**RESEARCH ARTICLE**

# Quantification of cytosolic plasmid DNA degradation using high-throughput sequencing: implications for gene delivery

Rahul Rattan[1,3]
Anna U. Bielinska[3]
Mark M. Banaszak Holl[1,2,3*]

[1]*Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA*

[2]*Department of Chemistry, University of Michigan, Ann Arbor, MI, USA*

[3]*Michigan Nanotechnology Institute for Medicine and Biological Sciences, University of Michigan, Ann Arbor, MI, USA*

*Correspondence to:
M. M. Banaszak Holl, Chemistry Department, 930 N. University Avenue, University of Michigan, Ann Arbor, MI 48109-1055, USA.
E-mail: mbanasza@umich.edu

## Abstract

**Background**    Although cytosolic DNA degradation plays an important role in decreasing transgene expression, the plasmid degradation pattern remains largely unexplored.

**Methods**    Illumina dye sequencing was employed to provide degradation site information for S1 and cytosolic nucleases. S1 nuclease provided a positive control for a comparison between the agarose gel method and sequencing approaches.

**Results**    The poly(A) region between the β-lactamase gene and the cytomegalovirus (CMV) promoter was identified as the most likely cut site for polyplex-treated cytosol. The second most likely site, at the 5' end of the β-lactamase gene, was identified by gel electrophoresis and sequencing. Additional sites were detected in the *OriC* region, the SV40/poly(A) region, the luciferase gene and the CMV promoter. Sequence analysis of plasmid treated with cytosol from control cells showed the greatest cut activity in the *OriC* region, the β-lactamase gene and the poly(A) region following the luciferase gene. Additional regions of cut activity include the SV40 promoter and the β-lactamase poly(A) termination sequence. Both cytosolic nucleases and the S1 nuclease showed substantial activity at the bacterial origin of replication (*OriC*).

**Conclusions**    High-throughput plasmid sequencing revealed regions of the luciferase plasmid DNA sequence that are sensitive to cytosolic nuclease degradation. This provides new targets for improving plasmid and/or polymer design to optimize the likelihood of protein expression. Copyright © 2014 John Wiley & Sons, Ltd.

**Keywords**    biotechnology; HeLa; plasmid design; plasmid gene delivery; polymer vector

## Introduction

Antisense DNA, silencing RNA and plasmid DNA (pDNA) remain promising therapeutics with possible applications including diagnosis, vaccines, and disease management [1]. The range of applications and potential impact on common, life-threatening diseases such as diabetes mellitus and metastatic

neoplasm [1] led to a predicted pharmaceutical industry of $16.8 billion/year by 2020–2025 [2]; however, almost 15 years after this prediction, gene therapy is only a small fraction of this multibillion dollar market projection. This lack of market availability and penetration arises primarily because of safety and efficiency issues for current nucleic acid delivery systems [3,4]. The development of nonviral delivery systems such as polycationic polymers presents a unique and unprecedented opportunity to genetic engineers [5,6]. Polymers allow a significant amount of customization, provide a high surface moiety to volume ratio and are relatively safe compared to viral systems [7,8]; however, they still have the major drawback of low transgene expression. Incomplete understanding of cellular uptake mechanisms [9,10], inefficient endosomal escape [11–13] and poorly designed and understood nuclear uptake [13,14] are considered to be the major reasons limiting optimization [3,13].

Recent work by Lechardaur *et al.* [15] and Pollard *et al.* [16] demonstrates that cyotosolic nucleases play a significant role in limiting expression efficiency. Our recent work has shown that differential transgene expression facilitated by jetPEI™ (Polyplus-transfection Inc., New York, NY, USA), linear poly(ethylenimine) and generation 5 poly(amidoamine) (G5 PAMAM) dendrimer can be explained by the activation of cytosolic nucleases that degrade the plasmid into a nonfunctional form [17]. Characterizing the activity pattern of these cytosolic nucleases is a key step in the development of better synthetic strategies for polymers capable of delivering, protecting and releasing the plasmid. Substantial efforts have been expended to make the plasmid less susceptible to nuclease degradation via backbone, codon and poly(A) modifications and by removing viral DNA components [18,19].

In the present study, high-throughput sequencing is used to identify and quantify labile sites on a pGL4.51 luciferase expressing plasmid that is susceptible to cytosolic nucleases. S1 nuclease was used both to validate the experimental protocols and to develop a comparison with the published cleavage pattern determined using agarose gel electrophoresis [18,19]. High-throughput sequencing not only identifies more labile sites than previously observed using agarose gel electrophoresis, but also provides information about the relative degree of degradation per site and substantially greater resolution of the site location in the plasmid sequence. Many of the cleavage sensitive sequences lay in homopurine and homopyrimidine regions of the synthetic/viral poly (A) regions and the plasmid's bacterial origin of replication (*OriC*); however, labile sites were also identified in the reporter and antibiotic resistance genes required for stable expression [20]. Many of the labile sites identified in the plasmid correspond to regions that are indispensible for plasmid functionality.

Human epithelial carcinoma cells (HeLa cells) were used based on previous studies exploring the role of cytosolic nucleases [15,18]. S1 nuclease was used as the control nuclease because it has been shown that customizing plasmid sequences on the basis of S1 nuclease activity can improve plasmid stability [18,19]. The luciferase expression plasmid was selected because it is widely used in gene expression studies in mammalian cells and understanding the susceptibility to cytosolic nucleases should be useful for a broad range of research groups. The plasmid also contains the commonly employed selective Neomycin gene to select for permanently transgene expressing mammalian cells.

