near the Active Site that Confer Enhanced 5-FC Sensitivity

To create yCD variants with increased activity to 5-FC, we subjected 11 codons within the most conserved region of yCD (T83, L84, Y85, T86, L88, S89, D92, M93, T95, G96 and I98) to regio-specific, partially randomizing mutagenesis as described in *Materials and Methods*. Completely conserved residues within this same region, assumed to be critical for activity, were omitted from randomization (T87, P90, C91 and C94) (Figure 5.1).

Identification of functional yCD variants with enhanced 5-FC activity involved a two-step selection procedure. First, functional yCD variants were selected for preservation of catalytic activity, based upon their ability to confer growth on cytosine-containing plates, under conditions that require CD activity for viability. Second, the functional variants were then screened on cytosine plates containing 5-FC. It was determined that the 5-FC dose at which wild-type yCD survives, or the sub-lethal dose, is 10μg/mL. Although wild-type pETHT:yCD will allow growth on plates

containing the sub-lethal dose of 5-FC, any mutant with increased 5-FC activity will not be viable. To identify the best variants within the library pool colonies were sequentially streaked onto plates containing 5-FC from 10μg/mL to 0.5μg/mL. From an estimated total of 34,000 transformants, 50 colonies (~0.15%) were identified as yCD positive based on their ability to grow on cytosine containing plates. Of the 50 yCD positive clones, only six conferred sensitivity at 0.5μg/mL 5-FC, the lowest effective 5-FC concentration in the negative selection.

In order to evaluate the library diversity, along with identifying which substitutions are tolerated, plasmid DNA from colonies on the non-selective and selective plates was isolated and sequenced. Sequence analysis of DNA isolated from colonies that grew on the non-selected plates revealed a broad spectrum of mutations in the regio-specific library (data not shown) and indicates that the oligonucleotides used to generate the library contained random sequences. Sequence analysis of the 6 variants that conferred the greatest sensitivity to *E. coli* revealed that two had a substitution at D92 to glutamic acid (D92E), two had a substitution at M93 to leucine (M93L) and two had a substitution of I98 to leucine (I98L). Additionally, the two I98L mutants have different nucleotide-level changes, suggesting these mutants were derived independently. Therefore, from a library of 34,000 transformants, three amino acid substitutions were identified that confer 5-FC sensitivity (D92E, M93L, and I98L) at the lowest 5-FC concentration in the selection scheme.

Construction of Combinations of yCD Mutants

Recently, we performed a computational design study aimed at improving the stability of the thermolabile yCD[152], in which mutations were generated in the enzyme's hydrophobic core, distant from the active site. From this study, two thermostable mutants ('yCD-double'; A23L/I140L and 'yCD-triple'; A23L/V108I/I140L) were generated that have apparent melting temperatures 6°C and 10°C greater than that of wild-type yCD, respectively. To investigate the combined effect of these mutations with those selected in and near the enzyme active site as described above, several new mutants were constructed by site-directed mutagenesis. These 'superimposed' mutants (yCD-triple/D92E, yCD-triple/M93L and yCD-triple/I98L), and their individual parental mutations, were then tested in a mammalian tumor cell line assay for increased 5-FC sensitivity.

***In Vitro* 5-FC Sensitivity Assays**

To determine the activity of these mutants towards 5-FC (and their ability to induce sensitivity to 5-FC) *in vitro*, mammalian expression vectors encoding the yCD variants were constructed and used to transfect rat C6 glioma cells (see *Materials and Methods*). Immunoblot analyses of lysates from the pools of transfectants show similar yCD expression levels for all of the mutants and wild-type yCD, with no detectable expression in vector control pools (data not shown). Pools of stable transfectants were assayed for 5-FC sensitivity over a drug range of 1 to 10mM.

Representative results of the 5-FC sensitivities displayed by the yCD mutants, wild-type yCD and a vector control are shown in Figure 5.2. Little to no toxicity was observed with vector alone at the lower 5-FC doses; however, above 6mM 5-FC an inherent cytotoxicity is observable in the glioma cell line. YCD-D92E displays the greatest reduction in IC_{50} (~30%) for 5-FC of the three regio-specific random mutants (Figure 5.2A). No difference in sensitivity was observed with the yCD-I98L substitution as compared to wild-type yCD. The yCD-M93L mutant displays an intermediate IC_{50} of 9.5mM, an estimated decrease of ~15% from wild-type yCD.

The computationally designed thermostablizing yCD-double and yCD-triple mutants also display increased sensitivity towards 5-FC in glioma cells (Figure 5.2B). The yCD-double mutant displays a similar IC_{50} (8mM) to yCD-D92E, an approximate 30% reduction in IC_{50} compared to wild-type yCD-transfected cells. The greatest enhancement in activity was observed with the yCD-triple mutant with an IC_{50} of approximately 6mM, an estimated 50% reduction relative to wild-type yCD. Unexpectedly, none of the superimposed mutants exhibit enhanced activities towards 5-FC (Figure 5.3). The yCD-triple/I98L and yCD-triple/M93L have an IC_{50} of 9mM while the yCD-triple/D92E has a similar IC_{50} to that of wild-type yCD transfected cells. These results indicate that the addition of the randomly generated substitutions to the designed mutant negates the effect of thermostability introduced by the triple substitutions.

5-FC Sensitivity Assayed using a Tumor Xenograft Model

The behavior of the two most active yCD constructs from the assays above (yCD-triple and yCD-D92E) were then further compared to wild-type enzyme and a negative control in a mouse xenograft model of prodrug gene therapy. As described in *Materials and Methods*, pools of rat C6 cells stably transfected with pCDNA containing yCD, yCD-triple, yCD-D92E or vector alone (0.5×10^6 cells) were subcutaneously injected into the flanks of nude mice (n= 5). When the tumor size reached 3-4mm, phosphate buffered saline (PBS) or 5-FC at 500mg/kg was injected intraperitoneally once a day for 18 days. Tumor size was monitored every other day. Tumor cells transfected with vector only (pCDNA) showed no statistical difference in tumor size between mice treated with PBS or prodrug (Figure 5.4A). Similarly, no statistical difference was observed in tumor volume in mice seeded with cells transfected with pCDNA:yCD that were injected with PBS or 5-FC (Figure 5.4B). The lack of difference in wild-type yCD tumor size between PBS and 5-FC treated mice at day 18 is likely a reflection of the low 5-FC dose administered and/or the duration of the prodrug treatment.

In contrast to tumors expressing wild-type yCD, the prodrug treated mice bearing yCD-triple or yCD-D92E expressing tumors elicited a strong response (Figures 5.4C and 5.4D). The mice bearing yCD-triple expressing tumors treated with 5-FC displayed the greatest restriction in tumor growth (mean tumor volume of $1076mm^3$ versus PBS tumor mean of $2833mm^3$; $P = 0.05$). The yCD-D92E tumor

bearing mice also showed restricted growth in the presence of 5-FC (733mm^3 versus 1942mm^3 for saline treated mice; $P = 0.035$). To compare the effects of 5-FC on tumors expressing the redesigned thermostabilized yCD-triple with that of the selected yCD- D92E, the mean tumor size of the prodrug treated group was divided by the mean tumor size of the prodrug treated group for each group on day 18. Both yCD-triple and yCD-D92E tumor bearing mice responded equally well (2.65 versus 2.64, respectively).

Enzyme Kinetics and Thermostability

The unexpected loss of sensitivity towards 5-FC when yCD-triple and yCD-D92E were combined, as compared to their individual behaviors, led us to investigate the catalytic and biophysical properties of these yCD variants in more detail. Using purified proteins, enzyme assays with cytosine and 5-FC as substrates were performed as described in *Materials and Methods.*

Overall, the relative catalytic efficiency (k_{cat}/K_m) towards cytosine of the individual yCD-triple or yCD-D92E mutants is not appreciably different from wild-type enzyme, whereas the yCD-triple/D92E is approximately 3-fold less efficient (Figure 5.5). This result is primarily caused by a reduction in the enzyme's turnover rate: k_{cat} (cytosine) for yCD-triple/D92E is approximately 1/8th that of wild-type enzyme. In contrast, the K_m values for cytosine observed with all mutant enzymes are all within approximately 2 to 2.5-fold of wild-type yCD values (Table 5.1).

When the prodrug 5-FC is instead used as the substrate in kinetic assays, the individual yCD-triple and yCD-D92E mutant enzymes again display overall catalytic efficiencies (k_{cat}/K_m) that are roughly equivalent to wild-type yCD (Table 5.1 and Figure 5.5). This corresponds to slightly higher K_m values (1.7- and 1.8-fold, respectively) and corresponding increases in k_{cat}. In contrast, although the combined yCD-triple/D92E enzyme has a slightly lower K_m, it displays a k_{cat} value half that of wild-type yCD. This gives the combined yCD-triple/D92E mutant enzyme a marginally impaired 5-FC catalytic efficiency (1.6-fold) compared to wild-type yCD, yCD-triple and yCD-D92E.

Inside a cell, cytosine and 5-FC compete for the active site of yCD. To take this into account, the relative specificities or substrate preference for 5-FC of all three mutant enzymes was compared (calculated as $[k_{cat}/K_{m\,(5FC)}/(k_{cat}/K_{m\,(5FC)} + k_{cat}/K_{m\,(cyt)})]$). With wild-type yCD set a relative specificity of 1.0, there is no considerable difference in substrate preference between yCD-wild-type (1), yCD-triple (0.947) or yCD-D92E (1.12) and only a modest 1.23-fold increase in preference for 5-FC displayed by the yCD-triple/D92E mutant. This modest shift in specificity towards 5-FC for the combined mutant is not correlated with what was observed in cell sensitivity assays, where yCD-triple/D92E performed no better than the wild-type enzyme.

In previous experiments we showed that the redesigned yCD-double and yCD-triple mutants display denaturation temperatures (T_m) that are 6°C and 10°C higher,

respectively, than that of the wild-type enzyme (wild-type yCD, T_m = 52°C; yCD-double and yCD-triple mutants, T_ms= 58°C and 62° C, respectively). The stabilization of yCD in the yCD-triple mutant corresponds to improved complementation of CD activity in *E. coli* at elevated temperatures[152], and presumably is also responsible for its improved performance in PGT assays in this study.

We therefore examined the effects of the individual yCD-D92E substitution and of the superimposed yCD-triple/D92E mutation on protein stability. Thermal denaturation experiments were performed on the wild-type, yCD-triple, yCD-D92E and yCD-triple/D92E proteins using circular dichroism spectroscopy as outlined in *Materials and Methods*. Although the yCD-D92E mutant was selected based on its ability to confer increased sensitivity to 5-FC in *E. coli*, this enzyme unexpectedly also displays an increase in its apparent denaturation temperature, corresponding to a T_m 4°C higher than the T_m for wild-type yCD. Combination of yCD-D92E with the previous yCD-triple mutations in the enzyme core results in a dramatic 16°C increase in apparent T_m as compared to wild-type yCD (6°C higher than that of the yCD-triple mutant) (Figure 5.6).

Structure Determination

A 2.3Å crystal structure was solved of yCD-triple/D92E bound to a transition state analogue (Table 5.2) and compared to both previously solved wild-type (PDB 1p6o) and triple mutant (PDB 1ysb) structures in order to identify any potential

structural changes that may explain the observed IC_{50} values as well as the thermodynamic and kinetic data. The yCD-D92E and the yCD-triple mutations appear far from the active site (Figure 5.7 and Figure 5.8A) and indeed the active site residues show no perturbations in either the yCD-triple or yCD-D92E/triple structures when compared to wild-type (Figure 5.8B). The yCD-D92E mutation, although close in sequence to active site residues C91 and C94, is actually involved in the nearby dimer interface of the homodimeric enzyme. In the wild-type enzyme both carboxyl oxygens in aspartate 92 make salt-bridge contact to a nearby arginine residue in the adjacent monomer (Figure 5.9A). In yCD-D92E the extra carbon in the glutamate is packed against a nearby lysine and isoleucine, resulting in a redistribution of contacts made by the carboxylate oxygens of the mutated side chain (Figure 5.9B). The resulting structural changes in the vicinity of the mutation are very subtle and the reasons that it results in such a significant stabilization are difficult to ascertain.

DISCUSSION

The initial premise of this study was that independent mutations in the active site and the hydrophobic core of yCD, that respectively influence substrate specificity and enzyme thermostability, could each be identified in separate experiments and then combined in a single construct for maximum benefit. Furthermore, whereas stabilizing interactions in a protein core can be improved through computational protein design[152], an enzyme's substrate specificity is still more effectively addressed

through incorporation of random mutations and subsequent genetic screens or directed evolution protocols. Thus, a two-pronged experimental approach to optimize the physical and catalytic properties of yCD was expected to yield a combination of enzyme mutations with optimal properties.

Instead, the experiments described above yielded unexpected results in three ways. First, the protocol of random mutagenesis and genetic screening generated a construct (yCD-D92E) whose primary characteristic relative to wild-type enzyme appears to be stabilization of the folded enzyme: the mutant displayed a significant increase in its T_m, but only a small (1.12-fold) specificity shift towards 5-FC relative to cytosine. This result appears to have been made possible through the choice of residues that were subjected to mutation, which extended beyond the substrate binding pocket, to nearby positions involved in dimerization. Residues were chosen for mutation prior to determination of the enzyme's structure, based on sequence conservation across the closest known homologues with deaminase function. In retrospect, strongly conserved residues in yCD (and most other enzymes) include not only those positions most critically involved in substrate recognition and catalysis, but also amino acids those that closely couple the architecture of the active site to the fold and stability of the enzyme tertiary structure (such as D92).

