Title

GeMS: an advanced software package for designing synthetic genes

Permalink

https://escholarship.org/uc/item/61x9k5vb

Journal

Nucleic Acids Research, 33(9)

ISSN

0305-1048

Authors

Jayaraj, Sebastian
Reid, Ralph
Santi, Daniel V

Publication Date

2005-05-16

DOI

10.1093/nar/gki614

Peer reviewed

GeMS: an advanced software package for designing synthetic genes

Sebastian Jayaraj, Ralph Reid and Daniel V. Santi*

KOSAN Biosciences Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

ABSTRACT

A user-friendly, advanced software package for gene design is described. The software comprises an integrated suite of programs—also provided as stand-alone tools—that automatically performs the following tasks in gene design: restriction site prediction, codon optimization for any expression host, restriction site inclusion and exclusion, separation of long sequences into synthesizable fragments, T_m and stem–loop determinations, optimal oligonucleotide component design and design verification/error-checking. The output is a complete design report and a list of optimized oligonucleotides to be prepared for subsequent gene synthesis. The user interface accommodates both inexperienced and experienced users. For inexperienced users, explanatory notes are provided such that detailed instructions are not necessary; for experienced users, a streamlined interface is provided without such notes. The software has been extensively tested in the design and successful synthesis of over 400 kb of genes, many of which exceeded 5 kb in length.

INTRODUCTION

To facilitate the emerging field of synthetic biology, the ability to efficiently synthesize long, accurate DNA sequences is becoming increasingly important. Rather than accept the increasing error frequency inevitable in single-step synthesis of long DNA sequences, we developed automated technology to synthesize large numbers of perfect 500–800 bp segments, termed synthons, from PCR assembly of short oligonucleotides together with efficient methods for cloning and connecting them into larger sequences of ∼5000 bases (1); these are then joined by conventional manual methods into sequences of ever-increasing size. Thus far, we reported the synthesis of a contiguous 32 kb gene cluster, but there is no theoretical upper size limit for preparing DNA by this approach.

Concomitant with developing gene-synthesis technology, we required computational software to optimize codons for expression, incorporate or eliminate restriction sites to allow modular exchange of segments, design the oligonucleotide (oligo) components and perform other operations integral to the synthetic procedure. Although several gene design programs were either available (2,3) or described (4,5), they were not appropriate for very long genes and did not contain all features required for our purpose. We, therefore, developed a new software package that automates codon optimization, restriction site placement, separation into synthons and design of oligo components (Figures 1 and 2). In the process, we expanded the flexibility of the software by modularizing and generalizing algorithms to be multi-purpose and included numerous user-defined parameters to control features of the software. The result described here is a user-friendly, advanced software package, called Gene Morphing System (GeMS), which has broad utility in the design of synthetic genes to be made by PCR assembly of short oligos.

PROGRAMMING

GeMS was written using the freely available open-source programming language Python version 2.3. (www.python.org). The web-based version uses the open-source Apache HTTP server version 2.0.49 (www.apache.org). DNA annealing temperature calculations are performed using the dan program (6,7) and stem–loops are detected using the palindrome program available in the EMBOSS suite version 2.10.0 (8). Codon usage tables are derived from the Kazusa Codon Usage Database (www.kazusa.or.jp/codon/). The software uses the codon-count program (http://www.kazusa.or.jp/codon/countcodon.html) to report codon distribution. The synthetic genes described in this article were designed on a 1.0 GHz Intel Pentium III PC with 512 MB of memory, running the RedHat 8.0 Linux operating system.

RESULTS

The software comprises a suite of integrated tools that provides an advanced program for gene design

