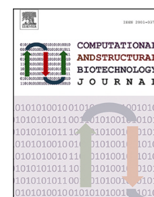
Computational and Structural Biotechnology Journal 29 (2025) 286–295

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj



Software/Web server article

CertPrime: a new oligonucleotide design tool for gene synthesis

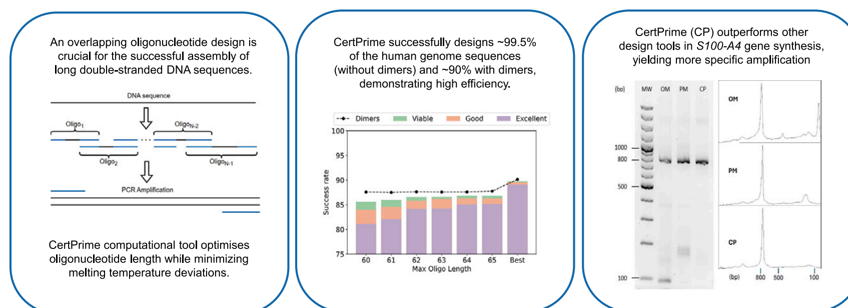
David Luna-Cerralbo [a,b,c], Ana Serrano [c], Irene Blasco-Machín [c], Fadi Hamdan [c], Juan Martínez-Oliván [c], Esther Broset [c,*], Pierpaolo Bruscolini [a,b,**]

[a] Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, c/ Pedro Cerbuna s/n, Zaragoza, 50009, España
[b] Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, c/ Mariano Esquillor s/n, Zaragoza, 50018, España
[c] Certest Pharma, Certest Biotec S.L, Polígono Industrial Río Gallego II, Calle J, 1, San Mateo de Gállego, Zaragoza, 50840, España

HIGHLIGHTS

- CertPrime improves oligonucleotide design by minimising melting-temperature deviations and spurious dimer formation.
- The new tool enhances gene-synthesis efficiency, allowing a precise control over oligonucleotide lengths and experimental parameters.
- CertPrime is a scalable tool capable of designing over 99 % of the human genome with optimised sequence assembly.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Gene synthesis
Gene assembly
Oligonucleotide design
Melting temperature

ABSTRACT

The design of oligonucleotides with uniform hybridisation temperatures is essential for successful gene synthesis. However, current computational tools for oligonucleotide design face significant limitations, including difficulties in processing long DNA sequences, poor adaptability to specific experimental conditions, limited control over oligonucleotide length, and challenges in minimising spurious dimer formation. To address these issues, we developed CertPrime (https://oligodesign.bifi.es), an innovative tool designed for scalable and efficient handling of long DNA sequences. CertPrime enables precise customisation of experimental parameters, provides flexibility to limit the maximum oligonucleotide length, and generates designs with reduced deviations in melting temperatures across overlapping regions compared to existing tools. We experimentally compared CertPrime designs with benchmark design methods for a complex DNA sequence and found that CertPrime designs led to more efficient gene assembly, significantly reducing the occurrence of non-specific bands. These results make CertPrime a powerful and versatile tool for oligonucleotide design in gene synthesis applications.

1. Introduction

Gene synthesis is a key technique in molecular biology, enabling the generation of long, custom-designed DNA sequences. This capability has driven significant progress in medicine, including the development of gene therapies [1], the production of recombinant proteins for targeted treatments [2,3], or the advancement of mRNA vaccines [4], which have

* Corresponding author.
** Corresponding author at: Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, c/ Pedro Cerbuna s/n, Zaragoza, 50009, España.
Email addresses: dluna@unizar.es (D. Luna-Cerralbo), aserrano@certest.es (A. Serrano), ibmachin@certest.es (I. Blasco-Machín), fhamdan@certest.es (F. Hamdan), jmartinez@certest.es (J. Martínez-Oliván), ebroset@certest.es (E. Broset), pier@unizar.es (P. Bruscolini).

https://doi.org/10.1016/j.csbj.2025.10.004
Received 17 March 2025; Received in revised form 1 October 2025; Accepted 1 October 2025
Available online 8 October 2025
2001-0370/© 2025 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

been pivotal in addressing global health challenges such as the COVID-19 pandemic.

The primary method for synthesising genes is based on Polymerase Chain Reaction (PCR) using overlapping oligonucleotides. These short, complementary oligonucleotides are designed with overlapping regions that anneal to adjacent sequences, providing a template for seamless assembly. Through iterative cycles of denaturation, annealing, and extension, DNA polymerase efficiently fills in and joins these overlapping fragments into a continuous DNA sequence. For this process to be effective, it is crucial that the melting temperatures T_m of all the overlapping complementary regions of adjacent oligonucleotides are roughly the same, and align with a target value (T_m^*) provided by the user, which is determined by the optimal operating temperature of the DNA polymerase used in the reaction [5]. This is a necessary (though not sufficient) condition to ensure that all overlapping oligonucleotides denature or anneal at once, yielding a uniform and efficient amplification of all parts of the gene. For this reason, all design tools require an estimation of the melting temperature of the overlapping regions, even if they use this information within different frameworks. The design of overlapping oligonucleotides can be categorised into two primary approaches: gapless PCR and gapped PCR. In the former, the design is such that oligos in the same strand are contiguous, to ensure high-fidelity synthesis. In the latter, neither strand is completely covered by the designed oligonucleotides, which present gaps between them; the complete gene sequence is recovered just because the overlapping oligos, in the opposite strand, span the gaps and provide the missing information. Gapped PCR provides greater flexibility and cost efficiency for specific applications.

Several computational tools and algorithms have been developed, implementing both design strategies. For gapless PCR, tools like TmPrime [6] and DNAWorks [7] are widely used, while Gene2Oligo [8], GeneDesign [9], Assembly PCR OligoMaker [10], Primerize [11], Integrated [12] and DeepFirstSearch [13] are applied for gapped PCR. However, these tools face significant challenges such as difficulties in processing long DNA sequences, poor adaptability to specific experimental conditions, limited control over oligonucleotide length, and challenges in minimising spurious dimer formation.

Regarding difficulties in handling long DNA sequences, tools such as Primerize and Integrated either lack the capability to process extended sequences or experience substantial increases in computational time as sequence length grows, reducing their practicality for large-scale gene synthesis projects. To overcome these difficulties, several experimental methods have been further developed to enable the synthesis of longer DNA sequences through PCR. These include techniques such as thermodynamically balanced inside-out (TBIO) [14,15], dual asymmetrical PCR (DA-PCR) [16], overlap extension PCR (O-PCR) [17,18], PCR-based two-step DNA synthesis (PTDS) [19,20] or PCR-based accurate synthesis (PAS) [21]. While these approaches primarily rely on modifying the PCR experimental process, often by adding additional steps, they do not address the computational challenges associated with oligonucleotide design, leaving a critical gap in scalability and precision for complex DNA assembly projects.

Another critical consideration in oligonucleotide design is the need to constrain the maximum length of oligos. Long oligonucleotides are challenging to synthesise chemically, often resulting in low yields and a higher probability of errors [22]. Tools like Primerize and PCR OligoMaker effectively address this requirement by allowing users to set maximum length constraints. In contrast, algorithms such as DeepFirstSearch and Integrated do not offer this flexibility and often default to designing shorter oligonucleotides, possibly to avoid the challenges associated with longer sequences. However, this bias results in a higher number of oligonucleotides being required to synthesise the same DNA sequence, as compared to tools like Primerize or PCR OligoMaker. While this approach does not compromise design quality, handling a larger number of oligonucleotides is less practical in laboratory workflows, increasing complexity and potentially impacting overall efficiency.

Experimental conditions, including DNA concentration, divalent ion levels and the concentration of dNTPs (deoxyribonucleotide triphosphate), can significantly influence oligonucleotide hybridisation temperatures and overall PCR performance [23]. Although PCR OligoMaker and Primerize offer limited options for adjusting DNA and monovalent ion concentrations, and others allow modifications to dNTP and divalent ion levels, these features are often insufficient to comprehensively accommodate the diverse requirements of different experimental setups. This mismatch between design parameters and actual laboratory conditions can lead to discrepancies in melting temperatures and hybridisation efficiency, often resulting in failed PCR assemblies. Consequently, users are frequently left to rely on trial-and-error adjustments, increasing experimental inefficiency and compromising reproducibility.