# Materials and methods

## Protocol for obtaining cell cytosol

HeLa cells (catalogue #CCL-2) were obtained from American Type Culture Collection (ATCC, Manassas, VA, USA). S1 nuclease and $10 \times$ S1 nuclease buffer, salmon sperm DNA (ssDNA) and luciferase plasmid (pGL4.51) were purchased from Promega Corp. (Madison, WI, USA). A circular map of the plasmid is provided in the Supporting information (Figure S1). The Serum free media (SFM) employed in the present study was minimal essential media (MEM) with glutamine and Earle salts obtained from Life Technologies (Life Technologies, Grand Island, NY, USA). Complete media was made by adding 50 ml of fetal bovine serum and 5 ml of $100 \times$ penicillin-streptomycin to 500 ml of MEM. The BCA protein assay kit was purchased from Thermo Fisher Scientific, Inc. (Waltham, MA, USA). PAMAM dendrimer was purchased from Dendritech, Inc. (Midland, MI, USA). In-house characterization revealed a molecular weight of 28 900 and a polydispersity index of 1.01. Titration showed 116 amines per dendrimer.

For transfection and cytosol extraction, 500 000 HeLa cells/well were cultured in a six-well plate with 1.5 ml of complete media at 37°C overnight. Subsequently, cells were treated in duplicate with a 10:1 nitrogen to phosphorus ratio of G5 PAMAM polyplexes (2 μg of ssDNA, 15.13 μg of G5 PAMAM, 100 μl of water) for 3 h in 0.9 ml of SFM. After 3 h, the transfection media was replaced with 1.5 ml of complete media and the cells were incubated for another 1 h. Cells were then washed with phosphate-buffered saline (PBS) without $Ca^{2+}$ and $Mg^{2+}$ and incubated with 200 μl of 0.1% trypsin for 5 min at 37°C. Then, 1.8 ml of cold PBS (4°C) with $Ca^{2+}$ and $Mg^{2+}$ ions was added to each well to stop trypsin activity. Duplicates were pooled together and cells were centrifuged at 200 *g*. To obtain cell lysates, the cell pellet was suspended in 200 μl of ice-cold cell lysis buffer and

incubated for 30 min on ice. Afterwards, the samples were centrifuged at 500 *g*. for 5 min at 4°C and the supernatants were collected. A BCA assay was used to assess and normalize protein concentration in subsequent experiments. The control lysates from nontreated cells were prepared identically, except that no polyplexes were added in the initial step.

## Sequencing and gel protocols

NP-40 lysis buffer was procured from Boston BioProducts Inc. (Worcester, MA, USA), protease inhibitor cocktail was obtained from Sigma-Aldrich (St Louis, MO, USA). The Gel extraction kit was purchased from Qiagen (Valencia, CA, USA) (catalogue number 28706) and used as directed. 10 × BlueJuice™ gel loading buffer was obtained from Life Technologies Corp. (catalogue number 0816–015). Three restriction endonucleases, *Sac*I (catalogue number R3156S), *Dra*III (catalogue number R3510S) and *Stu*I (catalogue number R0187S), were obtained from New England BioLabs, Inc. (Beverly, MA, USA). Nb.BtsI (catalogue number R0707S) and Nt.CviPII (catalogue number R0626S) were obtained from New England BioLabs Inc. An Illumina HiSeq 2000 (Illumina, Inc., San Diego, CA, USA) was employed for all sequencing experiments. Initial sequence alignment was completed with the help of BOWTIE (http://bowtie-bio.sourceforge.net/index.shtml). Integrated Genomic Viewer (http://www.broadinstitute.org/igv/) was used to view the aligned sequence. MATLAB™ (MathWorks, Natick, MA, USA) was employed to generate heat-maps of nuclease activity. The heat map cut ratios were obtained by normalizing each base readcount by the average readcount for that sample and then dividing by the normalized readcount from the untreated DNA control.

## Protocol for treating Luciferase pDNA with cell cytosol

In total, 20 μg of pDNA was incubated with 10 μg of the isolated cytosol in 100 μl of water for 30 min at 37°C. S1 nuclease cleavage sample were prepared by adding 10 units of S1 nuclease to 20 μg of pDNA in 100 μl of water and 1 × S1 nuclease buffer for 30 min at 37°C. The negative control was prepared using 20 μg of pDNA resuspended in 100 μl of water and employed using the same incubation conditions.

For Illumina dye sequencing, samples containing 100 μl of the plasmid reaction mixture were treated with cell extracts (polyplex-treated cells or untreated cells) or S1 nuclease and optimized using the Qiagen gel extraction kit (the 100 μl of reaction mixture was used in place of

excised gel) to change the buffering media to 10 mM Tris-Cl at pH 8.5 for downstream sequencing. In addition, a plasmid-only control was brought through the identical procedure.

For agarose gel electrophoresis, 10 μl of the plasmid reaction mixtures described above were added to each well in 1 × BlueJuice™ gel loading buffer. The gel electrophoresis was performed on a 0.9% agarose gel containing ethidium bromide at 60 V for 3 h.

