Scientific Research Publishing

Journal of Cancer Therapy, 2025, 16(7), 243-257
https://www.scirp.org/journal/jct
ISSN Online: 2151-1942
ISSN Print: 2151-1934

Development of a Protein Therapeutic That Targets the TMPRSS2:ERG4 Break Region in Prostate Cancer Cells: A Modular Design Approach

Alan Chant[1], Joshua L. Weiss[2], Christina M. Kraemer-Chant[2*]

[1]Department of Biochemistry, College of Agriculture and Life Sciences, University of Vermont, Burlington, VT, USA
[2]Department of Chemistry, Saint Michael's College, Colchester, VT, USA
Email: *alan.chant@med.uvm.edu

How to cite this paper: Chant, A., Weiss, J.L. and Kraemer-Chant, C.M. (2025) Development of a Protein Therapeutic That Targets the TMPRSS2:ERG4 Break Region in Prostate Cancer Cells: A Modular Design Approach. *Journal of Cancer Therapy*, **16**, 243-257.
https://doi.org/10.4236/jct.2025.167019

Received: June 14, 2025
Accepted: July 26, 2025
Published: July 29, 2025

Abstract

Prostate cancer is reported to be the second most common cancer in men. At this time, there exist over 20 FDA-approved drugs for the treatment of prostate cancer. These include alkylating agents, anchracenedione, anti-androgens, antimicrotubule agents, autologous cellular immunotherapy, conjugated estrogens, GnRH (Gonadotropin-Releasing Hormone) analogues and GnRH antagonists (Source: FDA.gov http://www.empr.com/fda-approved-prostate-cancer-drug-treatments/article/123619/#). With recent developments in the detection of prostate cancer, and with the growing population of men who are reaching the age where prostate cancer becomes a significant health issue, the necessity for a drug that targets prostate cancer cells with high specificity and selectivity has increased. The basic premise of our therapeutic, which is designed to bind tightly to a DNA sequence found only in prostate cancer cells, consists of a molecular motor, a small number of zinc finger proteins, and a C-terminal lysine tail. The irreversible binding of our therapeutic with the target DNA is predicted to regulate transcription of the gene and ultimately target the cell for death, thus reversing the proliferation of prostate cancer cells containing that sequence of DNA.

Keywords

Zinc Finger (ZF) Proteins, TMPRSS2:ERG4 Gene Fusion, ZiFit Binding Prediction Tool, Uracil DNA Glycosylase (UDG2), DNA Binding Specificity, Chromosomal Translocation/Breakpoint Region, Protein-DNA Interactions, Cancer Gene Fusions/Oncogenes

1. Introduction

Zinc finger (ZF) proteins account for 3% of the genes found in the human genome and have been shown to have roles in DNA replication, RNA packaging, transcriptional activation, regulation of apoptosis, protein folding and assembly, and lipid binding [1]. ZF motifs, or domains, are independently folded domains found within ZF proteins and contain a two-stranded antiparallel β-sheet as a β hairpin, an α helix and a zinc (II) ion. The zinc ion is tetrahedrally coordinated between the two cysteines (at the turn in the β hairpin) and two histidines (in the C-terminal part of the α helix) to form an inner hydrophobic core (X7-Cys-X4-Cys-X12-His-X3-His-Xn).

The ZF protein binds to its target DNA sequence by slotting the α helix into the major groove of the DNA. Each ZF domain recognizes, with a varying degree of specificity, a three-base segment on one strand of the DNA. Partners of many different ZF proteins have been determined [2]. Our therapeutic was designed to make use of several ZF domains that, when covalently linked, would be specific to the prostate cancer gene TMPRSS2:ERG4 [3] [4]. The ZF domains can be selected through the use of ZiFit, a program that determines ZF binding partners to specific sequences of DNA.

Multiple ZF domains have been shown to produce a greater degree of discrimination and complexity of binding to a specific DNA sequence [5] [6]. Therefore, the use of several ZF domains in tandem can promote tight, nearly (or completely) irreversible binding to specific DNA sequences. ZiFit does give several selections for each three-base sequence; our aim was to perform parallel binding studies to determine which combination of ZF domains are the most effective in binding to the TMPRSS2:ERG4 gene, as covalently linking two independent ZF domains can result in change in ability to predict binding to the DNA target [7]. We began by focusing on the use of five ZF proteins overlapping the breakpoint region, covalently bound together by short flexible linkers.

Human Uracil DNA Glycosylase, isoform 2 (UDG2) is a protein enzyme involved in the removal of uracil from DNA as part of the cell's DNA repair mechanism [8]. Two forms of UDG are generated from the UNG gene—the mitochondrial UDG, UDG1, which consists of 304 amino acids, and the nuclear UDG, UDG2, which consists of 315 amino acids. Roughly 90% sequence homology exists between the two isoforms; the difference at the N-terminus is for cellular localization, not for catalytic activity [9] [10]. We used UDG2 in this study, as we wished to use its nuclear localization sequence (NLS) to target the chromosomal DNA. The mechanism of UDG2 interaction with DNA appears to be dominated by DNA hopping between sequences of approximately 10 bp; fast 1-dimensional scanning along this length of sequence takes place before the enzyme hops to another segment of DNA in search of the specific uracil lesion to which it has high affinity. This nonspecific binding of UDG exhibits a very short residence time of approximately 5 ms [11]. The processivity of the *E. coli* form of UDG is affected by NaCl concentrations [12]. The enzyme must successfully bypass all of the potential blocks that it encounters, includ-

ing transcription factors and nucleosome remodeling.

GATA proteins are a family of transcription factors that bind, via one or more ZF motifs, to a canonical GATA DNA sequence found at promoter regions of key regulatory proteins [13] [14]. In addition to the ZF motifs, which are primarily responsible for the GATA binding to GATA DNA sequences, wild-type GATA proteins have a lysine-rich C-terminal tail that has been shown to be necessary for maintaining high DNA binding affinities [15] [16].