The formation of spurious dimers is another critical consideration in oligonucleotide design. Homodimer (self-interacting oligonucleotides) and heterodimer formation (unintended interactions between oligonucleotides from different regions of the sequence) often lead to non-specific DNA amplification, reducing the overall yield and accuracy of the synthesis process. While algorithms such as Integrated and DeepFirstSearch incorporate mechanisms to effectively prevent dimer formation, PCR OligoMaker lacks any functionality to address this issue. Similarly, Primerize does not actively prevent dimer formation but offers warnings about their potential presence and identifies the specific oligonucleotides affected, allowing users to manually adjust the design.

To address the limitations of existing design algorithms, we have developed CertPrime, a novel tool designed to overcome these challenges. CertPrime builds upon the principles of the DeepFirstSearch algorithm but revisits and reformulates each step to enhance performance. CertPrime exhibits excellent computational scalability, enabling it to handle long DNA sequences efficiently. It allows precise customisation of experimental parameters, including the concentrations of divalent and monovalent ions, DNA, and dNTPs, to align closely with laboratory conditions. Additionally, it enables users to set the maximum length for the oligonucleotides, offering flexibility to adapt to the constraints of chemical synthesis.

Notably, CertPrime achieves designs with reduced deviations in the melting temperatures of overlapping regions outperforming PCR OligoMaker and Primerize tools, while maintaining a similar number of oligonucleotides required for synthesis. When tested on the design of a complex DNA sequence with high GC% variations, CertPrime produced an efficient experimental gene assembly, with negligible occurrence of non-specific bands, and characterised by a better quality than the designs by PCR OligoMaker and Primerize. These features make CertPrime a robust and versatile tool for improving the precision, efficiency, and reproducibility of oligonucleotide-based gene synthesis workflows.

2. Materials and methods

2.1. Estimation of the melting temperatures

The melting (or denaturation) temperature T_m of the overlap region of two interacting oligonucleotides, is the fundamental parameter to determine if a design fulfills the desired requirements. T_m is the temperature at which half of the double-stranded structures denature, causing the strands to separate, and it depends non-trivially on the length and on the sequence of the strands, as well as on the characteristics of the solvent. Therefore, accurate estimations of T_m are quite challenging, due to the polymeric nature of the system and the numerous degrees of freedom involved. The challenge is even harder in the case of non-complementary interacting strands, which one should consider to prevent the formation of spurious dimers with a T_m in the PCR operating range. Finally, the T_m values also depend on the properties of the solution, as the presence of ions or other complex molecules can alter the hybridisation equilibrium and shift the melting temperature. To estimate T_m, most design tools use

the nearest-neighbour (NN) thermodynamic model [24,25] which offers a reliable approximation at a reasonable computational cost, when mismatches are absent or isolated along the sequence. This model calculates T_m as the temperature at which the free-energy difference between hybridised and free oligonucleotides becomes zero. The denaturation free energy is derived from the coupling energy of paired nucleotides, the stacking energy of neighbouring pairs and their entropies.

However, when two aligned oligonucleotides present two or more consecutive mismatches, conventional NN models lack the thermodynamic parameters required to accurately predict their effects. Thus, NN models are perfectly fit to calculate the T_m of complementary oligonucleotides, but are unsuitable for calculating the T_m of non-complementary dimers. To overcome this limitation, we took advantage of the temperature-estimation features of the Primer3 software, which implements and enhances the NN model's capabilities, providing an estimate of the melting temperatures for generic dimers. Primer3 is still based on the NN model, but unlike other tools, it can handle candidate complementary strands of different lengths, since it performs a sequence-alignment, seeking the best pairing, before estimating the T_m. In case of multiple consecutive mismatches, Primer3 calculates and outputs the T_m of the most stable paired subsequence. Although the predicted T_m may differ significantly from the experimental values, and more sophisticated interpolation schemes could be implemented to improve those predictions, these characteristics of Primer3 prompt for the detection of possible spurious homodimer and heterodimer interactions, since, in practice, the size of the overlap regions limits the magnitude of the estimation error.

Additionally, Primer3 allows users to specify the values of experimental parameters, such as the concentrations of monovalent and divalent ions, the concentration of dNTPs, as well as the DNA concentration. This flexibility enables a more accurate reproduction of experimental conditions, thereby improving the reliability of oligonucleotide design and minimising the risk of unintended interactions during synthesis.

2.2. Computational protocol for oligonucleotide design

Our method builds upon and enhances the Integrated strategy [12] that consists of four steps. In the first step, the DNA is divided into fragments with similar melting temperatures (of the dsDNA), all close to the target temperature T_m^* specified by the user. Notice that it is sufficient to consider a single strand of DNA, since the complementary sequence is completely determined. The above partition is done using a greedy algorithm that selects the (locally) best solution, starting from one end of the DNA and proceeding towards the other. The fragments thereby obtained represent the candidate regions from which the overlap regions will be extracted.

The second step involves an iterative process, analogous to a zero-temperature Monte Carlo algorithm. During this phase, the boundaries between segments are adjusted by shifting them by 0–4 nucleotides. These adjustments aim to reduce fluctuations in the T_m values of the segments, because the greedy algorithm does not guarantee a globally optimal outcome. Subsequently, between 0 and 5 nucleotides are removed from both the 5′ and 3′ ends of each segment, resulting in 36 candidate overlap regions for each segment. Finally, a dynamic programming algorithm is employed to select the optimal overlap region candidate for each segment. This approach systematically explores all possible combinations of the trimmed ends, identifying the overlap regions that minimise the melting-temperature deviations and enhance the overall efficiency of the assembly process.

This process is efficient but has several drawbacks. Specifically:

- The greedy algorithm employed in the first step yields a locally optimal outcome, but not a globally optimal one. When approaching the end of the strand, there is little control over the T_m of the last segment, which is often significantly lower than the others, potentially causing difficulties in the assembly process.
- The subsequent iterative algorithm, which redistributes the positions of the ends of the fragments, only accepts changes that improve the overall homogeneity of the T_m. However, all these temperatures, except for that of the last fragment, have already been adjusted to a local optimum in the previous stage. As a result, the initial iterations only shift the boundary between the last two segments; then, the other boundaries are moved to try to redistribute the residual mismatch between the actual and the target temperatures. This results in the melting temperature of the fragments being overall slightly lower than the target temperature specified by the user.
- The dynamic algorithm imposes no constraints on oligonucleotides length, prioritising segments with T_m closest to the target value. However, since the iterative algorithm produces slightly lower T_m than the desired value and base removal at the third step of the algorithm further decreases the T_m's, the base removal is actually discouraged, which results in almost gapless designs.

To cope with these issues, we began by modifying the initial greedy algorithm, to yield fragments with higher melting temperatures, by setting the initial target temperature to a $T_m' > T_m^*$, depending on the fraction of GC nucleotides in the strand. In the new procedure we also check, and enforce, that the segments' lengths fall in the range $[l_{min}^{ov}, l_{max}^{ov}]$ between the minimum and maximum overlap lengths, specified by the user. Since the T_m of a pair of hybridised strands is monotonically increasing with their length, the temperature T_m' is always set higher than the target T_m^* provided by the user. This ensures that shortening the regions S_i will bring their melting temperature $T_{m,i}$ closer to the desired target (Fig. 1).

Furthermore, we modified the greedy algorithm to prevent edge effects in the final fragments. Specifically, when we reach a position located between three and four l_{max}^{ov} from the end of the sequence, we determine the last three or four segments altogether. The exact number of segments depends on the fragments already set, and is determined by the need to have an odd total number of fragments (and hence, of overlap regions), ensuring that the final number of oligonucleotides is even. This constraint is related to technical reasons, to cope with the experimental setup and reconstruct the dsDNA from the oligo-design in a single PCR run, using an excess concentration of the terminal oligonucleotides. An even number of oligonucleotides in the design ensures that the first and last ones are located on opposite strands. This arrangement allows the entire sequence to be duplicated effectively.