Second, the stabilized mutant yCD constructs produced by repacking of the hydrophobic core (yCD-double and yCD-triple) were both significantly improved in their performance in PGT assays, with the latter construct actually outperforming the

best mutant construct selected solely on the basis of 5-FC sensitivity. This result, considered in concert with the properties of yCD-D92E described above (improved stability and nominal changes in catalytic behavior), clearly indicate that the primary limiting behavior of wild-type yCD for prodrug gene therapy is its stability, which presumably is coupled to steady-state enzyme levels in transfected cells and enhanced conversion of 5-FC to 5-FU.

Finally, the combination of two sets of physically separate mutations (yCD-triple and yCD-D92E) that were independently generated failed to act in a synergistic or additive manner in PGT assays when combined into a single construct. On the surface, an explanation for this behavior seems straightforward: the combined construct (yCD-triple/D92E) is significantly more thermostable than are any of the individual yCD constructs (with an apparent increase in T_m of 16°C), which might lead to a corresponding reduction in catalytic efficiency at physiological temperatures due to reduced conformational dynamic flexibility of the protein backbone. However, the difference in kinetic rate and catalytic efficiency towards 5-FC exhibited by yCD-triple or yCD-D92E compared to yCD-triple/D92E (which performs similarly to wild-type enzyme) is not dramatic: the combined mutant is perhaps 1.5-fold less active towards the prodrug than are the individual mutants. This result implies that there exists a quite sharp 'threshold' of activity towards 5-FC deamination in the cell that leads to considerably more sensitivity towards low concentrations of the prodrug. Furthermore, the specific activity of the wild-type enzyme under standard protocols of

transfection and expression must be somewhat close to that critical point of metabolic sensitivity.

Variable Routes to Protein Thermostabilization

Two entirely separate forms of thermostabilized yCD enzyme have been generated in these studies, one of which was generated in a screen for increased sensitivity to 5-FC, the other in a structure-based redesign protocol intended to produce a more robust form of the enzyme. The independent sequence changes that lead to stabilization (increased van der Waals packing of each subunit's hydrophobic core, versus more subtle alteration of contacts in the homodimer interface) are structurally dissimilar from one another, providing a clear demonstration that multiple evolutionary routes can act on protein stability through considerably different structural mechanisms.

This result reinforces a myriad of studies of protein stability conducted over the past ten years, where the structures of homologous proteins from various biological sources (ranging from cryophiles to mesophiles to thermophiles) have been directly compared in attempts to catalogue the determinants of protein stability[208-213]. Although a variety of factors have been implicated in thermostability (including hydrophobic packing, formation of additional salt-bridges within the protein structure, modifications of surface charge distribution and shifts in amino acid composition, to name just a few), each mechanism appears to be relatively unique to specific structural

families of proteins. In addition, individual proteins can be made thermostable either by reducing their relative free energy in the folded state (termed 'thermodynamic stabilization') or by decreasing the rate of unfolding through modifications that increase the energy barrier of the unfolding transition state (termed 'kinetic stabilization'). This latter mechanism for protein stabilization has been observed for a variety of proteins, such as a-lytic protease[214]. A significant aspect of such proteins is the identification of residues at points distant from the hydrophobic core, and often at domain interfaces, that are highly sensitive to mutational effects on protein stability[215]. Given the nature of the yCD-D92E mutation at the dimer interface, which does not appear to obviously increase favorable contacts in the folded enzyme, it might be possible that this mutation stabilizes yCD via this latter mechanism.

yCD and 5-FC and Enzyme Engineering in Recent Gene Therapy Trials and the Role of Optimized Enzyme Constructs

Between 1990 and 2004, 405 gene therapy trials were approved in the US for cancer treatment, of which 95 contain various elements of enzyme/prodrug therapeutic strategies; 74 of these trials were active as of April 2005. Three of these trials, all involving the combination of thymidine kinase and ganciclovir, have advanced to phase III multicenter programs[8,216]. A large percentage of these trials have targeted prostate cancer, due to the relatively low percentage of these cancers that have associated detectable metastatic disease (<10%) at the time of diagnosis[217]. Of these

latter trials, many have made use of cytosine deaminase enzymes and 5-FC, most in combination with additional enzyme/prodrug activities. Results from phase I trials against prostate cancer and pancreatic cancer have recently been reported for combined PGT therapies involving yeast cytosine deaminase. In each study a replication-competent adenovirus armed with wild-type yCD fused to a modified HSV-thymidine kinase (SR39), in combination with 5-FC/ganciclovir prodrug administration and radiation, was utilized[217]. In both cases, the trimodal approach (oncolytic virus, suicide gene/prodrug, and radiation) significantly increased tumor control beyond that displayed by individual modalities, with a corresponding reduction in morbidity and toxicity as compared to traditional chemotherapy. As well, positron emission tomography (PET) studies indicated strict localization of the transfected gene and its product in the targeted tissue.

Despite the promise of such studies, prodrug gene therapy is limited pharmacokinetically by the ability to deliver the suicide gene efficiently to tumor cells whereas traditional chemotherapeutic strategies are perhaps equally limited by the ability to specifically target the effects of cytotoxic agents to tumor cells at levels over normal proliferating cell populations in the body. The ability of PGT to effectively ablate tumor cells is also reliant on the inherent sensitivity that cell type displays towards a particular gene/prodrug combination. Additionally, when poor responsiveness necessitates repeat therapeutic intervention, rapid immune clearance of cells expressing a previously seen foreign protein may serve to further limit treatment

efficacy. As such, various enzyme/prodrug combinations will be required to accommodate these various treatment scenarios. To offset the significant hurdles associated with PGT, the pharmacokinetic properties of the enzyme (its stability, half-life and kinetic activity), the prodrug (its metabolism and toxicity) and combination of the two (their uniqueness to transfected cells) must be optimized for maximum therapeutic efficacy. In particular, the catalytic efficiency of enzymatic activation of the prodrug, and the relative level of activation of the prodrug in transfected cells must be optimized for several enzyme/prodrug combinations. For the most commonly used enzyme/prodrug combination under clinical study to date (HSV thymidine kinase and ganciclovir), random mutagenesis and genetic selection experiments have generated a variant of HSV thymdine kinase (SR39), that has led to a 294-fold increase in sensitivity to ganciclovir when compared to wild-type enzyme, as well as improved tumor growth inhibition[15,218].

In the case of cytosine deaminase and 5-FC, the availability of two separate enzyme scaffolds and folds (yCD and bCD) further increases the parameters that can be optimized and combined with other enzymes; each enzyme displays unique characteristics of size, oligomeric structure, behavior when 'tethered' to additional enzymes, metal dependence, stability and substrate specificity profiles. Given that a variety of studies that use two enzyme/prodrug combinations have shown synergistic improvements in efficacy over individual enzyme systems[92,93,219,220], the utility of kinetic and biophysical optimization of each enzyme seems clear. Thus, side-by-side

optimization of each enzyme source for PGT is likely to illuminate the balance of properties that are most critical for performance in gene therapy applications.

Table 5.1 Kinetic measurements Yeast Cytosine Deaminase Variants against Cytosine and 5-Fluorocytosine

	Substrate	Km (mM substrate)	Vmax (M prod/sec)	kcat (M prod/sec M E)	kcat/Km (1/sec M E)
Wild-Type	5FC	0.083	1.87E-05	18.8	2.26E+05
Triple	5FC	0.14	2.95E-05	29.5	2.09E+05
D92E	5FC	0.15	3.17E-05	31.7	2.16E+05
D92E+Triple	5FC	0.068	9.51E-06	9.5	1.39E+05
Wild-Type	Cytosine	1.17	1.70E-04	170	1.45E+05
Triple	Cytosine	0.57	8.62E-05	86.2	1.53E+05
D92E	Cytosine	0.52	5.18E-05	51.8	1.00E+05
D92E+Triple	Cytosine	0.47	6.53E-05	21.8	4.59E+04

Table 5.2 Crystallographic Statistics for Yeast Cytosine Deaminase Structures

Data Collection	WT Bound to DHP	D92E Triple Bound to DHP	Triple
Resolution	20-1.14 (1.18-1.14)	50-2.3 (2.38-2.30)	50-1.7
Completeness	92.6 (61.6)	100 (100)	93.9 (86.1)
Redundancy		7.2 (7.3)	2.2
Rmerge	0.047 (.283)	.042 (.087)	0.047
Average I/o	34.3 (3.8)	20.7 (7.5)	19 (9.5)
Refinement Statistics			
Resolution	10-1.14	50-2.3 (2.33-2.3)	50-1.7
Number of Reflections	92,220	13204	28869
R_{work}	11	20.4 (20.6)	16.7
R_{free}	15.2	25.7 (26.7)	19.4
Rms deviations			
Bonds	0.011	0.006	0.0039
Angles	1.99	1.1	1.1

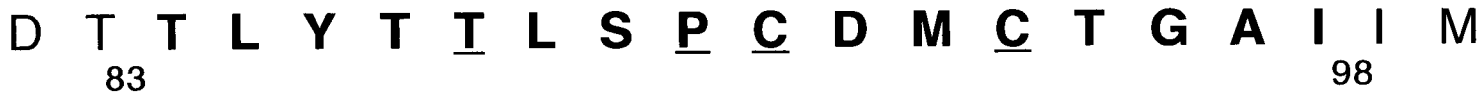


Figure 5.1 Yeast Cytosine Deaminase Amino Acid Sequence Targeted for Regiospecific Random Mutagenesis. The amino acid region targeted for regiospecific random mutagenesis is highlighted in black. Highly conserved residues are underlined and were not targeted for mutagenesis.

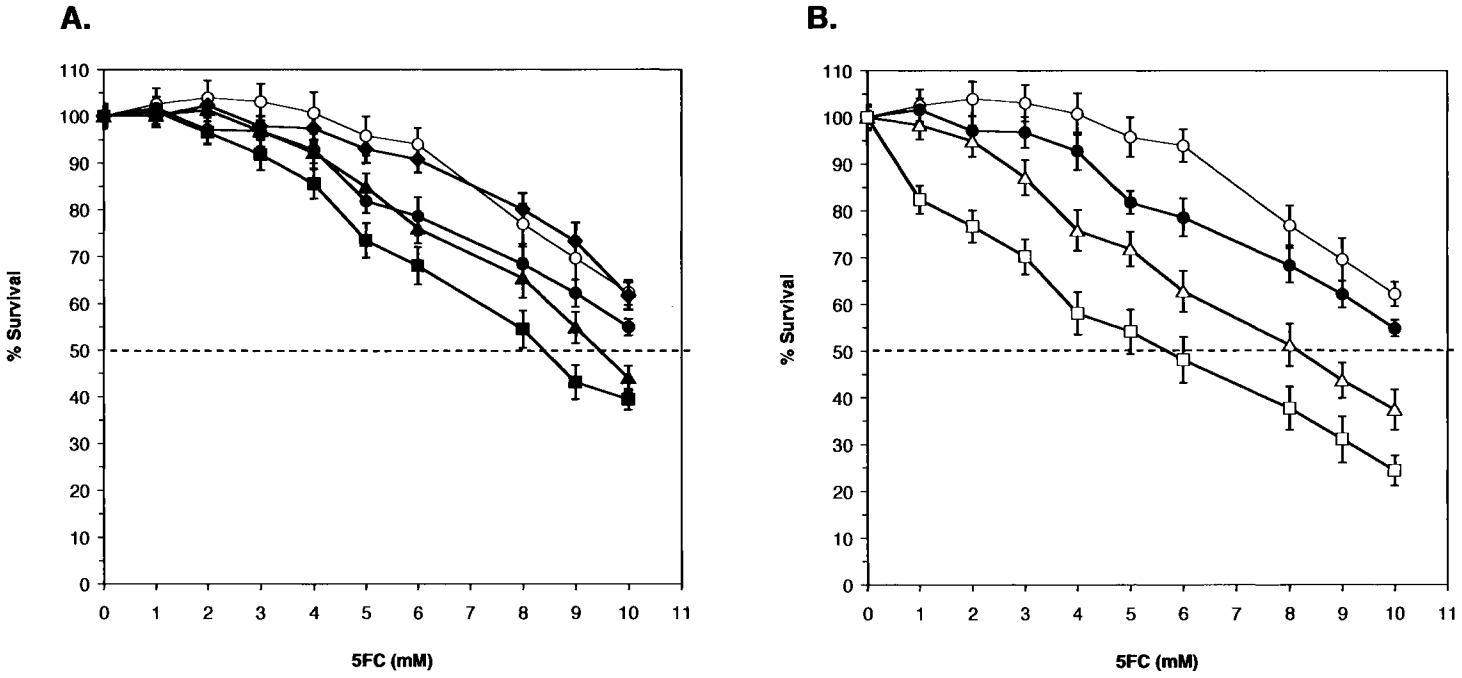


Figure 5.2 5-Fluorocytosine Sensitivity of Rat C6 Glioma Cells Transfected with Yeast Cytosine Deaminase Variants. Pools of stable transfectants containing vector only (pCDNA), wildtype yeast cytosine deaminase (yCD), **A)** regio-specific random mutants (D92E, M93L or I98L) or **B)** computationally designed thermostable mutants (double and triple) were used to transfect rat C6 glioma cells and evaluated for 5FC sensitivity as described in *Materials and Methods*. After six days of 5FC treatment, cell survival was determined using Alamar Blue according to the manufacturer's instructions. Each data point (mean ± SEM, n=3 performed with at least fifteen replicates) is expressed as a percentage of the value for control wells with no 5FC treatment.