It has been reported that there are 358 gene fusions known to cause tumorigenesis, which are accountable for 20% of the human cancer morbidity [17]-[20]. Translocation of chromosomal DNA causes rearrangements of genetic material that can lead to the formation of oncogenes. Most prostate cancer cells have DNA that has been cleaved and rejoined in such a way as to create a breakpoint region. Cancer caused by this manipulation of DNA stems from the fact that the growth regulation factors of the cell are damaged or removed; this leads to unregulated cell growth and lifetimes [21] [22]. The effects of this alteration of the cellular DNA could potentially interfere with a number of biological systems, including (but not limited to) overexpression of specific proteins, alteration of apoptotic pathways (preventing programmed cell death), or cell growth (stimulating unregulated cellular expansion). The breakpoint region that is produced by the co-joining of two different genes is typically a unique sequence not found in normal cells. This provides an opportunity for drug targeting to this region of genomic DNA that is specific to prostate cancer cells only, leaving normal cells unaffected.

2. Materials and Methods

2.1. Expression Vector

```
atgatcggccagaagaccctgtacagcttcttcagccccagccccgccagaaagagacac
gcccccagccccgagcccgccgtgcagggcaccggcgtggccggcgtgcccgaggagagc
ggcgacgccgccgccatccccgccaagaaggcccccgccggccaggaggagcccggcacc
ccccccagcagccccctgagcgccgagcagctggacagaatccagagaaacaaggccgcc
gccctgctgagactggccgccagaaacgtgcccgtgggcttcggcgagagctggaagaag
cacctgagcggcgagttcggcaagccctacttcatcaagctgatgggcttcgtggccgag
gagagaaagcactacaccgtgtacccccccccccaccaggtgttcacctggacccagatg
tgcgacatcaaggacgtgaaggtggtgatcctgggccaggacccctaccacggccccaac
caggcccacggcctgtgcttcagcgtgcagagacccgtgccccccccccccaggcctggag
aacatctacaaggagctgagcaccgacatcgaggacttcgtgcaccccggccacggcgac
ctgagcggctgggccaagcagggcgtgctgctgctgaacgccgtgctgaccgtgagagcc
caccaggccaacagccacaaggagagaggctgggagcagttcaccgacgccgtggtgagc
tggctgaaccagaacagcaacggcctggtgttcctgctgtggggcagctacgcccagaag
aagggcagcgccatcgacagaaagagacaccacgtgctgcagaccgcccaccccagcccc
ctgagcgtgtacagaggcttcttcggctgcagacacttcagcaagaccaacgagctgctg
cagaagagcggcaagaagcccatcgactggaaggagctgcccggcgagaagccctacaag
tgccccgagtgcggcaagagcttcagcgagagaagccacctgagagagcaccagagaacc
cacagcagacccggcgagagacccttcatgtgcacctggagctactgcggcaagagattc
accagaagcgacgacctgcagagacacaagagaacccacgacgtgatcaagaagagaaac
agacccggcgagaagccctacaagtgccccgagtgcggcaagagcttcagcagaagcgac
cacctgaccaaccaccagagaacccacagcagacccggcgagaagccctacaagtgcccc
gagtgcggcaagagcttcagcaccaccggcaacctgaccgtgcaccagagaacccacgac
gtgatcaagaagagaaacagacccggcgagaagccctacaagtgccccgagtgcggcaag
agcttcagcaccaccggcgccctgaccgagcaccagagaacccacgacgtgatcaagaag
agaaacaga
```

Figure 1. Complete DNA sequence for the Prostate Cancer Therapeutic. This gene includes sequences for Uracil DNA Glycosylase and five Zinc-Finger Protein (ZFP) domains, as well as the two types of linkers that connect them.

The gene sequence for our proposed prostate cancer therapeutic (PCT) was synthesized and cloned into the pET-24a(+) vector system using the BamHI and NdeI cloning sites by Twist Bioscience. This system is an overexpressing construct containing the kanamycin resistance gene. Codon optimization was carried out for *E. coli* codon usage using the Twist Bioscience sequence optimization software. Verification of the gene insert was carried out using Sanger sequencing (Twist Bioscience). Sequence data showed that the gene had the correct restriction sites and the gene sequence was correct. The full DNA sequence for the PCT can be seen in Figure 1; Figure 2 shows the protein sequence of the entire designed therapeutic (54.54 kDa calculated molecular weight; bioinformatics.org).

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEES

GDAAAIPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAA

ALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVA

EERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQDPYHG

PNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHG

DLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDA

VVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQT

AHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKELPG

EKPYKCPECGKSFSERSHLREHQRTHSRPGERPFMCTWSY

CGKRFTRSDDLQRHKRTHDVIKKRNRPGEKPYKCPECGK

SFSRSDHLTNHQRTHSRPGEKPYKCPECGKSFSTTGNLTV

HQRTHDVIKKRNRPGEKPYKCPECGKSFSTTGALTEHQR

THDVIKKRNR

Figure 2. Protein sequence for the Prostate Cancer Therapeutic (PCT). Amino acids are color-coded as follows: Uracil DNA Glycosylase (UDG)—black letters; Zinc-Finger (ZF) domains—green letters; SR linkers—red letters; Bridge-linkers/basic tail sequences—orange letters.

2.2. Transformation of PCT Gene Construct

BL21 (DE3) cells (Invitrogen) were transformed by adding 1.0 ng of the pET 24a(+) vector containing the gene sequence for PCT. Cells were then heat-shocked as follows: Cells were incubated on ice for 1 hour, placed in a 40˚C hot water bath for 30 s, then immediately placed on ice for 5 min. 500 µL of 37˚C SOC media (National Diagnostics) was added and incubated at 37˚C for 1 h. The cells were spun down and spread out on a LB-agar plate containing the kanamycin antibiotic. The plate was incubated overnight at 37˚C.