---

*To whom correspondence should be addressed. Tel: +1 510 731 5204; Fax: +1 510 732 8401; Email: santi@kosan.com
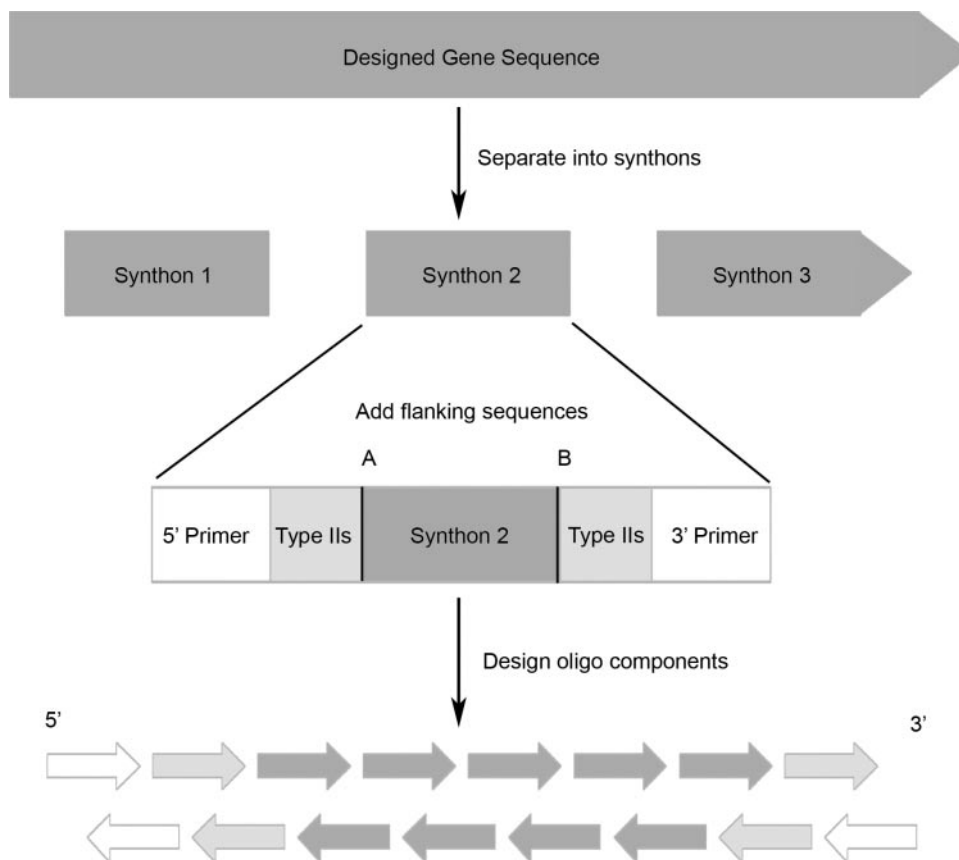
Figure 1. Overview of the synthon design process showing the placement of Type IIs restriction sites and flanking primers. A and B indicate sites of cleavage at synthon edges by Type IIs enzymes to form cohesive ends for seamless ligations.

(http://software.kosan.com/GeMS). The component tools are also accessible for individual use, as are codon tables and restriction site lists (Table 1). The software is divided in four sections: (i) sequence inputs and restriction site design (ii) parameter definition for oligo design (iii) program execution and (iv) report generation. The ensuing description follows the order of operation of GeMS (Figure 2), coupled with commentary.

The Home page provides a brief overview of the software features, along with navigation buttons leading to individual tools and a support page (Supplementary Figures 1S–8S). The user first selects either a guided or advanced interface. The guided interface is for inexperienced users and contains detailed information on how to navigate the software; the advanced interface omits explanations and allows for rapid data entry. In either, if the user is willing to accept GeMS default values, it is necessary to enter parameters in only six fields, each specified by a red asterisk.

Sequence inputs and restriction site design

Data input. Upon selection of an interface, the user is taken to the sequence input form. For gene design involving restriction site manipulations and codon changes, the input is a protein sequence, either provided directly or translated from an entered DNA sequence. The software will also accept a DNA sequence not to be modified and simply design the

oligo components needed for synthesis (see below; Figure 2). If the user chooses to place an NdeI site (CATATG) at the start codon, the software automatically places the His codon, CAT, 5′ to the ATG start codon. The user is also allowed to append DNA sequences to the 5′ and 3′ ends of the gene. After entries and choosing options, activation opens the appropriate window.

Restriction site inclusion and exclusion. For a codon-optimized gene, the user chooses whether to have restriction sites automatically spaced along the gene or to manually place them at specified locations. With average-sized genes, unique restriction sites can be automatically spaced to allow cassette mutagenesis with synthetic oligos. We recommend that no more than 30 unique sites be spaced along a gene, since purging a large number of undesirable sites from a randomly optimized sequence is computationally intensive. GeMS provides a default of 40 ± 15 bp separation of sites as a convenient size for oligo synthesis of cassettes for genes of ~1 kb.