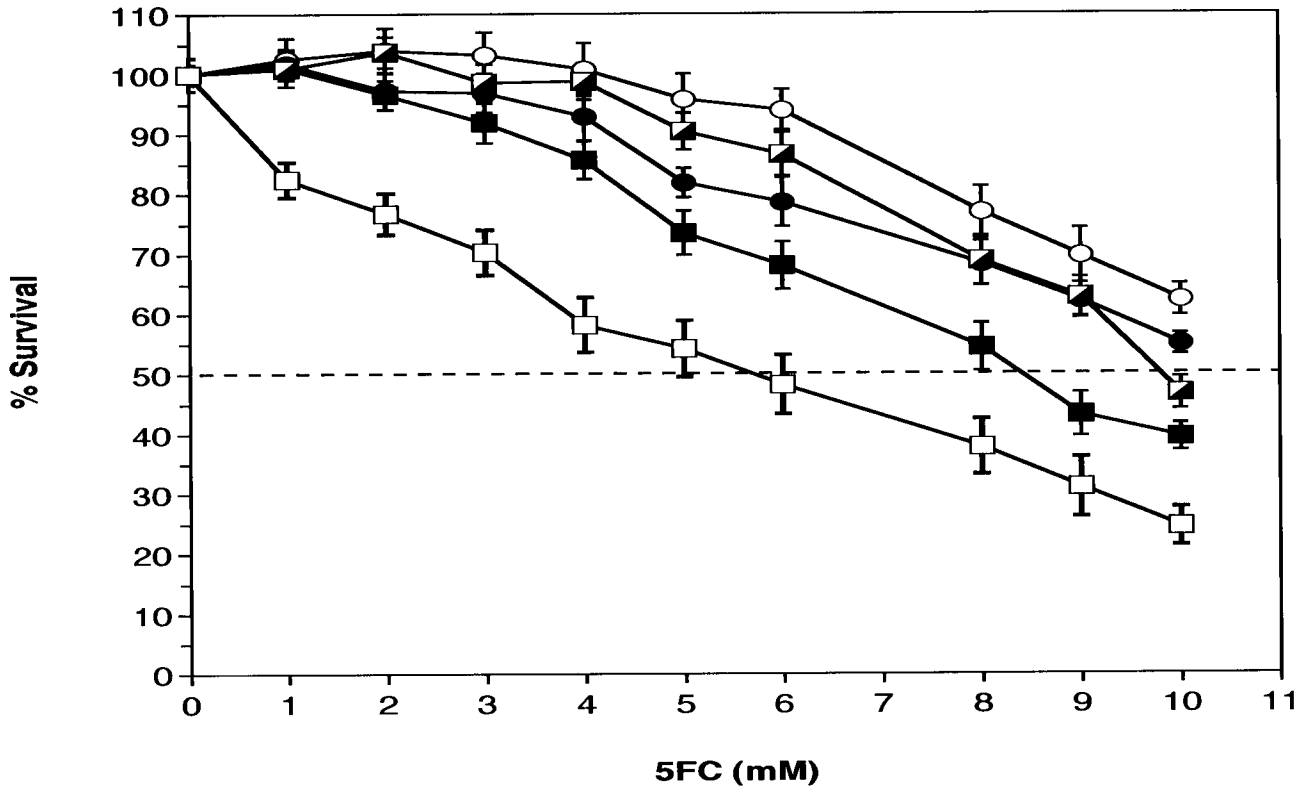


Figure 5.3 5-Flurocytosine Sensitivity of Rat C6 Glioma Cells Transfected with D92E-Triple Yeast Cytosine Deaminase. Pools of stable transfectants containing vector only (pCDNA), wildtype yeast cytosine deaminase (yCD) and the superimposed mutants (triple/D92E, triple/M93L or triple/I98L) were used to transfect rat C6 glioma cells and evaluated for 5-FC sensitivity as described in *Materials and Methods*. After six days of 5-FC treatment, cell survival was determined using Alamar Blue according to the manufacturer's instructions. Each data point (mean ± SEM, n=3 performed with at least fifteen replicates) is expressed as a percentage of the value for control wells with no 5-FC treatment.

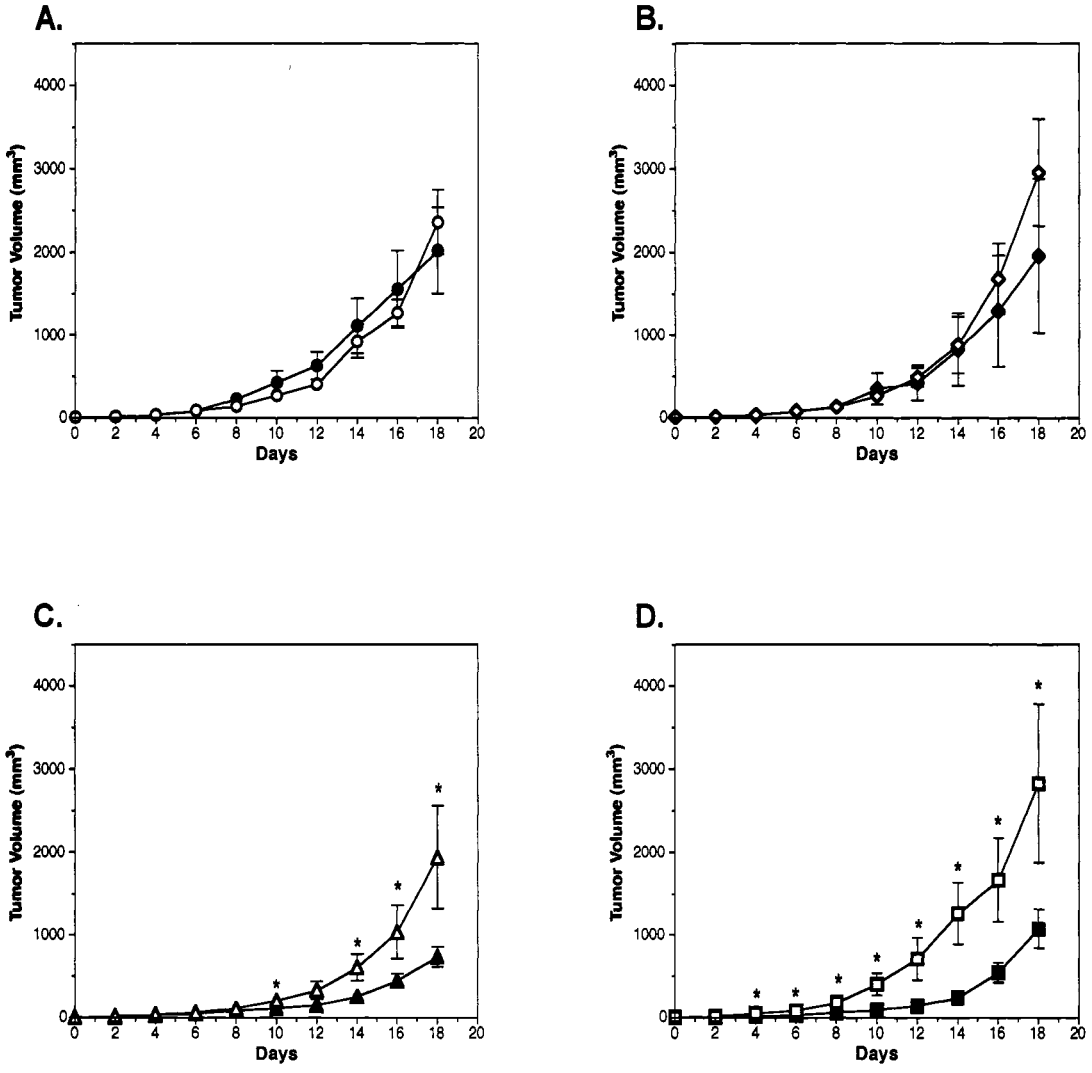


Figure 5.4 Rat C6 Glioma Mouse Tumor Xenograph Model. Pools of rat C6 glioma cells transfected with **A)** vector, **B)** wildtype yCD, **C)** triple or **D)** D92E were used to seed tumors in nude mice. When tumor size reached 3-4mm, PBS or 5-FC (500mg/kg) was administered once a day for 18 days. During this period tumor growth was measured every other day (n = 5 for each group). Tumor volume was calculated using the formula 4/3 ((width x length x height)/2), plotted and analyzed for statistical significance using the Student's T-test. Asterisks denote statistical significance ($P = 0.05$ or lower).

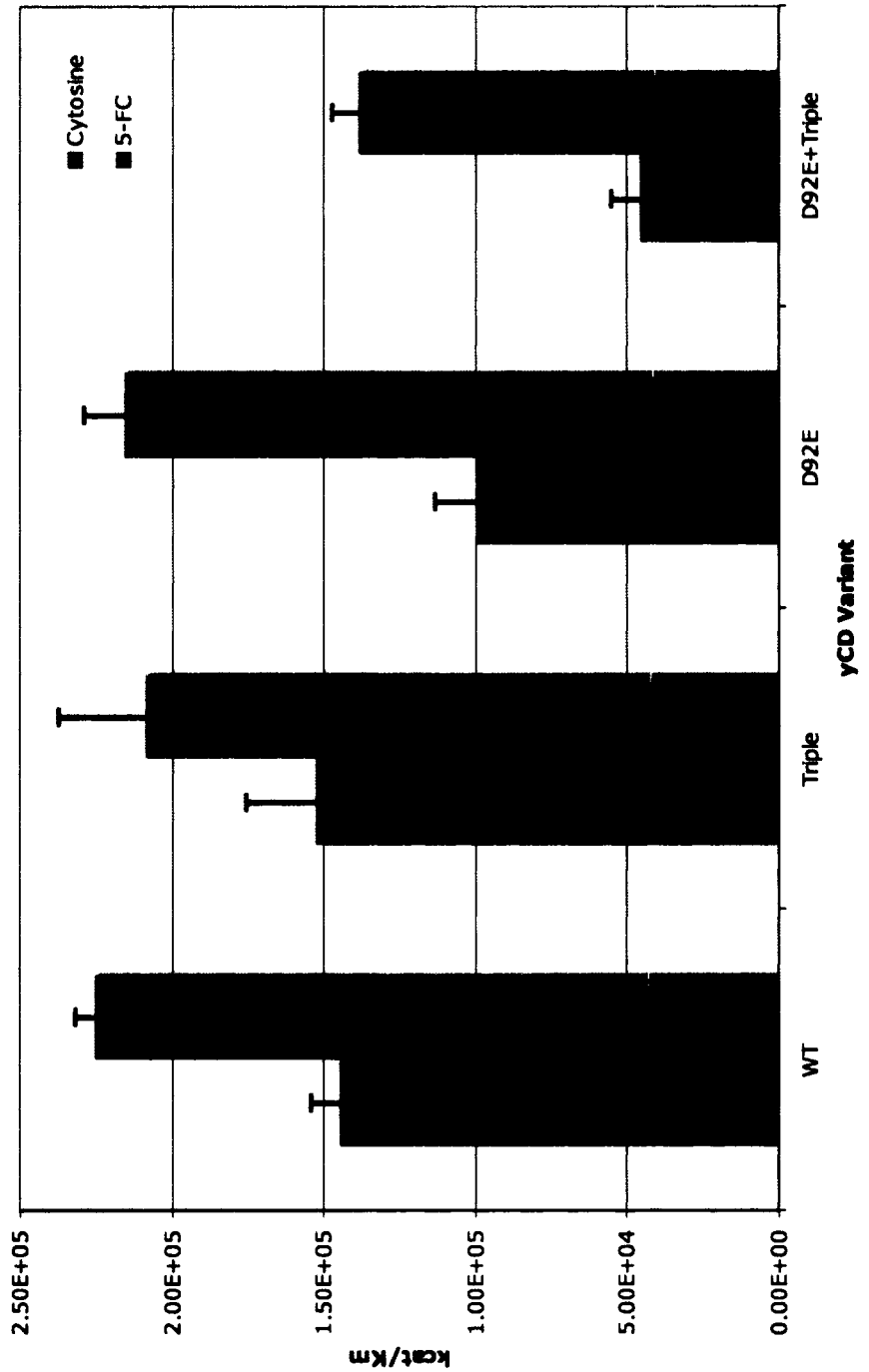


Figure 5.5 Catalytic Efficiencies of Yeast Cytosine Deaminase Variants for both Cytosine and 5-Fluorocytosine.