For the restriction nuclease analysis, agarose gel electrophoresis was performed using 10 μl of the S1 nuclease and luciferase plasmid reaction mixture in 10 parallel lanes (under the conditions used for this mixture in Figure 1, lane 5). The linear topology 6.4-kb form of the luciferase plasmid was then cut out from gel and purified using a Qiagen gel extraction kit. Next, 0.2 μg of the isolated 6.4-kb linear plasmid was incubated with either 2 units of *Sac*I, *Dra*III or *Stu*I restriction endonucleases in the supplied reaction buffer for 60 min at 37°C. Control digests were performed with these restriction endonucleases and using the luciferase plasmid under same conditions. *Sac*I, *Dra*III and *Stu*I endonucleases were used because they each have a single restriction site in the plasmid sequence, which provides sufficient resolution to map the pDNA. The endonuclease treated samples were analyzed on 0.9% agarose gel containing ethidium bromide at 60 V for 3 h.
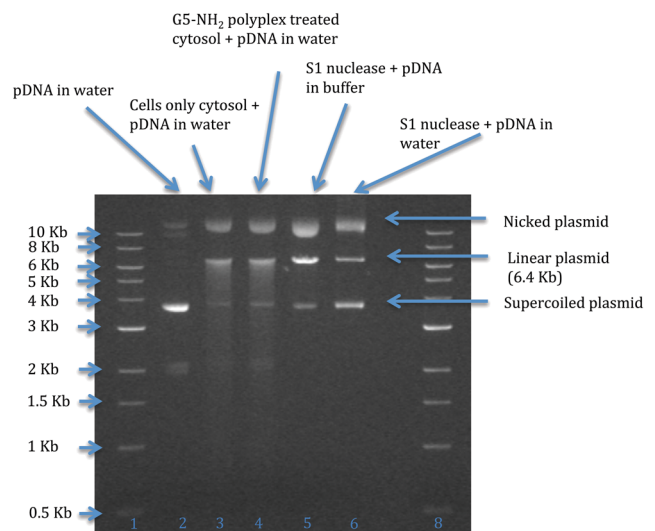


Figure 1. Agarose gel electrophoresis showing the DNA cleavage pattern of HeLa cytosol and S1 nuclease treated luciferase pDNA. Lanes 1 and 8: 10-kb DNA ladder. Lane 2: pDNA in water. Lane 3: pDNA treated with cells only cytosol in water. Lane 4: pDNA treated with cytosol in water from cells exposed to G5 polyplexes. Lane 5: pDNA treated with S1 nuclease in S1 nuclease buffer. Lane 6: pDNA treated with S1 nuclease in water. In lanes 3–6, both the nicked and linear topology of the plasmid can be seen.

# Results and Discussion

## Cleavage sites as determined by agarose gel electrophoresis

The degradation pattern resulting from treatment of luciferase plasmid with S1 nuclease or the nucleases present in cytosol was first studied using agarose gel electrophoresis. Figure 1 illustrates that upon treatment with either S1 or cytosolic nucleases the supercoiled pDNA degrades into two major forms, a linear plasmid of 6.4 kb and nicked plasmids appearing as approximately 10-kb fragments compared to the linear DNA size ladder (Figure S2). The gel also shows substantial streaking, suggesting the presence of other sites susceptible to nuclease cleavage. The major band of the 6.4-kb linear plasmid was extracted from the gel and incubated with either *Sac*1, *Dra*III or *Stu*I endonucleases (Figure 2). These three restriction endonucleases, which cut at the 735, 2006 and 3212-bp sites, respectively, provide the redundancy necessary to define the primary S1 cut site. The pattern obtained with all three endonucleases indicated a S1 cut site located at the $5000 \pm 500$ bp position of the pDNA. This site is close to the homopurine/homopyrimidine rich plasmid *OriC* region, consistent with previous work that showed homopurine regions are susceptible to S1 nuclease
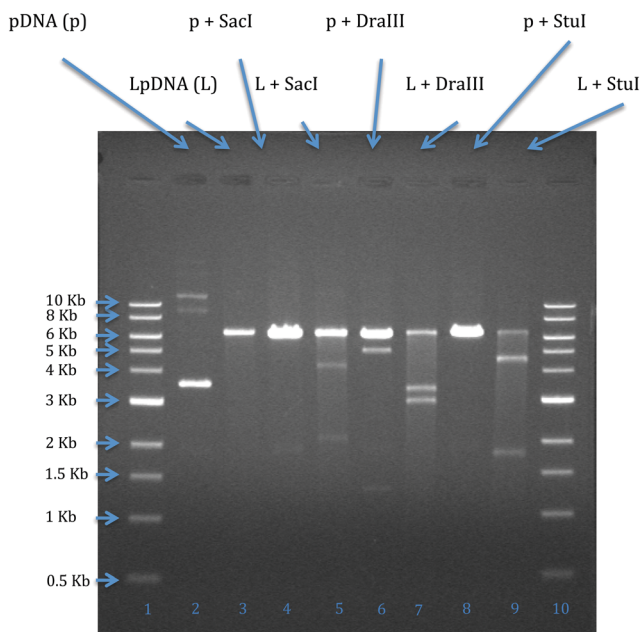
degradation [19]. In previous work, both the poly(A) and *OriC* regions were noted to be 'hot spots' for S1 cleavage [19].