The protein sequence is back-translated to all possible codons and from these, GeMS reports all possible six- and eight-base palindromic restriction sites in the GeMS library or entered by the user that are acceptable for inclusion in the gene. In the automatic mode of assigning sites, the software selects sites spaced at the designated distances, if possible. If these are insufficient or unacceptable, the user may add sites from the available list or remove sites from the automatically

GeMS Flowchart
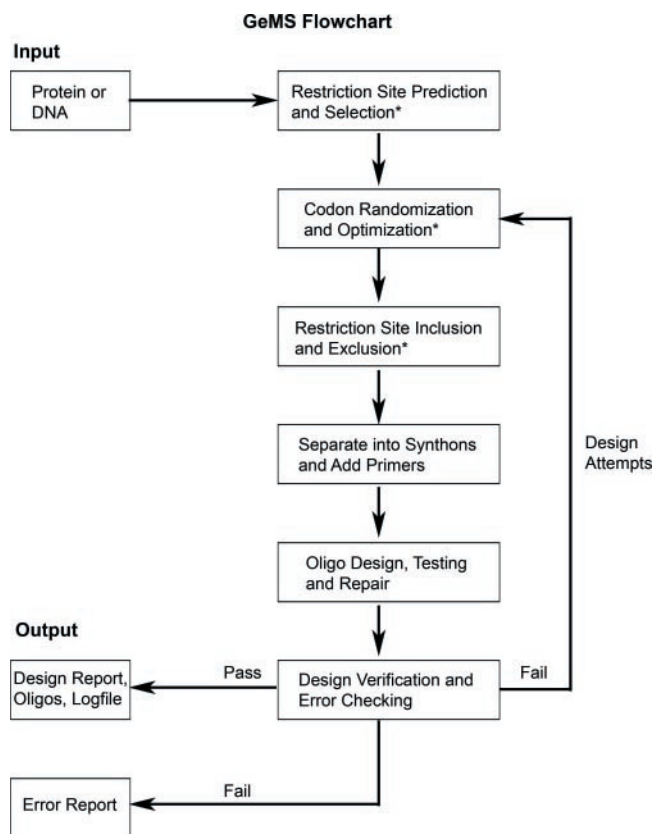
Figure 2. Sequence of operations during the course of automated gene design; a failure triggers an attempt to redesign the gene with a randomly chosen set of codons. Asterisks indicate steps that are skipped when designing native DNA sequences.

Table 1. Stand-alone tools available in the GeMS software package

GeMS software tool	Function
DNA–protein	
Translation	Translates a DNA sequence into a protein sequence
Reverse complement	Reports the reverse complement of a DNA sequence
Sequence alignment	Aligns protein or DNA sequences based on local similarities
DNA stem–loops	Reports inverted repeats in a DNA sequence
Assign vectors for LBS	Assigns vectors to synthons for ligation-by-selection
Restriction sites	
Restriction sites in DNA	Reports restriction sites in a DNA sequence
Predict all possible sites from protein	Reports all possible restriction sites
Automatic spacing of unique sites	Spaces unique sites over a designed gene sequence
Exclude sites	Excludes restriction sites while preserving protein sequence
Codon usage	
Custom codon table	Creates a custom codon preference table
Hybrid codon table	Creates a custom codon preference table for multiple organisms
Codon optimize	Optimizes codons in a DNA sequence for a desired organism
Oligo design	
T_m	Calculates melting temperature of complementary DNA sequences
Design 40mers	Designs 40mer oligo components of a DNA sequence
Design 40mers + primers	Designs 40mer oligo components of a DNA sequence with primers

selected sites. The software prevents selected sites in the final list from being used elsewhere in the gene. A map showing all possible sites in the sequence is available for viewing the spacing of sites.

Because it is often necessary to reserve certain restriction sites for cloning or use in the vector, it is also desirable to be able to exclude sites from the gene sequence. For this, the restriction site library is again displayed, and the user may choose those to exclude from the designed sequence; as before, a site absent from the library may be defined and excluded from the gene sequence.

