DMSO and Betaine Greatly Improve Amplification of GC-Rich Constructs in De Novo Synthesis

Michael A. Jensen*, Marilyn Fukushima, Ronald W. Davis
Stanford Genome Technology Center, Stanford University, Palo Alto, California, United States of America

Abstract
In Synthetic Biology, de novo synthesis of GC-rich constructs poses a major challenge because of secondary structure formation and mispriming. While there are many web-based tools for codon optimizing difficult regions, no method currently exists that allows for potentially phenotypically important sequence conservation. Therefore, to overcome these limitations in researching GC-rich genes and their non-coding elements, we explored the use of DMSO and betaine in two conventional methods of assembly and amplification. For this study, we compared the polymerase (PCA) and ligase-based (LCR) methods for construction of two GC-rich gene fragments implicated in tumorigenesis, IGF2R and BRAF. Though we found no benefit in employing either DMSO or betaine during the assembly steps, both additives greatly improved target product specificity and yield during PCR amplification. Of the methods tested, LCR assembly proved far superior to PCA, generating a much more stable template to amplify from. We further report that DMSO and betaine are highly compatible with all other reaction components of gene synthesis and do not require any additional protocol modifications. Furthermore, we believe either additive will allow for the production of a wide variety of GC-rich gene constructs without the need for expensive and time-consuming sample extraction and purification prior to downstream application.

Introduction
Since the de novo synthesis of the suppressor transfer RNA gene was first reported three decades ago [1], our ability to engineer and assemble synthetic gene constructs has revolutionized the field of biomedicine [2–5]. Yet, despite our many achievements from assembling multi-kilobase plasmids to whole genomes [6,7], de novo synthesis of GC-rich fragments remains a major obstacle namely because of secondary structure formation. Sequences populated with G repeats produce complex inter and intrastrand folding due to increased hydrogen bonding with neighboring guanines at their N-7 ring positions [8]. In PCR, this phenomenon is marked by the appearance of shorter bands following gel electrophoresis. These truncated versions of the target amplicon are primarily the consequence of arrest sites (hairpins) introduced into the template causing premature termination to polymerase extension [9]. In addition, mispriming and mis-annealing between template and compliment strands due to high melting temperature (Tm) of less GC-rich sequences, synthesis of longer oligodeoxynucleotides (ODN)s is often required, and may necessitate costly purification.

As a cheap and effective approach to disrupting secondary structure formation and minimizing high Tm ODN overlaps in de novo synthesis, we explored the use of the more popular and often referenced chemical agents, Dimethyl Sulfoxide (DMSO) [23,24] and betaine [25,26] during both the assembly and PCR amplification steps in conventional gene synthesis. These iso-stabilizing agents facilitate strand separation of double helix DNA by altering its melting characteristics. For example, betaine, an amino acid analog with both positive and negative charges close to neutral pH, acts to equilibrate the differential Tm between AT and GC base pairings; DMSO on the other hand, acts by disrupting inter and intrastrand re-annealing.

In this study, we compared the effects of these additives in the construction of two GC-rich gene fragments implicated in tumorigenesis, the Insulin-like Growth Factor 2 Receptor (IGF2R) and V-raf murine sarcoma viral oncogene homolog B1(BRAF) [29–31]. DMSO and betaine were also chosen because of their previously reported success in PCR amplification of the IGF2R gene fragment from a vector [25]. However, for our
purposes, IGF2R and BRAF were chemically synthesized and assembled in vitro by pooling overlapping, single-stranded ODNs using two conventional methods, the Polymerase Chain Assembly (PCA) [32] and the Ligase Chain Reaction (LCR) [33]. For a typical PCA reaction, assembly is done with one or two pre-PCR steps where single-stranded ODNs prime off each other, building up to the full-length product; 40 bp ODNs are designed (no gaps) with 20 bp overlap between template and compliment strands where a 5’ recess allows for polymerase binding and strand propagation. ODNs for LCR are the same as those for PCA except that each strand is 5’ phosphorylated for ligation. In this case, complimentary ODNs are denatured and annealed over several cycles for optimum strand alignment. A final round of PCR is then employed in both methods to amplify the target product using outside primers.

Here we report that DMSO and betaine greatly improve de novo synthesis of IGF2R and BRAF gene fragments generated from both PCA and LCR methods of assembly. Though we only tested two genes, incorporation of either additive could aid in the construction of most GC-rich sequences. Protocol manipulation of standard conditions is also unnecessary due to the isostabilizing properties of these additives. Even without the need for nucleotide conservation, this application saves a great deal of end-user time not having to re-design and codon optimize ODNs prior to synthesis. As such, the possibility of manually introducing sequence error is also limited; one mismatch, deletion or insertion could lead to a frame-shift or other gene lethality. Furthermore, DMSO and betaine are very inexpensive, easily obtainable and highly compatible with other biological agents, which make them ideal for any gene synthesis assay.