In the second step of the Integrated approach, the iterative process is replaced by a Simulated Annealing (SA) method to displace the fragment ends, and obtain a more uniform distribution of melting temperatures around the target value T_m'. Additionally, throughout the SA process, we ensure that the lengths of the segments remain within the specified bounds, l_{min}^{ov} and l_{max}^{ov}. At this point, the outcome is a gapless design.

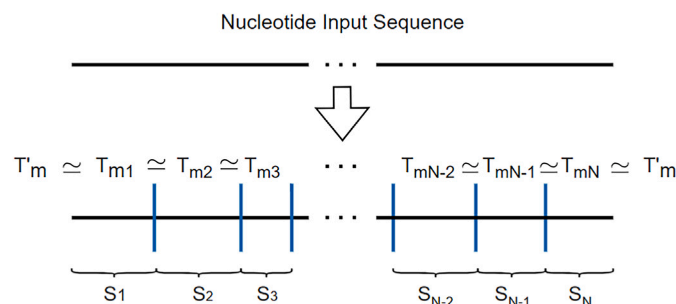


Fig. 1. Scheme of the first step in the design of certprime oligos. In this process, we start with an input sequence and obtain N segments, each with similar T_m, close to an increased target temperature T_m'. The number of segments N is forced to be odd, so that the final number of oligonucleotides, $N-1$, is even.

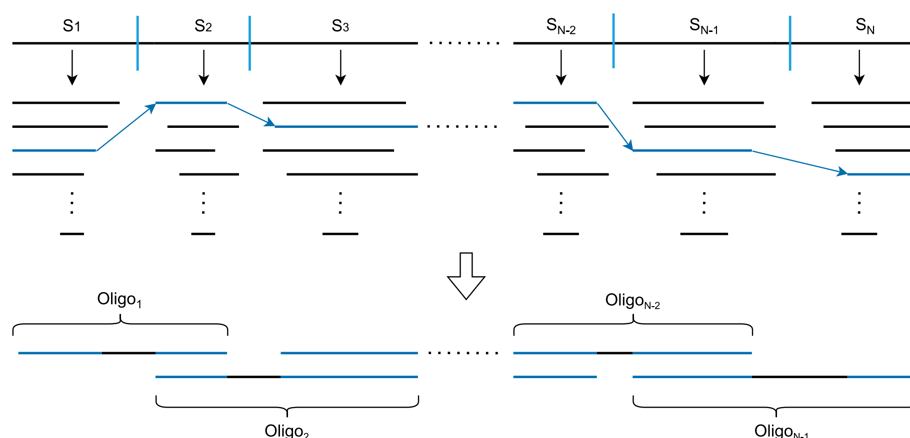


Fig. 2. Schematic representation of the final step performed by CertPrime. The process starts with the N fragments obtained with the gap-free partition described in Fig. 1. Candidate overlap regions (black segments) are obtained upon removing i nucleotides from the ends of the S_j fragments. The choice of these segments (blue lines) is performed to minimise Eq. (2) while keeping all the final oligonucleotide lengths below $M.O.L$.

In the following step, gaps are introduced to improve the overall quality of the design. As in the Integrated approach, we remove nucleotides from both the 5′ and 3′ ends of each segment, but without restricting the removal to just 5 nucleotides. Instead, we generate all possible fragments that satisfy the constraints on the overlap length (Fig. 2).

Finally, a selection of the candidate overlap regions among the latter fragments is made, which produces the final oligonucleotide design, according to the following criteria:

- The length of the selected fragments falls inside the user-specified range for overlap regions $[l^{ov}_{min}, l^{ov}_{max}]$, and the oligonucleotides spanned by any pair of neighbouring fragments are shorter than the maximal oligonucleotide length $M.O.L$, specified by the user.
- The overall deviation of their melting temperatures T^i_m from T^*_m is minimised.

The length criterion above is a hard constraint, stronger than the requirement on the melting temperatures, ensuring that the resulting lengths of the oligonucleotides and overlap regions always fall within the specified ranges, irrespective of temperature deviations. Based on these criteria, we apply a dynamic programming algorithm that finds the best design from the set of fragments generated in the previous step. This approach minimises the deviations in T_m while assigning an infinite deviation to solutions that do not meet the required oligo-length criterion. Once the design is complete, a few steps are still necessary to check the results and translate them into a proper output for the user:

- The strand of the even-numbered oligonucleotides is changed, returning their reverse complementary sequence.
- The presence of homodimers and heterodimers is checked, by calculating the melting temperatures for each possible interaction between pairs of oligonucleotides in the design.
- The overlap regions are explicitly extracted, and their melting temperature is determined.
- The %GC content for both the overlap regions and the oligos is calculated.

The input parameters that must be specified by the user, along with the gene sequence, are reported in Table 1, together with the values used in all the computational tests in this article.

2.3. Short-primers design

The PCR experimental protocol relies on two distinct primers (the sense and antisense one) that define the region to be amplified by annealing to the 3′ ends of the template and complementary strands. Thus, in order to amplify the whole gene and prevent partial or incomplete amplification, it is important that the primers anneal to the first and last oligonucleotides of the design with the highest probability. Instead of simply using an excess concentration of the terminal oligonucleotides, an alternative effective strategy involves designing ‘short’ oligo primers, with a melting temperatures of the oligo as a whole (T_m(S.O.)) close to a target value (T^*_m(S.O.)), and a given minimum length (m.O.L.(S.O.)). These short primers are complementary to part of the first and last oligonucleotides in the design, thereby enhancing both efficiency and specificity.

CertPrime allows the design of such short oligos, using the following protocol: it takes the first or last k bases, with k equal to the minimum oligo length for short oligonucleotides (m.O.L.(S.O.)), from either the initial or final oligo, depending on which S.O. is being designed. It then continues adding bases until the distance between the melting temperature of the short oligo (T_m(S.O.)) and the target value is minimised.

Table 1
Input parameters. List of user-specified input parameters for CertPrime. The reported values are those used in the computational tests. T^*_m is the target melting temperature for the overlap regions, M.O.L. is the maximum length for any oligonucleotide in the design, $[l^{ov}_{min}, l^{ov}_{max}]$ specifies the range of possible lengths for the overlap region. S.O. indicates short oligos, and m.O.L. (S.O.) is their minimum length. Brackets indicate concentrations; dNTPs: deoxyribonucleotide triphosphates. [Na^+], [Mg^{+2}], [DNA] and [dNTPs] are needed by Primer3 to calculate the melting temperature of each oligo sequence.

[Na^+]	50 mM	[dNTPs]	0.6 mM	$(l^{ov}_{min}, l^{ov}_{max})$	(16–30) bp
[Mg^{+2}]	1.5 mM	T^*_m	60 °C	T^*_m (S.O.)	60 °C
[DNA]	300 nM	M.O.L.	60–65 bp	m.O.L. (S.O.)	17 bp

2.4. Method-assessment protocol

2.4.1. Human genome database preparation

To perform an *in-silico* evaluation of the performance of the method, the ‘Homo Sapiens’ database was downloaded from Ensemblr [26] on 10 April 2024. This database contains 122,676 sequences, with lengths ranging from 10 to 10,976 nucleotides (Figure S.I. 1). Two subsets were extracted from the database: one for the synthesis-quality test (DS1) and another for the computational-scaling test (DS2). For the former, sequences were filtered to encompass lengths ranging from 150 to 1350 nts, resulting in a dataset of 73.424 sequences, to be used to assess how many human genes can be reliably designed by CertPrime, For the latter, a dataset of 960 sequences was built, sampling 10 sequences from

the human database for each length, in multiples of 50, ranging from length $L = 250$ to $L = 5000$ nucleotides. This dataset was used to test the dependence of the computational cost on sequence length.