We note that many programs, such as GCG, MacVector (from Accelrys) and EMBOSS, provide separate tools to back-translate protein to DNA and to map restriction sites in the DNA. In GCG, the backtranslate program can translate protein to an ambiguous DNA sequence, which can be used as input for the tool called map to produce a list of possible restriction sites. Similar tools are also available in MacVector. With this procedure, amino acids having subsets of codons in different boxes of the codon table (i.e. Leu, Arg and Ser) are assigned too many codons, and non-existent or 'phantom' restriction sites may be predicted that are incompatible with the protein sequence. Thus, the standard back-translations are Leu → YTN, Arg → MGN and Ser → WSN, with Y standing for C or T, and N for A, G, C or T. The DNA sequences YTN or MGN each refer to eight codons, only six of which code for Leu or Arg, respectively. Similarly, WSN refers to 16 codons,

only 6 of which code for Ser. By back-translating these amino acids into codons rather than nucleotide sequences, GeMS avoids such 'phantom' restriction sites.

It is not possible to insert a unique restriction site into or exclude a site from all sequences: there may exist restriction sites that cannot be made to be unique with a desired protein sequence. For example, in one of the sequences that can encode the tripeptide AlaArgAla, a BssHII site, GCGCGC, is present at two different positions, beginning at base 1 and base 3 of the first codon. Neither of these can be present as a unique site; either both must be inserted or both must be removed. As a second example, a request to avoid the sequence TGGATG fails if its encoded dipeptide TrpMet appears in the protein sequence.

Restriction site design proposal. The software next presents a proposed plan for the restriction sites to be included in the designed gene. The user is informed if intended unique sites are unique and displays a list of the sites that have been included or excluded, as well as a restriction site map of the final design. The user may either return to the previous window and make modifications or approve the design.

Parameter definition for oligonucleotide design

Codon harmonization. After the user approves the restriction site design, a window is provided for entry of other gene design and synthesis parameters. The user selects one of the several codon usage tables of organisms, or provides one, to be used in codon harmonization of the synthetic

gene. The codon preference for each amino acid is calculated as described by Withers-Martinez *et al.* (5) using the fractional preference for each codon within the set encoding each amino acid. The user may delete one or more codons, or provide a cut-off frequency that eliminates all codons below that value; the preferences of the remaining codons are then adjusted proportionately such that the sum of codon preferences for each amino acid again equals 1.0.

The software provides a stand-alone tool for creating codon preference tables weighted for codon usage in multiple organisms, and several of the most useful hybrid tables are provided as selectable options. When using the tool to construct other such tables, the user deletes codons considered particularly undesirable for any of the organisms and the software readjusts the preferences in each table. The user then enters the weighting factor desired for codons of each organism and calculates the combined codon preference as follows:

$$P_{hybrid} = \sum_{i=1}^{n}(WF_i) \times P_i,$$

$$\text{where } \sum_{i=1}^{n}(WF_i) = 1, \quad \text{and} \quad WF_i \geqslant 0 \text{ for each } i,$$

where WF_i is the weighting factor and P_i is the codon preference for the individual organism. The software multiplies codons of each amino acid by the provided factors, sums the weighted preferences for each codon and creates a multi-organism preference table to be used in the gene design.

Assembly design. If specified, the software will separate long coding sequences into smaller fragments or synthons to be synthesized (1), which then need to be ligated together (Figure 1). In one approach, the gene is separated into synthons by size or by number. If the user chooses this approach, the $3'$ and $5'$ edges of synthons to be joined will be automatically designed to overlap by 6 bp when cut to allow compatible cut sites produced by Type IIs enzymes placed outside of and adjacent to the synthon ends (Figure 1). The [BsaI] and [BbsI + G] sites are positioned adjacent to the synthon to result in cutting the 6 bp synthon edge sequences between positions 1/2 and 5/6, leaving 4 bp complementary overhangs that can be ligated to give a seamless connection. In the second approach, the gene is separated at unique restriction sites chosen by the user, and the software places the chosen sites at the appropriate $3'$ and $5'$ ends of the synthons to be ligated.

If the Type IIs sites are used to connect multiple synthons, we recommend using ligation-by-selection (1,9). Here, synthons are made in two specialized vectors that each possesses a unique pair of selectable markers. The vectors are cleaved and re-ligated to regenerate a vector containing the ligated synthons and a different unique pair of selectable markers. The 2-synthon products of two such ligations can be likewise connected to four, and so on, such that multiple ligations can be performed without isolating intermediate fragments. A stand-alone tool is provided that supports synthon cloning and assembly by ligation-by-selection. GeMS calculates the identifying number for each of the two vectors needed, assigns them to the appropriate synthon, and generates the plan for ligation steps in the form of a dendrogram (1). The dendrogram may also be used as an input to a robot to automate the ligation process.