Materials and Methods

IGF2R and BRAF gene fragment designs

Sequences (5’→3’) for IGF2R (bases 32-548) and BRAF (bases 1-512) were taken from the National Center for Biotechnology Information database (ACCESSION: NM_000876 and NM_004333, respectively). They were then entered into Gene2Oligo (http://berry.engin.umich.edu/gene2oligo/index.html), which cut both constructs up into 40 bp fragments with 20 bp hybridizable overlap between the +/- strands [14]. Though this program has the option of calculating the optimum length of overlap given a target uniform Tm, no such parameters were defined for either construct. ODN Tm values were calculated with Gene2Oligo using the Nearest Neighbor model.

Synthesis of IGF2R and BRAF constructs

ODN synthesis of both genes was done in-house (Stanford Genome Technology Center) with a 3900 DNA synthesizer (Applied Biosystems) using 1000 A CPG columns (Biosearch Technologies) for a 50 n mole-scale synthesis. Cycle conditions were similar to the manufacturer’s recommended protocol, which included the following reagents: deblock (3% TCA/DCM) (Acic), acetonitrile, 0.02 M oxidizing solution, cap A/B, 0.1 M solutions of dA, dC, dG and dT (Proligo), and 0.25 M 5-Benzylthio-1H-tetrazole (Glen Research). Post-synthesis steps included ODN cleavage from the support followed by base-deprotection overnight at 55 ºC with ammonium hydroxide (28–30%) (J.T. Baker). After lyophilization, ODNs were resuspended and the optical density for each was measured at 260 nm using a Spectramax 384 Plus 96-well plate reader. All ODNs were then normalized to 100 µM in water and analyzed for purity using reverse-phase HPLC (Transgenomic WAVE system).

Assembly and PCR amplification

For PCA, unmodified +/- strands were pooled together (100 µM), where 1 µl was added to High Fidelity (HF) Advantage polymerase mix (Clontech) according to the manufacturer’s recommended protocol (DMSO and betaine were not included in this kit); samples were then run on a Veriti 96-well thermal cycler (Applied Biosystems) through two iterations using the following parameters: 94 ºC/3 min | 20* [94 ºC/15 sec | 55 ºC/30 sec | 68 ºC/60 sec], where 1 µl from the first reaction was transferred to the second step (PCR components were replenished to a 20 µl final volume). Five µl from the last step were then taken for PCR amplification.

For LCR, ODNs were pooled separately into +/- strands (100 µM). Each set was enzymatically 5’ phosphorylated by adding 3 µl DNA to 41 µl water, 5 µl 10X T4 DNA ligase buffer with ATP and 10 U T4 Polynucleotide Kinase (NEB). Samples were incubated at 37 ºC for 30 min then heat-inactivated at 60 ºC for 20 min. Twenty-five µl each of +/- strands were desalted using Micro Bio-Spin 6 Chromatography columns (Bio-Rad), then pooled together. Two µl (~12 pmols) of the phosphorylated product were added to 41 µl water, 5 µl Ampligase 10X Reaction Buffer and 2 µl (10 U) of Ampligase (Epigence). The ligation reactions were cycled at 21* [95 ºC/1 min | 70 ºC/4 min] then cooled to 4 ºC (-1ºC per cycle). Following assembly with both PCA and LCR, target product was then selected for through PCR amplification using the forward and reverse primers (5’→3’) TCCCCGGTCCGTTGCTACCCCTCGC | ACAGGAAAGCCAA-TGCTGCTCTGGA (IGF2R) [25] and CGCCTCCCTTCC-CCCTCCCCC | ACTTGGGGTTGCTCCGTGC (BRAF). PCR conditions using HF Advantage were as follows: 94 ºC/5 min | 25* [94 ºC/15 sec | 55 ºC/30 sec | 68 ºC/60 sec] 68 ºC/5 min. For gel analysis of final product, 2 µl 6x Orange Loading Dye (Fermentas) were added to 10 µl of each PCR sample. Five µl O’GeneRuler 1 kb DNA Ladder Plus (Fermentas) were used in all cases for band size comparison. Samples were electrophoresed through a 1.25% SeaKem LE Agarose (Lonza) gel at 80 V for 45 min, stained with ethidium bromide, then visualized using a standard UV imager at 302 nm.