2.3. Isopropyl *β*-D-1-Thiogalactopyranosid (IPTG) Induction

A single colony was picked and inoculated in 5 mL of LB media containing 50 μg/mL kanamycin in a 15-mL conical tube. The resulting solution was incubated overnight at 37˚C in a shaking incubator. Cells were diluted 1:100 in 200 mL LB media containing 50 μg/mL kanamycin and grown at 37˚C in a shaking incubator until the optical density at 600 nm reached 0.6 OD units. At this point, 1 mM IPTG was added and the cells were left to shake for 3 h before being removed. The cell suspension was centrifuged at 3000×*g* for 30 min at 4˚C. The cell pellet was then suspended in liquid nitrogen for 1 min, then placed in a −80˚C freezer for storage.

2.4. Protein Expression and Purification

PCT proteins were isolated from crude cellular extracts by resuspending 10 g of wet cell pellets in 40 mL metal chelate (MC) binding buffer solution (5 mM imidazole, 250 mM NaCl, and 20 mM Tris-HCl (pH 7.9). A lysis buffer was made by adding 0.1% NP-40 (to promote solubility), 0.5 mM PMSF, and 1.0 mM benzamidine (to inhibit any endogenous proteases) to the 40 mL MC buffer containing the resuspended cells. Cells were lysed by subjecting them to a single round of a freeze/thaw cycle (−20˚C for 24 h), followed by 10 repetitions of a 10-s sonication with a 1-min cooling cycle on ice. Sonicated cell extracts were then centrifuged at 3000×*g* for 40 min at 4˚C. The supernatant containing the PCT proteins was collected and applied to a 7-mL Nickel-NTA HisPur™ Econo column (Bio Rad/Thermo Scientific). Prior to loading the cell extract, the affinity column was equilibrated in the MC metal chelate-binding buffer. The supernatant was loaded onto the column at a flow rate of 1 mL/min. The column was washed with 4 column volumes of MC buffer before starting the elution step. The bound protein was eluted over 3 column volumes of metal chelate elution buffer (1 M imidazole, 500 mM NaCl, and 20 mM Tris-HCl (pH 7.9)). Fractions were analyzed on 12% SDS-PAGE gels. Each protein eluted as a single peak within 3 - 5 fractions at an imidazole concentration of approximately 350 mM.

2.5. Protein Identification by Liquid Chromatography-Tandem Mass Spectroscopy (LC-MS/MS)

LC-MS/MS was performed using the same conditions as previously described [23]. Analysis was performed by the Proteomics Department at the University of Vermont (Burlington, VT).

3. Results

Our results show a proof of concept for the modular design of a protein cancer therapeutic.

3.1. Drug Design and Process

3.1.1. Building a Specific Zinc Finger Array for Binding to the TMPRSS2:ERG4 Gene

The Zinc Finger Consortium (http://www.zincfingers.org/default2.htm) has stud-

ied binding of individual ZF modules to their target DNA sequences [2] [24]. The ZiFiT (Zinc Finger Targeter) program [25] [26], which has been developed by the Consortium, facilitates the design of custom zinc finger proteins (ZFPs) by identifying suitable DNA target sites and matching them with validated zinc finger modules using a modular assembly approach. We utilized ZiFiT's "Modular Assembly" method, which aims to construct ZFPs by linking three individual pre-characterized ZF modules, each recognizing three consecutive and specific 3-bp DNA subsites, together. The ZiFiT program accepts as input uploaded DNA sequences, which it then divides into 9-bp regions composed of three consecutive 3-bp triplets. It then compares each triplet of triplets against a library of experimentally-validated ZF modules. Because three ZF proteins are required to bind to 9-bp regions, the selectivity of the designed ZFP system is increased and the chance of off-target binding is decreased. Only compatible triplets (especially GNNs) are retained for reliable design. Each potential 9-bp target site is evaluated based on **GNN Score** (preference is given to sites with more GNN triplets, as they yield higher functional success) and **Affinity Score** (a predictive metric indicating DNA-binding strength, where lower values suggest stronger binding and better functionality). Individual ZF–to–triplet affinities (e.g., a numeric K_d for a single ZF, such as ZF83 binding to AGC) have not been published or centrally curated, and these data are not reported in ZiFDB or Consortium notes. However, it has been reported that assembled multi-finger proteins typically bind with high affinity in the low nanomolar range ($K_d \approx 1 - 20$ nM) and with high specificity [27].

For our study, we uploaded the **TMPRSS2/ERG** DNA sequence of interest into ZiFiT and selected target sites with high GNN content and favorable Affinity Scores. We selected five promising ZF proteins that should, when expressed as a covalently-linked system, bind with high specificity to DNA triplets found around the breakpoint region of the prostate cancer gene TMPRSS2:ERG4 [3] [4], specifically AGC (for the first zinc finger) and GCG (for the second) in the TMPRSS2 exon 1 portion of the breakpoint region, and AGG (for the third zinc finger), AAT (for the fourth) and CTT (for the fifth) on the ERG4 portion of the breakpoint region. Based on these results, we selected ZF83, ZF23, ZF84, ZF77 and ZF103 as modules for the ZF array using the ZF numbering introduced by the Zinc Finger Consortium (see **Figure 3**). **Figure 4** shows a schematic for the overall concept for the design of the PCT protein.

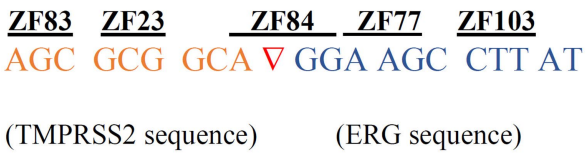

Figure 3. TMPRSS2/ERG DNA sequence. The TMPRSS2 sequence is shown in orange, and the ERG sequence is shown in blue. The red triangle indicates the break-region. Zinc Finger Proteins are shown at their respective triple base-pair binding sites (ZF numbering is from the Zinc-finger Consortium) and include ZF83 (AGC); ZF23 (GCG); ZF84 (AGG); ZF77 (AAT); and ZF103 (CTT).