2.4.2. Indicators of design quality

For each design, we calculate the following quality indicators:

- The temperature range:

$$\Delta T_m = T_m^{max} - T_m^{min} \tag{1}$$

- The distance from target:

$$\rho = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(T_m^* - T_m^i)^2} \tag{2}$$

- The temperature spread:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(\overline{T}_m - T_m^i)^2} \tag{3}$$

Here, n represents the number of oligonucleotides in the design; T_m^i denotes the melting temperature of the overlap (hybridisation) region between oligonucleotides i and $i+1$; $\overline{T}_m = (n-1)^{-1}\sum_{i=1}^{n-1} T_m^i$ is their average melting temperature, while $T_m^{\min}$, $T_m^{\max}$ correspond to the minimum and maximum melting temperatures, respectively, appearing in the oligo design. Additionally, T_m^* represents the desired (user-specified) melting temperature for the hybridisation regions.

The first indicator, Eq. (1) highlights designs presenting hybridisations with very different melting temperatures, which could affect the assembly process. Eq. (2) reports on the overall distance of the T_m values from the target melting temperature T_m^*, ensuring consistency of the design with the desired experimental conditions. Finally, the standard deviation Eq. (3) measures how closely the different T_m values are clustered together, providing an assessment, complementary to Eq. (1), of uniformity across all overlaps.

The combined use of multiple indicators, rather than relying solely on the standard deviation Eq. (3), as done in other studies (e.g., Integrated, DeepFirstSearch or TmPrime), allows us to identify and mitigate potential issues related to melting temperature variability, thereby enhancing the efficiency and reliability of the gene synthesis process.

Based on the values of these three indicators, we propose a classification of the designs as:

- "Excellent" ($\rho < 1.5$ and $\sigma < 1$ and $\Delta T_m < 2$)
- "Good" ($\rho < 2.5$ and $\sigma < 2$ and $\Delta T_m < 3$)
- "Viable" ($\rho < 3.5$ and $\sigma < 3$ and $\Delta T_m < 4$)

In this way, we can effectively evaluate and select oligonucleotide configurations that are more suitable for successful gene synthesis.

In addition to these three indicators, we also check the potential formation of dimers, to avoid spurious interactions in the design. This will serve as another criterion for evaluating oligonucleotide designs, ensuring that unintended homodimer and heterodimer formations are minimised, to enhance the efficiency and specificity of the gene synthesis process.

2.4.3. Synthesis of DNA sequences

A mix of desalted oligodeoxynucleotides (Integrated DNA Technologies) was prepared in TE buffer 1X, with a ratio of outer/inner oligos of 40. The first PCR reaction (gene assembly) was performed in a final volume of 25 μL, with 40 pmol of each outer oligo and 1 pmol of each inner oligo, 100 mM KCl, 3 % DMSO and 1X PrimeSTAR Max Premix (Takara). The reaction mixture was subjected to PCR amplification under the following conditions: initial denaturation at 98 °C for 30 s, followed by 20 cycles of denaturation at 98 °C for 5 s, annealing at 60 °C for 10 s, and extension at 72 °C for 30 s. A final extension step at 72 °C for 5 min was performed after the last cycle.

For the second PCR reaction (gene amplification), 2 μL of the crude product from the first PCR was mixed with 40 pmol of each outer oligo, 100 mM KCl, 3 % DMSO and 1X PrimeSTAR Max Premix (Takara), and double-distilled water to a final volume of 25 μL. The thermal cycling conditions were: initial denaturation at 98 °C for 30 s; 30 cycles of denaturation at 98 °C for 5 s, annealing at 60 °C for 10 s, and extension at 72 °C for 30 s; followed by a final extension at 72 °C for 5 min. PCR products were purified using QIAquick PCR Purification Kit and then analysed using 1 % agarose gel electrophoresis.

3. Results

3.1. Performance of CertPrime in designing human genes

We generated oligonucleotide designs for all the sequences in the dataset DS1, using the parameters reported in Table 1, with a target melting temperature $T_m^* = 60$ °C. To explore the effect of oligonucleotide length on design efficiency and performance, we considered constraints on the maximal oligonucleotide length in the design ranging from 60 to 65 nucleotides. This approach resulted in six different designs for each sequence, allowing us to assess the versatility and robustness of CertPrime across various length constraints. Then, the designs were classified as "Excellent", "Good" or "Viable" as specified in the Methods section. The results, reported in Fig. 3, suggest that increasing the maximum length of the oligonucleotides enables us to design a higher percentage of the genome satisfactorily, as CertPrime gains more flexibility in the design process. Specifically, satisfactory designs were achieved for more than 97 % of the genome across all maximum lengths tested, this percentage exceeding 99 % when the maximum length was set to 65 base pairs. Lastly, when choosing the best design for each sequence irrespective of the maximal length, we were able to design 99.6 % of the genome at "Excellent" level. Notice that the best design for a certain maximal oligo length is not necessarily found when designing at longer lengths, due to the stochastic nature intrinsic to the design process. This highlights CertPrime's exceptional adaptability to the diversity of sequences found across the entire human genome.

To further evaluate CertPrime, we applied a "post-filtering" step to refine the designs by accounting for dimer formation imposing that all designs affected by homodimer or heterodimer pairings with a T_m greater than $(T_m^* - 5)$°C, are rejected, independently from how well they performed in terms of σ or ΔT_m.

Using this approach, we obtained the black dotted line in Fig. 4. The results indicate that the dimers rendered a fraction between 12 % and

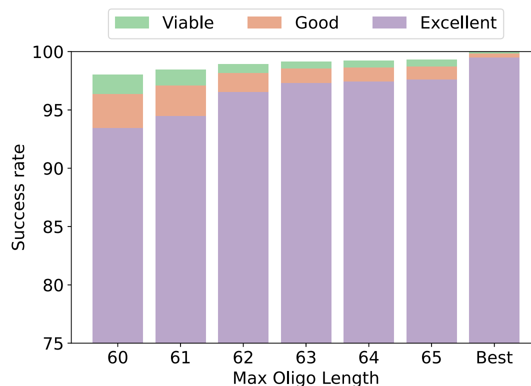


Fig. 3. The success rate and classification quality for designs with different constraints on the maximum oligo length. The choice of the lengths is related to the typical values managed in the experimental setups. The column "best" reports the best design for each sequence, irrespective of the maximum oligo length for which it was obtained.

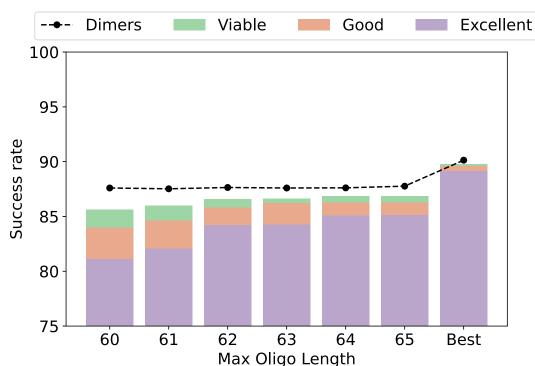


Fig. 4. Classification of the same designs as in Fig. 3 but discarding homodimers and heterodimers. The black dotted line "Dimers" marks the maximum possible success rate after discarding designs with dimers with a T_m of more than 55 °C. In the "Best" column, the best design has been selected for each sequence. The gap between the colored bars and the corresponding black dot represents sequences that are not dimer-prone, but for which no satisfactory design could be found.

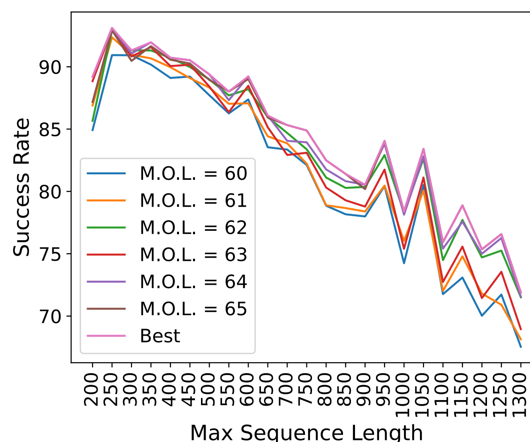


Fig. 5. Dependence of CertPrime success rate on sequence length. The fraction of successfully-designed sequences for each group of lengths is plotted as a function of the maximum sequence length in the group under consideration. The "Best" line considers the best design selected for each sequence, regardless of the maximum oligo length (M.O.L.) between 60 and 65 nts.