Additives

DMSO (99.9%) and betaine (5 M) (Sigma-Aldrich) were added in varying concentrations to select PCA, LCR and PCR steps. Where required, water was replaced with the necessary amount of either additive to generate 1–10% DMSO or 0.5–2.5 M betaine per 20 µl reaction (or 50 µl for LCR assembly).

Results

IGF2R was chosen as a template to determine if the same additives, DMSO and betaine used to successfully amplify the fully formed gene fragment [25] could also be employed to aid in building and amplifying it from a pool of overlapping, single-stranded ODNs. The BRAF gene fragment was also synthesized de novo for quality comparison given the same conditions as IGF2R. Particular attention was paid to the 5’ region of IGF2R (81.5% GC-rich between bases 1 and 260), which includes the non-coding element [34]. The rest of the gene fragment from bases 261 to 517 averages 44.4% GC. With respect to the hybridization map generated from Gene2Oligo [14], Tm values for the 20 bp overlaps average 87.0 ºC (first 260 bases) with a maximum of 92.6 ºC. For the BRAF gene, the first 183 bases are the most GC abundant at 78.1% (83.7% average Tm) and 43.2% for bases 184-512.

To determine at what concentration either DMSO or betaine improved full-length product generation of IGF2R and BRAF in
**Discussion**

We have shown that DMSO and betaine greatly improve de novo synthesis of two GC-rich gene fragments, IGF2R and BRAF, without having to modify nucleotide composition. This is particularly important when studying non-coding elements in cancer gene research where base conservation may be critical to structural function in malignant cell types. And because no optimum conditions for de novo synthesis of GC-rich genes have been previously reported, we compared both conventional methods of assembly, PCA and LCR in the presence and absence of DMSO and betaine during the construction of IGF2R and BRAF. We discovered that either chemical agent introduced into the assembly steps alone (PCA and LCR), despite concentration, had no effect on target generation when amplified with the stock polymerase mix; only when DMSO and betaine were added to the amplification step did they show a marked improvement, especially in the case of LCR.

For the LCR samples (Fig. 2B and D), increased target yield and specificity of both IGF2R and BRAF genes compared with PCA (Fig. 1B and D), is due to the higher stringency of the ligation method; only those overlapping fragments that are perfectly matched are ligated together [35]. Thus, protruding hairpin structures cause misalignment of template and complement strands, which subsequently do not tether to form a stable duplex. Proof-reading Exonuclease I activity in commercial polymerase mixes (such as HF Advantage) further degrades all unpaired, single-stranded ODNs, thereby “cleaning up” the reactions [36].
In contrast, multiple bands generated in PCA (Fig. 1B and D) are also the consequence of non-specific priming of overlapping ODNs during assembly, and are unavoidable without cycle optimization [37]. For these reasons, the added stability of LCR product accounts for the increased target band specificity and high yield amplification without the need for either DMSO or betaine during assembly.

Furthermore, our results for de novo synthesis of IGF2R are comparable to those obtained by Frackman et al., who PCR amplified IGF2R from a vector using the same additives [25]. It is the assembly from a pool of single-stranded, GC-rich ODNs that makes construction of genes with non-coding elements so challenging. There are any number of ways neighboring guanines can interact to form secondary structure such as when a single strand folds onto itself (unimolecular), two separate strands interact (bimolecular), or where four different molecules join to form a quadrimeric G-complex [38]. However, with introduction of either DMSO or betaine during the amplification process, much of the secondary structure caused by G-G interaction is disrupted. It is because of their isostabilizing properties that they are able to minimize if not eliminate problems of low yield and aberrant band formation associated with polymerase arrest and sequence misannealing. This in turn provides the end-user with a larger amount of working construct to maximize the number of reactions per experiment; and though target product was formed with both processes, LCR gave the highest yield with little to no background.

Therefore, with the ligase-based method one avoids having to gel extract and purify samples prior to their use in downstream application.