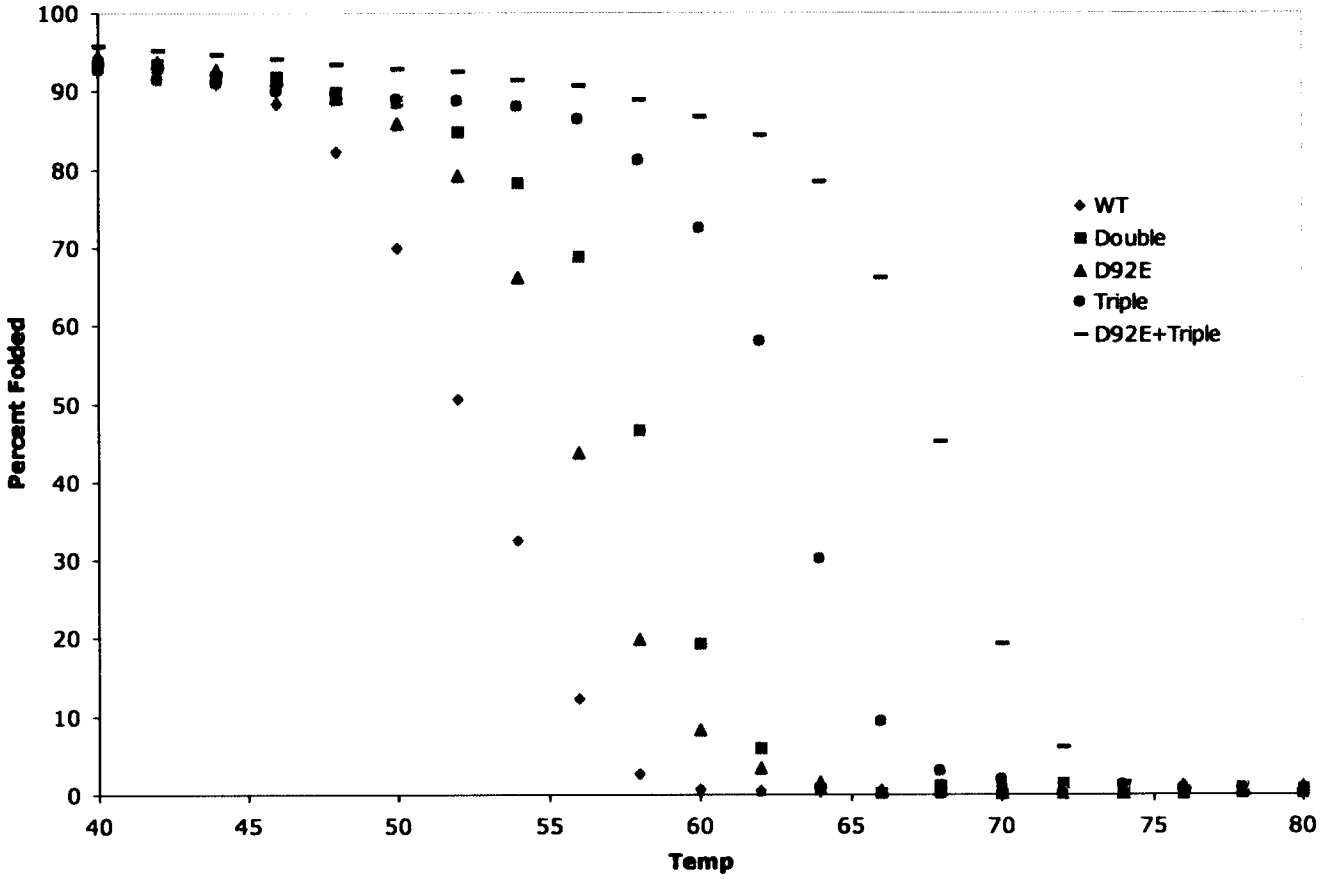


Figure 5.6. Thermal Denaturation Measurements of Yeast Cytosine Deaminase Variants. The temperature melt measures the change in signal at 220nm over a range of temperatures. All constructs show a folded baseline followed by a sigmoidal two-state transition to an unfolded baseline. The baseline plateaus correspond to an assignment of 100% folded protein.

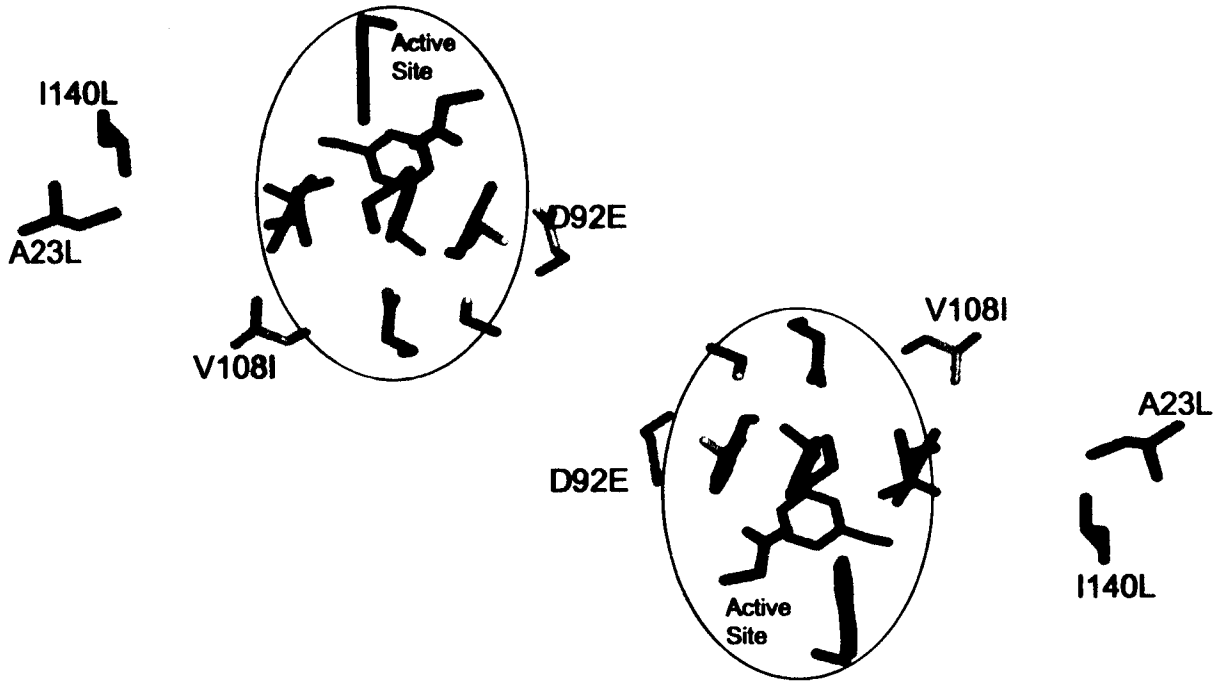


Figure 5.7 Location of Mutations Relative to the Active Site within the Yeast Cytosine Deaminase Structure. Triple mutant residues (A23L, V108I, I140L) highlighted in green, D92E mutant residue highlighted in purple and active site residues and substrate highlighted in pink.

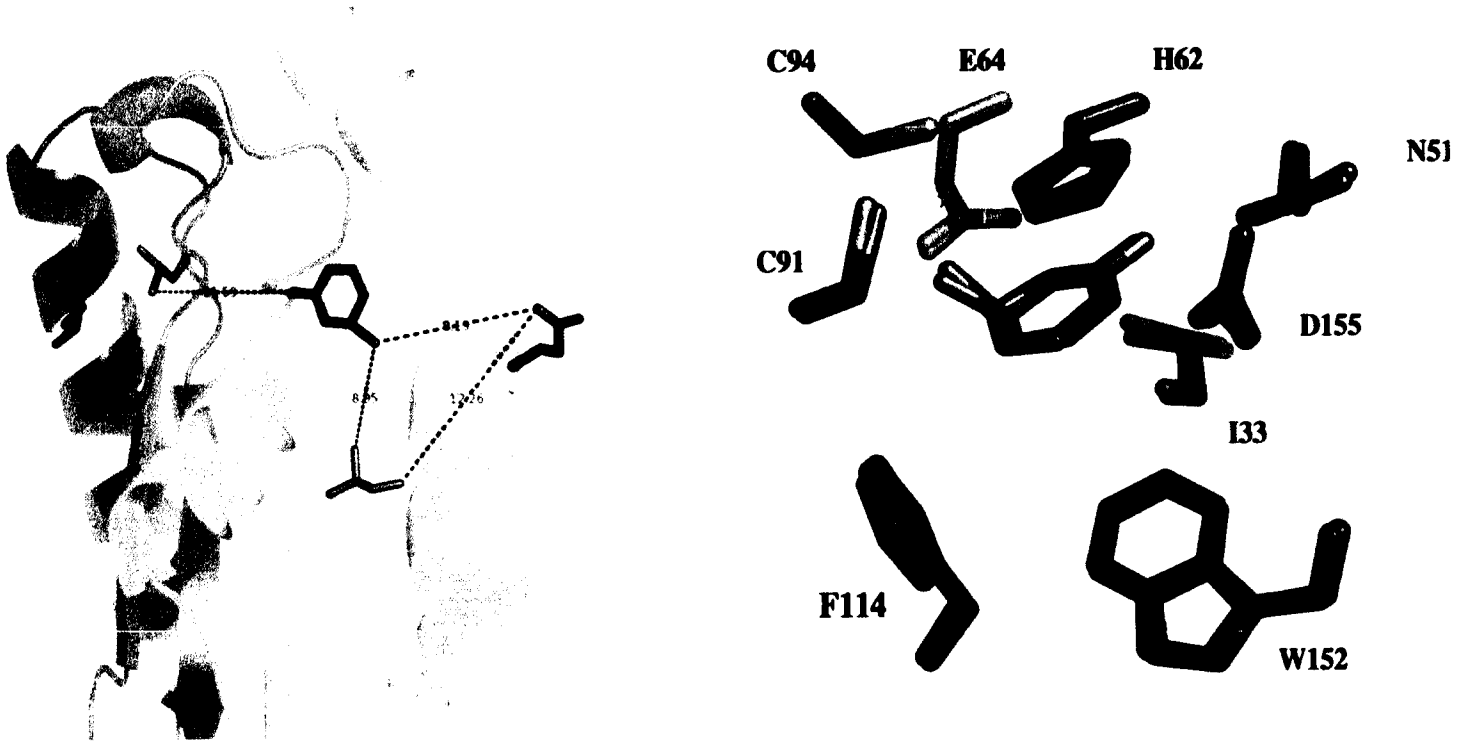


Figure 5.8 Crystal Structure of the Yeast Cytosine Deaminase D92E-Triple Active Site. **(A)** Relative distances of key residue interactions with a substrate analog (DHP). **(B)** Superimposition of wild-type and D92E active site residues.

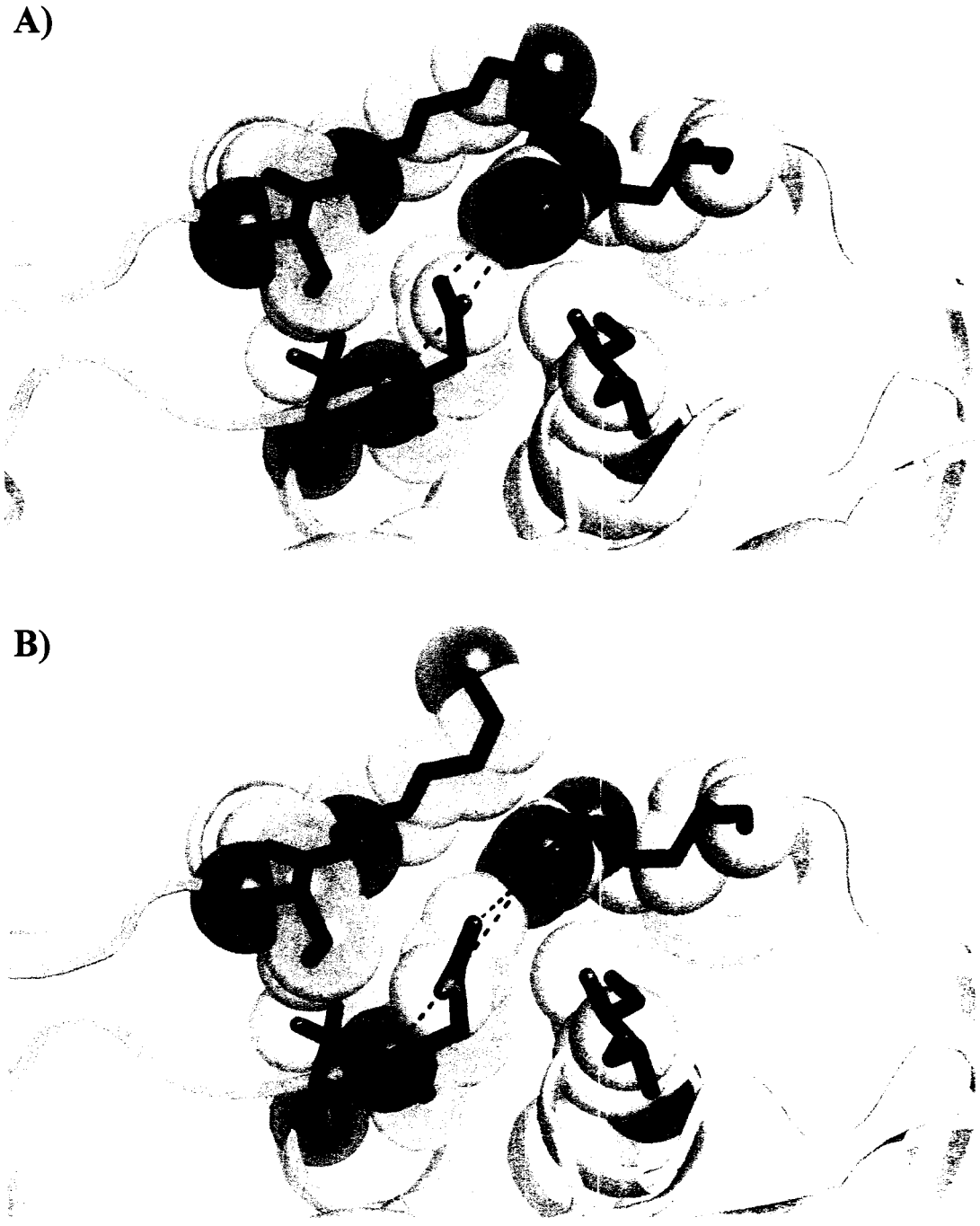


Figure 5.9 Interaction of D92E at the Dimer Interface. (**A**) wild-type crystal structure. (**B**) D92E-Triple crystal structure

Chapter Six: Discussion

A requirement of any effective enzyme for use in gene-directed enzyme/prodrug therapy (GDEPT) is its ability to catalyze the conversion of the prodrug to the active drug under physiological conditions. The therapeutic efficacy of the GDEPT enzyme yeast cytosine deaminase (yCD) is therefore compromised by its low catalytic efficiency and its low thermostability[13]. The goal of our research was to engineer yeast cytosine deaminase (yCD) in order to increase its GDEPT efficacy by improving the enzyme's catalytic turnover of the prodrug 5-fluorocytosine (5-FC) and increase the enzyme's thermostability.