Flanking sequences. The user enters sequences to be used to flank each synthon, which are not to be modified by the software. These are placed $5'$ of the Type IIs sites adjacent to the synthon and are used to construct PCR primer sites, cloning sites, etc. If Type IIs sites are not used, the sequences are directly fused to the synthon.

Oligonucleotide design. The user specifies, or accepts by default, a 'specificity threshold' (as described below) and the number of attempts the software will make to achieve that parameter. The software will create 40mer oligo components of synthons with 20 bp overlaps, with the exception of the $5'$ oligos of each strand, which can range from 20 to 40 bases. Then, each 20 base overlap is scanned against both strands; oligos that have exact $3'$ complements of the specified number of bases are analyzed for complementarity over the entire 20 bp. If any inappropriate 20mer overlap has greater than the user-designated percent complementarity, attempts will be made to correct the design; if unsuccessful, the entire synthon will be discarded and the codon randomization repeated.

T_m and stem–loop structure reports. The user may test the annealing temperature of each half-oligo using the nearest-neighbor model described by Breslauer (6,7). If an oligo falls outside a user-defined range, GeMS reports the oligos in the synthon that failed the T_m parameter for the synthon. The user may also choose to calculate stem–loop structures and report those sequences that fail to meet the criteria; however, the specificity threshold used in design of oligos effectively prevents most such structures. If all attempts fail to meet the specified T_m and stem–loop criteria, the user may choose to use one of the designs, reduce stringency of the failed parameters or increase the number of attempts at codon randomization.

Program execution

Codon harmonization. After choice of a codon table, a pool of codons is created for each amino acid that is equal to the number found in the protein, and that is in accord with their occurrence in the adjusted codon preference table. The software randomly shuffles the pool, randomly selects one of the designated codons and replaces the first natural codon of the open reading frame with it; the process is repeated sequentially to the C-terminus of the protein to complete optimization of the synthetic gene. The optimized gene thus contains codons in the exact proportions in which they are found in the adjusted codon preference table.

Restriction site placement. The software next eliminates all restriction sites and sequences entered in the user-specified removal list from the codon-optimized DNA sequence without changing the amino acid sequence. In the first pass, the program reports occurrences of all restriction sites, their positions and frames. In the second pass, the program moves to each location and attempts to eliminate the site by replacing one of the involved codons with an alternative. If the software encounters a single-codon amino acid, Met or Trp, an adjacent codon contained within the restriction site is modified; for example, with NdeI, the CAT of the hexanucleotide CATATG is replaced by an alternative His codon, leaving the ATG to encode Met. The codon preference is altered minimally by 'fixing' codons in this way. The user has the option of

checking the altered codon preference by using the codon-count program.

Because removal of one site can, by coincidence, create another, the program re-analyzes the purged sequence for undesired sites. If such sites cannot be eliminated, the program can enter long loops and this phase of the design process is aborted; the program discards the initially constructed DNA sequence, chooses an entirely independent set of codons and re-attempts the design. After a user-designated number of tries, the failures along with the offending restriction site and location are reported and the program terminates. The user can chose to discard the offending restriction site.

The program then inserts the previously designated restriction sites at the selected locations. For each site, the insertion algorithm replaces the bases at the appropriate position with the designated recognition sequence. If a codon for Leu, Arg or Ser—codons that fall in two different blocks of the codon table—is interrupted by a site that changes the encoded amino acid, the program attempts to alter the remainder of the codon to preserve the amino acid. The program then attempts to detect and correct undesired restriction sites that may have been coincidentally introduced by the insertion procedure. Some corrections may not be possible, as in the aforementioned incompatibility of a unique BssHII site with AlaArgAla, or may produce a new undesired site. If a correction attempt fails, the software reports the error and its location, chooses a new DNA sequence, re-initiates and completes the design process. In case of failure in repeated automatic design attempts, we encourage the user to remove the restriction site that fails or reduce the number of restriction sites in the list and try again. Also, with longer sequences a large number of requested unique or undesirable sites may cause repeated design failures that can be addressed by the reduction of the sizes of the unique and/or desired sites lists.