14 % of the designs invalid, depending on the maximum oligo length. Furthermore, as shown in the column "Best", we observed that in 9 % of cases, no valid design was identified. This suggests the existence of "dimer-prone" sequences, for which it is impossible to achieve dimer-free design even when allowing some flexibility in the oligonucleotide lengths.

Results also suggest that, especially for stricter limits on the maximum oligo-length, there were sequences that in principle were not affected by dimers, but for which CertPrime could not find any good design, according to the σ or ΔT_m criteria.

Next we determined whether the percentage of invalid designs, caused by either dimer formation or CertPrime constraints, varies with sequence length, with the aim to identify if there is an optimal sequence length that maximises the probability of a successful design.

For statistically reliable results, we considered the same database DS1 as before, and divided the sequences, according to their lengths, into 21 groups, spanning 50 lengths each (i.e., the first block contained sequences of 150–199 nts, and so on until the last group, with 1250–1299 nts).

Results showed that the fraction of successfully-designed sequences (with at least one valid design) in each group starts at values above 87 % and reaches 94 % (Fig. 5, "Best"). Additionally, as sequence length increases, the probability of success steadily declines. This trend is primarily due to the increase in the number of oligonucleotides required for synthesising longer sequences, which raises the probability of unwanted dimers formation. This suggests that dividing a long sequence into shorter segments and conducting two separate syntheses could be an effective strategy to enhance the quality of the design by minimising dimer formation. This approach can potentially improve experimental yield and overall efficiency, and we recommend it to practitioners, even beyond CertPrime's users: any algorithm will present an "optimal" sequence length, resulting from a trade-off between the need to keep the sequence short, to prevent dimer formation, and the fact that the shorter the sequence, the fewer the possible oligo arrangements, and hence, the good designs fulfilling the experimental requirements.

3.2. Computational cost dependence on DNA length

To study the length-dependence of the computational cost of CertPrime, we measured the time required to design each sequence contained in the human genome DS2 (Fig. 6). The results suggest a roughly quadratic relationship between the sequence length and the computational time, both when considering and measuring the time required for the process.

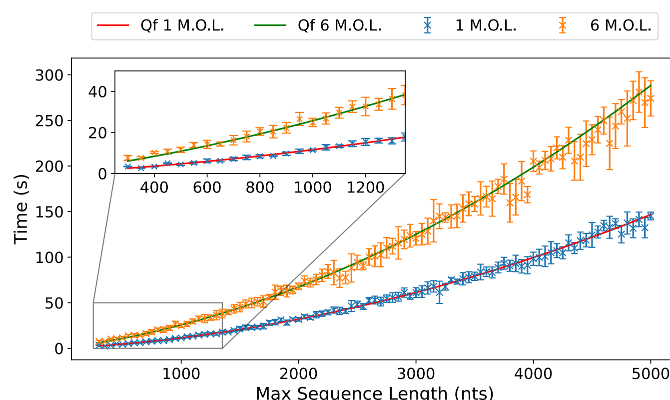


Fig. 6. Dependence of computation time on DNA sequence length. In the case of the "1 M.O.L." line (in blue) only one design *per* sequence was generated, with a maximum oligo length (M.O.L.) of 65 nts, while for the "6 M.O.L." line (in orange) six designs were made with maximum oligo lengths from 60 to 65 nts for each of the sequences. In addition, a quadratic fit (Q_f) was performed for each of the lines, yielding the red and green curves. The tests were performed on a computer with a 24-core Intel i913900K processor and 64 Gb RAM.

The results, reported in Fig. 6, suggest a roughly quadratic relationship between the sequence length and the computational time, both when considering a single design *per* sequence, with a specific maximum oligo length, or designs across all 6 lengths. For the single length case, we get:

$$\text{Time(s)} = 0.00712 \cdot L + 4.4404 \cdot 10^{-6} \cdot L^2 \quad (4)$$

while for the case of the 6 different maximum oligonucleotide lengths:

$$\text{Time(s)} = 0.01771 \cdot L + 7.9904 \cdot 10^{-6} \cdot L^2 \quad (5)$$

where L is the length of the sequence.

From these equations, it can be seen that the computational cost associated with considering several different cases for the maximal oligo length does not scale linearly with the number of cases. This non-linear behaviour arises because CertPrime is parallelised, to deal with each case independently. Consequently, all designs are performed simultaneously, with a significant reduction of the computational cost of the method. This parallelisation strategy enables CertPrime to handle long sequences in a reasonable time, approximately 140 s for lengths of 5000 nts.

Table 2

Characteristics of the sequences used for computational and experimental tests, and parameters of the corresponding oligonucleotide designs produced by CertPrime. Id. is the sequence identifier, L is the sequence length, %GC is the GC content, n is the number of oligonucleotides, l is the length of the shortest and longest oligonucleotides. %CG.Partial represents the GC content of the 50-nucleotide region in the sequence, with the highest and lowest GC content. The column Ov.L. reports the maximum and minimum lengths of the overlap regions, while %GC.Ov reports the maximum and minimum GC percentages of the overlap regions. The indicators ΔT_m, ρ, and σ were introduced in Eqs. (1)–(3) respectively. T_m Homo is the melting temperature of the oligonucleotide with the strongest self-interaction (thus forming the most stable homodimer), and T_m Hetero denotes the melting temperature of the pair of oligonucleotides that interact most strongly with each other (forming the most stable heterodimer).