The analysis presented in Figures 1 and 2 provides a useful start for understanding nuclease activity against plasmids in cells; however, two aspects of the experiment could benefit from substantial improvement. First, the inherent resolution of the gel electrophoresis method (approximately 500 bp) prevents a detailed sequence determination of where the pDNA was cut. This limits the extent to which one can understand the S1 and cytosolic nuclease activity and make comparisons between them. Second, it appeared unlikely, particularly giving the streaking in Figure 1, that a single cleavage sequence existed for either cytosolic nucleases or the S1 nuclease, suggesting that the gels did not have the desired level of sensitivity. In addition to these concerns, Sanger sequencing is limited by the choice of primers. Given this set of concerns and challenges, the high-throughput Illumina dye sequencing method was employed to explore the range of cellular nuclease cuts introduced on the pDNA [21]. We employed S1 nuclease as both a positive control, in comparison with the gel work, and to learn more about the S1 activity. There are number of advantages to using high-throughput Illumina dye sequencing for the cytosolic nuclease experiments. The development of specific sequencing primer sets, as well as the need to perform multiple gels for the analysis of restriction enzyme patterns for each cut site of interest, is avoided (as illustrated in Figure 2). In addition, the high-throughput sequencing also provides higher resolution of the cut site location. Lastly, high-throughput sequencing defines the degree of cutting over a region (*vida infra*), as opposed to the notion of a singular base pair location. This observation is important both for design of more robust plasmids and for the design of polymers that might protect these regions.

## Cleavage sites as determined by high-throughput sequencing

High-throughput Illumina dye sequencing was performed in three steps [21]. The first step was library preparation where the treated and control luciferase pDNA samples were randomly broken into smaller fragments with sonication, adenylated, connected to adapter oligonucleotides, size-selected and purified. Next, the materials underwent cluster generation where each library of fragments was clonally amplified. The final step was sequencing, which was performed for each cluster in parallel, using fluorescent deoxyribonucleotides.

Heat maps illlustrate both the location and extent of nuclease cut on the luciferase pDNA. Figures 3A, 4A, and 4D illustrate the full set of data. Figures 3B, 4B,



**Figure 2.  Endonuclease treatment of 6.4-kb linear plasmid (LP).** Lanes 1 and 10: 10-kb DNA ladder. Lane 2: pDNA in water. Lane 3: LP in water. Lanes 4, 6 and 8: pDNA treated with *Sac*I, *Dra*III and *Stu*I, respectively. Lanes 5, 7 and 9: LP treated with *Sac*I, *Dra*III and *Stu*I, respectively. The 6.4-kb LP is derived from treatment of luciferase pDNA with S1 nuclease.
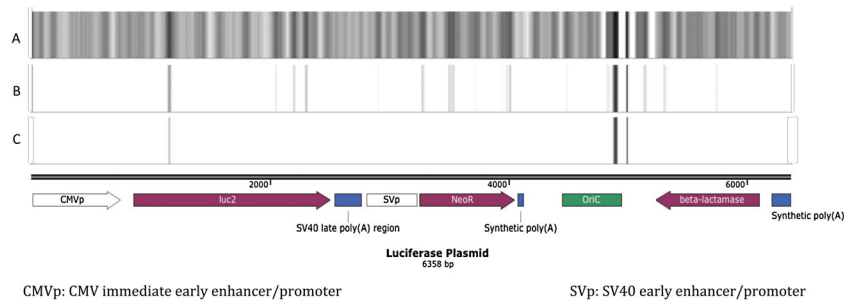
**Figure 3.** Heat maps showing the frequency of cuts on the luciferase pDNA resulting from the S1 nuclease digest in water. White color indicates that the percentage of plasmids cut is between zero and the lower part of the cut range listed. Black color indicates that the percentage cut is the maximum value for that cut range. Greyscale indicates cut percentages between these extrema. (A) The full range of labile sites. (B) Labile regions that have frequency of cuts between 25% and 45% (showing approximately 15 of the most cut regions). (C) Approximately top 5 of the cut regions between 28% and 45%. The luciferase functional map is provided in registry with the heat map to facilitate comparison of cuts sites to the different regions of the plasmid. CMVp, CMV immediate early enhancer/promoter; SVp, SV40 early enhancer/promoter.
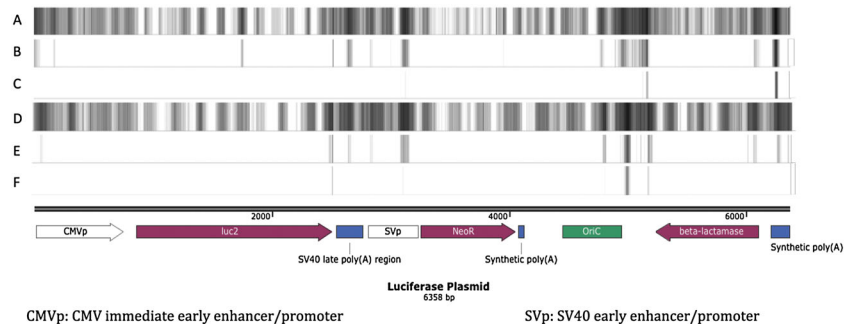


**Figure 4.** Heat maps showing frequency of cuts on the luciferase plasmid resulting from cytosolic nuclease treatments. White color indicates the percentage of cut is between zero and the lower part of the cut range listed. Black color indicates the percentage of cut is the maximum value for that cut range. Greyscale indicates cut percentages between these extrema. (A–C) Labile cut regions for G5 polyplex-treated cell cytosol treatment. (A) The full range of labile sites. (B) Labile regions that have a frequency of cuts between 41% and 65% (showing approximately 15 of the most cut regions). (C) Approximately top 5 of the cut regions between 52% and 65%. (D–F) Cells-only cytosol treatment. (D) The full range of labile sites. (E) Labile regions that have a frequency of cuts between 50% and 75% (showing approximately 15 of the most cut regions). (F) Approximately top 5 of the cut regions between 59% and 75%. CMVp, CMV immediate early enhancer/promoter; SVp, SV40 early enhancer/promoter.

and 4E focus on approximately the top 15 cut regions. Figures 3C, 4C, and 4F focus on the top 3–5 cut regions for easiest comparison with the agarose gel electrophoresis data. At the bottom of Figures 3 and 4, a luciferase gene functional map is provided in registry with the heat map to facilitate comparison of cut sites with the different regions of the pDNA plasmid. A black line indicates a cut extent equal to the highest value of the range. White regions indicate a cut frequency lower than or equal to the lowest value of the specified range. The greyscale vertical lines indicate the percentage of plasmids cut at that location.