While for this study we chose IGF2R and BRAF as representative genes with moderately high GC-rich content, there may be sequences of interest that contain more consecutive guanines per stretch (>90% GC). In this case, other chemical and biological additives are available that might prove more effective than either DMSO or betaine. These include formamide, glycerol, NP-40, Tween 20, trehalose, EcoSSB and 7-deaza-2’-deoxyguanosine 5’ triphosphate (d7GTP) [23,39–43]; d7GTP works differently than the other additives in that hydrogen bonding between neighboring guanines is minimized due to nitrogen displacement from position 7 of the base ring to position 8. It has also been reported that Thermococcus litoralis exo- (Vent exo-) DNA polymerase has helped resolve GC-rich sequences better than the Thermus aquaticus (Taq) DNA polymerase variety used in this study [44]. Because DMSO and betaine showed substantial improvement in amplification of the IGF2R and BRAF gene fragments, we limited our testing to these particular chemical agents. If on the other hand, either had proven inefficient, further exploration into any one of the aforementioned chemical agents/additives would have been warranted. Moreover, the PCA, LCR and PCR methods applied in this work have not been modified in any way to account for increased structure $T_m$. We wanted to see how each process would be affected by the addition of either DMSO or betaine alone. It is likely our results might have varied if we had altered any of the cycle parameters, especially those for PCA; increasing the annealing temperatures, for example, could have improved target yield and specificity by generating less spurious PCR product.

Conclusion

Secondary structure formation and mispriming in de novo synthesis of GC-rich constructs greatly inhibits our flexibility to investigate important genes and their non-coding elements, especially in cancer research. Because no application currently exists that defines a method for maintaining nucleotide composition for structural function in gene synthesis, we explored the use of cheap and readily available chemical additives, DMSO and betaine to aid in the production of GC-rich constructs. For this study, we chose two GC-rich gene fragments implicated in tumorigenesis, IGF2R and BRAF to determine the effectiveness of either DMSO or betaine in disrupting secondary structure formation and minimizing high $T_m$ ODN overlaps. While these additives had no benefit during the assembly steps of either process.
(PCA and LCR), we have shown they substantially improve gene target-specific amplification and yield when added to the PCR step alone. And because of their compatibility with routine gene synthesis constituents, DMSO and betaine may be well suited in aiding de novo synthesis of a wide range of GC-rich gene constructs and their non-coding regions. Furthermore, we demonstrated that LCR is the preferred method of assembly, yielding the highest amount of target product with the cleanest background. This in turn gives the end-user plenty of working material without the need for expensive and time-consuming sample extraction and purification between applications.

Acknowledgments

We would like to thank Drs. Kara Juneau, Maureen Hillemeyer, Henrik Persson and Mr. Keith Anderson for their generous comments and suggestions regarding this manuscript.

Author Contributions

Conceived and designed the experiments: MJ. Performed the experiments: MF. Analyzed the data: MJ. Contributed reagents/materials/analysis tools: MJ RWD. Wrote the paper: MJ. Principal investigator: RWD.