Id.	L	%CG	n	l	%CG.partial	Ov.L.	%CG.Ov.	ΔT_m	ρ	σ	T_m Hetero	T_m Homo
Seq_1	130	53.08	4	(42, 53)	(48.98, 60.87)	(17, 23)	(43.48, 64.71)	0.16	0.09	0.07	30.20	24.47
Seq_2	130	40.77	4	(44, 57)	(34.62, 50)	(20, 27)	(33.33, 55)	0.11	0.06	0.04	26.64	28.13
Seq_3	130	45.38	4	(42, 54)	(45.45, 50)	(17, 25)	(44, 58.82)	0.20	0.11	0.11	28.77	25.8
Seq_4	130	40.00	4	(48, 56)	(35.29, 47.06)	(23, 29)	(37.93, 47.83)	0.18	0.09	0.09	34.19	37.86
Seq_5	142	48.59	4	(44, 59)	(38, 59.09)	(16, 27)	(37.04, 68.75)	0.15	0.08	0.06	26.62	18.57
Seq_6	152	36.84	4	(54, 64)	(29.69, 48)	(24, 33)	(27.27, 45.83)	0.17	0.10	0.06	24.89	29.74
Seq_7	153	43.79	4	(49, 60)	(38.6, 48.98)	(18, 26)	(38.46, 61.11)	0.13	0.07	0.05	30.86	33.39
Seq_8	160	65.62	10	(40, 58)	(34.48, 80)	(13, 30)	(33.33, 84.62)	0.34	0.10	0.09	48.22	47.10
Seq_9	172	47.09	6	(44, 53)	(45.45, 51.11)	(20, 24)	(41.67, 55)	0.42	0.18	0.14	23.13	25.72
Seq_10	180	42.78	6	(42, 56)	(35.71, 54.76)	(19, 31)	(29.03, 57.89)	2.42	1.00	0.70	41.60	30.28
Seq_11	180	44.44	6	(44, 57)	(42, 47.83)	(20, 25)	(40, 55)	1.10	0.47	0.26	30.96	25.89
Seq_12	180	40.56	6	(39, 55)	(31.03, 53.85)	(14, 29)	(31.03, 71.43)	1.40	0.58	0.32	29.34	36.47
Seq_13	190	43.16	6	(47, 55)	(36.36, 51.06)	(21, 27)	(33.33, 52.38)	0.20	0.08	0.08	34.55	41.13
Seq_14	192	59.90	6	(39, 59)	(55, 62.71)	(17, 21)	(55, 64.71)	0.12	0.05	0.07	45.83	30.14
Seq_15	213	61.97	6	(41, 59)	(50.98, 68)	(15, 26)	(42.31, 80)	0.15	0.06	0.06	47.30	49.66
Seq_16	240	44.58	8	(44, 59)	(38.98, 64.71)	(21, 28)	(35.71, 54.55)	0.83	0.28	0.20	41.65	24.92
Seq_17	240	56.25	6	(51, 65)	(50, 58.46)	(19, 26)	(42.31, 57.89)	0.26	0.10	0.08	35.89	39.35
Seq_18	258	50.78	6	(56, 65)	(43.08, 70.59)	(17, 27)	(40.74, 64.71)	0.39	0.14	0.12	48.29	40.65
Seq_19	260	51.15	8	(41, 61)	(49.06, 57.14)	(19, 25)	(44, 57.89)	0.12	0.04	0.03	41.81	31.23
Seq_20	274	54.01	8	(42, 61)	(49.09, 60)	(18, 24)	(45.45, 66.67)	0.19	0.06	0.04	39.47	40.05
Seq_21	281	40.21	8	(54, 61)	(31.03, 50.85)	(20, 30)	(33.33, 60)	0.39	0.16	0.12	34.10	33.57
Seq_22	300	46.00	8	(48, 63)	(43.33, 55)	(19, 25)	(40, 57.89)	0.24	0.09	0.07	34.62	32.11
Seq_23	333	48.95	10	(45, 60)	(39.66, 55.32)	(20, 26)	(38.46, 55)	0.18	0.06	0.06	41.85	37.87
Seq_24	499	43.69	14	(47, 65)	(30.16, 57.89)	(19, 34)	(26.47, 57.89)	0.29	0.08	0.05	32.22	29.38
Seq_25	545	52.66	16	(45, 62)	(38.18, 60.42)	(15, 27)	(37.04, 73.33)	0.32	0.07	0.05	51.97	49.97
Seq_26	577	49.57	16	(45, 64)	(43.1, 57.89)	(17, 26)	(36, 64.71)	0.16	0.05	0.04	48.51	50.25
Seq_27	698	44.56	20	(48, 62)	(29.03, 60.34)	(17, 34)	(26.47, 64.71)	0.77	0.22	0.15	40.98	34.32
Seq_28	734	54.50	20	(47, 63)	(37.1, 63.33)	(15, 29)	(31.03, 73.33)	0.31	0.09	0.07	38.05	36.25

3.3. Assembly efficiency

To assess CertPrime's ability to generate experimentally viable designs, we used it to create optimal oligonucleotide sets for 28 different DNA sequences of different lengths, arbitrarily chosen from the database available at the Certest company; we evaluated the quality of the designs using the indicators introduced in Eqs. (1)–(3). Table 2 shows that the ΔT_m, ρ, and σ values are all close to 0, which indicates not only that all temperatures are clustered, but also that they are clustered around the target value. The designed oligos were then synthesised and subsequently evaluated in PCR amplification experiments. The quality of the resulting assemblies was then assessed through agarose gel electrophoresis, providing a visual representation of the efficiency and accuracy of the assembly process (Fig. 7). We observed that for all sequences, a single intense band corresponding to the expected size of the synthesised fragments was obtained. This indicates that the synthesis was successful and that no dimers or other undesired by-products were formed.

3.4. Comparison of CertPrime with other software tools

The performance of CertPrime was compared to that of other available design programs to evaluate its potential experimental advantages. Specifically, we designed oligonucleotides for the S100A4, PKB2, and GFPuv genes, whose sequences, reported in [6], can be found in Supplementary Information. These genes are frequently referenced in the literature as benchmarks for evaluating oligonucleotide design tools in gene synthesis studies. In spite of several difficulties, we could compare CertPrime's designs with those of Primerize [11], PCR OligoMaker [10], TmPrime [6], DeepFirstSearch [13], Integrated [12], DNAWorks [7] and Gene2Oligo [8]; the results are reported in Table 3. Gene2Oligo and DNAWorks were only partially assessed as their web servers appear as unavailable, and their source code is not publicly accessible. As a result, we report only the σ values extracted from their respective publications [7,8]. Similarly, DeepFirstSearch and Integrated lack public code or web servers, so parameters could only be retrieved for PKB2 from oligo designs provided in their related publications [12,13]. The server of TmPrime is no longer accessible, and no source code is available: the corresponding results in Table 3 are obtained using the oligo designs provided in the TmPrime reference [6]. Primerize and PCR OligoMaker were accessible but presented notable limitations. Primerize's computational scalability restricts input sequences longer than 1000 bp on its web server, although users can bypass this restriction by locally installing the software. The web server of PCR OligoMaker, on the other hand, is not actively maintained, and it is only compatible with outdated browser versions; also, its source code is not publicly available. Despite these constraints, we successfully installed Primerize and accessed the PCR OligoMaker web server, and we could produce designs for the S100A4, PKB2, and GFPuv genes with a maximum oligo length of 65 nucleotides and a T_m set to 60 °C. The designs for the three genes by the different tools are reported in Supplementary Information.

In terms of computational design quality, CertPrime consistently exhibited the lowest values for ΔT_m, ρ and σ across all tested genes (Table 3), serving as a robust indicator of the method's accuracy and reliability. Regarding the number of oligonucleotides required *per* design, Integrated, DeepFirstSearch, and TmPrime require more oligos compared to the other methods, whereas PCR OligoMaker and Primerize demonstrate a more efficient use of oligos. Notably, Primerize, DeepFirstSearch, and Integrated showed significantly higher ΔT_m values compared to their respective σ values. This suggests a scenario where all melting temperatures are well-clustered, with the exception of a few outliers that distort the overall design quality. This discrepancy poses a practical challenge in synthesis, as such deviations often lead to experimental failures. These issues arise because these methods focus solely

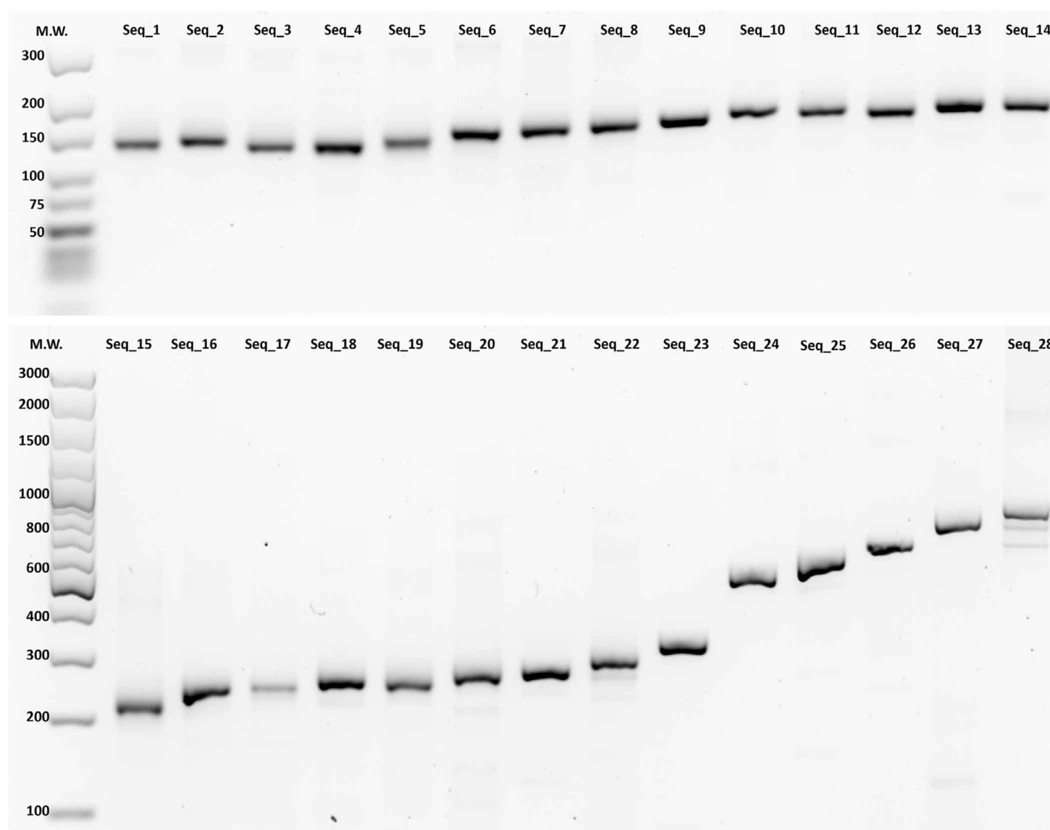


Fig. 7. Experimental tests of gene synthesis: Agarose gel electrophoresis of PCR amplicons for the 28 DNA sequences listed in Table 2. Top: Sequence 1–14; bottom: Sequence 15–28. A molecular-weight ladder (M.W.) is shown at the left of each panel for reference.