The heat map in Figure 3 illustrates S1 nuclease activity on luciferase pDNA in water. Figure 3A indicates the locations where 0–45% of the plasmids were cut. There are few 'white' regions corresponding to sequences where no

cuts were detected, indicating that the nuclease is promiscuous with respect to the types of base sequences suitable for a cleavage location. Figure 3B illustrates only sites in which 25–45% of the plasmids were cleaved and Figure 3C illustrates sites in which only 28–45% of the plasmids were cleaved. Information regarding each cut site, including percentage AT content, function of the plasmid region, cut location and width, and fraction of plasmids cut is summarized in Table 1 for the top 15 regions (Figure 3B). These regions include four cuts (S1–4) in the luciferase gene, a cut in the SV40 promoter/enhancer (S5), cuts in three locations on the Neomycin gene (S6–8), cuts at three sites on the origin of replication region (S9–11) and three cut sites on ampicillin gene (S13–15). Cuts S9–11 at *OriC* are consistent with those previously determined by agarose gel electrophoresis [19] and illustrated

**Table 1. Summary of the top 15 regions of the luciferase plasmid cut by S1 nuclease, as shown in Figure 3B.**

| Region | Percentage AT content | Function | Cut location and width (bp) | Percentage extent of cut (average, maximum) |
|---|---|---|---|---|
| S1 | 43 | Luciferase gene | 1142–1187 (46) | 29, 33 |
| S2 | 50 | '' | 2042–2067 (26) | 26, 27 |
| S3 | 36 | '' | 2195–2217 (22) | 27, 29 |
| S4 | 25 | '' | 2292–2319 (28) | 28, 30 |
| S5 | 55 | SV40 promoter | 2903–2913 (11) | 25, 26 |
| S6 | 39 | Neomycin gene | 3261–3296 (36) | 27, 29 |
| S7 | 32 | '' | 3489–3544 (56) | 28, 29 |
| S8 | 38 | '' | 3967–4018 (52) | 27, 29 |
| S9 | 53 | *OriC* | 4474–4484 (15) | 26, 26 |
| S10 | 54 | '' | 4812–4910 (99) | 31, 41 |
| S11 | 78 | '' | 4974–4990 (17) | 35, 38 |
| S12 | 48 | Close to *OriC* | 5062–5076 (14) | 25, 26 |
| S13 | 67 | Ampicillin gene | 5116–5145 (30) | 27, 29 |
| S14 | 47 | '' | 5283–5325 (43) | 26, 28 |
| S15 | 53 | '' | 5725–5741 (17) | 26, 28 |

in Figure 1. The homopurine/homopyrimidine rich sites like poly(A) and *OriC* have been reported as susceptible to S1 nucleases using gel electrophoresis [19,22]; however, using high-throughput sequencing, we have found many additional labile regions not previously identified. These findings call into question the long held hypothesis that S1 nuclease only attacks single-stranded secondary structures of DNA because many of the cleaved regions are considered to maintain a double-stranded DNA (dsDNA) structure [18,19,22]. Alternatively, if S1 nuclease is only active against single-stranded DNA, these data indicate regions of substantial structural fluxionality in regions generally considered to maintain a dsDNA structure.

Figure 4 illustrates the activity of cytosolic nucleases overlaid on the luciferase pDNA functional map. Figures 4A to 4C provide data for nuclease activity from cytosol obtained for cells treated with G5 polyplexes. In a previous study, we demonstrated that nuclease activity is polyplex dependent and is negatively related to the level of protein expression [17]. G5 polyplexes were selected because they induced the maximum level of nuclease activity and gave the minimum level of expression. We also prepared lysates from cells that had not been treated with polyplexes to establish the 'background' level of nucleases. The data presented in Figures 4D to 4F, although important and informative, must be treated with caution when ascribing nuclease levels in a normal, functional cell. Caveats include the neoplastic origin of HeLa cancer cell line and also that the cytosol was obtained by cell lysis, which is a procedure that may itself trigger nuclease activity [23,24].

Figures 4A and 4D show all the labile sites present in 0–65% and 0–75% extent of cut range for the G5 polyplex-treated and untreated cells, respectively. Similar to S1 nuclease, the cytosolic nucleases are promiscuous and attack numerous regions in addition to the most common cuts in the poly(A) and viral enhancer/promoter regions of the pDNA. Figures 4B and 4E represent the 15 most labile regions lying in the range of 41–65% and 50–75% for G5 polyplex-treated and untreated cell cytosolic samples, respectively. Figure 4C (52–65% extent of cut) and Figure 4F (59–75% extent of cut) represent the most labile regions and are most useful for comparison with gel electrophoresis data.