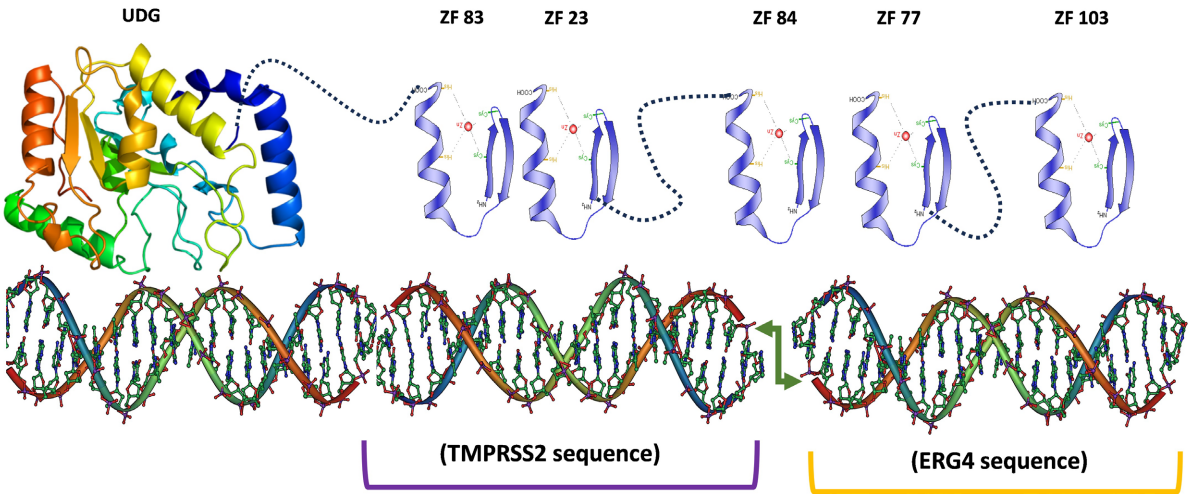

Figure 4. Schematic showing the PCT protein overlapping the TMPRSS2/ERG4 break region. The image shows the ribbon structure of UDG linked to five Zinc-Finger Proteins (ZFP). Each ZFP is aligned along the DNA in their approximate positions with respect to the break region.

3.1.2. Molecular Motor Design

UDG2 has an NLS, which should be useful *in vivo* to locate the system to the chromosomal DNA in an organized fashion as opposed to the ZF array randomly encountering its target sequence through diffusion. UDG2 can also act as a molecular motor, pulling the system along the DNA in search of its target sequence. One of UDG's functions is to unpackage, unwind and clear chromosomal DNA, which may assist in locating the system to the nucleus and increasing the accessibility of the DNA target. In addition, the presence of the UDG may serve to overcome nonspecific interactions of the zinc finger system to random DNA sequences, as the force of UDG hopping and short-range scanning may overcome any nonspecific, weaker interactions the ZF array may experience.

3.1.3. Linkers/Anchor Design

Wild-type GATA proteins have ZF motifs that are primarily responsible for binding to specific DNA sequences. They also contain a lysine-rich C-terminal tail (or anchor) that has been shown to be necessary for maintaining high DNA binding affinities [15] [16]. We predict that inclusion of a C-terminal tail found in the GATA1 protein (protein sequence DVIKKRNR) will improve the binding affinity of the UDG-containing fusion system to the TMPRSS2:ERG4 gene sequence. We have also used the DVIKKRNR sequence to bridge between sets of two zinc finger modules; each individual module is joined by short SR linkers (protein sequence SR). The fifth ZF motif is preceded by the longer DVIKKRNR sequence to introduce necessary flexibility for binding to the DNA target.

3.2. Protein Expression and Purification

3.2.1. Protein Expression

Protein expression was carried out according to the conditions outlined in Section 2.4. Colony growth on antibiotic-inclusive plates were excellent, with large, well-

distributed colonies present after 12 - 15 h incubation. Figure 5 shows the expression profile for the induction of the protein therapeutic PCT. When comparing the pre- and post-induction lanes, we see that they share a number of similar-sized protein bands. However, the post-induction has an additional heavy band at approximately 40 kD that is not present in the pre-induction lane. These results are not consistent with the expected molecular weight of 54 kD for the PCT protein.

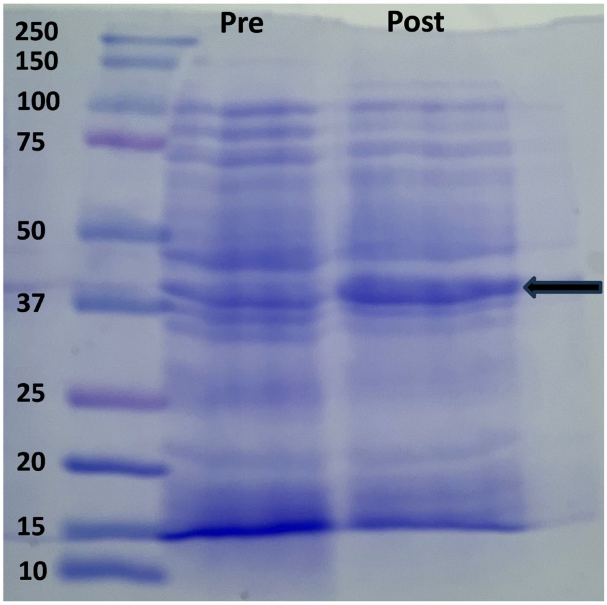

Figure 5. Gel of expression for the PCT/pET24a(+) construct. **PL** = protein ladder (MW values are labeled to the left of the gel); **Pre** is pre-induction; **Post** is post-IPTG induction after 4 hrs. The black arrow points to the heavy band at approximately 40 KD in the post-induction lane that indicates the overexpression of the *E. coli* OmpF protein, which is verified by LC-MS/MS as shown in Table 1.