Table 3
Comparison between designs by different methods for PKB2, GFPuv and S100A4 sequences. Here n is the number of oligos in the design, l is the length of the longest oligo, and ΔT_m, ρ and σ are the indicators introduced in Eqs. (1)–(3) respectively. We could perform direct comparisons just with Primerize and PCR OligoMaker and we had to resort to the literature for the rest: see text for details. For all designs performed with CertPrime, Primerize and PCR OligoMaker, the maximum allowed oligo length was set to 65 nts and $T_m^* = 60$ °C. The TmPrime, DNAWorks and Gene2Oligo designs were taken from the TMPrime article [6] where they were found to have $T_m^* = 65$ °C and DeepFirstSearch and Integrated were taken from Ref. [13] where the PKB2 and S100A4 were designed at $T_m^* = 65$ °C while the GFPuv design had $T_m^* = 55$ °C.

Tool	PKB2					GFPuv					S100A4				
	n	l	ΔT_m	ρ	σ	n	l	ΔT_m	ρ	σ	n	l	ΔT_m	ρ	σ
CertPrime	40	65	0.58	0.1	0.1	20	65	0.73	0.13	0.13	22	65	0.09	0.09	0.37
Primerize	32	65	13.64	3.88	3.09	18	65	3.42	1.29	0.95	18	65	15.25	4.88	3.74
PCR oligoMaker	30	65	3.76	2.45	0.97	16	64	4.43	1.43	1.16	16	65	5.24	2.15	1.43
TmPrime	60	57	5.2	2.03	1.2	36	48	6.8	2.43	1.73	30	66	8.7	3.04	2.85
DeepFirstSearch	62	58	9.72	1.45	0.38	–	–	–	–	–	–	–	–	–	–
Integrated	62	57	4.61	0.54	0.49	–	–	–	–	–	–	–	–	–	–
DNAWorks	–	–	–	–	3.7	–	–	–	–	1.8	–	–	–	–	3.1
Gene2Oligo	–	–	–	–	7.6	–	–	–	–	5.84	–	–	–	–	*

on minimising σ without accounting for the behavior of ΔT_m, allowing such errors to remain undetected.

The S100A4 gene presents a region with a very low GC% content, making its design more challenging than that of the PKB2 and GFPuv genes (Figure 2 in S.I.). Therefore, we performed experimental tests on S100A4 gene designs generated by PCR OligoMaker, Primerize, and CertPrime. The designs produced by PCR OligoMaker and Primerize exhibited dimer formation, as evidenced by bands around the 100 base pairs region on the agarose gel electrophoresis (Fig. 8, left panel). These dimers were further evidenced through densiometric intensity profiles corresponding to the gel image analysis (Fig. 8, right panel). In contrast, the design produced by CertPrime resulted in a single, more intense band with no detectable dimers at lower mass. This outcome indicates a successful synthesis and assembly process, underscoring CertPrime's effectiveness in preventing spurious interactions and producing high-quality oligonucleotide assemblies. To confirm these results, the assembly process was repeated a second time for all three methods, as shown in Figure SI-3. There, OM exhibits a strong band above 100 nt, indicating the presence of dimers. Although PM does not show any dimer band, the intensity of the band corresponding to the amount of final product is much lower than that for CP, confirming that CP yields the best outcome.

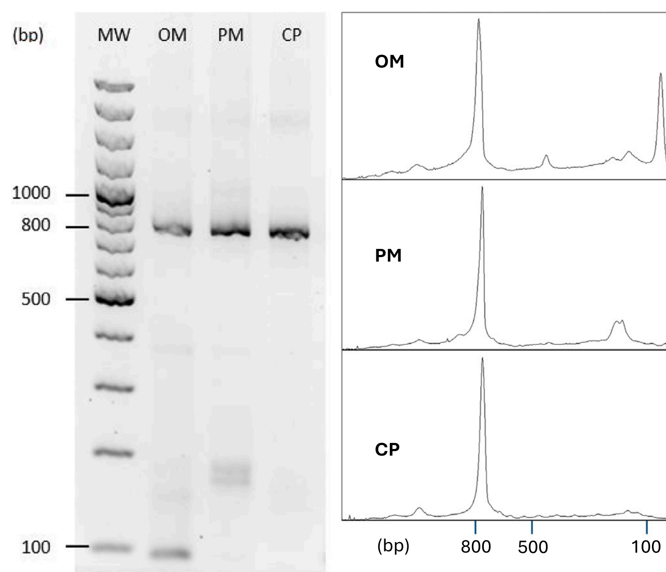


Fig. 8. Analysis of PCR amplification of the S100A4 gene using three different primer designs: PCR OligoMaker (OM) Primerize (PM) and CertPrime (CP). The left panel displays agarose gel electrophoresis results for each PCR experiment. The right panel shows densiometric intensity profiles derived from the corresponding gels, where the intensity of each band correlates with the height of the peak at the respective molecular weight. An optimal amplification is represented by a sharp, narrow peak aligned with the expected molecular weight of the S100A4 gene sequence.

3.5. CertPrime web-server

The CertPrime tool is accessible at the url: https://oligodesign.bifi.es upon registration. CertPrime is implemented in Python, using Streamlit as the web framework, and runs on a standard Ubuntu environment.

The web interface enables users to input the target sequence and specify experimental conditions; a simple manual is provided on the left of the web page. The design process is then executed automatically, in parallel, across all permitted values of maximum oligonucleotide length. Once the computation is completed, the resulting candidate designs are evaluated and summarised in a table, allowing users to review the relevant properties and select those of interest for download. We have opted for the simplest solutions both for the interface and for data management: for privacy and security reasons, uploaded sequences are not stored on our servers; they are automatically deleted once downloaded by the user, ensuring full compliance with legal requirements.

4. Conclusions

We have developed CertPrime as a fast, flexible, and highly accurate tool for oligonucleotide design in gene synthesis, with a particular strength in minimising T_m dispersion. By integrating key innovations at every step of the oligo design process, CertPrime achieves substantial improvements in computational efficiency while significantly increasing experimental success rates. One of CertPrime's prominent strengths lies in its flexibility, enabling precise specification of experimental conditions, that surpasses the capabilities of other available tools. It offers a range of customisable parameters to guide the design process, such as the maximum oligo length and overlap length. Additionally, CertPrime includes an advanced feature to minimise spurious oligonucleotide interactions, further improving the reliability of its designs.