A GDEPT enzyme's effectiveness is a direct measure of it is ability to catalyze the conversion of the prodrug to the active drug. The catalytic turnover of the prodrug 5-FC by cytosine deaminase is complicated by the fact that the enzyme's native substrate, cytosine, is in high abundance within the cell. Although the catalytic specificity for 5-FC is much higher in yCD than its GDEPT predecessor bCD,[13] its catalytic efficiency is low. In order to improve yCD's catalytic turnover of 5-FC we used a regiospecific random mutagenesis and selection protocol previously described by Black *et. al.*, 1996[14] in order to screen *E. coli* transformed with a codon specific randomized library of yCD mutants for increased sensitivity to 5-FC. The screen produced several distinct mutations that lead to an increase in 5-FC sensitivity relative to wild-type, the best of which was a point mutant D92E.

Another requirement of any effective GDEPT enzyme is that it must be stable and active under human physiological temperature and conditions. Previously in the literature it was reported that although yCD displays a much better catalytic profile than bCD for GDEPT using 5-FC, it's activity also displays a much lower half-life at 37° C[13]. To improve the thermostability of yCD we developed a novel protocol utilizing the computational protein design program RosettaDesign developed by Kuhlman *et. al.*, 2000[159] in order to select thermostabilizing mutations within yCD without disturbing the active site and catalytic function of the enzyme. The procedure generated many mutations from which we found 3 point mutations that when combined increased the thermostability of yCD by 10° C[152].

Both of these protocols generated yCD variants that had the desired characteristics as set by the design parameters: increased 5-FC sensitivity as a result of regiospecific random mutagenesis and increased thermostability as a result of computational protein design. To determine if these designs led to an improvement in efficacy for GDEPT, a 5-FC sensitivity assay was performed in both a *in vitro* cell culture and a *in vivo* mouse tumor xenograph model, using either wild-type enzymes or one of the following variants: the D92E mutant; the triple mutant; or a variant that had all four mutations (D92E-triple).

The cell culture assay showed significant decreases in 5-FC IC_{50} values for cancer cell lines transfected with D92E (2mM) or triple mutant yCD (1mM) relative to wild-type (>10mM). Similarly, the mouse tumor xenograph model showed

decreased tumor growth in tumors expressing D92E or triple mutant yCD relative to wild-type. The D92E-triple mutant, however, showed no change in IC_{50} relative to wild-type in the cell culture assay and as a result was not tested in the mouse tumor xenograph assay. The results of these two experiments demonstrate that both the D92E and the triple mutant alone show increased therapeutic efficacy for use in GDEPT over wild-type yCD but that when combined into a quadruple mutant the improvement was lost.

The results of the GDEPT efficacy experiments raised some important questions. Why would the D92E-triple mutant be less effective than either the D92E or triple mutants alone? We had hypothesized that both an increase in catalytic specificity and an increase in thermostability would lead to increased performance in GDEPT; why would layering these mutant variants lead to no change at all? To answer this we first had to determine how the mutant variants are different biophysically from wild-type yCD. To do this we examined enzyme kinetics, thermostabilities and crystal structures.

Catalytic efficiencies of wild-type yCD and all variants (D92E, triple and D92E-triple) against both the prodrug 5-FC and the native substrate cytosine were determined by measuring changes in UV absorption over time at room temperature (Table 5.1, Figure 5.5). As expected the results showed no effective change in activity or specificity for the triple mutant relative to wild-type. Unexpectedly however the D92E mutant shows very little change in catalytic efficiency with no change in 5-FC

kcat/K_m and only a very slight decrease in cytosine kcat/K_m. The layered mutant D92E-triple demonstrates a significant decrease in both 5-FC and cytosine catalytic efficiency with a slight improvement in specificity. Although the regiospecific random mutagenesis and selection procedure for 5-FC sensitivity selected HSV-TK mutants with significantly improved catalytic specificity for GCV,[14] the same does not appear to be true when the same procedure is applied to yCD. Since there appears to be no change in kinetics in the case of triple mutant, and only very slight changes in specificity in D92E, the significant differences in IC_{50} and tumor growth inhibition observed are unlikely to be a result of kinetic changes.

The thermostability of wild-type and the three yCD variants (D92E, triple and D92E-triple) were measured by monitoring change in circular dichroism over a range of temperatures (Figure 5.6). The results show a melting temperature of 52° C for wild-type, 62° C for triple, 56° C for D92E and 68° C for D92E-triple demonstrating that the D92E mutation is in fact a thermostabilizing mutation and that its increase in thermostability interacts synergistically with that of the triple mutant when combined into a quadruple mutant.

In order to determine what effect if any thermostabilization may have on the kinetics of the enzyme we determined the thermal activity profile of yCD by measuring the maximum kinetic turnover at saturating levels of cytosine (V_{max}) for both wild-type and the triple mutant over a range of temperatures (Figure 4.4). The results showed that the triple mutant has a slightly lower V_{max} than wild-type, as

previously observed at room temperature, but that the temperature at which V_{max} is at its maximum is slightly higher (42.5° C) than wild-type (40° C). The peak representing the temperature at which V_{max} is highest is also significantly broader in the triple mutant with changes in temperature away from optimal lead to less decrease in V_{max} to the point where V_{max} is actually significantly higher than wild-type at high temperature (>50° C).

In order to further investigate the cause of the observed changes in kinetics and thermostability of the enzyme variants, crystal structures were solved of both the triple mutant and the D92E-triple mutant (Table 5.2). The crystal structure of the D92E-triple mutant compared to the wild-type crystal structure shows no perturbations anywhere within the structure except in the immediate vicinity of the mutated residues. The triple mutant residues and D92E mutant residue are far apart (>10 angstroms) and are unlikely to have any influence upon each other (Figure 5.7 and 5.8a). This hypothesis is supported by the structure of the triple mutant alone where the A23L, V108I and I140L side-chains are in the exact same orientation as seen in the D92E-triple mutant structure. Therefore we hypothesis that the orientation of the D92E residue and its interactions with neighboring residues within the D92E-triple structure are representative of what a structure of the D92E mutation alone would be and therefore the D92E-triple mutant structure is sufficient to make inferences about the effect of the D92E mutation alone.

The yCD structures reveal that neither the mutations in the triple mutant (A23L, V108I, and I140L) or D92E are near the active site (Figure 5.8a). This result was expected for the triple mutant, which came from a structure-based design procedure that limited mutations to residues at least 4 angstroms from active site residues. The observation that D92 is not nearby the active site was not expected as D92 was selected along with other residues to be part of the random mutagenesis on the assumption that it was physically involved in the catalytic pocket. This assumption however was made based on data from multiple sequence alignments and not from structural information since at the time there was no crystal structure available. Interestingly although D92 is only one residue away from catalytic cysteine C91 the helical twist of the backbone points D92 away from the active site (Figure 5.8a) and towards the dimer interface where it forms a salt-bridge with an arganine on the neighboring monomer (Figure 5.9a). As a result neither the triple mutant nor the D92E-triple structures show any perturbation in the position of active site residues or the substrate relative to that of the wild-type yCD crystal structure (Figures 4.6b and 5.8b). These observations are in agreement with the lack of change in kinetics measured for both the D92E and triple mutants and demonstrate that the small loss in activity measured for D92E-triple cannot be attributed to any specific change within the enzyme active site.

Direct measurement of kcat and K_m values for the D92E-triple enzyme against both cytosine and 5-FC show that the reduction in catalytic efficiency against both

substrates is attributable to a significant reduction in kcat (kcat 8-fold lower for cytosine and 2-fold lower for 5-FC) (Table 5.1). A reduction in catalytic rate can be due to either slower catalytic turnover due to changes in the chemical environment of the active site or due to slower release of product from the active site. The observations that the active-site of D92E-triple appears unchanged and that the enzymes melting temperature (T_m) is 16° C higher than wild-type led to the hypothesis that the observed loss of activity in the D92E-triple mutant may be due to a loss in the dynamic flexibility of the structure relative to wild-type, leading to a reduction in the rate of product release at the temperature the kinetics measurements were made (room temperature). Such a reduction in the rate of product release would have an effect on overaly catalytic efficiency of the enzyme as it has been demonstrated that in the conversion of 5-FC by yCD product release is the rate-limiting step[206]. This hypothesis could be tested by measuring the on and off rates of the substrate and product for both wild-type and D92E-triple yCD. If correct we would expect to see slower off-rates in the hyperthermostabilized D92E-triple mutant relative to wild-type.

The crystal structures also give insight regarding the location of the mutations and their interaction with surrounding residues and how they might be thermostabilizing. The three mutations in the triple mutant substitute a hydrophobic residue for either a bulkier hydrophobic residue (A23L, V108I) or for a hydrophobic side-chain with altered topology in the case of I140L. The crystal structure of the triple mutant shows that these mutations lead to the burial of additional hydrophobic

surface area within the core of the protein. Improving hydrophobic packing within a protein core has been attributed numerous times to the thermostabilization of a protein and is the most likely cause of the thermostabilization seen in the triple mutant.

Unlike the mutations within the triple mutant D92E deals with the exchange of charged residues. The crystal structure containing the D92E mutation reveals that the addition of an extra carbon atom in the glutamate side-chain relative to aspartate appears to disrupt a salt-bridge formed across the dimer interface with nearby arganine 304 (Figure 5.9). Why such a mutation would be thermostabilizing is unclear from the crystal structure however stability is not just a factor of the folded state but rather of the equilibrium between the folded and unfolded states. It is possible that the D92E mutation shifts this equilibrium in such a way that favors the folded state perhaps by further destabilizing the unfolded state. As yCD is a dimer and D92E is located within the dimer interface the mutation may also have an influence on the equilibrium between the monomeric and dimeric states. This hypothesis could be tested by looking at the kinetics of folding of both wild-type and D92E-triple as well as the equilibrium between the monomeric and dimeric states over a range of concentrations.

Electrostatics are mathematically more complicated to model making them computationally expensive and as a result many computational design programs such as RosettaDesign rely on empirically generated models based on statistical inferences from the pdb. As a result the modeling of electrostatic interactions by computational design programs is significantly less robust than those of hydrophobic interactions.

Although the location of D92 nearby the active site excluded it from the original computational design experiment it seems unlikely that the program would register such a mutation as favorable. This underscores the importance of using a variety of design approaches including directed evolution approaches, which can find solutions which current computational design programs are unlikely to consider.

The fact that both the D92E and triple mutant are thermostabilizing yet show little to no change at all in kinetic profile relative to wild-type yCD suggest that this increase in thermostability is the cause of the observed improvement in therapeutic efficacy. This hypothesis is supported by the observation that a random mutagenesis and screening protocol designed to select for increased sensitivity to 5-FC selected a thermostabilizing mutation rather than a kinetically altered mutant. More support comes from the observation that *E. coli* made dependant on yCD for growth, by knocking out of the uracil catabolic pathway and rescuing it with yCD, show significantly elevated growth rates at 37° C when the rescue is done with triple mutant yCD relative to wild-type yCD (Figure 4.5). The hypothesis however does appear to be in conflict with the observed 52° C melting temperature of wild-type yCD and the thermal kinetic profile of the wild-type enzyme showing 37° C to be near the peak efficiency. Why would thermostabilizing an enzyme that appears to already be quite stable at 37° C *in vitro* have such a dramatic effect on the efficacy of therapy dependent on that enzyme *in vivo*?

There are several possible explanations. First, although there is no apparent difference in kinetics between wild-type and the yCD variants at room temperature perhaps at 37° C the catalytic efficiencies of the enzymes are significantly shifted in order to favor 5-FC conversion over cytosine. This explanation is not necessarily in conflict with the observed thermal profile data (Figure 4.4) since that data is measuring only V_{max} only for the triple mutant and wild-type yCD and only against cytosine. It is possible that the catalytic efficiency and/or specificity of these enzymes are shifted at 37° C relative to wild-type. In order to test this the enzyme kinetics would need to be remeasured at 37° C or a range of temperatures around and including 37° C.

Another possibility is that the improved efficacy in therapy observed is just due to an increase in the concentration of the mutant enzymes relative to that of wild-type yCD. This may be due to either increase in thermostability resulting in an increase enzyme half-life and subsequently higher enzyme concentration over time or just a higher level of enzyme expression of the yCD variants. If either of these were the case one would expect higher concentrations of thermostabilized yCD variant protein in cells relative to that of wild-type yCD. Western blots of cell lines expressing either wild-type, D92E or triple mutant yCD showed no significant difference in enzyme concentrations. This suggests that a change in enzyme levels within the cell is not the cause of improved efficacy in therapy observed however protein levels are difficult to quantitate using western blots and if there was a

difference in protein concentration a western-blot would not distinguish between an increase in expression level or an increase in protein half-life. A pulse-chase experiment measuring protein concentration levels of tagged protein over time would give a much more definitive result.

Further work will be needed to determine why the D92E mutant is thermostabilizing and why thermostabilization appears to have such a significant impact on therapeutic efficacy when *in vitro* yCD appears to be stable and active at 37° C. Nevertheless the observed decrease in 5-FC IC_{50} and tumor growth rate in cancer cells transfected with either the D92E and Triple mutant yCD variants suggest that these enzymes may have real world benefit in the clinic for use in GDEPT. Dr. Svend Freytag who is leading currently ongoing clinical trials of GDEPT utilizing the yCD/5-FC system at the Henry Ford medical center has expressed direct interest in using our improved yCD variants in future trials (personal communication)[221]. Although there is an established precedent in the Freytag group for utilizing protein engineered enzymes in GDEPT clinical trials[11] use of the triple mutant yCD variant would represent the very first example of a computationally engineered enzyme being used in human therapy.