Separation into synthons. The software proceeds to separate the gene into synthons of specified size or number, or at the user-designated restriction sites.

Oligonucleotide design. With the exception of the 5′ oligos, the upper and lower strand sequences of the synthons and flanking sequences are used to design 40 nt oligos having 20 bp overlaps with each of two oligos of the opposite strand. The 5′ oligos are designed to be 20–40 bp to allow adjustment for synthon length and are adjusted to keep them of similar size. If connected, the oligos of the upper and lower strands will usually exhibit short overhangs at each end. For example, if a synthon and flanking sequences are 465 bp, the upper strand might be composed of one 23mer and eleven 40mers, and the lower strand composed of a 5′ 22mer and eleven 40mers; if connected, there would be a 3 nt overhang at one end and a 2 nt overhang at the other.

The software then performs a specificity test to avoid hybridization of any oligo component of the gene with inappropriate oligo components for PCR assembly. In the first round of oligo extension each 40mer, the 3′ 20mers of the 40mer oligos are the productive primers in producing 60mers. In the second round of extension, the productive primers are the 3′ 20mers of the 60mers, equivalent to the 5′ 20mers of the 40mers. We, therefore, test all 20mers of the 40mers both 3′ and 5′ for primer specificity. Here, each 40 bp oligo of the

upper strand is split into two 20 bp half-oligos. The half-oligos are screened against the length of the upper and lower synthon strands. If a half-oligo has an exact match of a user-defined length at the 3′ end with any sequence other than the intended one, and the half-oligo has user-defined unacceptable homology with that sequence, a failure in specificity is triggered; the default values of the program are 3 bp at the 3′ end and 70% minimum homology. In certain cases, such as highly repetitive amino acid sequences, passing the specificity test may require decreasing the specificity threshold parameters.

A single half-oligo failure in specificity prompts the software to add random base pairs, two at a time, at the edge of a synthon sequence to allow further attempts to pass the specificity test. The addition of 2 bp of randomly picked sequence at the 5′ end, followed by reapplication of the algorithm has the effect of moving all the inter-oligo junctions toward the 5′ end by 1 bp. The change in junctions creates different oligo sequences and boundaries, hence modifying the specificity of half-oligos. After rechecking for half-oligo specificity, the process is repeated until the specificity test is passed, or the shifting junctions reach the initial junctions. If, for example, the oligo length is 2L, where L is the half-oligo length of 20, this allows up to 20 tries at passing the quality control by using each codon-optimized DNA sequence constructed. If repair of the oligo specificity is not possible, entire sets of synthon components of a sequence are automatically recreated by using a new codon-optimized DNA sequence, and specificity failures reported for each such attempt. The number of attempts at making oligos that satisfy desired conditions is preset by the user.

T_m and stem–loops. If the user chose to test the annealing temperature of each half-oligo, and an oligo falls outside a user-defined range, GeMS halts execution and reports a failed T_m for the synthon with the number of oligos that failed. The user may abort or accept one of the constructs.

Sequence reassembly and verifications. After all the synthons of an entered sequence have passed the quality control tests, they are reassembled to generate the larger DNA sequence. The sequence is translated into protein and checked against the original amino acid sequence for integrity. A final screen for undesired sites is performed; if such sites are found, the user is alerted and the program attempts a new run at designing the sequence. Upon passing this final screen, the sequence and corresponding oligo components to be used for synthesis are stored.

Oligonucleotide components. The software assigns the collection of oligos needed for all of the synthons composing the desired sequence to a series of 96-well plates. This information is used to generate order sheets for gene synthesis and can serve as instructions for the robot to select and combine the appropriate oligo components of each synthon.

Gene design report

Upon activation, the program presents a summary report of the design. The program reports whether a successful design was achieved and, if so, the number of attempts made to achieve it. If unsuccessful, we recommend more design attempts or changes in parameters. Then, details of the sequence design are presented, with hyperlinks that allow access to the final

codon distribution, T_m of overlaps between pairs of oligos, the designed DNA sequence and an alignment of the input versus final protein sequence to allow visual verification. There are two files that we recommend be opened and saved as permanent records. The first contains all details of the gene design; the second provides the oligo components in a spreadsheet that can be used to order them and as an input to a robot for automated gene synthesis. Finally, the report provides the sequence of each synthon with and without primers attached, the oligo components for each synthon and the complete sequence of the assembled gene.