References

1. Khurana HG (1979) Total synthesis of a gene. Science (New York, N Y) 203: 645–652.
2. Hakkaart GA, Dassa EP, Jacobs HT, Rustin P (2006) Allotopic expression of a mitochondrial alternative oxidase confers cytochrome resistance to human cell respiration. EMBO reports 7: 341–345.
3. D’Souza GG, Boldilatti SV, Weisig V (2007) Gene therapy of the other genome: the challenges of treating mitochondrial DNA defects. Pharm Res 24: 228–238.
4. Stahlner P, Beier M, Gao X, Hoheisel JD (2006) Another side of genetics: synthetic biology as a means for the exploitation of whole-genome sequence information. Journal of biotechnology 124: 206–212.
5. Bernre SA, Sizomur AM (2005) Synthetic biology. Nature reviews Genetics 6: 533–543.
6. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, et al. (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science (New York, N Y) 319: 1215–1220.
7. Smith HO, Hutchison CA, Pfannkoch C, Venter JC (2005) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proceedings of the National Academy of Sciences of the United States of America 100: 15440–15445.
8. Geller M, Lipsett MN, Davies DR (1962) Helix formation by guanyle acid. Proc Natl Acad Sci U S A 48: 2013–2018.
9. Sahdev S, Saini S, Tiwari P, Saxena S, Singh Saini K (2007) Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. Mol Cell Probes 21: 303–307.
10. McDowell DG, Burns NA, Parkes HC (1998) Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. Nucleic Acids Res 26: 3340–3347.
11. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415.
12. Jayaraj S, Reid R, Santi DV (2005) GeMS: an advanced software package for designing synthetic genes. Nucleic Acids Res 33: e31.
13. Hoover DM, Lukbiski J (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res 30: e43.
14. Rouillard JM, Lee W, Traun G, Zhao Q, et al. (2004) Gene2Oligo: oligonucleotide design for in vitro gene synthesis. Nucleic Acids Res 32: W176–W180.
15. Usdin K, Woodford KJ (1995) CGG repeats associated with DNA instability and advanced colorectal cancer patients. International journal of cancer Journal of molecular microbiology and biotechnology 3: 123–126.
16. Chou Q (1992) Minimizing deletion mutagenesis artifact during Taq DNA polymerase PCR by E. coli SSB. Nucleic acids research 20: 4371.
17. Simonsson T (2001) G-quadruplex DNA structures–variations on a theme. Biol Chem 382: 621–629.
18. Sarker G, Kapelner S, Sornson SS (1990) Formamidine can dramatically improve the specificity of PCR. Nucleic acids research 18: 7465.
19. Bachmann B, Luke W, Hunnam G (1990) Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucleic acids research 18: 1369.
20. Frey UH, Bachmann HS, Peters J, Siffert W (2000) PCR-amplification of GC-rich regions: ‘slowdown PCR’. Nature protocols 3: 1312–1317.
21. Weissman SM, Bakaran BN (2000) Amplification of Nucleic Acids. U.S. Patent 6114,150.
22. Richardson SM, Wheelan SJ, Yarrington RM, Boeke JD (2006) GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome research 16: 550–556.
23. Varadaraj K, Skinner DM (1994) Denaturants and cosolvents improve the specificity of PCR amplification of a G+C-rich DNA using genetically specific DNA polymerases. Gene 140: 1–5.
24. Winship PR (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic acids research 17: 1266.
25. Frackman S, Kobs G, Simpson D, Steorts D (1998) Betaine and DMSO: enhancing Agents for PCR. P romega Notes. 27 p.
26. Rees WA, Yager TD, Koert J, van Hiepp HL (1993) Betaine can eliminate the base pair composition dependence of DNA melting. Biochemistry 32: 137–144.
27. Kosimaa A, Evangelou K, Sideridou M, Kotzamanis G, Constantines C, et al. (2008) The 3′ UTR of EGFR-R2 and R2 variant is associated with increased tumor growth and advanced stages in non-small cell lung cancer. Cancer letters 259: 177–185.
28. Zervas AI, Pinpath W, Tu W, Cartosso V, Lam A, et al. (2003) Insulin-like growth factor II receptor gene-167 genotype increases the risk of oral squamous cell carcinoma in humans. Cancer research 63: 296–297.
29. Cohen Y, Xing M, Mambo E, Gao Z, Wu G, et al. (2003) BRAF mutation in papillary thyroid carcinoma. Journal of the National Cancer Institute 95: 625–627.
30. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, et al. (2002) Mutations of the BRAF gene in human cancer. Nature 417: 949–954.
31. Meyer P, Sergi G, Garbe C (2003) Polymorphisms of the BRAF gene predispose males to malignant melanoma. J Carcinog 2: 7.
32. Stemmer WP, Crameri A, Ha KD, Brennan TM, Heynek H (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxynucleotides. Gene 164: 49–53.
33. Luo LC, Yang FY, Yang WJ, Lo SH, Kao CF (1989) Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in Escherichia coli. Biochemical and biophysical research communications 214: 200–203.
34. Morgan DO, Edman JC, Standring DN, Fried VA, Smith MC, et al. (1987) Insulin-like growth factor II receptor as a multifunctional binding protein. Nature 329: 301–307.
35. Brody RS, Doherty KG, Zimmerman PD (1986) Procesivity and kinetics of the reaction of exonuclease I from Escherichia coli with polydeoxynucleotides. The Journal of biological chemistry 261: 7136–7143.
36. Don RH, Cox PT, Wainwright RJ, Baker K, Mattick JS (1991) ‘Touchdown’ PCR to circumvent spurious priming during gene amplification. Nucleic acids research 19: 4008.
37. Simonsson T (2001) G-quadruplex DNA structures–variations on a theme. Biol Chem 382: 621–629.
38. Spiess AN, Mueller N, Irvill R (2004) Trehalose is a potent PCR enhancer: lowering of DNA melting temperature and thermal stabilization of taq polymerase by the disaccharide trehalose. Clinical Chemistry 50: 1256–1259.
39. Chou Q (1990) Minimizing deletion mutagenesis artifact during Taq DNA polymerase PCR by E. coli SSB. Nucleic acids research 20: 4371.
40. McCoulogue L, Brow MA, Innis MA (1988) Structure-independent DNA polymerase activity of thermostable ligase. Clinical Chemistry 50: 1256–1259.
41. Kotsinas A, Evangelou K, Sideridou M, Kotzamanis G, Constantinides C, et al. (2003) BRAF mutation in papillary thyroid carcinoma. Journal of the National Cancer Institute 95: 625–627.
42. Chou Q (1989) Minimizing deletion mutagenesis artifact during Taq DNA polymerase PCR by E. coli SSB. Nucleic acids research 20: 4371.