3.2.2. Affinity Chromatograph

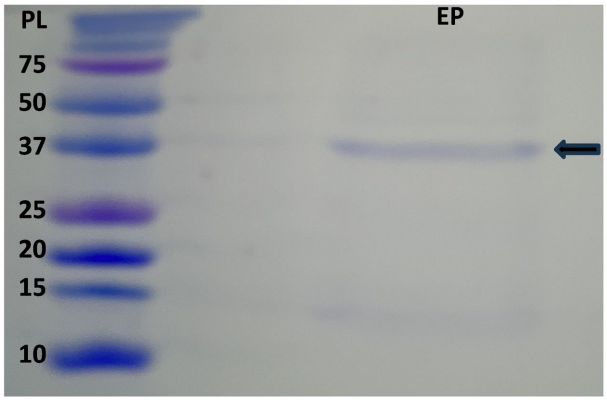

Figure 6. Gel of the eluted protein from Nickel-NTA purification. **PL** indicates protein ladder (MW values are labeled to the left of the gel); **EP** is the eluted protein lane. The black arrow points towards the band at approximately 40 KD in the eluted protein sample lane (EP) that indicates overexpression and elution of the *E. coli* OmpF protein as verified by LC-MS/MS (see Table 1).

Protein extracts were purified using a Nickel-NTA column under the conditions outlined in Section 2.4. The eluted protein sample was visualized on a 12% SDS-PAGE gel (Figure 6). The only band present is at the 40 kD mark, which corresponds to the overexpressed protein in Figure 3. In order to determine if this was a truncated PCT or a different protein entirely, the gel piece containing the band was removed and analyzed by LC-MS/MS.

3.2.3. Protein Identification by Liquid Chromatography-Tandem Mass Spectroscopy (LC-MS/MS)

From the LC-MS/MS data, it can be seen that the predominant protein found is an *E. coli* protein, Outer Membrane Porin F (Table 1). This protein was identified with high confidence and a high coverage (>10 peptides with a high peptide identification score). From the SDS-PAGE analysis, these results have been consistent throughout the overexpression optimization process for PCT. No PCT fragments were found in the LC-MS/MS results, indicating that protein expression of the target therapeutic was not successful.

Table 1. Sample Identification by Liquid Chromatography-Tandem Mass Spectroscopy (LC-MS/MS). The table shows the identity and information for the 25 most significant proteins detected from an excised protein sample as shown in Figure 6. The proteins are listed from the most to least abundant proteins in the sample analyzed.

Accession	Description	Coverage [%]	#Unique Peptides	#AAs	MW [kDa]	calc. pI	Score Sequest HT: Sequest HT
Q8XDF1	Outer membrane porin OmpF OS = Escherichia coli O157:H7 OX = 83334 GN = ompF PE = 3 SV = 1	31	11	362	39.3	4.96	70.55
P04264	Keratin, type II cytoskeletal 1 OS = Homo sapiens OX = 9606 GN = KRT1 PE = 1 SV = 6	19	11	644	66	8.12	32.95
P13645	Keratin, type I cytoskeletal 10 OS = Homo sapiens OX = 9606 GN = KRT10 PE = 1 SV = 6	19	10	584	58.8	5.21	30.64
P35908	Keratin, type II cytoskeletal 2 epidermal OS = Homo sapiens OX = 9606 GN = KRT2 PE = 1 SV = 2	20	8	639	65.4	8	30.02
P0A911	Outer membrane protein A OS = Escherichia coli O157:H7 OX = 83334 GN = ompA PE = 3 SV = 1	29	9	346	37.2	6.42	29.56
P35527	Keratin, type I cytoskeletal 9 OS = Homo sapiens OX = 9606 GN = KRT9 PE = 1 SV = 3	14	9	623	62	5.24	26.46
Q8XBD3	Outer membrane protein assembly factor BamC OS = Escherichia coli O157:H7 OX = 83334 GN = bamC PE = 3 SV =4	19	6	344	36.8	5.55	19.71
P0A6P3	Elongation factor Ts OS = Escherichia coli O157:H7 OX = 83334 GN = tsf PE = 3 SV = 2	21	6	283	30.4	5.29	13.32
A0A0H3JCB2	Lactose operon repressor OS=Escherichia coli O157:H7 OX = 83334 GN = lacI PE = 4 SV = 2	10	4	363	38.9	7.17	9.86
Q8XAA9	Outer membrane protein assembly factor BamB OS = Escherichia coli O157:H7 OX = 83334 GN = bamB PE = 3 SV = 1	13	4	392	41.8	4.91	9.83
P02533	Keratin, type I cytoskeletal 14 OS=Homo sapiens OX = 9606 GN = KRT14 PE = 1 SV = 4	6	2	472	51.5	5.16	8.8

Continued

P02538	Keratin, type II cytoskeletal 6A OS=Homo sapiens OX=9606 GN=KRT6A PE=1 SV=3	6	1	564	60	8	8.39
P65765	FKBP-type peptidyl-prolyl cis-trans isomerase FkpA OS = Escherichia coli O157:H7 OX = 83334 GN = fkpA PE = 3 SV = 1	8	2	270	28.9	8.47	7.7
P13647	Keratin, type II cytoskeletal 5 OS=Homo sapiens OX = 9606 GN = KRT5 PE = 1 SV = 3	5	1	590	62.3	7.74	6.67
P81605	Dermcidin OS = Homo sapiens OX = 9606 GN = DCD PE = 1 SV = 2	20	2	110	11.3	6.54	4.96
Q8X773	Uncharacterized protein OS=Escherichia coli O157:H7 OX = 83334 GN = ydgH PE = 4 SV = 1	9	2	314	33.9	9.28	4.63
P0A7Z6	DNA-directed RNA polymerase subunit alpha OS = Escherichia coli O157:H7 OX = 83334 GN = rpoA PE = 1 SV = 1	5	2	329	36.5	5.06	3.84
P0AA27	Thioredoxin 1 OS = Escherichia coli O157:H7 OX = 83334 GN = trxA PE = 1 SV = 2	11	1	109	11.8	4.88	3.37
Q86YZ3	Hornerin OS = Homo sapiens OX = 9606 GN = HRNR PE = 1 SV = 2	2	1	2850	282.2	10.04	2.83
E7EQ64	Trypsin-1 OS = Homo sapiens OX = 9606 GN = PRSS1 PE = 1 SV = 1	8	1	261	28.1	7.25	2.61
P69799	PTS system mannose-specific EIIAB component OS = Escherichia coli O157:H7 OX = 83334 GN = manX PE = 3 SV = 2	4	1	323	35	6.02	2.53
P0A9S4	Galactitol 1-phosphate 5-dehydrogenase OS = Escherichia coli O157:H7 OX = 83334 GN = gatD PE = 3 SV = 1	3	1	346	37.4	6.38	2.45
P0A9B4	Glyceraldehyde-3-phosphate dehydrogenase A OS = Escherichia coli O157:H7 OX = 83334 GN = gapA PE = 3 SV = 2	5	1	331	35.5	7.11	2.14
P0AAC2	Universal stress protein E OS=Escherichia coli O157:H7 OX = 83334 GN = uspE PE = 3 SV = 2	2	1	316	35.7	5.31	2.08
P58070	GTPase Era OS = Escherichia coli O157:H7 OX = 83334 GN = era PE = 3 SV = 1	2	1	301	33.7	7.65	2.07