Table 2 summarizes the top 15 most labile regions obtained with the cytosol from untreated cells (data in Figure 4E). Cytosolic nucleases in this cytosol produce four cuts in the viral cytomegalovirus (CMV) enhancer/promoter region, one cut in the luciferase gene, three cuts in the poly(A) regions, two cuts on *OriC* and three cuts on the ampicillin gene. Table 3 shows the top 15 labile sites for cytosol from G5 polyplex treated cells (data in Figure 4B). Three cuts are observed in the viral enhancer/promoter regions, one cut in the luciferase gene, two cuts in the poly (A) sequence, two cuts near *OriC*, one cut in the ampicillin resistance gene and one cut in the neomycin gene.

Cytosol nuclease activity from untreated cells and G5 polyplex treated cells display significant similarities. The top 15 labile regions for cytosolic samples can be divided into three major regions: the viral promoter/enhancer sequence, poly(A) sequences and *OriC*. There are also some cut regions in the luciferase reporter and the selective neomycin resistance gene. Cuts in viral promoter/enhancers, poly(A) sequences and the transgene can severely abrogate transgene expression [18,25]. The multiple cuts in the selective neomycin and ampicillin resistance genes point to the additional challenge in efforts to establish the cell lines with stable plasmid expression. The bacterial origin of replication (*OriC*) is necessary for plasmid replication in bacteria [26], although it has little functional role in eukaryotes [27]. The discovery that *OriC* is one of the most degraded regions for S1 (Table S1) and cytosolic nucleases (Tables S2 and S3) suggests that eukaryotic cells may recognize *OriC* as a marker for foreign DNA. In summary, we have detected a greater number of labile regions

**Table 2. Summary of the top 15 regions of the luciferase plasmid cut by cells only cytosolic nucleases (as shown in Figure 4E).**

| Region | Percentage AT content | Function | Cut location and width (bp) | Percentage extent of cut (average, maximum) |
|---|---|---|---|---|
| C1 | 67 | CMV promoter | 62–113 (52) | 51, 53 |
| C2 | 57 | '' | 321–327 (7) | 51, 53 |
| C3 | 50 | Luciferase gene | 2483–2520 (38) | 55, 63 |
| C4 | 62 | SV40 poly(A) | 2573–2580 (8) | 51, 53 |
| C5 | 73 | '' | 2633–2683 (51) | 53, 57 |
| C6 | 62 | SV40 promoter | 2831–2859 (29) | 52, 54 |
| C7 | 44 | '' | 3081–3160 (80) | 58, 61 |
| C8 | 45 | *OriC* | 4774–4804 (31) | 58, 61 |
| C9 | 50 | '' | 4915–5022 (103) | 58, 71 |
| C10 | 50 | Close to *OriC* | 5060–5063 (4) | 52, 53 |
| C11 | 62 | '' | 5075–5126 (52) | 51, 56 |
| C12 | 86 | Ampicillin gene | 5148–5196 (49) | 57, 62 |
| C13 | 58 | '' | 5991–6002 (12) | 52, 55 |
| C14 | 42 | '' | 6025–6066 (42) | 54, 57 |
| C15 | 70 | Synthetic poly(A) | 6217–6273 (57) | 54, 59 |

**Table 3. Summary of the top 15 regions of the luciferase plasmid cut by G5 polyplex-treated cells cytosolic nucleases (as shown in Figure 4B).**

| Region | Percentage AT content | Function | Cut location and width (bp) | Percentage extent of cut (average, maximum) |
|---|---|---|---|---|
| G1 | 57 | CMV promoter | −4 to 33 (37) | 45, 49 |
| G2 | 83 | '' | 38–55 (18) | 42, 44 |
| G3 | 37 | '' | 141–148 (8) | 41, 42 |
| G4 | 75 | '' | 150–181 (32) | 41, 45 |
| G5 | 56 | Luciferase gene | 1738–1764 (27) | 44, 48 |
| G6 | 73 | SV40 poly(A) | 2610–2686 (71) | 45, 55 |
| G7 | 54 | SV40 promoter | 2827–2850 (24) | 42, 44 |
| G8 | 25 | '' | 3002–3005 (4) | 42, 44 |
| G9 | 47 | '' | 3083–3156 (74) | 48, 54 |
| G10 | 33 | Neomycin gene | 3947–3949 (3) | 41, 41 |
| G11 | 33 | Close to *OriC* | 4359–4361 (3) | 41, 41 |
| G12 | 48 | *OriC* | 4748–4797 (50) | 43, 48 |
| G13 | 59 | '' | 4882–5182 (300) | 47, 59 |
| G14 | 51 | Ampicillin gene | 6051–6109 (51) | 44, 49 |
| G15 | 72 | Synthetic poly(A) | 6208–6275 (60) | 52, 61 |

susceptible to cleavage for cytosolic extracts than for S1 nuclease exposure. The greater number of cleaved regions likely arises from the presence of a number of different mammalian nucleases being present in the cytosol, although we cannot rule out the presence of a single class of nuclease that are simply less selective.

In other studies, treatment with G5 PAMAM polyplexes lead to the highest cytosolic nuclease activation leading to a decrease in transgene expression [17]. Figure 5 shows that G5 PAMAM polyplex treated cell cytosol has activity similar to control cell cytosol with several notable, and possibly significant differences, especially in the luciferase transgene. The labile region in the luciferase gene is one of the top 15 most labile regions for G5 polyplex treated cytosol; however, this is not the case for untreated cells cytosol. This may contribute to poor luciferase transgene expression facilitated by G5 PAMAM polyplexes. This is consistent with the hypothesis that G5 PAMAM polyplexes induce heightened activation of cytosolic nuclease, therefore decreasing the transgene expression [17]. When

making this comparison, it is important to note that all cells were lysed for the sequencing experiments, which may lead to changes in the activity of cytosolic nucleases.