DISCUSSION

Here, we describe a user-friendly, advanced software package for the design of synthetic genes. The software was developed to design synthetic genes to be made by the PCR assembly of short oligos, but it should be adaptable to ligation methods as well (10–12). The user provides a protein or DNA sequence and enters various parameters, and the software provides optimized oligos for the synthesis of the designed genes. The software contains most components found in other gene synthesis programs (2–5), as well as certain unique features.

Some of the differentiating features of GeMS are as follows. (i) GeMS offers automatic spacing or manual placement of unique or redundant restriction sites along the gene sequence. It will also prevent specified sites from occurring in the gene. For restriction site placement, most programs back-translate proteins into nucleotide sequences rather than into codons, and then search the back-translated sequence for restriction sites; as described in Results, this often generates phantom sites that change the sequence of the encoded protein. By back-translating amino acids into codons, GeMS avoids such phantom restriction sites. (ii) A tool is provided that allows creation of hybrid codon preference tables for optimizing gene expression in two or more organisms. (iii) If specified, the software will automatically separate long coding sequences into smaller fragments or synthons to be synthesized, which can be ligated together. The user can choose to add adjacent Type IIs sites to generate overlapping cohesive ends on adjacent synthons or to separate synthons at unique restriction sites for conventional ligations. In either case, the adjustments for design, such as creating overlapping sites on adjacent synthons, are made automatically. (iv) GeMS uses a unique algorithm to optimize oligo components of the synthons so that they do not misprime in PCR gene assembly. The combination of codon randomization with the specificity-checking step prevents the formation of most possible stem–loop structures, which we rarely or never observe in genes designed with GeMS. (v) GeMS calculates and reports, but does not adjust for, T_m values of component oligos. Although some workers have advocated this (2–4,13), we are unaware of objective data supporting it, and our failure in synthon assembly by PCR has thus far been negligible. Perhaps our algorithm for avoiding oligo mispriming overcomes possible advantages of iso-thermal T_m values.

To date, we have used GeMS to design over 400 kb of synthetic genes ranging from ∼0.5 to >6 kb in length. Our failure rate in synthesis has been negligible and we conclude that the software is suitable for the design of most genes of current interest.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Kodumal,S.J., Patel,K.G., Reid,R., Menzella,H.G., Welch,M. and Santi,D.V. (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl Acad. Sci. USA*, **101**, 15573–15578.
2. Hoover,D.M. and Lubkowski,J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30**, e43.
3. Rouillard,J.M., Lee,W., Truan,G., Gao,X., Zhou,X. and Gulari,E. (2004) Gene2Oligo: oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res.*, **32**, W176–W180.
4. Gao,X., Yo,P., Keith,A., Ragan,T.J. and Harris,T.K. (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.*, **31**, e143.
5. Withers-Martinez,C., Carpenter,E.P., Hackett,F., Ely,B., Sajid,M., Grainger,M. and Blackman,M.J. (1999) PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Eng.*, **12**, 1113–1120.
6. Baldino,F.Jr, Chesselet,M.F. and Lewis,M.E. (1989) High resolution *in situ* hybridization histochemistry. *Methods Enzymol.*, **168**, 761–777.
7. Breslauer,K.J., Frank,R., Blocker,H. and Marky,L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.
8. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
9. Kodumal,S.J. and Santi,D.V. (2004) DNA ligation by selection. *Biotechniques*, **37**, 34, 36, 38 passim.
10. Mandecki,W. and Bolling,T.J. (1988) FokI method of gene synthesis. *Gene*, **68**, 101–107.
11. Stemmer,W.P., Crameri,A., Ha,K.D., Brennan,T.M. and Heyneker,H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, **164**, 49–53.
12. Dillon,P.J. and Rosen,C.A. (1990) A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques*, **9**, 298, 300.
13. Xiong,A.S., Yao,Q.H., Peng,R.H., Li,X., Fan,H.Q., Cheng,Z.M. and Li,Y. (2004) A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. *Nucleic Acids Res.*, **32**, e98.