4. Discussion

In this study, we designed a system that theoretically would recognize and bind to a specific breakpoint region of the prostate cancer gene TMPRSS2:ERG4 [20]. Fusion of the TMPRSS2 gene to the ETS transcription factor ERG has been reported in up to two-thirds of prostate cancer cases (for example, see [3] [4] [20] [28]-[33]). It has been shown that translocations of TMPRSS2 to ERG can result in at least 17 distinctly structured fusion transcripts [20] [31] [34], but the most commonly detected transcript involved joining of exon 1 of TMPRSS2 to exons 4 or 5 of the ERG gene [20]. This leads to overexpression of the 3'-exons of the ERG gene. We initially focused on the joining with exon 4 of the ERG gene (TMPRSS2:ERG4), as it has been suggested that this fusion is the most abundant in prostate cancer samples [30].

Our objective was to design a protein therapeutic that would suppress the growth of prostate cancer cells that contain TMPRSS2:ERG4. Since this breakpoint region is found only in prostate cancer cells, no healthy cells will be affected. The therapeutic consisted of a UDG2 component, an array of ZF protein modules, and a lysine tail. The array contained five covalently-linked ZF protein modules that were designed to bind with high specificity and affinity to the TMPRSS2:ERG4 translocation, would block replication and trigger cell death, thus preventing its proliferation. The UDG2 component N-terminal to the array would locate the system to the nucleus in an efficient fashion, where it would unwind the DNA from histones and clear it of bound proteins and small molecules (thus making the DNA accessible to ZF binding) while also acting as a molecular motor to pull the system along the chromosomal DNA in search of the TMPRSS2:ERG4 region. However, the force of motion of UDG2 along the DNA may be strong enough to overcome the high binding affinity of the designed array. Inclusion of a lysine-rich tail such as is found in zinc finger-containing GATA proteins was included to potentially help to balance the force of motion of the UDG2, thereby rescuing some or all of the binding affinity of the zinc finger system.

Other applications have used multiple ZF proteins fused to nonspecific nucleases for down-regulation of genes in the treatment of genetic disorders; in those applications, the zinc fingers locate the gene of interest and a fused nonspecific nuclease introduces a double-strand breakage at a specific site [35].

5. Conclusions

In this article, we present a novel drug design using a modular approach to target the TMPRSS2:ERG4 gene fusion [36] [37]. The data in this paper highlight the barriers encountered when attempting to combine the functions of proteins from different species in order to utilize their unique characteristics; our therapeutic included Uracil DNA glycosylase (*Homo sapiens*); a linker sequence (*Aspergillus nidulans*); and synthesized zinc-finger domains (predominantly derived from plant species by the Zinc Finger Consortium). Although every effort was made to generate an optimized gene sequence with a high degree of confidence, there may still be unforeseen issues in the algorithm we used in terms of sequence complexity that is unacceptable to the *E. coli* cells overexpressing the therapeutic, rare codon usage, or other factors.

Another reason why the expression of the drug was unsuccessful could be a leaky promoter. pET24a(+) vectors are known to express low level amounts of the gene pre-induction. It is conceivable that the low levels of expressed PCT could competitively bind at the promoter region, preventing further expression, even after induction with IPTG. Because this system was designed to bind to DNA, this would suggest that the drug binds indiscriminately—something that is highly unfortunate in the design of a therapeutic meant for interaction with a specific DNA sequence.

It is interesting that the overexpression of the Outer Membrane Porin F protein

(OmpF) native to *E. coli* is reproducibly overexpressed; others have reported that expression of OmpF is downregulated under antibiotic stress [38] [39]. It is even more interesting that it is retained and detected at such high levels after metal-affinity column purification, as OmpF is not rich in histidine residues. However, it has been seen that BL21 (DE3) cell lines do have elevated levels of OmpF over-expression as compared to other *E. coli* strains [40] [41].

To address the lack of target PCT expression, future studies will include reducing the number of zinc-finger modules in varying arrangements to see the effects on protein expression. We are also in the process of examining the use of a eukaryotic vector-based system for protein expression in a mammalian cell line, which might be more tolerant of the protein therapeutic.

Acknowledgements

Research reported in this publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103449. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

The authors wish to acknowledge the previous work of our Saint Michael's College research students that, while not represented in this paper, did work on this project. These students include Sara Williams, Isaiah St. Pierre, Dana Bourne, Christopher Ricciardi, David Weiss, Megan Ackerman, Christopher Toomey, Samantha Delaney, and Tiana Dunne.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Laity, J.H., Lee, B.M. and Wright, P.E. (2001) Zinc Finger Proteins: New Insights into Structural and Functional Diversity. *Current Opinion in Structural Biology*, **11**, 39-46. https://doi.org/10.1016/s0959-440x(00)00167-6

[2] Wright, D.A., Thibodeau-Beganny, S., Sander, J.D., Winfrey, R.J., Hirsh, A.S., Eichtinger, M., *et al.* (2006) Standardized Reagents and Protocols for Engineering Zinc Finger Nucleases by Modular Assembly. *Nature Protocols*, **1**, 1637-1652. https://doi.org/10.1038/nprot.2006.259

[3] Mehra, R., Tomlins, S.A., Shen, R., Nadeem, O., Wang, L., Wei, J.T., *et al.* (2007) Comprehensive Assessment of TMPRSS2 and ETS Family Gene Aberrations in Clinically Localized Prostate Cancer. *Modern Pathology*, **20**, 538-544. https://doi.org/10.1038/modpathol.3800769