## No consensus cleavage sequence was identified

Apart from showing the position of labile sites, sequencing also provides a unique opportunity to evaluate homology between the labile sites. To study and quantify homology between the labile sequence for S1 and cytosolic sample, we attempted to develop a consensus sequence by performing a multiple sequence alignment of the most labile region of the cytosolic samples and the S1 nuclease sample (S10, C9, G13 in Tables 1, 2 and 3, respectively). The best consensus sequence identified, AGCTCTTGA-TCCGGCAAACAAACCACC, was then used to calculate similarity for other labile sequence locations. Tables S3, S4, and S5 show the analyses performed for S1 nuclease, control
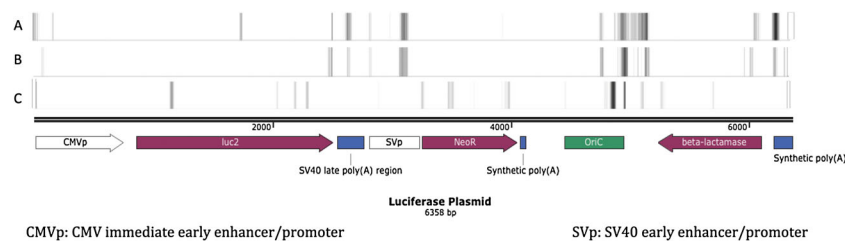
**Figure 5.** Heat map comparison between cytosol nuclease samples and S1 nuclease. (A) G5 polyplex treated cytosol cut regions with 41–65% cuts. (B) Cells-only cytosol cut regions with 50–75% cuts. (C) S1 nuclease cut regions with 25–45% cuts. CMVp, CMV immediate early enhancer/promoter; SVp, SV40 early enhancer/promoter.

cytoplasm and the G5 polyplex treated cells cytoplasm samples, respectively. This analysis shows that there exists very low sequence homology between the best identified 'consensus sequence' and labile regions for S1 nuclease and cytosolic samples. The lack of a consensus sequence for the cleavage locations supports the hypothesis that nucleases are not primary DNA sequence specific, and can be considered promiscuous with respect to the base pair sequence.

## Conclusions

Agarose gel electrophoresis and Illumina dye sequencing confirm that cytosolic nucleases cleave luciferase pDNA, rendering it a nonfunctional template for expression, and that S1 nuclease cuts at high homopurine/homopyrimidine regions of the plasmid [18,19]. To the best of our knowledge, this is the first time high-throughput sequencing has been used to study degradation effects of S1 and cytosolic nucleases on pDNA. Interestingly, the most susceptible cut region for S1 nuclease overlaps the susceptible region for cytosolic nucleases, consistent with the

hypothesis that S1 type nucleases are part of the cytosolic nuclease milieu. Lack of significant homology between labile sequences and a consensus sequence suggests that these nucleases are highly promiscuous. In addition, our agarose gel results are consistent with the hypothesis that S1 and cytosolic nucleases are more likely to cleave supercoiled pDNA than the linear form of DNA. Our data further indicates that, because many of susceptible sites lie in the indispensible functional units of pDNA, there is a limit to DNA sequence optimization. To overcome this challenge, further research is needed on polymers designed to protect these labile regions from degradation during the transport of pDNA into the cell and to the nucleus.

## Acknowledgements

## References

1. Scholz C, Wagner E. Therapeutic plasmid DNA versus siRNA delivery: common and different tasks for synthetic carriers. *J Control Release* 2012; **161**: 554–565.

2. Hess P, Cooper D. Impact of pharmacogenomics on the clinical laboratory. *Mol Diagnosis* 1999; **4**: 289–298.

3. Liu C, Zhang N. Nanoparticles in gene therapy principles, prospects, and challenges. *Prog Mol Biol Transl Sci* 2011; **104**: 509–662.

4. Lechardeur D, Lukas GL. Intracellular barriers to non-viral gene transfer. *Curr Gene Ther* 2002; **2**: 183–194.

5. Smedt SCD, Demeester J, Hennink WE. Cationic polymer based gene delivery systems. *Pharm Res* 2000; **17**: 113–126.

6. Benns JM, Choi J-S, Mahato RI, *et al*. pH-sensitive cationic polymer gene delivery vehicle: N-Ac-poly(L-histidine)-graft-poly(L-lysine) comb shaped polymer. *Bioconjug Chem* 2000; **11**: 637–645.

7. Luo D, Saltzman WM. Synthetic DNA delivery systems. *Nature Biotechnol* 2000; **18**: 33–37.

8. Lungwitz U, Breunig M, Blunk T, *et al*. Polyethylenimine-based non-viral gene delivery systems. *Eur J Pharm Biopharm* 2005; **60**: 247–266.

9. Hong S, Rattan R, Majoros IJ, *et al*. The role of ganglioside GM1 in cellular internalization mechanisms of poly (amidoamine) dendrimers. *Bioconjugate Chem* 2009; **20**: 1503–1513.

10. Qi R, Mullen DG, James R, Baker J, *et al*. The mechanism of polyplexes internalization into cells: testing the GM1/caveolin-1-mediated lipid raft mediated endocytosis pathway. *Mol Pharm* 2010; **7**: 267–279.