The performance of the tool is underscored by its ability to synthesise over 99.5 % of the human coding sequences with a length ranging from 150 to 1350 nucleotides when dimer formation is not considered, and approximately 90 % when dimers are taken into account. Approximately 9 % of human sequences remain intrinsically difficult to design due to highly repetitive or GC-rich motifs; in such cases, codon optimisation may provide an alternative strategy when the focus is on protein expression rather than strict preservation of the native DNA sequence. Computationally, CertPrime demonstrates a mildly quadratic relationship between sequence length and computational cost, which does not represent a real burden even at the longest sequences considered. However, CertPrime is not intended for the design of very long sequences as design quality declines with increasing length, primarily due to dimer formation. This limitation has also been observed experimentally, since mixtures containing an extremely high number of oligos often lead to random interactions. For this reason, long genes are routinely divided into smaller fragments in standard laboratory protocols. The efficient synthesis of long constructs typically relies on advanced assembly strategies such as Gibson [27] or recursive DNA synthesis [28]. These approaches are based on the iterative assembly of overlapping oligonucleotides into progressively longer DNA fragments. A critical requirement for their success is the availability of well-designed overlapping oligos with robust hybridisation, uniform melting temperatures, and minimal secondary structures. CertPrime directly addresses this need by algorithmically partitioning target DNA into optimised overlapping oligos that meet these constraints. Thus, even if at present CertPrime does not tackle recursive synthesis strategies, the outputs of our tool can serve as fundamental building blocks for recursive synthesis workflows.

CertPrime's computational achievements were supported by robust experimental validation. Two assays were conducted: the first assessed the feasibility of CertPrime designs for 28 sequences, all of which yielded the expected results without spurious interactions. The second assay compared designs of the S100A4 gene produced by CertPrime and other tools. While all the tested tools are able to correctly synthesise the gene, CertPrime was the only one to provide an efficient, dimer-free design. Even if there is no guarantee that CertPrime finds the best global design, nor that an arbitrary sequence can be synthesised effectively, our results support the value of the tool in typical cases, and its superiority over other available tools. In conclusion, CertPrime represents a significant advancement in oligonucleotide design for gene synthesis, combining computational efficiency, experimental reliability, and unparalleled flexibility to address the challenges of modern genomic applications.

CRediT authorship contribution statement

David Luna-Cerralbo: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Ana Serrano:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Conceptualization. **Irene Blasco-Machín:** Writing – review & editing, Visualization, Investigation. **Fadi Hamdan:** Supervision, Conceptualization. **Juan Martínez-Oliván:** Supervision, Funding acquisition, Conceptualization. **Esther Broset:** Writing – review & editing, Supervision, Conceptualization. **Pierpaolo Bruscolini:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the revision of this work, the authors used DeepL in order to improve the language quality of some sentences. After using this tool, the authors reviewed and edited the content as needed, rejecting the suggestions that did not reflect the original meaning of the sentences. The authors take full responsibility for the content of the publication.

Funding

Research reported in this publication was supported by Gobierno de Aragón (Spain) through project IDMF/2021/0009 (Nuevas tecnologías para el diseño y obtención de vacunas de ARN en Aragón) and E30_20R ("Supercomputación Y Física De Sistemas

Complejos Y Biológicos"—COMPHYS). P.B. acknowledges support through grant PID2020-113582GB-I00 and PID2023-147734NB-I00, funded by MCIN/AEI/10.13039/ 501100011033.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Juan Martinez-Olivan, Esther Broset, Fadi Hamdan, Irene Blasco-Machin, David Luna-Cerralbo, Ana Serrano reports that financial support was provided by Certest Biotec SL. Juan Martinez-Olivan reports that financial support was provided by Government of Aragón. Pierpaolo Bruscolini reports that financial support was provided by the Government of Spain Ministry of Science and Innovation. Pierpaolo Bruscolini reports a relationship with Certest Biotec SL that includes: consulting or advisory. Juan Martinez-Olivan, Fadi Hamdan, Irene Blasco-Machin, David Luna-Cerralbo, Ana Serrano reports a relationship with Certest Biotec SL that includes: employment. Esther Broset reports a relationship with Certest Biotec SL that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors express their gratitude to all the members of Certest Biotec S.L.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.csbj.2025.10.004.

Data availability

The sequences for the proteins reported in Table 3 and their %CG profiles are included in the Supplementary Information. The oligo designs used in Fig. 8 are included as Supplementary Files.

References

[1] Braendstrup P, Levine BL, Ruella M. The long road to the first FDA-approved gene therapy: chimeric antigen receptor T cells targeting CD19. Cytotherapy 2020;22(2):57–69.
[2] Riggs AD. Making, cloning, and the expression of human insulin genes in bacteria: the path to Humulin. Endocr Rev 2021;42(3):374–80.
[3] Irving SJ. Human insulin from recombinant DNA technology. Science 1983;219(4585).
[4] Barbier AJ, Jiang AY, Zhang P, Wooster R, Anderson DG. The clinical progress of mRNA vaccines and immunotherapies. Nat Biotechnol 2022;40(6):840–54.
[5] Richard DA. 4 - thermostable DNA polymerases. In Innis MA, Gelfand DH, Sninsky JJ, editors. PCR strategies. Academic Press; 1995. pp. 39–57.
[6] Bode M, Khor S, Hongye Y, Mo-Huang L, Ying JY. TmPrime: fast, flexible oligonucleotide design software for gene synthesis. Nucleic Acids Res 2009;37 :W214–W221.
[7] Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res 2002;30(10).
[8] Rouillard J-M, Lee W, Truan G, Gao X, Zhou X, Gulari E. Gene2oligo: oligonucleotide design for in vitro gene synthesis. Nucleic Acids Res 2004;32 :W176–180.
[9] Richardson SM, Wheelan SJ, Yarrington RM, Boeke JD. GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome Research 2006;16(4): 550–6.
[10] Rydzanicz R, Zhao XS, Johnson PE. Assembly PCR oligo maker: a tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production.. Nucleic Acids Res 2005;33 :W521–W525.
[11] Tian S, Das R. Primerize-2d: automated primer design for RNA multidimensional chemical mapping. Bioinformatics 2017;33(9):1405–6.
[12] Fang G, Liang H. An integrated algorithm for designing oligodeoxynucleotides for gene synthesis. Front Genet 2022;13(836108).
[13] Liang H, Chen Z, Fang G. A depth-first search algorithm for oligonucleotide design in gene assembly. Front Genet 2022;13(1023092).
[14] Ragan TJ, Vincent HA. Pcr-based assembly of gene sequences by thermodynamically balanced inside-out (TBIO) gene synthesis. Methods Mol Biol 2023;2633:65–79.
[15] Gao X. 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 2003;31(22).
[16] Sandhu GS, Aleff RA, Kline BC. Dual asymmetric PCR: one-step construction of synthetic genes. BioTechniques 1992;12(1):14–6.
[17] Nelson MD, Fitch DHA. Overlap extension PCR: an efficient method for transgene construction. Methods Mol Biol 2011;772:459–70.
[18] Hilgarth RS, Lanigan TM. Optimization of overlap extension PCR for efficient transgene construction. MethodsX 2020;7(100759).
[19] Xiong A-S, Yao Q-H, Peng R-H, Xian L, Fan H-Q, Cheng Z-M, et al. A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. Nucleic Acids Res 2004;32(12).
[20] Young L, Dong Q. Two-step total gene synthesis method. Nucleic Acids Res 2004;32(7).
[21] Xiong A-S, Yao Q-H, Peng R-H, Duan H, Xian L, Fan H-Q, et al. PCR-based accurate synthesis of long DNA sequences. Nat Protoc 2006;1(2):791–7.
[22] Song L-F, Deng Z-H, Gong Z-Y, Lu-Lu L, Bing-Zhi L. Large-scale de novo oligonucleotide synthesis for whole-genome synthesis and data storage: challenges and opportunities. Front Bioeng Biotechnol 2021;9.
[23] von Ahsen N, Wittwer CT, Schütz E. Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg2+, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. Clin Chem 2001;47(11):1956–61.
[24] SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor Thermodynamics. In: Proceedings of the National Academy of Sciences of the United States of America; vol. 95. 1998. p. 1460–5(4.
[25] SantaLucia J, Hicks D. The Thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct 2004;33:415–40.
[26] Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, et al. Ensembl 2024. Nucleic Acids Res 2024;52 :D891–D899.
[27] Gibson DG, Young L, Ray-Yuan Chuang JCV, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods May 2009 ;6(5):343–5.
[28] Linshiz G, Yehezkel TB, Kaplan S, Gronau I, Ravid S, Adar R, et al. Recursive construction of perfect DNA molecules from imperfect oligonucleotides. Mol Syst Biol 2008;4(1).