[4] St. John, J., Powell, K., Conley-Lacomb, M.K. and Chinni, S.R. (2012) TMPRSS2-ERG Fusion Gene Expression in Prostate Tumor Cells and Its Clinical and Biological Significance in Prostate Cancer Progression. *Journal of Cancer Science & Therapy*, **4**, 94-101. https://doi.org/10.4172/1948-5956.1000119

[5] Persikov, A.V., Osada, R. and Singh, M. (2008) Predicting DNA Recognition by Cys2His2 Zinc Finger Proteins. *Bioinformatics*, **25**, 22-29.

https://doi.org/10.1093/bioinformatics/btn580

[6] Wolfe, S.A., Nekludova, L. and Pabo, C.O. (2000) DNA Recognition by Cys2His2 Zinc Finger Proteins. *Annual Review of Biophysics and Biomolecular Structure*, **29**, 183-212. https://doi.org/10.1146/annurev.biophys.29.1.183

[7] Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G. and Zhang, F. (2012) A Transcription Activator-Like Effector Toolbox for Genome Engineering. *Nature Protocols*, **7**, 171-192. https://doi.org/10.1038/nprot.2011.431

[8] Krokan, H. and Urs Wittwer, C. (1981) Uracil DNA-Glycosylase from Hela Cells: General Properties, Substrate Specificity and Effect of Uracil Analogs. *Nucleic Acids Research*, **9**, 2599-2614. https://doi.org/10.1093/nar/9.11.2599

[9] Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skorpen, F., *et al.* (1997) Nuclear and Mitochondrial Uracil-DNA Glycosylases Are Generated by Alternative Splicing and Transcription from Different Positions in the UNG Gene. *Nucleic Acids Research*, **25**, 750-755. https://doi.org/10.1093/nar/25.4.750

[10] Otterlei, M., Haug, T., Nagelhus, T.A., Slupphaug, G., Lindmo, T. and Krokan, H.E. (1998) Nuclear and Mitochondrial Splice Forms of Human Uracil-DNA Glycosylase Contain a Complex Nuclear Localisation Signal and a Strong Classical Mitochondrial Localisation Signal, Respectively. *Nucleic Acids Research*, **26**, 4611-4617. https://doi.org/10.1093/nar/26.20.4611

[11] Porecha, R.H. and Stivers, J.T. (2008) Uracil DNA Glycosylase Uses DNA Hopping and Short-Range Sliding to Trap Extrahelical Uracils. *Proceedings of the National Academy of Sciences*, **105**, 10791-10796. https://doi.org/10.1073/pnas.0801612105

[12] Bennett, S.E., Sanderson, R.J. and Mosbaugh, D.W. (1995) Processivity of *Escherichia coli* and Rat Liver Mitochondrial Uracil-DNA Glycosylase Is Affected by NaCl Concentration. *Biochemistry*, **34**, 6109-6119. https://doi.org/10.1021/bi00018a014

[13] Ko, L.J. and Engel, J.D. (1993) DNA-Binding Specificities of the GATA Transcription Factor Family. *Molecular and Cellular Biology*, **13**, 4011-4022. https://doi.org/10.1128/mcb.13.7.4011-4022.1993

[14] Yang, H. and Evans, T. (1992) Distinct Roles for the Two Cgata-1 Finger Domains. *Molecular and Cellular Biology*, **12**, 4562-4570. https://doi.org/10.1128/mcb.12.10.4562-4570.1992

[15] Manfield, I.W., Reynolds, L.A., Gittins, J. and Kneale, G.G. (2000) The DNA-Binding Domain of the Gene Regulatory Protein Area Extends beyond the Minimal Zinc-Finger Region Conserved between GATA Proteins. *Biochimica et Biophysica Acta* (*BBA*)—*Gene Structure and Expression*, **1493**, 325-332. https://doi.org/10.1016/s0167-4781(00)00197-4

[16] Starich, M.R., Wikström, M., Arst, H.N., Clore, G.M. and Gronenborn, A.M. (1998) The Solution Structure of a Fungal AREA Protein-DNA Complex: An Alternative Binding Mode for the Basic Carboxyl Tail of GATA Factors. *Journal of Molecular Biology*, **277**, 605-620. https://doi.org/10.1006/jmbi.1998.1625

[17] Mitelman, F., Johansson, B. and Mertens, F. (2007) The Impact of Translocations and Gene Fusions on Cancer Causation. *Nature Reviews Cancer*, **7**, 233-245. https://doi.org/10.1038/nrc2091

[18] Schröder, F.H. (2007) Re: Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer. *European Urology*, **51**, 1443-1444. https://doi.org/10.1016/j.eururo.2007.02.021

[19] Teixeira, M.R. (2006) Recurrent Fusion Oncogenes in Carcinomas. *Critical Reviews™ in Oncogenesis*, **12**, 257-271. https://doi.org/10.1615/critrevoncog.v12.i3-4.40

[20] Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X., *et al.* (2005) Recurrent Fusion of *tmprss2* and ETS Transcription Factor Genes in Prostate Cancer. *Science*, **310**, 644-648. https://doi.org/10.1126/science.1117679

[21] Chen, J., Cooper, D.N., Férec, C., Kehrer-Sawatzki, H. and Patrinos, G.P. (2010) Genomic Rearrangements in Inherited Disease and Cancer. *Seminars in Cancer Biology*, **20**, 222-233. https://doi.org/10.1016/j.semcancer.2010.05.007

[22] Aplan, P.D. (2006) Causes of Oncogenic Chromosomal Translocation. *Trends in Genetics*, **22**, 46-55. https://doi.org/10.1016/j.tig.2005.10.002

[23] Toomey, C.G., Weiss, D., Chant, A., Ackerman, M., Ahlers, B.A., Lam, Y., *et al.* (2017) Development and Applications of a Calmodulin-Based Fusion Protein System for the Expression and Purification of WW and Zinc Finger Modules. *Advances in Biological Chemistry*, **7**, 89-106. https://doi.org/10.4236/abc.2017.72006