BIBLIOGRAPHY

1. Jain, R.K. Delivery of molecular and cellular medicine to solid tumors. *Adv Drug Deliv Rev* **46**, 149-68 (2001).

2. Melo, J.V. The molecular biology of chronic myeloid leukaemia. *Leukemia* **10**, 751-6 (1996).

3. Druker, B.J. et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* **344**, 1031-7 (2001).

4. Sawyers, C.L. et al. Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase II study. *Blood* **99**, 3530-9 (2002).

5. Kaelin, W.G., Jr. Gleevec: prototype or outlier? *Sci STKE* **2004**, pe12 (2004).

6. Moolten, F.L. Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy. *Cancer Res* **46**, 5276-81 (1986).

7. Portsmouth, D., Hlavaty, J. & Renner, M. Suicide genes for cancer therapy. *Mol Aspects Med* **28**, 4-41 (2007).

8. Dachs, G.U., Tupper, J. & Tozer, G.M. From bench to bedside for gene-directed enzyme prodrug therapy of cancer. *Anticancer Drugs* **16**, 349-59 (2005).

9. Freytag, S.O. et al. Efficacy and toxicity of replication-competent adenovirus-mediated double suicide gene therapy in combination with radiation therapy in an orthotopic mouse prostate cancer model. *Int J Radiat Oncol Biol Phys* **54**, 873-85 (2002).

10. Freytag, S.O. et al. Phase I study of replication-competent adenovirus-mediated double-suicide gene therapy in combination with conventional-dose three-dimensional conformal radiation therapy for the treatment of newly diagnosed, intermediate- to high-risk prostate cancer. *Cancer Res* **63**, 7497-506 (2003).

11. Barton, K.N. et al. Second-generation replication-competent oncolytic adenovirus armed with improved suicide genes and ADP gene demonstrates greater efficacy without increased toxicity. *Mol Ther* **13**, 347-56 (2006).

12. Freytag, S.O., Stricker, H., Movsas, B. & Kim, J.H. Prostate cancer gene therapy clinical trials. *Mol Ther* **15**, 1042-52 (2007).

13. Kievit, E. et al. Superiority of yeast over bacterial cytosine deaminase for enzyme/prodrug gene therapy in colon cancer xenografts. *Cancer Res* **59**, 1417-21 (1999).

14. Black, M.E., Newcomb, T.G., Wilson, H.M. & Loeb, L.A. Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc Natl Acad Sci U S A* **93**, 3525-9 (1996).

15. Black, M.E., Kokoris, M.S. & Sabo, P. Herpes simplex virus-1 thymidine kinase mutants created by semi-random sequence mutagenesis improve prodrug-mediated tumor cell killing. *Cancer Res* **61**, 3022-6 (2001).

16. Huber, B.E., Richards, C.A. & Krenitsky, T.A. Retroviral-mediated gene therapy for the treatment of hepatocellular carcinoma: an innovative approach for cancer therapy. *Proc Natl Acad Sci U S A* **88**, 8039-43 (1991).

17. Schepelmann, S. & Springer, C.J. Viral vectors for gene-directed enzyme prodrug therapy. *Curr Gene Ther* **6**, 647-70 (2006).

18. Warrington, K.H., Jr., Teschendorf, C., Cao, L., Muzyczka, N. & Siemann, D.W. Developing VDEPT for DT-diaphorase (NQO1) using an AAV vector plasmid. *Int J Radiat Oncol Biol Phys* **42**, 909-12 (1998).

19. Kanazawa, T. et al. Gamma-rays enhance rAAV-mediated transgene expression and cytocidal effect of AAV-HSVtk/ganciclovir on cancer cells. *Cancer Gene Ther* **8**, 99-106 (2001).

20. Fukui, T. et al. Suicide gene therapy for human oral squamous cell carcinoma cell lines with adeno-associated virus vector. *Oral Oncol* **37**, 211-5 (2001).

21. Westphal, E.M., Ge, J., Catchpole, J.R., Ford, M. & Kenney, S.C. The nitroreductase/CB1954 combination in Epstein-Barr virus-positive B-cell lines: induction of bystander killing in vitro and in vivo. *Cancer Gene Ther* **7**, 97-106 (2000).

22. Huebner, R.J., Rowe, W.P., Schatten, W.E., Smith, R.R. & Thomas, L.B. Studies on the use of viruses in the treatment of carcinoma of the cervix. *Cancer* **9**, 1211-8 (1956).

23. Whyte, P., Ruley, H.E. & Harlow, E. Two regions of the adenovirus early region 1A proteins are required for transformation. *J Virol* **62**, 257-65 (1988).

24. Tollefson, A.E., Ryerse, J.S., Scaria, A., Hermiston, T.W. & Wold, W.S. The E3-11.6-kDa adenovirus death protein (ADP) is required for efficient cell death: characterization of cells infected with adp mutants. *Virology* **220**, 152-62 (1996).

25. Jia, W. & Zhou, Q. Viral vectors for cancer gene therapy: viral dissemination and tumor targeting. *Curr Gene Ther* **5**, 133-42 (2005).

26. Zhang, W.W. Development and application of adenoviral vectors for gene therapy of cancer. *Cancer Gene Ther* **6**, 113-38 (1999).

27. Bischoff, J.R. et al. An adenovirus mutant that replicates selectively in p53-deficient human tumor cells. *Science* **274**, 373-6 (1996).

28. Hu, Q.J., Dyson, N. & Harlow, E. The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. *Embo J* **9**, 1147-55 (1990).

29. Houweling, A., van den Elsen, P.J. & van der Eb, A.J. Partial transformation of primary rat cells by the leftmost 4.5% fragment of adenovirus 5 DNA. *Virology* **105**, 537-50 (1980).

30. Heise, C. et al. An adenovirus E1A mutant that demonstrates potent and selective systemic anti-tumoral efficacy. *Nat Med* **6**, 1134-9 (2000).

31. Kao, C.C., Yew, P.R. & Berk, A.J. Domains required for in vitro association between the cellular p53 and the adenovirus 2 E1B 55K proteins. *Virology* **179**, 806-14 (1990).

32. White, E. Tumour biology. p53, guardian of Rb. *Nature* **371**, 21-2 (1994).

33. Kirn, D., Hermiston, T. & McCormick, F. ONYX-015: clinical data are encouraging. *Nat Med* **4**, 1341-2 (1998).

34. Kirn, D. Clinical research results with dl1520 (Onyx-015), a replication-selective adenovirus for the treatment of cancer: what have we learned? *Gene Ther* **8**, 89-98 (2001).

35. Ganly, I. et al. A phase I study of Onyx-015, an E1B attenuated adenovirus, administered intratumorally to patients with recurrent head and neck cancer. *Clin Cancer Res* **6**, 798-806 (2000).

36. Freytag, S.O. et al. Phase I study of replication-competent adenovirus-mediated double suicide gene therapy for the treatment of locally recurrent prostate cancer. *Cancer Res* **62**, 4968-76 (2002).

37. Raper, S.E. et al. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol Genet Metab* **80**, 148-58 (2003).

38. Walther, W. & Stein, U. Viral vectors for gene transfer: a review of their use in the treatment of human diseases. *Drugs* **60**, 249-71 (2000).

39. McNeish, I.A. et al. Virus directed enzyme prodrug therapy for ovarian and pancreatic cancer using retrovirally delivered E. coli nitroreductase and CB1954. *Gene Ther* **5**, 1061-9 (1998).

40. Green, N.K. et al. Sensitization of colorectal and pancreatic cancer cell lines to the prodrug 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB1954) by retroviral transduction and expression of the E. coli nitroreductase gene. *Cancer Gene Ther* **4**, 229-38 (1997).

41. Pawelek, J.M., Low, K.B. & Bermudes, D. Tumor-targeted Salmonella as a novel anticancer vector. *Cancer Res* **57**, 4537-44 (1997).

42. Bermudes, D., Low, B. & Pawelek, J. Tumor-targeted Salmonella. Highly selective delivery vectors. *Adv Exp Med Biol* **465**, 57-63 (2000).

43. Niculescu-Duvaz, I. & Springer, C.J. Introduction to the background, principles, and state of the art in suicide gene therapy. *Mol Biotechnol* **30**, 71-88 (2005).

44. Freeman, S.M. et al. The "bystander effect": tumor regression when a fraction of the tumor mass is genetically modified. *Cancer Res* **53**, 5274-83 (1993).

45. Huber, B.E., Austin, E.A., Richards, C.A., Davis, S.T. & Good, S.S. Metabolism of 5-fluorocytosine to 5-fluorouracil in human colorectal tumor cells transduced with the cytosine deaminase gene: significant antitumor effects when only a small percentage of tumor cells express cytosine deaminase. *Proc Natl Acad Sci U S A* **91**, 8302-6 (1994).

46. Dilber, M.S. et al. Gap junctions promote the bystander effect of herpes simplex virus thymidine kinase in vivo. *Cancer Res* **57**, 1523-8 (1997).

47. Freeman, S.M., Ramesh, R. & Marrogi, A.J. Immune system in suicide-gene therapy. *Lancet* **349**, 2-3 (1997).

48. Kuriyama, S. et al. Cytosine deaminase/5-fluorocytosine gene therapy can induce efficient anti-tumor effects and protective immunity in immunocompetent mice but not in athymic nude mice. *Int J Cancer* **81**, 592-7 (1999).

49. Kuriyama, S. et al. Cancer gene therapy with HSV-tk/GCV system depends on T-cell-mediated immune responses and causes apoptotic death of tumor cells in vivo. *Int J Cancer* **83**, 374-80 (1999).

50. Vile, R.G., Nelson, J.A., Castleden, S., Chong, H. & Hart, I.R. Systemic gene therapy of murine melanoma using tissue specific expression of the HSVtk gene involves an immune component. *Cancer Res* **54**, 6228-34 (1994).

51. Gagandeep, S. et al. Prodrug-activated gene therapy: involvement of an immunological component in the "bystander effect". *Cancer Gene Ther* **3**, 83-8 (1996).

52. Wilson, K.M. et al. HSV-tk gene therapy in head and neck squamous cell carcinoma. Enhancement by the local and distant bystander effect. *Arch Otolaryngol Head Neck Surg* **122**, 746-9 (1996).

53. Dilber, M.S. & Smith, C.I. Suicide genes and bystander killing: local and distant effects. *Gene Ther* **4**, 273-4 (1997).

54. Kuriyama, S. et al. Tissue-specific expression of HSV-tk gene can induce efficient antitumor effect and protective immunity to wild-type hepatocellular carcinoma. *Int J Cancer* **71**, 470-5 (1997).

55. Yamamoto, S., Suzuki, S., Hoshino, A., Akimoto, M. & Shimada, T. Herpes simplex virus thymidine kinase/ganciclovir-mediated killing of tumor cell

induces tumor-specific cytotoxic T cells in mice. *Cancer Gene Ther* **4**, 91-6 (1997).

56. Patterson, A. & Harris, A.L. Molecular chemotherapy for breast cancer. *Drugs Aging* **14**, 75-90 (1999).

57. Freeman, S.M. et al. The treatment of ovarian cancer with a gene modified cancer vaccine: a phase I study. *Hum Gene Ther* **6**, 927-39 (1995).

58. Herman, J.R. et al. In situ gene therapy for adenocarcinoma of the prostate: a phase I clinical trial. *Hum Gene Ther* **10**, 1239-49 (1999).

59. Sterman, D.H. et al. Adenovirus-mediated herpes simplex virus thymidine kinase/ganciclovir gene therapy in patients with localized malignancy: results of a phase I clinical trial in malignant mesothelioma. *Hum Gene Ther* **9**, 1083-92 (1998).

60. Shand, N. et al. A phase 1-2 clinical trial of gene therapy for recurrent glioblastoma multiforme by tumor transduction with the herpes simplex thymidine kinase gene followed by ganciclovir. GLI328 European-Canadian Study Group. *Hum Gene Ther* **10**, 2325-35 (1999).

61. Rainov, N.G. et al. Immune response induced by retrovirus-mediated HSV-tk/GCV pharmacogene therapy in patients with glioblastoma multiforme. *Gene Ther* **7**, 1853-8 (2000).

62. Rainov, N.G. A phase III clinical evaluation of herpes simplex virus type 1 thymidine kinase and ganciclovir gene therapy as an adjuvant to surgical resection and radiation in adults with previously untreated glioblastoma multiforme. *Hum Gene Ther* **11**, 2389-401 (2000).

63. Wallace, H., Clarke, A.R., Harrison, D.J., Hooper, M.L. & Bishop, J.O. Ganciclovir-induced ablation non-proliferating thyrocytes expressing herpesvirus thymidine kinase occurs by p53-independent apoptosis. *Oncogene* **13**, 55-61 (1996).

64. Caruso, M. et al. Regression of established macroscopic liver metastases after in situ transduction of a suicide gene. *Proc Natl Acad Sci U S A* **90**, 7024-8 (1993).

65. Kokoris, M.S., Sabo, P., Adman, E.T. & Black, M.E. Enhancement of tumor ablation by a selected HSV-1 thymidine kinase mutant. *Gene Ther* **6**, 1415-26 (1999).

66. Kussmann-Gerber, S., Kuonen, O., Folkers, G., Pilger, B.D. & Scapozza, L. Drug resistance of herpes simplex virus type 1--structural considerations at the molecular level of the thymidine kinase. *Eur J Biochem* **255**, 472-81 (1998).

67. Vogt, J. et al. Nucleoside binding site of herpes simplex type 1 thymidine kinase analyzed by X-ray crystallography. *Proteins* **41**, 545-53 (2000).