11. Varkouhi AK, Scholte M, Storm G, *et al*. Endosomal escape pathways for delivery of biologicals. *J Control Release* 2011; **151**: 220–228.

12. Cho YW, Kim J-D, Park K. Polycation gene delivery systems: escape from endosomes to cytosol. *J Pharm Pharmacol* 2003; **55**: 721–734.

13. Lechardeur D, Verkman AS, Lukacs GL. Intracellular routing of plasmid DNA during non-viral gene transfer. *Adv Drug Deliv Rev* 2005; **57**: 755–767.

14. Dean D, Strong D, Zimmer W. Nuclear entry of nonviral vectors. *Gene Ther* 2005; **12**: 881–890.

15. Lechardeur D, Sohn KJ, Haardt M, *et al*. Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. *Gene Ther* 1999; **4**: 482–497.

16. Pollard H, Toumaniantz G, Amos J-L, *et al*. Ca$^{2+}$-sensitive cytosolic nucleases

prevent efficient delivery to the nucleus of injected plasmids. *J Gene Med* 2001; **3**: 153–164.

17. Rattan R, Vaidyanathan S, Wu G, *et al*. Polyplex-induced cytosolic nuclease activation leads to differential transgene expression. *Mol Pharm* 2013; **10**: 3013–3022.

18. Azzoni AR, Ribeiro SC, Monteiro GA, *et al*. The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. *J Gene Med* 2007; **9**: 392–402.

19. Ribeiro SC, Monteiro GA, Prazeres DMF. The role of polyadenylation signal secondary structures on the resistance of plasmid vectors to nucleases. *J Gene Med* 2004; **6**: 565–573.

20. Condreay JP, Witherspoon SM, Clay WC, *et al*. Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc Natl Acad Sci U S A* 1999; **96**: 127–132.

21. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010; **2010**: pdb.prot5448.

22. Hoffman EK, Trusko SP, Murphy M, *et al*. An S1 nuclease-sensitive homopurine/homopyrimidine domain in the c-Ki-ras promoter interacts with a nuclease factor. *Proc Natl Acad Sci U S A* 1990; **87**: 2705–2709.

23. Hallick RB, Chelm BK, Gray PW, *et al*. Use of aurintricarboxylic acid as an inhibitor of nucleases during nucleic acid isolation. *Nucleic Acids Res* 1977; **4**: 3055–3064.

24. Solymosy F, Fedorcsák I, Gulyás A, *et al*. A new method based on the use of diethyl pyrocarbonate as a nuclease inhibitor for the extraction of undegraded nucleic acid from plant tissues. *Eur J Biochem* 1968; **5**: 520–527.

25. Rao BS. Regulation of DNA replication by homopurine/homopyrimidine sequences. *Mol Cell Biochem* 1996; **2**: 163–168.

26. Mott ML, Berger JM. DNA replication initiation: mechanisms and regulation in bacteria. *Nature Rev Microbiol* 2007; **5**: 343–354.

27. Nasheuer H-P, Smith R, Bauerschmidt C, *et al*. Initiation of eukaryotic DNA replication: regulation and mechanisms. *Prog Nucleic Acid Res Mol Biol* 2002; **72**: 41–93.

# Supporting information

Additional supporting may be found in the online version of this article at the publisher's web site.

**Figure S1**. Circular plasmid map of pGL4.51.

**Figure S2**. Agarose gel electrophoresis comparing Luciferase plasmid digestion pattern for S1 nuclease with a known nicking endonuclease (Nb.BtsI, two sites) and a known degradation nuclease (CviPII, > 500 sites). Lanes 1 and (6) 10-kb DNA ladder. Lane (2) pDNA in water. Lane (3) pDNA treated with 0.2 units of S1 nuclease. Lane (4) pDNA treated with 1.0 unit of Nb.BtsI. Lane (5) pDNA treated with 1.0 unit of CviPII. 0.2 µg pDNA used for all lanes. Endonuclease treatments performed with supplied buffers for 1 h at 37°C. To prove the band appearing at approximately 10 kb is a nicked plasmid arising from the luciferase pDNA, we performed digestions with Nb.BtsI, a nicking endonuclease that acts at two locations, and CviPII endonuclease that nicks at >500 locations leading to complete digestion of the plasmid. The S1 nuclease results in three bands, including the nicked plasmid at approximately 10 kb as observed in Fig. 1. Treatment with the Nb.BtsI yields only the nicked plasmid at approximately 10 kb. Treatment with CviPII degrades the pDNA into <500-bp fragments. The total amount of signal at the 10-kb band location increase by approximately 10% upon treatment with S1 nuclease and approximately 45% upon treatment with Nb.BtsI. The sum of these data is consistent with assignment of the 10-kb band as nicked plasmid.

**Table S1**. Top 20 most cut sites for S1 nuclease in luciferase plasmid. S10_X represents the cut site belonging to the S10 region from Table S1.

**Table S2**. Top 20 most cut sites for S1 nuclease in luciferase plasmid. C9_X represents the cut site belonging to the C9 region from Table S3.

**Table S3**. Top 20 most cut sites for G5 polyplex treated cells cytosolic nuclease. G15_X and G13_X represents the cut site belonging to the G15 and G13 regions, respectively, from Table S5.

**Table S4**. S1 nuclease treatment: pairwise alignment using consensus sequence.

**Table S5**. Cytoplasmic extract from HeLa cells only treatment: pairwise alignment using consensus sequence.

**Table S6**. Cytoplasmic extract from G5 polyplex HeLa cells treatment: pairwise alignment using consensus sequence.