[24] Ramirez, C.L., Foley, J.E., Wright, D.A., Müller-Lerch, F., Rahman, S.H., Cornu, T.I., *et al.* (2008) Unexpected Failure Rates for Modular Assembly of Engineered Zinc Fingers. *Nature Methods*, **5**, 374-375. https://doi.org/10.1038/nmeth0508-374

[25] Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K. and Dobbs, D. (2010) ZiFiT (Zinc Finger Targeter): An Updated Zinc Finger Engineering Tool. *Nucleic Acids Research*, **38**, W462-W468. https://doi.org/10.1093/nar/gkq319

[26] Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F. and Dobbs, D. (2007) Zinc Finger Targeter (ZiFiT): An Engineered Zinc Finger/Target Site Design Tool. *Nucleic Acids Research*, **35**, W599-W605. https://doi.org/10.1093/nar/gkm349

[27] Mandell, J.G. and Barbas, C.F. (2006) Zinc Finger Tools: Custom DNA-Binding Domains for Transcription Factors and Nucleases. *Nucleic Acids Research*, **34**, W516-W523.

[28] Perner, S., Demichelis, F., Beroukhim, R., Schmidt, F.H., Mosquera, J., Setlur, S., *et al.* (2006) TMPRSS2:ERG Fusion-Associated Deletions Provide Insight into the Heterogeneity of Prostate Cancer. *Cancer Research*, **66**, 8337-8341. https://doi.org/10.1158/0008-5472.can-06-1482

[29] Jhavar, S., Reid, A., Clark, J., Kote-Jarai, Z., Christmas, T., Thompson, A., *et al.* (2008) Detection of TMPRSS2-ERG Translocations in Human Prostate Cancer by Expression Profiling Using Genechip Human Exon 1.0 ST Arrays. *The Journal of Molecular Diagnostics*, **10**, 50-57. https://doi.org/10.2353/jmoldx.2008.070085

[30] Clark, J., Merson, S., Jhavar, S., Flohr, P., Edwards, S., Foster, C.S., *et al.* (2006) Diversity of TMPRSS2-ERG Fusion Transcripts in the Human Prostate. *Oncogene*, **26**, 2667-2673. https://doi.org/10.1038/sj.onc.1210070

[31] Hermans, K.G., van Marion, R., van Dekken, H., Jenster, G., van Weerden, W.M. and Trapman, J. (2006) TMPRSS2:ERG Fusion by Translocation or Interstitial Deletion Is Highly Relevant in Androgen-Dependent Prostate Cancer, but Is Bypassed in Late-Stage Androgen Receptor-Negative Prostate Cancer. *Cancer Research*, **66**, 10658-10663. https://doi.org/10.1158/0008-5472.can-06-1871

[32] Iljin, K., Wolf, M., Edgren, H., Gupta, S., Kilpinen, S., Skotheim, R.I., *et al.* (2006) TMPRSS2 Fusions with Oncogenic ETS Factors in Prostate Cancer Involve Unbalanced Genomic Rearrangements and Are Associated with HDAC1 and Epigenetic Reprogramming. *Cancer Research*, **66**, 10242-10246. https://doi.org/10.1158/0008-5472.can-06-1986

[33] Cerveira, N., Ribeiro, F.R., Peixoto, A., Costa, V., Henrique, R., Jerönimo, C., *et al.* (2006) TMPRSS2-ERG Gene Fusion Causing ERG Overexpression Precedes Chromosome Copy Number Changes in Prostate Carcinomas, Paired HGPIN Lesions. *Neoplasia*, **8**, 826-832. https://doi.org/10.1593/neo.06427

[34] Wang, J., Cai, Y., Ren, C. and Ittmann, M. (2006) Expression of Variant TMPRSS2/ERG Fusion Messenger RNAs Is Associated with Aggressive Prostate Cancer. *Cancer Research*, **66**, 8347-8351. https://doi.org/10.1158/0008-5472.can-06-1966

[35] Cathomen, T. and Keith Joung, J. (2008) Zinc-Finger Nucleases: The Next Generation Emerges. *Molecular Therapy*, **16**, 1200-1207. https://doi.org/10.1038/mt.2008.114

[36] Angermueller, C., Lee, H., Reardon, B., Chen, Y., Linder, J., Friedensohn, S., *et al.* (2023) A Universal Deep-Learning Model for Zinc Finger Design Enables Transcription Factor Reprogramming. *Nature Biotechnology*, **41**, 1343-1353. https://doi.org/10.1038/s41587-023-01762-4

[37] Shao, L., Chang, D.C., Yan, W., Ye, H., Zhang, Y., Li, R., *et al.* (2021) Targeting the TMPRSS2/ERG Fusion mRNA Using Liposomal Nanovectors Enhances Docetaxel Treatment in Prostate Cancer. *International Journal of Nanomedicine*, **16**, 5433-5450. https://doi.org/10.2147/IJN.S308694

[38] Liu, Y., Jia, Y., Yang, K. and Tong, Y. (2012) OmpF Deletion Increases Antibiotic Resistance in *Escherichia coli*. *Microbial Pathogenesis*, **52**, 293-296. https://doi.org/10.1016/j.micpath.2012.01.007

[39] Viveiros, M., Martins, A., Couto, I., Rodrigues, L., Spengler, G., Martins, M. and Amaral, L. (2007) New Methods for the Identification of Efflux-Mediated MDR Bacteria, with a Focus on *E. coli*. *Drug Resistance Updates*, **10**, 255-265. https://doi.org/10.1016/j.drup.2007.10.001

[40] Miroux, B. and Walker, J.E. (1996) Over-Production of Proteins in *Escherichia coli*: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology*, **260**, 289-298. https://doi.org/10.1006/jmbi.1996.0399

[41] Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., *et al.* (2008) Tuning *Escherichia coli* for Membrane Protein Overexpression. *Proceedings of the National Academy of Sciences*, **105**, 14371-14376. https://doi.org/10.1073/pnas.0804090105