68. Sulpizi, M., Schelling, P., Folkers, G., Carloni, P. & Scapozza, L. The rational of catalytic activity of herpes simplex virus thymidine kinase. a combined biochemical and quantum chemical study. *J Biol Chem* **276**, 21692-7 (2001).

69. Mullen, C.A., Kilstrup, M. & Blaese, R.M. Transfer of the bacterial gene for cytosine deaminase to mammalian cells confers lethal sensitivity to 5-fluorocytosine: a negative selection system. *Proc Natl Acad Sci U S A* **89**, 33-7 (1992).

70. Santi, D.V., McHenry, C.S. & Sommer, H. Mechanism of interaction of thymidylate synthetase with 5-fluorodeoxyuridylate. *Biochemistry* **13**, 471-81 (1974).

71. Ghoshal, K. & Jacob, S.T. Specific inhibition of pre-ribosomal RNA processing in extracts from the lymphosarcoma cells treated with 5-fluorouracil. *Cancer Res* **54**, 632-6 (1994).

72. Kuriyama, S. et al. Bystander effect caused by cytosine deaminase gene and 5-fluorocytosine in vitro is substantially mediated by generated 5-fluorouracil. *Anticancer Res* **18**, 3399-406 (1998).

73. Kievit, E. et al. Yeast cytosine deaminase improves radiosensitization and bystander effect by 5-fluorocytosine of human colorectal cancer xenografts. *Cancer Res* **60**, 6649-55 (2000).

74. Cobb, L.M. et al. 2,4-dinitro-5-ethyleneiminobenzamide (CB 1954): a potent and selective inhibitor of the growth of the Walker carcinoma 256. *Biochem Pharmacol* **18**, 1519-27 (1969).

75. Knox, R.J., Friedlos, F., Jarman, M. & Roberts, J.J. A new cytotoxic, DNA interstrand crosslinking agent, 5-(aziridin-1-yl)-4-hydroxylamino-2-

nitrobenzamide, is formed from 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB 1954) by a nitroreductase enzyme in Walker carcinoma cells. *Biochem Pharmacol* **37**, 4661-9 (1988).

76. Anlezark, G.M. et al. The bioactivation of 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB1954)--I. Purification and properties of a nitroreductase enzyme from Escherichia coli--a potential enzyme for antibody-directed enzyme prodrug therapy (ADEPT). *Biochem Pharmacol* **44**, 2289-95 (1992).

77. Bridgewater, J.A. et al. Expression of the bacterial nitroreductase enzyme in mammalian cells renders them selectively sensitive to killing by the prodrug CB1954. *Eur J Cancer* **31A**, 2362-70 (1995).

78. Weedon, S.J. et al. Sensitisation of human carcinoma cells to the prodrug CB1954 by adenovirus vector-mediated expression of E. coli nitroreductase. *Int J Cancer* **86**, 848-54 (2000).

79. Chung-Faye, G. et al. Virus-directed, enzyme prodrug therapy with nitroimidazole reductase: a phase I and pharmacokinetic study of its prodrug, CB1954. *Clin Cancer Res* **7**, 2662-8 (2001).

80. Searle, P.F. et al. Nitroreductase: a prodrug-activating enzyme for cancer gene therapy. *Clin Exp Pharmacol Physiol* **31**, 811-6 (2004).

81. Green, N.K., Kerr, D.J., Mautner, V., Harris, P.A. & Searle, P.F. The nitroreductase/CB1954 enzyme-prodrug system. *Methods Mol Med* **90**, 459-77 (2004).

82. Clarke, L. & Waxman, D.J. Oxidative metabolism of cyclophosphamide: identification of the hepatic monooxygenase catalysts of drug activation. *Cancer Res* **49**, 2344-50 (1989).

83. Jounaidi, Y. & Waxman, D.J. Combination of the bioreductive drug tirapazamine with the chemotherapeutic prodrug cyclophosphamide for P450/P450-reductase-based cancer gene therapy. *Cancer Res* **60**, 3761-9 (2000).

84. Pawlik, T.M. et al. Prodrug bioactivation and oncolysis of diffuse liver metastases by a herpes simplex virus 1 mutant that expresses the CYP2B1 transgene. *Cancer* **95**, 1171-81 (2002).

85. Salmons, B., Lohr, M. & Gunzburg, W.H. Treatment of inoperable pancreatic carcinoma using a cell-based local chemotherapy: results of a phase I/II clinical trial. *J Gastroenterol* **38 Suppl 15**, 78-84 (2003).

86. Chang, T.K., Weber, G.F., Crespi, C.L. & Waxman, D.J. Differential activation of cyclophosphamide and ifosphamide by cytochromes P-450 2B and 3A in human liver microsomes. *Cancer Res* **53**, 5629-37 (1993).

87. Uckert, W. et al. Double suicide gene (cytosine deaminase and herpes simplex virus thymidine kinase) but not single gene transfer allows reliable elimination of tumor cells in vivo. *Hum Gene Ther* **9**, 855-65 (1998).

88. Kim, J.H., Kim, S.H., Brown, S.L. & Freytag, S.O. Selective enhancement by an antiviral agent of the radiation-induced cell killing of human glioma cells transduced with HSV-tk gene. *Cancer Res* **54**, 6053-6 (1994).

89. Bruso, C.E., Shewach, D.S. & Lawrence, T.S. Fluorodeoxyuridine-induced radiosensitization and inhibition of DNA double strand break repair in human colon cancer cells. *Int J Radiat Oncol Biol Phys* **19**, 1411-7 (1990).

90. Koyama, F. et al. Adenoviral-mediated transfer of *Escherichia coli* uracil phosphoribosyltransferase (UPRT) gene to modulate the sensitivity of the human colon cancer cells to 5-fluorouracil. *Eur J Cancer* **36**, 2403-2410 (2000).

91. Chung-Faye, G.A. et al. In vivo gene therapy for colon cancer using adenovirus-mediated transfer of the fusion gene cytosine deaminase and uracil phosphoribosyltransferase. *Gene Ther* **8**, 1547-1554 (2001).

92. Adachi, Y. et al. Experimental gene therapy for brain tumors using adenovirus-mediated transfer of cytosine deaminase gene and uracil phosphoribosyltransferase gene with 5-fluorocytosine. *Hum Gene Ther* **11**, 77-89 (2000).

93. Erbs, P. et al. In vivo cancer gene therapy by adenovirus-mediated transfer of a bifunctional yeast cytosine deaminase/uracil phosphoribosyltransferase fusion gene. *Cancer Res* **60**, 3813-22 (2000).

94. Mesnil, M. & Yamasaki, H. Bystander effect in herpes simplex virus-thymidine kinase/ganciclovir cancer gene therapy: role of gap-junctional intercellular communication. *Cancer Res* **60**, 3989-99 (2000).

95. Chen, S.H. et al. Combination suicide and cytokine gene therapy for hepatic metastases of colon carcinoma: sustained antitumor immunity prolongs animal survival. *Cancer Res* **56**, 3758-62 (1996).

96. Ju, D.W., Wang, B.M. & Cao, X. Adenovirus-mediated combined suicide gene and interleukin-2 gene therapy for the treatment of established tumor and induction of antitumor immunity. *J Cancer Res Clin Oncol* **124**, 683-9 (1998).

97. Palu, G. et al. Gene therapy of glioblastoma multiforme via combined expression of suicide and cytokine genes: a pilot study in humans. *Gene Ther* **6**, 330-7 (1999).

98. Hamstra, D.A., Rice, D.J., Fahmy, S., Ross, B.D. & Rehemtulla, A. Enzyme/prodrug therapy for head and neck cancer using a catalytically superior cytosine deaminase. *Hum Gene Ther* **10**, 1993-2003 (1999).

99. Encell, L.P., Landis, D.M. & Loeb, L.A. Improving enzymes for cancer gene therapy. *Nat Biotechnol* **17**, 143-7 (1999).

100. Black, M.E. Enzyme and pathway engineering for suicide gene therapy. *Genet Eng (N Y)* **23**, 113-27 (2001).

101. Pantuck, A.J. et al. Optimizing prostate cancer suicide gene therapy using herpes simplex virus thymidine kinase active site variants. *Hum Gene Ther* **13**, 777-89 (2002).

102. Christians, F.C., Scapozza, L., Crameri, A., Folkers, G. & Stemmer, W.P. Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat Biotechnol* **17**, 259-64 (1999).

103. Race, P.R. et al. Kinetic and structural characterisation of Escherichia coli nitroreductase mutants showing improved efficacy for the prodrug substrate CB1954. *J Mol Biol* **368**, 481-92 (2007).

104. Turner, N.J. Directed evolution of enzymes for applied biocatalysis. *Trends Biotechnol* **21**, 474-8 (2003).

105. Johannes, T.W. & Zhao, H. Directed evolution of enzymes and biosynthetic pathways. *Curr Opin Microbiol* **9**, 261-7 (2006).

106. Reetz, M.T. et al. Enhancing the enantioselectivity of an epoxide hydrolase by directed evolution. *Org Lett* **6**, 177-80 (2004).

107. van Loo, B. et al. Directed evolution of epoxide hydrolase from A. radiobacter toward higher enantioselectivity by error-prone PCR and DNA shuffling. *Chem Biol* **11**, 981-90 (2004).

108. Cho, C.M., Mulchandani, A. & Chen, W. Altering the substrate specificity of organophosphorus hydrolase for enhanced hydrolysis of chlorpyrifos. *Appl Environ Microbiol* **70**, 4681-5 (2004).

109. Gould, S.M. & Tawfik, D.S. Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* **44**, 5444-52 (2005).

110. Castle, L.A. et al. Discovery and directed evolution of a glyphosate tolerance gene. *Science* **304**, 1151-4 (2004).

111. Wong, D.W., Batt, S.B., Lee, C.C. & Robertson, G.H. High-activity barley alpha-amylase by directed evolution. *Protein J* **23**, 453-60 (2004).

112. Solbak, A.I. et al. Discovery of pectin-degrading enzymes and directed evolution of a novel pectate lyase for processing cotton fabric. *J Biol Chem* **280**, 9431-8 (2005).

113. Garrett, J.B. et al. Enhancing the thermal tolerance and gastric performance of a microbial phytase for use as a phosphate-mobilizing monogastric-feed supplement. *Appl Environ Microbiol* **70**, 3041-6 (2004).

114. Palackal, N. et al. An evolutionary route to xylanase process fitness. *Protein Sci* **13**, 494-503 (2004).

115. Seng Wong, T., Arnold, F.H. & Schwaneberg, U. Laboratory evolution of cytochrome p450 BM-3 monooxygenase for organic cosolvents. *Biotechnol Bioeng* **85**, 351-8 (2004).

116. van den Berg, S., Lofdahl, P.A., Hard, T. & Berglund, H. Improved solubility of TEV protease by directed evolution. *J Biotechnol* **121**, 291-8 (2006).

117. McLoughlin, S.Y., Jackson, C., Liu, J.W. & Ollis, D. Increased expression of a bacterial phosphotriesterase in Escherichia coli through directed evolution. *Protein Expr Purif* **41**, 433-40 (2005).

118. Cadwell, R.C. & Joyce, G.F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **2**, 28-33 (1992).

119. Stemmer, W.P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-91 (1994).

120. Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* **91**, 10747-51 (1994).

121. Leung, D.W., Chen, E. & Goeddel, D.V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**, 11-15 (1989).

122. Crameri, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288-91 (1998).

123. Matsumura, M. & Aiba, S. Screening for thermostable mutant of kanamycin nucleotidyltransferase by use of a transformation system for a thermophile, *Bacillus stearothermophilus*. *J Biol Chem* **260**, 15,298-15,303 (1985).

124. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the viron surface. *Science* **228**, 1315-1317 (1985).

125. Samuelson, P., Gunneriusson, E., Nygren, P.A. & Stahl, S. Display of proteins on bacteria. *J Biotechnol* **96**, 129-154 (2002).

126. He, M. & Taussig, M.J. Ribosome display: Cell-free protein display technology. *Brief Funct Genomic Proteomic* **1**, 204-212 (2002).

127. Parsell, D. & Sauer, R. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. *J Biol Chem* **264**, 7590-7595 (1989).

128. Sieber, V., Pluckthun, A. & Schmid, F.X. Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* **16**, 955-960 (1998).

129. Miyazaki, K., Wintrode, P.L., Grayling, R.A., Rubingh, D.N. & Arnold, F.H. Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol* **297**, 1015-1026 (2000).

130. Street, A.G. & Mayo, S.L. Computational protein design. *Structure Fold Des* **7**, R105-9 (1999).

131. Kraemer-Pecore, C.M., Wollacott, A.M. & Desjarlais, J.R. Computational protein design. *Curr Opin Chem Biol* **5**, 690-5 (2001).

132. Venclovas, C., Zemla, A., Fidelis, K. & Moult, J. Assessment of progress over the CASP experiments. *Proteins* **53 Suppl 6**, 585-95 (2003).

133. Park, S., Yang, X. & Saven, J.G. Advances in computational protein design. *Curr Opin Struct Biol* **14**, 487-94 (2004).

134. Kuhlman, B. & Baker, D. Exploring folding free energy landscapes using computational protein design. *Curr Opin Struct Biol* **14**, 89-95 (2004).

135. Gordon, D.B., Marshall, S.A. & Mayo, S.L. Energy functions for protein design. *Curr Opin Struct Biol* **9**, 509-13 (1999).

136. Mendes, J., Guerois, R. & Serrano, L. Energy estimation in protein design. *Curr Opin Struct Biol* **12**, 441-6 (2002).

137. Dahiyat, B.I. In silico design for protein stabilization. *Curr Opin Biotechnol* **10**, 387-90 (1999).

138. Luo, P. et al. Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening. *Protein Sci* **11**, 1218-26 (2002).

139. Malakauskas, S.M. & Mayo, S.L. Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* **5**, 470-5 (1998).

140. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* **332**, 449-60 (2003).

141. Filikov, A.V. et al. Computational stabilization of human growth hormone. *Protein Sci* **11**, 1452-61 (2002).

142. Benson, D.E., Haddy, A.E. & Hellinga, H.W. Converting a maltose receptor into a nascent binuclear copper oxygenase by computational design. *Biochemistry* **41**, 3262-9 (2002).

143. Reina, J. et al. Computer-aided design of a PDZ domain to recognize new target sequences. *Nat Struct Biol* **9**, 621-7 (2002).

144. Shifman, J.M. & Mayo, S.L. Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* **323**, 417-23 (2002).

145. Berg, D.T. et al. Engineering the proteolytic specificity of activated protein C improves its pharmacological properties. *Proc Natl Acad Sci U S A* **100**, 4423-8 (2003).

146. Looger, L.L., Dwyer, M.A., Smith, J.J. & Hellinga, H.W. Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185-90 (2003).

147. Dwyer, M.A. & Hellinga, H.W. Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr Opin Struct Biol* **14**, 495-504 (2004).

148. Kuhlman, B. et al. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-8 (2003).

149. Dwyer, M.A., Looger, L.L. & Hellinga, H.W. Computational design of a biologically active enzyme. *Science* **304**, 1967-71 (2004).

150. Lassila, J.K., Keeffe, J.R., Oelschlaeger, P. & Mayo, S.L. Computationally designed variants of Escherichia coli chorismate mutase show altered catalytic activity. *Protein Eng Des Sel* **18**, 161-3 (2005).

151. Oelschlaeger, P. & Mayo, S.L. Hydroxyl groups in the (beta)beta sandwich of metallo-beta-lactamases favor enzyme activity: a computational protein design study. *J Mol Biol* **350**, 395-401 (2005).

152. Korkegian, A., Black, M.E., Baker, D. & Stoddard, B.L. Computational thermostabilization of an enzyme. *Science* **308**, 857-60 (2005).

153. Dunbrack, R.L., Jr. & Cohen, F.E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-81 (1997).

154. Dehouck, Y., Gilis, D. & Rooman, M. A new generation of statistical potentials for proteins. *Biophys J* **90**, 4010-7 (2006).

155. Vizcarra, C.L. & Mayo, S.L. Electrostatics in computational protein design. *Curr Opin Chem Biol* **9**, 622-6 (2005).

156. Thomas, P.D. & Dill, K.A. Statistical potentials extracted from protein structures: how accurate are they? *J Mol Biol* **257**, 457-69 (1996).

157. Kortemme, T., Morozov, A.V. & Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* **326**, 1239-59 (2003).

158. Morozov, A.V. & Kortemme, T. Potential functions for hydrogen bonds in protein structure prediction and design. *Adv Protein Chem* **72**, 1-38 (2005).

159. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-8 (2000).

160. Metropolis, N., Rosenbluth, A. Rosenbluth, M. Teller, A. and Teller, E. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087-1092 (1953).

161. Voigt, C.A., Gordon, D.B. & Mayo, S.L. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J Mol Biol* **299**, 789-803 (2000).

162. Desmet, J., De Maeyer, M., Hazes, B. & Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539-542 (1992).

163. Goldstein, R.F. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys J* **66**, 1335-40 (1994).

164. Looger, L.L. & Hellinga, H.W. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J Mol Biol* **307**, 429-45 (2001).

165. Desjarlais, J.R. & Clarke, N.D. Computer search algorithms in protein modification and design. *Curr Opin Struct Biol* **8**, 471-5 (1998).

166. Neria, E., Fischer, S. & Karplus, M. Simulation of activation free energies in molecular systems. *J Chem Phys* **105**, 1902-1921 (1996).

167. Lazaridis, T. & Karplus, M. Effective energy function for proteins in solution. *Proteins* **35**, 133-52 (1999).

168. Simons, K.T. et al. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* **34**, 82-95 (1999).

169. Hayes, R.J. et al. Combining computational and experimental screening for rapid optimization of protein properties. *Proc Natl Acad Sci U S A* **99**, 15926-31 (2002).

170. Moore, G.L. & Maranas, C.D. Modeling DNA mutation and recombination for directed evolution experiments. *J Theor Biol* **205**, 483-503 (2000).

171. Kurtz, J.E. & Black, M.E. Enhancement of suicide gene prodrug activation by random mutagenesis. *Methods Mol Med* **90**, 331-44 (2004).

172. Schramm, V.L. Enzymatic transition states and transition state analog design. *Annu Rev Biochem* **67**, 693-720 (1998).

173. Kraut, D.A., Carroll, K.S. & Herschlag, D. Challenges in enzyme mechanism and energetics. *Annu Rev Biochem* **72**, 517-71 (2003).

174. Daniel, R.M., Dunn, R.V., Finney, J.L. & Smith, J.C. The role of dynamics in enzyme activity. *Annu Rev Biophys Biomol Struct* **32**, 69-92 (2003).

175. Eijsink, V.G. et al. Rational engineering of enzyme stability. *J Biotechnol* **113**, 105-20 (2004).

176. Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. & Engel, P.C. Protein thermostability in extremophiles. *Biochimie* **80**, 933-41 (1998).

177. Eidsness, M.K., Richie, K.A., Burden, A.E., Kurtz, D.M., Jr. & Scott, R.A. Dissecting contributions to the thermostability of Pyrococcus furiosus rubredoxin: beta-sheet chimeras. *Biochemistry* **36**, 10406-13 (1997).

178. Rees, D.C. & Adams, M.W. Hyperthermophiles: taking the heat and loving it. *Structure* **3**, 251-4 (1995).

179. Sterner, R. & Liebl, W. Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol* **36**, 39-106 (2001).

180. Dahiyat, B.I. & Mayo, S.L. De novo protein design: fully automated sequence selection. *Science* **278**, 82-7 (1997).

181. Ireton, G.C., Black, M.E. & Stoddard, B.L. The 1.14 A crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure (Camb)* **11**, 961-72 (2003).

182. Katsuragi, T., Sonoda, T., Matsumoto, K., Sakai, T., and Tonomura, K. Purification and some properties of cytosine deaminase from baker's yeast. *Agric. Biol. Chem.* **53**, 1313-1319 (1989).

183. Greco, O. & Dachs, G.U. Gene directed enzyme/prodrug therapy of cancer: historical appraisal and future prospectives. *J Cell Physiol* **187**, 22-36 (2001).

184. Hubbard, S. Naccess 2.1: Atomic Solvent Accessible Area Calculations. http://wolf.bms.umist.ac.uk/naccess/ (University College, London, 1996).

185. Ireton, G.C. & Stoddard, B.L. Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr* **60**, 601-5 (2004).

186. Otwinowski, Z., and Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).

187. Kissinger, C.R., Gehlhaar, D.K. & Fogel, D.B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr D Biol Crystallogr* **55 (Pt 2)**, 484-91 (1999).

188. Brunger, A.T. Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr* **49**, 24-36 (1993).

189. Brunger, A.T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-21 (1998).

190. McRee, D.E. A visual protein crystallographic software system for X11/XView. *J. Molecular Graphics* **10**, 44-46 (1992).

191. DeLano, W.L. The PyMol Molecular Graphics System. (DeLano Scientific, San Carlos, Ca, USA, 2002).

192. Nishiyama, T. et al. Antineoplastic effects in rats of 5-fluorocytosine in combination with cytosine deaminase capsules. *Cancer Res* **45**, 1753-61 (1985).

193. Ireton, G.C., McDermott, G., Black, M.E. & Stoddard, B.L. The structure of Escherichia coli cytosine deaminase. *J Mol Biol* **315**, 687-97 (2002).

194. Kanamaru, R., Kakuta, H., Sato, T., Ishioka, C. & Wakui, A. The inhibitory effects of 5-fluorouracil on the metabolism of preribosomal and ribosomal RNA in L-1210 cells in vitro. *Cancer Chemother Pharmacol* **17**, 43-6 (1986).

195. Pinedo, H.M. & Peters, G.F. Fluorouracil: biochemistry and pharmacology. *J Clin Oncol* **6**, 1653-64 (1988).

196. Thomas, D.M. & Zalcberg, J.R. 5-fluorouracil: a pharmacological paradigm in the use of cytotoxics. *Clin Exp Pharmacol Physiol* **25**, 887-95 (1998).

197. Diasio, R.B., Lakings, D.E. & Bennett, J.E. Evidence for conversion of 5-fluorocytosine to 5-fluorouracil in humans: possible factor in 5-fluorocytosine clinical toxicity. *Antimicrob Agents Chemother* **14**, 903-8 (1978).

198. Lawrence, T.S. et al. Preferential cytotoxicity of cells transduced with cytosine deaminase compared to bystander cells after treatment with 5-flucytosine. *Cancer Res* **58**, 2588-93 (1998).

199. Hoganson, D.K., Batra, R.K., Olsen, J.C. & Boucher, R.C. Comparison of the effects of three different toxin genes and their levels of expression on cell growth and bystander effect in lung adenocarcinoma. *Cancer Res* **56**, 1315-23 (1996).

200. Nishihara, E. et al. Treatment of thyroid carcinoma cells with four different suicide gene/prodrug combinations in vitro. *Anticancer Res* **18**, 1521-5 (1998).

201. Trinh, Q.T., Austin, E.A., Murray, D.M., Knick, V.C. & Huber, B.E. Enzyme/prodrug gene therapy: comparison of cytosine deaminase/5-fluorocytosine versus thymidine kinase/ganciclovir enzyme/prodrug systems in a human colorectal carcinoma cell line. *Cancer Res* **55**, 4808-12 (1995).

202. Lawrence, T.S., Davis, M.A. & Maybaum, J. Dependence of 5-fluorouracil-mediated radiosensitization on DNA-directed effects. *Int J Radiat Oncol Biol Phys* **29**, 519-23 (1994).

203. Hamstra, D.A. et al. Combined radiation and enzyme/prodrug treatment for head and neck cancer in an orthotopic animal model. *Radiat Res* **152**, 499-507 (1999).

204. Kaliberov, S.A. et al. Mutation of Escherichia coli cytosine deaminase significantly enhances molecular chemotherapy of human glioma. *Gene Ther* (2007).

205. Stackhouse, M.A. et al. Fractionated radiation therapy in combination with adenoviral delivery of the cytosine deaminase gene and 5-fluorocytosine enhances cytotoxic and antitumor effects in human colorectal and cholangiocarcinoma models. *Gene Ther* **7**, 1019-26 (2000).

206. Yao, L., Li, Y., Wu, Y., Liu, A. & Yan, H. Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase. *Biochemistry* **44**, 5940-7 (2005).

207. Mahan, S.D., Ireton, G.C., Stoddard, B.L. & Black, M.E. Alanine-scanning mutagenesis reveals a cytosine deaminase mutant with altered substrate preference. *Biochemistry* **43**, 8957-64 (2004).

208. Davies, G.J., Gamblin, S.J., Littlechild, J.A. & Watson, H.C. The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. *Proteins* **15**, 283-9 (1993).

209. Yip, K.S. et al. The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147-58 (1995).

210. Rice, D.W. et al. Insights into the molecular basis of thermal stability from the structure determination of Pyrococcus furiosus glutamate dehydrogenase. *FEMS Microbiol Rev* **18**, 105-17 (1996).

211. Harris, G.W. et al. Structural basis of the properties of an industrially relevant thermophilic xylanase. *Proteins* **29**, 77-86 (1997).

212. Wallon, G. et al. Crystal structures of Escherichia coli and Salmonella typhimurium 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from Thermus thermophilus. *J Mol Biol* **266**, 1016-31 (1997).

213. Russell, R.J., Ferguson, J.M., Hough, D.W., Danson, M.J. & Taylor, G.L. The crystal structure of citrate synthase from the hyperthermophilic archaeon pyrococcus furiosus at 1.9 A resolution. *Biochemistry* **36**, 9983-94 (1997).

214. Baker, D., Sohl, J.L. & Agard, D.A. A protein-folding reaction under kinetic control. *Nature* **356**, 263-5 (1992).

215. Kelch, B.A. & Agard, D.A. Mesophile versus thermophile: insights into the structural mechanisms of kinetic stability. *J Mol Biol* **370**, 784-95 (2007).

216. Norris, J.S. et al. The present and future for gene and viral therapy of directly accessible prostate and squamous cell cancers of the head and neck. *Future Oncol* **1**, 115-23 (2005).

217. Freytag, S.O. et al. Phase I trial of replication-competent adenovirus-mediated suicide gene therapy combined with IMRT for prostate cancer. *Mol Ther* **15**, 1016-23 (2007).

218. Wiewrodt, R. et al. Adenovirus-mediated gene transfer of enhanced Herpes simplex virus thymidine kinase mutants improves prodrug-mediated tumor cell killing. *Cancer Gene Ther* **10**, 353-64 (2003).

219. Rogulski, K.R., Kim, J.H., Kim, S.H. & Freytag, S.O. Glioma cells transduced with an Escherichia coli CD/HSV-1 TK fusion gene exhibit enhanced metabolic suicide and radiosensitivity. *Hum Gene Ther* **8**, 73-85 (1997).

220. Willmon, C.L., Krabbenhoft, E. & Black, M.E. A guanylate kinase/HSV-1 thymidine kinase fusion protein enhances prodrug-mediated cell killing. *Gene Ther* **13**, 1309-12 (2006).

221. Freytag, S.O. (February 8, 2006).

VITA

Aaron Korkegian was born and raised in San Diego, CA. He earned his Bachelor of Arts in Molecular Biology at Pomona College in 2001. In 2007 he earned a Doctor of Philosophy at the University of Washington in Molecular Biology.